Advances in Geographic Information Science

Jean-Claude Thill *Editor*

Spatial Analysis and Location Modeling in Urban and Regional Systems

Springer

Advances in Geographic Information Science

More information about this series at http://www.springer.com/series/7712

Jean-Claude Thill
Editor

Spatial Analysis and Location Modeling in Urban and Regional Systems

Springer

Editor
Jean-Claude Thill
Geography & Earth Sciences
University of North Carolina at Charlotte
Charlotte, NC, USA

ISSN 1867-2434 ISSN 1867-2442 (electronic)
Advances in Geographic Information Science
ISBN 978-3-642-37895-9 ISBN 978-3-642-37896-6 (eBook)
https://doi.org/10.1007/978-3-642-37896-6

Library of Congress Control Number: 2017960160

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer-Verlag GmbH Germany
The registered company address is: Heidelberger Platz 3, 14197 Berlin, Germany

Contents

Innovations in GIS&T, Spatial Analysis, and Location Modeling 1
Jean-Claude Thill

Part I ESDA and CSDA

A Bayesian Interpolation Method to Estimate Per Capita GDP at the Sub-Regional Level: Local Labour Markets in Spain 9
Domenica Panzera and Ana Viñuela

Discovering Functional Zones in a City Using Human Movements and Points of Interest 33
Nicholas Jing Yuan, Yu Zheng, and Xing Xie

Is Your City Economic, Cultural, or Political? Recognition of City Image Based on Multidimensional Scaling of Quantified Web Pages 63
Jae Soen Son and Jean-Claude Thill

Spatial Hedonic Modeling of Housing Prices Using Auxiliary Maps 97
Branislav Bajat, Milan Kilibarda, Milutin Pejović, and Mileva Samardžić Petrović

Modeling and Simulation of Cohesion Policy Funding and Regional Growth Diffusion in an Enlarged European Union 123
Sebastien Bourdin

Part II Geography as a Spatial System

Analyzing and Simulating Urban Density Exploring the Difference Between Policy Ambitions and Actual Trends in the Netherlands 145
Eric Koomen, Jasper Dekkers, and Dani Broitman

A Decision Support System for Farmland Preservation: Integration of Past and Present Land Use 173
Eduardo Corbelle-Rico, Inés Santé-Riveira, and Rafael Crecente-Maseda

Aggregate and Disaggregate Dynamic Spatial Interaction Approaches to Modeling Coin Diffusion 193
Marion Le Texier and Geoffrey Caruso

Comprehensive Evaluation of the Regional Environmental and Economic Impacts of Adopting Advanced Technologies for the Treatment of Sewage Sludge in Beijing 223
Guofeng Zhang, Xiaojing Ma, Jingjing Yan, Jinghua Sha, and Yoshiro Higano

Part III Spatial Systems Optimization

Locational Modeling in Spatial Analysis: Development and Maturity of Concepts 265
M.E. O'Kelly

Spatial Uncertainty Challenges in Location Modeling with Dispersion Requirements 283
Ran Wei and Alan T. Murray

The Nearest Neighbor Ant Colony System: A Spatially-Explicit Algorithm for the Traveling Salesman Problem 301
Jean-Claude Thill and Yu-Cheng Kuo

Large-Scale Energy Infrastructure Optimization: Breakthroughs and Challenges of CO_2 Capture and Storage (CCS) Modeling 323
Jordan K. Eccles and Richard S. Middleton

The Recharging Infrastructure Needs for Long Distance Travel by Electric Vehicles: A Comparison of Battery-Switching and Quick-Charging Stations 341
Linda Christensen, Sigal Kaplan, Thomas C. Jensen, Stefan Røpke, and Allan Olsen

Transport Strategies in Reverse Logistics for Establishing a Sound Material-Cycle Society 363
JongJin Yoon and Shigeru Morichi

Index 383

Innovations in GIS&T, Spatial Analysis, and Location Modeling

Jean-Claude Thill

Abstract Spatial analysis can be portrayed as the scientific analysis of data, wherein the spatial position –whether absolute or relative, or both – of data records is explicitly accounted for. It is fundamentally a modeling exercise that relates data entities together across the geographic space, as well as across multiple attribute spaces, and increasingly along the temporal dimension. In this chapter, we provide a brief epistemology of spatial analysis in relation to geography, regional science, and geographic information science, and situate it in the proper historical context of geography's quantitative revolution. We also restate its lineage to data science, statistics, and location analysis. The fundamental principles that cut across most of this field of research are introduced. In addition, we focus on the existing and potential contributions of this line of research to the knowledge and understanding of urban and regional systems. The chapters of this edited volume are situated within this context.

Keywords Spatial analysis • Location modeling • Spatial dependence • Spatial heterogeneity • Urban and regional systems

Spatial analysis can be portrayed as the scientific analysis of data, wherein the spatial position –whether absolute or relative, or both – of data records is explicitly accounted for. While its distinctiveness from other fields of research and scholarship such as geography, regional science, location analysis, geographic information science and others, can be quite difficult to ascertain and remains rather contested today, its intellectual roots are somewhat easier to trace. Surely, some pioneering work that espoused the line of reasoning and some of the rudimentary methods of spatial analysis can be in found in the 19th and early part of the twentieth century, such as John Snow's (1855) pioneering epidemiological work on the sources of London's 1854 cholera outbreak or Johann Heinrich von Thünen's (1826) study of agricultural land use and crop distribution around a market place. However, the

J.-C. Thill (✉)
Department of Geography and Earth Sciences, University of North Carolina, Charlotte, NC 28223, USA
e-mail: jean-claude.thill@uncc.edu

J.-C. Thill (ed.), *Spatial Analysis and Location Modeling in Urban and Regional Systems*, Advances in Geographic Information Science,
https://doi.org/10.1007/978-3-642-37896-6_1

earnest beginnings of spatial analysis through the paradigm of "spatial thinking" are usually placed in the late 1950s and early 1960s (Fotheringham and Rogerson 2009; Johnston 2009), which coincides with the advent of the so-called quantitative revolution in geography. Disciplinary research practices changed then by seeking to adopt positivist principles in the study of phenomena occurring on the surface of the earth and to customize the methods of data analysis in several existing disciplines such as statistics, operations research and computer science to the imperatives of spatial positioning. While the primary protagonists of this paradigm shift were geographers (Brian Berry, Duane Marble, Waldo Tobler, Art Getis, Michael Dacey and others), they positioned their research at the interface of multiple established disciplines which enabled intellectual cross-fertilization as well as the adoption of the spatial analytic perspective across disciplines as diverse as biology and ecology, genetics, planning, economics, transportation and health sciences. Two of the most iconic and influential contributions of spatial analysis at the time were Haggett (1965) and Abler et al. (1971), and these volumes remain classic references for scholars on spatial analysis and allied fields to this day.

The reason for being of spatial analysis rests on the premise that data has a spatial component. It has been said indeed that 80% of all data in the modern world have some spatial description (Franklin and Hane 1992). Measurement of the spatial aspects of each data entity is also an essential element of the analytical process. Measurement issues of data entities pertain to properties of distance, shape, perimeter, area, volume, direction, and others, of individual entities or collections thereof. While they are sometimes referred to as "spatial analysis", they are in fact outside of spatial analysis, sensu stricto, and nowadays are largely subsumed by basic operations of Geographic Information Systems (GIS). Spatial analysis encompasses various symbolic models of reality based on suitable measurements of spatial properties of data entities (Longley 2009). Thus, spatial analysis is fundamentally a modeling exercise that relates data entities together across the geographic space, as well as across multiple attribute spaces, and increasingly along the temporal dimension.

The attention garnered by spatial analysis inside the discipline of geography, its status as one of the pillars of Geographic Information Science (Longley et al. 2009), and its success at permeating into the analytical toolbox of a number of social, natural and life sciences reflect that the geographic space is a critical frame of reference that is not neutral in how social, physical, and natural processes operate on the surface of the earth. This translates into patterns, but importantly processes, that are marked by a few distinctive properties (Haining 2009) that constitute the central paradigm of spatial analysis and have driven the research agenda of spatial analysis for the past few decades. One of these properties is that of spatial dependence, which posits that the semantic information associated with a given location informs on the information carried by a nearby location. This statistical dependence relationship, also known as spatial autocorrelation when it pertains to a single variable, is often ascribed to some form of spatial spillover process. It was enshrined by Tobler (1970) as the First Law of Geography, stipulating that 'new things are more related than distant things' (Miller 2004). Overlooking spatial autocorrelation is known to

result in biased estimates in regression models (Haining 2003). More fundamentally however, doing so would deny the relevance of spatial relationships in interfacing with social and natural processes, and ultimately that geography matters.

Another core principle of spatial analysis is that of spatial heterogeneity according to which univariate and/or multivariate realities may not be stationary across the geographic space. As a result, phenomena may exhibit local hotspots and coldspots (Anselin 1995), while the parametrization of multivariate models may drift across the territorial expanse under study to reflect local variations around the "global" model calibrated for the study area at large (Fotheringham et al. 2000). In a nutshell, a strong case has been made in support of the position that "spatial data is special" (Anselin 1990). Yet, a rigid interpretation has often resulted in the postulate that the geographic space exists objectively and independently of the social and natural processes that operate across the spatial expanse and in the conceptual and operational separation of spatial and semantic information in spatial models. This view has been called into question, most recently by Thill (2011), as too dogmatic and disconnected from the reality of constructed spaces.

A particularly prominent part of spatial analysis called location analysis and modeling (Horner 2009) traces its theoretical foundations to Alfred Weber's location theory of industrial location. The overall scientific question here is that of the efficacy of the decisions leading to the selection of facility locations within a given region and of the quality of access that consumers have to the goods and services provided by these facilities. Perspectives may be focused on a single facility or a multi-facility organization (like a retail chain or a location school system) to serve recognized managerial or policy objectives (such as profit or market share maximization, or welfare or equity optimization) within set constraints. Location models may adopt a descriptive approach to actual urban and regional systems, but are also known to be very effective as normative frameworks in the analysis and planning of these systems.

Thus spatial analysis intersects with mainstream data analytics such as statistics, econometrics, operations research, network science, and various data science perspectives, such as machine learning and data mining. It draws on the principles of these fields and leverages them to the extent they enable a deeper apprehension of the role of spatial principles in structuring social and natural outcomes. The cohesion of the body of scholarly research that espouses these principles and transcends domains of application (social or economic versus natural or physical realm) has prompted some scholars (see e.g. Johnston 2009) to coin the term "Spatial Science" that appears to better signify the singularities of spatial analysis.

In addition to geography, spatial analysis has some well-defined lineage with regional science. Emerging in the early 1950s, regional science had established itself as the field of study of the region, as the spatial expression of many social, economic, and political structures that frame the operational relationships that maintain the functional cohesiveness of regions (Thill 2017). Since its inception, regional science has espoused a neo-positivism approach to knowledge discovery. Just as regional science was prompt to adopt the emerging paradigm of geospatial data integration embedded in geographic information science and technologies (GIS&T) (Fischer

and Nijkamp 1992; Thill 2017; Yeh and Batty 1990) to study the complex structures of urban and regional systems and the space-time processes hat shape them, it expanded the methodological frontier of spatial analysis, through both modalities of exploratory spatial data analysis (ESDA) (Anselin 1998) and confirmatory spatial data analysis (CSDA) (Anselin and Getis 1992).

Urban and Regional systems are spatial systems that operate among a number of objects, entities and agents in ways that may be thematically very diverse; the net outcome of the interactions and functional relationships among them is some form of order expressed through "spatial organization". This order extends over space (such as the Christallerian system of cities, towns and villages), and may change over time (for instance the emergence of polycentric urban systems, or socio-economic convergence of regions comprising a country's economic space). It may also exhibit properties that persist over certain ranges of scales and granularities, and morph into others on other ranges. Thus, the study of urban and regional systems and their dynamics is at the core of regional science and their complexity can advantageously be unraveled through the use of methods of spatial analysis.

This book aims at contributing to both Urban and Regional Sciences and Spatial Analysis through a collection of fifteen distinctive research contributions by noted scholars of urban and regional systems. All the chapters of this book are original pieces of research presenting an innovative development of spatial analysis in the context of urban and regional systems or a novel application of existing spatial analytic methods to this domain; some were presented at one of several scientific meetings of the Regional Science Association International in the United States, Canada, and Europe. Others were solicited by the editor to complement the themes selected as areas of emphasis of this volume. Each chapter was subjected to a rigorous peer-review process and was also reviewed by the editors of the volume. They were selected for their complementary focus on spatial analysis; as such, they serve as good examples of the state-of-the-art in spatial analysis applied to urban and regional systems.

The book is organized in three parts that apply spatial analytic methods in increasingly integrated and systemic fashions. The first part contains five chapters that encompass contributions to ESDA and CSDA at various spatial resolutions, spanning from neighborhoods in a single city to regions of the European Union, and up to world city regions across the globe. The analytic methods that are utilized in these studies include Bayesian statistics, multidimensional scaling, hedonic regression modeling, spatial data mining, and spatial simulations for well recognized spatial analytic purposes as diverse as multivariate prediction and forecast under alternative scenarios to spatial interpolation, and spatial classification and generalization.

The second part contains four chapters that methodologically incorporate some of the standard elements of ESDA and CSDA while explicitly recognizing the territorial expanse of a study area as an integrated spatial system. In this perspective, spatial entities are not limited to pairwise and direct relationships; multiple interactions and interdependencies may occur simultaneously, directly as well as indirectly. This raises the possibility for simulations of proposed policies and

possible futures based on rules and relationships calibrated on historical data layers, while allowing for cross-scalar effects, including the to-down and bottom-up propagation of constraints and preference structures.

The third and final part contains six chapters that also follow a spatial systems approach, but along normative principles rather than with the intent of describing real-world phenomena and extrapolating from them on future states of the system. Spatial optimization serves as a framework for the efficient allocation of infrastructure and operational resources to system components scattered across a geographic space. Areas of application discussed in these pieces of research pertain to transportation, logistics, and energy facility planning. These models and applications squarely fit in the location analysis tradition that has flourished within spatial analysis.

The chapters that form this volume reflect some of the latest developments in spatial analysis and location modeling for urban and regional systems. Research conducted over the past 30 years has brought this field to maturity. While the early years of spatial analysis were marked by research in modeling frameworks heavily constrained by data granularity and computational limits, it has now become possible to redirect priorities away from these constraints and refocus our scholarship on the concepts and behaviors at the core of urban and regional operations as well as on the data describing the state and dynamics of these systems. The next frontier lies where spatial analysis can fully leverage computational power of distributed computing systems applied to the massive volumes of heterogeneous and disaggregated data that populate the space-time data cube. The cross-fertilization between spatial analysis, GIS&T and the foundational principles of data science has been a watershed point in the recent account of spatial analysis. We contend that it has reasserted the cutting-edge scholarship contributed by spatial analysis to a host of social and natural sciences for the decades to come.

References

Abler, R., Adams, J., & Gould, P. (1971). *Spatial organization – The geographers view of the world*. Englewood Cliffs: Prentice-Hall.

Anselin, L. (1990). What is special about spatial data? Alternative perspectives on spatial data analysis. In D. Griffith (Ed.), *Spatial statistics past, present and future* (pp. 63–77). Ann Arbor: Institute of Mathematical Geography.

Anselin, L. (1995). Local idicators of spatial association – LISA. *Geographical Analysis, 27*, 93–115.

Anselin, L. (1998). Exploratory spatial data analysis in a geocomputational environment. In P. A. Longley, S. M. Brooks, R. McDonnell, & W. Macmillian (Eds.), *Geocomputation: A primer* (pp. 77–94). New York: Wiley and Sons.

Anselin, L., & Getis, A. (1992). Spatial statistical analysis and geographic information systems. *Annals of Regional Science, 26*(1), 19–33.

Fischer, M. M., & Nijkamp, P. (1992). Geographic information systems and spatial analysis. *Annals of Regional Science, 26*, 3–17.

Fotheringham, A. S., & Rogerson, P. (2009). Introduction. In A. S. Fotheringham & P. Rogerson (Eds.), *The SAGE handbook of spatial analysis* (pp. 1–4). Los Angeles: Sage.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2000). *Quantitative geography: Perspectives on spatial data analysis*. London: Sage.

Franklin, C., & Hane, P. (1992). An introduction to GIS: Linking maps to databases. *Database, 15*(2), 17–22.

Haggett, P. (1965). *Location analysis in human geography*. London: Edward Arnold.

Haining, R. (2003). *Spatial data analysis: Theory and practice*. Cambridge: Cambridge University Press.

Haining, R. (2009). The special nature of spatial data. In A. S. Fotheringham & P. Rogerson (Eds.), *The SAGE handbook of spatial analysis* (pp. 5–23). Los Angeles: Sage.

Horner, M. W. (2009). Location analysis. In R. Kitchin & N. Thrift (Eds.), *International encyclopedia of human geography* (pp. 263–269). Oxford: Elsevier.

Johnston, R. (2009). Spatial science. In R. Kitchin & N. Thrift (Eds.), *International encyclopedia of human geography* (pp. 384–395). Oxford: Elsevier.

Longley, P. A. (2009). Spatial analysis and modeling. In M. Madden (Ed.), *Manual of geographic information systems* (pp. 587–590). Bethesda: American Society for Photogrammetry and Remote Sensing.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2009). In M. Madden (Ed.), *Geographic information science* (pp. 19–24). Bethesda: American Society for Photogrammetry and Remote Sensing.

Miller, H. (2004). Tobler's first law and spatial analysis. *Annals of the Association of American Geographers, 94*, 284–289.

Snow, J. (1855). *On the mode of communication of cholera* (2nd ed.). London: John Churchill.

Thill, J.-C. (2011). Is spatial really that special? A tale of spaces. In V. Popovich, C. Claramunt, T. Devogele, M. Schrenk, & K. Korolenko (Eds.), *Information fusion and geographic information systems: Towards the digital ocean* (pp. 3–12). Heidelberg: Springer.

Thill, J.-C. (2017). Regional science. In N. Castree, M. Goodchild, W. Liu, A. Kobayashi, R. Marston, & D. Richardson (Eds.), *International encyclopedia of geography* (pp. in press). Wiley-Blackwell.

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography, 46*, 234–240.

von Thünen, J. H. (1826). Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie, Hamburg, Perthes. English translation by C.M. Wartenberg (1966). *The Isolated State*. Oxford: Pergamon Press.

Yeh, A. G. O., & Batty, M. (1990). Applications of geographic information systems in urban and regional planning. *Environment and Planning. B, Planning & Design, 17*(4), 369–374.

Part I
ESDA and CSDA

A Bayesian Interpolation Method to Estimate Per Capita GDP at the Sub-Regional Level: Local Labour Markets in Spain

Domenica Panzera and Ana Viñuela

Abstract Although economic data is increasingly available for almost any topic, there is still a dearth of data at very high levels of spatial disaggregation. GDP data (total or per capita) is available for the European NUTS II regions and the Spanish National Statistics Office (INE) regularly publishes such information at the NUTS III level (which, in the Spanish case, corresponds to provinces). In this research, we move toward a higher level of spatial disaggregation. We estimate the per capita GDP for the 804 Local Labor Market Areas (LLMs) into which the Spanish territory can be divided by applying the Bayesian Interpolation Method (BIM) introduced by Palma and Benedetti (Geograph Sys 5:199–220, 1998), and considering spatial dependence between observations. Before proceeding with the estimation, we test the methodology by estimating per capita GDP for provinces (NUTS III level), as if such information is only known at a more aggregated level (NUTS II regions) and compare the results to actual provincial values. We then derive per capita GDP values for LLMs given the observed data at the provincial level. The results obtained reveal a high level of internal heterogeneity in GDP per capita within the Spanish administrative regions, and highlight the importance of agglomeration economies and relative location.

Keywords Spatially disaggregated data • Bayesian interpolation method • Local labor market areas

D. Panzera (✉)
Department of Economic Studies, University "G. d'Annunzio" of Chieti-Pescara, Chieti, Italy
e-mail: domenicapanzera@yahoo.it

A. Viñuela
Applied Economics Department, University of Oviedo, Oviedo, Spain
e-mail: avinuela@uniovi.es

J.-C. Thill (ed.), *Spatial Analysis and Location Modeling in Urban and Regional Systems*, Advances in Geographic Information Science,
https://doi.org/10.1007/978-3-642-37896-6_2

Introduction

A significant volume of research has explored the underlying causes of income disparities across sub-national regions and the concomitant dispersion in economic outcomes. Nowhere is this more apparent than in Spain. When investigating these issues, regional economists are constrained by the use of data which, at their most disaggregate, are available at NUTSIII level,[1] which corresponds to provinces in the Spanish case. However, given the internal heterogeneity within the regions at various scales, the challenge becomes one of trying to estimate income values for regional systems based on economic rather than politico-administrative criteria.

Differences in economic outcomes within administrative regions reflect differences in productivity, employment levels and wages (Moretti 2011). While in Spain there are employment data available at the local (municipality) level – which can be used by industry level or in total to demonstrate the internal heterogeneity of Spanish regions – still, no data exists on GDP (or productivity) at local levels.

Yet the Local Labor Markets (LLMs) present themselves as a more suitable *economic* unit of analysis. This is based upon two ideas: (1) that workers and firms interact primarily in Local Labor Markets, the size of which are much smaller than that of the national market; and (2) that few people move from one market to another (Armstrong and Taylor 1993; Bartik 1996; Hughes and McCormick 1994; Topel 1986). In most countries, perhaps only a relatively small number of large LLMs (cities) account for most of the country's output. However, economic researchers still face the problem of a lack of GDP figures, either in total or per capita, at this level of disaggregation.

The purpose of this paper is to estimate per capita GDP for the Local Labor Markets (hereafter LLMs) into which the Spanish territory is divided.

In section "Disaggregating data spatially: The *local* unit of analysis and methodologies", we briefly explain the choice of the LLM as the appropriate sub-regional unit for estimating per capita GDP and the problem of disaggregating data from the regional level in the general framework of areal interpolation. Section "The Bayesian Interpolation Method" presents the proposed methodology – the Bayesian Interpolation Method (hereafter, BIM) – to disaggregate per capita GDP. The solution to this problem requires formulating a hypothesis regarding the probability distribution of the *original* process and exploits information about the *spatial dependence* between observations, as well as auxiliary information observed at the sub-regional level. Prior to estimating per capita GDP for the Spanish LLMs, in section "Estimating Per Capita GDP for Spanish Provinces" the performance of the BIM is tested when disaggregating per capita GDP for *Autonomous Communities* (NUTS II divisions) into per capita GDP for *Provinces* (NUTS III divisions), while treating the latter as unknown. In section "Estimating Per Capita GDP for Spanish

[1]The Nomenclature of Territorial Units for Statistics (NUTS) was established by Eurostat in order to provide a single uniform delineation of territorial units for the production of regional statistics which is based upon the administrative divisions applied in the Member States.

LLMs", the results of the per capita GDP estimated at LLM level using the BIM are presented and discussed. Finally, section "Conclusions and future research" suggests future extensions of this work and a research agenda for this topic.

Disaggregating Data Spatially: The *Local* Unit of Analysis and Methodologies

Local Unit of Analysis: Local Labor Markets

What is the proper unit of analysis for estimation of per capita GDP at the local level? This question is closely related to the definition of an *economic region* in the regional economics literature. The hallmark of an economic region is the existence of an especially high degree of interdependence among individual incomes within the area (Parr 2008).

A number of researchers in the United States during the 1960s (Fox and Kumar 1965), and subsequently in Europe during the 1970s (Smart 1974), designed quantitative techniques or regionalization schemes for identifying such regions. Different names have been applied to these areas, such as Functional Economic Areas and Labor Market Areas (and, as used here, Local Labor Markets or LLMs), but they all referred to a region with an increasingly interwoven and internally interactive area that encompasses the home-to-work daily journeys of its residents.

Originally designed to set the limits of a *city*, these *regions* reflect functional relationships between workers and jobs.[2] With an urban center in the core and fringe municipalities surrounding this core, the LLMs have formed increasingly interwoven and internally interactive areas, in their own right, over time. At the present time in Europe, jobs are increasingly created in the fringe areas, but people prefer to live in the city centers, resulting in two-way journey-to-work flows. Based upon commuting patterns, LLMs demarcate the borders of labor catchment areas. In practice, this means that at least 75% of residents work in the LLM and that 75% of those who work in the LLM also live there. To qualify as an LLM, an area must have a minimum of 3500 residents that work in the area.

The regionalization procedure to establish the borders of an LLM is based upon an algorithm originally developed by Coombes et al. (1986). Starting with the municipal administrative unit and combining the data on the resident employed population, total employed population and journey-to-work commuting patterns, boundaries of the LLMs are defined through a multi-stage aggregation process.[3] This methodology, slightly modified to meet the specific characteristics of countries,

[2]For a discussion on the increasing tendency to treat the region as the city and a city as a region in urban and regional analysis, see Parr (2008).

[3]For a description of the previous method, see Smart (1974). For a discussion of problems that arose with that method, see Ball (1980) and Coombes and Openshaw (1982).

has been applied in several European countries. The EU Department of Employment, Social Affairs & Inclusion has defined the LLMs (so-called Travel-To-Work-Areas or TTWAs) for Great Britain, for Italy (Sforzi et al. 1997), for Denmark (Andersen 2002), and for Spain (Boix and Galleto 2006).

Based on the idea that workers and firms interact mainly within their respective LLMs, these economic regions may be appropriate spatial units of analysis for studying topics such as: the underlying causes of spatial income disparities across a country; the existence of agglomeration economies and diseconomies and their effect on location decisions, economic structure and growth between larger and smaller cities (Behrens and Thisse 2007); or the relationship between urban agglomeration and productivity (Melo et al. 2009). However, these types of studies or analyses cannot be carried out for most European countries, due to the lack of per capita GDP figures at a further level of disaggregation than NUTS III regions.

Approaches to the Lack of Spatially Disaggregated Data

The European System of National and Regional Accounts (ESA 95) provides its users with a consistent and reliable quantitative description of the economic structure of the European Member States and their regions. The European regional accounts are closely connected to the National Accounts and are based upon the transactions that occur within a specific NUTS-level region.[4] Some attempts have been made by the European Union to provide economic indicators (including GDP figures) for some European cities, as well. The Urban Audit Project provides comparable statistics collected every three years for 321 cities in the 27 countries of the EU, along with 36 additional cities in Norway, Switzerland and Turkey.[5]

In Spain, the National Statistics Institute (INE), which is responsible for the elaboration of the national and regional accounts, provides GDP figures for the NUTS II regions (*Autonomous Communities)* and also for the NUTS III regions (*Provinces*). However, GDP figures at the next level of spatial disaggregation, such as LLMs, are not available.

Situations in which observable regional data are decomposed into their respective smaller spatial units can be viewed as a special case of areal interpolation. Areal interpolation refers to the process of estimating one or more variables for a set of *target* zones, based upon the known values in the set of *source* zones. Target units can be either finer-scale, or misaligned, with respect to the source units.

[4]For a discussion on the regionalization method implemented by Eurostat for deriving the regional accounts or conceptual difficulties that constrain the economic variables available at NUTS level – such as the existence of many productive activities crossing regional boundaries, units of production operating in different regions, inter-firm transfers or people living in one region, but working in another, see Eurostat, System of Regional Accounts, ESA1995, (http://circa.europa.eu/irc/dsis/nfaccount/info/data/ESA95/en/titelen.htm)

[5]For more details see http://www.urbanaudit.org/

A number of solutions for deriving data not available at the desired scale of spatial resolution have been proposed in the regional economics literature.[6] Applied to disaggregate Spanish per capita GDP for Autonomous Communities, and treating actual provincial per capita GDP (at the NUTS III level) as not observed, Peeters and Chasco (2006) compared two different approaches to disaggregate information from aggregated data: the classical OLS (ordinary least squares) model and a new Generalized Cross-Entropy (GCE) model. The latter involves the assumption of spatial heterogeneity across observations, and yields predictions of the unobserved that "are *superior* in terms of accuracy".[7] Another recent contribution dealing with the prediction of disaggregated regional data has been provided by Polasek et al. (2010). In this work, the authors extended the Chow-Lin method (Chow and Lin 1971), originally proposed to predict high-frequency time series data from related series, by its application to regional cross-sectional data. For this purpose, the authors specified a spatial autoregressive model (SAR) for the disaggregated data, thereby introducing the spatial dependence effect in the analysis. The spatial dependence parameter is estimated by running the SAR model on the aggregated data; GLS estimators for the parameters associated with regressors observed at the disaggregated level are derived and the prediction of the unobserved dependent variable is carried out.

In the present study, the Spatial Chow-Lin procedure is developed within a Bayesian framework, and it is applied to predict the GDP per capita for Spanish Provinces (assumed to be unobserved), given available indicators at the provincial level and the aggregated data observed for Autonomus Communities. Taking the actual values of provincial GDP as a benchmark, in this empirical application the authors find a significant improvement in the estimates when moving from the classical OLS to the spatial autoregressive specification.

In the next section, we propose the use of the BIM to spatially disaggregate per capita GDP using Spanish data. Bearing in mind that the aim is to estimate per capita GDP for the LLMs defined in section "Local unit of analysis: Local labor markets", in section "Estimating Per Capita GDP for Spanish Provinces" we will test the proposed methodology by disaggregating Autonomous Communities' per capita GDP into Provincial per capita GDP, treating the latter as an unobserved variable. Using the same methodology, in section "Estimating Per Capita GDP for Spanish LLMs" we will estimate per capita GDP for a higher level of disaggregation (LLMs) and verify the existence of intraregional disparities in economic welfare.

[6]For a review on this topic see e.g. Flowerdew and Green (1992), and for a general framework on the areal interpolation of socio-economic data see Goodchild et al. (1992).

[7]In their GCE approach, Peeters and Chasco (2006) develop the prediction model considering the aggregated data to be a weighted geometric mean of the disaggregated data and incorporate the spatial heterogeneity effect through the introduction of unit-specific coefficients for provinces. Unlike the GCE model, the BIM that will be proposed in this work focuses on the spatial dependence effect, which is incorporated into the covariance structure of the data generating process, allowing for the identifcation of a model characterized by fewer parameters according to the parsimony principle.

The Bayesian Interpolation Method

The problem of deriving data that are not available at the desired scale of interest is here formalized by assuming the areal data as a realization of a spatial stochastic process or random field.

Consider a set of n areal units $\{\omega_i, i = 1, 2, \ldots, n\}$ which form a partition Ω over a geographical domain, and denote by Z a variable of interest. The set of random variables $\{Z(\omega_i), \ i = 1, \ldots, n\}$, indexed by their locations, defines a random field. We refer to the set $\{Z(\omega_i), \ i = 1, \ldots, n\}$ as the *original process*.

Consider a partition Ω^* of the same geographical domain in $m < n$ areal units $\left\{\omega_j^*, \ j = 1, \ldots, m\right\}$ obtained by grouping the areal units ω_i, $i = 1, \ldots, n$, into larger areas. Changing the support of Z creates a new variable Z^*, and data collected on Z^* can be viewed as a realization of the *derived* or *aggregated process* $\left\{Z^*\left(\omega_j^*\right), \ j = 1, \ldots, m\right\}$.

Now assume that data on the variable of interest are available with reference to the partition Ω^*, while we are interested in the level of disaggregation corresponding to the partition Ω. In the described framework, the areal data conversion problem consists of reconstructing the realizations of the original process by starting with the knowledge, based upon data, of the derived process.

As a possible solution to the areal data conversion problem we propose the Bayesian Interpolation Method (BIM) introduced by Palma and Benedetti (1998). According to the BIM the joint distribution of the random vector $\boldsymbol{Z} = (Z(\omega_1), \ldots, Z(\omega_n))'$ is assumed to be a multivariate normal distribution, and $\boldsymbol{Z}$ is expressed in an additive form, that is:

$$\boldsymbol{Z} = \boldsymbol{S} + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{S} = (S(\omega_1), S(\omega_2), \ldots, S(\omega_n))'$ refers to the variable of interest at the n sites ω_i, and $\boldsymbol{\varepsilon}$ is a random vector of error terms.

As an additional assumption, the random vectors $\boldsymbol{S}$ and $\boldsymbol{\varepsilon}$ are modeled through the Conditional Autoregressive (CAR) specification (Besag 1974). CAR models satisfy a Markov property in the space where the value of a random variable in a region, given the observations in all the other regions, depends only on the observations in neighboring regions. Assuming CAR models for the random vectors $\boldsymbol{S}$ and $\boldsymbol{\varepsilon}$ does not entail any loss of generality, since any Gaussian process on a finite set of sites can be modeled according to this specification (Ripley 1981, p.90).

The CAR model for $\boldsymbol{S}$ is specified by the set of full conditional distributions:

$$S(\omega_i) \mid S(\omega_{-i}) \sim N\left(T_i + \rho \sum_{j=1}^{n} w_{ij}\left(S\left(\omega_j\right) - T_j\right), \sigma_i^2\right) \qquad i, j = 1, \ldots, n \tag{2}$$

where $\omega_{-i} = \{\omega_j : j \neq i, j = 1, \ldots, n\}$, T_i is the expected value of $S(\omega_i)$, σ_i^2 is its conditional variance, ρ is a scalar parameter measuring the spatial autocorrelation, and w_{ij} is the (ij)-th entry of a $n \times n$ spatial weights matrix $\boldsymbol{W}$, with $w_{ii} = 0$ for all i.

Under some specified regularity conditions (Besag 1974), the full conditional distributions in (2) uniquely determine the following joint distribution:

$$\boldsymbol{S} \sim MVN\left(\boldsymbol{T}, \boldsymbol{\Sigma}_{\boldsymbol{S}}\right) \tag{3}$$

where $\boldsymbol{T} = (T_1, \ \ldots, T_n)^{'}$ and $\boldsymbol{\Sigma}_{\boldsymbol{S}}$ is the covariance matrix expressed as:

$$\boldsymbol{\Sigma}_{\boldsymbol{S}} = (\boldsymbol{I} - \rho \boldsymbol{W})^{-1} \boldsymbol{M} \tag{4}$$

with $\boldsymbol{I}$ denoting a n-dimensional identity matrix and $\boldsymbol{M} = diag\left(\sigma_1^2, \ldots, \sigma_n^2\right)$.

The matrix $\boldsymbol{W}$ in (4) is commonly specified by normalizing a proximity matrix $\boldsymbol{C}$, whose elements c_{ij} are defined as:

$$c_{ij} = \begin{cases} 1 \ if \ \omega_j \in N(i), \\ 0 \ if \ i = j, \\ 0 \ \ otherwise \end{cases} \qquad i, j = 1, \ \ldots, n.$$

with $N(i)$ denoting the set of neighbors of the areal unit ω_i identified according to any proximity criterion. Let $c_{i+} = \sum_{j=1}^{n} c_{ij}$ denote the cardinality of $N(i)$, for $i = 1, \ \ldots, n$. Constructing the n-dimensional diagonal matrix $\boldsymbol{D} = \text{diag}\ (c_{1+}, \ \ldots, c_{n+})$, the matrix $\boldsymbol{W}$ can be expressed as follows:

$$\boldsymbol{W} = \boldsymbol{D}^{-1} \boldsymbol{C}. \tag{5}$$

The number of neighbors can vary for the areal units. However, since the CAR specification requires the symmetry condition expressed by $\sigma_j^2 w_{ij} = \sigma_i^2 w_{ji}$, for all $i, j = 1, \ \ldots, n$, when the weighting scheme in (5) is used, the conditional variances have to be inversely proportional to c_{i+}, for $i = 1, \ \ldots, n$ (Clayton and Berardinelli 1992; Wall 2004). Then, by setting $\sigma_i^2 = \sigma_S^2 / c_{i+}$, for $i = 1, \ \ldots, n$, where σ_S^2 is a scalar parameter measuring the overall variability of $\boldsymbol{S}$, the covariance matrix in (4) becomes:

$$\boldsymbol{\Sigma}_{\boldsymbol{S}} = \sigma_S^2 (\boldsymbol{D} - \rho \boldsymbol{C})^{-1} \tag{6}$$

where the matrices $\boldsymbol{D}$ and $\boldsymbol{C}$ are defined as above. The specification of the covariance matrix in (6) characterizes the specification that is referred to as CAR(ρ, σ^2) by Gelfand and Vounatsou (2003), and a sufficient condition for its positive definiteness is $|\rho| < 1$.

Similar considerations hold for the random vector $\boldsymbol{\varepsilon}$ which is modeled through a zero-centered CAR, so that its joint distribution is given by:

$$\boldsymbol{\varepsilon} \sim MVN\left(0, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\right) \tag{7}$$

where 0 denotes a $n \times 1$ vector of zeros and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \sigma_{\varepsilon}^2 (\boldsymbol{D} - \rho_{\varepsilon} \boldsymbol{C})^{-1}$, with σ_{ε}^2 and ρ_{ε} denoting scalar parameters.

From the formulated assumptions it follows that the joint distribution of the original process can be expressed as:

$$\boldsymbol{Z} \sim MVN\left(\boldsymbol{S}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\right) \tag{8}$$

where $\boldsymbol{S}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ are specified as above.

The transformation of the n-dimensional random vector $\boldsymbol{Z} = (Z(\omega_i), \ \ldots, Z(\omega_n))'$, related to the original or disaggregated process, into the m-dimensional random vector $\boldsymbol{Z}^* = \left(Z^*\left(\omega_1^*\right), \ldots, Z^*\left(\omega_m^*\right)\right)'$, related to the derived or aggregated process, can be formalized by introducing a linear operator $\boldsymbol{G}$. The transformation operator $\boldsymbol{G}$ is constructed as a $m \times n$ matrix whose elements can be specified according to any sum or averaging operations, so that:

$$\boldsymbol{Z}^* = \boldsymbol{G}\boldsymbol{Z} = \boldsymbol{G}\left(\boldsymbol{S} + \boldsymbol{\varepsilon}\right) = \boldsymbol{G}\boldsymbol{S} + \boldsymbol{G}\boldsymbol{\varepsilon}.$$

Since the observed aggregated data derive from the unobserved disaggregated data through the operator $\boldsymbol{G}$, Palma and Benedetti (1998) give a solution to the areal data conversion problem based on identifying the posterior probability distribution of $\boldsymbol{S}|\boldsymbol{Z}^*$. This posterior probability distribution can be derived by the Bayes' rule as follows:

$$P\left(\boldsymbol{S}|\boldsymbol{Z}^*\right) \propto P\left(\boldsymbol{S}\right) P\left(\boldsymbol{Z}^*|\boldsymbol{S}\right) \tag{9}$$

where $P(\boldsymbol{S})$ denotes the prior probability distribution of $\boldsymbol{S}$, and $P(\mathbf{Z}^* \,|\, \mathbf{S})$ is its likelihood on the basis of the observed data.

Based upon the assumption in (3), $\boldsymbol{S}$ has a multivariate normal distribution. Furthermore, the conditional distribution of $\boldsymbol{Z}^*|\boldsymbol{S}$ can be derived from the distribution of $\boldsymbol{Z}$ as follows (Anderson 1958, p. 26):

$$Z^* \mid S \sim MVN\left(GS, G\Sigma_{\varepsilon}G'\right).$$

From the Gaussian nature of the distributions of $\boldsymbol{S}$ and $\boldsymbol{Z}^*|\boldsymbol{S}$, it follows that $\boldsymbol{S}|\boldsymbol{Z}^*$ also has a multivariate normal distribution.

Under the additional hypothesis of a known covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ the Bayesian approach described leads us to the following result (Pilz 1991):

$$\boldsymbol{S} \mid \boldsymbol{Z}^* \backsim MVN\left(\widehat{\boldsymbol{S}}, \boldsymbol{V}_{\widehat{\boldsymbol{S}}}\right)$$

where $\widehat{\mathbf{S}}$ and $\mathbf{V}_{\widehat{S}}$ are BIM estimators specified as follows:

$$\boldsymbol{V}_{\widehat{\boldsymbol{S}}} = \left[\boldsymbol{G}'\left(\boldsymbol{G}\frac{(\boldsymbol{D} - \rho_{\varepsilon}\boldsymbol{C})}{\sigma_{\varepsilon}^2}\boldsymbol{G}'\right)^{-1}\boldsymbol{G} + \frac{(\boldsymbol{D} - \rho\boldsymbol{C})}{\sigma_{\boldsymbol{S}}^2}\right]^{-1} \tag{10}$$

$$\widehat{\boldsymbol{S}} = \boldsymbol{V}_{\widehat{\boldsymbol{S}}}\left[\frac{(D - \rho C)}{\sigma_{\boldsymbol{S}}^2}\boldsymbol{T} + \boldsymbol{G}'\left(\boldsymbol{G}\frac{(D - \rho_{\varepsilon}C)}{\sigma_{\varepsilon}^2}\boldsymbol{G}'\right)^{-1}\boldsymbol{Z}^*\right] \tag{11}$$

Any inference on the original process can be carried out by the posterior distribution with parameters defined by (10) and (11). Point estimates for $\boldsymbol{S}$ can be obtained using $\widehat{\boldsymbol{S}}$, which is its *Maximum A Posterior* (MAP) estimate, as shown in Benedetti and Palma (1994). Confidence intervals and hypothesis tests can be performed in the standard way using multivariate normal distributions.

Additional issues are raised by the pycnophylactic, or mass preserving, property which consists of finding an estimate of $\boldsymbol{S}$ such that, by applying the transformation operator $\boldsymbol{G}$, the observed data are again obtained (Tobler 1979). In order to preserve this property, the posterior distribution of $\boldsymbol{S}|\boldsymbol{Z}^*$ is conditioned to the linear constraint $\boldsymbol{GS} = \boldsymbol{Z}^*$, so that the constrained version of the BIM estimators is obtained as follows:

$$\widetilde{\boldsymbol{S}} = \widehat{\boldsymbol{S}} + \boldsymbol{V}_{\widehat{\boldsymbol{S}}}\boldsymbol{G}'\left[\boldsymbol{G}\boldsymbol{V}_{\widehat{\boldsymbol{S}}}\boldsymbol{G}'\right]^{-1}\left(\boldsymbol{Z}^* - \boldsymbol{G}\widehat{\boldsymbol{S}}\right) \tag{12}$$

$$\boldsymbol{V}_{\widetilde{\boldsymbol{S}}} = \boldsymbol{V}_{\widehat{\boldsymbol{S}}} - \boldsymbol{V}_{\widehat{\boldsymbol{S}}}\boldsymbol{G}'\left[\boldsymbol{G}\boldsymbol{V}_{\widehat{\boldsymbol{S}}}\boldsymbol{G}'\right]^{-1}\boldsymbol{G}\boldsymbol{V}_{\widehat{\boldsymbol{S}}} \tag{13}$$

Point estimates of $\boldsymbol{S}$ can be obtained by using $\widetilde{\boldsymbol{S}}$; confidence intervals and hypothesis tests can be performed in the standard way using multivariate normal distributions.

Estimating Per Capita GDP for Spanish Provinces

Before proceeding to the estimation of per capita GDP for LLMs, the performance of the method is tested by disaggregating per capita GDP for NUTS II regions (17 Spanish *Autonomous Communities*) into NUTS III regions (50 *Provinces*), while treating the latter as not observed. Both figures are published by the National Statistics Institute. The availability of actual data at provincial level provides a benchmark for evaluation of the estimates and methodology.

According to the BIM, per capita GDP values for the 50 Spanish Provinces can be derived by applying the procedure described in section "The Bayesian Interpolation Method", to disaggregate data observed for Autonomous Communities. Specifically, the estimates for per capita GDP at provincial level are computed as the BIM estimates, given in Eqs. 11, 12. The computation of the unconstrained and constrained BIM estimates requires specifiying the prior mean vector $\boldsymbol{T}$ and the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{S}}$, as well as the aggregation matrix $\boldsymbol{G}$.

In the presence of covariates observed at provincial level, a natural way to express the prior mean of $\boldsymbol{S}$ is:

$$\mathbf{T} = \mathbf{X}\widehat{\boldsymbol{\beta}} \tag{14}$$

Table 1 Regression coefficient estimates for data at autonomous community (NUTS II) level

Coefficients	Model 1		Model 2		Model 3		Model 4	
Constant	−19.194	***	−30.913	***	−23.466	***	−47.533	***
(standard errors)	(3.764)		(6.716)		(4.777)		(9.667)	
Employed population	89.108	***			102.479	***		
(p-values)	(9.312)				(13.208)			
Active population			102.380	***			143.490	***
(p-values)			(14.417)				(22.747)	
Foreign population					−30.286		−67.900	*
(p-values)					(21.817)		(30.973)	
R^2	0.8592		0.7707		0.8763		0.8293	

Significance levels: '***' 0.001, '**' 0.01, '*' 0.05

where $\mathbf{X}$ is a $n \times k$ matrix which includes $k-1$ explanatory variables and an initial column of ones, and $\widehat{\boldsymbol{\beta}}$ is a $k \times 1$ vector of estimated regression parameters, including an intercept term.

There is great availability of socioeconomic variables at the provincial level that can be included as explanatory variables of the variable under study (per capita GDP). However, our final goal is to reconstruct per capita GDP figures for LLMs, and only a few socioeconomic variables -such as total population, population by age and gender, foreign population, active population, and employed population- are available for the Spanish LLMs. Therefore, to ensure coherence with the analysis carried out later at the LLM level, only some of these socioeconomic variables are included in our analysis.

The explanatory variables used to estimate the prior mean vector come from the 2001 Population and Household Census published and administrated by the INE, and the regression coefficients associated with these variables are estimated based upon data observed for the Autonomous Communities.

Table 1 shows results for different regression models estimated with the Autonomous Communities data. The coefficients associated with the explanatory variables included in Models 1 and 2 are both positive and highly significant, providing evidence of the contribution of the employed and active population in explaining per capita GDP. The coefficient associated with the foreign population is negative and not significant for Model 3 and negative and significant at the 0.05 level for Model 4. Comparing Models 1 and 2 we can see that the best fit is reported for Model 1, which exhibits a higher R^2.

The coefficients estimated for Model 1 are then used in the specification of the prior mean vector $\boldsymbol{T}$. The regression residuals' variance is used to estimate the scalar parameter σ_S^2 included in the covariance matrix $\boldsymbol{\Sigma}_S$. Three different values of the autocorrelation parameter, namely $\rho = 0.06$, $\rho = 0.2$ and $\rho = 0.8$, are in turn assumed. The proximity matrix $\boldsymbol{C}$ is constructed according to the 5 nearest neighbors criterion, and is conveniently symmetrized.[8]

[8]Estimations were also run using different k values for the C matrix. No significant differences were found.

The aggregation matrix $\boldsymbol{G}$ expresses the aggregation of NUTS III (*Provinces*) into NUTS II (*Autonomous Communities*). The entries of $\boldsymbol{G}$ are specified as follows:

$$g_{ji} = h_{ji}\frac{P_i}{P_j} \qquad i = 1,\ldots,n, \quad j = 1,\ldots,m \tag{15}$$

where P_i and P_j denote the Total Population in the *i-th* Province and in the *j-th* Autonomous Community respectively, and h_{ji} takes the value 1 if the *i-th* Province belongs to the *j-th* Autonomous Community, and 0 otherwise. The ratio $\frac{P_i}{P_j}$ is needed so that *per capita* values for Provinces can be aggregated into per capita values for Autonomus Communities.

The BIM estimates are computed in their unconstrained version – Eq. (11) – and constrained version – Eq. (12). As explained above, the constrained solutions preserve the pycnophylactic property (which ensures that the values estimated for the target zones after the aggregation yield a total equal to the total of the values observed for the source zones); the pycnophylactic property is not considered in the unconstrained version of the BIM estimates. Table 2 shows the evaluation of the accuracy of the estimates – that is, the comparison between the estimated and the actual values – using indices from the forecasting literature such as as the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE).[9]

Table 2 Accuracy of constrained and unconstrained per capita GDP estimates for provinces (prior mean specified by Model 1)

ρ	Type	Correlation	RMSE	MAE	MAPE
0.06	*Unconstrained*	0.89107	1.68583	1.39997	0.09493
0.06	*Constrained*	0.94021	1.25207	0.85685	0.05952
0.2	*Unconstrained*	0.89107	1.68583	1.39997	0.09493
0.2	*Constrained*	0.93966	1.25697	0.85770	0.05961
0.8	*Unconstrained*	0.89107	1.68582	1.39997	0.09493
0.8	*Constrained*	0.92973	1.36126	0.92428	0.06611

[9]The considered accuracy measures are computed as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y_i}|$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y_i}}{y_i}\right|$$

As expected, the constrained estimates perform better than the unconstrained versions (the calculated values of the accuracy measures are lower). The constrained solutions also exhibit a higher correlation with actual data. The values of the correlation coefficient for the RMSE, MAE and MAPE appear to be essentially independent of the hypothesis regarding ρ, which reveals the robustness of the estimates with respect to the autocorrelation parameter.

Results of the estimated per capita GDP for Provinces are reported in Table 3. The results assuming Model 1 and Model 2 in the specification of the prior mean and variance are compared with the actual values. There is an improvement in the estimates when the prior mean is specified according to Model 1, which is confirmed by the computed accuracy measures.

The estimated per capita GDP values suggest that the BIM performs well in converting data from the NUTS II to the NUTS III level. This testing exercise validates the application of the BIM to higher levels of spatial disaggregation. Results for the estimation of per capita GDP for the Spanish LLMs are presented in the following section.

Estimating Per Capita GDP for Spanish LLMs

The 806 LLMs into which the Spanish territory is divided (Table 4) show high disparities in size, in terms of both population and number of municipalities. The largest LLMs in terms of population are the Madrid and Barcelona LLMs, which comprise 20.51% of the total Spanish population. More than 85% of the LLMs can be considered rural (i.e., less than 50,000 inhabitants), but only 23.23% of the population live in these rural LLMs, which are typically located in inaccessible/poorly-connected areas characterized by hilly terrain -such as those in the northern part of the country.[10]

In this section, the BIM is applied to disaggregate per capita GDP data for the 50 Spanish provinces in order to determine per capita GDP for the 804 LLMs.[11] Data on per capita GDP for the 804 Spanish LLMs are computed as the constrained BIM estimates given in (12). The computation of the BIM estimates requires specifying the prior mean vector and the covariance matrix of $\boldsymbol{S}$ as well as the aggregation matrix $\boldsymbol{G}$.

where y_i and $\widehat{y}_i$ denote real and estimated values respectively. RMSE and MAE can range from 0 to $+\infty$ and are negatively oriented values, so that lower values of these indices means better estimates.

[10] As expected, the delimitation of LLMs based upon commuting criteria does not match the boundaries of the administrative regions; consequently, there are a few cases in which municipalities included in the same LLM belong to different NUTS III regions. As will be explained below, this fact has been taken into account when defining the aggregation matrix.

[11] Ceuta and Melilla LLMs are not considered in the analysis due to their particular status of autonomous cities.

Table 3 Disaggregation of per capita GDP (EUR) for Spanish provinces

	Actual values	Prior mean specified by Model 1			Prior mean specified by Model 2		
		$\rho = 0.06$	$\rho = 0.2$	$\rho = 0.8$	$\rho = 0.06$	$\rho = 0.2$	$\rho = 0.8$
Álava	23,164	22,003	21,876	21,023	21,550	21,312	19,901
Albacete	12,530	12,994	13,013	12,906	14,036	14,068	14,000
Alicante	15,017	15,684	15,666	15,291	15,398	15,264	14,311
Almería	15,584	18,975	18,983	18,688	17,925	17,969	17,928
Ávila	13,346	13,131	13,085	12,457	12,003	11,869	10,788
Badajoz	10,710	9281	9222	8620	9759	9796	10,115
Balears	20,875	20,875	20,875	20,875	20,875	20,875	20,875
Barcelona	20,212	20,487	20,468	20,303	20,969	20,963	20,929
Burgos	18,787	18,759	18,702	18,310	18,661	18,558	17,932
Cáceres	10,805	13,131	13,226	14,207	12,353	12,292	11,773
Cádiz	12,594	9333	9235	8414	11,556	11,469	10,927
Castellón	19,039	19,779	19,686	19,031	17,779	17,642	16,768
Ciudad Real	13,277	10,780	10,748	10,465	10,739	10,690	10,423
Córdoba	10,923	10,960	10,883	10,332	13,154	13,136	13,087
Coruña, A	13,327	12,359	12,381	12,512	12,840	12,802	12,789
Cuenca	12,843	12,407	12,386	12,091	10,383	10,314	9997
Girona	20,832	21,382	21,399	21,481	20,458	20,548	21,187
Granada	11,119	10,793	10,883	11,586	11,535	11,672	12,503
Guadalajara	15,671	17,388	17,350	17,170	16,886	16,873	16,904
Guipúzcoa	21,649	21,970	21,964	21,756	21,262	21,257	21,143
Huelva	13,057	11,448	11,494	12,293	12,969	12,892	12,190
Huesca	17,688	17,607	17,647	17,827	15,419	15,524	16,244
Jaén	10,849	9941	9992	10,135	9074	9098	9498
León	13,957	13,566	13,601	13,890	13,549	13,658	14,438
Lleida	20,816	20,038	20,099	20,577	17,352	17,562	18,921
Rioja, La	18,881	18,881	18,881	18,881	18,881	18,881	18,881
Lugo	12,591	13,961	13,937	14,037	12,321	12,432	13,347
Madrid	22,377	22,377	22,377	22,377	22,377	22,377	22,377
Málaga	12,683	14,226	14,247	14,215	13,197	13,264	13,823
Murcia	13,901	13,901	13,901	13,901	13,901	13,901	13,901
Navarra	21,349	21,349	21,349	21,349	21,349	21,349	21,349
Ourense	12,095	10,927	10,907	10,911	9558	9563	9813
Asturias	14,160	14,160	14,160	14,160	14,160	14,160	14,160
Palencia	15,850	15,324	15,359	15,429	15,229	15,184	14,986
Palmas	17,014	16,804	16,752	16,513	16,852	16,755	16,314
Pontevedra	13,195	14,263	14,254	14,054	14,836	14,837	14,398
Salamanca	13,940	13,789	13,679	12,957	14,672	14,533	13,558

(continued)

Table 3 (continued)

	Actual values	Prior mean specified by Model 1			Prior mean specified by Model 2		
		$\rho = 0.06$	$\rho = 0.2$	$\rho = 0.8$	$\rho = 0.06$	$\rho = 0.2$	$\rho = 0.8$
Sta. Cruz de Tenerife	15,729	15,961	16,018	16,281	15,908	16,014	16,500
Cantabria	15,979	15,979	15,979	15,979	15,979	15,979	15,979
Segovia	16,914	17,135	17,120	16,859	15,121	15,144	14,999
Sevilla	12,580	13,392	13,396	13,684	11,645	11,592	11,205
Soria	17,202	19,069	19,144	19,952	16,350	16,140	14,724
Tarragona	21,406	19,189	19,287	20,229	17,833	17,673	16,544
Teruel	18,090	14,691	14,766	15,627	11,261	11,127	10,220
Toledo	13,110	14,645	14,681	15,165	14,867	14,918	15,306
Valencia	16,070	15,468	15,500	15,891	16,094	16,212	17,031
Valladolid	17,143	17,613	17,631	17,899	19,036	19,104	19,647
Vizcaya	19,355	19,458	19,494	19,835	19,996	20,060	20,487
Zamora	12,030	11,792	11,933	12,754	10,528	10,725	11,950
Zaragoza	17,698	18,235	18,214	18,040	19,278	19,273	19,240
Correlation		0.940	0.940	0.930	0.891	0.888	0.856
RMSE		1.252	1.257	1.361	1.706	1.728	1.966
MAE		0.857	0.858	0.924	1.180	1.201	1.407
MAPE		5.95%	5.96%	6.61%	7.87%	7.99%	9.19%

Source: 2001 Spanish Regional Accounts, INE; and 2001 Population and Household Census, INE (2007)

Table 4 Distribution of LLMs by population size (2001)

Names or numbers of LLMs of this size		Number of municipalities	% of total population
2,500,000 $\leq$ Inhabitants	Madrid	153	20.51%
	Barcelona	151	
500,000 $\leq$ Inhabitants <2,500,000	Valencia	52	16.49%
	Sevilla	39	
	Bilbao	59	
	Malaga	20	
	Zaragoza	96	
	Palmas de Gran Canaria	15	
	Sabadell	17	
	Sta. Cruz Tenerife	17	
100,000 $\leq$ Inhabitants <500,000	60	2102	31.20%
50,000 $\leq$ Inhabitants <100,000	50	666	8.56%
Inhabitants <50,000	686	4822	23.23%
Total population: 40,393,173	806	8106	40,847,371
	LLMs	municipalities	Inhabitants

Source: 2001 Population and Household Census, INE (2007)

Table 5 Economic and socio-economic variables for the 804 Spanish LLMs classified by population size

LLM population size	% active population	% employed population	% foreign population
2,500,000 ≤ Inhabitants	15.63%	16.04%	33.03%
500,000 ≤ Inhabitants <2,500,000	18.07%	17.67%	10.92%
100,000 ≤ Inhabitants <500,000	33.61%	33.75%	22.48%
50,000 ≤ Inhabitants <100,000	9.23%	9.18%	13.07%
Inhabitants <50,000	23.45%	23.36%	20.49%
Total population: 40,393,173 inhabitants	17,499,182	14,976,512	1,572,013

Source: 2001 Spanish Population and Household Census, INE (2007)

Table 6 Regression coefficient estimates for data at provincial level

Coefficients	Model 1		Model 2		Model 3		Model 4	
Constant	−13.902	***	−21.785	***	−15.912	***	−23.328	***
(standard errors)	(2.169)		(4.497)		(2.519)		(5.611)	
Employed population	75.975	***			82.705	***		
(standard errors)	(5.586)				(6.995)			
Active population			83.083	***			87.109	***
(standard errors)			(9.914)				(13.201)	
Foreign population					−20.515		−9.169	
(standard errors)					(13.157)		(19.640)	
R^2	0.794		0.594		0.804		0.596	

Significance levels: '***' 0.001, '**' 0.01, '*' 0.05

The prior mean vector is defined as in (14). For explanatory variables used to define the prior mean vector, we consider employed population, active population and foreign population. Data on these variables are available at the LLM level, and come from the 2001 Population and Household Census (INE 2007). Data on active population, employed population and foreign population for LLMs grouped by LLM population size are summarized in Table 5.

The coefficients associated with the explanatory variables are estimated based upon provincial data. Results for different specifications of the regression model estimated with per capita provincial data are displayed in Table 6. The regression residuals' variance is used to estimate the scalar parameter σ_S^2.

Given the expected relationship between labor variables (employment and active population) and economic activity, Models 1 and 2 show high coefficients of determination. A higher $R^2 = 0.794$ is reported for Model 1, which only considers the Employed Population in explaining per capita GDP. The regression coefficient associated with this independent variable is positive and highly significant, as is the coefficient associated with the Active Population in Model 2. The results related to Models 3 and 4 reveal that coefficients associated with the Foreign Population are negative but not significant. Therefore, in order to define the prior mean vector $\boldsymbol{T}$

we will focus on Model 1, which appears perform best[12] at the provincial as well as at the Autonomous Community level (see section "Estimating Per Capita GDP for Spanish Provinces").

In the conversion of provincial data to the LLM level, some important issues have to be considered. First, given that the disaggregate (LLM-level) data are not available, the accuracy of any predicted value simply cannot be verified. Secondly, as mentioned in section "Conclusions and future research", several LLMs go beyond the provincial limit, so that some part of the *i-th* LLM can belong to a given province *j* and rest to one or more other provinces (see Fig. 1).

In order to take into account this feature, the elements of the aggregation matrix $\boldsymbol{G}$, which expresses the LLMs' aggregation into provinces, will be constructed according to a weighted sum. A weight h_{ji} is assigned to each LLM i belonging to the province j according to the following scheme:

$$h_{ji} = \begin{cases} 1 \text{ if the LLM } i \text{ belongs only to the province } j \\ k_{ji} \text{if the LLM } i \text{ belongs to the province } j \\ 0 \text{ otherwise} \end{cases}$$

where k_{ji} is a real value chosen to be proportional to the active population in the municipalities included in the *i-th* LLM, and such that $0 < k_{ji} < 1$ and $\sum_{j=1}^{m} k_{ji} = 1$, for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. The entries of the aggregation matrix $\boldsymbol{G}$ are therefore specified as follows:

$$g_{ji} = h_{ji} \frac{P_i}{P_j} \qquad i = 1, \ldots, n, \quad j = 1, \ldots, m \tag{16}$$

where P_i and P_j denote the total population in the *i-th* LLM and in the *j-th* province, respectively, in order to take into account the aggregation of *per capita* figures.

The BIM estimates are then carried out in their constrained version on the basis of the above specified prior mean vector, covariance and aggregation matrices. Different values for the spatial autocorrelation parameter, namely $\rho = 0.06$, $\rho = 0.2$ and $\rho = 0.8$, are considered. The proximity matrix $\boldsymbol{C}$ is specified according to the k nearest neighbours method ($k = 5$), and is conveniently symmetrized.

The obtained BIM estimates defines the per capita GDP for LLMS. The results for each of the 804 LLMs, obtained by setting different values for ρ, can be easily

[12] Regressions with additional variables from the *Anuario Economico de España* (La Caixa 2011), with data available at local level, were also estimated: commercial activities, number of banks, number of fixed telephone lines, vehicles available, etc. However, the coefficients for these variables were not significant and the models showed a goodness of fit inferior to the models displayed in Table 3.

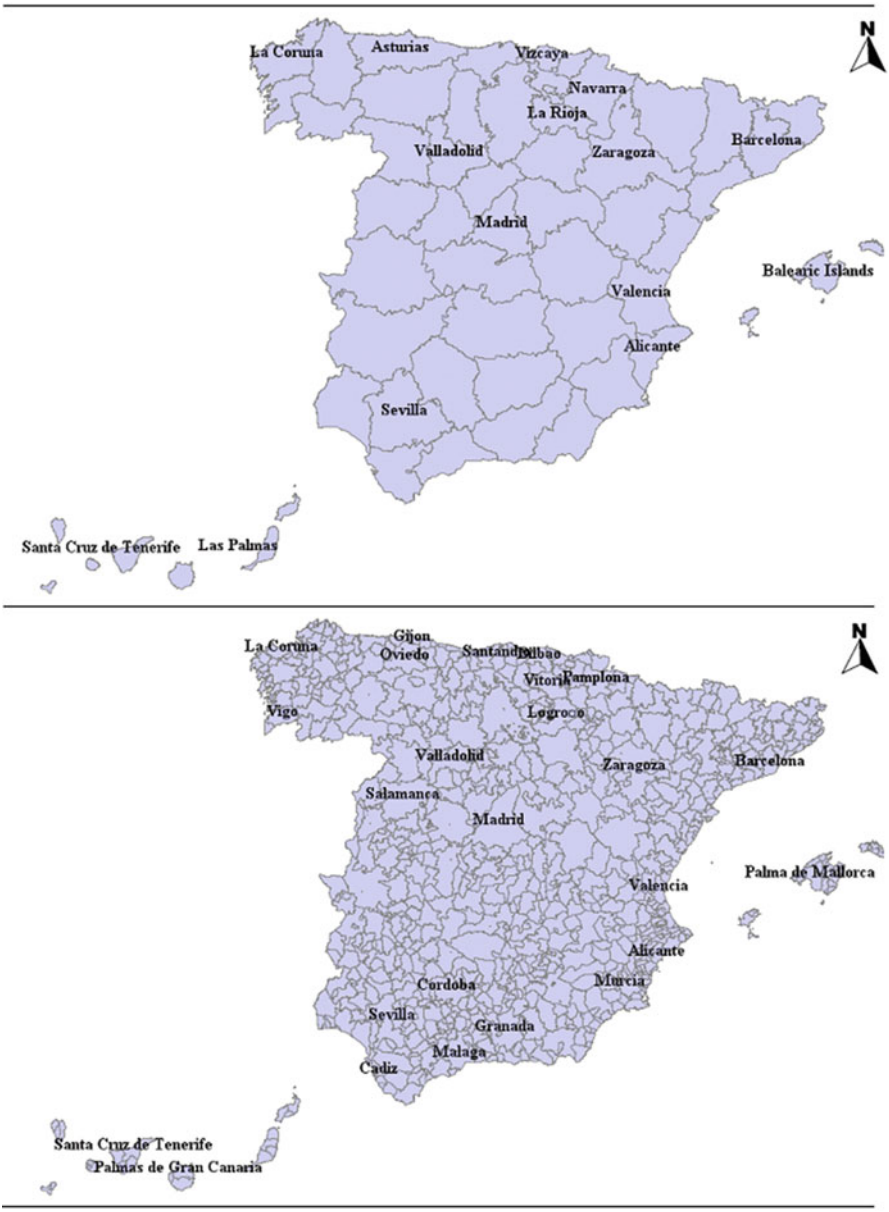

Fig. 1 Provinces and LLMs in Spain

Table 7 Estimated per capita GDP (EUR) for Spanish LLMs classified by population size

Population size	Population	Estimated per capita GDP		
		$\rho = 0.8$	$\rho = 0.2$	$\rho = 0.06$
Madrid LLM	*5,287,342*	*22,488*	*22,487*	*22,487*
Barcelona LLM	*3,027,950*	*19,422*	*19,701*	*19,733*
2,500,000 ≤ Inhabitants		21,412	21,480	21,486
500,000 ≤ Inhabitants <2,500,000		15,940	15,847	15,837
100,000 ≤ Inhabitants <500,000		16,192	16,173	16,170
50,000 ≤ Inhabitants <100,000		16,304	16,299	16,304
Inhabitants <50,000		14,338	14,328	14,327

Source: 2001 Spanish Population and Household Census, INE (2007)

visualized in a map or aggregated in a table according to size[13] as displayed in Table 7.

The results appear to confirm the existence of internal spatial disparities in per capita GDP between the metropolitan areas – LLMs comprised of a core city and surrounding municipalities with a total population greater than 2.5 million inhabitants – and LLMs in the rest of the country, comprised of small and medium size cities and rural areas. The larger the size (quantified by population) of the LLM area, as defined in Table 7 above, the higher its estimated per capita GDP. This suggests the importance of agglomeration economies. Besides the positive effects derived from large-scale production and positive externalities associated with size, agglomeration economies associated with urban concentration lead to lower recruitment and training costs and increased, knowledge spillovers and competition (Porter 1990; Beardsell and Henderson 1999; Rosenthal and Strange 2004). In other words, large metropolises stimulate the exchange of knowledge, productivity, and economic growth, and therefore economic welfare. In his hierarchy of central places, Christaller (1966) states that not all central regions or cities are equal – there are higher-order centers, with a greater concentration of economic activity, as well as other lower-order places.

Another relevant factor in regional studies is the importance of distance, and the manner in which transportation costs can affect the business location, level of economic activity, employment, employment rates and also per capita GDP. As suggested by Polèse and Shearmur (2004) for Canada and by Polèse et al. (2006) for Spain, not only the larger cities benefit from agglomeration economies but areas close to them share in these benefits as well. For a graphical demonstration of this related to the present study, see Fig. 2, where estimated per capita GDP values (with $\rho = 0.06$) above and below the Spanish average are depicted for provinces and also for LLMs.

[13]For each LLM we have the estimated per capita GDP and also actual population, so we can calculate their estimated GDP and an average per capita GDP figure for different size categories of LLMs.

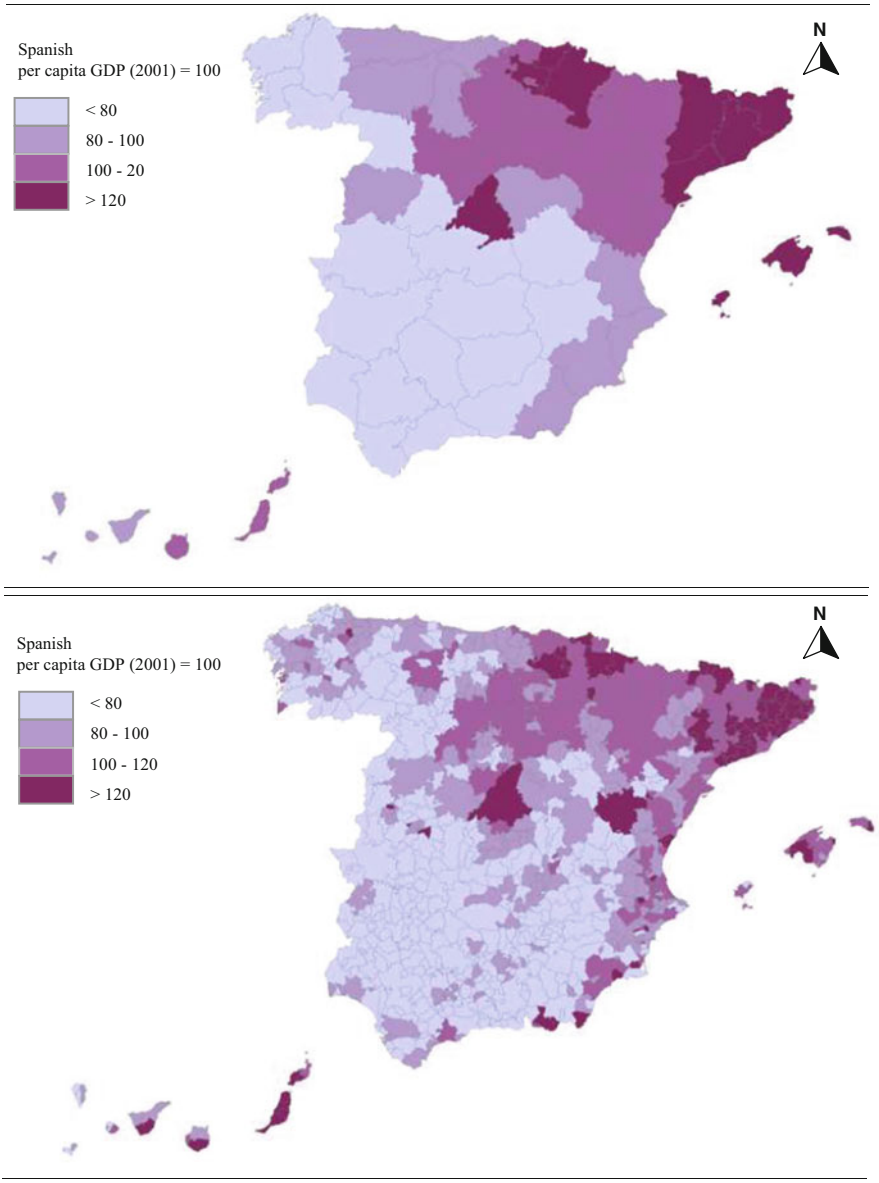

Fig. 2 *Actual* per capita GDP (provinces) and *estimated* per capita GDP for LLMs ($\rho = 0.06$)

As Polèse (2009) assessed, relative location matters. The location of a local labor market must be considered not only in regard to the national urban system, but also in the context of international connections. Proximity to international borders with important trade flows could be relevant, and this appears to be confirmed in

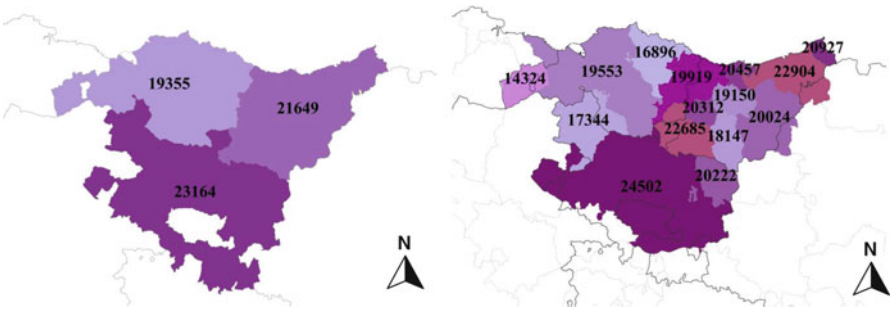

Fig. 3 *Actual* per capita GDP (provinces) and *estimated* per capita GDP for Basque LLMs ($\rho = 0.06$)

the Spanish case. In Spain, there is a northeastern concentration of the highest per capita GDP values, both by provinces and by LLM. However, by working with disaggregated data, we reveal the internal spatial disparities and the importance of the largest cities, as well as the dichotomy between the rural and the urban areas. As an example, Fig. 3 shows detailed results for LLMs that conform the Basque Country Autonomous Community (NUTS II), whose average per capita GDP was EUR 20,618 according to the 2001 Spanish Regional Accounts (published by the INE).

Divided into three provinces or NUTS III regions (Alava, Guipuzcoa and Vizcaya), the Basque Country is a highly industrialized and urbanized autonomous community with a strategic location beside the French border. In the Basque Country, there are severe intra-regional disparities in terms of per capita GDP that cannot be clearly observed at the provincial level. The severity of intra-regional differences in GDP per capita that is readily apparent at the LLM level (Fig. 3) appears to be far less pronounced when aggregated at the province level – where the GDP per capita differences for Alava, Guipuzcoa, and Vizcaya (EUR 23,164, 21,649 and 19,355, respectively) do not tell the full story of the region's disparity in terms of this measure. Previously, researchers could not verify or test the provincial-level results, as GDP per capita figures were not available at a highly disaggregate level. Therefore, the availability of this LLM-level dataset opens the door to these and many other questions related to, for instance, the impact of regional policies at the local level or the weight of cities in regional economies.

Conclusions and Future Research

Regional economists use administrative units as a proxy of the *region* in their empirical analyses, either due to the lack of alternatives or the impossibility of having a region consistent with the theoretical assumptions of regional economics.

Such a basic economic variable as per capita GDP is only available in Spain at the NUTS III level, which corresponds to provinces. However, we believe that the ideal unit of analysis is the further disaggregated LLM. Defined by the daily commuting patterns of its residents, the larger LLMs (larger cities and their surrounding areas) are presumably the national engines of growth, where most of the national GDP is concentrated. However, no data on GDP – either total or per capita – are available at the LLM level of disaggregation.

In this paper, we have estimated per capita GDP for Spanish LLMs using the Bayesian Interpolation Method. As expected, our results confirm the existence of wide disparities in per capita GDP within the administrative regions – the units commonly used in any regional study of this type – and the importance of agglomeration economies and relative location.

The proposed method was also applied to estimate per capita GDP for provinces (NUTS III level) using the data available for Autonomous Communities (NUTS II regions). Comparisons with the actual GDP per capita data available for provinces have revealed good performance by the proposed method, thereby validating its application to estimating per capita GDP at a higher level of spatial disaggregation (i.e., LLMs).

Although preliminary, the results obtained lead us to believe that this line of research should be explored in more detail. Directions for future research could include using additional auxiliary variables to explain per capita GDP, considering different specification of the neighborhood structure for LLMs or applying this methodology to some other geographical scenarios.

Acknowledgments This work is a result of the efforts of G.J.D. Hewings (University of Illinois at Urbana-Champaign) in bringing people together to prove that one plus one is greater than two. To him we give our infinite gratitude. The authors would also like to thank R. Benedetti (University of Pescara) for his useful comments and explanations on technical issues, and P. Postiglione (University of Pescara) for his suggestions.

A preliminary version of this chapter was presented at the RSAI 2012 (Timisorara, Romania) and also at a seminar of the Regional Economic Analysis Laboratory (REGIOlab) of the University of Oviedo. The authors would like to thank participants at both events for the comments and suggestions received. Finally, the chapter has greatly benefitted from the comments of the anonymous referees.

References

Andersen, A. K. (2002). Are commuting areas relevant for the delimitation of administrative regions in Denmark? *Regional Studies, 36*(8), 833–844.

Anderson, T. W. (1958). *An introduction to multivariate statistical analysis*. New York: John Wiley.

Armstrong, H., & Taylor, J. (1993). *Regional economics*. London: Harvester Wheatsheaf.

Ball, R. M. (1980). The use and definition of travel-to-work areas in great Britain: Some problems. *Regional Studies, 14*(2), 125–139.

Bartik, T. J. (1996). The distributional effects of local labor demand and industrial mix: Estimates using individual panel data. *Journal of Urban Economics, 40*, 150–178.

Beardsell, M., & Henderson, V. (1999). Spatial evolution of the computer industry in the USA. *European Economic Review, 43*(2), 431–456.

Behrens, K., & Thisse, J. F. (2007). Regional economics: A new economic geography perspective. *Regional Science and Urban Economics, 37*, 457–465.

Benedetti, R., & Palma, R. (1994). Markov random field-based image subsampling method. *Journal of Applied Statistics, 21*(5), 495–509.

Besag, J. (1974). Spatial iteraction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society B, 36*, 192–225.

Boix, R., & Galleto, V. (2006). Sistemas industriales de trabajo y distritos industriales marshallianos en España. *Economia Industrial, 359*, 165–184.

Caixa, L. (2011). *Anuario Económico de España*. Barcelona: Servicio de Estudios de la Caja de Ahorros y Pensiones de Barcelona.

Chow, G. C., & Lin, A. (1971). Best linear unbiased interpolation, distribution and extrapolation of time series by related series. *The Review of Economics and Statistics, 53*(4), 372–375.

Christaller, W. (1966). *Central places in southern Germany: Translated from Die zentralen Orte in Süddeutschland*. Englewood Cliffs: Prentice Hall.

Clayton, D., & Berardinelli, L. (1992). Bayesian methods for mapping disease risk. In P. Elliott, J. Cuzick, D. English, & R. Stern (Eds.), *Geographical and environmental epidemiology: Methods for small-area studies* (pp. 205–220). Oxford: Oxford University Press.

Coombes, M. G., & Openshaw, S. (1982). The use and definition of travel to work areas in Great Britain: Some comments. *Regional Studies, 16*, 141–149.

Coombes, M. G., Green, A. E., & Openshaw, S. (1986). An efficient algorithm to generate official statistical reporting areas: The case of the 1984 travel-to-work areas revision in Britain. *Journal of the Operational Research Society, 37*(10), 943–953.

Flowerdew, R., & Green, M. (1992). Developments in areal interpolation methods and GIS. *The Annals of Regional Science, 26*, 67–78.

Fox, K. A., & Kumar, T. K. (1965). The functional economic area: Delineation and implications for economic analysis and policy. *Papers in Regional Science, 15*(1), 57–85.

Gelfand, A. E., & Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics, 4*(1), 11–25.

Goodchild, M. F., Anselin, L., & Deichmann, U. (1992). A framework for the areal interpolation of socioeconomic data. *Environment and Planning A, 25*, 383–397.

Hughes, G., & McCormick, B. (1994). Did migration in the 1980s narrow the north-south divide? *Economica, 61*, 509–527.

INE. (2007). 2001 population and household census (*Censo de Población y Viviendas*, 2001), Instituto Nacional de Estadística, Madrid. Available online at www.ine.es.

INE, Spanish Regional Accounts. (2001). *Contabilidad Regional.* Instituto Nacional de Estadistica, Madrid. Available online at www.ine.es.

Melo, P. C., Graham, D. J., & Nolan, R. B. (2009). A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics, 39*(3), 332–342.

Moretti, E. (2011). Chapter 14 – Local labor markets. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics*, *Volume 4, Part B*. Amsterdam: Elsevier.

Palma, D., & Benedetti, R. (1998). A transformational view of spatial data analysis. *Geographical System, 5*, 199–220.

Parr, J. (2008). Cities and regions: Problems and potentials. *Environment and Planning A, 40*, 3009–3026.

Peeters, L., & Chasco, C. (2006). Ecological inference and spatial heterogeneity: An entropy-based distributionally weighted regression approach. *Papers in Regional Science, 85*(2), 257–276.

Pilz, J. (1991). *Bayesian estimation and experimental design in linear regression models*. New York: John Wiley.

Polasek, W., Llano, C., & Sellner, R. (2010). Bayesian methods for completing data in spatial models. *Review of Economic Analysis, 2*, 194–214.

Polèse, M. (2009). *The wealth and the poverty of regions: Why cities matters*. Chicago: University of Chicago Press.

Polèse, M., & Shearmur, R. (2004). Is distance really dead? Comparing the industrial location patterns over time in Canada. *International Regional Science Review, 27*(4), 1–27.

Polèse, M., Shearmur, R., & Rubiera, F. (2006). *Observing regularities in location patterns. An analysis of the spatial distribution of economic activity in Spain*. Montreal: INRS-Internal Document.

Porter, M. (1990). *The competitive advantage of nations*. New York: Free Press.

Ripley, B. D. (1981). *Spatial statistics*. New York: John Wiley.

Rosenthal, S. S., & Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. *Handbook of Regional and Urban Economics, 4*, 2119–2171.

Sforzi, F., Openshaw, S., & Wymer, C. (1997). Le procedura di identificazione dei sistemi locali del lavoro [The procedure to identify local labour market area]. In F. Sforzi (Ed.), *I sistemi locali del lavoro 1991* (pp. 235–242). Rome: ISTAT.

Smart, M. W. (1974). Labour market areas: Uses and definitions. *Progress in Planning, 2*, 239–353.

Tobler, W. R. (1979). Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association, 74*, 519–530.

Topel, R. H. (1986). Local labor markets. *The Journal of Political Economy, 94*(3), S111–S143.

Wall, M. M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference, 121*, 311–324.

Discovering Functional Zones in a City Using Human Movements and Points of Interest

Nicholas Jing Yuan, Yu Zheng, and Xing Xie

Abstract The development of a city gradually fosters different functional zones, such as educational areas and business districts. Typically, a city is segmented into disjointed regions by major roads, such as highways and urban expressways. In this chapter, we propose a framework for discovering functional zones in a city through the analysis of human mobility among regions and points of interest (POIs) within regions. Specifically, we infer the functions of each region from the results of a topic-modeling-based approach, which regards a region as a *document*, a function as a *topic*, categories of POIs (e.g., restaurants and shopping malls) as *metadata* (such as *authors*, *affiliations*, and *keywords*), and human mobility patterns (origins/destinations and arrival/departure times) as *words*. As a result, a region can be represented by a distribution of functions. This type of representation enables *functional zones* to be identified, which are comprised of clusters of regions with similar distributions of functions. We then further identify the intensity of each functional zone type occurring in different locations. We evaluated our method using large-scale, real-world datasets, consisting of two POI datasets of Beijing (in 2010 and 2011) and two three-month GPS trajectory datasets (representing human mobility); these trajectories were generated by over 12,000 taxicabs in Beijing in 2010 and 2011 respectively. The results demonstrate the advantages of our approach over baseline methods which solely use POIs or human mobility.

Keywords Functional zones • Urban computing • Taxi trajectories • Human mobility

Introduction

Modern cities develop by the formation of different *functional zones*, which serve the citizens through various urban functions in order to satisfy various needs and make possible diverse socioeconomic activities-e.g., Manhattan is a well-known

N.J. Yuan (✉) • Y. Zheng • X. Xie
Microsoft Research Asia, No. 5 Danling District, Beijing, China
e-mail: nicholas.yuan@microsoft.com; yuzheng@microsoft.com; xing.xie@microsoft.com

J.-C. Thill (ed.), *Spatial Analysis and Location Modeling in Urban and Regional Systems*, Advances in Geographic Information Science,
https://doi.org/10.1007/978-3-642-37896-6_3

financial district in New York city, and Silicon Valley is a typical high-technology business district in the San Francisco Bay Area. These functional zones can either be preemptively designed by urban planners, or naturally formulated according to the actual lifestyles of the population. Meanwhile, both the extents and functions of these zones could be reshaped during the evolution of a city.

An urban functional zone is canonically characterized by its high frequency of intra-regional economic interactions (Karlsson and Olsson 2006), its agglomeration of activities, its intra-regional transport infrastructure, and the mobility of people and inputs/outputs within and across its borders (Karlsson 2007). Although the exact definition of an urban functional zone may vary in general, existing approaches for identifying urban functional zones are often based upon population density and commuting travel-to-work data (Flórez-Revuelta et al. 2008), which are often obtained by census or transportation surveys (Cörvers et al. 2009).

Recently, the proliferation of ubiquitous sensing technologies, smart cities, and location-based services has led to the growing availability of human trajectories. For example, in big cities like New York, Munich and Beijing, most taxis are equipped with GPS devices for dispatching and security management. These taxis regularly report their locations to the data center with a certain reporting frequency. Hence, a large number of taxi trajectories are accumulated every day. In addition to movements, these trajectories imply the activities of citizens, at various locations and times, since the activities are the essential reasons, which mobilize people to travel between different places.

In this chapter, we aim to discover functional zones in urban areas through the analysis of human mobility and points of interest (POIs). Here, a city is partitioned into individual regions by major roads, such as highways and ring roads (Fig. 1a); in this analysis of functional zones, each region is considered to be a unit of

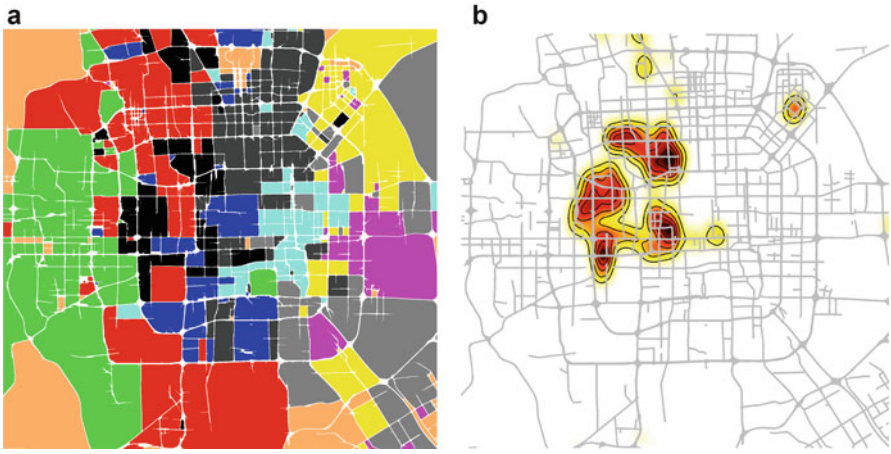

Fig. 1 The objectives of this work: (**a**) Functional zones; (**b**) Intensity of a function (Bing Map data: 2011)

study. Human mobility can be represented by individual movement trajectories (i.e., cell-tower traces in a cellular network), or trajectories of driving routes, or a sequence of posts (like geo-tweets, geo-tagged photos, or check-ins) in location-based services (Zheng and Zhou 2011). A POI is associated with a coordinate and an associated *category*, such as restaurants and shopping malls. Specifically, regions with the same function are assigned the same color (Fig. 1a) and the intensity of each function is identified at its respective locations of occurrence. For example, Fig. 1b presents the functionality intensity of the function type termed developed commercial/entertainment areas in Beijing, where darker shades suggest a higher intensity. This step is motivated by the observations that sometimes, only part of a region contributes to a function-yet, on the other hand, a function may be distributed across several individual regions (e.g., a shopping street). Finally, each function is annotated in a semi-manual way based upon the output of our method.

Discovering functional zones can enable a variety of valuable applications. First, it can provide a quick understanding of a complex city (such as New York City, Tokyo, or Paris), as well as social recommendations. For example, tourists can easily differentiate scenic areas from business districts given these functional zones, thereby reducing the efforts for trip planning. Local people can also expand their knowledge about a city by finding additional regions that have similar functions (e.g., entertainment areas) to those with which they are already familiar. It is very common for local people to lack knowledge of parts of a metropolis even if they have been residing in the metropolis for a few years. Second, these functional zones can calibrate the urban planning of a city and contribute to its future planning to some extent. It is not surprising when a city does not evolve as originally planned, given the complexity of urban planning itself and the difficulty in predicting the development of a city. Third, these functional zones would also benefit in terms of location choices (site selection) for businesses or outdoor advertising. For instance, when building a supermarket the distance to residential areas must be considered, and outdoor advertising for a training course could be best sited by considering the geospatial intensity of the educational function.

In order to identify the functions of a region, both POIs located in a region and human mobility among regions must be taken into account. Therefore, the following two aspects are the focus of this analysis:

(1) *POI data*: On the one hand, POIs may represent the function of a region. For example, a region containing a number of universities and schools is likely to be an educational area. On the other hand, a region usually contains a variety of POIs, thereby having compound functions instead of a single function. Some regions may serve as both business districts and entertainment areas in a city. In addition, the information from the POI data used in this analysis cannot differentiate the quality of different venues or reflect the interactions between functional zones. For instance, restaurants are everywhere in a city-however, they could denote different functions. Some small restaurants were built just for satisfying local residents' daily needs, while a few famous restaurants attracting

many people might be regarded as a feature of an entertainment area. As a result, sometimes two regions sharing a similar distribution of POIs could still have different functions.

(2) *Human mobility*: The function of a region has a strong correlation with the traveling behavior of people who visit the region from outside of its borders. The knowledge that human mobility contributes revealing reveal the functions of a region, is the product of two components. One is the time that an individual arrives at or departs from a region. The other is where people come from and leave for. Intuitively, on a workday people usually leave a residential area in the morning and return in the evening. The major time when people visit an entertainment area, however, is during the evening on workdays, or at any given time on non-workdays. Furthermore, regions are correlated by function types in the context of human mobility. For instance, people arriving at an entertainment area have a high probability of having departed from a working area (on a workday), and a residential area (on non-workday). As a result, two regions are more likely to have similar functions, if people travel to the two regions from similar functional zones, or depart for similar zones.

This chapter offers the following contributions:

- We propose a topic model-based method to identify the functions of individual regions, which are obtained using a morphological image segmentation approach. The proposed method regards a region as a *document*, deems a function as a *topic*, denotes human mobility related to the region as *words*, and treats POIs located in a region as *metadata* (such as titles, authors, affiliations, and keywords). As a result, a region is represented by a distribution of *topics* (functions), and each *topic* is a denoted by a distribution of *words* (human mobility). This model fits the research intuitively, reducing the problem of data sparsity, e.g., the mobility data associated with a particular POI might be quite sparse.
- We infer the territory of these functions by clustering regions into groups according to the topic (function) distribution of each region. Regions from the same cluster form a *functional zone* similar functions, and different clusters represent different functions. Then, for each cluster of regions, we identify the intensity of functions in the regions (belonging to the cluster), using Kernel Density Estimation which employs human mobility as samples.
- We evaluated our method using large-scale, real-world datasets consisting of two POI datasets of Beijing (for years 2010 and 2011) and two three-month GPS trajectory datasets (representing human mobility) generated by over 12,000 taxicabs in Beijing for years 2010 and 2011, respectively. The results confirm the advantages of our approach over baseline methods which solely take into account either POI or human mobility aspects (but not both). In addition, these powerful datasets allow us to study not only the current functional zones in a city, but also the evolution of the city over time.

Discovery of Region Topics

In this section we first segment the urban area of a city into spatial units (here termed regions), with reference to the boundaries formed by the network of major roads. We then infer the distribution of functions in each region unit using a topic model-based method.

Map Segmentation

An urban major road network is generally comprised of highways and ring roads; while these roadways provide connectivity, they also partition the city into areas which we term regions. For example, as presented in Fig. 2, the red segments denote freeways and urban expressways in Beijing, and blue segments represent urban arterial roads. These three types of roads are coded as road levels, 1, and 2, respectively, in the road network database used in this analysis. This road network forms a concomitant segmentation of the urban area of Beijing. We term each segmented region a *formal region* from this point forward, following the definition proposed in Bednarz et al. (1994). A formal region is a spatial unit fulfilling social-economic functions, the need for which stems from the following traits of such regions. First, people live in formal regions and POIs are located in regions. Second, the need to travel between these formal regions-the origins and destinations of trips-is the root cause of human mobility.

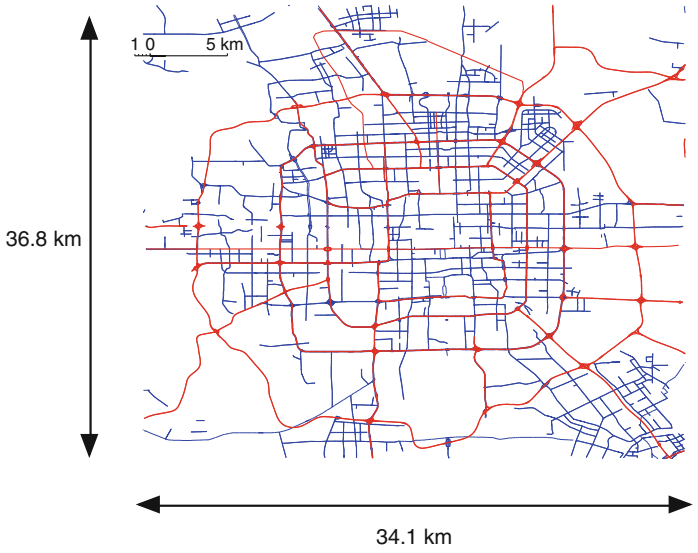

Fig. 2 Beijing road network. red: level-0 level-1 roads; blue: level-2 roads (Bing Map data: 2011)

In this analysis, a raster-based model is used to represent the road network and enable utilization of morphological image processing techniques in order to address the task of map segmentation. Typically, in a Geographic Information System (GIS), there are two models which can be used to represent spatial data: a *vector*-based model and a *raster*-based model. Vector-based models use geometric primitives such as points, lines and polygons to represent spatial objects referenced by Cartesian coordinates, whereas raster-based models quantize an area into discrete grid cells. Each of the two models has advantages and disadvantages, depending upon the specific applications. For instance, vector-based method is more powerful for precise identification of shortest paths; however, intensive computation is required when performing topological analysis, such as the problem of map simplification, which NP-complete (Estkowski 1998). On the other hand, raster-based modeling is more computationally efficient and succinct for territorial analysis, but its accuracy is limited by the number of cells used for discretizing the road networks.

Specifically, in this analysis, the raster-based map uses a binary image (with 2400×2400 grid cells) to delineate the road network (e.g., a grid cell may be coded either 0, which represents road segments, or 1, which represents non-road space). In order to remove the unnecessary details, such as the lanes of a road and the overpasses (see Fig. 3a), we first perform a *dilation* operation to thicken the roads. Through this operation, we can fill the small holes and smooth out unnecessary details, demonstrated in Fig. 3b. Second, we obtain a skeletal representation of the road network by performing a *thinning* operation based upon the algorithm proposed in Lam et al. (1992), as depicted in Fig. 3c. This operation recovers the full size of a region, which was reduced by the dilation operation, while maintaining the connectivity between regions. The last step is to perform connected component labeling (CCL) that finds individual regions by clustering grid cells labeled 1, using the method proposed in Shapiro and Stockman (2001). The result of this process is displayed in Fig. 3d.

After the connected component labeling, we obtain the segmented regions as well as their labels (termed as *region IDs*) for the entire study area. Here, all of the road segments form a region, the grid cells of which are labeled 0. Therefore, given a GPS point represented by latitude and longitude, we can determine in which region it is situated (by examining the region ID of the grid cell which the GPS point occupies). However, in this case, instead of mapping isolated points, we need to transform a whole trajectory into a region ID sequence. For example, trajectory $\mathscr{T} : O \rightarrow p_1 \rightarrow p_2 \cdots \rightarrow p_5 \rightarrow D$ in Fig. 4 is a trajectory originating in region R_1 and concluding in region R_5. Note that some points may be located along roadways (such as p_1) while some points are situated in the non-road regions (such as p_5). In order to obtain meaningful mapping of the raw trajectory – placing it within the context of the region IDs – for each point x of the trajectory, we apply the following rules:

- If x is the origin/destination point, we first retrieve its k nearest neighbour grid cells $N(x)$. Then we assign to x the most frequent label (region ID) in $N(x)$ which

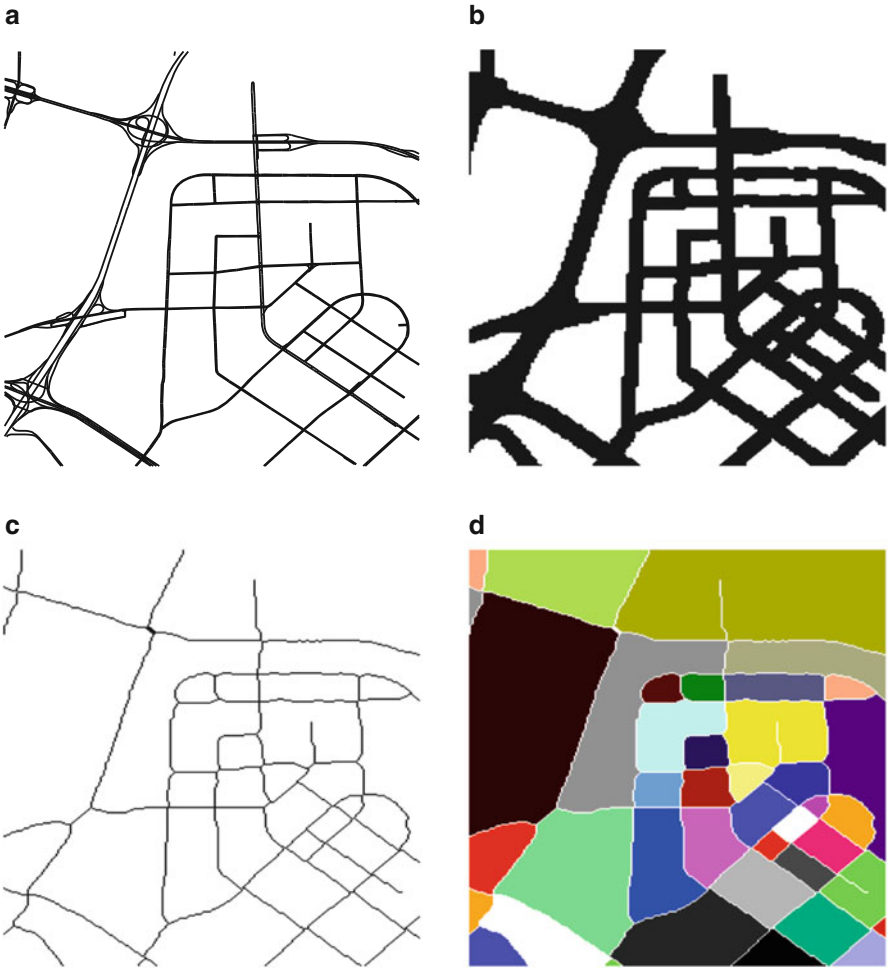

Fig. 3 Map segmentation process: (**a**) Original, (**b**) Step 1: Dilation, (**c**) Step 2: Thinning, (**d**) Step 3: CCL (connected component labeling) (Bing Map data: 2011)

is non-zero (not on road segments). As presented in Fig. 4b, d, respectively, the origin O is assigned to region R_1 and the destination D is assigned to region R_5.

- If x is not the origin/destination point, and if all its k nearest neighbour grid cells $N(x)$ have an identical label (region ID), we assign to x with the same regin ID, as depicted in Fig. 4c; otherwise, we label x with 0 (indicating that the point is along the road).

As a result, $\mathscr{T}$ is mapped to the region ID sequence $R_1 \rightarrow 0 \rightarrow 0 \rightarrow 0 \rightarrow 0 \rightarrow R_4 \rightarrow R_5$.

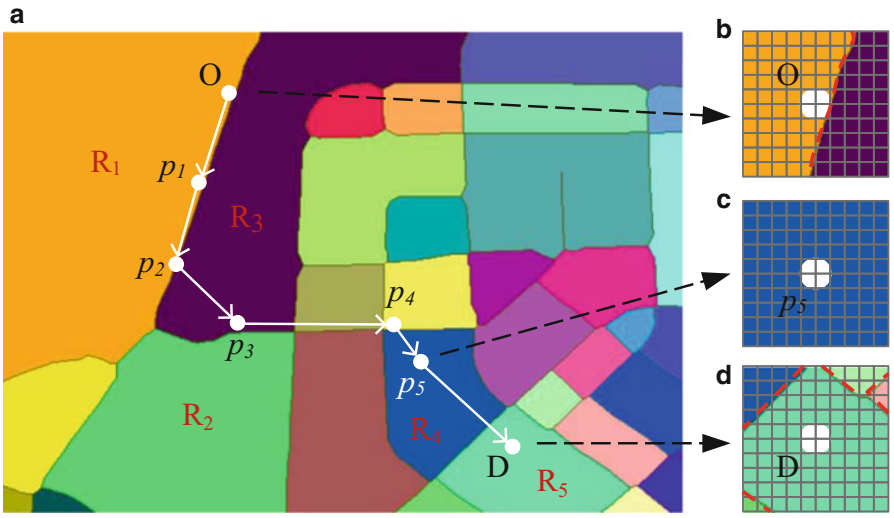

Fig. 4 Trajectory mapping example: (**a**) Overview of trajectory, (**b**) Origin point O assigned to region R_1, (**c**) Point p_5 assigned to region R_4, (**d**) Destination point D assigned to region R_5

Based upon the above approach, we map a large number of GPS trajectories of taxicabs during three months of both 2010 and 2011 (Yuan et al. 2012) between the segmented regions, and then extract the origin-destination (OD) pairs. Figure 5 presents the top 200 most frequent OD pairs for the time window between 8:00–10:00 a.m on an average of weekdays; darker colors indicate higher frequencies. Figure 6 further shows the number of transitions (OD pairs) changing over the rank (k) of OD pairs (log-log plot), which suggests a typical Zipf distribution.

Topic Discovery

Preliminary

Definition 1 (Transition) A *transition Tr* is a quaternion containing the following four items: origin region ($Tr.r_O$), leaving time ($Tr.t_L$), destination region ($Tr.r_D$) and arrival time ($Tr.t_A$). Here, $Tr.r_O$ and $Tr.r_D$ are spatial features while the others are temporal features.

Definition 2 (Mobility Pattern) A *mobility pattern M* is a triple extracted from a transition. Given a transiton $Tr = (Tr.r_O, Tr.r_D, Tr.t_L, Tr.t_A)$, we obtain two mobility patterns: the *leaving mobility pattern* $M_L = (Tr.r_O, Tr.r_D, Tr.t_L)$, and the *arriving mobility pattern* $M_A = (Tr.r_O, Tr.r_D, Tr.t_A)$.

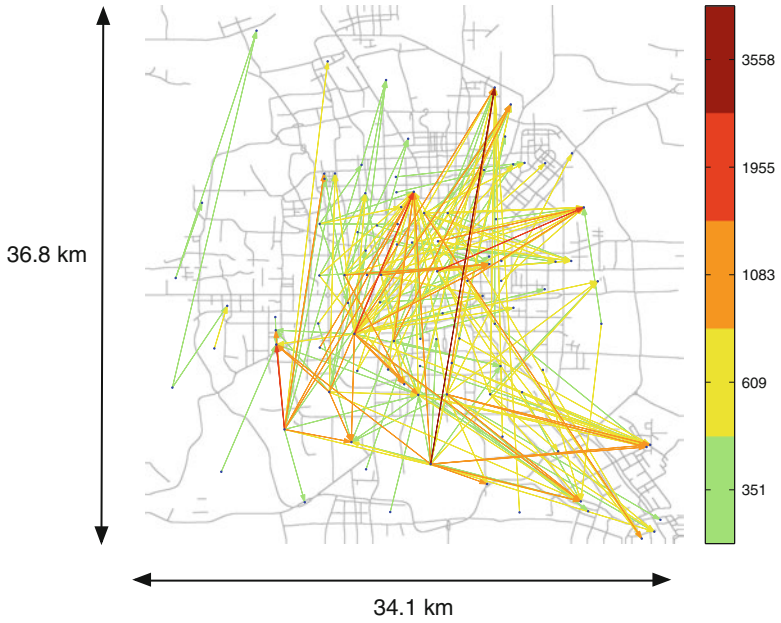

Fig. 5 Top 200 frequent OD pairs on an average of weekdays during a time window from 8:00 to 10:0 a.m

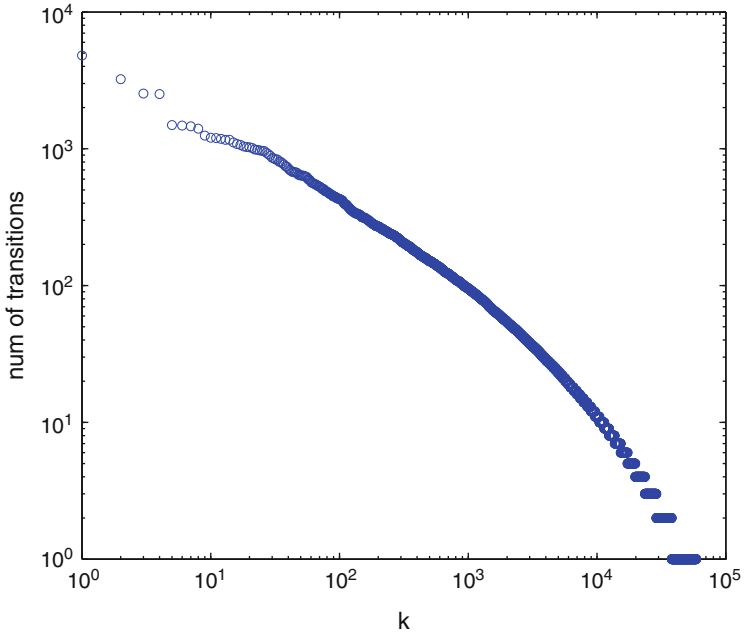

Fig. 6 Number of transitions with respect to k (rank of O-D pair)

Definition 3 (Transition Cuboids) A *transition cuboid C* is an $R \times R \times T$ cuboid, where R is the number of regions and T is the number of time bins. Since there exist two types of mobility patterns, we define two types of transition cuboids: *leaving cuboid* C_L and *arriving cuboid* C_A. The cell with index (i, j, q) of the leaving cuboid records the number of mobility patterns that leave region r_i for region r_j at time t_q, i.e.,

$$C_L(i, j, q) = \|\{M_L = (x, y, z) | x = r_i, y = r_j, z = t_q\}\|. \tag{1}$$

Similarly,

$$C_A(i, j, q) = \|\{M_A = (x, y, z) | x = r_i, y = r_j, z = t_q\}\|. \tag{2}$$

We project each trajectory representing human mobility on the segmented region units, turning a trajectory into a transition. Then, we discretize time of day into time bins, in each of which we deposit the transitions and formulate the window-specific mobility patterns. Here, we do not differentiate different weekdays but differ the time bins in terms of weekdays versus weekend days. For example, using two-hour time bins, we will have 24 bins (12 for weekdays and 12 for weekend days) in total. Later, two transition cuboids are built using the mobility patterns.

Concepts of Topic Models. Probabilistic topic models have been successfully used for extracting the hidden semantic structure in large archives of documents (Blei 2011). In this model, each *document* of a *corpus* (a collection of documents) exhibits multiple *topics*, and each *word* of a document supports a certain topic. Given all the words of each document in a corpus as observations, a topic model is trained to infer the hidden semantic structure behind the observations. Latent Dirichlet Allocation (LDA) is a generative model that includes hidden variables. The intuition behind this model is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over a particular set of words (Blei et al. 2003). Let α and η be the prior parameters for the Dirichlet document-topic distribution and topic-word distribution respectively. Assume there are K topics and β is a $K \times V$ matrix where V is the number of words in the vocabulary (all of the words in the corpus D). Each β_k is a distribution over the vocabulary. The topic proportions for the dth document are θ_d, where $\theta_{d,k}$ is the topic proportion for topic k in the dth document. The topic assignments for the dth document are z_d, where $z_{d,n}$ is the topic assignment for the nth word in the dth document. Finally, the observed words for document d are w_d, where $w_{d,n}$ is the nth word in document d, which is an element in the vocabulary of mobility patterns.

Using the above notations as presented in Fig. 7, the generative process can be described as follows:

1. For each topic k, draw $\beta_k \sim Dir(\eta)$.
2. Given the dth document d in corpus D, draw $\theta_d \sim Dir(\alpha)$.
3. For the nth word in the dth document $w_{d,n}$:

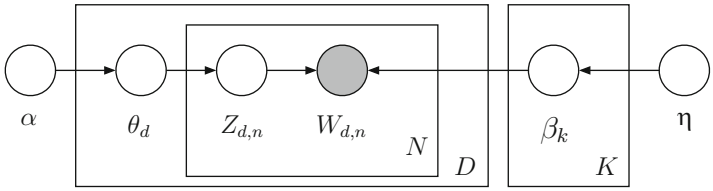

Fig. 7 Graphic model of LDA

Table 1 Analogy from region-functions to document-topics

Document − topicterms		Region − functionequivalents
transition cuboids	⟶	vocabulary
formal regions	⟶	documents
function of a region	⟶	topic of a document
mobility patterns	⟶	words
POI feature vector	⟶	metadata of a document

(3)

a. Draw $z_{d,n} \sim Mult(\theta_d)$;
b. Draw $w_{d,n} \sim Mult(\beta_{z_{d,n}})$.

Here, $Dir(\cdot)$ is the Dirichlet distribution and $Mult(\cdot)$ is the multinomial distribution. The estimation of LDA can be implemented using the Expectation Maximization (EM) algorithm; and the most commonly used inference method of LDA is Gibbs sampling. See (Blei et al. 2003) for a variational inference method and detailed discussions of LDA.

Topic Modeling

As shown in Table 1, we draw an analogy between discovering functions of a region and the topic discovery of a document. Specifically, we regard a formal region as a document and a function as a topic. In other words, a region with multiple functions is like a document containing a variety of topics. Meanwhile, we deem the mobility patterns associated with a region as words and POIs as metadata of a document.

Figure 8 further details the topic discovery analogy using an example. In this proposed methodology, given the mobility dataset, we build the arriving and leaving cuboids respectively according to Definition 3 (section "Conclusion"). For a specific region r_i, the mobility patterns associated with r_i are counted by $C_A(1{:}R, i, 1{:}T)$ and $C_L(i, 1{:}R, 1{:}T)$, which are two *slices* extracted from the arriving cuboid and the leaving cuboid (termed as the arriving matrix and the leaving matrix respectively). The right side of Fig. 8 presents a *document* that we compose for region r_1, in

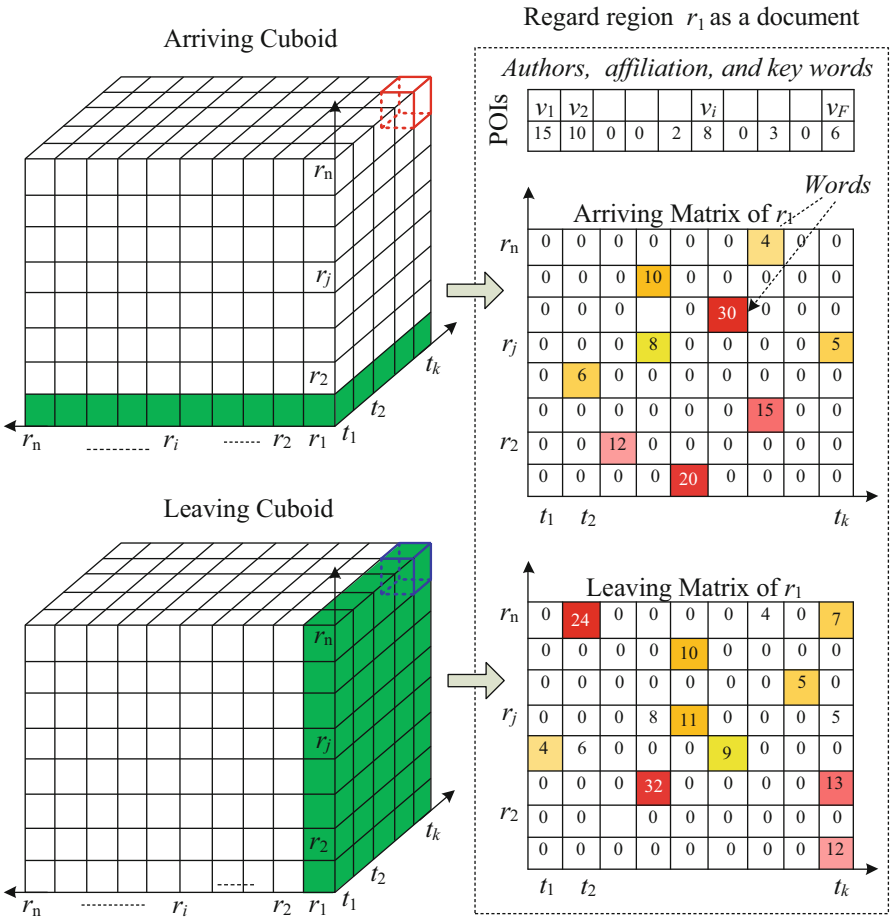

Fig. 8 Analogy between mobility model and topic model, based upon transition cuboids

which a particular cell (in the matrices) represents a specific mobility pattern and the numbers in the cell denote the frequency of occurrence of the pattern. For example, in the rightmost column of the arriving matrix, the cell containing the value 5 indicates that on average, the mobility pattern that arrived at r_1 after leaving from r_j in time bin t_k occurred 5 times per day. A POI is recorded in a POI database as a tuple consisting of a POI category (as listed in Table 2), name and a geo-coordinates (latitude, longitude). For each formal region r, the number of POIs in each POI category can be counted. The *frequency density* v_i of the ith POI category in r is calculated by:

$$v_i = \frac{\text{Number of the POIs of the } i\text{th POI category}}{\text{Total of the areas of the region } r \text{ (measured by grid cells)}}, \tag{4}$$

Table 2 POI category taxonomy

Code	POI category	Code	POI category
1	Car service	16	Banking and insurance service
2	Car sales	17	Corporate business
3	Car repair	18	Street furniture
4	Motorcycle service	19	Entrance/bridge
5	Café/Tea Bar	20	Public utilities
6	Sports/stationery shop	21	Chinese restaurant
7	Living service	22	Foreign restaurant
8	Sports	23	Fastfood restaurant
9	Hospital	24	Shopping mall
10	Hotel	25	Convenience store
11	Scenic spot	26	Electronic products store
12	Residence	27	Supermarket
13	Governmental agencies and public organizations	28	Furniture building materials market
14	Science and education	29	Pub/bar
15	Transportation facilities	30	Theaters

and the *POI feature vector* of r is denoted by $x_r = (v_1, v_2, \ldots, v_F, 1)$ where F is the number of POI categories and the last value 1 is a default feature to account for the mean value of each function, as explained in Mimno and McCallum (2008). The *POI feature vector* is regarded as the metadata of each region, which is an analogue of the observed features of a document such as author, email address, and institution.

Using the basic LDA model, region functions can be discovered using the mobility patterns. However, as stated in section "Introduction", the region functions are products of both the POIs and mobility patterns. In order to combine the information from both of these aspects, we utilize a more advanced topic model based upon LDA and Dirichlet Multinomial Regression (DMR) (Mimno and McCallum 2008).

The DMR-based topic model (simply termed DMR from this point forward) takes into account the influence of the observable metadata in a document by using a flexible framework, which supports arbitrary features (Mimno and McCallum 2008). Compared to other models designed for specific data such as Author-Topic models and Topic-Over-Time models (both members of the supervised-LDA family of topic models Blei 2011), DMR achieves similar or improved performance, while it is also more computationally efficient and succinct in implementation (Mimno and McCallum 2008).

As presented in Fig. 9, the generative process of the DMR model is as follows:

1. For each region function k,
   a. Draw $\lambda_k \sim \mathcal{N}(0, \sigma^2 I)$;
   b. Draw $\phi_k \sim Dir(\beta)$.

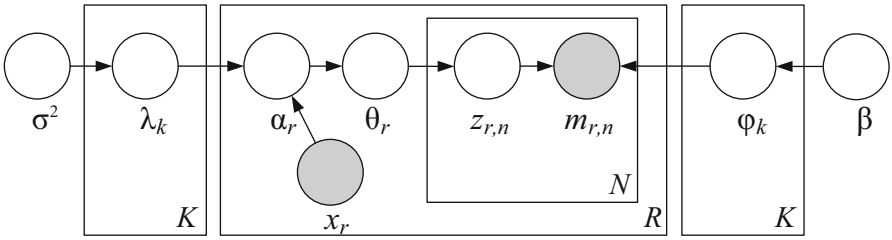

Fig. 9 DMR-based topic model

2. Given the *r*th region,

   a. For each region function k, let $\alpha_{r,k} = \exp(x_r^T \lambda_k)$;
   b. Draw $\theta_r \sim Dir(\alpha_r)$;
   c. For the *n*th mobility pattern in the *r*th region $m_{r,n}$,

      i. Draw $z_{r,n} \sim Mult(\theta_r)$;
      ii. Draw $m_{r,n} \sim Mult(\phi_{z_{r,n}})$.

Here, $\mathcal{N}$ is the Gaussian distribution with σ as a hyper parameter, and λ_k is a vector with the same length as the POI feature vector. The *n*th observed mobility pattern of region r is denoted as $m_{r,n}$. ϕ_k is the mobility pattern distribution of function k derived based upon hyper parameter β. Other notations are similar to the previous LDA model. This model can also be estimated using EM and inferred using Gibbs sampling. In a divergence from the basic LDA model, presented in Fig. 7, here, the Dirichlet prior α is now specified to individual regions (α_r) based upon the observed POI features of each region. Therefore, for different combination of POI category distributions, the resulting α values are distinct. Thus the semantic region function distributions extracted from the data are induced by both the POI features and mobility patterns. As a result, by applying DMR, given the mobility patterns and POI features, we obtain the topic assignment for each region and the mobility pattern distribution of each function.

Territory Identification

Region Aggregation

This step aggregates similar formal regions in terms of region function distributions through the use of a clustering algorithm. Regions in the same cluster have similar functions, and different clusters represent different functions. For region r, after parameter estimations based on the DMR model, the function distribution is a K dimensional vector $\theta_r = (\theta_{r,1}, \theta_{r,2}, \ldots, \theta_{r,K})$, where $\theta_{r,k}$ is the proportion of function k for region r. We perform the k-means clustering method on the K-dimensional

points $\theta_r,\ r \in 1, 2, \ldots, R$. The number of clusters can be predefined according to the needs of an application or determined using the average *silhouette* value as the criterion (Rousseeuw 1987). The silhouette value of a point i in the dataset, denoted by $s(i)$, is in the range of $[-1, 1]$, where an $s(i)$ approximating 1 indicates that the point is appropriately clustered and very distant from its neighboring clusters; an $s(i)$ approximating 0 indicates that the point is not distinctly clustered into one cluster or another; an $s(i)$ approximating –1 indicates that the point is probably assigned to the wrong cluster. The average silhouette value of a cluster measures how tightly the data in this cluster are grouped, and the average silhouette of the entire dataset reflects how appropriately all of the data has been clustered. In practice, we perform cross-validation on the dataset for different k, multiple times, and choose an appropriate k with the maximum overall silhouette value. Consequently, we aggregate the regions into k clusters, each of which is termed as a *functional zone*.

Estimation of Functional Intensity

On the one hand, the intensity of a functional zone is generally not uniformly distributed across the entire region. On the other hand, sometimes, the core functional area may span multiple formal regions and may have an irregular shape, e.g., a hot shopping street crossing several formal regions. In order to reveal the degree of intensity and glean the essential territory of a functional zone, we estimate the *functional intensity* for each aggregated functional zone (a cluster of formal regions).

Intuitively, the number of visits implicitly reflects the popularity of a certain functional zone. In other words, mobility patterns imply the functional intensity. Therefore, we enter the origin and destination of each mobility (represented by latitude and longitude) into a Kernel Density Estimation (KDE) model, in order to infer the functional intensity in a functional zone. Note that the actual place that an individual visited may not be the destination that we can obtain from a mobility dataset. For example, the drop-off points of taxi trajectories may not be the final destinations for these individual trips; they may not accord exactly with the final destination of interest, such as a shopping mall. However, it would not be expected that a taxi rider would pay the cost of a taxi – a sizable portion of which is minimum fare – and then exit a taxi at a point that would require the time cost of a long walk, in addition to taxi fare. Typically, the farther distance from an actual destination location to the drop-off point, the lower probability that people would visit the destination.

Given n points $x_1, x_2, \ldots, x_n$ located in a two-dimensional spatial space, KDE estimates the intensity at location s by a kernel density estimator, defined as:

$$\lambda(l) = \sum_{i=1}^{n} \frac{1}{nh^2} K\left(\frac{d_{i,l}}{h}\right), \tag{5}$$

where $d_{i,l}$ is the distance from x_i to l, h is the bandwidth and $K(\cdot)$ is the kernel function, the value of which decays as d_i, h increases, as in the Gaussian, Quartic, Conic and negative exponential function types. The choice of the bandwidth usually determines the smoothness of the estimated density – a large h achieves smoother estimation whereas a small h reveals more detailed peaks and valleys. In our case, we chose the Gaussian function as the kernel function, i.e.,

$$K\left(\frac{d_{i,l}}{h}\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{d_{i,l}^2}{2h^2}\right). \tag{6}$$

and the bandwidth h is determined according to MISE (mean integrated squared error) criterion (Wand and Jones 1995).

Region Annotation

In this step, given the results we obtained, we try to annotate each cluster of regions with some semantic terms, which can contribute to the understanding of the actual functions of the cluster. Region annotation is a very challenging problem in both traditional urban planning and document processing. Essentially, this annotation problem is the visualization problem of topic model, which is listed as a future direction of topic modeling in the recent survey paper by Blei (2011). A makeshift method currently employed is to use the most frequent words in a discovered topic in order to annotate a document. However, in the case of the current research, listing the most frequently observed mobility patterns (the analogue to words) is far from sufficient for naming a functional zone.

In the methodology of the current research, we annotate a functional zone by considering the following four aspects:

(1) The POI configuration in a functional zone. We calculate an average POI feature vector across the regions in a functional zone. According to the average frequency density in the calculated POI feature vector, we rank each POI category in a functional zone (termed internal ranking) and rank all functional zones for each POI category (termed external ranking). We provide an example, as presented in Fig. 12.
(2) The most frequently observed mobility patterns in each functional zone.
(3) The functional intensity. We analyze the representative POIs located in each functional kernel – e.g., a functional region could be considered an educational area if its kernel contains many universities and schools.
(4) Human-labeled regions. People may recognize the functions of a few well-known regions (e.g., the region that contains the Forbidden City is an area of historic interest). After the clustering analysis is performed, an investigation of the human labeling of regions help us to understand the other regions in a cluster. (Refer to the experiments detailed in section "Evaluation" for the detailed results and analysis.)

Evaluation

Settings

In the topic-model based analysis of functional zones based upon POIs and mobility patterns, the datasets listed in Table 3 were employed.

As noted in Table 3, the counts of POIs in each category are presented in Fig. 10, and the statistics for Beijing taxi trajectories in 2010 and 2011, as well as road networks, are presented in Table 3. Only occupied taxi trips were selected from the data for use in this analysis, as identified by the taxi meter information. Though there are more than 30 cities globally which can claim over 10,000 taxicabs, Beijing is notable among these cities, in that it has over 67,000 taxis. In Beijing, taxi trips represent a significant portion of individual urban mobility. According to the report of Beijing Transportation Bureau, the taxi trips occupy over 12 percent of traffic flows on road surfaces (Yuan et al. 2011b). Of course, other mobility datasets such as mobile phone traces can be used in our framework.

Table 3 Statistics of taxi trajectories and road networks

	Year	2010	2011
Trajectories	#taxis	12,726	13,597
	#occupied trips	21,678,203	8,202,012
	#days in dataset	112	92
	Average trip distance(km)	7.22	7.47
	Average trip duration(min)	15.98	16.1
	Average sampling interval(sec)	74.46	70.45
Roads	#road segments	150,357	162,246
	Percentage of major roads	18.9%	17.1%
	#segmented formal regions	565	554
	Size of "vocabulary" (non-0 items)	3,318,331	3,244,901

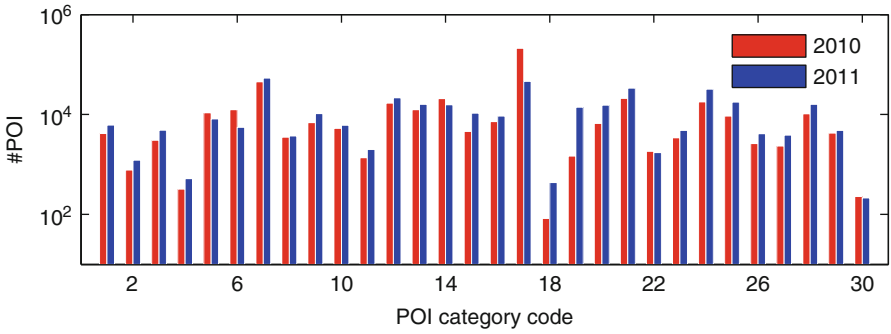

Fig. 10 POI counts for each POI category in 2010 and 2011 (Refer to Table 2 for POI category codes)

Table 4 Overall efficiency

Operation	Time(min)
Map segmentation	0.325
Building transition cuboids	40.1
Estimate topic model (1,000 iterations)	1353
Region aggregation	0.124
Total	1394

The methodology was implemented on a 64-bit server with a Quad-Core 2.67 G CPU and 16 GB RAM. The model was trained with ten topics for 1,000 iterations, optimizing the parameters every 50 iterations. For k-means clustering, the average silhouette value was incorporated in order to determine the k, and the average results based upon a 5-fold cross-validation were used. The average efficiency on averageis presented in Table 4.

We compare our method with two baselines:

- **TF-IDF-based Clustering**, which solely uses the POI data. This method employs the *term frequency-inverse document frequency* (TF-IDF) to measure the importance of a POI in a formal region. Specifically, for a given formal region, we formulate a POI vector, $\langle v_1, v_2, \ldots, v_F \rangle$ where v_i is the TF-IDF value of the i-th POI category, given by:

$$v_i = \frac{n_i}{N} \times \log \frac{R}{\|\{r|\text{the } i\text{-th POI category} \in r\}\|}, \tag{7}$$

  where n_i is the number of POIs belonging to the i-th category and N is the number of POIs located in this region. The idf term is calculated by computing the quotient of the number of regions divided by the number of regions which have the i-th POI category, then taking the logarithm of that quotient. Later, a k-means clustering is used to cluster the formal regions into k functional zones.
- **LDA-based Topic Model**, which uses only the mobility data. Similar to our analogy of regions as documents, this method feeds the mobility patterns (the analogue to words) into an LDA model. Later, a k-means clustering is performed, similar to the method used for grouping all of the formal regions based upon their function distributions learned from LDA. The parameters, such as number of iterations, number of topics are set identically as in our approach.

We used the following three validation methods to evaluate the effectiveness of our framework and methodology, though this proved challenging. (1) We invited some local people (who had been in Beijing for over six years) and requested that they label two representative regions for each kind of function. We then checked whether the regions which were given the same labels had been assigned to the same functional zone, and whether the regions with different labels had been improperly clustered into one functional zone. (2) We investigated the evolution of Beijing by comparing the results from 2010 to 2011, then assessed whether the differences made sense. (3) We compared our results with land use plans for Beijing.

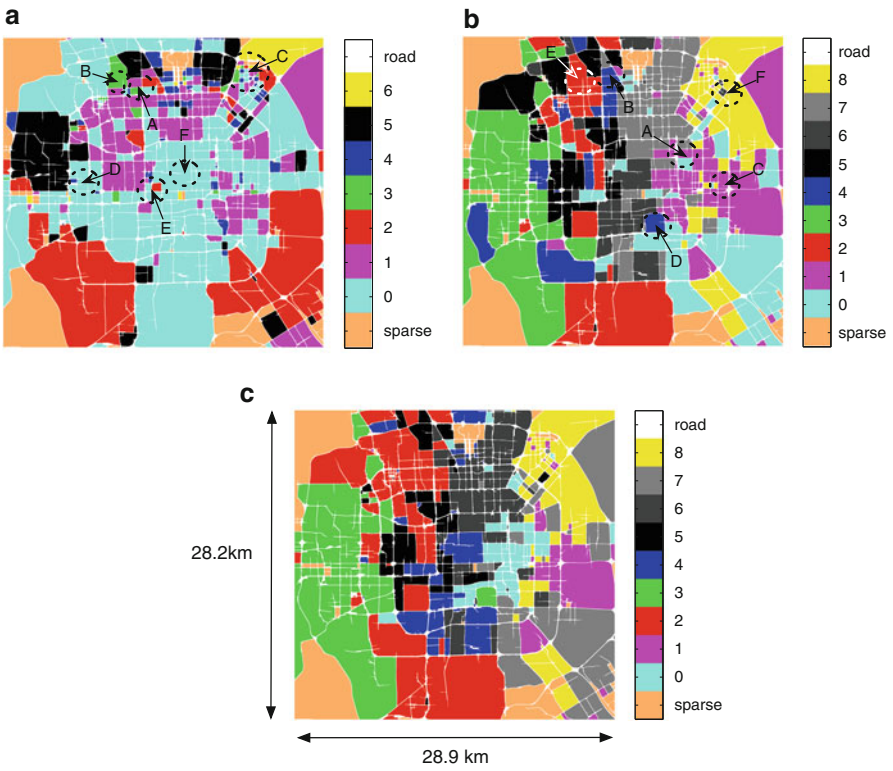

Fig. 11 Functional zones discovered by different methods: (**a**) TF-IDF, (**b**) LDA, (**c**) DMR

Results

Discovered Functional Zones

Figure 11 displays the aggregated functional zones discovered by different methods, in which different colors indicate different functions. Note that in different figures, the same color may represent different functions. As demonstrated, TF-IDF-based method forms seven clusters while LDA-based method and DMR each form 9 functional zones.

Note that in different figures, the same color may stand for different functions. For example, as shown in Fig. 11a, Region *B* is an university, which should be clustered with Region *A* (another university) and Region *D* (a high school). Meanwhile, region *F* (the Forbidden City) is not distinguished from other commercial areas like region *E* (Xidan). Another example is the Wangjing area (*C*), which is an emerging residential area with some business buildings and tertiary sector businesses, such as apartment complexes, Unfortunately, the TF-IDF method improperly divides this area into many small functional zones as this method only considers POI distributions.

Basically, the LDA-based and DMR-based methods produced similar results in terms of functional zones. However, there were several exemplary regions in which the DMR-based method clearly outperformed LDA-based method obviously. For example, region *F* in Fig. 11b is a developing commercial/entertainment region in Wangjing region. But LDA aggregates it with the Forbidden City (Region *F* in Fig. 11a), which is a region of historical interests; Region *B* (China Agricultural University) and Region *E* (Tsinghua University) are typical science and education areas where LDA fails to correctly cluster them together; Region *A* around Sanlitun is a well-known diplomatic district of Beijing, which is mixed with a developing commercial Region *C*; Region *D* is a park which LDA fails to cluster it with other park/mountain areas (the green regions displayed in Fig. 11b). The LDA-based method, only using human mobility, overlooks the feature of POIs, thereby failing behind the DMR-based method (displayed in Fig. 11c) in terms of clustering performance.

Region Annotation

We calculate an average POI feature vector across the regions in a functional zone. According to the average frequency density in the calculated POI feature vector, we rank each POI category in a functional zone (termed internal ranking) and rank all functional zones for each POI category (termed external ranking).

Figure 12 shows the average POI feature vector of each region cluster (c_0–c_8, recalling that our method generated 9 clusters) and the corresponding internal and external rankings, where the external rank is represented by the depth of the color (a rank of 1 indicated by the darkest shade, and a rank of 4 indicated by the lightest shade). It appears that region clusters c_0, c_2, c_5, c_6, c_8 are more mature and more developed areas as compared to other clusters, as they feature more high ranked POI categories, which are annotated as follows:

Diplomatic and embassy areas (c_0). The most commonly occurring POI category in this functional zone was the governmental agencies and public organizations, with a significantly higher frequency density than other functional zones. Actually, most of the embassies are located within these regions, which are well-configured/appointed for the diplomatic function-e.g., they have the highest external ranking of Pubs/Bars and transportation facilities, and the second highest rankings of residential buildings, hospitals, and hotels, among all of the clusters.

Education and science areas (c_2). This region cluster contains the most science and education POIs (e.g., Tsinghua university and Beijing university). In addition, a famous high-technology industry zone in Beijing named *ZhongguanCun*, known as the Silicon Valley of China, is located in this functional zone.

Developed residential areas (c_6). This region cluster is clearly a mature residential area with the most residential buildings, service-based businesses, hospitals, hotels,

POI	c0		c1		c2		c3		c4		c5		c6		c7		c8	
	FD	IR	FD	IR	FD	IR	FD	IR	FD	IR	FD	IR	FD	IR	FD	IR	FD	IR
CarServ	0.046	25	0.016	23	0.052	26	0.044	18	0.060	17	0.028	25	0.056	24	0.091	13	0.053	21
CarSale	0.009	28	0.005	27	0.061	24	0.006	27	0.009	27	0.005	28	0.021	27	0.015	26	0.006	27
CarRepa	0.021	26	0.011	24	0.062	23	0.042	19	0.051	20	0.023	27	0.062	23	0.057	18	0.039	25
MotServ	0.002	30	0.003	28	0.004	28	0.001	28	0.002	29	0.004	29	0.001	29	0.001	29	0.003	28
Caf/Tea	0.226	14	0.121	9	0.226	12	0.066	15	0.113	13	0.252	6	0.237	13	0.052	19	0.153	10
StaStor	0.135	17	0.037	20	0.127	17	0.037	20	0.058	18	0.080	19	0.100	19	0.073	15	0.072	17
LivServ	1.289	1	0.581	2	1.322	2	0.399	1	0.698	1	0.780	2	1.345	2	0.430	2	0.886	2
Sports	0.054	23	0.035	21	0.092	21	0.030	22	0.041	22	0.033	23	0.080	20	0.035	20	0.093	16
Hospital	0.244	13	0.088	13	0.222	13	0.069	14	0.137	12	0.144	15	0.246	12	0.070	16	0.194	8
Hotel	0.202	15	0.063	16	0.115	18	0.058	16	0.071	16	0.086	18	0.211	15	0.059	17	0.049	22
SceSpo	0.048	24	0.007	26	0.032	27	0.012	25	0.016	25	0.029	24	0.044	25	0.012	27	0.031	26
Residen	0.795	3	0.230	5	0.638	6	0.203	5	0.323	5	0.398	5	0.797	4	0.221	4	0.440	3
Gov/Pub	0.442	7	0.103	11	0.276	11	0.094	10	0.188	9	0.169	12	0.375	7	0.177	6	0.150	11
Sci/Edu	0.315	11	0.139	7	1.084	3	0.109	9	0.323	6	0.251	8	0.530	6	0.124	9	0.266	6
TrasFac	0.459	6	0.115	10	0.397	7	0.091	11	0.150	11	0.191	11	0.364	8	0.113	10	0.257	7
Bank/Fina	0.376	9	0.128	8	0.383	8	0.078	13	0.107	14	0.197	10	0.320	10	0.083	14	0.135	12
CopBusi	1.128	2	0.593	1	1.947	1	0.334	2	0.348	4	0.548	4	1.738	1	0.475	1	0.977	1
StrFur	0.002	29	0.000	30	0.001	30	0.001	30	0.000	30	0.001	30	0.000	30	0.001	30	0.000	30
Entr/Bri	0.296	12	0.065	14	0.210	14	0.081	12	0.160	10	0.160	14	0.228	14	0.133	7	0.097	15
PubUti	0.405	8	0.101	12	0.285	9	0.112	8	0.238	7	0.209	9	0.314	11	0.132	8	0.132	13
ChiRes	0.692	5	0.252	4	0.926	5	0.294	3	0.399	3	0.813	1	0.829	3	0.235	3	0.370	4
ForRes	0.098	18	0.050	17	0.054	25	0.010	26	0.009	26	0.163	13	0.063	21	0.018	25	0.101	14
FasRes	0.095	19	0.046	18	0.141	16	0.034	21	0.050	21	0.126	16	0.132	17	0.026	22	0.057	20
ShopMal	0.724	4	0.268	3	0.929	4	0.242	4	0.476	2	0.559	3	0.734	5	0.203	5	0.306	5
ConvStor	0.370	10	0.157	6	0.281	10	0.128	7	0.234	8	0.251	7	0.362	9	0.108	11	0.160	9
EOStor	0.056	21	0.017	22	0.107	20	0.029	23	0.037	23	0.037	22	0.063	22	0.018	24	0.040	23
SupMar	0.055	22	0.008	25	0.065	22	0.020	24	0.025	24	0.042	21	0.040	26	0.021	23	0.040	24
FurBuil	0.086	20	0.065	15	0.151	15	0.192	6	0.093	15	0.088	17	0.142	16	0.099	12	0.064	19
Pub/Bar	0.179	16	0.043	19	0.114	19	0.044	17	0.053	19	0.060	20	0.120	18	0.031	21	0.071	18
Theater	0.011	27	0.001	29	0.002	29	0.001	29	0.006	28	0.025	26	0.007	28	0.002	28	0.002	29

Fig. 12 Overall POI feature vector ranking of functional zones by DMR method (FD: frequency density; IR: internal ranking; external rank indicated by depth of color: rank of 1 indicated by the darkest shade, and rank of 4 indicated by the lightest shade)

etc. In this kind of areas, tertiary sector establishments cater to needs of individuals' daily lives, such as the restaurants, shopping malls, banking services, schools, sports centers.

Emerging residential areas (c_8). This area is named the emerging residential area, as it has a balanced POI configuration, in terms of service-based businesses, residential buildings, sports centers, hospitals and some corporate centers, etc.

Figure 13 compares the arriving/leaving transitions of (region clusters) c_6 and c_8 with that of other clusters. The x-axes are time of day (by hour) and y-axes are the functional zones from which transition flows arrive to clusters c_6 and c_8 (left subfigures) and for which transition flows leave clusters c_6 and c_8 (right subfigures). Both c_6 and c_8 follow the trend that most *leaving* transitions in the morning (around 8:00–9:00 a.m., for the morning commute to work) and most arriving transitions in the early evening (around 5:00–6:00 p.m., for the evening commute home), which is a typical pattern for a residential area. However, in terms of the absolute quantity, c_8 is much lower than c_6, which shows that there are more people living in c_6.

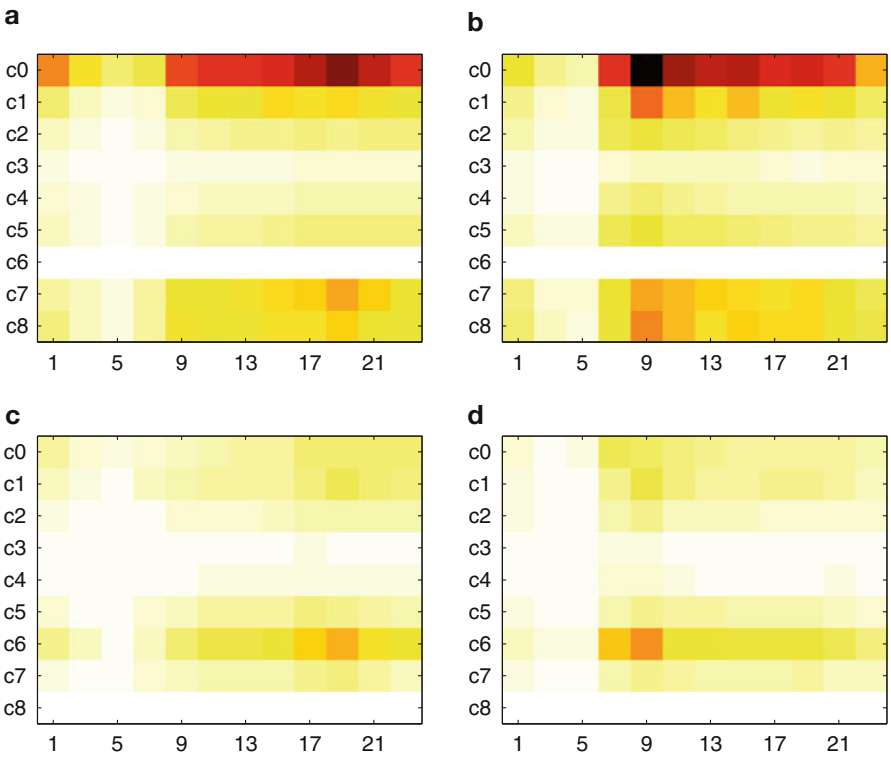

Fig. 13 Arriving and leaving transition flows by time of day and origin/destination cluster: (**a**) Transition flows arriving to c_6, (**b**) Transition flows leaving from c_6, (**c**) Transition flows arriving to c_8, (**d**) Transition flows leaving from c_8

Developed commercial/entertainment areas (c_5). The cluster comprises typical entertainment areas, featuring the highest external rank for theaters, foreign restaurants and café/tea bars, along with many shopping malls, Chinese restaurants and convenience stores. Figure 14 demonstrates the difference between arriving transitions c_5 on weekdays versus weekends. It is clear that people reach this kind of commercial/entertainment areas (c_5) from the residential regions (c_6) much earlier on weekends (about 9:00–11:00 a.m.) than on weekdays (7:00–9:00 p.m.).

With regard to region clusters other than c_5, c_6, and c_8, since the frequency densities of POIs are much lower for that of the above functional zones, we identify their semantic functions with more consideration given to the functional intensity and frequent mobility patterns derived for each functional zone in addition to the POI configurations.

Developing commercial/business/entertainment areas (c_1). The POI configuration of this cluster is similar to cluster c_5 and c_7, but in terms of the absolute quantity of POIs, c_1 has fewer than c_5, but has more than c_7. A certain number

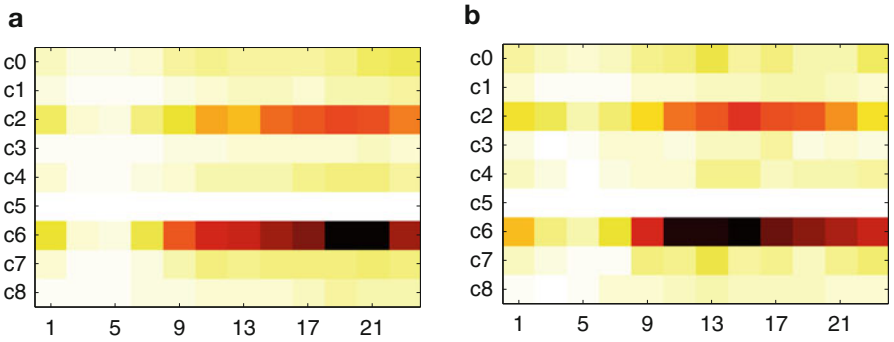

Fig. 14 Arriving transition flows by time of day and origin/destination cluster – weekdays versus weekend days: (**a**) Transition flows arriving to c_5, weekdays, (**b**) Transition flows arriving to c_5, weekend days

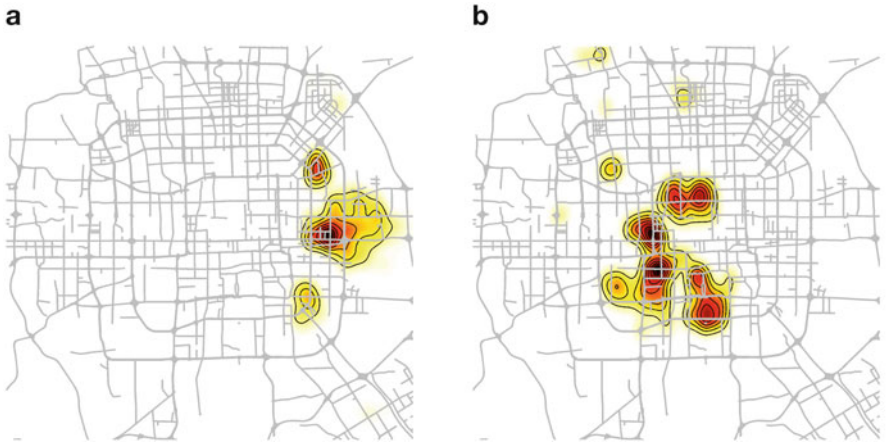

Fig. 15 Functional intensity of: (**a**) Functional zone c_1, (**b**) Functional zone c_4

of shopping malls, restaurants and banking services position this cluster as a developing commercial/ business/ entertainment functional zone (as any of these characterizations is possible/accurate). In the meantime, the functional intensity provides further corroboration for this characterization. As depicted in Fig. 15a, the core of this functional zone is the new CBD (central business district) of Beijing.[1]

Regions under construction (c_7). The region may potentially become like c_1 or c_8 since the POI configuration produces a rudiment of the commercial/residential area with related supporting services. Figure 16 validates the degree of development of this cluster with respect to c_5, c_1.

[1] http://en.wikipedia.org/wiki/Beijing

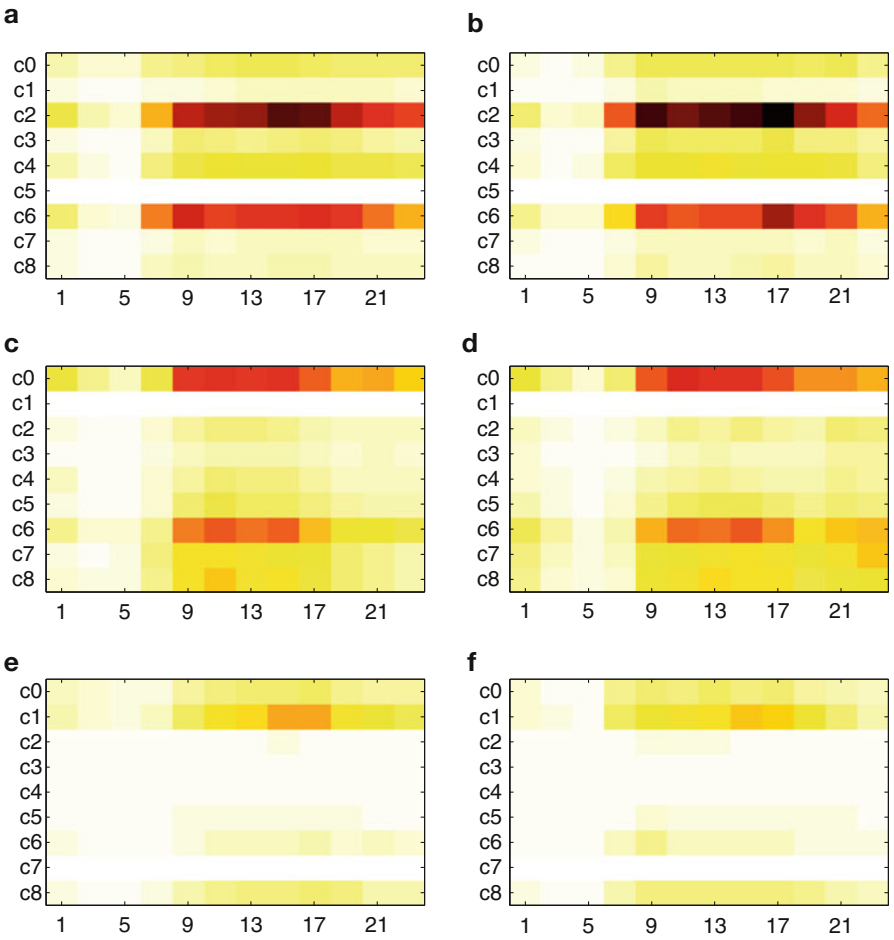

Fig. 16 Arriving and leaving transition flows by time of day and origin/destination cluster: (**a**) Transition flows arriving to c_5, (**b**) Transition flows leaving from c_5, (**c**) Transition flows arriving to c_1, (**d**) Transition flows leaving from c_1, (**e**) Transition flows arriving to c_7, (**f**) Transition flows leaving from c_7

Areas of historic interest (c_4). If we only consider the POI configuration- which contains some public utilities, entrances/bridges, government organizations, science and education areas – the character of this cluster is not immediately apparent. However, by considering the functional intensity estimated from the mobility patterns, it is readily apparent that these are areas of historic interests in Beijing. As presented in Fig. 15b, the famous historical places such as Forbidden City and the Temple of Heaven are located in these areas.

Nature and parks (c_3). These areas have the fewest POIs in most of the POI categories. Many forests and mountains comprise this cluster, e.g., the Xishan Forest

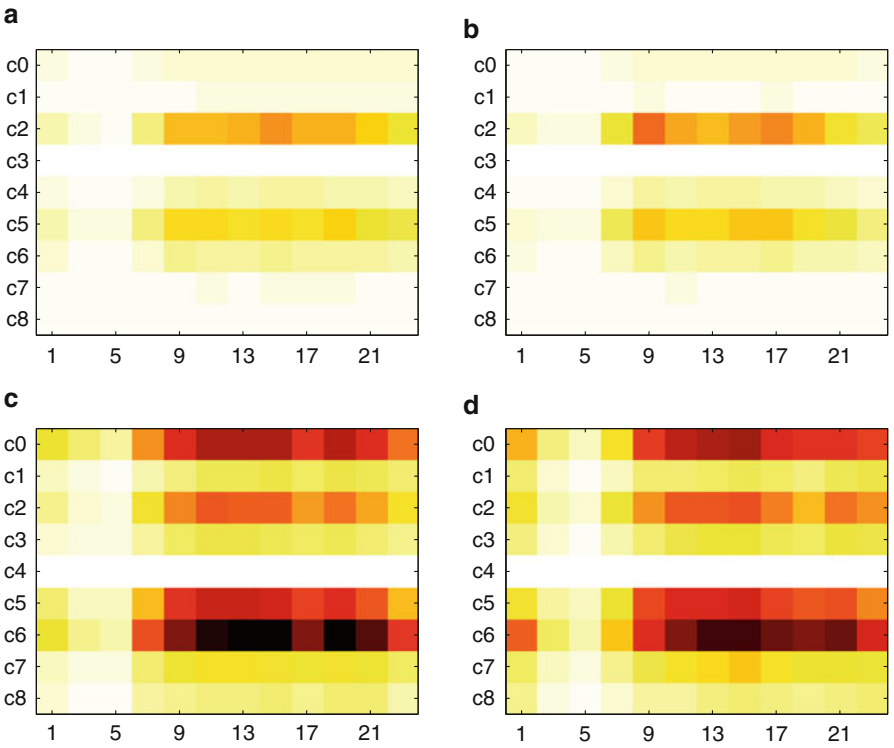

Fig. 17 Arriving and leaving transition flows by time of day and origin/destination cluster: (**a**) Transition flows arriving to c_3, (**b**) Transition flows leaving from c_3, (**c**) Transition flows arriving to c_4, (**d**) Transition flows leaving from c_4

Park, Century Forest Park, Baiwang Moutain, etc. Figure 17 demonstrates that transition flows to this functional zone followed similar temporal patterns with c_4 (the historical areas), but the quantity of transitions was observed to be some what lower than that of c_4, since many POIs in c_4 are very famous scenic spots.

Results in Different Years

We applied our method to the data (road network, POI and mobility pattern for 2010 and 2011 respectively). The functional zones discovered for 2010 are similar to those discovered for 2011, with slight differences in some regions. Figure 18 presents a detailed comparison of the east areas of the Forbidden City. For example, region *A* (Qianmen Street) became a developed commercial area in 2011 from a nature/park area. This region was developing in 2010 after a major repair in 2009. Similar to Region *A*, Region *B* (close to the new CBD of Beijing) became a developing commercial area in 2011, from an under construction area in

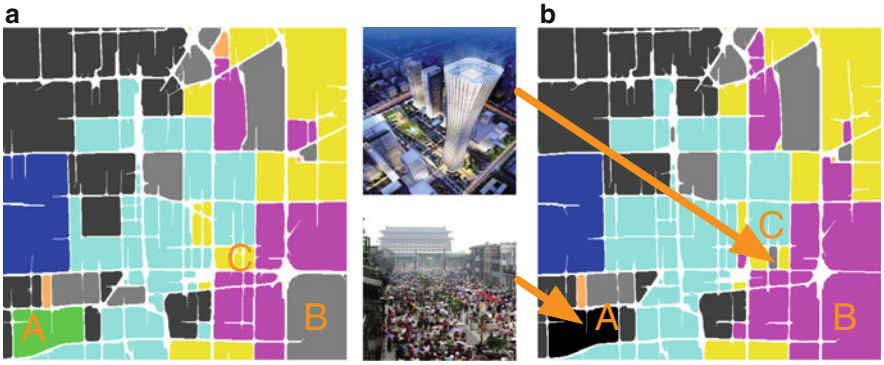

Fig. 18 The east area of Forbidden City: (**a**) in 2010, (**b**) in 2011

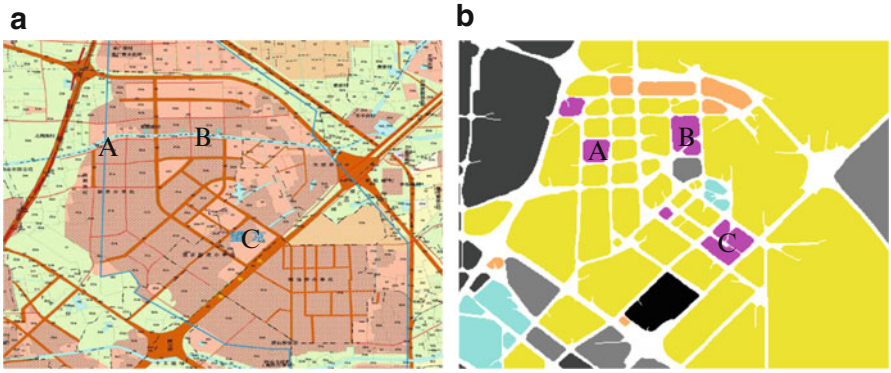

Fig. 19 (**a**) Governmental land use plan (2002–2010), (**b**) Functional zones identified for 2011

2010. Intriguingly, Region *C* became an under construction area in 2011, from an developing commercial area in 2010! We have discovered that the tallest building in Beijing will be constructed in this region in 2015, which explains the observed relocation work in 2011.

Calibration for Urban Planning

The functional zones identified by this analysis provide a means of calibration and reference for urban planning. For example, Fig. 19 presents the comparison between the governmental land use plan (2002–2010) and the results of our method in 2011. The entire study area forms an emerging residential area under the government plan, while some small regions become developing commercial areas-such as Regions *A*, *B* and *C*-after 2 years' development.

Related Work

Urban Computing with Taxicabs

In recent decades, urban computing has emerged as a concept in which every sensor, device, person, vehicle, building, and street in urban areas can be used as a component to probe urban dynamics and further enable city-wide computing to serve individuals and their cities. The increasing availability of GPS-embedded taxicabs provides us with an unprecedented wealth of data to assist in understanding human mobility in a city, thereby enabling a variety of novel urban computing research as of late. For example, Ge et al. (2010, 2011) and Yuan et al. (2011b) respectively studied strategies for improving taxi drivers' income by analyzing the pick-up and drop-off behavior of taxicabs in different locations. Yuan et al. (2011a) aimed to find the fastest driving route to a destination according to a large number of taxi trajectories, and Zheng et al. (2011) gleaned the problems with urban planning in a city from a set of taxi trajectories. Based upon the traffic flow represented by taxi trajectories, the technology for detecting traffic anomalies in urban areas has been reported in Liu et al. (2011).

The work presented in this chapter is also a step towards urban computing. Our work here differs from the above-mentioned research, however, we focus on the discovery of functional zones in a city, which represents a unique conceptual framework and use of taxicab trajectory data.

Discovery of Functional Zones

Functional zones (Antikainen 2005) have been studied in traditional fields of GIS and urban planning for years, as their identification can benefit policy making, resource allocation, and related research. As early as 1970, Goddard (1970) provided a case study on functional regions within the Central London area, using surveyed data of taxi flows collected in 1962, which is part of the London Traffic Survey. Recently, a number of data-driven approaches for identifying functional zones have been suggested; most of these are based upon clustering of commuting patterns (OD pairs) between zones (Farmer 2009). Some algorithms classify regions in an urban area based upon remote-sensor data, as thoroughly investigated in Vatsavai et al. (2011). Other network-based clustering algorithms (e.g., spectral clustering), however, employed interaction data, such as economic transactions, communication records (Ratti et al. 2010) and human movement between regions. In Ratti et al. (2010), the authors exploited telecommunication data to partition the Great Britain into regions.

Karlsson and Olsson (2006) summarized existing methods of identifying functional zones into three categories: (1) the local labor market approach, which is based upon the comparison of commuting frequency between various urban

cores (Sweden 1991); (2) the commuting zone approach (Killian et al. 1993), which is built from the existing mutual dependency of municipalities and measures the strength of the two-way commuting ties between zones; and (3) the accessibility approach (Karlsson and Olsson 2006), which utilizes two kinds of accessibility: employers' access to workers and workers' access to jobs. Our method is motivated by both the local labor market approach and the commuting zone approach. By combining the commuting data (human mobility) and the existing land use (POI allocation), we tackle the problem of identifying urban functional zones from a data mining perspective.

As the capital of China, Beijing has experienced profound changes, especially during the past two decades. Gaubatz (1995) presented a historical review of urban planning in Beijing, with a focus on the period from 1979 to 1995, during which the author indicates that new urban-planning ideas, complex land use and transportation patterns are integrated into the evolving city form. The work reported in this chapter, however, focuses on contemporary Beijing and potentially enables calibration for urban planners in the near future.

Discussion

We note that although the use of the analogue of topics of a document to represent functions of a region enables an effective data-driven approach to study urban functional zones from a data mining perspective, there still exist several limitations in the proposed model. For example, by considering the regions as *documents*, we do not explicitly model the topology of the connections between regions. Although this topology is implicitly represented by the arriving and leaving matrix of each region, the lack of explicit representation of the topological connections may reduce the information available regarding spatial dependency among the functional regions. Second, the temporal dependency between arriving and leaving mobility patterns are only partially modeled. Recall the definition of mobility pattern (Definition 2), which only involves one time stamp (arriving time or leaving time) instead of both, thus the actual duration of a transition from an origin to a destination is in directly leveraged in this model. However, it should be noted that the travel time between origin and destination regions is in fact implicitly represented by the distance between these two regions, an aspect which is already encoded as *words* of a document. The reason that we do not incorporate another temporal feature to the mobility patterns is that the size of the vocabulary would be further increased, causing a subsequent increase in the data sparsity problem.

Therefore, we suggest that the proposed method should be performed in concert with existing methodologies in urban planning and regional science, so as to delineate more comprehensive understanding of urban functional zones.

Conclusion

This chapter proposes a framework for discovering regions of various functions (e.g., educational areas, entertainment areas, and regions of historic interests) in a city, using both human mobility and POIs. The identification of functional zones assists in providing a readily understandable guide to a complex metropolitan area, which can benefit through a variety of applications, such as urban planning, business site selection, advertising placement, and tourism recommendations. We evaluated this framework through the use of a two-year Beijing POI dataset (2010 and 2011) and GPS trajectory datasets generated by over 12,000 taxis in 2010 and 2011. According to the comparative analysis, the proposed DMR-based method outperforms the two baseline methods (which solely used either POIs or mobility data) for effectively identifying functional zones. We also compared the results of our method in 2010 with that of 2011, discovering the evolution of Beijing. In addition, by comparing the identified functional zones to the Beijing land use planning (2002–2010), we not only validate our framework, but also find interesting results.

In the future, we will further study the effectiveness of our method by changing the scale of the data that we use. At the same time, we will add other mobility data sources, such as cell-tower tracing and check-ins for location-based services.

References

Antikainen, J. (2005). The concept of functional urban area. Findings of the Espon project 1(1).

Bednarz, S. W., et al. (1994). *Geography for life: National geography standards*. ERIC. https://eric.ed.gov/?id=ED375073

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84.

Blei, D., Ng, A., & Jordan, M. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research, 3*, 993–1022.

Cörvers, F., Hensen, M., & Bongaerts, D. (2009). Delimitation and coherence of functional and administrative regions. *Regional Studies, 43*(1), 19–31.

Estkowski, R. (1998). No Steiner point subdivision simplification is NP-complete. In *Proceedings of 10th Canadian Conference on Computational Geometry*, Citeseer.

Farmer, C. J. (2009). Data driven functional regions. In *Proceedings of 10th International Conference on GeoComp*.

Flórez-Revuelta, F., Casado-Díaz, J. M., & Martínez-Bernabeu, L. (2008). An evolutionary approach to the delineation of functional areas based on travel-to-work flows. *International Journal of Automation and Computing, 5*(1), 10–21.

Gaubatz, P. (1995). Changing Beijing. *Geographical Review, 85*(1), 79–96.

Ge, Y., Xiong, H., Tuzhilin, A., Xiao, K., Gruteser, M., & Pazzani, M. (2010). An energy-efficient mobile recommender system. In *Proceedings of KDD'10* (pp. 899–908).

Ge, Y., Liu, C., Xiong, H., & Chen, J. (2011). A taxi business intelligence system. In *Proceedings of KDD'11* (pp. 735–738).

Goddard, J. B. (1970). Functional regions within the city centre: A study by factor analysis of taxi flows in central London. *Transactions of the Institute of British Geographers*, 161–182.

Karlsson, C. (2007). Clusters, functional regions and cluster policies. JIBS and CESIS Electronic Working Paper Series (84).

Karlsson, C., & Olsson, M. (2006). The identification of functional regions: Theory, methods, and applications. *The Annals of Regional Science, 40*(1), 1–18.

Killian, M. S., Tolbert, C. M., Singelmann, J., & Desaran, F. (1993). Mapping social and economic space: The delineation of local labor markets in the united states. *Inequalities in labor market areas* (pp. 69–79). Boulder: Westview Press.

Lam, L., Lee, S., & Suen, C. (1992). Thinning methodologies-a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 14*(9), 869–885.

Liu, W., Zheng, Y., Chawla, S., Yuan, J., & Xing, X. (2011). Discovering spatio-temporal causal interactions in traffic data streams. In *Proceedings of KDD'11* (pp. 1010–1018).

Mimno, D., & McCallum, A. (2008). Topic models conditioned on arbitrary features with dirichlet-multinomial regression. In *Uncertainty in Artificial Intelligence* (pp. 411–418).

Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., Claxton, R., & Strogatz, S. H. (2010). Redrawing the map of great Britain from a network of human interactions. *PLoS One, 5*(12), e14,248.

Rousseeuw, P. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65.

Shapiro, L., & Stockman, G. (2001). *Computer vision*. Upper Saddle River: Prentice Hall.

Sweden, S. (1991). Local labour markets and employment regions: A new regional division of labour market areas in Sweden grounded on commuting statistics. *Labour Information 7*.

Vatsavai, R.R., Bright, E., Varun, C., Budhendra, B., Cheriyadat, A., & Grasser, J. (2011). Machine learning approaches for high-resolution urban land cover classification: A comparative study. In *Proceedings of COM.Geo'11* (pp. 11:1–11:10).

Wand, M., & Jones, M. (1995). *Kernel smoothing* (Vol. 60). London/New York: Chapman & Hall/CRC.

Yuan, J., Zheng, Y., Xie, X., & Sun, G. (2011a). Driving with knowledge from the physical world. In *Proceedings of KDD'11* (pp. 316–324).

Yuan, J., Zheng, Y., Zhang, L., Xie, X., & Sun, G. (2011b). Where to find my next passenger. In *Proceedings of Ubicomp'11* (pp. 109–118).

Yuan, J., Zheng, Y., & Xie, X. (2012). Discovering regions of different functions in a city using human mobility and pois. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM.

Zheng, Y., & Zhou, X. (2011). *Computing with spatial trajectories*. New York: Springer.

Zheng, Y., Liu, Y., Yuan, J., & Xie, X. (2011). Urban computing with taxicabs. In *Proceedings of Ubicomp'11* (pp. 89–98).

Is Your City Economic, Cultural, or Political? Recognition of City Image Based on Multidimensional Scaling of Quantified Web Pages

Jae Soen Son and Jean-Claude Thill

Abstract This research extends city image research to the communication space afforded by the World Wide Web through the efficient use of web-based information and the analysis of the thematic dimensionality of city images embedded in the discourse conveyed on this space. A large-scale database is built from web crawling results for each of 264 cities recognized to be significant in the world; textual contents are extracted from the crawled web pages. The thematic dimensionality of the textual contents denotes how they are expressed, discussed, and shared on the Web; it is measured by quantified content analysis and multidimensional scaling based on the relative preponderance of the economic, cultural and political themes that cut across these contents. The web-based discourse on cities is found to be highly structured and to be responding to the geographies, histories, and socio-political context of each city. The analysis demonstrates the merit of the new interdisciplinary approach that integrates web-crawling, quantified content analysis, and multidimensional scaling to extract, assess, and visualize the image of cities from Web contents.

Keywords City image • Global city • Web pages • Multidimensional scaling • Quantified content analysis

J.S. Son (✉)
Project Mosaic, University of North Carolina, Charlotte, NC 28223, USA
e-mail: jaesoenson@gmail.com

J.-C. Thill
Department of Geography and Earth Sciences, University of North Carolina, Charlotte, NC 28223, USA
e-mail: jean-claude.thill@uncc.edu

J.-C. Thill (ed.), *Spatial Analysis and Location Modeling in Urban and Regional Systems*, Advances in Geographic Information Science,
https://doi.org/10.1007/978-3-642-37896-6_4

Introduction

It has been argued for over a decade that worldwide economic systems are increasingly framed by metropolitan gateways that form the focal points of thick trade and financial flows and house critical command and control business functions (Andersson and Andersson 2000). Metropolitan economies have positioned themselves in an economic space where national borders may have little bearing anymore. They are not only competing for a share of foreign direct investments, but also to attract innovative knowledge-based companies, international students, business schools, research and development centers, and highly skilled knowledge workers (Clark 2006). What factors offer comparative advantages that decisively differentiate cities in this race to the top? If cities offer similar structural and economic advantages, their image is bound to play an important role in attracting investment, knowledge workers, and business (Larsen 2014). This motivates cities to seek to establish a reputation for themselves and to seek a competitive edge in global markets through investment in branding campaigns (Anholt 2010). A starting point in such endeavor is the current status of the city image. In this research, we seek to better understand it through the composite image of cities as it is portrayed on the worldwide web.

A theoretical framework of city branding has been discussed among scholars in different disciplines. Previously, city branding has been facilitated by marketing theories developed from the perspectives of product and place marketing, but recently it has been more in line with principles of corporate branding (Kavaratzis 2009; Kavaratzis and Ashworth 2005). The similarity between corporate branding and city branding can be found in the components that lead to successful branding. Just as corporate branding consists of its core identity and of balanced relationships between strategic vision, organizational culture, and the stakeholders' image (Hatch and Schultz 2008), city branding is also structured by a city's identity and the relationships among strategic vision, its residents' culture, and external people's images. In both cases, the image is one of critical components to successful branding (Greenhalgh 2008; Kavaratzis and Ashworth 2005). Considering that the brand image refers to how the brand is perceived (Kavaratzis and Ashworth 2005), it is an affective component based on emotional experiences or preferences (Kwon and Vogt 2010).

The relationship between brand identity and brand image is reciprocal (Cai 2002; Florek et al. 2006; Qu et al. 2011). Personal experiences and preferences influence city image, and in turn the city image affects the city identity. There is a consensus that image development is utilized to facilitate word-of-mouth endorsements and destination branding can be conceptualized as forms of communication (Kavaratzis 2004). In this era of virtual information exchange, the Web is a space where communication happens; social media, especially, is fast becoming the most important channel for word-of-mouth communication (Conroy and Narula 2010).

It comes as no surprise to learn that the Internet is the center of our lives once you adapt to this modern technology. The Internet is not only worth being closely considered as a relevant source of information, but also because it has taken on the critical and active role of building the image of the treated information. Its efficiency and accessibility mean the fast accumulation of images, which are described, expressed, commented, and reviewed by an unprecedented number of potential users. In this respect, the image of the city is no exception. In other words, studying the image of the city through the lens of the Internet provides a multi-facetted portrayal of the city, that is, how contemporary citizens of the world think about the city. Although information on the Internet may be intentional (branding), or the result of third parties, or the results of happenstance (like good or bad news), all information is important to be considered because of the reciprocal effect between city identity and city image.

Cities are now compelled not only to advertise and market themselves but also to maintain an attractive and polished image. Having a positive public image is a valuable asset for a city, while a negative public image may be a major liability that is hard to overcome. Although the Internet plays an important role to create, disseminate, and modify the image of cities, Internet-based data have so far not been leveraged to portray the image of a broad cross-section of cities around the globe, and only a few case studies have focused on single cities. The image of cities is an intangible asset that remains to be apprehended through the quantification of Internet-based data.

The purpose of this research is to uncover the state of city images molded by the flourishing information exchanged among Internet users. The construction of the thematic dimensions that form the composite image of each city and the sorting of cities based on their similarity tendencies on each dimension enable us to apprehend the polarity, or alternatively the lack thereof, of the image of each city in relation to other cities around the globe. In turn, this portrayal can be related to world geographies and histories in which cities are embedded. The proposed data analytic approach based on internet programming, content analysis, and dimension reduction, provides a solution to the unstructured nature and large volume of organic information on the Internet. Considering that the branding strategy that a city may espouse to enhance its image starts at confirming its status, the distributional characteristics of the city image on the relative thematic space would be helpful to verify the status of the city image and provide a starting point to prepare a required branding strategy for individual cities.

The rest of the study is structured as follows. Section 2 reviews the modalities of emergence of city identify and city image, and the role of the city image on the Web as the foundation of a city branding strategy. The methodology and data sources of the study are introduced in Section 3. Results are presented and discussed in Section 4. Conclusions of this research are in Section 5.

City Identity, Image, and Branding on the Web

An image is a representation of the external form of a person or a thing in art; it means a simile or metaphor. It may also refers to the general impression that a person, organization, or product presents to the public. The latter view is in line with the purpose of this research because the aim here is to measure the standing of cities based on the quantified textual information, which is the representation of city images. Considering that a city is an integrated place for human interactions, the image of a city (or destination image) in the public arena is multifaceted since it results from the image formation of each individual (Gallarza et al. 2002). The image of a city is the composite of the impressions that the city presents to the public. Many different types of sources affect the image one may have of a city, such as reading a newspaper article, watching TV, personal visits, or word of mouth. The Internet on mobile devices is the fastest and most powerful information source nowadays, compared to more traditional media. As a medium of information exchange, the contents of the Internet like news portals, Really Simple Syndication (RSS) feeds, and social network web sites and mobile applications affect people while developing their image or view on a certain topic. In addition, contemporary media based on the Internet allows information providers to react to the users immediately while previous media generations were one-way mass communication. Thus, the process of information exchange between providers and users or among users reshapes the image that was once provided by the mass media.

According to Internet World Stats (2014), there are over three billion[1] Internet users (42.3% of the World population) in the World. The growth of the world total between 2000 and 2014 was 741.0%. Although Africa (6498.6%), the Middle East (3303.8%), and Latin America (1672.7%) show rapid expansion, their Internet users remain a small share of worldwide tallies (Africa, 9.8%; Middle East, 3.7%; Latin America, 10.5%); it is quite smaller than the shares of Asia (45.7%) and Europe (19.2%). This statistic reflects the inequity of the Internet infrastructure, but it also shows that the Internet is an important tool in contemporary societies. The development of wireless technologies and the popularization of smart phones have accelerated *Homo Interneticus*[2] (Barnes 2010). Searching, collecting, and distributing information on the Web have become the inherence of humans, like reading, listening, speaking, and writing. As of January 2015, there were 876,812,666 sites; the number of sites has increased without interruption, except during the period between December 2008 and December 2009, the period of the global economic downturn (Netcraft 2015). The rapid proliferation of web sites is further evidence that people accept the Web as a part of their lives, whether their purposes are personal, commercial (business), government, and nonprofit organization web sites.

[1]The number is 3,035,749,340 as of June 30, 2014.

[2]This is the title of 4th episode in *The Virtual Revolution*, which is the BBC documentary discussing the huge benefit and unforeseen downsides of the World Wide Web.

While traditional media (e.g. guidebooks, brochures, advertisements on TV and newspapers) exhibit comparative disadvantages, including being passive forms of communication, expensive to produce, whose effectiveness may be difficult to monitor and whose message often goes unheard (Kolb 2006), web sites help marketing providers overcome them. The web site of business and nonprofit organizations has three main purposes, which are direct selling, sales support and customer service, and advertising and public relations; similarly, the main purpose of a web site of a tourism office is to advertise the city's features and benefits (Kolb 2006). In addition, the reason why people use the Internet for getting information is that materials and contents on the Internet are more current, more thorough and richer than conventional promotional agents (Govers and Go 2003; Heung 2003). Thus, this platform provides a good branding opportunity to enhance the image of a city while also meeting the information needs of web site users.

'Branding' is the promotion of a particular product or company by means of advertising and distinctive design. City branding is a purposeful strategy of image building and shaping intended to enhance people's beliefs that a city has a positive unique something. City branding is typically motivated by economic considerations. Just as the corporate image needs to be updated for global recognition by promoting all offerings under one corporate brand (Balmer 2010; Erdogmus et al. 2010; Hatch and Schultz 2009), cities also try to brand themselves for economic benefits. Successfully branded cities are recognized for their positive image as an attractive and pleasant place to live, work, recreate, and invest.

The starting point of city branding is to ascertain the identity of the city. Whether the city likes or dislikes its current identity, the image of the city is molded by the interaction between identity and people. Thus, it is critically important to know these images in order to set up an effective branding strategy. As the influence of online digital information on image formation has become an important issue (Govers and Go 2005), information on the Web is important for city branding.

Achieving a desired city image through branding on the internet is a complex undertaking. For instance, Govers and Go (2005) used pictures and text from tourism-related web sites to find the projected images of Dubai. They concluded that information provided by the web is fragmented, lacks creativity and coherence, and just offers limited products through a few business sectors (i.e. only dining and shopping, no consumption of place). In addition, Choi et al. (2007) used the narrative and visual information on a sample of web sites in Macau. They pointed out that the image of Macau projected online varies according to the information sources due to their different communication objectives and targeted audiences. Both studies show the need for a master plan, including identities of the city, desired images, projected images, and solution for any arising discrepancy.

When we consider that the Web (including social media service) is the most popular space for sharing views, impressions, and images nowadays, the quantified textual contents of the Web can be very effective at giving a semantic meaning to the images of each city in relation to an externally established corpus of knowledge on cities that is synthesized by recognized basic thematic dimensions. The comparison between each such dimension and actual textual data on cities can help to find how

each dimension is projected on the textual contents of the Web. In other words, the result informs us about the thematic closeness or similarity of each city to each dimension. In turn, this can be used for measuring the similarity among cities in the discourse that is carried out by various agents on the communication space of the Internet. In addition, it can be used as the evidence on the basis of which to set up the strategy to enhance the image of a certain city.

Research Design

A representative set of large and globally recognized cities from all world regions is used in this study. A list of 264 cities (Appendix) was compiled on the basis of a comprehensive review of the literature[3] on global cities. The list includes cities labeled as global cities by two studies at least, including statistics from the United Nations. These cities are distributed geographically[4] as follows: 73 cities in Asia, 58 in Europe, 44 in North America, 33 in Africa, 21 in South America, 15 in the Middle East, 11 in Central America, 6 in Oceania, and 3 in the Caribbean.

Data about each study city was assembled from web pages written in English. Different conclusions could be reached if web pages written in other languages were also considered. English has been used for communicating internationally (lingua franca) for many decades, and it has deep and powerful influences. The usage of English has be instrumental in the process of globalization and all indications suggest it may well continue to do so in the future. Considering that English is the language most often used in Web contents −56.1% (W3Techs 2012), including much content assembled in and for countries where English is not the local population's native language, we are not meaningfully restricted in our ability to recognize the image of global cities from English web pages only. Data consist in the full text of a sample of web pages associated with each of the cities under study. While web pages can be downloaded manually, web crawling was used as this approach is more efficient to download a massive amount of web pages in a limited amount of time. Web crawlers require a list of seed URLs to initiate the search for relevant information. For each city under study, part of the seed URLs comes from the database of the Open Directory Project (ODP, http://www.dmoz.org), while other seeds come from Google Search results under the city name (and country). These seeds were collected between February and July 2012. Collected seed URLs were uploaded to a commercial-grade crawling web

[3]Literature includes Hymer (1972), Cohen (1981), Reed (1981), Friedmann (1986), Rimmer (1991), London Planning Advisory Council (1991), Sassen (1994), Friedmann (1995), Keeling (1995), Knox (1995), Petrella (1995), Finnie (1998), Beaverstock et al. (1999), Short and Kim (1999), Thrift (1999), Taylor and Catalano (2000), Taylor et al. (2002), Kearney (2008), Knox et al. (2008), Mastercard (2008), Kearney (2010), EIU (2012), & UN DESA (2012).

[4]This geography is based on geographical regions and compositions of the UN DESA Statistics Division (2013).

service (80legs.com). Downloading the main dataset of web pages was conducted from August to September 2012 with the following settings: 100,000 target URLS (nodes) per city, breadth first as crawl type, and 6 degrees for depth level. The total volume of downloaded data amounted to about 600GB of text, at the exclusion of other types of contents such as pictures, flash animations, videos, and others. The number of collected words for each city varies from thousands to millions. In order to remove the scale effect of total numbers the proportion of three thematic groups of words is used in this study.

The premise of our research design is that the textual contents of city web pages encompass necessary and sufficient information to establish the thematic profile of each city in the global communication space. Three challenges are associated with the use of city web pages as evidence of city images. First, text can be the real description or experience of a city or possibly its aspired image, like in advertisements and branding messages. It is argued here that images are intrinsically subjective and that the attitudes and motivations of the authors of city-related textual content can hardly be determined. Hence, textual contents collected from web pages will not be filtered before further processing. Second, the contents of city web pages include much generic and ubiquitous text as well as text data that may serve to depict the city's own image. For this purpose, we follow the general process of quantitative content analysis (QCA). The QCA includes the preparation of code schemes (i.e. the creation of dictionaries), frequency analysis, and the interpretation of frequency distributions based on code schemes. Third, given the large and diverse volume of data that city web pages encompass, the semantic dimensions present in the textual contents must be generalized and reduced to forms that are more readily interpretable and comparable across cities and groups thereof. The data reduction technique of multidimensional scaling (MDS) is used for this purpose. The multistep data processing approach proposed in this study provides a possible method for city research through the quantification of textual data from web pages and dimensional reduction.

The process of extraction of frequencies of keywords from web pages is depicted in Fig. 1. The processing of the textual content involves pre-processing such as tokenization, parsing, stemming, and lemmatization (Büttcher et al. 2010, pp. 84–90; Manning et al. 2009, pp. 22–35). These procedures generate a set of keywords for each city. These sets are used for the calculation of frequencies and categorization, which are based on a predefined code scheme. Keywords which are used for the predefined code scheme are extracted from articles relevant to globalization from The Global Policy Forum (2014a, b, c). Among the Global Policy Forum archives, the articles classified as globalization of the economy, globalization of cultures, and globalization of politics are used to build keyword dictionary for the economic, cultural, and political dimensions of the written discourse. The final selection of keywords extracted from these articles is made after a close review of the list of frequent keywords from each set of articles (Table 1). The word frequencies are calculated by counting how many times the keywords associated with each dimension correspond to the keyword set of each city. Therefore, the geographic matrix of word frequencies consists of cities and three dimensions.

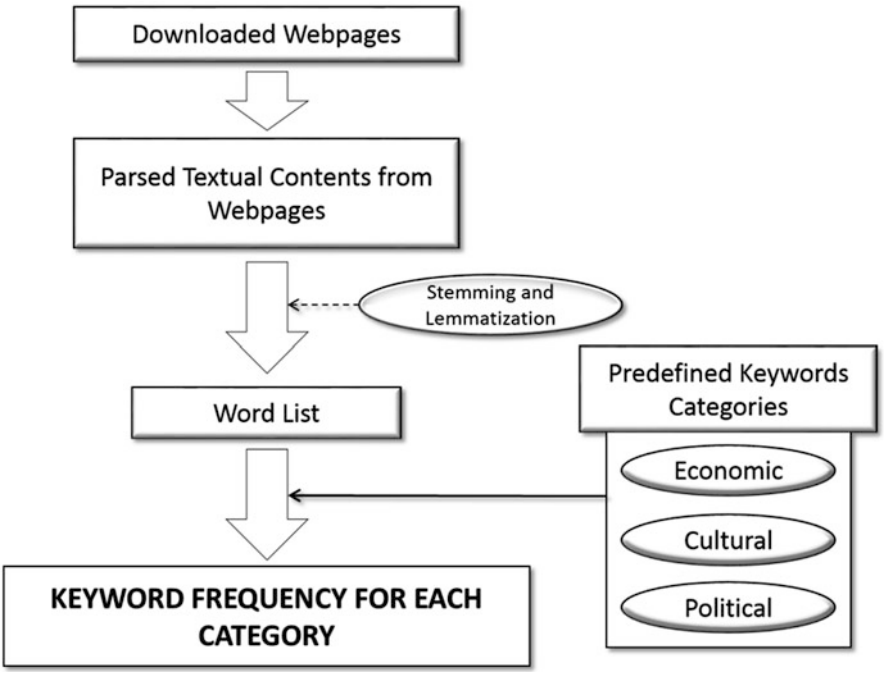

Fig. 1 Procedure of word frequency extraction

Table 1 Keywords for the predefined code scheme of three thematic dimensions

Dimension (number of words)	Keywords
Economic (44)	Bank, business, capital, company, corporate, currency, debt, demand, develop, dollar, economy, employer, export, finance, firm, flow, fund, gdp, goods, income, industry, inequality, interests, investment, investors, labor, market, monetary, money, pay, poverty, price, product, profit, rates, sectors, services, subsidies, supply, taxes, trade, union, wealth, work
Cultural (25)	Church, communication, community, culture, diversity, education, English-language, heritage, humanitarian, indigenous, information, intellectual, language, media, Olympic, religion, revolution, school, society, sports, subculture, television, UNESCO, war, web
Political (50)	Activist, administration, capitalism, capitalist, commission, conflict, congress, cooperation, corruption, council, crisis, davos, democracy, democratic, deregulation, diplomatique, elections, g20, governance, government, humanitarian, institutions, intervention, justice, leadership, leftwing, liberal, military, minister, nation-state, neoliberal, ngo, organiz(s)ation, peace, policy, policymakers, politics, president, protection, protests, protocol, public, reform, security, socialist, sovereignty, treaty, un-ngo, vote, war

Note: the stems of these words are used for frequency analysis

Lastly, the frequencies for each dimension of each city are transformed into the proportion of key words of each dimension to the total frequencies in order to compare the similarity of the component ratio among cities (See Appendix).

Multidimensional scaling (MDS) is helpful to reveal the similarity of cities based on the relative share of the three categories of keywords. MDS is an ordination technique that extracts a limited number of dimensions from a multidimensional dataset based on similarities or dissimilarities exhibited by observations on a number of original characteristics. It creates a space of lower-dimensionality, which can be visualized to explore patterns. MDS estimates the coordinates in a space of specified dimensionality that comes from the distance between pairs of objects (cities) (Deun and Delbeke 2000; Clark 2004). Generally, MDS refers to a class of models by which information contained in a set of data is represented by a set of points in an attribute space. For this analysis, metric MDS is used, as implemented in R (R Core Team 2013). The MDS procedure provides the visualization that allows researchers to analyze the relationships among variables (i.e., categories) and among objects (i.e., cities) more readily.

Results

The analysis of the textual contents of city web pages produces a scatterplot (or map) of all cities where cities with similarly categorized word frequencies are positioned in the same vicinity. Fig. 2 is a 2-dimensional MDS map, where the proportions of each of the three categories of keywords are used as input attributes. The similarities (i.e. distances) are calculated by the metric MDS algorithm. Stress (i.e. goodness of fit) is 1 for 2 dimensions. The relative strength of each thematic dimension is also represented on the MDS map in Fig. 3 to facilitate the interpretation of the positioning of cities in the space of image dimensions. The direction from the center of the plot to the top-left of the plot points towards a higher share of economic keywords; the direction from the center of the plot to the bottom-right corresponds to a higher share of cultural keywords, and an increasing share of political keywords is associated with the direction from the center of the plot to the top-right. Thus, the distribution of cities fits on a triangle whose three corners represent the highest shares of citations on each of the three thematic dimensions. Worldwide, the MDS map reveals that city images balance off the three primary thematic dimensions of the analysis, i.e., the economic, the political, and the cultural, as cities exhibit a strong tendency to lump towards the center of the scatterplot. Some differentiation across cities is noticeable however.

A few cities are overwhelming known on the web-based communication space for their economic qualities. This is the case of a number of Chinese cities, including Foshan, Dalian, Zhengzhou, Dongguan, Shenzhen, Shantou, Xiamen, and Shanghai, as well as Luxembourg, all of which are positioned in the top-left of the plot. The political theme prevails among a second group of cities that are all located in less developed countries, including Khartoum, Damascus, Islamabad, Kinshasa, Addis

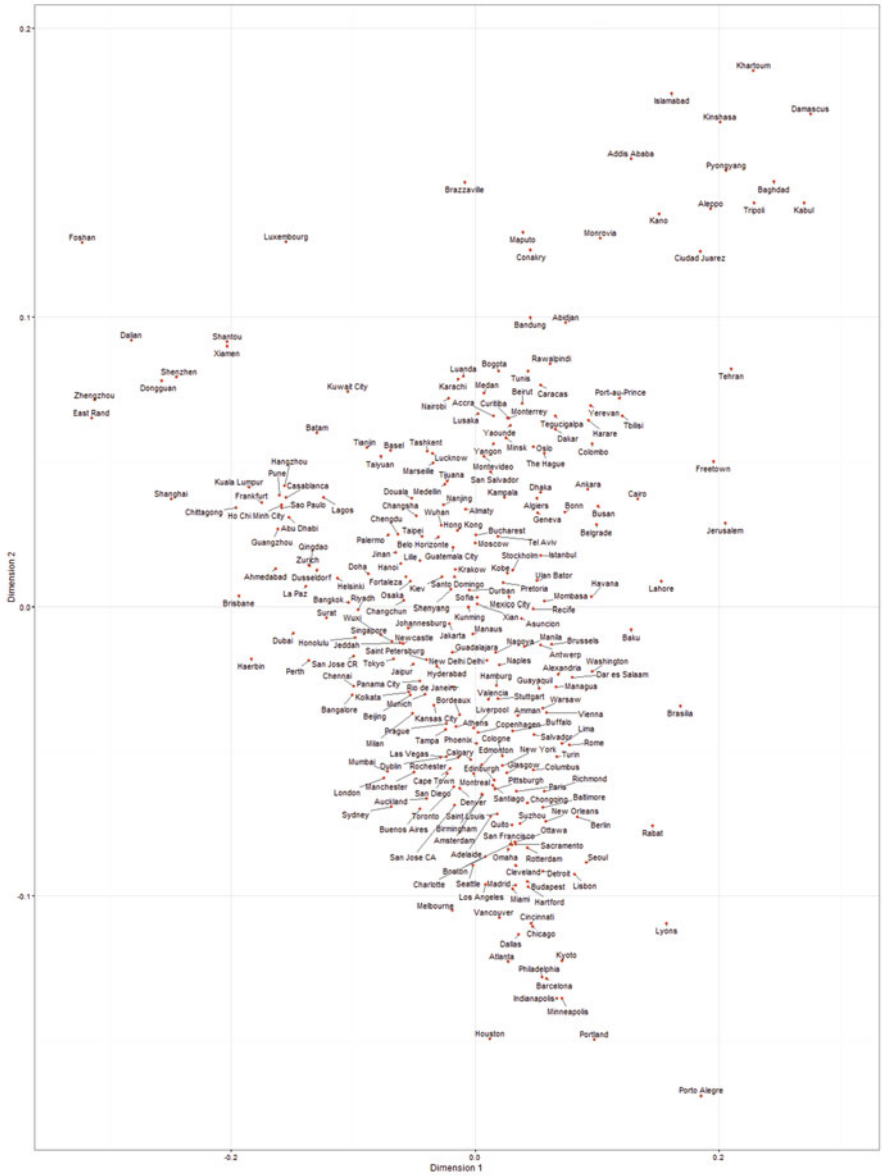

Fig. 2 Multidimensional scaling map of cities based on keyword distribution across themes

Ababa, Pyongyang, Baghdad, Tripoli, Kabul, Aleppo, Kano, Ciudad Juarez, and Monrovia. All these cities are found in the top-right of the plot. We can also include Brazzaville and Maputo in this group, but these cities have more economic tendency compared to the first group of political cities. Finally, the web-based discourse on

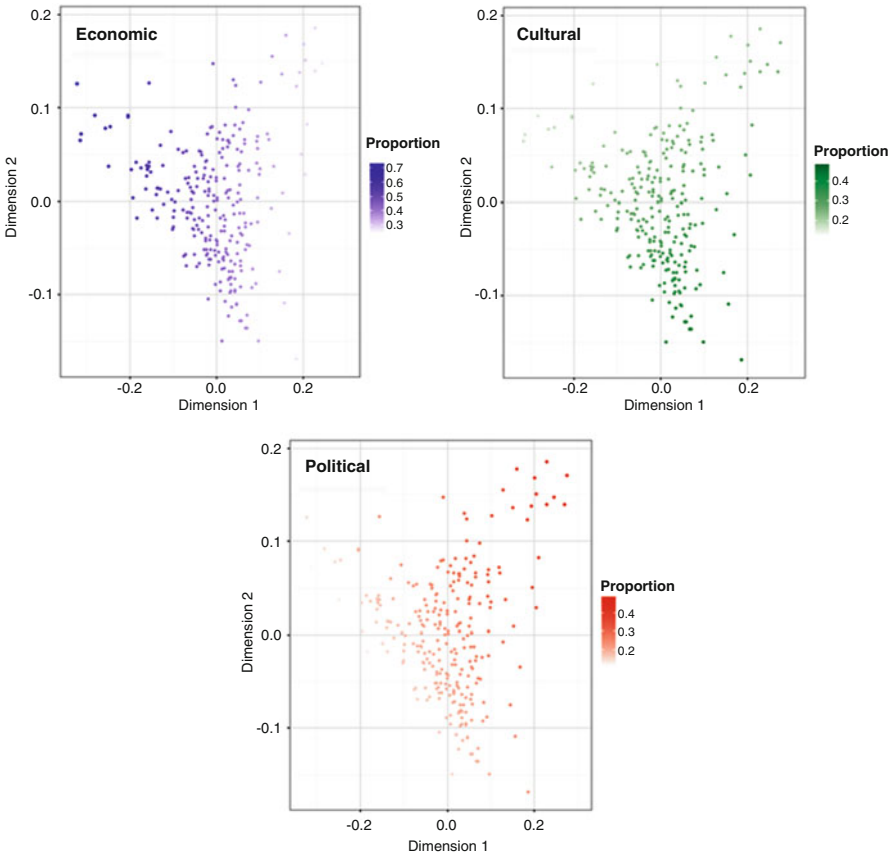

Fig. 3 Distributions of the proportion of keywords of each thematic dimension

a small group of cities is dominated by a cultural tendency. This group includes Houston, Portland, Indianapolis, Minneapolis, Atlanta, Chicago, Barcelona, and Kyoto.

The geography of the web-based discourse on cities can also be approached more directly on a continent-by-continent basis. This results in nine distinct maps (Figs. 4, 5, 6, 7, 8, 9, 10, 11 and 12). These maps provide evidence for the geographic differences among cities on the basis of textual data. Continents are classified into 9 geographic ensembles: Africa, Asia, Middle East, Oceania, Europe, North America, Central America, South America, and the Caribbean.

African Cities (Figure 4): East Rand, Lagos, and Casablanca show a strong tendency for the economic. Rabat shows the strongest tendency for the cultural, and Cape Town, Alexandria, and Dar es Salaam are next. However, it is on the political dimension that African cities stand out most clearly: Khartoum, Kinshasa, Addis Ababa, Tripoli, and Kano do so most strongly, while Brazzaville and Maputo

Fig. 4 Distribution of African cities on the MDS map of image dimensions

are next. The remaining cities are distributed between the political and the cultural poles. Overall, African cities show a strong tendency towards either the political or the cultural, or a combination of these dimensions. This is highly closely with the political issues that plague African countries on an ongoing basis such as wars and political unrest, as well as to the resource-based economy of many African countries. These themes are routinely discussed and commented on through internet-based media and dominate the discourse on Africa.

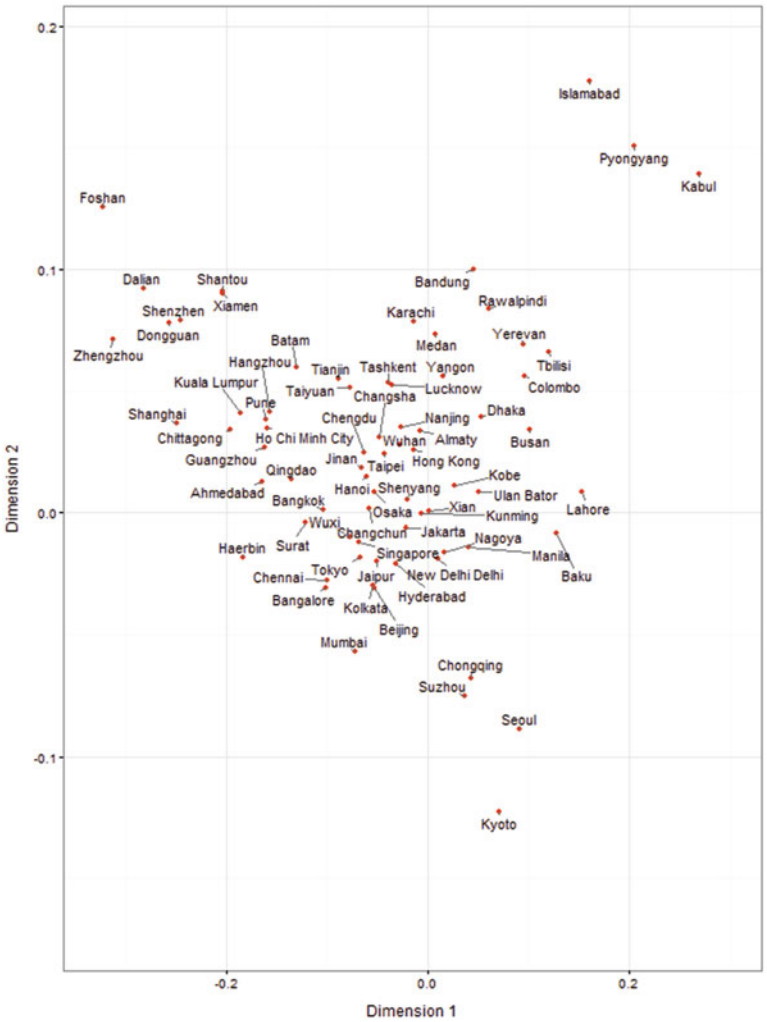

Fig. 5 Distribution of Asian cities on the MDS map of image dimensions

Asian Cities (Figure 5): Three groups of cities have the strongest tendency towards each pole of the web-based discourse. For the economic, 7 Chinese cities including Foshan, Dalian, Zhengzhou, Dongguan, Shenzhen, Shantou, and Xiamen are grouped together. Islamabad, Pyongyang, and Kabul are associated with a strong political tendency. As for the cultural dimension, we find Kyoto, Seoul, Suzhou, and Chongqing being grouped together. More cities are distributed along the imaginary line connecting the economic and cultural poles. The plot also shows that cities towards the political pole are more dispersed. Many cities are distributed around the origin of the coordinates, which denotes the low thematic polarity of many

Fig. 6 Distribution of Middle Eastern cities on the MDS map of image dimensions

Asian cities. In sum, three small groups of cities associated with each content dimension and the low polarity of other cities are the distributional characteristics of the worldwide discourse on Asian cities.

Middle Eastern Cities (Figure 6): Damascus, Baghdad, and Aleppo are associated with a strong political tendency; Kuwait City shows a moderately strong tendency towards the economic theme. Abu Dhabi, Riyadh, Dubai, Doha, and Jeddah form a group located between the economic and the cultural poles, while Tehran and Jerusalem exhibit bipolarity that blends the political and cultural poles. Amman

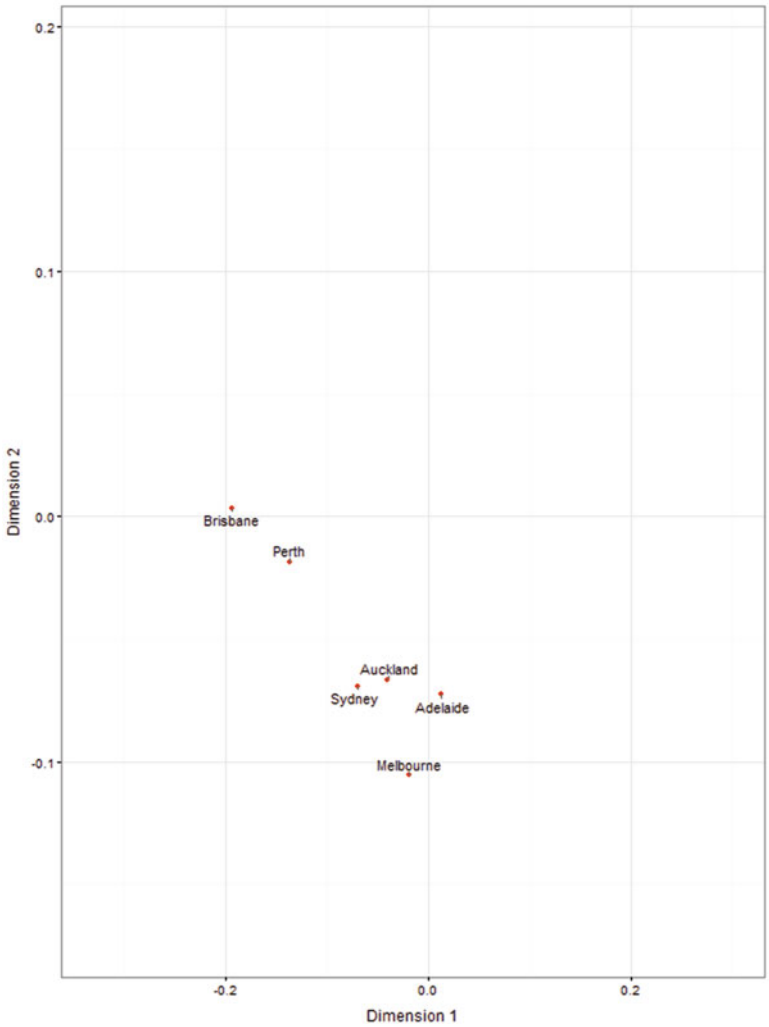

Fig. 7 Distribution of Oceania cities on the MDS map of image dimensions

shows a weak tendency towards the cultural. Tel Aviv and Beirut are the cities of the Middle East whose web-based discourse is closest to achieving a balance between the three thematic dimensions. The distributional characteristics of Middle Eastern cities are marked by the separation between some cities with strong political tendency and some cities with economic tendency. The cultural emphasis is surprisingly weak among this regional group of cities.

Oceania Cities (Figure 7): All 6 cities are distributed on the imaginary line between the economic and cultural poles. Brisbane, Perth, and Adelaide exhibit some tendency towards the economic. Other cities including Sydney, Auckland, and Melbourne form a group trending more towards the cultural theme. Without any city

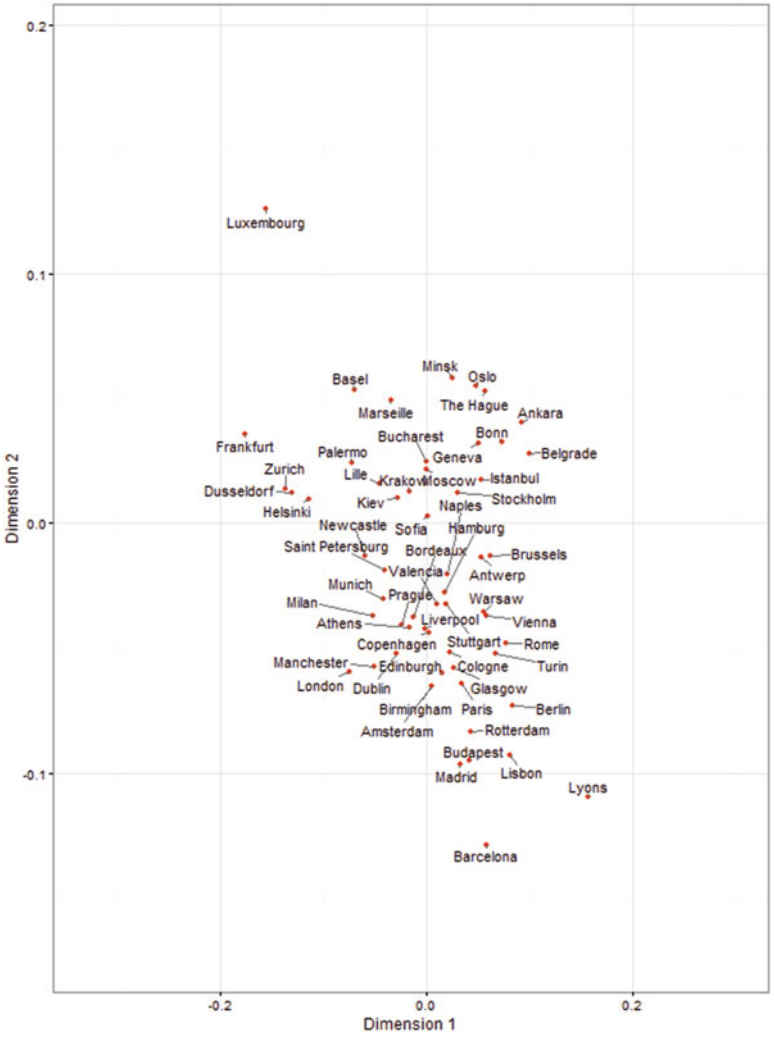

Fig. 8 Distribution of European cities on the MDS map of image dimensions

with a political tendency, the mixed tendency of the economic and the cultural forms the distributional characteristic of Oceania cities.

European Cities (Figure 8): European cities are distributed around the center of the map and extend towards the cultural pole (i.e. the bottom-right of the plot), except for Luxembourg. This denotes that European cities have fairly balanced images. Luxembourg has the strongest tendency for the economic theme, while Frankfurt is next, owing to their strong position as financial gateways in the European space and beyond. Along with Zurich, Düsseldorf, and Helsinki, the discourse on Frankfort is in fact well balanced between the economic and the cultural themes. Minsk, Oslo, Belgrade, The Hague, and Ankara form a group with

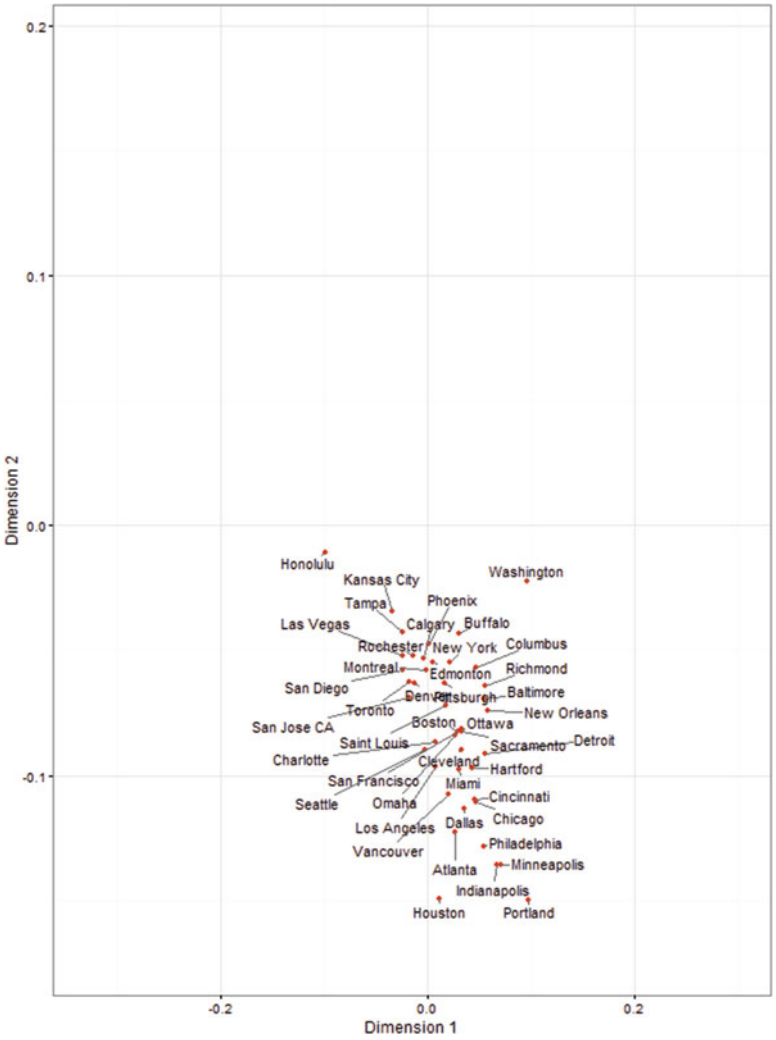

Fig. 9 Distribution of North American cities on the MDS map of image dimensions

some political tendency. Barcelona shows the strongest tendency for the cultural, while Madrid, Lisbon, Rotterdam, and Budapest can be seen as the next group with strong tendency for culture.

North American Cities (Figure 9): The web-based discourse on North American cities has the commonality of a strong tendency for culture so that they are distributed predominantly at the top of the plot. Of all North American cities, Honolulu is the only one showing some tendency towards the economic; reflecting its function as a national capital, Washington is somewhat more associated with political discourse than other North American cities. Strong similarity among North

Fig. 10 Distribution of Central American cities on the MDS map of image dimensions

American cities means that the web-based discourse on them is rather similar, and this discourse is strongly identified with the cultural theme. This rather lopsided trend calls for further analysis. It may well be related to the imperatives that North American cultural institutions found themselves in to reach out to potential patrons through well-articulated media and advertising campaigns to secure their financial future given the limited financial support from governmental sources. No systematic differences exist between US and Canadian cities.

Central American Cities (Figure 10): Central American cities are distributed rather centrally on the plot of thematic dimensions of the web-based discourse on

Fig. 11 Distribution of South American cities on the MDS map of image dimensions

cities, except for Ciudad Juarez, which has a strong tendency for the political. Cities are similar to each other in terms of their lack of emphasis on the economic and cultural dimensions. Some minor differentiation among these cities can be detected, with the discourse on Monterrey, San Salvador, and Tegucigalpa being noticeably more political than for other cities.

South American Cities (Figure 11): São Paulo and La Paz exhibit the strongest economic tendency, with a leaning towards the cultural. Bogotá, Caracas, Curitiba, Medellin, and Montevideo show a relatively strong tendency towards the political.

Fig. 12 Distribution of the Caribbean cities on the MDS map of image dimensions

Of the remaining cities, Quito, Buenos Aires, Salvador, Porto Alegre, and Lima have a discourse leaning towards the cultural. Overall, South American cities are more homogeneous than African, Central American, or Asian cities; the web-based discourse is rather well balanced between all three thematic dimensions.

The Caribbean Cities (Figure 12): All three Caribbean cities are the object of fairly similar discourses that trend towards the political and show weakness on the economic and the cultural. Port-au-Prince exhibits a relatively strong political tendency compared to Havana and Santo Domingo.

Finally, the web-based discourse on US cities and Chinese cities is presented as a focused case study. This analysis is of interest because, of all the countries

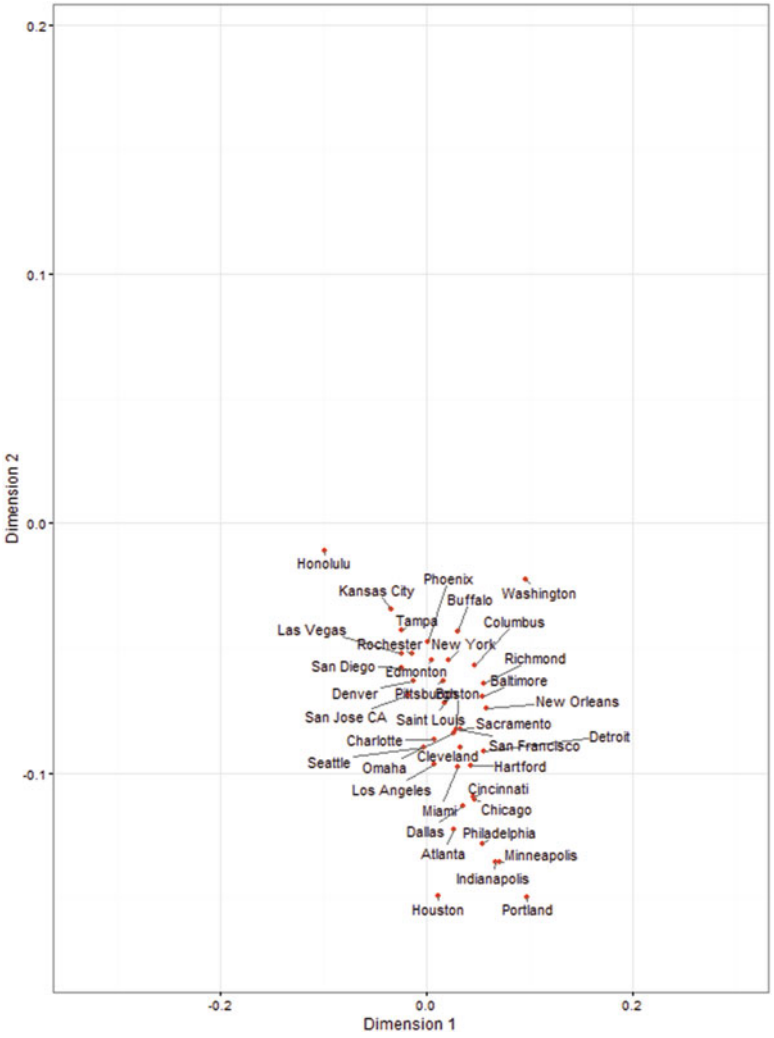

Fig. 13 Distribution of US cities on the MDS map of image dimensions

represented in our dataset, the United States and China encompass the largest numbers of cities. They also represent the two largest economic powers in the world and considerable trade seals they mutual interdependence. Finally, many Chinese cities have tailored ambitious branding and marketing programs to establish themselves as global cities in a contemporary world that is still largely dominated by Western metropolitan economies.

US Cities (Figure 13): The MDS map for US cities is not much different from the map for North American cities (Fig. 8) because most North American cities are in

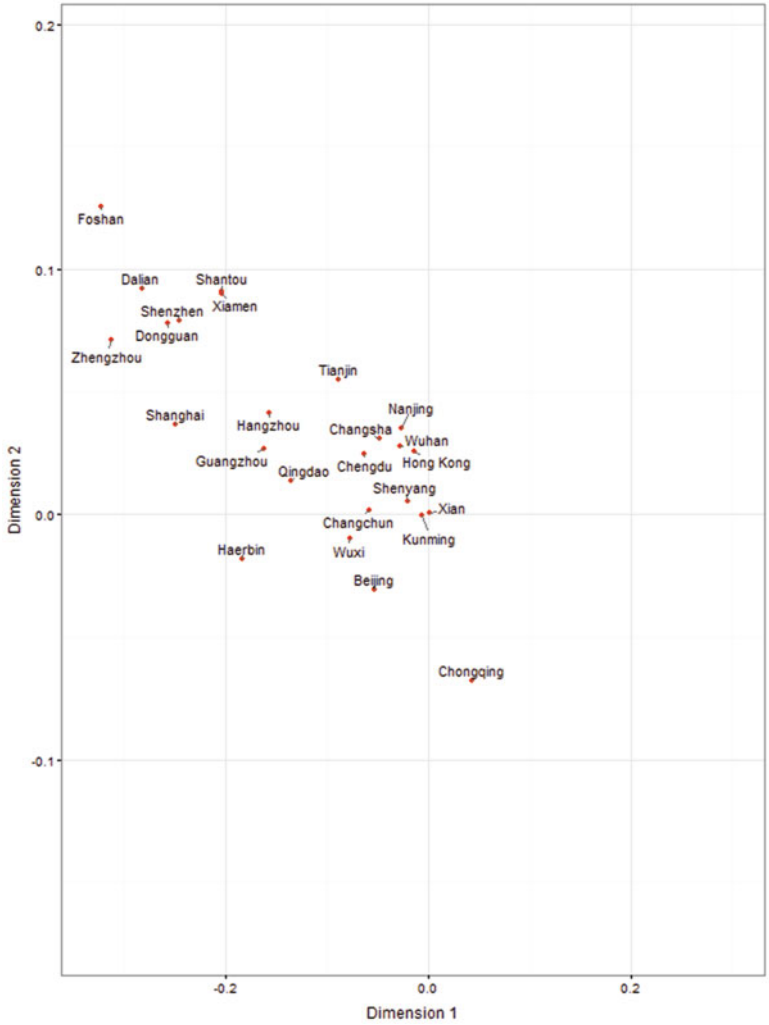

Fig. 14 Distribution of Chinese cities on the MDS map of image dimensions

the United States. US cities are highly concentrated towards the cultural pole even though Honolulu and Washington also exhibit some weak tendency towards the economic and the political, respectively. Overall, US cities have a strong similarity to the cultural aspect of cities. Political stability may explain the lack of overall emphasis on political matters.

Chinese Cities (Figure 14): Chinese cities are distributed along the axis between the economic and cultural poles. Foshan, Dalian, Shantou, Shenzhen, Zhengzhou,

Dongguan, and Xiamen are distributed in the top-left of the plot, which shows a very strong tendency for the economic. Shanghai, Hangzhou, and Tianjin form the next group of strong tendency for the economic. On the other hand, Chongqing is the sole Chinese city exhibiting a strong cultural tendency. Other cities fall between these two extremes. The political theme is characteristically absent from the discourse pertaining to Chinese cities, which can be directly ascribed to the lack of open debate in the Chinese political system.

A comparison between US cities and Chinese cities reveals several interesting points. First, we note the absence of a city having strong tendency for the political theme in either country. Cities in both countries are situated on the imaginary axis between the economic and cultural poles. Second, several differences are noticeable. The first difference concerns the degree of differentiation among cities of each country. US cities are more homogeneous than Chinese cities, which could be ascribed to the greater maturity of American cities and to the diversity of growth paths followed by Chinese cities since the 1978 adoption of the national economic reform policy. The second difference relates to the polar tendencies of cities in the two countries. While US cities are mostly situated in the plot area of strong cultural emphasis, Chinese cities are distributed near the plot area of strong economic polarity. As far as web-based discourse is concerned, Chinese and US cities have little in common.

Conclusions

This research aimed at revealing the state of city image from the massive Web-based contents, thus enabling us to verify the status of the city image and provide a starting point for the future branding strategy of each city. City branding has become an important tool to enhance cities' competitiveness to draw people, capital, and businesses. It is important for a city to recognize what kinds of image external people have in order to enhance city's brand power. The Web is the most important channel for word-of-mouth communication nowadays, which exchanges information and knowledge for image formation fast and extensively. Although the Web information is important in the field of city image research, the massiveness and dynamicity of Web data and the difficulty of creating general thematic dimensions have hindered the utilization of this data source. Internet programming and content analysis were proposed in this study as potential methods to resolve these issues.

The outcome of this study supports the merit of the new interdisciplinary approach that integrates web-crawling, QCA, and MDS to extract, assess, and visualize the image of cities from Web contents. Thus this research framework opens the door to the new world of information research which can support urban science research to reflect contemporary lifestyles and communication trends.

From the specific perspective of the analytical results of the study, we find that Chinese cities and Luxembourg show strong tendency for economic qualities, cities in less developed countries show strong tendency for the political theme, and a small number of cities located in developed countries trend towards the cultural dimension. Continent-by-continent analysis of the web-based discourse on cities confirmed geographic differences among cities that can be linked back to their histories and socio-political context. Finally, from the focused case study, US cities versus Chinese cities, we found the common absence of political polarity as well as the differences–more homogeneous distribution of US cities, and cultural polarity of US cities vs. economic polarity of Chinese cities.

The distribution of cities on the MDS map mirrors their image as they are expressed, discussed, and shared on the Web. Cities having high tendency for one of the three themes are strongly connected with that theme in the internet communication space. However, the hybrid topical mapping of cities does not reveal a simple relationship between global city ranking and polarity, or lack thereof. For instance while it could be surmised that well-balanced cities are better positioned to be globally significant cities, it turns out that the three cities that routinely make it to the top of the global rankings, London, New York, and Tokyo, are not located in the center of the MDS map, but between the economic and the cultural poles. This study suggests much greater nuances, so that we can only use this information to confirm the broad status of a city's image and how similar cities are to each other, but not which cities are higher up on the world rankings. Nevertheless, this study provides a benchmark against which a city's image status on the global communication space of the Web can be measured. A city can leverage this result as a benchmark so as to more effectively craft a branding strategy for enhancing its image.

In spite of the above contributions that demonstrates the scientific value of web-based big data analytics for urban and regional science research, this particular study has limitations. First, as a communication platform, the Web is highly dynamic. By the time data has been collected, processed, and analyzed, it already reflects a past reality. Continuous data collection assorted with automated processing would enhance our ability to monitor and trace longitudinal changes in city discourse. Another limitation is that this research is only based on web pages written in English language. Considering that different languages provide perceptions of cities framed in different cultural contexts, language-specific analyses would bring complementarity to the study of city images conducted here. It could reveal the linguistic characteristics of Web data and human perception differences of peoples who use these languages and share different cultural affinities. Finally, as the world is increasingly communicating using social media platforms, this fast expanding communication space may increase the scope of city image studies. Further research along these lines of considerations would enhance web-based city image research.

Appendix: List of Study Cities and Word Frequency Proportions

City	Country	Continent	Economic keywords	Cultural keywords	Political keywords
Abidjan	Cote d'lvoire	Africa	0.405	0.258	0.337
Abu Dhabi	United Arab Emirates	Middle East	0.578	0.244	0.178
Accra	Ghana	Africa	0.449	0.263	0.288
Addis Ababa	Ethiopia	Africa	0.380	0.230	0.390
Adelaide	Australia	Oceania	0.537	0.300	0.164
Ahmadabad	India	Asia	0.558	0.264	0.178
Aleppo	Syria	Middle East	0.289	0.269	0.443
Alexandria	Egypt	Africa	0.400	0.341	0.260
Algiers	Algeria	Africa	0.415	0.292	0.292
Almaty	Kazakhstan	Asia	0.460	0.287	0.254
Amman	Jordan	Middle East	0.416	0.347	0.238
Amsterdam	Netherlands	Europe	0.467	0.332	0.202
Ankara	Turkey	Europe	0.392	0.297	0.311
Antwerp	Belgium	Europe	0.401	0.333	0.265
Asunción	Paraguay	South America	0.398	0.342	0.260
Athens	Greece	Europe	0.455	0.336	0.209
Atlanta	United States, GA	North America	0.423	0.408	0.169
Auckland	New Zealand	Oceania	0.479	0.345	0.176
Baghdad	Iraq	Middle East	0.273	0.269	0.457
Baku	Azerbaijan	Asia	0.360	0.337	0.303
Baltimore	United States, MD	North America	0.393	0.384	0.223
Bandung	Indonesia	Asia	0.419	0.247	0.334
Bangalore	India	Asia	0.524	0.306	0.170
Bangkok	Thailand	Asia	0.540	0.279	0.181
Barcelona	Spain	Europe	0.405	0.412	0.183
Basel	Switzerland	Europe	0.511	0.268	0.221
Batam	Indonesia	Asia	0.560	0.223	0.217
Beijing	China	Asia	0.484	0.322	0.195
Beirut	Lebanon	Middle East	0.433	0.271	0.296
Belgrade	Serbia	Europe	0.409	0.286	0.305
Belo Horizonte	Brazil	South America	0.479	0.279	0.241
Berlin	Germany	Europe	0.427	0.363	0.210
Birmingham	United Kingdom	Europe	0.446	0.353	0.201
Bogotá	Colombia	South America	0.454	0.252	0.294
Bonn	Germany	Europe	0.386	0.307	0.307
Bordeaux	France	Europe	0.474	0.325	0.201
Boston	United States, MA	North America	0.432	0.373	0.196
Brasília	Brazil	South America	0.317	0.361	0.321
Brazzaville	Congo (Rep. of)	Africa	0.475	0.201	0.324

(continued)

City	Country	Continent	Economic keywords	Cultural keywords	Political keywords
Brisbane	Australia	Oceania	0.579	0.272	0.149
Brussels	Belgium	Europe	0.405	0.334	0.261
Bucharest	Romania	Europe	0.489	0.266	0.246
Budapest	Hungary	Europe	0.422	0.376	0.202
Buenos Aires	Argentina	South America	0.488	0.347	0.166
Buffalo	United States, NY	North America	0.439	0.339	0.222
Busan	Korea (Rep. of)	Asia	0.366	0.326	0.308
Cairo	Egypt	Africa	0.357	0.322	0.321
Calgary	Canada	North America	0.459	0.345	0.195
Cape Town	South Africa	Africa	0.463	0.352	0.185
Caracas	Venezuela	South America	0.421	0.268	0.311
Casablanca	Morocco	Africa	0.573	0.239	0.188
Changchun	China, Jilin	Asia	0.492	0.288	0.220
Changsha	China, Hunan	Asia	0.530	0.261	0.209
Charlotte	United States, NC	North America	0.446	0.374	0.180
Chengdu	China, Sichuan	Asia	0.509	0.279	0.212
Chennai	India	Asia	0.532	0.297	0.171
Chicago	United States, IL	North America	0.393	0.415	0.191
Chittagong	Bangladesh	Asia	0.607	0.232	0.161
Chongqing	China	Asia	0.413	0.371	0.216
Cincinnati	United States, OH	North America	0.442	0.374	0.184
Ciudad Juarez	Mexico	Central America	0.336	0.260	0.404
Cleveland	United States, OH	North America	0.410	0.395	0.195
Cologne	Germany	Europe	0.439	0.340	0.221
Colombo	Sri Lanka	Asia	0.441	0.300	0.259
Columbus	United States, OH	North America	0.428	0.365	0.207
Conakry	Guinea	Africa	0.438	0.242	0.320
Copenhagen	Denmark	Europe	0.446	0.341	0.213
Curitiba	Brazil	South America	0.430	0.268	0.302
Dakar	Senegal	Africa	0.404	0.279	0.317
Dalian	China, Liaoning	Asia	0.685	0.165	0.150
Dallas	United States, TX	North America	0.413	0.401	0.186
Damascus	Syria	Middle East	0.255	0.258	0.488
Dar es Salaam	Tanzania	Africa	0.388	0.345	0.267
Denver	United States, CO	North America	0.462	0.352	0.186
Detroit	United States, MI	North America	0.401	0.394	0.206
Dhaka	Bangladesh	Asia	0.423	0.295	0.282
Doha	Qatar	Middle East	0.523	0.277	0.200
Dongguan	China, Guangdong	Asia	0.663	0.183	0.155
Douala	Cameroon	Africa	0.496	0.263	0.241
Dubai	United Arab Emirates	Middle East	0.568	0.275	0.156

(continued)

City	Country	Continent	Economic keywords	Cultural keywords	Political keywords
Dublin	Ireland	Europe	0.506	0.339	0.155
Durban	South Africa	Africa	0.459	0.309	0.232
Düsseldorf	Germany	Europe	0.555	0.264	0.181
East Rand	South Africa	Africa	0.672	0.213	0.115
Edinburgh	United Kingdom	Europe	0.441	0.355	0.204
Edmonton	Canada	North America	0.447	0.350	0.203
Fortaleza	Brazil	South America	0.500	0.274	0.226
Foshan	China, Guangdong	Asia	0.721	0.129	0.150
Frankfurt	Germany	Europe	0.597	0.231	0.172
Freetown	Sierra Leone	Africa	0.299	0.326	0.375
Geneva	Switzerland	Europe	0.416	0.299	0.285
Glasgow	United Kingdom	Europe	0.463	0.302	0.234
Guadalajara	Mexico	Central America	0.479	0.294	0.226
Guangzhou	China, Guangdong	Asia	0.577	0.255	0.168
Guatemala City	Guatemala	Central America	0.503	0.277	0.219
Guayaquil	Ecuador	South America	0.403	0.342	0.255
Haerbin	China, Heilongjiang	Asia	0.607	0.264	0.129
Hamburg	Germany	Europe	0.456	0.316	0.228
Hangzhou	China, Zhejiang	Asia	0.585	0.235	0.180
Hanoi	Vietnam	Asia	0.502	0.281	0.217
Harare	Zimbabwe	Africa	0.418	0.279	0.303
Hartford	United States, CT	North America	0.419	0.393	0.188
Havana	Cuba	The Caribbean	0.379	0.337	0.284
Helsinki	Finland	Europe	0.543	0.271	0.186
Ho Chi Minh City	Vietnam	Asia	0.575	0.243	0.182
Hong Kong	China	Asia	0.465	0.291	0.244
Honolulu	United States, HI	North America	0.519	0.295	0.186
Houston	United States, TX	North America	0.419	0.430	0.150
Hyderabad	India	Asia	0.478	0.312	0.210
Indianapolis	United States, IN	North America	0.391	0.425	0.184
Islamabad	Pakistan	Asia	0.362	0.219	0.419
Istanbul	Turkey	Europe	0.413	0.315	0.272
Jaipur	India	Asia	0.484	0.313	0.203
Jakarta	Indonesia	Asia	0.466	0.308	0.226
Jeddah	Saudi Arabia	Middle East	0.514	0.291	0.195
Jerusalem	Israel	Middle East	0.295	0.348	0.356
Jinan	China, Shandong	Asia	0.501	0.281	0.217
Johannesburg	South Africa	Africa	0.495	0.301	0.204
Kabul	Afghanistan	Asia	0.255	0.281	0.464
Kampala	Uganda	Africa	0.451	0.268	0.281
Kano	Nigeria	Africa	0.345	0.256	0.400

(continued)

City	Country	Continent	Economic keywords	Cultural keywords	Political keywords
Kansas City	United States, MO	North America	0.479	0.320	0.200
Karachi	Pakistan	Asia	0.484	0.247	0.269
Khartoum	Sudan	Africa	0.301	0.238	0.461
Kiev	Ukraine	Europe	0.484	0.283	0.234
Kinshasa	Congo (Dem. Rep. of)	Africa	0.313	0.240	0.447
Kobe	Japan	Asia	0.438	0.305	0.257
Kolkata	India	Asia	0.494	0.315	0.191
Krakow	Poland	Europe	0.470	0.290	0.240
Kuala Lumpur	Malaysia	Asia	0.603	0.229	0.169
Kunming	China, Yunnan	Asia	0.445	0.323	0.232
Kuwait City	Kuwait	Middle East	0.544	0.221	0.235
Kyoto	Japan	Asia	0.370	0.422	0.208
La Paz	Bolivia	South America	0.613	0.241	0.146
Lagos	Nigeria	Africa	0.603	0.220	0.178
Lahore	Pakistan	Asia	0.340	0.341	0.319
Las Vegas	United States, NV	North America	0.476	0.333	0.191
Lille	France	Europe	0.504	0.262	0.234
Lima	Peru	South America	0.392	0.363	0.245
Lisbon	Portugal	Europe	0.378	0.395	0.227
Liverpool	United Kingdom	Europe	0.445	0.330	0.226
London	United Kingdom	Europe	0.531	0.318	0.151
Los Angeles	United States, CA	North America	0.442	0.380	0.178
Luanda	Angola	Africa	0.465	0.248	0.286
Lucknow	India	Asia	0.504	0.260	0.236
Lusaka	Zambia	Africa	0.471	0.257	0.271
Luxembourg	Luxembourg	Europe	0.613	0.150	0.237
Lyons	France	Europe	0.366	0.373	0.261
Madrid	Spain	Europe	0.418	0.388	0.193
Managua	Nicaragua	Central America	0.422	0.335	0.243
Manaus	Brazil	South America	0.455	0.314	0.232
Manchester	United Kingdom	Europe	0.503	0.330	0.167
Manila	Philippines	Asia	0.364	0.369	0.267
Maputo	Mozambique	Africa	0.436	0.214	0.350
Marseille	France	Europe	0.495	0.257	0.248
Medan	Indonesia	Asia	0.457	0.259	0.285
Medellin	Colombia	South America	0.474	0.261	0.265
Melbourne	Australia	Oceania	0.482	0.345	0.173
Mexico City	Mexico	Central America	0.429	0.318	0.254
Miami	United States, FL	North America	0.426	0.383	0.191
Milan	Italy	Europe	0.485	0.326	0.189
Minneapolis	United States, MN	North America	0.391	0.425	0.185
Minsk	Belarus	Europe	0.442	0.273	0.286

(continued)

City	Country	Continent	Economic keywords	Cultural keywords	Political keywords
Mombasa	Kenya	Africa	0.448	0.310	0.242
Monrovia	Liberia	Africa	0.401	0.245	0.354
Monterrey	Mexico	Central America	0.437	0.266	0.297
Montevideo	Uruguay	South America	0.440	0.274	0.286
Montreal	Canada	North America	0.460	0.351	0.190
Moscow	Russia	Europe	0.452	0.298	0.250
Mumbai	India	Asia	0.501	0.332	0.168
Munich	Germany	Europe	0.491	0.315	0.194
Nagoya	Japan	Asia	0.456	0.313	0.232
Nairobi	Kenya	Africa	0.481	0.248	0.270
Nanjing	China, Jiangsu	Asia	0.471	0.280	0.249
Naples	Italy	Europe	0.456	0.312	0.232
New Delhi	India	Asia	0.438	0.330	0.233
New Orleans	United States, LA	North America	0.390	0.393	0.217
New York	United States, NY	North America	0.425	0.364	0.210
Newcastle	United Kingdom	Europe	0.494	0.302	0.204
Omaha	United States, NE	North America	0.428	0.380	0.191
Osaka	Japan	Asia	0.489	0.294	0.217
Oslo	Norway	Europe	0.430	0.278	0.292
Ottawa	Canada	North America	0.427	0.382	0.191
Palermo	Italy	Europe	0.502	0.281	0.217
Panama City	Panama	Central America	0.472	0.327	0.200
Paris	France	Europe	0.413	0.364	0.223
Perth	Australia	Oceania	0.544	0.298	0.158
Philadelphia	United States, PA	North America	0.424	0.394	0.182
Phoenix	United States, AZ	North America	0.457	0.352	0.191
Pittsburgh	United States, PA	North America	0.431	0.361	0.208
Port-au-Prince	Haiti	The Caribbean	0.365	0.287	0.348
Portland	United States, OR	North America	0.371	0.437	0.192
Porto Alegre	Brazil	South America	0.362	0.378	0.260
Prague	Czech Republic	Europe	0.473	0.331	0.196
Pretoria	South Africa	Africa	0.444	0.303	0.254
Pune	India	Asia	0.575	0.238	0.187
Pyongyang	Korea (D. P. R. of)	Asia	0.304	0.261	0.435
Qingdao	China, Shandong	Asia	0.554	0.267	0.179
Quito	Ecuador	South America	0.388	0.388	0.224
Rabat	Morocco	Africa	0.333	0.401	0.266
Rawalpindi	Pakistan	Asia	0.439	0.255	0.306
Recife	Brazil	South America	0.427	0.328	0.245
Richmond	United States, VA	North America	0.409	0.367	0.224
Rio de Janeiro	Brazil	South America	0.468	0.315	0.216
Riyadh	Saudi Arabia	Middle East	0.539	0.264	0.197

(continued)

City	Country	Continent	Economic keywords	Cultural keywords	Political keywords
Rochester	United States, NY	North America	0.468	0.344	0.188
Rome	Italy	Europe	0.388	0.361	0.251
Rotterdam	Netherlands	Europe	0.397	0.384	0.219
Sacramento	United States, CA	North America	0.475	0.332	0.193
Saint Petersburg	Russia	Europe	0.424	0.373	0.203
Salvador	Brazil	South America	0.491	0.306	0.204
San Diego	United States, CA	North America	0.407	0.365	0.228
San Francisco	United States, CA	North America	0.470	0.338	0.192
San Jose	United States, CA	North America	0.416	0.383	0.201
San José	Costa Rica	Central America	0.459	0.358	0.183
San Salvador	El Salvador	Central America	0.534	0.295	0.170
Santiago	Chile	South America	0.441	0.269	0.290
Santo Domingo	Dominican Republic	The Caribbean	0.489	0.305	0.206
São Paulo	Brazil	South America	0.446	0.300	0.254
Seattle	United States, WA	North America	0.578	0.241	0.181
Seoul	Korea (Rep. of)	Asia	0.435	0.376	0.189
Shanghai	China	Asia	0.366	0.400	0.234
Shantou	China, Guangdong	Asia	0.652	0.216	0.131
Shenyang	China, Liaoning	Asia	0.644	0.176	0.180
Shenzhen	China, Guangdong	Asia	0.462	0.306	0.233
Singapore	Singapore	Asia	0.666	0.176	0.158
Sofia	Bulgaria	Europe	0.506	0.300	0.194
St. Louis	United States, MO	North America	0.455	0.303	0.242
Stockholm	Sweden	Europe	0.434	0.298	0.269
Stuttgart	Germany	Europe	0.437	0.332	0.231
Surat	India	Asia	0.548	0.273	0.179
Suzhou	China, Jiangsu	Asia	0.413	0.377	0.210
Sydney	Australia	Oceania	0.506	0.335	0.159
Taipei	China (Rep. of; Taiwan)	Asia	0.489	0.282	0.229
Taiyuan	China, Shanxi	Asia	0.513	0.257	0.229
Tampa	United States, FL	North America	0.474	0.333	0.193
Tashkent	Uzbekistan	Asia	0.504	0.250	0.246
Tbilisi	Georgia	Asia	0.363	0.302	0.334
Tegucigalpa	Honduras	Central America	0.396	0.288	0.316
Tehran	Iran	Middle East	0.294	0.310	0.396
Tel Aviv	Israel	Middle East	0.439	0.297	0.264
The Hague	Netherlands	Europe	0.413	0.289	0.298
Tianjin	China	Asia	0.534	0.244	0.221
Tijuana	Mexico	Central America	0.476	0.274	0.250
Tokyo	Japan	Asia	0.536	0.309	0.155
Toronto	Canada	North America	0.487	0.338	0.176
Tripoli	Libya	Middle East	0.292	0.264	0.444
Tunis	Tunisia	Africa	0.437	0.257	0.305

(continued)

City	Country	Continent	Economic keywords	Cultural keywords	Political keywords
Turin	Italy	Europe	0.390	0.367	0.243
Ulan Bator	Mongolia	Asia	0.411	0.322	0.267
Valencia	Spain	Europe	0.453	0.336	0.211
Vancouver	Canada	North America	0.421	0.384	0.195
Vienna	Austria	Europe	0.403	0.353	0.244
Warsaw	Poland	Europe	0.414	0.343	0.243
Washington	United States, DC	North America	0.373	0.358	0.269
Wuhan	China, Hubei	Asia	0.471	0.289	0.240
Wuxi	China, Jiangsu	Asia	0.512	0.305	0.183
Xiamen	China, Fujian	Asia	0.618	0.192	0.190
Xi'an	China, Shaanxi	Asia	0.443	0.314	0.243
Yangon	Myanmar	Asia	0.447	0.274	0.278
Yaoundé	Cameroon	Africa	0.436	0.273	0.292
Yerevan	Armenia	Asia	0.402	0.276	0.323
Zhengzhou	China, Henan	Asia	0.702	0.173	0.126
Zurich	Switzerland	Europe	0.558	0.264	0.178

References

Andersson, A. E., & Andersson, D. E. (Eds.). (2000). *Gateways to the global economy*. Northampton: Edward Elgar.

Anholt, S. (2010). *Places: Identity, image and reputation*. New York: Palgrave MacMillan.

Balmer, J. M. T. (2010). Explicating corporate brands and their management: Reflections and directions from 1995. *Journal of Brand Management, 18*, 180–196.

Barnes, R. (Producer). (2010, February 20). *The virtual revolution (episode 4: Homo Interneticus?)* [Television broadcast]. London: BBC.

Beaverstock, J. V., Smith, R. G., & Taylor, P. J. (1999). A roster of world cities. *Cities, 16*, 445–458.

Büttcher, S., Clarke, C. L. A., & Cormack, G. V. (2010). *Information retrieval: Implementing and evaluating search engines*. Cambridge: The MIT Press.

Cai, A. (2002). Cooperative branding for rural destinations. *Annals of Tourism Research, 29*(3), 720–742.

Choi, S., Lehto, X. Y., & Morrison, A. M. (2007). Destination image representation on the web: Content analysis of Macau travel related websites. *Tourism Management, 28*, 118–129.

Clark, V. (2004). *SAS/STAT 9.1: User's guide*. Cary: SAS Publishing.

Clark, G. (2006, November). City marketing and economic development. In *Paper presented at the International City Marketing Summit, Madrid, Spain*. Available at: http://www.madrid.es/UnidadWeb/Contenidos/EspecialInformativo/RelacInternac/MadridGlobal/Ficheros/InformesGenerales/Greg_Clark.pdf. Accessed 12 Jan 2015.

Cohen, R. B. (1981). The new international division of labour, multi-national corporations and urban hierarchy. In M. Dear & A. J. Scott (Eds.), *Urbanization and urban planning in capitalist society* (pp. 287–317). New York: Methuen.

Conroy, P., & Narula, A. (2010). *A new breed of brand advocates: Social networking redefines consumer engagement*. Deloitte Development LLC.

Deun, K. V., & Delbeke, L. (2000). *Multidimensional scaling. Open and distance learning*. Leuven, Belgium: University of Leuven. Available at: http://www.mathpsyc.uni-bonn.de/doc/delbeke/delbeke.htm. Accessed 1 May 2012.

Economist Intelligence Unit (EIU). (2012). *Hot spots: benchmarking global city competitiveness*. A report from the Economist Intelligence Unit. Available at: http://www.managementthinking.eiu.com/hot-spots.html. Accessed 12 July 2012.

Erdogmus, I. E., Bodur, M., & Yilmaz, C. (2010). International strategies of emerging market firms. *European Journal of Marketing, 44*(9–10), 1410–1436.

Finnie, G. (1998). Wired cities. *Communications Week International, 18*, 19–22.

Florek, M., Insch, A., & Gnoth, J. (2006). City council websites as a means of place brand identity communication. *Place Branding, 2*(4), 276–296.

Friedmann, J. (1986). The world city hypothesis. *Development and Change, 17*, 69–83.

Friedmann, J. (1995). Where we stand: A decade of world city research. In P. L. Knox & P. J. Taylor (Eds.), *World cities in a world-system* (pp. 21–47). Cambridge: Cambridge University Press.

Gallarza, M. G., Saura, I. G., & García, H. C. (2002). Destination image: Towards a conceptual framework. *Annals of Tourism Research, 29*(1), 56–78.

Global Policy Forum. (2014a). *General analysis on globalization of the economy*. Retrieved from http://www.globalpolicy.org/globalization/globalization-of-the-economy2-1/general-analysis-on-globalization-of-the-economy.html.

Global Policy Forum. (2014b). *Globalization of culture*. Retrieved from http://www.globalpolicy.org/globalization/globalization-of-culture.html.

Global Policy Forum. (2014c). *General analysis on globalization of politics*. Retrieved from http://www.globalpolicy.org/globalization/globalization-of-politics/generalanalysis-on-globalization-of-politics.html.

Govers, R., & Go, F. M. (2003). Deconstructing destination image in the information age. *Information Technology & Tourism, 6*(1), 13–29.

Govers, R., & Go, F. M. (2005). Projected destination image online: Website content analysis of pictures and text. *Information Technology & Tourism, 7*(2), 73–89.

Greenhalgh, P. (2008). An examination of business occupier relocation decision making: Distinguishing small and large firm behavior. *Journal of Property Research, 25*(2), 107–126.

Hatch, M. J., & Schultz, M. (2008). *Taking brand initiative: How to align strategy, culture and identity through corporate branding*. San Francisco: Jossey-Bass/Wiley.

Hatch, M. J., & Schultz, M. (2009). Of bricks and brands: From corporate to enterprise branding. *Organizational Dynamics, 38*(2), 117–130.

Heung, V. C. S. (2003). Internet usage by international travelers: Reasons and barriers. *International Journal of Contemporary Hospitality Marketing, 15*(7), 370–378.

Hymer, S. (1972). The multinational corporation and the law of uneven development. In J. Bhagwati (Ed.), *Economics and world order from the 1970s to the 1990s* (pp. 113–140). New York: Macmillan.

Internet World Stats. (2014). *Internet usage statistics – The big picture: World internet users and population stats*. Internet World Stats, 30th June. Available at: http://www.internetworldstats.com/stats.htm. Accessed 2 Feb 2015.

Kavaratzis, M. (2004). From city marketing to city branding: Towards a theoretical framework for developing city brands. *Place Branding, 1*(1), 58–73.

Kavaratzis, M. (2009). Cities and their brands: Lessons from corporate branding. *Place Branding and Public Diplomacy, 5*(1), 26–37.

Kavaratzis, M., & Ashworth, G. J. (2005). City branding: An effective assertion of identity or a transitory marketing trick? *Tijdschrift voor Economische en Sociale Geografie, 96*(5), 506–514.

Kearney, A. T. (2008). *The 2008 global cities index. Executive agenda*. Chicago: A.T. Kearney.

Kearney, A. T. (2010). *The urban elite: The A.T. Kearney global cities index 2010*. Chicago: A.T. Kearney.

Keeling, D. J. (1995). Transport and the world city paradigm. In P. L. Knox & P. J. Taylor (Eds.), *World cities in a world-system* (pp. 115–131). Cambridge: Cambridge University Press.

Knox, P. L. (1995). World cities in a world-system. In P. L. Knox & P. J. Taylor (Eds.), *World cities in a world-system* (pp. 3–20). Cambridge: Cambridge University Press.

Knox, P. L., Agnew, J., & McCarthy, L. (2008). *The geography of the world economy* (5th ed.). London: Hodder Education.

Kolb, B. M. (2006). *Tourism marketing for cities and towns: Using branding and events to attract tourists*. Burlington: Butterworth-Heinemann.

Kwon, J., & Vogt, C. A. (2010). Identifying the role of cognitive, affective, and behavioral components in understanding residents' attitudes toward place marketing. *Journal of Travel Research, 49*(4), 423–435.

Larsen, H. G. (2014). The emerging Shanghai city brand: A netnographic study of image perception among foreigners. *Journal of Destination Marketing & Management, 3*, 18–28.

London Planning Advisory Council. (1991). *London: World city moving into the twenty first century*. London: HMSO.

Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An introduction to information retrieval* (online Ed.). Cambridge: Cambridge University Press. Available at: http://nlp.stanford.edu/IR-book/pdf/irbookprint.pdf.

Mastercard. (2008). *Insights: Worldwide centers of commerce index 2008*. Purchase: Mastercard Worldwide.

Netcraft. (2015). *January 2015 web server survey*. Netcraft. Available at: http://news.netcraft.com/archives/2015/01/15/january-2015-web-server-survey.html. Accessed 2 Feb 2015.

Petrella, R. (1995). A global agora vs. gated city-regions. *New Perspectives*, Winter, 21-22.

Qu, H., Kim, L. H., & Im, H. H. (2011). A model of destination branding: Integrating the concepts of the branding and destination image. *Tourism Management, 32*, 465–476.

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available from http://www.R-project.org/.

Reed, H. C. (1981). *The pre-eminence of international financial centers*. New York: Praeger.

Rimmer, P. J. (1991). The global intelligence corps and world cities: Engineering consultancies on the move. In P. W. Daniels (Ed.), *Services and metropolitan development: International perspectives* (pp. 66–106). London: Routledge.

Sassen, S. (1994). *Cities in a world economy*. Thousand Oaks: Pine Forge Press.

Short, J. R., & Kim, Y. H. (1999). *Globalization and the city*. London: Longman.

Taylor, P. J., & Catalano, G. (2000). *World city network: the basic data (data set 11)*. Globalization and World Cities (GaWC). Available from http://www.lboro.ac.uk/gawc/datasets/da11.html. Accessed 26 June 2012.

Taylor, P. J., Walker, D. R. F., & Beaverstock, J. V. (2002). Firms and their global service networks. In S. Sassen (Ed.), *Global network, linked cities* (pp. 93–115). New York: Routledge.

Thrift, N. (1999). Cities and economic change: Global governance? In J. Allen, D. Massey, & M. Pryke (Eds.), *Unsettling cities* (pp. 283–338). New York: Routledge.

UN DESA, Population Division. (2012). *World urbanization prospects: The 2011 revision* (CD-ROM ed.). United Nations, Department of Economic and Social Affairs, Population Division.

UN DESA, Statistics Division. (2013). *Composition of macro geographical (continental) regions, geographical sub-regions, and selected economic and other groupings*. United Nations Department of Economic and Social Affairs, Statistics Division. Available at: http://unstats.un.org/unsd/methods/m49/m49regin.htm. Accessed 20 June 2016.

W3Techs. (2012). *Content languages for websites (on April 1, 2012)*. W3Tecs, 2nd April. Available at: http://w3techs.com/technologies/overview/content_language/all. Accessed 2 Apr 2012.

Spatial Hedonic Modeling of Housing Prices Using Auxiliary Maps

Branislav Bajat, Milan Kilibarda, Milutin Pejović, and Mileva Samardžić Petrović

Abstract The latest applications of hedonic dwelling price models have included recent advances in spatial analysis that control for spatial dependence and heterogeneity. The study of spatial aspects of hedonic modelling pertains to spatial econometrics, which is relevant to this study because it clearly accounts for the influence and peculiarities related by space in real estate price modeling analysis.

The research reported herein introduces regression-kriging as a geostatistical method for obtaining econometric models in the analysis of real estate. The aim of this study is to compare the efficacy of regression-kriging (RK) with common regression and geographically weighted regression (GWR) methods of econometric modelling.

The spatial predictors, given as raster maps, were used as auxiliary inputs necessary for regression modeling. In addition to standard environmental predictors, some socio-economic data such as distribution, ages and income of inhabitants, were prepared in the same manner enabling their use in a GIS supported environment. Based upon global and local spatial analysis (Moran's indices), we inspected spatial pattern and heterogeneity in model residuals for all considered methods. The obtained results indicate a similar spatial pattern of model residuals for RK and GWR methods. A spatial-econometric hedonic dwelling price model was developed and estimated for the Belgrade metropolitan area based on cross-sectional and georeferenced transaction data.

Keywords Hedonic price model • Spatial econometrics • Geographically weighted regression • Regression kriging

B. Bajat (✉) • M. Kilibarda • M. Pejović • M. Samardžić Petrović
Faculty of Civil Engineering, Department of Geodesy and Geoinformatics, University of Belgrade, Bulevar kralja Aleksandra 73, 11000, Belgrade, Serbia
e-mail: bajat@grf.bg.ac.rs; kili@grf.bg.ac.rs; mpejovic@grf.bg.ac.rs; mimas@grf.bg.ac.rs

J.-C. Thill (ed.), *Spatial Analysis and Location Modeling in Urban and Regional Systems*, Advances in Geographic Information Science,
https://doi.org/10.1007/978-3-642-37896-6_5

Introduction

Nowadays, the residential market is a major component of the overall real estate market. Over time, methods and research related to this field of economics has shifted from classical econometrics to spatial econometrics (Anselin 1988).

The appearance of hedonic price models, derived mostly from Lancaster's (Lancaster 1966) consumer theory and Rosen's (Rosen 1974) model, became the milestone in econometric theory related to the real estate market. A hedonic price model decomposes the price of a good into separate components that determine the price. Basically, the hedonic equation is a regression of expenditures (rents or values) on housing characteristics of the unit that determine that rent or value. Other pricing models related to hedonic price indices include repeat-sales models (Wang and Zorn 1997) or hybrid models, which combine the elements of hedonic price and repeat-sales models (Quigley 1995). Meese and Wallace (1997) provide comprehensive research comparing the advantages and limitations of all mentioned models. The main drawback of the conventional hedonic model is that it is not capable of taking into account of spatial effects on housing prices even when locational variables are taken into consideration.

One of most important issues related to observed data in hedonic modelling is spatial autocorrelation. Basu and Thibodeau (1998) outlined two main reasons for spatial autocorrelation of housing prices. The first reason that housing prices are found to be spatially autocorrelated is that most of the dwellings in neighborhoods were built at the same time with similar structural characteristics such as dwelling size, year built, interior and exterior design features, etc. The second reason that housing prices are found to be spatially autocorrelated is a consequence of sharing the same neighborhood amenities such as proximity to public transportation, schools, markets, etc. Since hedonic house price parameters are usually estimated using ordinary least squares procedures – which assume independent observations from residuals that are spatially autocorrelated – the resulting parameter estimates often produce incorrect confidence intervals for estimated parameters and for predicted values. The importance of spatial relationships was recognized in recent hedonic studies by introducing spatial lag and spatial error models (Anselin and Bera 1998).

This problem could also be solved using spatial statistical techniques like GWR or RK that incorporate the observed spatial relationships between sample data. Geostatistics has become an essential tool in diverse environmental studies performed during the last few decades. It is imposed particularly in the field of spatial data analysis and in the prediction of numerous natural phenomena. Spatial econometrics, geostatistics, and spatial statistics share many similarities since these fields all deal spatial autocorrelation and spatial heterogeneity (Anselin 1999). Increased interest in use of geostatistics has resulted in numerous improvements and modifications that are essentially extensions to the fundamental kriging theory. Extended versions of kriging have been adopted to deal with non-normality (lognormal kriging, disjunctive kriging), while others address nonstationarity, i.e.

varying trend or drift (universal kriging, kriging with external drift, IRF-k kriging and stratified kriging) (McBratney et al. 2000). The common characteristics of all geostatistical applications are that they were initially used for spatial modelling of diverse natural (i.e. non-anthropogenic) phenomena. Although hedonic regression models (Can and Megbolugbe 1997; Kim et al. 2003; Osland 2010) prevail in real estate appraisal applications, the role of geostatistics has increased in importance recently (Dubin 1998; Yoo and Kyriakidis 2009; Fernández-Avilés et al. 2012).

Most of those spatial statistical techniques are already available in most geographical information systems (GIS) operational environments. The use of GIS technology in spatial econometrics studies started in the mid-1990s. Soon after, the advantages of using GIS applications for hedonic price modelling were recognized in a number of studies (Lake et al. 1998; Anselin 1998; Lovett and Bateman 2001).

In this paper, a geostatistical method regression-kriging (RK) is presented as a method for spatial prediction and mapping of housing prices. Although the RK technique has not been extensively used in hedonic price modeling, there are certain examples where forms of RK were applied under different names. In the literature, the terms used for geostatistical methodologies can be confusing due to the different terms used for the same or very similar techniques (Hengl 2009). Yoo and Kyriakidis (2009) used the term area-to-point kriging with external drift (A2PKED). Chica-Olmo (2007) tested the performance of two co-kriging methods for prediction of housing location prices, in which the authors used a heterotropic version of co-kriging that is very similar to RK.

Usually, most input data for hedonic modelling is acquired from multiple listings databases. In this study, both transaction data and explanatory data are organized as auxiliary maps in a raster format. The majority of required spatial data is already available in GIS formats, thereby minimizing the effort required for pre-processing. The results of spatial prediction are produced as a raster GIS layer in the same resolution as input maps; the resulting map may be of interest to appraisers, real-estate companies and governmental agencies, as it provides an overview of location prices.

The performance of an alternative method, Geographically Weighted Regression (GWR), was also examined in this study; this method is intended for the local analysis of relationships in multivariate data sets. The comparison of the proposed approach using GWR reveals RK as the model of choice for spatial prediction of housing prices in combination with auxiliary data.

The results of spatial prediction are exported as a standard raster GIS layer, or in HTML format enabling simple creation of rich interactive Web-based maps. The user-created maps can assist in facilitating communication between appraisers, real-estate companies, governmental agencies, and other interested parties.

The overall objectives of this study were: (1) to compare two alternative regression techniques (OLS in implicit and error model form and GWR) against the RK geostatistical model for evaluating dwelling prices; (2) to compare the performance of all models using global statistics; and (3) to evaluate the performance of all models in terms of spatial distribution and clustering of residuals using local indicators of spatial association (LISA).

The next section contains a brief description of the theoretical foundations of hedonic price models, geographically weighted regression and regression-kriging methodologies, as well as some details about the *R* software environment used in this study. Included in this study is a concise representation of real estate market characteristics in the city of Belgrade as well as the description of the transaction and auxiliary data layers used. The important issues related to multicollinearity and attribute selection are discussed in the following section, along with detailed explanatory data analyses and mapping. The final section concludes the study.

Methodology

Hedonic Price Models

The basic hedonic price function can be represented as (Can and Megbolugbe 1997):

$$\mathrm{Y} = f\left(\mathrm{S}\beta, \mathrm{N}\gamma\right) + \varepsilon \tag{1}$$

where **Y** is a vector of observed housing values; **S** is a matrix of structural characteristics of properties; **N** is a matrix of neighborhood characteristics, including measures of socioeconomic conditions, environmental amenities and public accommodations for area residents; β and γ are the parameter vectors corresponding to **S** and **N**; and ε is vector of random error terms. The basic form of hedonic regression assumes that each parameter is fixed in space, which means that each identified attribute has the same intrinsic contribution throughout the study area. The given formula can be expressed like a common regression function:

$$\mathrm{Y} = \mathrm{X}\beta + \varepsilon \tag{2}$$

where $\mathrm{Y}_{n\times 1}$ represents vector of observed sale prices of n dwellings, $\mathrm{X}_{n\times k}$ is a vector of k explanatory variables characterizing housing units. $\beta_{k\times 1}$ is a vector of unknown coefficients and $\varepsilon_{n\times 1}$ is a vector representing the error term. By using ordinary least squares (OLS), the unknown coefficients are solved as:

$$\widehat{\beta} = \left(\mathrm{X}^{\mathrm{T}}\mathrm{X}\right)^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{Y} \tag{3}$$

The basic assumption for OLS usage is the indepedence of observations, which is often violated due to spatial autocorrelation in data, leading to a biased estimation of standard errors of model parameters and misleading significance tests. The above given regression formula has particular modified versions that are often used for house price modelling: the spatial-lag model (also known as spatial autoregressive model) and spatial error model (Dubin 1988), (Kim et al. 2003), (Osland 2010):

$$\begin{aligned} &\text{Spatial lag model :} && \mathrm{Y} = \mathrm{X}\beta + \phi\mathrm{W}\mathrm{Y} + \varepsilon \\ &\text{Spatial error model :} && \mathrm{Y} = \mathrm{X}\beta + \xi + \lambda\mathrm{W}\varepsilon \end{aligned} \tag{4}$$

where ε is the vector of errors terms and W represents the weights matrix that specifies the assumed spatial structure or connections between the observations. The elements of the weights matrix can be based upon contiguity (i.e., shared borders) or distance. The parameter φ is often referred to as the spatial correlation parameter giving the intensity of the dependence between neighboring prices. The term ξ represents random error and λ is a spatial autoregressive parameter.

If data exhibits a spatial lag process the target variable is affected by the values of the target variables in nearby places. The OLS hedonic model omits φwY, and this becomes part of the error which leads to biased parameters estimates (Anselin 1988). If spatial lag is in true functional form (Anselin and Bera 1998), then parameter estimates should be inefficient as well. Nevertheless, some studies showed that although spatially correlated errors are presented in data, non-spatial hedonic models (implicit OLS) may provide results useful for policy analysis (Mueller and Loomis 2008).

Geographically Weighted Regression

The geographically weighted regression is a relatively new method used in spatial modeling and was developed as an alternative format for spatial analysis that is local rather than global in its analytical design (Fotheringham et al. 2002). Although this method is useful in a wide range of applications, its widest practical application still is in the mass assessment of real property (Crespo et al. 2007; Yrigoyen et al. 2008; Hanink et al. 2010). Increasing application of this technique is made possible by GIS database processing tools and publicly available databases on the Internet. GWR represents the extension of a conventional multiple regression framework by addressing the issue of non-stationary processes (Fotheringham et al. 2002):

$$y_i = \beta_0 (u_i, v_i) + \sum_{k=1}^{m} \beta_k (u_i, v_i) x_{ik} + \varepsilon_i, \quad i = 1, \ldots, n \tag{5}$$

where (u_i,v_i) are the coordinates for i-th point; $\beta_k(u_i,v_i)$ are the realizations of continuous function $\beta_k(u,v)$ at the same location; $x_{i1}, x_{i2}, \ldots x_{im}$ are the explanatory variables at point i; and ε_i is the error term.

$$\widehat{\beta}(i) = \left(\mathrm{X}^T \mathrm{W}(i) \mathrm{X}\right)^{-1} \mathrm{X}^T \mathrm{W}(i) \mathrm{y} \tag{6}$$

W(i) is a matrix of weights for particular location i, such that observations nearer to i are given greater weight than observations further away.

$$\mathrm{W}(i) = diag\,[w_{i1}, w_{i2}, \ldots, w_{in}] \tag{7}$$

w_{ij} is the weight related to data point j for the estimate of the local parameters at location i. Several types of parameterized weight functions may be used. A common choice also used here is the Gaussian distance-decay curve has the form:

$$w_{ij} = \exp\left[-\frac{d_{ij}^2}{2b^2}\right] \tag{8}$$

where d_{ij} is a distance between location i and location j, while parameter b is a range to be determined. The weight's value would decay gradually with distance to the extent that when $d_{ij} = b$, the weight reaches the value of 0.5. In the event that the spatial distribution of sampled variables is spatially homogeneous, this parameter is taken as a constant value. However, a spatially variable (adaptive) parameter of the range should be used in the event that the spatial distribution of the variables is heterogeneous.

To calculate the parameters associated with a weighting function, such as the bandwidth and *Nth* nearest neighbors considered, the GWR methodology utilizes a calibration process. This calculates the parameter so as to form an appropriate trade-off between bias and standard error in the prediction of the overall model. Commonly used approaches include minimizing the cross-validation scores (CV) or Akaike Information Criterion (AIC_c) (Fotheringham et al. 2002). The optimal values for b and N reported here were obtained by minimizing the cross-validation scores.

Regression-Kriging

Regression-kriging (RK) is a geostatistical technique that combines the regression of the target variable on explanatory variables with kriging of the regression residuals. In the literature, this interpolation technique is termed as kriging with external drift (KED). Let the measured values of the target variables be symbolized as $Y(s_i)$, $i = 1, \ldots n$, where s_i represents spatial location and n number of realized measurements. The system of equations from which the estimated values of target variables $\hat{Y}(s_0)$ are obtained is:

$$\begin{aligned} &\hat{Y}(s_0) = \hat{m}(s_0) + \hat{e}(s_0) \\ &\hat{Y}(s_0) = \sum_{k=0}^{p} \hat{\beta}_k \cdot q_k(s_0) + \sum_{i=1}^{n} w_i(s_0) \cdot e(s_i); \\ &q_0(s_0) = 1 \end{aligned} \tag{9}$$

where $\hat{m}(s_0)$ is the fitted deterministic part; $\hat{e}(s_0)$ is the interpolated residual; $\hat{\beta}_k$ is estimated deterministic model coefficient; and w_i represents ordinary kriging weights resolved by the spatial structure of residuals $e(s_i)$ (Hengl 2009). The essential difference between RK and KED is explained as follows: while KED

weights are solved within extended matrix taking into consideration trend and residuals at the same time, the RK drift model coefficients are computed separately and the residuals interpolated by ordinary kriging (OK) are summed back to the drift estimates using simple summing of predicted drift and residual surfaces. Despite the differences in the computational steps used, the resulting predictions and prediction variances are the same, given the same point set, auxiliary variables, regression functional form, and regression fitting method (Hengl et al. 2007). Regression coefficients $\hat{\beta}_k$ could be obtained by different fitting methods such as ordinary least squares (OLS) or generalized least squares (GLS) which is more often recommended:

$$\hat{\beta}_{GLS} = \left(\mathrm{q}^{\mathrm{T}} \cdot \mathrm{C}^{-1} \cdot \mathrm{q}\right)^{-1} \cdot \mathrm{q}^{\mathrm{T}} \cdot \mathrm{C}^{-1} \cdot \mathrm{Y} \tag{10}$$

where $\hat{\beta}_{GLS}$ is the vector of estimated regression coefficients; **C** is the covariance matrix of residuals; **q** is the matrix of predictors at the sampling location; and **Y** is the vector of measured values of target variable. The estimated $\hat{\beta}_{GLS}$ coefficients basically present a special case of geographically weighted regression coefficients. The predicted variable value $\hat{Y}(s_0)$ at the location s_0, obtained by regression-kriging is commonly written in matrix notation:

$$\hat{Y}_{RK}(\mathrm{s}_0) = \mathrm{q}_0^{\mathrm{T}} \cdot \hat{\beta}_{GLS} + \lambda_0^{\mathrm{T}} \cdot \left(y - \mathrm{q} \cdot \hat{\beta}_{GLS}\right) \tag{11}$$

where $\mathbf{q_0}$ is the vector of $p + 1$ predictors and λ_0 is the vector of n kriging weights used for interpolation of residuals.

RK explicitly separates trend estimation from residual interpolation, thereby allowing the use of arbitrarily complex forms of regression rather than using the simple linear techniques that can be used with KED. Besides, RK allows the separate interpretation of the two interpolated components, which reinforces the advantage of the RK approach (Hengl et al. 2007).

Spatial Residual Analysis

We used global and local analysis methods (i.e., global and local Moran's I indices) to evaluate spatial distribution and heterogeneity in model residuals. Spatial autocorrelation of residuals was calculated by *Moran's I* statistic (O'Sullivan and Unwin 2003):

$$MI = \frac{n}{\sum_{i=1}^{n}(e_i - \overline{e})^2} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(h)\,(e_i - \overline{e})\left(e_j - \overline{e}\right)}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \tag{12}$$

Where e_i and e_j are calculated values of the residuals at the locations i and j; the term $\bar{e}$ is a mean value; and w_{ij} is the spatial weight within a given distance or bandwidth that was determined as $h = 750$ m, according to the variogram of the OLS model residuals (Fig. 5). The weighted matrix was row standardized in order to overcome the problem of uneven spatial distribution of observations, which causes up-weighted values for locations with more neighbors. The value of the Moran's I statistic ranges from near $+1$ indicating clustering of the e values to near -1 indicating dispersed pattern of the e values. In the Global Moran's I statistic, the results of the analysis are always interpreted within the context of its null hypothesis, which states that the variable (residuals in our case) being analyzed is randomly distributed among the locations in our study area; or better said, the spatial processes promoting the observed pattern of values is random chance. The results of Moran's I statistic with significant p-values and positive Z-scores indicates spatially clustered datasets. However, at the same time, significant p-values and negative Z-scores depict that spatial pattern is more spatially dispersed than what would be expected to result from random spatial processes.

Local Indicators of Spatial Association (LISA) (Anselin 1995) that are based on the local Moran's I test enable the assessment of significant local spatial clustering around an individual location- thereby providing details of (1) the degree of spatial clustering; (2) an estimate of detailed variations of clustering in the locally defined area; and (3) identification of the locations of the spatial clusters. The local version of Moran's I at location i is given by:

$$MI_i = \frac{(e_i - \bar{e})}{S_i^2} \sum_{j=1, j\neq i}^{n} w_{ij}(h)\left(e_j - \bar{e}\right) \tag{13}$$

where w_{ij} is the spatial weight within a given distance or bandwidth (h), as stated in Eq. 12 and S_i^2 is calculated as:

$$S_i^2 = \frac{\sum_{j=1, j\neq i}^{n} w_{ij}}{n-1} - \bar{e}^2 \tag{14}$$

This local indicator represents a disaggregated measure of autocorrelation that depicts the extent to which the residuals for particular areal locations are similar to, or differ from, neighboring locations. The *Local Moran's I* statistic is used to detect possible non-stationarity of the data – i.e., a clustered pattern – among sub-regions. A positive local MI_i indicates a cluster of similar residual values of the same sign around i, while a negative MI_i indicates that the value of the residual at location i has a sign opposite that of its neighbors. If values for e_i and e_j are close to $\bar{e}$, then local MI_i indicates the absence of spatial autocorrelation.

Case Study: Belgrade Metropolitan Area

Belgrade, the capital of the Republic of Serbia, is situated at the confluence of the Sava and Danube Rivers. The administrative boundary encompasses a 3223 km^2 area of nearly 2 million inhabitants. Its territory is divided into 17 municipalities, comprised of 157 settlements. The urban core, an area of 360 km^2, includes the 10 urban municipalities which constitute the study area for this research. According to official census data records, there are approximately 470,000 households with 1,300,000 inhabitants in the study area, which represents metropolitan area of the city (Statistical Office of the Republic of Serbia 2011).

Due to political issues, and the fact that most apartments in large cities were under social ownership, the number of individual transactions was negligible prior to privatization. Between 1991 and 2000, most housing units were privatized. The housing market in Serbia started developing rapidly and housing prices spiked after privatization in the 1990s. Housing prices peaked in 2008, just prior to the present financial crisis. As of the time of publication, this market remains depressed and housing prices are falling along with declining construction activity.

According to local property experts, the housing prices are expected to fall by another 20% in the year 2013. Average housing prices in Belgrade vary considerably depending upon apartment location and structural characteristics; in some cases, average prices may be 60% higher in one Serbian municipality or town, as compared to another (Cvijanović 2006). However, housing studies regarding Belgrade and other cities in Serbia are limited because of the short history of the free housing market.

The original dataset used in this study consists of 747 records of apartment transactions referring to real estate sales in the year 2010. The dataset used was provided by several real estate trading companies because a unique database of real estate transactions has not been compiled for the Belgrade area. Selected transaction records include total transaction value (EUR), covered flats size (m^2) and addresses. Since additional information regarding internal living space, age of building, and parking amenities was available only for some records, these attributes were not included in the analysis. A geographic information system (GIS) was used to match street addresses of the transactions with the official dataset of building geographic coordinates in order to geocode observations into the study area (Fig. 1).

Target Variable

Housing price (expressed in EUR/m^2) is the target variable to be spatially evaluated. A study area map with corresponding housing price ranges is shown in Fig. 1 and the descriptive statistics of housing price are displayed in Fig. 2. The distribution is positively skewed, with a mean of EUR 1633 EUR/m^2.

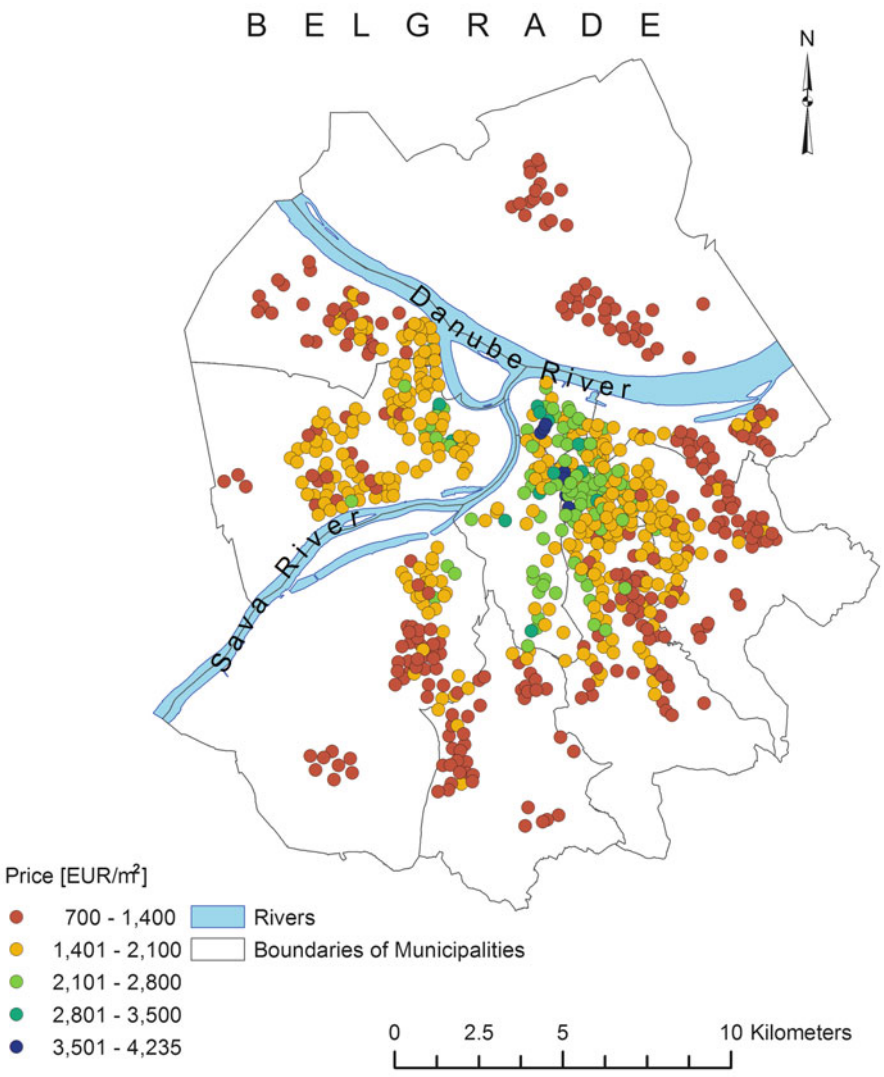

Fig. 1 Locations and values of observed transactions in Belgrade in the year 2010

Mapped housing price values (Fig. 1) indicate that the most expensive apartments are located in the city center, while the prices are lower in peripheral areas. The variability of the prices is more pronounced in the central urban core, where the price depends upon the specific location of the building, the quality of interior infrastructure and other amenities (Cvijanović 2006).

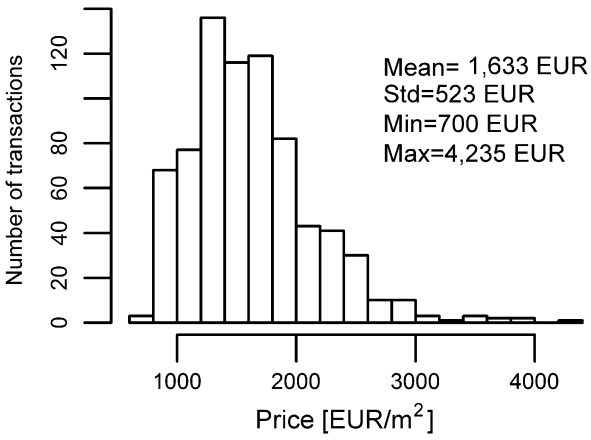

Fig. 2 Histogram of observed housing prices

Exogenous Variables

In applying the hedonic price model to the real estate market, the determinants of housing prices can be divided into four groups (Lake et al. 1998): (1) structural variables (e.g., age, the number of rooms in each house, etc.); (2) accessibility variables (e.g., the proximity of schools, bus routes, railway stations, shops, parks, and the Central Business District); (3) neighborhood variables (e.g., local unemployment rates); (4) environmental variables (e.g., road noise and visibility impact). The accessibility characteristics primarily consist of site-related factors.

In this study we were confined to predictors that incorporate accessibility, neighborhood and environmental variables (Table 1). The explanatory variables referring to accessibility and environment could be considered as spatial determinants referred to as the distance variables (Koramaz and Dokmeci 2012). They were arranged as input maps/grids of 20 m resolution by using a proximity function within the SAGA (System for Automated Geoscientific Analyses) GIS environment (http://www.saga-gis.org/). The values assigned to grid cells are calculated Euclidean distances between each cell and the input features (roads, schools, parks, etc.); these values are included in each cell in the grid map. Researchers have previously used travel times to the central business district and to highways, shopping centers, schools, and universities, instead of Euclidean distances to these features, in order to compute accessibility measures for hedonic models (Des Rosiers et al. 2000; Adair et al. 2000). Two neighborhood variables (illiteracy and income) are based upon census data (Statistical Office of the Republic of Serbia 2011). The illiteracy layer/map was generated as a factorial variable referenced to each municipality, where each cell represents the proportion of illiterate inhabitants in a particular municipality in regard to the whole city of Belgrade. The income variable represents the average income in each municipality so that every grid cell within each municipality area has the same value.

Table 1 The list of explanatory variables used in study with corresponding VIF (variance inflation factor) values

Variables	Description	Type	VIF
dist_ Airport	Prox. (Euclidean distance) to airport	Accessibility	4.138
dist_ Highway	Prox. to highway	Environmental	7.022*
dist_ Culture	Prox. to museums, theatres	Accessibility	4.708
dist_ Main road	Prox. to main roads/ boulevards	Environmental	2.951
dist_ Sciences facilit	Prox. to University/science facilities	Accessibility	6.018*
dist_ Schools	Prox. to elementary/high schools	Accessibility	2.045
dist_ Parks	Prox. to parks/playgrounds	Accessibility	8.110*
dist_ Market	Prox. to green markets	Accessibility	3.201
dist_ Industry	Prox. to industrial objects	Environmental	1.567
dist_ River	Prox. to river banks	Environmental	7.264*
dist_ Recreation	Prox. to big green areas/forest	Accessibility	1.458
dist_ Sport	Prox. to sport stadiums	Accessibility	1.903
dist_ public transport	Prox. to station of public transport	Accessibility	1.470
dist_ Shopping	Prox. to Shopping centers	Accessibility	9.458*
dist_ Main streets	Prox. to main streets	Accessibility	3.292
dist_ Religious	Prox. to Religious facility	Accessibility	1.498
dist_ Kindergarten	Prox. to kindergarten	Accessibility	1.837
dist_ Hospital	Prox. to ambulance/hospitals	Accessibility	1.813
dist_ Railway	Prox. to railway	Environmental	3.985
Illiteracy	Percentage of inhabitants who are illiterate	Neighborhood	2.688
Income	Average income in municipality	Neighborhood	5.724*

*VIF > 5; variable indicates high multicollinearity

R Language Environment

The *R* open source language (Development Core Team 2008) contains the base system that allows statistical computation, linear algebra computation, graphics creation and the like. Most of the computations related to spatial autocorrelation estimation and model testing are utilized through the *spdep* package (Bivand et al. 2008). It includes a number of features such as tools for the creation of spatial weights matrix, a collection of tests for spatial autocorrelation, including global and local Moran's I and and Getis/Ord G statistics, and functions for estimating spatial simultaneous autoregressive (SAR) lag and error models. There is also a set of developed *R* packages that are especially interesting for geoscientists.

All utilized methods were implemented using the open-source R statistical computing environment with *gstat* and *spgwr* packages (Bivand et al. 2008) intended for modelling and prediction, as well as the *sp.* package which provides classes and methods for dealing with spatial data in *R* (Pebesma 2004). The results obtained in *R* can easily be converted into any of the standard GIS formats, which enables the manipulation and analysis of the results in commercial GIS packages afterwards.

There are also several open-source and commercially available software packages with associated GWR methods. Unfortunately, GWR is a time-consuming computational procedure, especially in the case of large data sets. However, it is possible to solve this problem by using grid computing (Harris et al. 2010). The *spgwr* package used in *R* environment has also been adopted for use on grid based systems.

The recently developed *R* package *plotGoogleMaps (*Kilibarda and Bajat 2012*)*, designed to automatically create web maps by combining users' data and Google Maps layers as a base map, map was also used in this study to improve insight into predicted housing price layers.

Model Fitting and Evaluation

Before regression analyses were performed, an indicator of multicollinearity between exogenous variables was examined. The variance inflation factor (VIF) test (Fox 2008) indicates the presence of multicollinearity between predictors (Table 1). Principal components analysis (PCA) is often used with the aim of transforming a dataset with many correlated variables into a dataset consisting of a smaller number of uncorrelated variables, known as principal components (PCs) (Lake et al. 1998). PCA can assist when multicollinearity exists between predictor variables and can also assist in collapsing a large set of variables into a more efficient set of uncorrelated components. However, the main drawback of using PCA is that the newly generated components complicate the interpretation of the influence of the original variables.

After the performing PCA upon the set of explanatory variables, 21 PC predictors were derived. Because the predictors are now known to be independent, we can reduce their number by using step-wise regression (Draper and Smith 1981) based on AIC or *t-statistics* (Table 2).

By looking at the specific *t*-values of coefficients, we can infer which predictors are significant and useful in further analysis. Because the target variable is positively skewed, we used lognormal transformation to improve linearity. The results obtained show that eight predictors are highly significant, seven are marginally significant and six predictors can be removed from the list.

In order to check for the presence of spatial correlation we performed the Lagrange Multiplier test (Anselin 1988; Bivand et al. 2008).This test is often used in spatial econometrics as a supplement to Moran's I and is designed to determine whether the input data generate a spatial lag (LMlag) or spatial error (LMerr) model (Eq. 4). In addition, the robust LM error (RLMerr) statistics tests for error dependence in the possible presence of a missing lagged dependent variable while its counterpart, the robust RLMlag statistics test, works the other way round. Based on the obtained test results (Table 3), it appears that the spatial correlation problem is more of the "spatial error type" rather than of the "spatial lag type".

Nevertheless, model fitting statistics indicates that the introduction of the OLS modified ,spatial error data "version (OLS.ERR) could not be of benefit to improve

Table 2 Results of step-wise regression analysis of predictors

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	Signif. code
(Intercept)	7.05202	0.02392	−294.85	<2.00E-16	***
PC1	0.09051	0.00875	10.35	<2.00E-16	***
PC2	0.04272	0.00837	5.1	4.30E-07	***
PC3	0.03479	0.00679	5.12	3.90E-07	***
PC5	−0.04602	0.01229	−3.75	0.00019	***
PC6	0.02117	0.01214	1.74	0.08148	.
PC8	0.03983	0.01304	3.05	0.00234	**
PC9	−0.07971	0.02301	−3.46	0.00056	***
PC10	−0.17372	0.01599	−10.87	<2.00E-16	***
PC11	0.0473	0.02013	2.35	0.01905	*
PC12	−0.07523	0.01819	−4.14	4.00E-05	***
PC13	0.03499	0.01782	1.96	0.05004	.
PC15	−0.05511	0.02258	−2.44	0.01488	*
PC17	0.18078	0.02624	6.89	1.20E-11	***
PC19	0.0357	0.02291	1.56	0.11954	
PC21	0.09761	0.04082	2.39	0.01705	*

[a]Significant codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '•' 0.1 '' 1,
[b]Residual standard error: 0.184 on 731 degrees of freedom,
[c]Multiple R-squared: 0.646, Adjusted R-squared: 0.636,
[d]F-statistic: 88.4 on 15 and 731 DF, p-value: < 2.2e-16.

Table 3 Lagrange Multiplier test on OLS residuals

Test	LMerr	RLMerr	LMlag	RLMlag
Statistic	65.43107	60.40537	6.161991	1.136293
p-value	5.55E-16	7.77E-15	0.013052	0.286437

Table 4 Model fitting statistics for utilized modeling techniques

Model	R^2	SSE	RMSE	F-Test[a]	p.value
OLS	0.56	89077395	349.0801		
OLS.ERR	0.56	90550113	351.9539	0.986253	0.850
RK	0.63	75277050	320.9021	1.182	0.022
GWR	0.61	79413549	329.1511	1.123563	0.111

[a] Hypothesis test for testing the improvement of model fitting over OLS

the performance of implicit OLS (Table 4). Both models had the same R^2 values (0.56), while the error sum of squares (SSE) and the root mean square error (RMSE) values as well as the Fisher statistics for homogeneity of variance indicate better model fitting performance of implicit OLS over the spatial error data OLS model. For that reason we left the implicit OLS model as the benchmark model in further analysis.

Even though quantitative evaluation measures demonstrate the similar performance for RK and GWR methods, the RK method provides slightly better

performance. Both methods achieved improvement in performance, this performance was not significantly improved over the results achieved via standard OLS hedonic modeling.

Results and Discussions

The model's residuals were obtained through the leave one out cross-validation process of all referred spatial models. Table 5 and Fig. 3 show that the model's residuals have similar characteristics. All models demonstrate that the dataset is characterized by positive skewness and large positive kurtosis. Figure 4 shows strong linear relationship within OLS, RK and GWR residuals.

Table 5 Characteristics of model residuals for utilized modeling techniques

	Mean	Std	Skewness	Kurtosis	Min	Median	Max
OLS	29.758	344.267	1.393	3.761	−733.956	−18.767	2020.549
RK	25.219	316.655	1.220	3.547	−885.708	−8.546	1874.712
GWR	31.070	324.786	1.167	3.551	−892.572	−6.445	1915.185

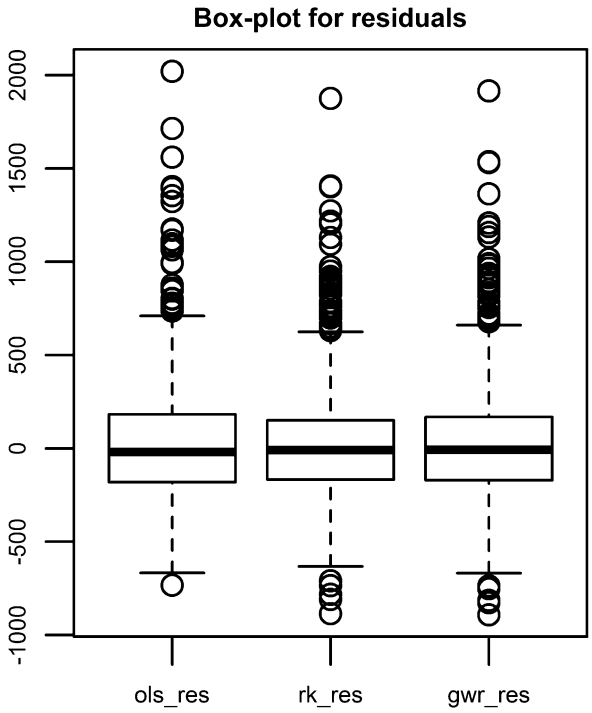

Fig. 3 Box plot of the model's residuals

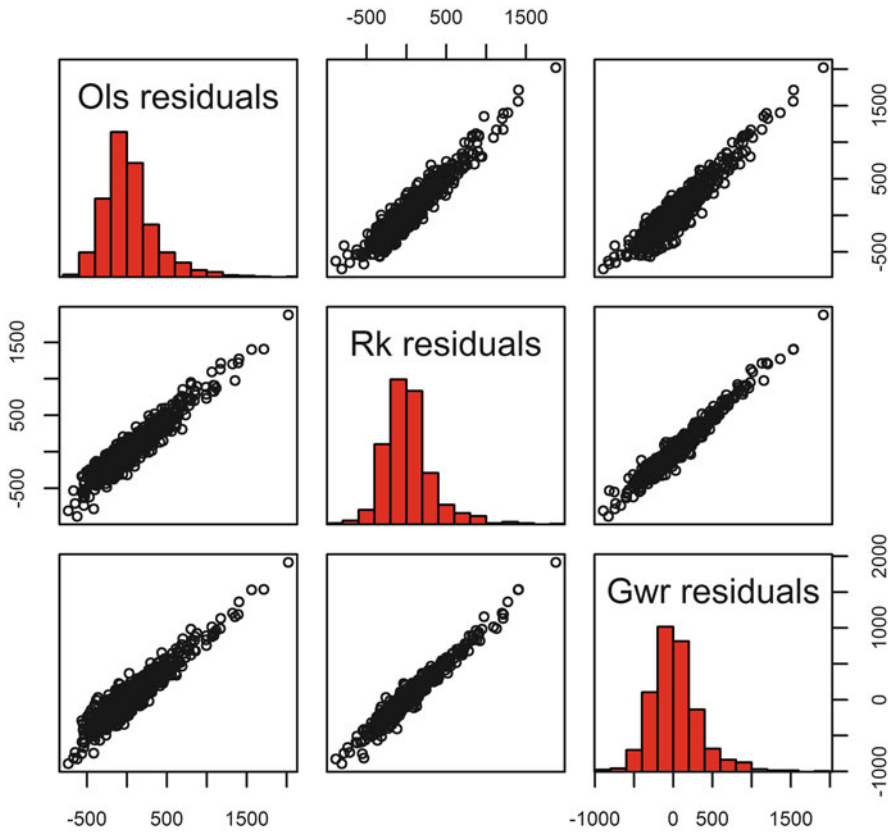

Fig. 4 Matrix plot of residual relationship between referred models

A semivariogram of OLS residuals was created in order to examine the spatial distribution of the model's residuals. Specifically, the semivariogram determines the range of the residuals, i.e. the distance at which the spatial correlations between observations fall to zero (Fig. 5). Based upon this information, we built a spatial (row standardized) weights matrix W (Eqs. 12 and 13) in which all buildings that are located within 750 m from one another are considered as neighbors. At same time the experimental variogram of GWR residuals exhibits practically no autocorrelation.

The global Moran's *I* (Eq. 12) was computed for all three models (Table 6). The global *MI* for the OLS residuals is significantly positive (Z.value > 1.96) indicating that model residuals tend to be similar across the space. In contrast, the global *MI* for RK's and GWR's are significantly negative (Z.value < −1.96), which means their residuals are clustered with dissimilar values (positive are surrounded with negative, and vice versa).

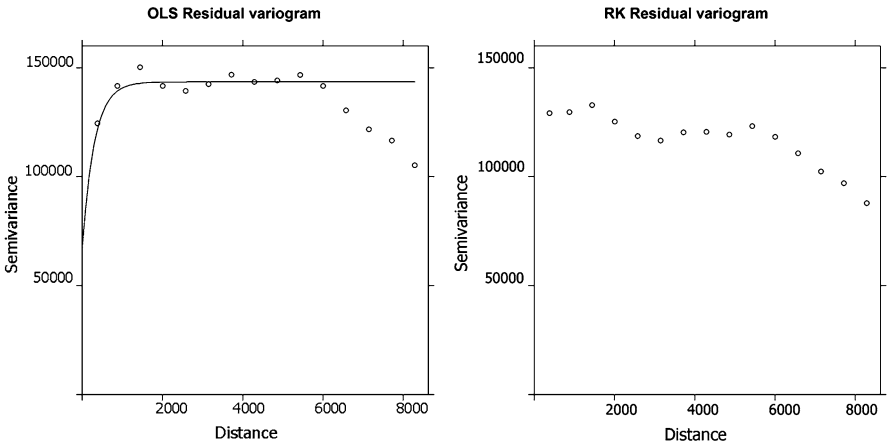

Fig. 5 Semivariograms of the OLS (left) and GWR (right) residuals

Table 6 Global Moran (*MI*) for utilized modeling techniques

Model	MI	Z.value[a]	Z.value[b]
OLS	0.111	6.484	6.468
RK	−0.050	−2.831	−2.824
GWR	−0.041	−2.274	−2.269

[a]Standard normal test based on randomization assumption
[b]Standard normal test based on normal assumption

Table 7 Local Moran indexes (MI_i) for utilized modeling techniques

	Mean	Std	Skewness	Kurtosis	Min	Median	Max
OLS	0.112	0.531	2.534	20.293	−2.702	0.021	4.570
RK	−0.050	0.333	−1.762	29.343	−3.267	−0.008	2.313
GWR	−0.041	0.345	−0.594	26.449	−2.993	−0.007	2.747

The local indicator of spatial association MI_i was used to determine "hot spots" (positive autocorrelation or similarity) and "cold spots" (negative autocorrelation or dissimilarity) of residual values of all utilized models (Anselin 1995). Table 7 indicates that local MI_i of RK and GWR residuals have similar characteristics while MI_i of OLS residuals are considerably different. Obviously, RK and GWR produced more frequently negative local MI_i with stronger linear relationship in contrast to OLS (Fig. 6).

Similar patterns were obtained for calculated Z-values of the local Moran (MI_i) (Table 8).

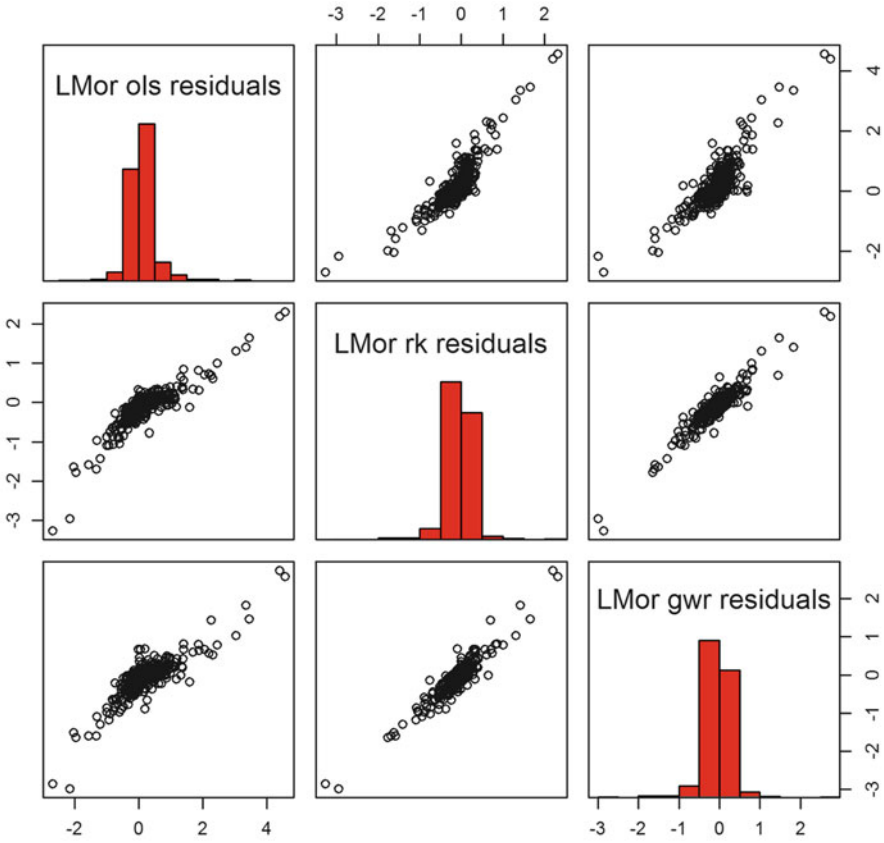

Fig. 6 Matrix plot of local Moran's indexes relationship between referred models

The local Z-values for the local MI_i were also evaluated for the significance level 0.05 ($Z_{\alpha=}1.96$). Table 9 indicates that RK and GWR produced a similar percentage of significant Z-values, where higher percentages among significant values are negative (76% and 71%, respectively). These results indicate the clustering of the model residuals, whereby a large residual tends to be surrounded by smaller neighboring residuals and vice versa when a small residual is surrounded by larger ones. On the other hand the majority of significant positive Z-values indicate opposite behavior of the OLS model (71%) – i.e., it generates more clusters of either positive or negative model residuals. That means that OLS produced sub-areas with underestimated (positive residuals) or overestimated (negative residuals) prediction.

Table 8 Z-values of the local Moran (MI_i) for utilized modeling techniques

	Mean	Std	Skewness	Kurtosis	Min	Median	Max
OLS	0.374	1.841	2.563	18.48	−7.549	0.072	15.324
RK	−0.183	1.106	−1.257	22.32	−10.329	−0.023	7.758
GWR	−0.139	1.166	−0.424	23.29	−10.481	−0.018	8.777

Table 9 Significant Z-values for local Moran's indexes

	n = 747	Among the significant Z-values	
Model	No. of sign. $\|Z\| \geq 1.96$ (%)	$Z \leq -1.96$ (%)	$Z \geq 1.96$ (%)
OLS	94 (12)	27 (29)	67 (71)
RK	45 (6)	34 (76)	11 (24)
GWR	51 (7)	36(71)	15 (29)

Number in parenthesis is the percentage

The visual inspection of the mapped residuals may give us more detailed insight into the performance of the model in the case of particular observations (Figs. 7, 8 and 9). Spatial patterns of Z-values in terms of size, sign and clustering are apparently similar for GWR and RK.

Generated local *MI* Z-values can also be completely mapped by *plotGoogleMaps* package (Kilibarda and Bajat 2012) in HTML format (available at http://osgl.grf.bg.ac.rs/en/materials/hedonic/), which has become a standard medium for cartographic communication.

Finally, based upon the estimated model parameters, we produced raster maps that depict predictions for dwelling prices over the whole case study area. The raster maps were generated in the same resolution (20 m) as input maps/grids (Figs. 10 and 11). Visual inspection of spatially predicted maps clearly emphasizes the limitation of the GWR method (Fig. 10). At first glance, the artefacts (areas with high price values) generated at the peripheral areas (north-western and northern) are easily distinguished on the GWR map. Those artefacts are caused by the spatial pattern of observed transactions (Fig. 1) because most of the transactions are grouped in the inner part of the city and peripheral parts are apparently lacking data. The peripheral parts of the case study area did not have any data on transactions and therefore the predicted values are a result of extrapolation.

Conclusions

This chapter demonstrates how GWR and RK techniques- which are already widely used for spatial prediction in various environmental studies – be useful tools for the prediction and mapping of housing prices. The primary objective of this research was to compare the performance of recently introduced spatial modelling methods in the hedonic prediction of residential property values in city of

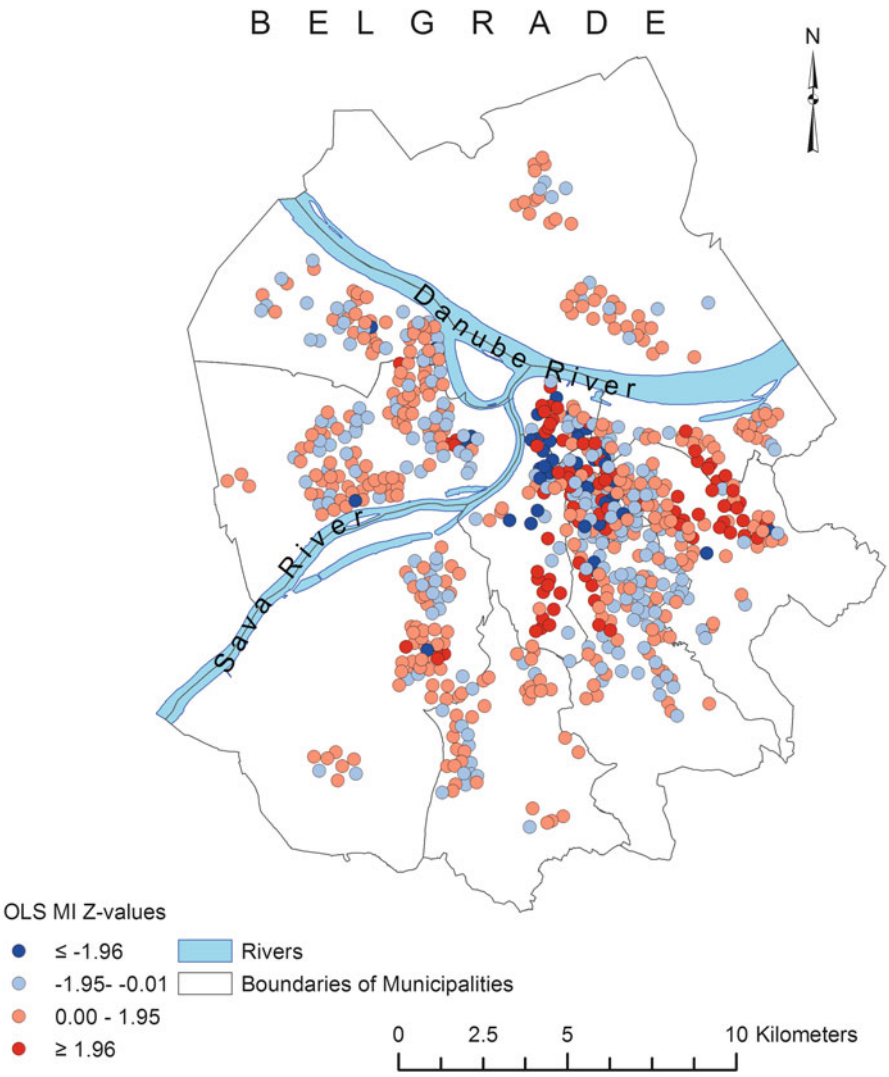

Fig. 7 Mapped local MI values of OLS residuals

Belgrade, based upon examination of spatial distribution and heterogeneity in model residuals. Spatial models including geographically weighted regression (GWR) and regression-kriging (RK) were estimated in contrast to standard hedonic price modelling based upon an ordinary least squares (OLS) technique. Although similar values for quantitative evaluation measures were obtained for by each model, spatial patterns of residuals for the RK and GWR models were found to be different in contrast to the OLS model. The values erroneously predicted by the GWR

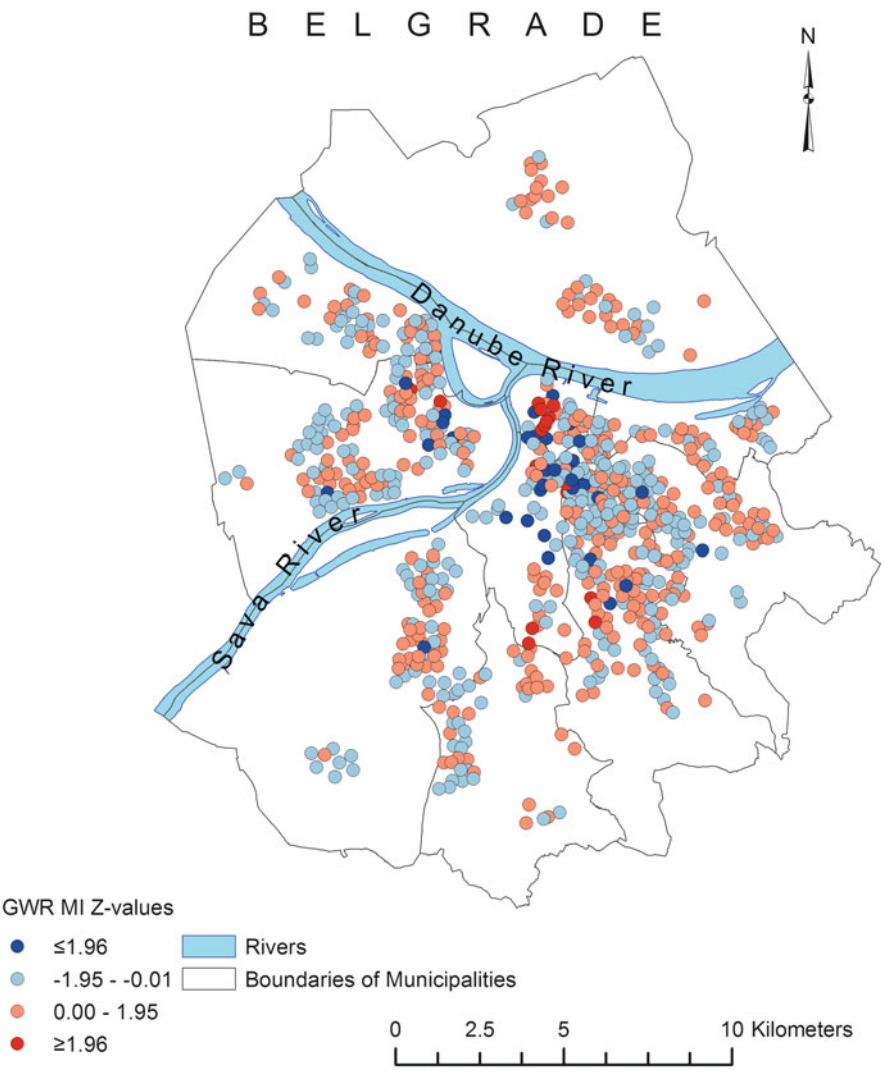

Fig. 8 Mapped local MI values of GWR residuals

model (located in the peripheral part of the city) indicate that this technique cannot correctly handle the samples pattern used in this study.

Coupling this methodology with GIS and the Web 2.0 environment facilitates large-scale housing price appraisal in the framework of collaborative GIS – thereby enabling platforms for evaluation and spatial decision support. The use of a Web-based GIS tools enables authorities to combine the different spa-

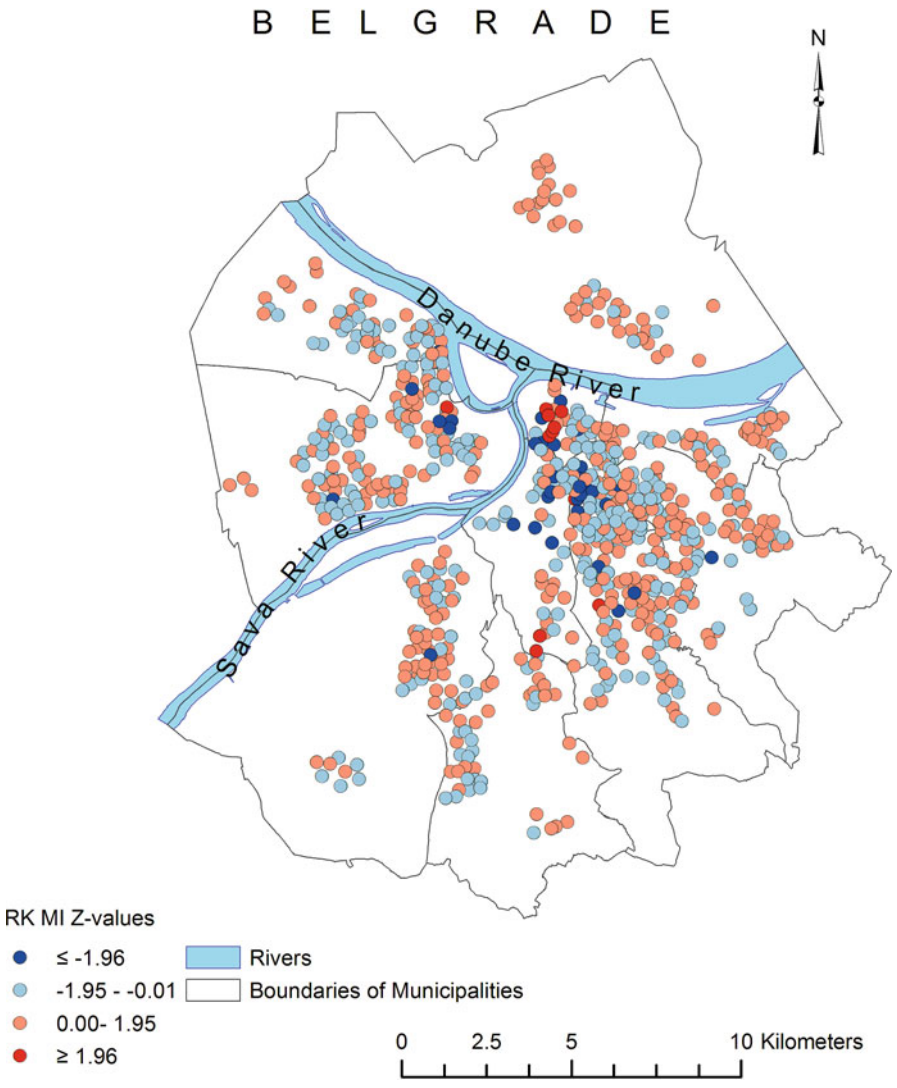

Fig. 9 Mapped local MI values RK residuals

tial layers, particularly socioeconomic datasets provided as raster maps, to spatially model distribution of housing values. This methodology provides a reliable view to spatial distribution of housing price and can be useful in hedonic price modelling.

Further research could focus on the application of spatiotemporal geostatistical techniques in hedonic price modelling. By combining the growing number of transaction database records dating from different time periods with the great variety

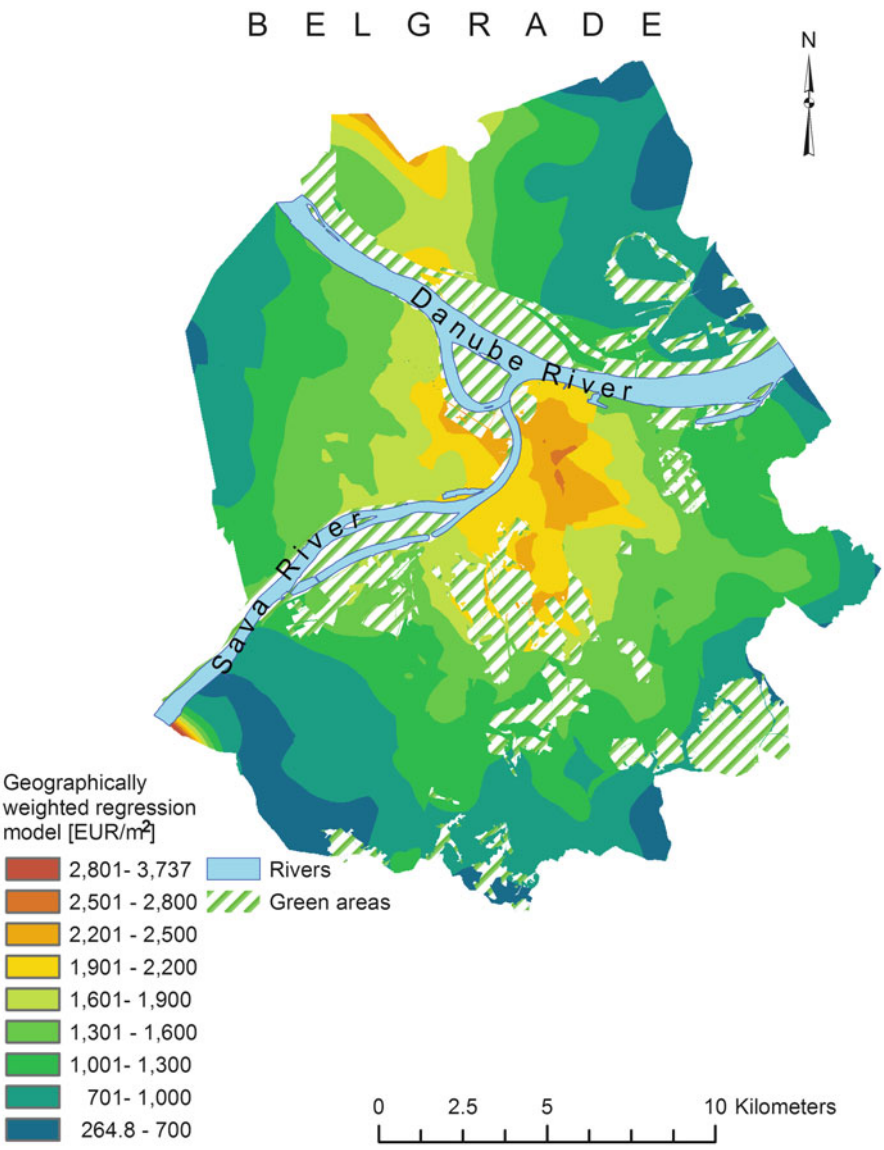

Fig. 10 The map of housing prices predicted by GWR (also available in HTML format at: http://osgl.grf.bg.ac.rs/static/materials/bajat/GWR_RK_htm/gwr.htm)

of publically accessible spatial data in GIS formats, housing price models can be developed for spatial and temporal domains. The application of GWR and RK should be reinforced in hedonic price modelling in regard to the latest developments in the field of spatio-temporal modelling (Huang et al. 2010; Cressie and Wikle 2011).

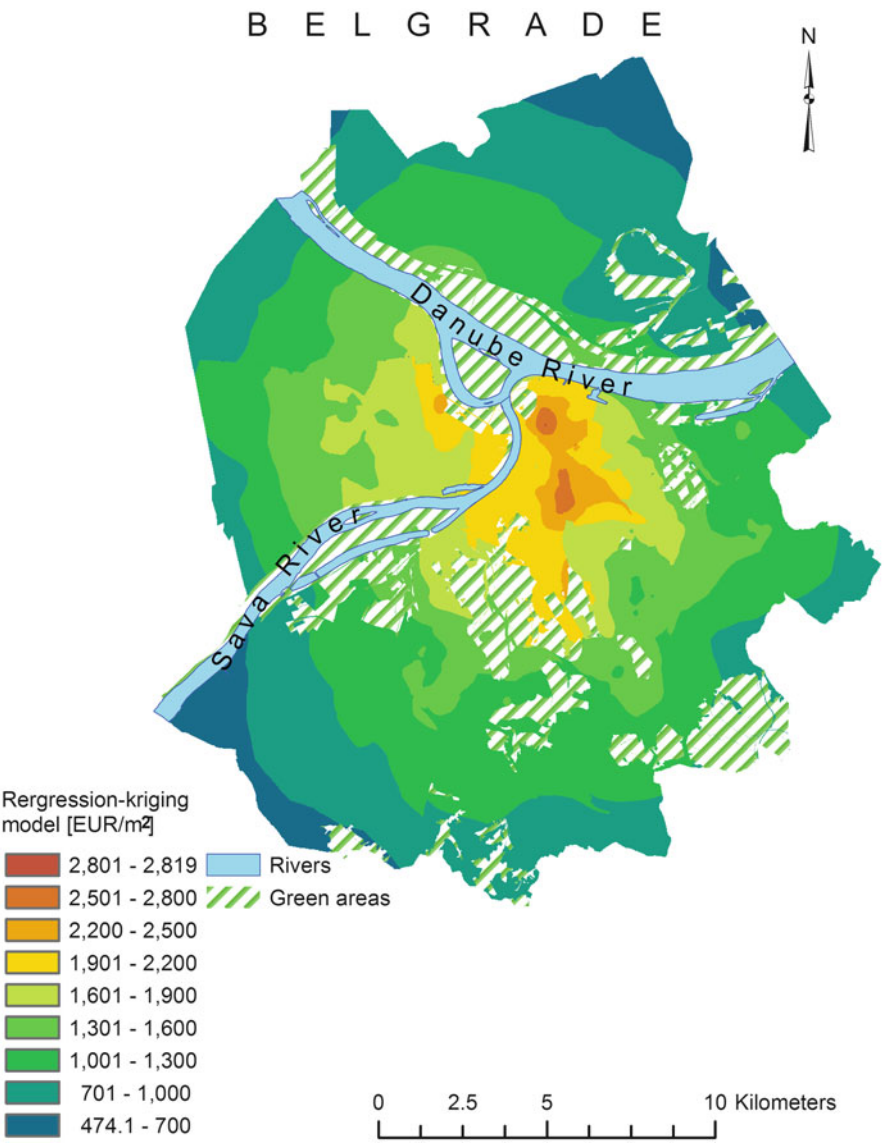

Fig. 11 The map of housing prices predicted by RK (also available in HTML format at: http://osgl.grf.bg.ac.rs/static/materials/bajat/GWR_RK_htm/rk.htm)

Acknowledgments This work was supported by the Ministry of Science of the Republic of Serbia (Contracts No. III 47014, TR 36035 and TR 36009).

References

Adair, A., McGreal, S., Smyth, A., Cooper, J., & Ryley, T. (2000). House prices and accessibility: The testing of relationships within the Belfast urban area. *Housing Studies, 15*(5), 699–716.

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Dordrecht: Kluwer Academic Publishers.

Anselin, L. (1995). Local indicators of spatial association-LISA. *Geographical Analysis, 27*, 93–115.

Anselin, L. (1998). GIS research infrastructure for spatial analysis of real estate markets. *Journal of Housing Research, 9*(1), 113–133.

Anselin, L. (1999). Spatial econometrics www.csiss.org/learning_resources/content/papers/baltchap.pdf. Accessed 20 Sept 2011

Anselin, L., & Bera, A. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah & D. Giles (Eds.), *Handbook of applied economic statistics* (pp. 237–290). New York: Marcel Dekker.

Basu, S., & Thibodeau, T. G. (1998). Analysis of spatial autocorrelation in house prices. *The Journal of Real Estate Finance and Economics, 17*(1), 61–85.

Bivand, R., Pebesma, E., & Rubio, V. (2008). *Applied spatial data analysis with R*, *Use R Series*. Heidelberg: Springer.

Can, A., & Megbolugbe, I. (1997). Spatial dependence and house price index construction. *The Journal of Real Estate Finance and Economics, 14*, 203–222.

Chica-Olmo, J. (2007). Prediction of housing location price by a multivariate spatial method: Cokriging. *Journal of Real Estate Research, 29*(1), 95–114.

Crespo, R., Fotheringham, A. S., & Charlton, M. E. (2007). *Application of geographically weighted regression to a 19-year set of house price data in London to calibrate local hedonic price models. Proceedings of the 9th International Conference on Geocomputation 2007*. Maynooth: National University of Ireland Maynooth.

Cressie, N., & Wikle, C. K. (2011). *Statistics for spatio-temporal data*. Hoboken: John Wiley.

Cvijanović, D. (2006). Tržište stambenih nekretnina u Srbiji. (in Serbian) Kvartalni monitor 5: 66–77. http://www.fren.org.rs/sites/default/files/qm/036_km5-00-ceo.pdf. Accessed 20 June 2011

Des Rosiers, F., Thériault, M., & Villeneuve, P. (2000). Sorting out access and neighborhood factors in hedonic price modeling. *Journal of Property Investment & Finance, 18*(3), 291–315.

Draper, N., & Smith, H. (1981). *Applied regression analysis* (2nd ed.). New York: John Wiley & Sons.

Dubin, R. A. (1988). Estimation of regression coefficients in the presence of spatially autocorrelated error terms. *The Review of Economics and Statistics, 70*, 466–474.

Dubin, R. A. (1998). Predicting house prices using multiple listings data. *The Journal of Real Estate Finance and Economics, 17*(1), 35–59.

Fernández-Avilés, G., Minguez, R., & Montero, J. (2012). Geostatistical air pollution indices and spatial hedonic and models: the case of Madrid Spain. *Journal of Real Estate Research, 34*(2), 243–274.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. Chichester: Wiley.

Fox, J. (2008). *Applied regression analysis and generalized linear models* (2nd ed.). Los Angeles: Sage.

Hanink, D. M., Cromley, R. G., & Ebenstein, A. Y. (2010). Spatial variation in the determinants of house prices and apartment rents in China. *The Journal of Real Estate Finance and Economics, 45*(2), 347–363.

Harris, R., Singleton, A., Grose, D., Brunsdon, C., & Longley, P. (2010). Grid-enabling geographically weighted regression: A case study of participation in higher education in England. *Transactions in GIS, 14*(1), 43–61.

Hengl, T. (2009). *A practical guide to geostatistical mapping* (2nd ed.). Amsterdam: University of Amsterdam. www.lulu.com.

Hengl, T., Heuvelink, G. B. M., & Rossiter, D. G. (2007). About regression-kriging: From equations to case studies. *Computational Geosciences, 33*(10), 1301–1315.

Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science, 24*(3), 383–401.

Kilibarda, M., & Bajat, B. (2012). plotGoogleMaps: The R-based web-mapping tool for thematic spatial data. *Geomatica, 66*(1), 37–49.

Kim, C. W., Phipps, T. T., & Anselin, L. (2003). Measuring the benefits of air quality improvement: A spatial hedonic approach. *Journal of Environmental Economics and Management, 45*, 24–39.

Koramaz, T. K., & Dokmeci, V. (2012). Spatial determinants of housing price values in Istanbul. *European Planning Studies, 20*(7), 1221–1237.

Lake, I. R., Lovett, A. A., Bateman, I. J., & Langford, I. H. (1998). Modelling environmental influences on property prices in an urban environment. *Computers, Environment and Urban Systems, 22*(2), 121–136.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy, 74*, 132–157.

Lovett, A. A., & Bateman, I. J. (2001). Economic analysis of environmental preferences: Progress and prospects. *Computers, Environment and Urban Systems, 25*, 131–139.

McBratney, A. B., Odeh, I. O. A., Bishop, T. F. A., Dunbar, M. S., & Shatar, T. M. (2000). An overview of pedometric techniques for use in soil survey. *Geoderma, 97*, 293–327.

Meese, R. A., & Wallace, N. E. (1997). The construction of residential housing price indices: A comparison of repeat-sales, hedonic-regression and hybrid approaches. *The Journal of Real Estate Finance and Economics, 14*(1–2), 51–73.

Mueller, J. M., & Loomis, J. B. (2008). Spatial dependence in hedonic property models: Do different corrections for spatial dependence result in economically significant differences in estimated implicit prices? *Journal of Agricultural and Resource Economics, 33*(2), 212–231.

O'Sullivan, D., & Unwin, D. (2003). *Geographical information analysis*. Hoboken: John Wiley & Sons.

Osland, L. (2010). An application of spatial econometrics in relation to hedonic house price modelling. *Journal of Real Estate Research, 32*(3), 289–320.

Pebesma, E. J. (2004). Multivariable geostatistics in S: The gstat package. *Computational Geosciences, 30*, 683–691.

Quigley, J. M. (1995). A simple hybrid model for estimating real estate price indexes. *Journal of Housing Economics, 4*(1), 1–12.

R Development Core Team. (2008). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy, 82*, 35–55.

Statistical Office of the Republic of Serbia. (2011). *2011 census of population, households and dwelings in the republic of Serbia-first results*. Belgrade: Statistical Office of the Republic of Serbia.

Wang, F. T., & Zorn, P. M. (1997). Estimating house price growth with repeat sales data: What's the aim of the game? *Journal of Housing Economics, 6*, 93–118.

Yoo, E. H., & Kyriakidis, P. C. (2009). Area-to-point Kriging in spatial hedonic pricing models. *Journal of Geography, 11*, 381–406.

Yrigoyen, C. C., Otero, J. V., & Rodríguez, I. G. (2008). Modeling spatial variations in household disposable income with geographically weighted regression. *Estadística Española, 50*(168), 321–360.

Modeling and Simulation of Cohesion Policy Funding and Regional Growth Diffusion in an Enlarged European Union

Sebastien Bourdin

Abstract The dilemma between equity and competitiveness has created concerns about the future of redistribution of European regional policy funding. The objective of this chapter is to estimate the spatial expression of convergence and regional growth in the European Union. After contextualizing the EU enlargements of 2004 and 2007, this study uses spatial statistics and the simulation platform GeoCells, the goal of which is to analyze two alternatives for future economic development of the EU.

Keywords Cellular automata • Cohesion policy • Regional disparities • Equity and competitiveness

Introduction

On May 1st, 2004, the most important extension of the European Union (EU) in history took place. Ten countries became full EU members: in the north, the three Baltic States (Estonia, Latvia and Lithuania), the four countries of Central Europe (Hungary, Poland, the Czech Republic and Slovakia), a country of south-west area (Slovenia) and two islands (Cyprus and Malta). Two countries in South-East Europe (Bulgaria and Romania) integrated the EU on January 1st, 2007. Consequently, the level of prosperity in the EU declined significantly. However, because of the long process of transformation of post-soviet societies, this event was generally received with enthusiasm.

Numerous geographical issues arose from this policy of openness in the Central and Eastern European Countries (CEECs). What territorialized management of the cohesion policy was required with the arrival of ten new countries? The community economic frame was disrupted by the last two enlargements which provoked an unprecedented increase in the economic gap between developed regions and those

S. Bourdin (✉)
Department of Regional Economics & Sustainable Development, Normandy Business School, Métis Lab, 9 rue Claude Bloch, 14052 Caen, France
e-mail: sbourdin@em-normandie.fr

J.-C. Thill (ed.), *Spatial Analysis and Location Modeling in Urban and Regional Systems*, Advances in Geographic Information Science,
https://doi.org/10.1007/978-3-642-37896-6_6

lagging behind. This situation requires the member states to revise the objectives regarding cohesion in order to prevent increasing economic, social and territorial fragmentation of the Union.

The inclusion of the CEECs, countries with far less economic development than the poorest of the EU-15 (Italy, Spain, Greece and Portugal) reopened the question of the ability of Europe to promote socioeconomic and territorial cohesion. In light of the results of our simulations, our prospective approach proposes two possible scenarios of economic development for the EU of tomorrow by demonstrating the dilemma between equity and competitiveness (Lackenbauer 2006).

The purpose of this research is to understand the process of convergence by using the simulation platform GeoCells (Elissalde et al. 2009) coupled with spatial statistics. An application of this model demonstrates the economic performance of European regions according to the variation in aid granted by the European Union, as well as neighborhood effects. Taking into account the regional disparities, GeoCells analyzes European regions' relative positions from the angle of macroeconomic and budgetary indicators. The cellular automaton GeoCells allows an assessment of the overall effectiveness of regional policy and measures the influence of modification of granting rules.

The introduction of simulation and forecasting methods, along with spatial statistics, in EU regional policy debates is not an attempt to find the one and only response to the problem of European regions' unequal development. Instead, it suggests a range of credible options as a decision support tool for territorial solidarity – as well as economic and social cohesion – in a European space which is in perpetual evolution. Even though European regions belong to an interdependent group, they each have their specific trajectories, in which reaction times and pace of change vary strongly from one to another. These various trajectories build a European regional mosaic, making it difficult for policy makers to override initially planned regional policies (Cohesion Policy, Cohesion Funds, etc.) with budgetary adjustments. Overarching policies are enacted for these separate states/regions in their separate trajectories – but these policies may actually prevent, curtail, or disproportionately power certain trajectories, and in fact may disable newer "corrective" policy/fiscal mechanisms from assisting.

Methodologically, the GeoCells cellular automaton is based upon interactions between variables (e.g. time periods, growth rates in the GDP per head, flows of public investments) and three geographical levels (European level, national level and regional level). Due to the role of spatial interactions and contiguity effects in regional trajectories, in this research, a regional growth diffusion parameter was added to the above variables ratified by the European Commission. Though many regional growth models analyze the region as a stand-alone unit and ignore spatial interaction phenomena linked to proximity, neighborhood, or contiguity effects, the spatial dynamic parameter was added to the variables to underline the role of growth diffusion in regional development.

What Is the Role of European Regional Policy in Reducing Disparities in the EU?

The issue of the solidarity effort between Member States and regions (NUTS 2), as well as their adherence to the cohesion principles defined in the European texts and treaties, is at the center of the debates on European regional policy. The European Union's regional policy seeks to reduce structural disparities between EU regions, foster balanced development throughout the EU and promote real equal opportunities for all. Based on the concepts of solidarity and economic and social cohesion, it achieves this in practical terms by means of a variety of financing operations, principally through the Cohesion Policy (European Regional Development Fund (ERDF) and the Cohesion Fund). For the period 2007–2013, the European Union's regional policy is the EU's second largest budget item, with an allocation of EUR 348 billion. The objective of economic and social cohesion was introduced in 1986 with the adoption of the Single European Act. The policy was finally incorporated into the EC Treaty itself (Articles 158–162) with the Maastricht Treaty (1992).

The main question is in regard to the ability of Cohesion Policy to reduce disparities produced by the single market. How can we improve redistribution and territorial equity in a Union with low economic growth? In such an economic context, should we limit the solidarity efforts of wealthy countries or, on the contrary, increase it in order to accelerate the economic advancement of regions in an earlier stage of economic development?

The implicit deal between the EU and CEECs of opening new markets against the backdrop of the promise of a rising standard of living for relevant populations also implies that this development is achieved by offering newcomers Cohesion Policy. The results of EU policies in helping regions to economically advance are very difficult to assess accurately.

The evaluation of effectiveness of Cohesion Policy in promoting regional development raises methodological problems (Fayolle and Lecuyer 2000). Even if the distribution of Cohesion Policy is proportional to the economic development level, and regions lagging behind are catching up with wealthier regions, it is difficult to determine whether these outcomes are due to Cohesion Policy or other factors. In addition, there is no guarantee that the Cohesion Policy constitutes an explanatory factor of the regional convergence, even though their correlation is significant (ESPON Project 2.2.1. 2005). Indeed, we cannot rule out the possibility that a natural convergence process is simply an outcome of developmental progress.

Following the integration of ten CEECs in the EU in 2004, a debate on the development of the poorest regions emerged in the mid-2000s. The European Commission (2006, 2008a) hoped to invest massive resources in order to help them to develop more quickly. Nevertheless, Gorzelak et al. (2010) argues that the development through a massive injection of money in poor regions is ineffective. The transfer of more than EUR 1 billion euros did not meet expectations in southern Italy and former East Germany (*ibid.*). In addition, this method of massive

investment has a perverse effect: it can create a situation in which the inhabitants of these regions become dependent upon aid they receive. Aghion and Cohen (2004) have shown that the only effective regional investment for poor regions is investment in education. However, these Funds are traditionally invested by the new Member Countries (a decision-making power which the EU has allowed) in other infrastructures such as transport. It is therefore understandable that policies in southern Italy or in Extremadura have not been fruitful.

The EU has to face to another structural obstacle. The EU must accept that the regional disparities in Eastern Europe have existed for centuries. It is very difficult to change these disparities in the time frame outlined in the Cohesion Policy program (i.e., 2000–2006 or 2007–2013). In Poland, for example, Coudroy de Lille (2009) highlighted the fact that regional contrasts and their spatial inscription were created in the nineteenth century. Stryjakiewicz (2007) also explains that metropolization has accentuated the regional disparities during economic transition. Thus, the CEECs are fragmented within their own borders, with disparities between cities and the countryside and between West and East. These differences are reinforced in the historical distribution of wealth, the post-Soviet transformation, the values of society and the efficiency of government.

Finally, before making decisions about fund allocation, it is necessary to consider where to invest. One might think that for ethical reasons, that aid should go to the poorest regions. However, studies show that investing in cities has much more of an impact than investing in rural communities (European Commission 2008b). The analysis of successive generations of European aid to CEECs highlights the dilemma between equity (investment in rural areas with the goal of convergence) and competitiveness (investment in cities with the goal of growth). The economist Williamson (1965) studied the contradiction between a strong GDP growth rate and the increase of regional disparities. These studies were recently replicated in the EU by Ezcurra and Rapún (2006), who also came to the conclusion that an increase in financial support for CEECs would produce simultaneous convergences between the growth rates of CEECs and member countries of the EU, while increasing regional disparities within the CEECs. According to Bergs (2001) inter-regional convergence could take place over time, but at the expense of the national growth potential of new members. The latest report from the European Commission on Economic and Social Cohesion seems to confirm this prediction. If the disparities in the GDP per capita are decreasing between countries, they are increasing in each country. This is the case for both EU-15 Member States and the new Members States (European Commission 2006, 2008a; European Parliament 2007, 2008). Thus the problem of competitiveness and equity is posed (Fayolle and Lecuyer 2000): should we help the least developed regions in order to help them to catch up?

Although the EU structural policy remains an important instrument of cohesion and solidarity at European level, its effectiveness at the EU regional policy level needs to be considered. However, because of the myriad of factors that come into play, it is impossible to assess categorically the true impact of the Cohesion Policy on European Spatial Planning (Dühr et al. 2009) and territorial cohesion (Jouen 2008; Kilper 2009). It is also difficult to know what beneficiary regions would look

like today if the funds had not been granted. It is for this reason that the modeling and simulation of EU Cohesion Policy based upon the configuration of regional economic disparities could contribute to the evaluation of European policies.

Toward Modeling the Cohesion Policy and Its Effects Upon Regional Economic Dynamics

With the aim of investigating possible solutions for reducing the development gap – a gap which increased significantly with the progressive transition from 15 to 27 Member States in the European Union – we have developed a cellular automaton. The simulation platform GeoCells is use to determine under which conditions (in terms of budgetary redistribution settings) and according to which goals (of reduction, convergence, or adjustment), European solidarity policies could be effective.

The Need for Modeling and Simulation to Understand the Issues of European Regional Policy

Economic theory has various tools for clarifying and analyzing the issue of the role of European cohesion policy in the convergence process:

(i) growth theories allow for an analysis of the mechanisms of economic growth as well as the outlook for divergence or convergence of economies;
(ii) theories on geographic economy allow for a study of agglomeration mechanisms in economic activity and the spatial structure of economic disparities;
(iii) econometric methods present tools for an evaluation of convergence phenomena in conjunction with cohesion policy.

With the development of simulation methods, several macroeconomic models have been created in order to understand the role of European regional policy in reducing regional disparities. Such simulations allow for an evaluation of what would have been the current situation of GDP in the absence of cohesion policy. These models also permit ex-ante or ex-post analyses and offer scenarios according to budgetary stance. Such a model undeniably has certain benefits. It is mainly for these reasons that the European Commission bases itself on work carried out within the framework of the HERMIN (Bradley et al. 1995, 2003 and Bradley and et Untiedt 2007) and QUEST (Roeger and Jan in 't Veld 1997, 2004; Varga and Jan in 't Veld 2011) models in its European Funds assessment reports. It draws some rather flattering estimations on the role of regional European policy in short term growth (Kelber 2010) for the HERMIN model, whereas the QUEST model makes some slight references to its long term impact (Magnier 2004).

Several publications have, however, highlighted their limitations. For Sjef Ederveen and Gorter (2002), Ederveen et al. 2006), the application of models such as QUEST et HERMIN only gives a glimpse of the potential effects of cohesion policy in the sense that these Funds have numerous parameters of efficiency. Nevertheless, according to these same authors, regional policy appears to be more successful in a environment which is conducive to growth. The example of the "Irish Miracle" is a clear illustration of this. Furthermore, Philippine Cour and Nayman (1999) note that the simulations only assess what the economic situation would have been in the absence of European regional policy (see for example "Panorama Inforegio", n°33, 2010) in a short term analysis. Finally, numerous underlying assumptions are made and their generality is problematical (Cappelen et al. 2003). For example, it is taken for granted that the collected Funds are systematically allocated to productive public investments, an assumption which is far from being systematically verified. The HERMIN model is based on the assumption that States are open economies (Bradley 2002), which is not the case everywhere in Europe. One of the major limitations of this macroeconomic model is that it can only be applied on a national level. The regional declination is overlooked in this model due to insufficiently comprehensive databases. Moreover, amongst the assessments of the role of cohesion policy in regional growth and convergence, a number of authors (Le Gallo 2004; Rey and Janikas 2005; Ertur and Le Gallo 2008) have demonstrated the role of the effects of neighborhood and spatial dependency on the efficiency of European Funds. The effects of diffusion of regional growth have not been taken into account in either model.

In this context, we consider that modeling by cellular automaton enables a clarification of the issues of convergence and European regional integration. The model that we have developed allows for the effects of neighborhood and diffusion of regional growth to be taken into account. In addition, modeling by simulation is useful in that it reveals the processes and mechanisms (i) and serves as a decision support tool (ii).

(i) Cellular automaton simulation is constructive as it takes into account the complexity of the relationship between decision making (budgetary stance, duration of European regional policy programming periods), economic factors (growth and convergence) and spatial aspects (interaction between regions/Member States).
(ii) Simulation is helpful when it is not a question of finding the optimal solution but of exploring a wide range of possible scenarios in order to identify the parameters that would significantly improve the efficiency of European cohesion policy.

GeoCells, A Multi-Layered Hierarchical Automaton

GeoCells is a simulation platform based upon layers of geographic information. Its main engine is a meta-model based upon spatial agents or a topologic cellular agent.

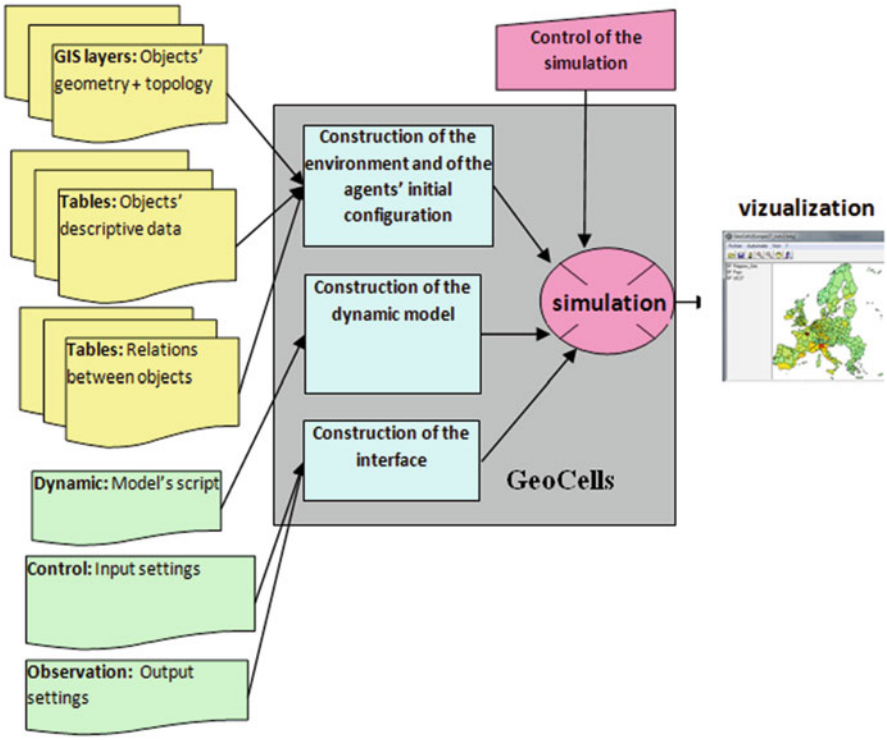

Fig. 1 GeoCells functioning principle

GeoCells is used to model the evolution of GDP per capita in the EU-27, and the simultaneous influence of different types of aid under the cohesion policy, and the effects of growth diffusion by neighborhood. The general operating principles for GeoCells are displayed in Fig. 1.

The system is based upon a group of geographic information layers (Fig. 2). Each layer (EU (1); member-state (2); region (3)) consists of features from the same class. Each layer is made up of cells (EU, countries, regions). A cell's main function is to own, in addition to the feature's physical components (location, shape, size . . .), the knowledge of its neighborhood and above all a behavior dynamic.

Each layer owns behavior rules giving to the cells of its class the same function in the system (region, member-state, EU), properties and attributes (perimeter, surface, budget of the cell) and relations with cells from other layers of the system.

The system takes into account the hierarchical relationships existing between layers (Fig. 2); a region (Layer 1) belongs to a country (Layer 2) – inclusion link – and a country is made up of regions – containing link.

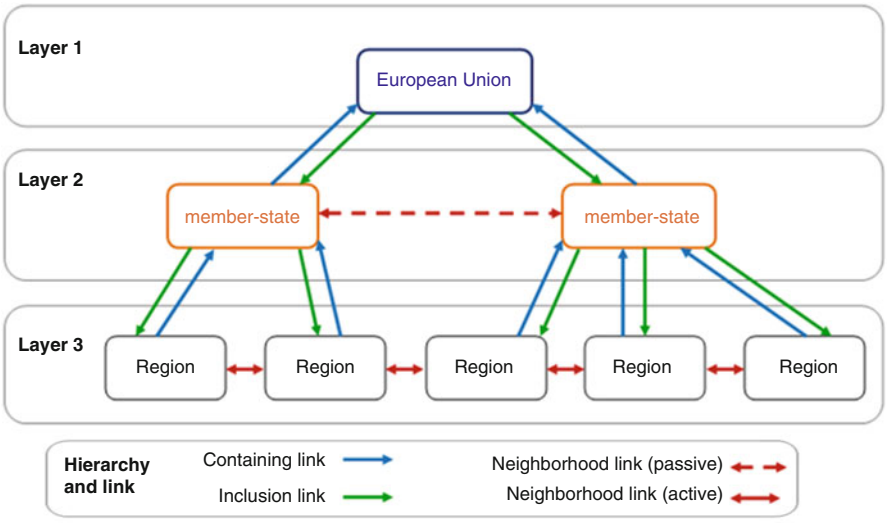

Fig. 2 Hierarchy of cellular layers

The **generated automaton characteristics** are described below:

Cellular interactions: A cell layer interacts "naturally" with its sister cells (its topological neighbors) with its mother cell (above in the hierarchy) or its daughter cells (below in the hierarchy), but can also interact with cells of any other layer, through explicit links. These links represent cellular exchanges.

State: Each cell's state represents an attribute that is likely to change during the simulation. Each state has a semantics which can represent information (such as the budget of its neighbor) or an amount of material or energy (such as its own wealth, or population).

Phases of Life: One of the difficulties of this type of mechanism is to maintain the temporal coherence between all cellular layers. Every cell performs four steps when it receives inflow:

- Reading of its inputs (inflow from outside);
- Implementation of its program of action (behavior);
- Writing of its outputs (outflow exchange);
- Storage of its context (each cell must maintain at least the contents of the previous context). The context is defined here as the previous state of the cell and the recording of the state variables of neighboring cells.

Capacity – Every cell has its attributes (or state variables) but the rules of behavior are collective (because they are shared by all elements of its class). Each cell generates actions that depend upon its inputs and its state at a given time. The action taken is the result of a choice of the cell. This choice depends on the evaluation of the relevance of the rules of actions that may apply. In other words, the cell can have "smart" behavior comparable to that of an *agent* (we nevertheless retain the term *cell*).

Communication canals – A bidirectional communication canal exists to combine the system's multilayer nature. Each cell owns the input and output references relating to the canals that concern it. For this reason, the cell knows its environment and enters into dialogue with it.

The Possible Simulation Settings

Given the data available for the group of regions NUTS2 of the EU-27, the model generated, as the main indicator, the variation in GDP per capita of each European region. Within Geocells, policy variables are adjusted for each simulation, while population remains constant. A user interface provides an opportunity at the beginning of the simulation for the user to enter a value for each policy variable. The settings which can be varied within Geocells are described below.

The Article 160 of the Treaty establishing the European Community (in its consolidated version in 2002) provides that the European Regional Development Fund (ERDF) is intended to help to redress the main regional imbalances in the Community. The ERDF therefore contributes to reducing the gap between the levels of development of the various regions and the extent to which the least favoured regions, including rural and urban areas, declining industrial regions, areas with a geographical or natural handicap, such as islands, mountainous areas, sparsely populated areas and border regions, are lagging behind. Rules of allocation of Cohesion Policy as defined in the Treaty have been implemented in Geocells. The **GDP variation rate** is, either specific to the region or identical to the group of regions of the same country or identical for the whole of EU. The terms of public intervention include the mechanisms relating to **contributions** (Countries and EU), to the aid linked to regional policy, such as **eligibility thresholds** (75% of the average GDP per capita of the EU) for Cohesion Policy. The **European budget weight** is taken into account. The EU budget is stabilized around a threshold of 1% of the total European GDP (threshold reached since 1984 with the Single European Act). The EU had an agreed budget of EUR120.7 billion for the year 2007 and EUR 864.3 billion for the period 2007–2013, representing 1.05% of the overall wealth of the EU-27's. From this average budget, simulations were able to make the Eurropean budget weight vary from 0,5 to 3% of the EU total GDP. The principle of **additionality** between the States and the European Union in the Cohesion Policy financing was also taken into account. According to this principle, EU funds can only be paid in addition to a contribution from the member states, not instead of it. The variability of the relative importance of regional policy in the **EU budget expenditures** is also one of the simulation settings. The ERDF and the Cohesion Fund make up one of the largest items of the budget of the EU. The overall budget for the period 2007–2013 is EUR 271 billion and represent 30,4% of total EU expenditures. In addition to these principles officially ratified by the Treaty establishing the European Community, we have added to our model a **spatial dynamic parameter**: the hypothesis of the role of spatial interactions and of contiguity effects in the regions' trajectories.

The diffusion by contact with neighboring regions, made possible by the functioning of the cellular automaton, is carried out therefore naturally in one way or another. With GeoCells, what is happening in the neighboring regions is not ignored. Several researchers (Baumont et al. 2002; Islam 2003; Le Gallo 2004; Rey and Janikas 2005; Dall'erba and Le Gallo 2008; Dall'erba et al. 2009; Dall'erba and Hewings 2009; Ertur and Le Gallo 2008) have shown that most studies consider the regions as isolated entities, as if their geographical location and their potential inter-linkages were not important. However, the geographical distribution of growth phenomena at the regional level is rarely random: the economic performances of neighboring regions are often similar (Getis 1991). The impact of the unequal distribution of economic activities in space upon the territories' economic growth was underlined in particular by Baumont (1998). While a situation of spatial competition between activities and between territorial units exists, the taking into account of contagion, of mimicry phenomena linked to neighborhood effects proves to be necessary.

Growth-Diffusion Model for European Regions

We have attempt to model a complex diffusion process in real life by choosing a specific diffusion mechanism. The diffusion by contact with neighboring regions was highlighted especially by Elissalde et al. (2009) and Bourdin (2013) who has shown for example that regions of Central Europe (eastern Germany, the western parts of the Czech Republic, Slovakia, Hungary and Slovenia) have a low level of GDP per capita compared to the EU15 average, but a geographic environment which is more favorable than the regions further to the east in the EU. In this context, a catching-up of regions of Central Europe is explained in part by a growth diffusion process by neighboring. The proximity of regions of Central Europe to the border of the EU15 gives to these regions a high development potential compared to regions further east. This suggests that the distribution of regional growth occurs more neighbor to neighbor.

We will now clarify the unique diffusion model that we have used. The term X_i represents the GDP of the region i, P_i its population and $Y_i = X_i/P_i$ its GDP per capita at a moment t.

We present the following hypothesis. Each cell has the aim to homogenize, through time, its standard of living Y in relation to its neighbors. The attempt to homogenize standard of living is the policy goal of the Territorial Cohesion. The main aim of the Territorial Cohesion policy is to contribute to a balanced distribution of economic and social resources among the European regions with the priority on the territorial dimension. This means that resources and opportunities should be equally distributed among the regions and their populations. But, in our model, standard of living is not capable of diffusing like a flow. It is through the variation of wealth (X) symbolized by the GDP (by internal growth and by diffusion) or through

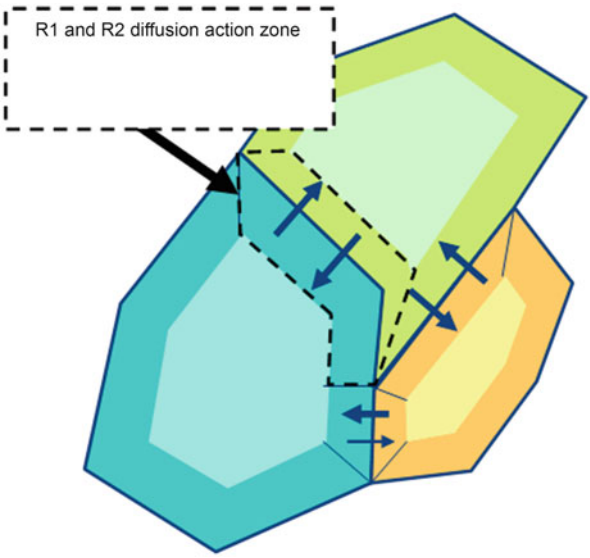

Fig. 3 Practical implementation of growth-diffusion rate

the variation of population (P) (also by internal growth or by migrations) that each region can work in order to achieve its goal. The diffusion mechanism only relies on the variation of X.

Another hypothesis is to consider that a small fringe close to the borderline (area in dotted line, Fig. 3) takes part in the diffusion of wealth, by the leveling-out of standards of living of the two neighboring border fringes (Fig. 3). Since we do not have any information on the spatial distribution of the populations inside a region, we must put forward the hypothesis of a uniform distribution. Consequently, we use a simple proportionality parameter, called the diffusion rate, the value of which can be set within the user interface. This parameter rate k (of surface area, population, and wealth) is all at once, since we consider them as uniformly distributed over the region's surface area.

In order to model the diffusion between two regions i and j, we then introduce the coefficient k_{ij} which is the surface area's proportion i matching the intersection between the border fringe defined by k and the proportion p_{ij} of its borderline land shared by the region j, defined by $p_{ij} = \frac{l_{ij}}{\sum_{k \in Vois(i)} l_{ik}}$, where l_{ij} is the borderline's length between i and j.

We then have: $k_{ij} = k \,.\, p_{ij}$.

If the wealth on the two sides of the border fringe between i and j was evenly distributed like connected areas, we would obtain a leveled-out standard of living (which is not the average of the two previous standards), defined by:

$$Y_{ij} = \frac{k_{ij}.X_i + k_{ji}.X_j}{k_{ij}.P_i + k_{ji}.P_j} \tag{1}$$

We can then define the variation dX_{ij} (positive if it emits or negative if it receives) of the diffusion from the region i towards the region j during a short lapse of time dt as being proportional to the concerned population ($k_{ij}P_i$) and proportional to the difference between the current standard of living (Y_i) and the (local) aim of leveling-out (Y_{ij}) of standards of living i and j. This can be translated into the following equation:

$$\frac{dX_{ij}}{dt} = K.k_{ij}.P_i\left(Y_i - Y_{ij}\right) \tag{2}$$

The value of K is set internally (since we can already play on k).

By adding the border fringes of the region i, we note:

$$dX_i = \sum_{j \in Vois(i)} dX_{ij} \tag{3}$$

One should notice that this diffusion is, by construction, preservative of the mean $\sum_{i=1}^{n} X_i$ (because one can verify easily that for any couple (i, j) we have: $dX_{ij} + dX_{ji} = 0$).

Moreover, the variable X_i is subjected to an a priori exponential internal growth, $\frac{dX_i}{dt} = C_i X_i$.

Internal growth is adjustable, either individually region by region through the attribute table, either on the whole as being the same for all regions with the help of a setting determined by the user within.

The final growth-diffusion equation is thus given by:

$$X_i\left(t + dt\right) = X_i(t) + \left(C_i.X_i(t) + K.k_{ij}P_i\left(Y_i - Y_{ij}\right)\right)dt \tag{4}$$

The lapse of time for the discretization of growth and diffusion processes are small compared to redistributing flows, because they correspond to continuous processes. We have selected the month as lapse of time, that also matches the time unit that we chose, so $dt = 1$. (C_i is then the twelfth of the annual growth rate).

The equation with this lapse of time is then written:

$$X_i\left(t + dt\right) = \left(1 + C_i\right)X_i(t) + K.k_{ij}P_i\left(Y_i - Y_{ij}\right) \tag{5}$$

Europe 2025: Which Scenario from Which Policy?

To assess the weight of political cohesion in regional trajectories, simulations were performed with the GeoCells platform. These simulations were based on the one hand on the settings of allocations Funds and, on the other hand on neighborhood effects. The two scenarios presented below ask questions about the effectiveness of the cohesion policy and the dilemma between competitiveness and equity. This

Table 1 Indicators in Scenarios of Cohesion Policy in Europe in 2025.

Scenario:	Simu 1: Competitiveness	Simu 2: Solidarity
Cohesion Policy – percentage of the EU budget	10%	35%
Threshold of allocation of Cohesion Policy (GNI/gross national income as percent of EU average)	90%	75%
Beta convergence	−0,2654	−0,6818
Sigma convergence	0,0021	−0,0012
GDP diffusion rate	30%	10%
Gini index	0,21	0,14
Moran index	0,43	0,37
Doctrine	Liberalism	Planned economy
Priorities for the cohesion policy	Competitiveness	Convergence Integration of the less economically developed regions
Public policies	Renationalization of aid	Increase Cohesion Policy total budget
Mechanism for promoting cohesion	Liberalization and competition increased	Strong regulation

dilemma can be read in the Treaty of Rome (1957) and the Single European Act (1987) where it says that the EU has to support the growth and the job creation in Member states and least developed regions.

The first scenario (simu 1) is the one of free competition between regions without the intervention of Cohesion Policy (Table 1). It is tantamount to abolishing European "interventionism" and to "renationalizing" aid, just as recommended in the Sapir Report (2003). "An Agenda for a Growing Europe", also called The Sapir Report, is a report on the economy of the European Union edited by a panel of experts under the direction of André Sapir and published in July 2003. The report follows an initiative by Romano Prodi, President of the European Commission, notably to analyze the Lisbon Strategy. According to the experts of this report, Cohesion Policy and other community interventions do not contribute in an easily measurable way to the convergence of the regions. The results obtained by the countries of the EU remain dependent on their good governance, which leads the experts of this report to write the following recommendation: "there is a solid argument for the new EU convergence policy to focus on countries, rather than on regions". Considering the European budgetary constraints, the report recommends an important reduction of Funds intended for the Cohesion Policy. The simulations include a low percentage of Cohesion Policy in the EU budget. Almost all regions can apply for the Cohesion Policy because the threshold of allocation of Cohesion Policy of 90% of the average GDP per capita of the EU. We observe

that disadvantaged regions catch up slowly and the sigma convergence indicates divergence[1]. The distribution of wealth is more non-egalitarian than the scenario of equity.

The second scenario (Simu 2) has as its goal territorial equity (Table 1). Territorial equity includes ideas of parity of treatment, equality of access, and, more generally, solidarity between regional organizations in terms of public action, especially by implementing corrective measures as far as resources and facilities are concerned. The scenario consists of endowing each region with a measure of autonomy and the necessary conditions for development. Cohesion Policy are used alone, by increasing the percentage devoted to regional policy to 35% of EU budget, and by retaining the actual threshold of allocation of Cohesion Policy to 75% of the average GDP per capita of the EU.

The measure of convergence based on the evolution of the standard deviation (sigma convergence) gives the most valuable result for the scenario of equity based upon increasing the budget for regional policy, and the prospect of catching-up (beta convergence) is more credible with the scenario of equity as well. With this policy orientation, every region of each country reacts positively to territorial solidarity programs. In accordance with the results in terms of beta and sigma convergence, simulation 2 brings out a better result in terms of Territorial Cohesion, mitigating significatively regional disparities across EU.

The cartography of these scenarios gives concrete expression to the impact on geographic distribution of growth chosen by each parameter setting (Fig. 4). We have measured local concentrations through the Getis-Ord statistics. This index allows the identification of spatial clusters (or "local pockets"). A positive value will indicate a spatial concentration of GDP per capita (called a "hot spot"), while a negative value of that index is associated with spatial concentration of low value of GDP per capita (a "cold spot"). Two main patterns of clusters can be shown. Within the competitive scenario, the "Pentagon" (cluster of prosperous regions) and regions bordering this cluster are strongly linked to each other; unfortunately many regions of formerly socialist countries remain far behind. This scenario produces the phenomena of the clustering of prosperous regions very often from metropolitan regions (South of England, Parisian Basin, North West of Italy) whereas poor regions do not manage to progress of their backwardness. Representative of a non-egalitarian growth, this phenomena reveals a certain effectiveness at national level, but establishes itself as less homogeneous at European level. Growth takes place by clusters of regions, but the development gaps are not on the whole being closed (low beta convergence). On the other hand, the scenario of equity highlights the progress of the convergence of GDP. It allows CEECs regions to catch up while allowing the Pentagon to continue to grow. This hypothesis gives a negative sigma convergence

[1]The sigma convergence refers to a reduction in the dispersion of levels of income across economies. Here there is an increase of disparities among regions because of positive result (0,0021).

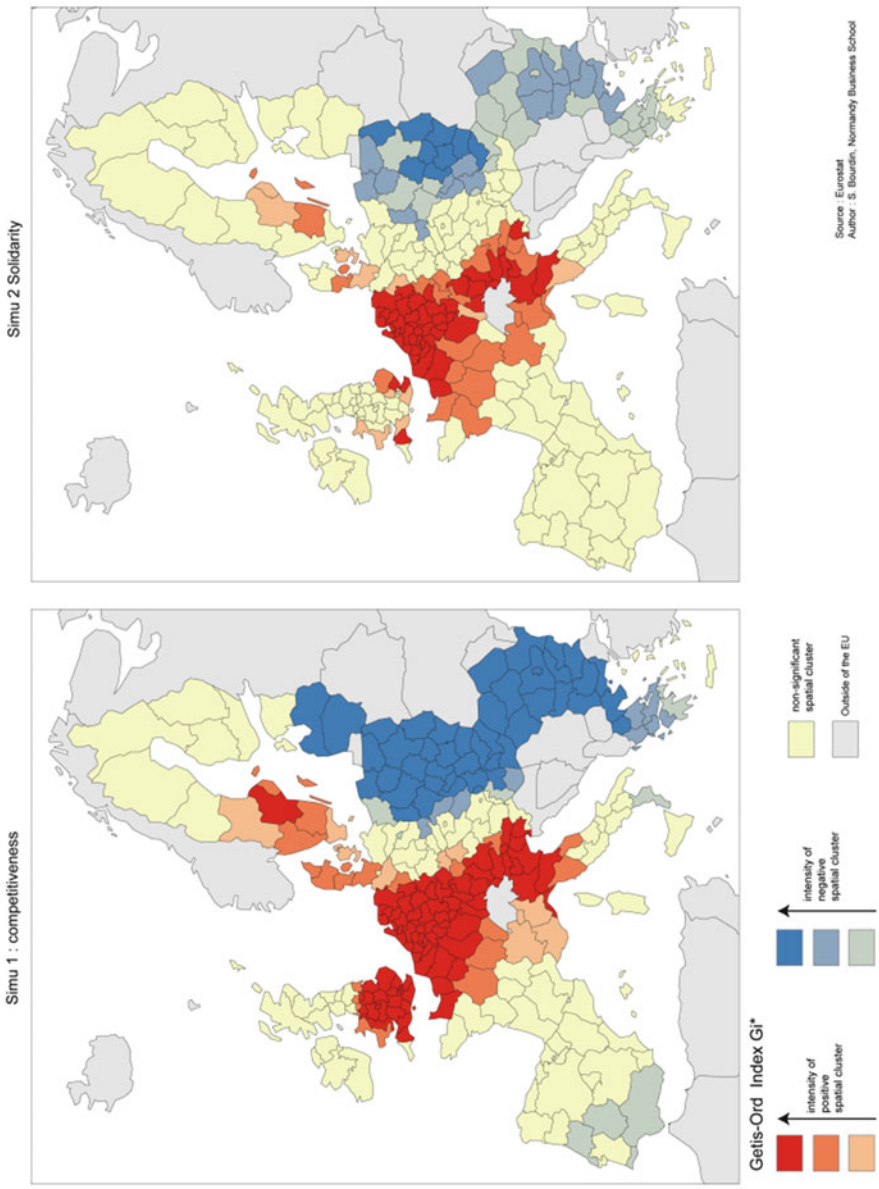

Fig. 4 Cartography of Scenario 1 and 2 – Spatial statistics

with a low dispersion of incomes between regions, since poorer regions saw their GDP per capita rise, but, not at the same rhythm. CEECs regions located closest to the former Iron Curtain seem to be progressing faster.

In addition to the two indicators of convergence (beta and sigma) – in theory complementary and often referenced in the literature in regard to regions' convergence – the introduction of a contiguity-based growth propagation variable changed the expected scenarios which stood as a basis for EU policies. This introduction of spatial interaction by neighborhood transforms the deterministic projections of the EU policies into a system of regional units reacting according to a multi-scalar complexity. The process accounting neighborhood effects reveals the potential for a spatial diffusion process to occur under the assumptions given in each scenario.

The budget of the European regional policy has always been the second largest item of expenditure in the EU, far behind that of the Common Agricultural Policy. With the new programming period 2007–2013, the budget was brought to the forefront because of the efforts related to the 2004 and 2007 enlargements. Achieving competitiveness of the regions included in the Lisbon strategy requires building development strategies that enhance regional strengths and overcome weaknesses and regional gaps. To meet the challenges of globalization, the EU has included the concept of competition in the 2007–2013 programming period for the Cohesion Policy. Meanwhile, the EU continues to pursue its objective of solidarity between regions and countries. This dilemma can be answered by the concept of polycentric development[2]. This concept refers to a development of a polycentric and balanced urban system, and strengthening of the partnership between urban and rural areas, so as to create a new urban-rural relationship. It includes the promotion of integrated transport and communication, which support the polycentric development of the EU territory, so that there is gradual progress towards parity of access to infrastructure and knowledge. Implicitly, this principle implies the presence of "centers" that spread their prosperity to their neighborhoods (hence the need to introduce neighborhood effects in GeoCells) while continuing to help the less economically developed regions to be competitive vis-à-vis the wealthier[3]. This would combine greater European competitiveness with an increase in prosperity of peripheral regions in order to catch up. The spatial dimension of European public action is an opportunity to resolve these contradictions. The territorialization of public policies for regional development (which consists of differentiating policy applications for different regions) coupled with a polycentric planning can allow a difficult compromise between equity and competitiveness.

[2]"Promote a harmonious and well-balanced development of the EU's territory", European Commission (1998).

[3]Sapir Report advocates this but stopping aid to regions in an earlier stage of economic development, thus not allowing these regions to be competitive vis-a-vis the wealthiest

Conclusion

The objective of this chapter was not to provide an answer on how the Cohesion Policy should be used (axiological neutrality) but to clarify issues for the future of European cohesion policy. This clarification is necessary to understand the geographic organization of economic inequality and regional developement in a specific context where the EU asks for each region to define its Smart Specialisation Strategy (Foray et al. 2009; Capello 2014). The two scenarios that have been demonstrated in this study show that the political choices between equity and competitiveness have a profound impact on territorial development. These choices in structural funding investment produce very different economic and spatial configurations. Not only the political orientation can influence outcome, but other factors can have a significant impact on territorial cohesion. Both pre-determined (i.e. programming policies, historical factors) and random (neighborhood effects, diffusion of regional growth) factors affect the dynamics of regional growth and convergence. Because each region has a unique trajectory based not only upon Cohesion Policy but also upon random factors, it is impossible to directly link Cohesion Policy alone to regional economic growth.

At this stage of our research, it would be helpful to use an input-output model as an extension for future work. The input-output model would represent the sectoral diffusion of the funding (underlying processes) and the simulation could represent the resulting geographic diffusion/interactions. The goal would be to explore the logical consequences of assumptions based on neighborhood effects, to complete them with the simulation results so get to know the reality and act more effectively on it.

Acknowledgements This work is part of a collaborative project led with Bernard Elissalde, Dominique Goyat and Patrice Langlois, all of whom we wish to thank deeply.

References

Aghion, P., & Cohen, E. (2004). *Education et Croissance*. Paris: La Documentation Française.

Baumont, C. (1998). Economie géographique et intégration régionale: Quels enseignements pour les Pays d'Europe Centrale et Orientale ? *LATEC – Economics Working Paper (1991–2003), 1998–11*.

Baumont, C., Ertur, C., & Le Gallo, J. (2002). The European regional convergence process, 1980–1995: Do spatial regimes and spatial dependence matter? *Economics Working Paper Archive – Econometrics 0207002*.

Bergs, R. (2001). EU regional and cohesion policy and economic integration of the accession countries. *Discussion Paper, Policy Research & Consultancy*.

Bourdin, S. (2013). Une mesure spatiale locale de la sigma-convergence pour évaluer les disparités régionales dans l'Union européenne. *Région et Développement, 37*, 1–18.

Bradley, J. (2002). An examination of the ex-post macroeconomic impacts of CSF 1994–99 on objective 1 countries and regions. Greece, Ireland, Portugal, Spain, East Germany and Northern Ireland. *Final Report, ESRI & GEFRA*, Dublin.

Bradley, J. et Untiedt, G. (2007). Do economic models tell us anything useful about Cohesion Policy impacts? A comparison of HERMIN, QUEST and ECOMOD. *Working Paper, 3, GEFRA – Gesellschaft fuer Finanz- und Regionalanalysen.*

Bradley, J., Herce, J.-A., & et Leonor, M. (1995). Modelling in the EU periphery: The HERMIN project. *Economic Modelling, 12*(3), 219–222.

Bradley, J., Morgenroth, E. et Untiedt, G. (2003). Macro-regional evaluation of the Structural Funds using the HERMIN modelling framework. *ERSA Congress University of Jyväskylä – Finland.*

Cappelen, A., Castellacci, F., Fagerberg, J., & et Verspagen, B. (2003). The impact of EU regional support on growth and convergence in the European Union. *Journal of common market studies, 41*, 641–644.

Capello, R. (2014). Smart specialisation strategy and the new EU cohesion policy reform: Introductory remarks. *Scienze Regionali*.

Coudroy de Lille, L. (2009). Les nouveaux territoires polonais. In Y. Jean & G. Baudelle (Eds.), *L'Europe, aménager les territoires*. Paris: Armand Collin.

Cour, P., & et Nayman, L. (1999). Fonds structurels et disparités régionales en Europe. *La lettre du CEPII, 177*, 1–14.

Dall'erba, S., & Hewings, G. J. D. (2009). European regional development policies: The trade-off between efficiency-equity revisited. *Connections, 5*, 73–84.

Dall'erba, S., & Le Gallo, J. (2008). Regional convergence and the impact of European Cohesion policy over 1989–1999: A spatial econometric analysis. *Papers in Regional Science, 87*, 219–244.

Dall'erba, S., Guillain, R., & Le Gallo, J. (2009). Impact of cohesion policy on regional growth: How to reconsider a 9-year-old black box. *Région et Développement, 30*, 77–100.

Dühr, S., Nadin, V., & Colomb, C. (2009). *European spatial planning: territorial development, cooperation and EU spatial policy*. Londres: Routledge.

Ederveen, S. et Gorter, J. (2002). Does European cohesion policy reduce regional disparities? An empirical analysis. *CPB Working Paper, 15*, CPB Netherlands Bureau for Economic Policy Analysis.

Ederveen, S., Groot, H. L. F., & et Nahuis, R. (2006). Fertile soil for structural funds? A panel data analysis of the conditional effectiveness of european cohesion policy. *Kyklos, 59*(1), 17–42.

Elissalde, B., Goyat, D., & Langlois, P. (2009). GeoCells model: European cohesion policy and regional interactions, which convergences for the European regions? *Revue électronique Cybergeo*.https://doi.org/10.4000/cybergeo.22388.

Ertur, C., & Le Gallo, J. (2008). Regional growth and convergence : heterogenous reaction versus interaction in spatial econometric approaches. oai: hal.archives-ouvertes.fr:hal-00463274.

ESPON Project 2.2.1. (2005). *The territorial effects of the Cohesion policy, Final report*. Esch-sur-Alzette: Espon Coordination Unit and Nordregio.

European Commission (1998). *Resolution on regional planning and the European spatial development perspective*, Brussels.

European Commission. (2006). *Cohesion policy and cities: The urban contribution to growth and jobs in the regions*, Brussels.

European Commission. (2008a). *Growing regions, growing Europe, 5th cohesion report*. Brussels.

European Commission. (2008b). *Green Paper on territorial cohesion*. Brussels: DG Regio.

European Parliament. (2007). *Regional disparities and cohesion: what strategies for the future ?* Brussels.

European Parliament. (2008). *Shrinking regions: a paradigm shift in demography and territorial development*, Brussels.

Ezcurra, R., & Rapún, M. (2006). Regional disparities and national development revisited, the case of western Europe. *European Urban and Regional Studies, 13*(4), 355–369.

Fayolle, J., & Lecuyer, A. (2000). Croissance régionale, appartenance nationale et fonds structurels européens. *Un bilan d'étape. Revue de l'OFCE, 73*, 161–196.

Foray, D., David, P. A., & Hall, B. (2009). Smart specialisation-the concept. *Knowledge Economists Policy Brief*, *9*(85), 100.

Getis, A. (1991). Spatial interaction and spatial autocorrelation: a cross-product approach. *Environment and Planning, 23*, 1269–1277.

Gorzelak, G., Maier, G., & Petrakos, G. (2010). *Integration and transition in Europe: The economic geography of interaction*. London: Routlege.

Islam, N. (2003). What have we learnt from the convergence debate ? *Journal of Economic Surveys, 17*, 309–362.

Jouen, M. (2008). La cohésion territoriale, de la théorie à la pratique. *Notre Europe.*

Kelber, A. (2010). La politique de cohésion et les nouveaux États membres de l'Union européenne. *Bulletin de la Banque de France, 181*(3), 1–13.

Kilper, H. (Ed.). (2009). *New disparities in spatial development in Europe*. *German Annual of Spatial Research and Policy*, Springer.

Lackenbauer, J. (2006). *Equity, efficiency, and perspectives for cohesion policy in the enlarged European Union*. Bamberg: BERG Velag.

Le Gallo, J. (2004). Space-time analysis of GDP disparities among European regions: A Markov chains approach. *International Regional Science Review, 27*, 138–163.

Magnier, P. (2004). Fonds structurels européens et politiques régionales. *Commissariat général du plan, La Documentation française*.

Panorama. (2010). Evaluating regional policy: Insights and results. *Inforegio, 33*.

Rey, S. J., & Janikas, M. V. (2005). Regional convergence, inequality and space. *Journal of Economic Geography, 5*(2), 155–176.

Roeger, W., & Jan in 't Veld (1997). QUEST II. A multi-country business cycle and growth model. *European Economy – Economic Papers, 123*, Directorate general economic and monetary affairs, European Commission.

Roeger, W., & Jan in 't Veld. (2004). Some selected simulation experiments with the European commission's QUEST model. *Economic Modelling, 21*(5), 785–832.

Sapir, J., Aghion, P., Bertola, G., Hellwig, M., Pisani-Ferry, J., Rosati, D., Viñals, J., Wallace, H., Buti, M., Nava, M., & Smith, P. M. (2003). *An Agenda for a growing Europe: The sapir report*. Oxford: Oxford University Press.

Stryjakiewicz, T. (2007). La nouvelle géographie de la Pologne dans le contexte de la transition économique et sociale et de l'élargissement de l'Union européenne. *L'Information Géographique, 71*(4), 100–120.

Varga, J., & Jan in 't Veld. (2011). A model-based analysis of the impact of cohesion policy expenditure 2000–06: Simulations with the QUEST III endogenous R&D model. *Economic Modelling, 28*(1–2), 647–663.

Williamson, J. G. (1965). Regional inequality and the process of national development: A description of the patterns. *Economic Development and Cultural Change, 13*, 1–84.

Part II
Geography as a Spatial System

Analyzing and Simulating Urban Density Exploring the Difference Between Policy Ambitions and Actual Trends in the Netherlands

Eric Koomen, Jasper Dekkers, and Dani Broitman

Abstract This chapter explores the potential of geospatial analysis to characterize land use dynamics in urban areas and the surrounding urban fringe. It focuses on the difference between policy ambitions and reality with respect to urban densification. Policy ambitions for the containment of urban development are ambitious in the Netherlands, as is evident from the many local objectives to concentrate residential development within existing urban areas. These ambitions are formulated as target shares of the total net increase in housing stock that should be realized within designated urban area boundaries. Following a GIS-based analysis of local changes in housing stock between 2000 and 2008, we are able to describe actual intensification shares and residential densities in newly-developed urban areas. We observe that, especially in the already densely-populated western part of the country, the realized urban intensification shares are below the specified policy goals/ambition levels. Using these observations along with scenario-based projections of the regional increases in housing stock, we are able to determine the demand for new urban land in 2020. Using a land use simulation model, we simulate the urban development processes until 2020 according to two scenarios and two policy alternatives. In the first policy alternative, growth in the number of house units is based upon policy ambitions, while in the second scenario growth is based upon observed trends over the past decade. The results show that current policy ambitions related to urban intensification levels will greatly help in containing the substantial urban growth that is projected in the case of a high economic and population growth scenario. If urban development follows past trends, unabated, thus demonstrating relatively low urban intensification shares, large-scale urban extensions are likely to occur.

Keywords Urbanization • Spatial analysis • Urban planning • Urban growth boundary

E. Koomen (✉) • J. Dekkers • D. Broitman
Faculty of Architecture and Town Planning, Technion – Israeli Institute of Technology, Haifa, Israel
e-mail: e.koomen@vu.nl

J.-C. Thill (ed.), *Spatial Analysis and Location Modeling in Urban and Regional Systems*, Advances in Geographic Information Science,
https://doi.org/10.1007/978-3-642-37896-6_7

Introduction

Spatial Planning and the Urbanization Paradox

In planning, policymakers define objectives and implement various types of instruments to realize specified ambitions. In the domain of spatial planning, typical instruments include urban concentration policies and/or policies set to protect open space, zoning regulations with regard to flood-prone locations, climate adaptation policies and financial measures and zoning regulations to attract new businesses or to reduce the amount of vacant office space (Kuijpers-Linde 2011). While ex-ante evaluation of proposed strategies and polices is common in the legal framework of environmental impact assessment, ex-post evaluation is much less common. Yet, this type of evaluation can be very insightful when new policies are proposed. Looking back at the degree of success of past initiatives can distinguish potentially effective alternatives from less effective ones.

This chapter aims to contribute to the urban land management debate by reflecting on past and future urban developments in relation to spatial policies that aim to steer them. We present an ex-post evaluation of the degree of success of urban containment policies and use this information to show how future developments may potentially deviate from current ambitions.

The Netherlands was selected as a case study for this analysis. In this country, strategic spatial policy is initially formulated in national-level policy documents. Subsequently, this framework is used to formulate more concrete policy ambitions for the regional (i.e., provincial) and local (i.e., municipal) levels. Prior to the establishment of the new national strategic vision for infrastructure and the environment (I&M 2011), spatial plans for the lower levels had to correspond with the ambitions set at a higher level. However, with the approval of the new national vision, the trend towards decentralization of decision-making power in urban planning has continued, and the relationship between urban land use planning at various scales is further changing. The nature of this change and its potential effect of moving from national-level, development-oriented planning towards the setting of the legal role for development-related decision-making at the regional and local levels, is discussed extensively elsewhere (Roodbol-Mekkes et al. 2012).

In our analysis of the effectiveness of urban containment policies we focus on the ambitions for urban intensification that we define as the process of increasing the number of housing units within existing urban area. The national strategic urban intensification policy goal was previously set at a minimum realization of 40 percent of the addition to the housing stock within the existing urban area of 2000 (VROM et al. 2004). This was a net ambition, representing the absolute change in housing stock comprised of the balance of newly constructed housing units minus demolished units while excluding renovated units. However, at the regional level policy ambitions seem to differ substantially. In order to obtain insight about the degree to which these policy ambitions differ regionally, we developed an inventory

of current policy ambitions with regard to urban intensification. The most recent relevant strategic visions were collected for each region or municipality. Appendix 1 presents the results of this inventory at the COROP-level.[1]

An interesting pattern emerges from these intensification ambitions. In the Randstad, which is the most urbanized western part of the country, the intensification ambitions are typically more strict than in most other parts of the country. To illustrate, in the Greater Amsterdam region and the areas around Rotterdam and The Hague (see Fig. 1), the intensification ambition is for 60 percent of the new development to occur within the urbanized area (Keers et al. 2011), while the goal for the metropolitan area around The Hague (i.e., Haaglanden) is for 80 percent of new house units to be constructed within the existing urban area. Therefore, in the urban areas that already have the highest densities, the ambition levels are also set highest. In the regions with population decline at the edges of the country, ambitions regarding urban intensification are high for a completely other reason: more compact urban forms are expected to help maintain current urban facility levels (i.e. shops, restaurants, public services, etcetera).

With the approval of the new national strategic vision for infrastructure and the environment (I&M 2011), the national government does not retain the national policy goal of 40 percent urban intensification (40 percent of new development must occur within urbanization boundaries) anymore. Instead, the national government has adopted a three-step approach with regard to urbanization, recognizing the differences between regions and the need for tailor-made policies. First, demand for specific developments is examined. Next, the priority is to determine whether existing urban areas or existing buildings can be used to accommodate the demand. Only when an extension of the urban area is truly necessary, will this be allowed. With this new policy, the national government effectively gives provinces and municipalities the mandate to set their own policies with regard to urbanization and maintaining open space.

Prior analysis shows that over the past decade, a larger share of the additional housing stock (construction minus demolition) was built as urban extensions outside existing urban areas than was called for in the relevant planning documents (PBL 2010; Keers et al. 2011). Although substantial regional differences were observed in this study, the general suggestion was that the provinces comprising the Randstad had greater difficulty in meeting their intensification goals. The apparent discrepancy between high intensification goals and low degrees of success in implementing them in the most urbanized regions of the country is an interesting urbanization paradox that provides the motivation for this study. With the use of actual data for urban intensification scores, we aim to contribute to evidence-based planning. Such information can be used to adjust existing plans, or to draft new plans and policies that may prove to be more feasible.

[1]The Netherlands has 40 COROP-regions. This administrative division is often used to represent statistical data and is identical to the European NUTS 3 division.

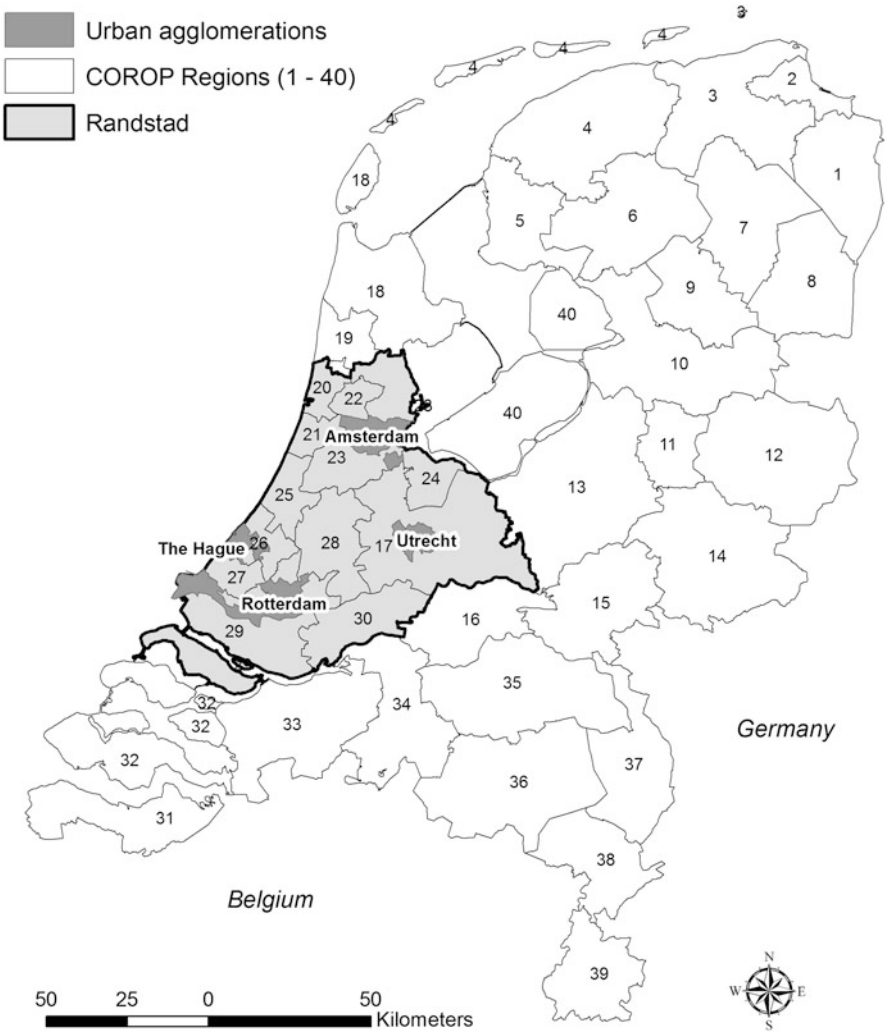

Fig. 1 Locations of the Randstad, the administrative (COROP) regions used in our analysis and the four largest urban agglomerations of the Netherlands (Source: VU-SPINlab, 2013)

Future Urbanization Scenarios

To show the potential impact of various urban intensification policy levels upon future urbanization patterns, we apply a land use simulation model. Such models are useful tools to support the analysis of the impact of spatially relevant policies (Koomen et al. 2008a). They can, for example, simulate potential autonomous spatial developments outside of the urbanized area and show the possible consequences of various policy alternatives.

In this study we apply a land use model to simulate potential urban development patterns according to two scenarios taken from the *Welfare, prosperity and quality of the living environment* scenario study (CPB et al., CPB et al. 2006). In that study, four scenarios are introduced which follow the well-known Special Report on Emissions Scenarios, or SRES-scenarios, of the International Panel on Climate Change (IPCC 2000). The scenarios differ in their emphasis on basic societal trends, and the balance of globalization versus regionalization and economic versus environmental values. To limit the number of results we have selected the two most divergent scenarios from the aforementioned study: the Global Economy (GE) and Regional Communities (RC) scenario.

The Global Economy scenario is part of the A1 scenario family in the SRES terminology; the scenario is based upon substantial population growth and strong economic growth. In the Regional Communities scenario (based upon the B2 scenario family of SRES) the population remains more or less stable, with modest economic growth and a higher unemployment rate. This scenario expects the population in the Randstad area to continue growing, but assumes a population decrease in certain peripheral areas. The two selected scenarios offer the most extreme outcomes in terms of the key variables that steer urban development (demographic change and economic growth), and thus bracket the range of potential changes. In this respect, they help in visualizing the most extreme situations that policymakers are likely to encounter. The inclusion of additional scenarios with intermediate values for key variables could potentially overburden policymakers with too much information.

For this study we translated the regional population projections associated with the two scenarios into a demand for urban land based upon two alternatives: (1) current policy ambitions related to urban intensification (i.e., share of net new housing stock to be facilitated within existing urban areas and prescribed dwelling densities in number of housing units per hectare); and (2) observed intensification results over the past decade. The empirical analysis underlying the latter alternative, and the resulting land demands, are discussed in subsequent sections. By combining two socio-economic scenarios (Global Economy/GE and Regional Communities/RC) with two alternative intensification options, we are able to produce four alternative simulations of future urbanization patterns that allow us to assess the potential impact of current policy ambitions under different demographic conditions (Fig. 2).

Figure 3 puts the projected growth in number of inhabitants and households in a historic perspective. The graph shows a continuous growth in the number of inhabitants and households since 1970, a trend which appears to slow down by the end of the century. The growth, according to the GE scenario, is mainly caused by more open migration policies in line with historic development. The RC scenario, on the other hand, shows a trend breach, assuming a stabilizing population following from an aging population and limited immigration (Hilbers and Snellen

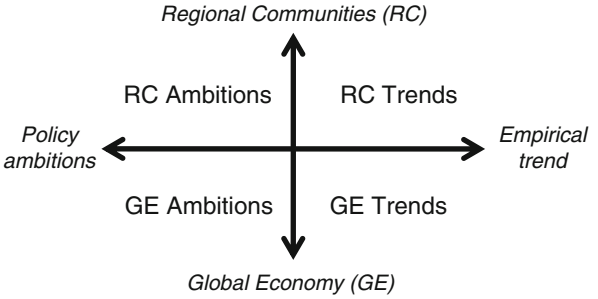

Fig. 2 Combination of two socio-economic scenarios Global Economy (GE) and Regional Communities (RC) with two urban intensification alternatives (policy ambitions vs. empirical results) yields four possible futures

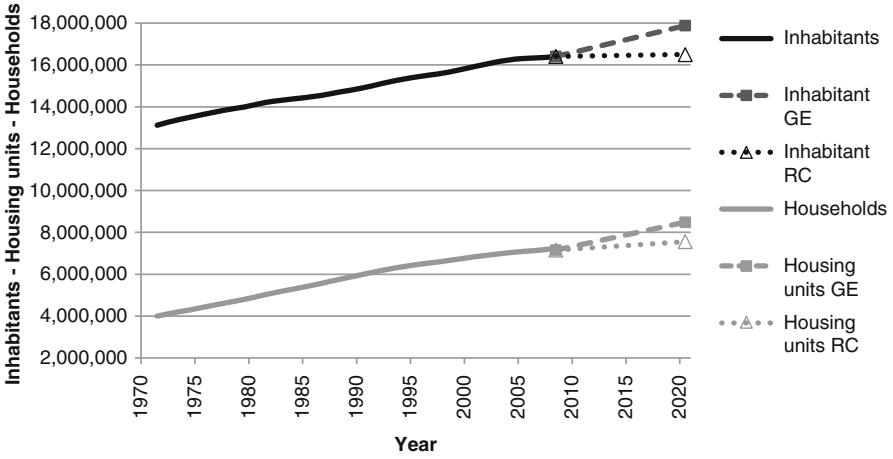

Fig. 3 Growth in number of inhabitants, households and housing units in the Netherlands, 1970–2020 (Source: Statistics Netherlands 2012 for 1970–2008; TIGRIS XL model of the PBL Netherlands Environmental Assessment Agency for 2008–2020)

2010, pp.25–28). Figure 3 also shows that average household density has been decreasing substantially over the past decades, causing the number of households to increase almost as fast as the total population. This particular trend is assumed to continue in the GE scenario, further reducing household size from 2.27 persons (2008) to 2.07 in 2020. Household size is expected to remain stable at 2.24 in the RC scenario.

The remainder of this chapter is organized as follows. In section "Methodology and datasets", we present an overview of the applied methodologies and datasets – In particular the spatial analysis techniques used, and our land use simulation model,

Land Use Scanner, is discussed. Subsequently, we analyze past developments in the density of new urban areas, and the share of new dwellings facilitated in existing urban areas, in section "Analyzing urban development". Then, in section "Simulating urban development", we switch from empirical analysis to scenario simulation by taking two socio-economic world views and combining these with either Dutch urban policy ambitions or an empirical trend extrapolation. Section "Discussion and conclusion" then ends the chapter with conclusions and reflections on the results.

Methodology and Datasets

This section describes the methods and data sets we apply in our analysis. First, we discuss the spatial analysis approach we apply to describe local changes in residential density between 2000 and 2008. Then, we introduce the Land Use Scanner model used to simulate land use changes.

Spatial Analysis Techniques

Spatial analysis and detailed geographical data form a powerful combination to assess the effectiveness of spatial policies. By 1992 Longley and others had used this combination to assess the influence of Green Belt zoning regulations in Britain by comparing the geometry of settlements which were subject to such policy to those which were not (Longley et al. 1992).

In our first analysis, we focused on residential density change establishing differences between urban development within and outside of existing urban areas. To this end, we employed several detailed raster datasets with a 100-meter resolution which described the urban area and contained the number of housing units in 2000 and 2008. In this analysis urban area is defined to include account all types of urban land use (mainly residences, but also associated elements such as shops, public buildings, parks, sports fields et cetera) except large industrial sites. Housing units (in this chapter also called dwellings) refer to individual residential objects that may be contained within larger buildings (e.g. apartments) or that differ in other respects from regular houses (e.g. houseboats). The datasets containing number of housing units per 100-meter grid cell are built based on a combination of sources related to the spatial location of built structures and households, as described in Evers et al. (2005). Due to slight data definition changes over time and the fact that each of the included data sources is aimed to fulfil different objectives the resulting number of housing units can differ from the official figures of Statistics

Netherlands (CBS). The 2000–2008 period was selected for analysis because it is longest possible time span without substantial data definition issues. The datasets with numbers of housing units, provided by the PBL Netherlands Environmental Assessment Agency, contain nearly the same municipal housing totals as Statistics Netherlands; there are differences of less than 0.3 percent of the total number of housing units per municipality between both datasets.

By comparing the detailed housing stock datasets of both years in a geographical information system (GIS) we can analyze local changes in number of housing units. The analysis distinguishes between: (1) intensification share (percentage of net addition to the housing stock within the 2000 urban area); (2) density of urban extension (expressed in number of housing units per hectare of newly developed urban area between 2000 and 2008); and (3) rural housing share (percentage of net addition to the housing stock outside the urban area of 2008). For the first case we made use of the boundary of existing built-up area in 2000 that was developed by the Ministry of Infrastructure and Environment in order to evaluate urban development processes. An extensive description of this dataset is provided by Odijk et al. (2004). To discern the new urban areas developed between 2000 and 2008 we compared the two land use datasets of these years provided by Statistics Netherlands. These datasets were used as base maps in the subsequent land use simulation efforts.

We use zonal statistics to sum cell values of the raster-based datasets within different regions. By comparing the total number of dwellings on an aggregate level for different years, we can calculate growth in housing stock in absolute and relative terms, in order to obtain insight about the net growth in housing stock and the changes in urban density. Zonal statistics can be created via standard GIS procedures available in most GIS-software packages.

Land Use Scanner

The second part of our study involves the simulation of future spatial patterns. The model we apply here, Land Use Scanner, is rooted in economic theory and has been applied in a large number of policy-related research projects in the Netherlands, in Europe and in other regions and countries. The model includes all types of land use, not only urban but also natural and agricultural functions. The Land Use Scanner is a GIS-based model that simulates future land use. It uses an expected regional land demand for different types of land use along with a local definition of suitability for these types of land use in order to create an integrated map of future land use (See Fig. 4 for an overview of the main components of the model).

The regional demand is typically derived from external sector-specific models or specialized sources. Local suitability is defined within the model using a large number of detailed geo-datasets that define the local conditions for each land use

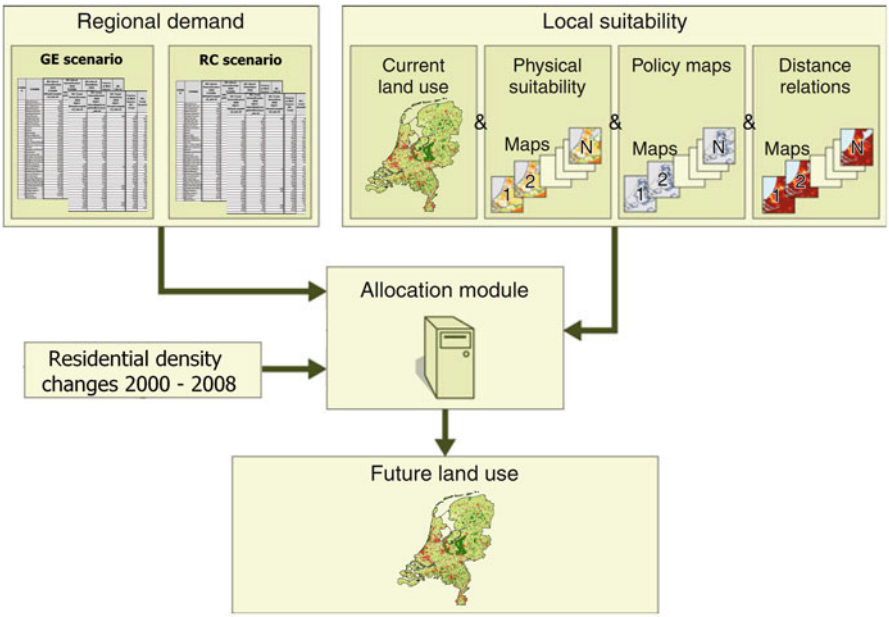

Fig. 4 Basic layout of the Land Use Scanner model

type, and for each 100 by 100 m grid cell in the country. The land use for the base year of the simulation is used as an input here. Biophysical properties of the land, such as soil type and groundwater level, offer additional input for defining land suitability, in particular for agricultural and natural functions. In addition, policy maps such as municipal zoning plans, nature conservation areas and buffer zones are included. Finally, thematic maps are included, generally expressing distance relations (e.g., accessibility or distance from similar land uses assumed to influence commercial and residential developments). The local suitability factors included represent a positive or a negative influence on the suitability value of a land use type.

The allocation module uses an efficient algorithm which determines the optimal future land use based upon a given set of local suitability inputs and regional demands. Both aforementioned input components are processed by the allocation module of the model. As a result, the model allocates land in fractions per grid cell to the different land use types. These fractions function as probabilities signifying the inclination of a cell to adopt a particular type of land use. These probabilities reflect the demand for various types of land use and the scarcity of available land suitable for certain uses. The algorithm applied here is a doubly-constrained logit model that, in its simplest form, can be formulated as:

$$M_{cj} = a_j{}^* b_c{}^* e^{Scj}$$

in which:

M_{cj} is the proportion of land in cell c expected to be used for land use type j;
a_j is the demand balancing factor (Constraint 1), which ensures that the total amount of allocated land for land use type j equals the sector-specific demand;
b_c is the supply balancing factor (Constraint 2), which ensures that the total amount of allocated land in cell c does not exceed the amount of land that is available for that particular cell;
S_{cj} is the suitability of cell c for land use type j based upon its physical characteristics, operative spatial policies and neighborhood relations.

The allocation results from an iterative process that mimics the bidding process of competing land use types. This process assures that all types of use fulfill their regional demand while trying to obtain their preferred (most suitable) locations. Land use types with a relatively high demand compared to the amount of land that is initially allocated to them, will offer the highest bid on the available land.

In the context of strategic, scenario-based national planning, the model proved to be a very useful tool to inform policymakers about potential future developments (Borsboom-van Beurden et al. 2007; Dekkers and Koomen 2007; Kuhlman et al. 2013). The model also proved useful for providing ex-ante evaluations of policy alternatives in national (Van der Hoeven et al. 2009) and regional contexts (Koomen et al. 2011; Jacobs et al. 2011) in the Netherlands and abroad (Hoymann 2010; Te Linde et al. 2011). The basics of the model and the process of its calibration have been described extensively elsewhere (Koomen and Borsboom-van Beurden 2011; Loonen and Koomen 2009).

In the application of this model presented in this chapter, we specifically focused on the impacts of policy ambitions related to urban development. The regional demand for urban area was based upon our own analysis and assumptions, as discussed in section "Simulating urban development". Suitable locations for urban area were determined by: proximity to current urban areas; known urban development plans as incorporated in municipal or regional planning documents were included as well. This information is taken from an extensive survey performed by PBL Netherlands Environmental Assessment Agency.

In addition, a demand for natural areas was taken from the national plans for development of the National Ecological Network. In doing so, we followed the most recent national ambitions documented in the new national spatial strategy (I&M 2011; Elings et al. 2011). These revised plans foresee a more limited network of natural areas, for which new natural areas will only be developed on already acquired agricultural land, with plans for its eventual conversion. Data regarding the amount of land involved and its location is taken from a recent inventory of ratified natural areas plans (DLG 2010). For more information about these claims and definitions of suitable locations for natural areas, we refer to Dekkers et al. (2012).

For all other land use types (e.g., agriculture, greenhouses, industry, recreation) the future demand is set to the amount of land present in 2008. The suitability for these land use types is based solely on their 2008 location. By doing this we keep the non-urban land use patterns stable and, thus, maintain our focus on the impacts of the policy ambitions related to urban development.

As an increase in urban area is expected in almost all regions, land will have to be provided in areas currently allocated to other types of land use. In decades past the space required for urbanization was almost exclusively provided by grassland and arable farmland (Koomen et al. 2008b), therefore the model allows less land to be allocated to these types of land use. Furthermore, land for development is also provided by the land use type known as building lots (i.e. sites that have already been prepared for building activities).

Analyzing Urban Development

The observed residential density changes between 2000 and 2008 are summarized at the COROP level in Fig. 5 and included in Appendix 1. The appendix also includes the names of the COROP regions; in combination with Fig. 1, this serves as a reference for the reader as we discuss the trends hereafter.

The intensification share, which represents the percentage of new housing units built within the urban area existing in 2000, shows a distinct regional pattern. High values are generally encountered in the south, in parts of the Randstad and, to a lesser degree in the northeast of the country, whereas the lowest intensification shares (below 40 percent) are mainly found in the major urban areas of the Randstad (e.g. around The Hague, Rotterdam and Utrecht). Another area with a strikingly low intensification share is the province of Flevoland that consists of former lake areas that were reclaimed in the twentieth century. The existing urban areas in this region are typically less than 30 years old and apparently do not offer much opportunity for intensification. Moreover, this province has been designated to accommodate a substantial part of urban growth in the Randstad region in the past decades in new urban extensions. Another remarkable result is found for the COROP region of Delfzijl (in the far northeast of the map at left in Fig. 5) where the total number of housing units within the existing urban area declined while some urban development took place outside the existing urban area. Apparently, urban shrinkage and urban extension occurred simultaneously here between 2000 and 2008, indicating how developments may take place in other regions which may transition into areas of population decline in the following decades.

Appendix 1 shows that the policy ambitions for the intensification shares of new dwellings are in general higher than what is empirically observed. This is especially true for many urban areas in the Randstad where the ambitions are highest. As can be expected, the share of rural housing (denoting the newly developed, dispersed

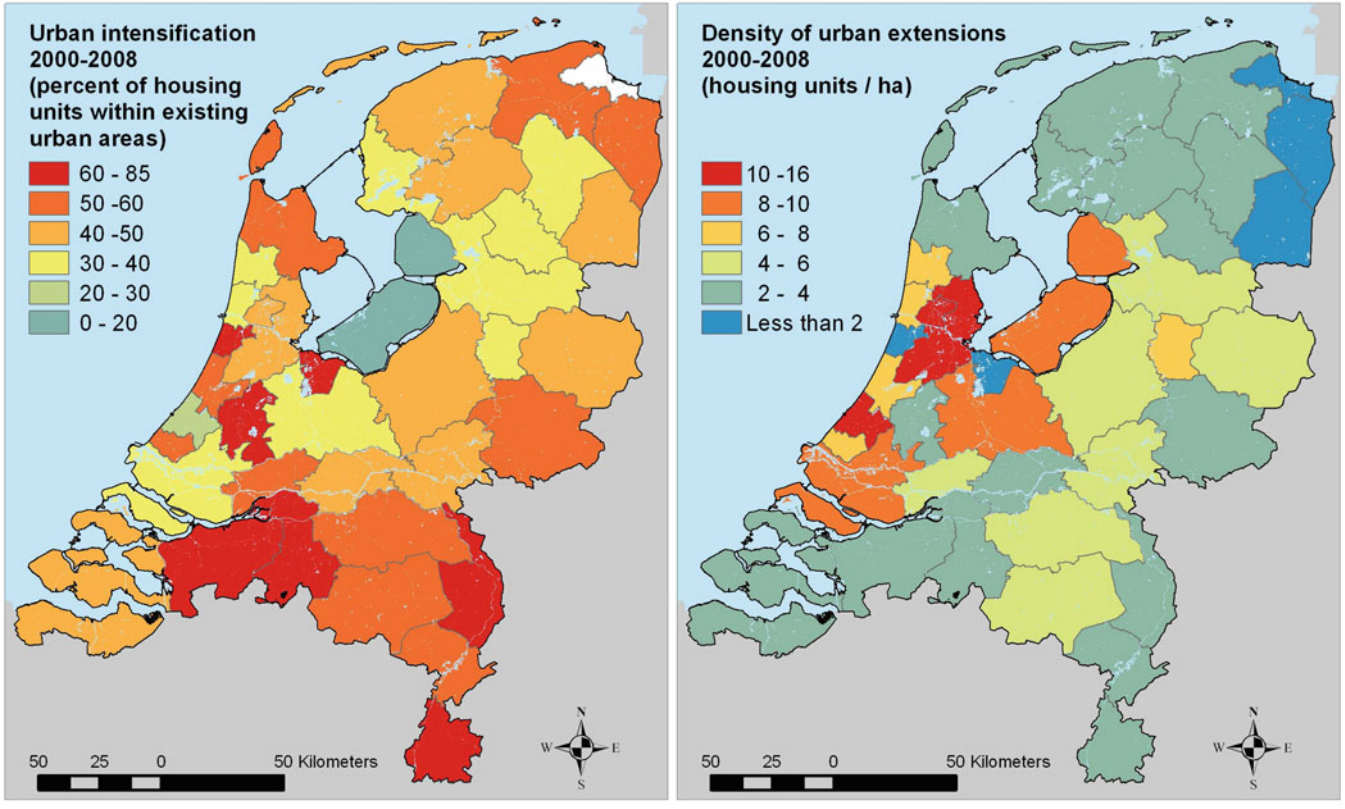

Fig. 5 Urban intensification (percent of the total number of new dwellings added within the urban area contour of 2000) and densities of urban extensions areas for the period of 2000–2008. The COROP region of Delfzijl (in the far northeast) is left blank in the map at *left* as the total number of housing units within the existing urban area has declined

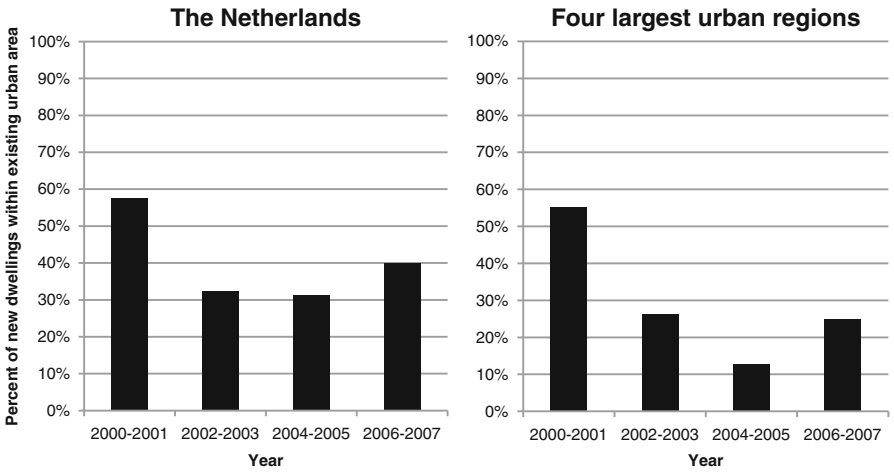

Fig. 6 Observed intensification share (percent of the total number of new dwellings added within the urban area contour of 2000), during the periods of 2000–2001, 2002–2003, 2004–2005, and 2006–2007 for the Netherlands and a group of the four largest urban regions encompassing Amsterdam, Rotterdam, The Hague and Utrecht (Source: Statistics Netherlands 2012)

dwellings that can be found outside of the mapped urban area of 2008) is typically low in the Randstad area (around 10 percent) and higher in the peripheral parts of the country (around 20 percent).

The intensification shares in the observed period are, of course, no guarantee for future developments. An obvious limitation might be that over time, it becomes increasingly difficult to find locations available for intensification. This process is hinted at in Fig. 6, which shows regional statistics on the intensification share for four subsequent 2-year periods. The figure compares the aggregate national intensification shares with the regional intensification shares for the COROP regions that include the four largest cities (Amsterdam, Rotterdam, The Hague and Utrecht). It appears to indicate that the intensification share, especially in the COROP regions encompassing the four largest cities, shows an overall decline from 2000 to 2007. In each time period, the intensification share of the group of four regions is lower than the share of the nation as a whole; as well, the group of four regions consistently shows intensification shares that are well below the high ambition levels (60–80% intensification) for these regions.

New urban areas created after 2000 are, in general, characterized by low densities, especially in the peripheral zones of the country. The lowest figures (less than 2 housing units per hectare) were observed in the northeast. Most northern areas and areas in the south-east and south-west of the country show densities in the range of 2 to 5 housing units per hectare. In most parts of the Randstad, the densities are

much higher (more than 10 housing units per hectare in the Amsterdam metropolitan region and in The Hague). The difference between those two extremes reflects the relative scarcity of land available for development in highly urbanized areas of the Randstad region, as compared with the country's periphery. The residential densities we observe are lower than typically described for urban extension plans. This is caused by the fact that we take into account all urban space required for complete urban extensions. In addition to the space taken by residences, this refers to elements such: local infrastructure, parks, schools, shops, businesses et cetera.

To a large extent, the empirically observed densities in the newly-formed urban areas mirror the intensification shares: high densities are found in regions with relatively low intensification shares, and vice versa. This may be related to the available amount of land available for urban development in relation to the demand for new dwellings. In peripheral regions, the demand for new dwellings is relatively low and can, to a large extent, be accommodated within existing urban areas that still have space available. When urban extensions are created in these regions, there is no strong incentive to build in high density as sufficient space is available. Conversely, in the more densely populated Randstad areas, where the demand for new dwellings is high, space is relatively scarce both within and outside of existing urban areas. This implies that options are limited to further intensify urban development within existing urban areas, and that urban extensions will also have little space available to accommodate a large demand.

It should be noted, however, that diversions from these general trends exist. Some regions within the Randstad are characterized by high intensification shares and show low densities in their urban extensions (e.g., Haarlem and Het Gooi). This may be partly related to the limited number of new housing units that were added to these relatively affluent commuting towns, which lack a strong increase in new inhabitants. Furthermore, the few dwellings added in these relatively wealthy regions are most likely larger than those of typical urban extensions.

Simulating Urban Development

Policy Ambitions Versus Trend-Based Extrapolations

The simulation of future urban development is based on the projected developments in total housing stock through 2020 according to the Global Economy (GE) and Regional Communities (RC) scenarios that were discussed in section "Future urbanization scenarios". Figure 3 shows that the net demand for new dwellings is much higher in GE than in the RC scenario; reflecting the higher population growth and the continuing decrease in average number of people per household in the GE-scenario.

These national projections were employed by PBL Netherlands Environmental Assessment Agency to obtain regional household projections for 2020 as part of their Delta Scenarios Study (Rijken et al. 2013). For each scenario, the net regional increase in housing stock was provided by the TIGRIS XL model, which calculates housing demand via a system of model components that include dynamic interactions (Significance and Bureau Louter 2007; Zondag and Geurs 2011). In essence, the model is comprised of labor market model that interacts with both a commercial real estate market model and a land market model via a transport model. The model uses information on regional workforces, labor demand, demographic data, data on commercial sites and numbers of households, housing units, persons, firms and jobs. For more information about the model components and variables we refer to Zondag and De Jong (2005).

For each of the two socio-economic scenarios, we then created two urban intensification alternatives in which the additional demand for urban areas is based upon either the observed intensification shares during the 2000–2008 period ('Trends' alternative) or the regional policy ambitions related to urban intensification ('Ambitions' alternative). To obtain an additional urban area demand from the projected regional changes in housing stock we apply the following steps based on the COROP region-specific outcomes of our empirical analysis (as listed in Appendix 1). First, we subtract the number of housing units that are likely to be built outside of formal urban areas, and that are thus not simulated as part of the urban land use category. The number of units belonging to this category is based upon the share of rural housing observed in the housing stock increase in the 2000–2008 period.

The intensification share (based upon either the observed values for the 2000–2008 period or the policy ambitions) is then subtracted in order to account for those dwellings that are expected to be facilitated within the urban area that exists in the base year of simulation. The remaining share of housing units is expected to form new urban areas, and the size of these areas is determined by dividing the corresponding number of units by the residential densities observed in the urban extensions formed between 2000 and 2008. The resulting additional urban demand for 2020 is then used as input for the Land Use Scanner model.

For the (near) future, growth is still expected in many parts of the country, but the projections for individual regions can differ substantially (PBL 2011). In some peripheral regions in the Netherlands, a population decline is expected – a trend that is already observed in the most peripheral regions in the northeast, southeast and southwest. The high-growth regions are mainly located in densely-populated and already highly urbanized areas.

Figure 7 shows the projected urban density at the national scale for the different simulations. The presented figures result from: (1) the projected net additional demand for dwellings, which differs for each scenario (GE versus RC), (2) the intensification shares that differ per alternative (policy ambitions versus empirical results); and (3) the observed density in the urban extensions formed between 2000 and 2008 (equal for all simulations). In both 'Trends' alternatives, the aggregate national urban density is lower than in the equivalent 'Ambitions' alternatives,

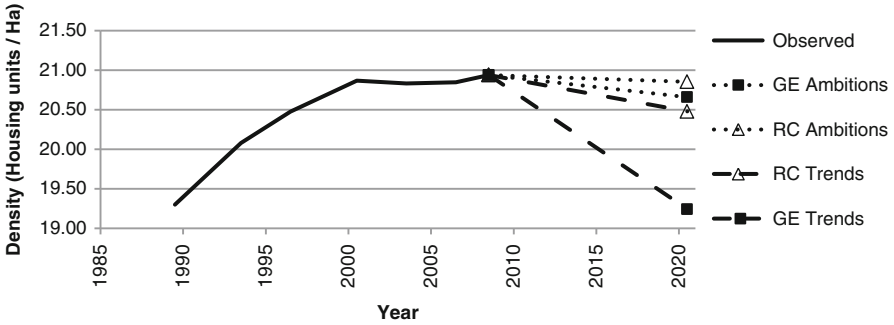

Fig. 7 Observed urban density (dwellings/ha) in the Netherlands through 2008 (Source: Statistics Netherlands 2012). The expected densities in 2020 are based on the combination of the scenarios and urban intensification alternatives (policy ambitions versus empirical results)

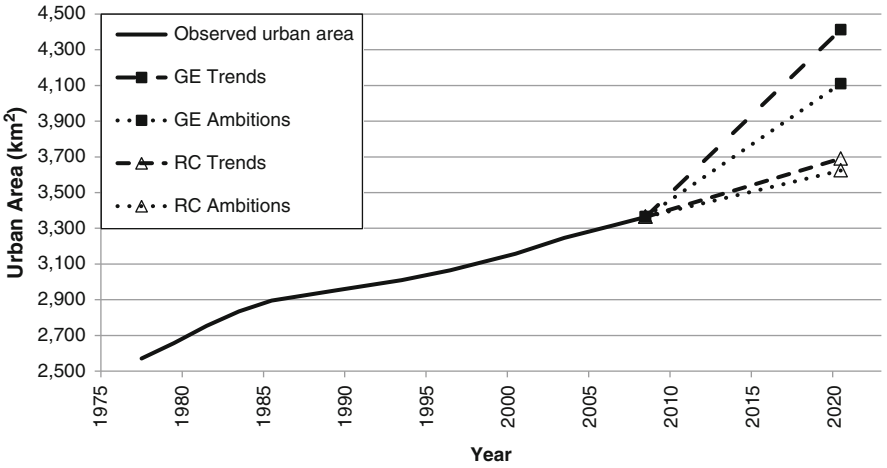

Fig. 8 Observed change in the amount of urban area in the Netherlands (in km^2) through 2008 (Source: Statistics Netherlands 2012) compared to our simulation alternatives

showing that ambitions are indeed needed to maintain current density levels. In the GE scenario overall density levels are expected to drop because of the addition of many new housing units in urban extensions with relatively low densities.

The expected growth in urban area for 2020 which results from the aforementioned calculations is displayed in Fig. 8, together with the observed growth between 1977 and 2008. From the figure it is clear that the major difference in expected amount of urban area is more between the GE and RC scenarios, than between the Ambitions/Trends alternatives (4412 km^2, 4109 km^2 for GE Trends and GE Ambitions respectively, as opposed to 3690 km^2 and 3624 km^2 for RC Trends and RC Ambitions). In examining the demand for each of the four simulations, we observe that the difference in intensification shares indeed appears to be associated

with a difference in demand for urban area between the two GE scenario alternatives and the two RC scenario alternatives. This suggests the need for analysis and discussion of the potential impact of the alternatives under different socio-economic conditions.

Simulation Results

The main results of our simulations are displayed in Fig. 9. We stress that these simulation exercises are obvious simplifications of reality. For instance, we have only included a limited number of suitability maps and land use types for these simulations, and no additional land use demand for land use types other than urban area and nature. The upper-left map presents the main land uses in the central part of the Netherlands in 2008. The urban footprint of the Randstad cities is displayed clearly.

The simulation result of the GE Trends alternative is displayed in the upper-right map (Fig. 9). This is the most extreme scenario in terms of new urban extensions, since it assumes a high population growth rate and a decreasing number of persons per household, along with relatively high shares of new dwellings located outside of existing urban areas. The differences between the GE Trends simulation result and the 2008 land use map are considerable. We see that urban settlements grow together and an almost fully-urbanized crescent encompassing the main Randstad cities emerges. In parallel, a large land conversion process takes place in the Green Heart (the open green area between the major urban agglomerations of the Randstad), where relatively small cities increase their urban footprint by means of additional extensions.

The lower maps (Fig. 9) display the likely urban extension areas in the Ambitions alternatives (with RC Ambitions at middle-left and GE Ambitions at lower-left) and the Trends alternatives (with RC Trends at middle-right and GE Trends at lower-right). The expected increase in urban areas is demonstrated via a color intensity that reflects the probability of urban development (in fractions per grid cell, see section "Land use scanner"). Urbanization probabilities between 0.2 and 0.4 are considered low, higher than 0.4 but below 0.6 are defined as medium, and above 0.6 probabilities of urban development are categorized as high. The maps clearly show that in the GE scenario, there are more urban extensions than in the RC scenario. The GE scenario shows a clear difference between the Ambitions and Trends alternatives. Current policy ambitions related to urban intensification will greatly assist in containing the large urban growth that is expected in the GE scenario. Under the much more moderate conditions of the RC scenario, this difference is less clearly visible.

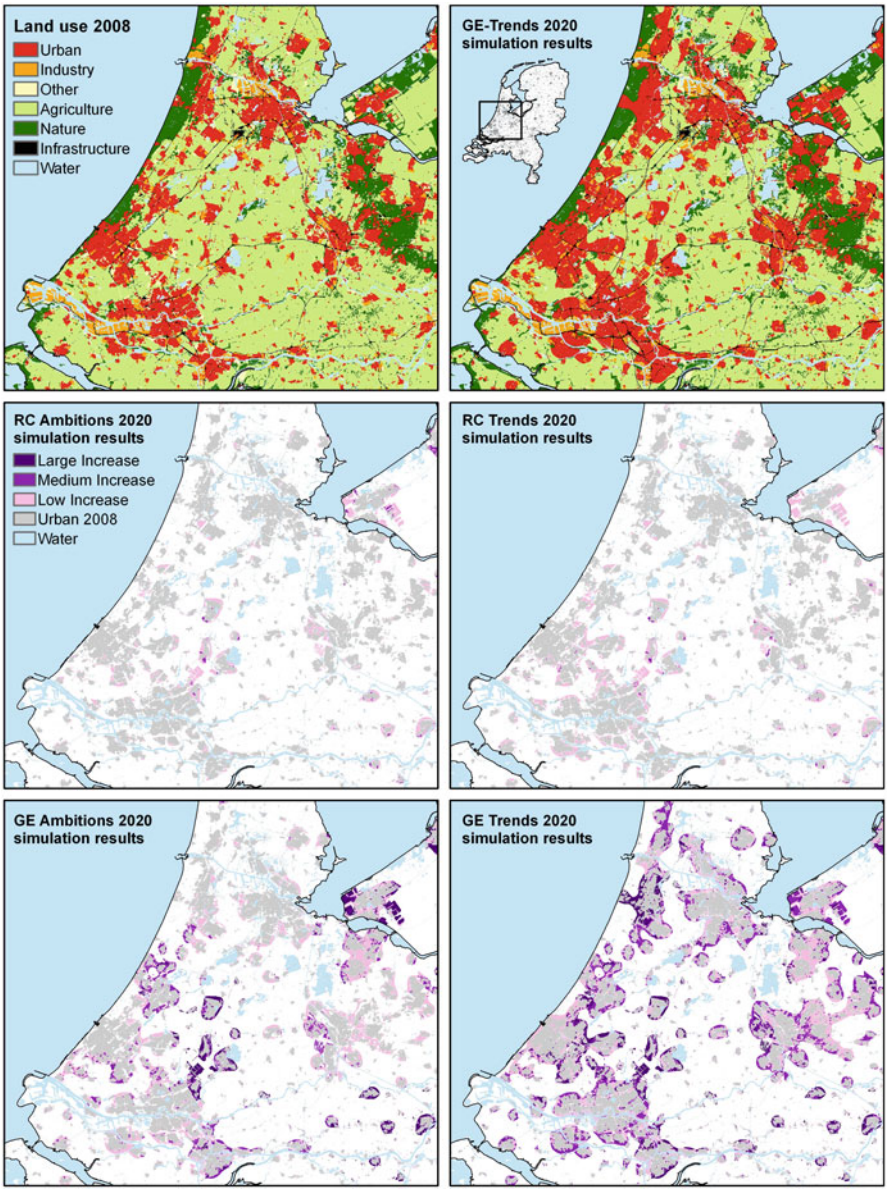

Fig. 9 Current (2008, upper-left) and simulated (2020, upper-right) land use according to the GE Trends alternative. The maps below display the simulated increase in urban area according to each of the four scenario and intensification alternative combinations: RC Ambitions (middle-left); RC Trends (middle-right); GE Ambitions (lower-left); and GE Trends (lower-right). Existing urban areas in 2008 are depicted in grey, while the yellow/red colors represent diverse grades of expected urbanization pressures according to the simulations

Discussion and Conclusion

Conclusions from the Empirical Analysis

One of the goals of urban planning policies in the Netherlands is to control the spatial footprint of urbanized areas in order to maintain relatively compact cities. An important policy instrument to achieve this objective is to specify urban intensification targets, defined by the share of new dwellings to be developed within actual urban boundaries. Ex-post evaluation of such policies can shed light on the degree of success of current policy ambitions. Therefore, we used geospatial analysis to perform ex-post evaluation of urban intensification policies. Using the evaluation results, we subsequently simulated plausible future urban development scenarios and compared policy ambitions and scenario outcomes.

In general, policy ambitions regarding urban intensification share were met in the past, but large regional differences exist. Relatively low intensification shares prevail in the larger urban areas of the Randstad where ambition levels are typically very high. These low values seem to coincide with high residential densities in the urban extensions. The combination of high intensification ambitions and relatively poor realization was introduced as the 'urbanization paradox' in the beginning of this chapter, and our analysis has at least partially confirmed its existence.

A probable explanation is that this phenomenon is caused by shortage of space within urban areas in the already densely populated cities. The most feasible building locations are likely to have been used for intensification in the past decade, making it more difficult for policymakers in these regions to realize their high ambitions. This explanation is supported by the fact that low intensification shares in the Randstad typically coincide with high residential densities in urban extensions, which is also assumed to result from the limited amount of available space. Some regional exceptions to this general trend exist in the Randstad area. Higher intensification rates and lower densities in extensions are found in COROP regions. This is often the case where relatively few new housing units are added.

Conclusions from the Simulation Exercise

Based upon the empirical analysis we are able to show that current ambition levels are needed to prevent extensive loss of open (non-urbanized) space. Should urban development (in terms of intensification share and urban extension densities) follow past trends, large-scale urban extensions are likely to occur. This is especially true when future socio-economic conditions resemble the Global Economy scenario. Whether or not this scenario with its associated strong increase in number of

households will become reality is beyond the sphere of influence of local or regional decision makers, but depends on (global) macro economic conditions and European and national regulations related to, for example, immigration. Local restrictions on urban development could, however, divert growth to neighboring towns.

Our simulation can easily be expanded to accommodate additional spatial restrictions or other types of development (e.g., agriculture, industry/commerce, recreation). Such simulations have been performed as part of other studies focusing on, for example, the environmental impacts of a new regional strategic vision (Koomen et al. 2011) or the potential for specific biofuel crops in the country (Kuhlman et al. 2013).

Discussion

Some general drawbacks should be considered in relation to the analysis presented in this chapter -the most important issue being that past developments offer no guarantee on the trajectory of future developments. This relates, for example, to the fact that policy objectives for intensification shares may not be realistic anymore because the most feasible urban redevelopment locations have already been put to use, as discussed earlier. But other (external), currently underexplored options or changing societal conditions may become important influences upon the possibilities for urban intensification and thus limit the validity of our simulation outcomes. We will discuss several of these potential developments here.

First, we may underestimate the option for transformation of existing vacant buildings like factories or offices into dwellings. Especially office buildings may offer opportunities for urban intensification as about 14 percent of the total office building stock was estimated to be empty in 2011 (DTZ Zadelhoff 2011), giving the Netherlands the highest office vacancy rates in Europe (Seebus 2012). A substantial portion of this empty building stock is expected to remain empty because the condition or location does not match current market preferences. Extensive office parks at poorly accessible locations thus run the risk of becoming the brownfields of the twenty-first century and are in need of redevelopment programs similar to those that helped to revitalize the former docklands in cities like Amsterdam and Rotterdam. Transformation clearly has advantages, but there are also challenges to be overcome – like the changing of local zoning designations, differences in building regulations for offices and dwellings, the relatively high costs involved in renovation and transformation and the lack of government funding in the current political and financial climate. The increased attention for the historical-cultural value of certain buildings and locations ensures at least some government funding for transformation of, for instance, old factories into dwellings. Also, over the coming decade 1100

church buildings are expected to become vacant (ND 2012). One way to preserve these often monumental buildings is to change them into apartments or give them another function, for instance, public services (RCE 2012). As the transformation of historic buildings and churches typically is a time-consuming and costly process it is not likely to become widespread practice, and it will probably be confined to popular cities with a high demand for attractive housing.

Second, since the establishment of the new national strategic vision on infrastructure and environment (I&M 2011), specific national restrictive zoning regulations (such as Buffer zones[2]) and nationally designated urban development zones (called Bundling zones) are effectively abolished, leaving the responsibility for urban containment with regional and local governments. This changing planning (discussed extensively elsewhere – see Koomen and Dekkers 2013; Roodbol-Mekkes et al. 2012) may have considerable impact upon urban development in the Netherlands. In particular, urban intensification processes may become more difficult to realize as local and regional authorities may lack the financial means and other resources necessary to make more expensive and complex inner-city redevelopment projects possible. Especially for the regeneration of brownfield sites, development costs are generally higher than for greenfield sites because land must often be sanitized first (ECORYS 2005; Van Hoek et al. 2011). Also, the complex legal and spatial planning procedures may hinder inner-city development (Warbroek 2011). Local authorities may also lack the knowledge or power to resist market forces that will favor the development of less complex and more profitable extensification projects. In fact, municipalities were often willing to sell land for commercial development as it provided, at least until the financial crisis of 2008, a welcome source of revenue (Van Hoek et al. 2011; Seebus 2012). Such developments may, however, be more costly for society as a whole than urban intensification projects, as is evidenced in societal cost-benefit analysis of extensification versus intensification projects (ECORYS 2005; Van Hoek et al. 2011).

Moreover, local governments are often tempted to out-compete each other in the provision of building lots in order to attract commercial and residential development and companies. This behavior was one of the causes for the over-supply in offices (Seebus 2012). Regional cooperation between local governments is thus preferred to avoid this kind of mutual competition that incurs substantial costs to society (KBB 2012; Kuhlman et al. 2012). An interesting suggestion is to apply the legal framework of ex-ante evaluation for infrastructure projects on housing development as well, thus providing a broader view of, and more knowledge about, the societal costs and benefits of projects to local policymakers. Also, in a recent policy debate in the Netherlands, scholars and policy makers discussed the imposition of a

[2]These green corridors were designated to maintain open spaces between major agglomerations and are discussed extensively by, for example, Koomen et al. (Koomen et al. 2008b) and Van Rij et al. (2008).

development tax on building projects outside existing urban areas, arguing that current policy instruments generate disproportionate negative external effects upon welfare. By replacing the policy instruments with a relative simple price mechanism such as a tax, these external effects would be avoided, while simultaneously making inner-city brownfield development more price-competitive in relation to urban extensification. Korthals Altes (2009) reviews this debate and explains why the tax instrument was, in the end, not implemented in the Netherlands.

Obviously, the current financial crisis has slowed or stopped most urban development. Should this crisis remain, developments may be more likely to correspond to the Regional Communities scenario. In fact, this scenario assumes a considerable population decrease in certain peripheral areas that may call for urban restructuring rather than urban growth. This implies that certain neighborhoods may face the demolition or redevelopment of particular parts of the housing stock. These developments are outside of the scope of this study and cannot be addressed in the current version of our land use model, but we anticipate on incorporating such aspects in our modeling framework as a part of ongoing research projects. One of these projects will explore potential changes in population distribution, associated spatial patterns, potential solution strategies and related impacts. The development of spatial planning and design concepts in this case will focus on accommodating population decline, while paying attention to the relationship between population and necessary facility provision levels. The Land Use Scanner simulation model will be adapted to accommodate the analysis and simulation of these phenomena.

Another related topic explored in a second project is the fact that the city has become more popular as a residential destination within the last decade, certainly among specific groups of people (De Groot et al. 2010). However, these groups may have specific location preferences and housing needs. If there is a mismatch between demand and supply in terms of housing and facilities, these groups may decide to relocate outside the city or to another city. We, therefore, initiated an analysis and simulation of location preferences of highly-educated knowledge industry workers. For the simulation of households in this project, the adapted Land Use Scanner will also be used.

Acknowledgements This research is funded by the research programme Urban Regions in the Delta (URD), part of the VerDuS-programme ('Verbinding Duurzame Steden') of the Netherlands Organization for Scientific Research (NWO). We would also like to thank the PBL Netherlands Environmental Assessment Agency for providing spatial data that was incorporated in both the analyses and the simulations discussed in this chapter. Finally, we want to thank our colleague Ronnie Lassche for the construction of the basic layout of the simulation model used in this chapter.

Appendix 1: Empirical Analysis Results

This table includes the basic results of our empirical analysis and the figures used in the subsequent calculation of additional urban area demand applied in land use simulation

COROP region		Housing stock change	Intensification (%)		Density in urban extensions	Rural housing (%)	Expected housing stock change (2008–2020)		Additional urban area (hectares)			
#	Name	(2000–2008)	Trends	Ambitions	(Dwellings per Ha)	(2000–2008)	GE	RC	GE trend	GE ambitions	RC trend	RC ambitions
1	Oost-Groningen	1770	54	80	0.73	28	6739	0	1640	717	0	0
2	Delfzijl en omgeving	−707	0[3]	100	1.19	8	1199	804	922	9	618	618
3	Overig Groningen	11,056	58	80	2.88	11	32,348	10,870	3427	1635	1152	549
4	Noord-Friesland	7917	44	37	3.41	16	20,404	5463	2371	2662	635	713
5	Zuidwest-Friesland	2961	31	40	3.07	23	6935	2139	1027	897	317	277
6	Zuidoost-Friesland	5547	40	40	3.03	22	13,341	4108	1671	1671	514	514
7	Noord-Drenthe	6725	37	40	3.84	13	15,193	6490	2003	1896	856	810
8	Zuidoost-Drenthe	3652	41	60	1.89	23	8457	2601	1630	1107	501	341
9	Zuidwest-Drenthe	3084	35	40	3.72	22	9366	2697	1086	997	313	287
10	Noord-Overijssel	12,160	34	40	5.47	19	29,308	14,989	2536	2309	1297	1181
11	Zuidwest-Overijssel	5272	34	45	6.85	14	11,300	2608	849	709	196	164
12	Twente	14,356	49	60	4.79	13	43,096	15,443	3422	2667	1226	956
13	Veluwe	18,052	48	40	4.51	15	52,290	19,351	4349	4984	1609	1844
14	Achterhoek	9930	58	40	3.22	18	22,639	11,523	1680	2428	855	1236
15	Arnhem/Nijmegen	19,455	48	56	5.24	13	70,862	26,300	5253	4455	1950	1653
16	Zuidwest-Gelderland	6109	49	40	3.22	17	20,707	4358	2202	2606	463	549
17	Utrecht	44,704	39	66	8.13	11	122,305	43,861	7490	4172	2686	1496
18	Kop van Noord-Holland	12,347	51	51	3.13	11	29,303	12,590	3595	3576	1544	1537

(continued)

(continued)

COROP region		Housing stock change	Intensification (%)		Density in urban extensions	Rural housing (%)	Expected housing stock change (2008–2020)		Additional urban area (hectares)			
#	Name	(2000–2008)	Trends	Ambitions	(Dwellings per Ha)	(2000–2008)	GE	RC	GE trend	GE ambitions	RC trend	RC ambitions
19	Alkmaar en omgeving	8738	35	57	6.70	9	18,332	4924	1534	1017	412	273
20	IJmond	4558	31	100	6.19	12	15,469	1638	1423	21	151	2
21	Agglomeratie Haarlem	3297	83	99	0.51	6	17,149	0	3772	226	0	0
22	Zaanstreek	5608	41	97	11.49	11	15,718	4213	656	33	176	9
23	Groot-Amsterdam	42,261	49	82	10.49	11	100,297	20,463	3838	1389	783	283
24	Het Gooi en Vechtstreek	4365	82	78	1.65	6	21,539	3498	1585	1915	257	311
25	Aggl. Leiden&Bollenstr.	7897	57	60	6.59	6	40,036	9158	2214	2073	506	474
26	Aggl. ‘s-Gravenhage	30,460	21	80	15.18	6	70,369	18,643	3369	856	893	227
27	Delft en Westland	5710	52	60	6.12	11	21,911	8483	1320	1092	511	423
28	Oost-Zuid-Holland	6005	64	60	2.38	16	39,250	11,802	3409	3772	1025	1134
29	Groot-Rijnmond	32,942	30	60	8.34	18	90,329	24,165	5651	3230	1512	864
30	Zuidoost-Zuid-Holland	7388	51	60	4.07	17	30,183	4555	2365	1931	357	291
31	Zeeuws-Vlaanderen	2156	46	75	2.36	10	4087	930	769	353	321	321
32	Overig Zeeland	8253	45	75	3.32	22	15,463	4842	1575	712	493	223
33	West-Noord-Brabant	17,213	61	62	3.74	15	40,844	16,007	2692	2603	1055	1020
34	Midden-Noord-Brabant	13,049	66	90	2.61	17	39,894	17,295	2739	794	1187	344
35	Noordoost-Noord-Brab.	17,801	54	66	4.20	18	55,307	20,446	3746	2741	1385	1013
36	Zuidoost-Noord-Brabant	21,421	57	70	5.43	10	65,997	21,979	4009	2771	1335	923
37	Noord-Limburg	8125	61	93	2.26	19	19,886	5215	1763	315	462	83
38	Midden-Limburg	5747	59	71	2.50	12	10,386	3136	1204	856	876	876
39	Zuid-Limburg	11,915	66	90	2.34	9	32,773	4570	3544	1048	494	146
40	Flevoland	23,928	18	3	8.35	11	52,224	19,948	4455	5281	1702	2017

[3] Delfzijl lost dwellings within its 2000 urban area, but some units were added in urban extensions

References

Borsboom-van Beurden, J. A. M., Bakema, A., & Tijbosch, H. (2007). A land-use modelling system for environmental impact assessment. Recent applications of the LUMOS toolbox. Chapter 16. In E. Koomen, J. Stillwell, A. Bakema, & H. J. Scholten (Eds.), *Modelling land-use change; progress and applications* (pp. 281–296). Dordrecht: Springer.

CPB, MNP, & RPB. (2006). *Welvaart en Leefomgeving. Een scenariostudie voor Nederland in 2040*. Den Haag: Centraal Planbureau, Milieu- en Natuurplanbureau en Ruimtelijk Planbureau.

De Groot, H. L. F., Marlet, G. A., Teulings, C., & Vermeulen, W. (2010). *Stad en land*. Den Haag: Centraal Planbureau.

Dekkers, J. E. C., & Koomen, E. (2007). Land-use simulation for water management: application of the land use scanner model in two large-scale scenario-studies. Chapter 20. In E. Koomen, J. Stillwell, A. Bakema, & H. J. Scholten (Eds.), *Modelling land-use change; progress and applications* (pp. 355–373). Dordrecht: Springer.

Dekkers, J. E. C., Koomen, E., Jacobs-Crisioni, C. G. W., & Rijken, B. (2012). *Scenario-based projections of future land use in the Netherlands*. VU University, Amsterdam. Spinlab Research Memorandum SL-11.

DLG (2010) Natuurmeting op Kaart 2010. De realisatie van Ecologische Hoofdstructuur (EHS) en Recreatie om de Stad (RodS) op 1–1-2010. Dienst Landelijk Gebied, Utrecht.

DTZ Zadelhoff. (2011). *Van veel te veel; De markt voor Nederlands commercieel onroerend goed*. Utrecht: DTZ Zadelhoff v.o.f.

ECORYS. (2005). *Maatschappelijke kosten en baten IBO Verstedelijking. Input voor Interdepartementaal Beleidsonderzoek*. Rotterdam: ECORYS Nederland B.V.

Elings, C., Zijlstra, R., Koomen, E., & De Groot, S. (2011). *Milieueffectrapport Ontwerp Structuurvisie Infrastructuur en Ruimte in opdracht van Ministerie van Infrastructuur en Milieu*. Amsterdam/Nijmegen: Geodan/Royal Haskoning.

Evers, W., Vries, L., De Man, R., & Schotten, C. G. J. (2005). *Woning- en populatiebestanden in het dataportaal. Overzicht van de basisbestanden, bewerkingen en kwaliteitsacties*. Bilthoven: RIVM.

Hilbers, H. D., & Snellen, D. (2010). *Bestendigheid van de WLO scenario's. 500161003*. Den Haag: Planbureau voor de Leefomgeving.

Hoymann, J. (2010). Spatial allocation of future residential land use in the Elbe River Basin. *Environment and Planning. B, Planning & Design, 37*(5), 911–928.

I&M. (2011). *Ontwerp structuurvisie infrastructuur en ruimte; Nederland concurrerend, bereikbaar, leefbaar en veilig*. Den Haag: Ministerie van Infrastructuur en Milieu.

IPCC. (2000). *Emission scenarios. Special report of the intergovernmental panel on climate change*. Cambridge: Cambridge University Press.

Jacobs, C. G. W., Koomen, E., Bouwman, A. A., & Van der Burg, A. (2011). Lessons learned from land-use simulation in regional planning applications. Chapter 8. In E. Koomen & J. Borsboom-van Beurden (Eds.), *Land-use modelling in planning practice* (pp. 131–149). Heidelberg: Springer.

KBB. (2012). *Digitaal Basisboek Bevolkingsdaling.Vragen en antwoorden voor gemeenten: Waarom is concurrentie niet verstandig?* Kenniscentrum voor Bevolkingsdaling en Beleid. http://www.bevolkingsdaling.nl/Default.aspx?tabid=1286. Last accessed 27 Nov 2012.

Keers, G., Smeulders, E., & Teerlink, T. (2011). *Onderzoek: Toekomst voor de stedelijke woningbouw? P15360*. RIGO Research en Advies/Bouwfonds Ontwikkeling.

Koomen, E., & Borsboom-van Beurden, J. (Eds.). (2011). *Land-use modeling in planning practice*, *Geojournal library, volume 101*. Heidelberg: Springer.

Koomen, E., & Dekkers, J. E. C. (2013). The impact of land-use policy on urban fringe dynamics; Dutch evidence and prospects. In D. Czamanski, I. Benenson, & D. Malkinson (Eds.), *Cities and nature* (pp. 9–35). Berlin: Springer.

Koomen, E., Dekkers, J., & Van Dijk, T. (2008a). Open-space preservation in the Netherlands: Planning, practice and prospects. *Land Use Policy, 25*(3), 361–377.

Koomen, E., Rietveld, P., & De Nijs, T. (2008b). Modelling land-use change for spatial planning support; Editorial. *Annals of Regional Science, 42*(1), 1–10.

Koomen, E., Koekoek, A., & Dijk, E. (2011). Simulating land-use change in a regional planning context. *Applied Spatial Analysis and Policy, 4*(4), 223–247.

Korthals Altes, W. (2009). Taxing land for urban containment: Reflections on a Dutch debate. *Land Use Policy, 26*(2), 233–241.

Kuhlman, T., Agricola, H. J., De Blaeij, A., De Hoop, J. G., Michels, R., Smit, A. B., & Vogelzang, T. A. (2012). *Landbouw en recreatie in krimpregio's: knelpunten en kansen. LEI-rapport 2012–001*. Den Haag: LEI Wageningen UR.

Kuhlman, T., Diogo, V., & Koomen, E. (2013). Exploring the potential of reed as a bioenergy crop in the Netherlands. *Biomass and Bioenergy, 55*, 41–52.

Kuijpers-Linde, M. (2011). A policy perspective of the development of Dutch land-use models. Chapter 10. In E. Koomen & J. Borsboom-van Beurden (Eds.), *Land-use modelling in planning practice* (pp. 177–189). Heidelberg: Springer.

Longley, P., Batty, M., Shepherd, J., & Sadler, G. (1992). Do green belts change the shape of urban areas? A preliminary analysis of the settlement geography of south east England. *Regional Studies, 26*(5), 437–452.

Loonen, W., & Koomen, E. (2009). *Calibration and validation of the Land Use Scanner allocation algorithms*. Bilthoven: PBL Netherlands Environmental Assessment Agency.

ND. (2012). *Dossier Kerksluiting, Nederlands Dagblad*. http://www.nd.nl/dossiers/kerk-en-religie/kerksluiting. Last accessed 30 Nov 2012.

Odijk, M., Van Bleek, B., & Louwerse, P. (2004). *Begrenzing Bebouwd Gebied 2000*. Den Haag: Ministerie van VROM.

PBL. (2010). *Balans van de Leefomgeving. 500206001*. Bilthoven: Planbureau voor de Leefomgeving.

PBL. (2011). *Nederland in 2040: een land van regio's*. Den Haag: Planbureau voor de Leefomgeving.

RCE. (2012). Een toekomst voor kerken. Handreiking voor het herbestemmen van vrijkomende kerkgebouwen. Rijksdienst voor het Cultureel Erfgoed. In *Ministerie van Onderwijs*. Den Haag: Cultuur en Wetenschap.

Rijken, B., Bouwman, A., Van Hinsberg, A., Van Bemmel, B., Van den Born, G. J., Polman, N., Lindenhof, V., & Rijk, P. (2013). *Regionalisering en kwantificering verhaallijnen Deltascenario's 2012. Technisch achtergrondrapport*. Den Haag: Planbureau voor de Leefomgeving/LEI Wageningen UR.

Roodbol-Mekkes, P. H., Van der Valk, A. J. J., & Korthals Altes, W. K. (2012). The Netherlands spatial planning doctrine in disarray in the 21st century. *Environment and Planning A, 44*(2), 377–395.

Seebus, J. (2012). Conversion gains ground – Slowly. *Property EU (European Union), 4*, 38–45.

Significance and Bureau Louter. (2007). *Toepassen TIGRIS XL binnen de studie 'Nederland Later'*. Bilthoven: PBL Netherlands Environmental Assessment Agency.

Statistics Netherlands. (2012). *Statline, the central database of Statistics Netherlands*. http://www.statline.nl. Last accessed 1 Aug 2012.

Te Linde, A. H., Bubeck, P., Dekkers, J. E. C., De Moel, H., & Aerts, J. C. J. H. (2011). Future flood risk estimates along the river Rhine. *Natural Hazards and Earth System Sciences, 11*(2), 459–473.

Van der Hoeven, E. M. M. M., Aerts, J., van der Klis, H., & Koomen, E. (2009). An integrated discussion support system for New Dutch flood risk management strategies. Chapter 8. In S. Geertman & J. C. H. Stillwell (Eds.), *Planning support systems: Best practices and new methods. GeoJournal library* (pp. 159–174). Berlin: Springer.

Van Hoek, T., Koning, M. A., & Mulder, M. (2011). *Succesvol binnenstedelijk bouwen. Een onderzoek naar maatschappelijke kosten en baten van mogelijkheden tot optimalisatie van binnenstedelijk bouwen*. Amsterdam: EIB Economisch Instituut voor de Bouw.

Van Rij, E., Dekkers, J., & Koomen, E. (2008). Analysing the success of open space preservation in the Netherlands: The Midden-Delfland case. *Tijdschrift voor Economische en Sociale Geografie, 99*(1), 115–124.

VROM, LNV, V&W and EZ. (2004). *Nota Ruimte. Ruimte voor ontwikkeling. Ministeries van Volkshuisvesting, Ruimtelijke Ordening en Milieubeheer, Landbouw, Natuur en Voedselkwaliteit, Verkeer en Waterstaat en Economische zaken*. Den Haag: SDU uitgeverij.

Warbroek, B. (2011*). Binnenstedelijk bouwen kan veel goedkoper, Binnenlands Bestuur*. http://www.binnenlandsbestuur.nl/ruimte-en-milieu/nieuws/binnenstedelijk-bouwen-kan-veel-goedkoper.648361.lynkx. Last accessed 27 Nov 2012.

Zondag, B., & De Jong, G. (2005). *The development of the TIGRIS XL model: A bottom-up to transport, land-use and the economy*. Paper presented at the Economic Impacts of Changing Accessibility conference at Napier University (27–28 October). http://www.tri.napier.ac.uk/Events/Freight/Abstracts/zondagdeJong.pdf.

Zondag, B., & Geurs, K. (2011). Coupling a detailed land-use model and a land-use and transport interaction model. Chapter 5. In E. Koomen & J. Borsboom-van Beurden (Eds.), *Land-use modeling in planning practice* (pp. 79–95). Dordrecht: Springer.

A Decision Support System for Farmland Preservation: Integration of Past and Present Land Use

Eduardo Corbelle-Rico, Inés Santé-Riveira, and Rafael Crecente-Maseda

Abstract Preservation of agricultural land is a common need in many States and regions in the European Union and worldwide. Underlying reasons for this need are diverse, as are the approaches chosen to facilitate farmland preservation. Zoning is very often the main instrument used to protect agricultural land from urbanization and afforestation. This chapter presents a multi-criteria decision analysis system intended to support zoning decisions. In this study, parcels are used as the unit of analysis. The system of analysis integrates biophysical criteria related to the productivity of land, structural factors related to the efficiency of farming activities and landscape configuration as defined by present and past land use. As an intermediate step to define landscape configuration, a methodology for classification of historical aerial photographs is proposed, based upon object-oriented classification of individual land parcels with the aid of supervised decision trees and ancillary textural information. The resulting decision support system takes the form of a parcel rating system, allowing end users to identify agricultural areas for which protective zoning should be implemented by selecting progressively lower scoring parcels until the desired total area (ideally based upon demand estimates) is satisfied.

Keywords Decision support system • Land abandonment • Geographic information systems • Automatic classification • Farmland preservation

Introduction

The development of strategies to protect agricultural land has been a common issue in planning literature for the past three or four decades (Singer et al. 1979; Nellis and Maca 1986; Daniels and Reed 1988; Bunce 1998). In the North American context, alleged urban sprawl was perceived as the most relevant threat for the preservation of agricultural land; both scientific and governmental reports initially

E. Corbelle-Rico (✉) • I. Santé-Riveira • R. Crecente-Maseda
Land laboratory, Deptartment of Agricultural and Forest Engineering, Universidade de Santiago de Compostela, Santiago de Compostela, Galicia, Spain
e-mail: eduardo.corbelle@usc.es

J.-C. Thill (ed.), *Spatial Analysis and Location Modeling in Urban and Regional Systems*, Advances in Geographic Information Science,
https://doi.org/10.1007/978-3-642-37896-6_8

demonstrated this concern with expressive – if not necessarily accurate – titles such as "Foodland: Preservation or Starvation" (Ontario Institute of Agrologists 1975), "The Vanishing Land" (MacGregor 1980) or *Where Have all the Farmlands Gone* (National Agricultural Lands Study 1980; Bunce 1998). As a result of this concern, a large variety of strategies for farmland preservation are currently implemented in the United States, from the Purchase or Transfer of Development Rights (PDR, TDR) to the creation of agricultural districts or Cluster Development (Brabec and Smith 2002; Tulloch et al. 2003; Bengston et al. 2004).

Industrial-era attempts to protect agricultural land were specifically motivated by the preservation of national food production capacity. This perspective is still prevalent in the developing world, as in the case of China (Skinner et al. 2001; Lichtenberg and Ding 2008; Tan et al. 2009). On the contrary, most industrialized countries are currently more concerned about the overall ecological and social consequences of urban sprawl than about its implications for food security in particular. As a result, preservation of agricultural land is often addressed in the more general context of preservation of open space, which may include not only agricultural areas but also semi-natural areas or even forests. The apparently dated concept of farmland preservation has often been redefined as "countryside preservation" in, for example, the United Kingdom, the Netherlands and Germany (Koomen et al. 2008).

In the European context, urbanization and development of transportation infrastructure have comprised the most substantial changes in land cover currently affecting agricultural land. In Europe, approximately 13,000 km^2 of former agricultural land changed to other land uses between 1990 and 2000. Nearly 8000 km^2 of this transitioned to urban, industrial and infrastructural uses, particularly concentrated in Belgium, the Netherlands, northern Germany and specific locations along the coastal areas of the European Union (EEA 2006). Nevertheless, although less influential than urbanization in terms of total area transitioned, abandonment of agricultural activities (i.e., transitions from agricultural land use to forest or semi-natural areas) also involves important social, environmental, and economic implications (MacDonald et al. 2000). This specific cause of land use transition has recently received attention from the European Commission. Marginalization of agriculture, subsequent abandonment of agricultural land use has been on the European agenda since the first reforms of the Common Agricultural Policy (CAP) in the 1990s, when peripheral or mountainous regions were expected to be those most affected (Baldock et al. 1996; Pinto Correia and Breman 2008). Even though afforestation may be associated to some positive effects (e.g., carbon sequestration), the overall balance of this particular kind of land use transition is strongly dependent on the type of pre-existing landscape. The consequences of the abandonment of agricultural land are considered to be negative not only from the environmental perspective (Suárez Seoane et al. 2002; DLG 2005), but also from the cultural (Höchtl et al. 2005), economic (Gellrich et al. 2006) and social (Soliva 2006) points of view – especially in cultural landscapes that resulted from "the combined works of nature and of man" (UNESCO 2008) or "High Nature Value farming" (EEA 2004). In Mediterranean regions, a particular negative effect

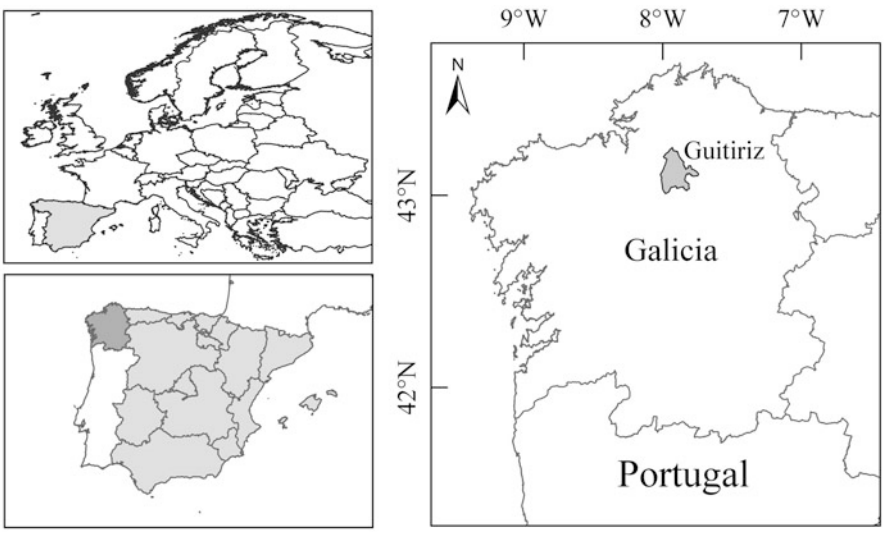

Fig. 1 Location of Galicia (right, with Guitiriz highlighted) within Spain and Europe. (Spatial data source: Global Administrative Areas, http://www.gadm.org)

of abandonment of agriculture is an increase in volume and continuity of biomass resulting from the encroachment of spontaneous vegetation, or from afforestation of former agricultural land, which creates ideal conditions for large wildfires (Moreira et al. 2001; Romero Calcerrada and Perry 2004; Millington 2007).

As one of such regions affected by wildfires, Galicia – an autonomous region located in the northwest of Spain (Fig. 1) – has recently passed several laws requiring the zoning of protected agricultural land. The delineation of protected agricultural land designates areas in which land use changes, including afforestation, may require special permits or simply would not be allowed. Examples of such legislation include the passing of Regional Law 3/2007, regarding wildfires, and Regional Law 7/2007 regarding the regional Land Bank.

This chapter describes a multi-criteria decision analysis (MCDA) system developed to support the process of land use planning, particularly the zoning of protected agricultural land. The use of MCDA systems is common practice in the support of planning activities (Malczewski 1999; Geneletti 2007; Geneletti and van Duren 2008). The purpose of the system defined here is to help planners in the selection of parcels that should be part of the protected agricultural land at the municipal scale, according to social, environmental and economic criteria. More specifically, the objectives include the development of a MCDA system that: (a) takes into account both current and past land use; (b) is adaptable to social reality (variable demand for agricultural land); and (c) can be used for multiple purposes (not only for the support of land use planning, but also for land banking, open space preservation, or land consolidation projects).

Decision Support Systems for Planning Purposes

A decision support system (DSS) may be defined as "an interactive system that helps decision-makers in the use of data and models to solve unstructured problems" (Gorry and Morton 1971). As Witlox (2005) points out, spatial location problems are typically unstructured or semi-structured problems, resulting from difficulties to "specify the relevant criteria (variables) at the outset of the problem, specify the weights to be assigned to criteria, establish site-specific constraints before knowing the resulting consequences, and capture all relevant data with sufficient accuracy" (p. 440). DSSs developed to solve location problems are usually referred to as Spatial Decision Support Systems (SDSSs), and they should consist of at least of three main components (Makowski and Wierzbicki 2000): a database, a model or set of models (very often a MCDA system), and an interface with which end users interact (usually a geographic information system).

In the last years, there has been a very interesting debate on the usefulness of SDSSs. For example, McCown (2002) argues that "the pure DSS idea is elegant – easy-to-use software on a computer readily accessible to a manager to provide interactive assistance in the manager's decision process... But review of 30 years of DSS R&D shows that the reality has been more chaotic than elegant" (p. 19). Among the suggested reasons for the lack of real use, many authors point out that much attention has been usually paid to the technical aspects (algorithms, decision rules, models, etc.) and comparatively less attention has been given to the integration of the system in the decision-making process (Matthews et al. 2008; McCown 2002; Uran and Janssen 2003).

Study Area

To test the proposed MCDA system we selected a Spanish rural municipality (Guitiriz, Fig. 1). This municipality was selected because its local master plan was being developed at the time, and the authors of this work could be in close contact with the planning team. Guitiriz has a total area of 294 km^2 and a hilly topography with altitudes ranging from 400 to 800 m above sea level. As it is the case of all rural municipalities in the region, population has been diminishing since the 1950s (11,500 inhabitants in 1950 and 5727 in 2011; INE 2010b). Population density in Guitiriz (19 inhabitants/km^2) is much lower than the Spanish and Galician averages (87 and 94 inhabitants/km^2, respectively). In addition, the population of Guitiriz has a large proportion of elderly people – 31% of total population is 65 or older (INE 2010b). High population dispersion, with more than 310 different human settlements within the municipality, is another important aspect which demonstrates the complexity of the planning process.

Agriculture has traditionally been the main source of employment, but currently involves about one-third of the working population (INE 2010b). Most of these

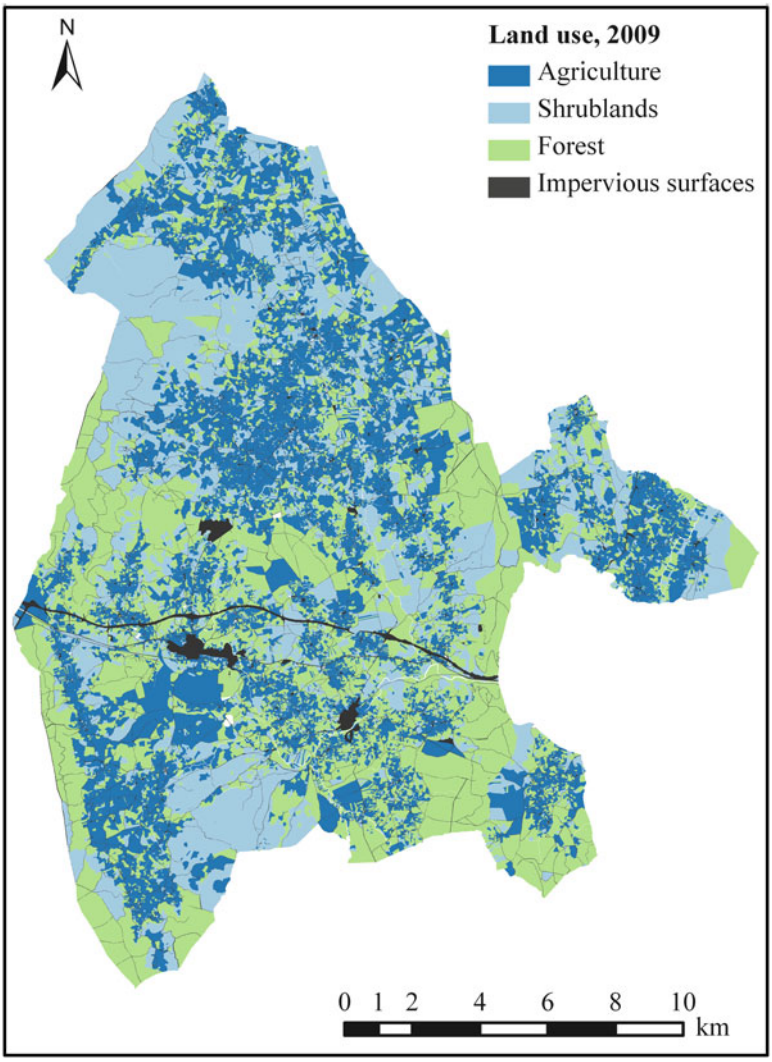

Fig. 2 Land use in the study area (municipality) of Guitiriz in 2009 (Spatial data source: Spanish Land Parcel Information System, http://sigpac.mapa.es/fega/visor)

jobs are generated by small dairy farms that almost exclusively use the labour of relatives; these family farms very often cease their activities as older farmers reach the age of retirement. There were 2000 farms in Guitiriz in 1962, but only 1259 in 1999 (INE 1963, 2010a). In recent years many parcels formerly used by farms have been abandoned or afforested, thus preventing them from being transferred (by lease or sale) to farms still active in the area. Figure 2 shows current (2009) distribution of land uses within the municipality.

Materials

The most important data source for this research is the Spanish Land Parcel Information System (SIGPAC), created in 2004 to comply with European Union (EU) regulations (Council Regulation 1593/2000). SIGPAC is based upon the cadastral map at 1:5000 scale, it includes information about the land use, structure and characteristics of the 71,208 agricultural and forested parcels in the municipality. For this study, we used a version of SIGPAC updated for 2009 (Fig. 2).

To gather information about the historical use of land (Fig. 3), we employed aerial photographs taken in 1957. The end of the 1950s is commonly accepted as a turning point in the abandonment of the traditional agricultural system in Galicia (Bouhier 1979) and the changes in the demographic structure in Spain (Collantes Gutiérrez and Pinilla Navarro 2011).

Information about the biophysical quality of land was taken from a map published by Díaz-Fierros Viqueira and Gil Sotres (1984), as it is the only land suitability map available covering the study area that follows the methodology proposed by the Food and Agriculture Organization of the United Nations (FAO 1976). However, this map was prepared at the regional scale (1:200,000), and for this reason it was deemed reasonable to supplement it at the parcel level with detailed data about slope derived from topographic maps at 1:5000 scale.

Methodology

The form of the MCDA proposed here is essentially that of system of rating land parcels, loosely inspired by the Land Evaluation and Site Assessment system (LESA). LESA was introduced by the USDA Soil Conservation Service in 1981 (Steiner et al. 1994) and has been used for the preservation of agricultural land by a number of state agencies in the United States and Canada (Androkovich 2013). For this study, each of the 71,208 rural parcels in the municipality was rated according to its estimated suitability for agricultural use, based upon biophysical, structural, and land use (past and present) criteria. It is worth noting that the proposed system is not intended to provide a fixed demarcation of protected agricultural land; instead, zoning of the protected area should be preceded by an external estimate of the total area demanded by local farms for agricultural activities, which is the input that enables the MCDA to proceed by adding the most suitable parcels until the desired total area is reached. Such an approach follows the recommendations of the FAO (1976) for suitability assessment and land use plans, by taking into account not only the biophysical quality of land (soil, climate, topography), but also social variables related to the structure of property and the demand for agricultural land (which is determined externally).

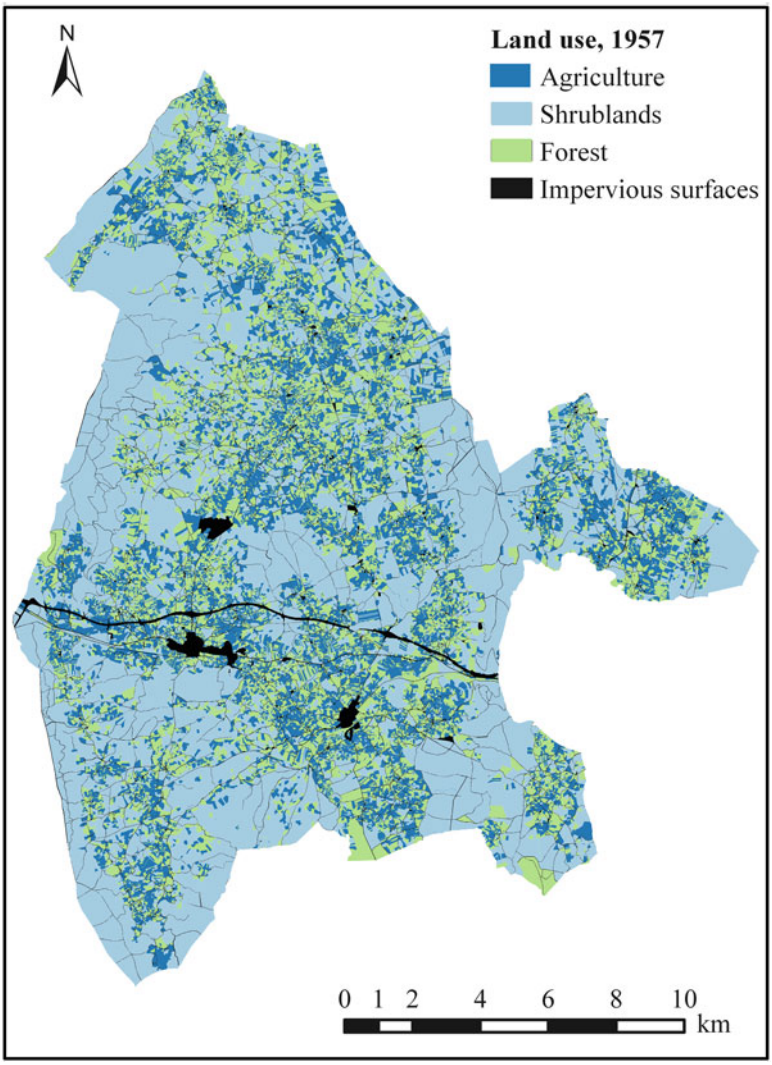

Fig. 3 Automatic classification of past (1957) land use in the municipality of Guitiriz. (Spatial data source: Spanish Land Parcel Information System, http://sigpac.mapa.es/fega/visor)

The proposed system can be considered innovative because it establishes a departure from the usual raster-based approach in decision support systems (Bojórquez-Tapia et al. 2001; Ceballos-Silva and López-Blanco 2003), and integrates the past land use of the area in the current planning activities. The development of this SDSS is a unique yet simple attempt to address the problem of scarce implementation that often plagues MCDA systems (Uran and Janssen 2003; Matthews et al. 2008). The system is also an attempt to recover part of the landscape structure

Table 1 Variables selected and alternative weight sets

Variable group	Weight sets			Variable and variable weight within group	
	1	2	3		
Biophysical factors	0.50	0.34	0.25	Soil productivity	(0.50)
				Average slope of parcel	(0.50)
Structural factors	0.25	0.33	0.25	Parcel area	(0.34)
				Parcel shape	(0.33)
				Accessibility from roads	(0.33)
Land use factors	0.25	0.33	0.50	Present land use	(0.50)
				Past land use	(0.50)

existing in the 1950s, characterized by the existence of blocks of parcels used for agriculture – called "agras" by local inhabitants. Forest plantations made since then often disrupted these landscape units (Calvo et al. 2011), a trend that the system proposed in this study tries to reverse – at least in part.

The structure of the proposed MCDA system takes the form of a simple weighted linear combination (WLC), one of the most common approaches in the implementation of MCDA in geographic information systems (Malczewski 2006). Three main groups of variables have been selected to estimate the degree to which each land parcel fulfils the requirements to become protected agricultural land: biophysical quality of land, structural characteristics of parcels, and current and past land use. The intention behind the selection of these three criteria groups is to protect the best quality soils, minimize erosive processes and facilitate agricultural production. For each of the criteria groups, several variables were selected (Table 1) and their values calculated for each individual parcel. Values of the variables were then transformed into relative scores ranging from 0 (minimum desirability) to 100 (maximum desirability), according to the contribution of the original values to the achievement of the objectives of the MCDA system. Different weights were assigned to each variable in an attempt to express their relative importance for the objectives of the zoning process (Table 1).

The first group of relative weights for the variables (Weight Set 1) stresses the importance of protecting the best soils for agricultural use (50% of the overall weight in the proposed scheme) and leaves the consideration of current land use and parcel characteristics in a secondary place (25% each in the proposed scheme). The final score (V) assigned to each parcel results from the weighted summation of the rated criterion layers, according to Eq. 1,

$$V = \sum_{i=1}^{n} w_i a_i \tag{1}$$

where w is the weight assigned to each variable (the product of the weight assigned to the group of variables and the weight of each variable inside its group) and a is the value of the transformed (rated) variable.

The weighting scheme was developed in consultation with the team of technicians (architects and rural engineers) in charge of the municipal plan, who deemed it acceptable. It is clear, though, that this procedure does not account for the validity of the weights proposed; although the team generally agreed on the scheme, subsequent discussion showed that a considerably large range of importance of each variable was possible, and that more than slightly different outcomes could have been equally held as acceptable. In order to present the range of possible contributions of each criteria group, a sensitivity analysis was performed using two alternative weight sets. The alternative weights represent an approach giving equal importance to the three sets of variables (Weight Set 2) and another one giving more importance to variables of present and past land use (Weight Set 3). Keeping in mind that the objective of the MCDA system is the ranking of parcel suitability for agricultural use, we looked for changes in the rank of each parcel, as was the focus of Geneletti and van Duren (2008). Small changes in the relative ranking of a particular parcel would mean that the inclusion of that parcel in the preserved area is not strongly dependent upon the weight set used (importance ranking of variables) and therefore is expected to be more readily accepted by planners and local residents.

Rating of Biophysical Variables

To prevent the most productive soils from being occupied by other land uses, data from a soil productivity map was included in the pool of variables. The map offers estimation for five categories of suitability and two agricultural uses (arable land and pastures) according to the system proposed by FAO (1976). In this system, the categorization scheme assigns S1 for the most suitable soils, S2 and S3 for those less suitable, and N1 and N2 for unsuitable soils. The values of the variable for each type of parcel were rated from 100 to 0 according to the scheme of Table 2.

The map takes into account many biophysical characteristics of soils (texture, minerals, water reserve, stoniness and slope). Although a bit dated at the time of writing this study and with a rather small scale (1:200,000), it still is the most comprehensive source of information about soil productivity for the most areas in Galicia. To complement that information, the average slope of each parcel was separately included, as derived from much more detailed topographic maps (1:5000). The link between slope and suitability for agricultural use depends upon the potential for facilitation of mechanization on a particular parcel. Local practice considers any form of mechanised farming to be inadvisable above a slope of 25% – although such parcels are still commonly used as pastures. In an area of similar characteristics in the Pyrenees, arable land was found to have a mean slope of approximately 5%, whereas pastures were located in areas with a mean slope of 20% (Mottet et al. 2006). Following this reference, parcel slope was rated at 100 when lower than 5% and 0 when higher than 25%, decreasing in a linear relationship between both values.

Table 2 Rating scheme for the variable "soil productivity"

Suitability class		Assigned rating
As arable land	As pastureland	
S1	(Any)	100
S2	S1	100
	S2	90
	S3	80
S3	S1	70
	S2	60
	S3	50
N1	S1	40
	S2	30
	S3	20
N2	S1	10
	(Others)	0

(Source: Suitability classes as defined in Díaz-Fierros Viqueira and Gil Sotres 1984)

Rating of Structural Variables

Parcel characteristics such as size, shape, and possibility of direct access from roads were calculated from the cadastral map, using a geographic information system. Shape was measured through an area/perimeter ratio as expressed by Eq. 2.

$$Shape.index = 100\frac{4\sqrt{Area\,(m^2)}}{Perimeter(m)} \tag{2}$$

The shape index has a value of 100 in the case of a perfectly square shape (though the value could be larger than 100 if the parcel approximated an oval or circular shape); the value is less than 100 in the case of elongated or complex shapes. Because values resulting from Eq. 2 are comprised between 0 and 100, they can be used directly as part of the rating system.

Parcel size was rated using the average size of parcels leased in the Land Bank of Galicia (1 ha) as a reference. The Land Bank is a public service that acts as mediator in many land rental agreements. We interpreted the average size of parcels leased under this system as a conservative estimation of the optimum size of parcels that farmers would be willing to cultivate. Accordingly, parcels larger than 1 ha were rated at the maximum (100), and smaller parcels were rated linearly to zero. The third variable in this group, direct access from public roads, was represented as a logical variable (true/false), and parcels were rated 100 when direct access existed and zero if not.

Rating of Land Use Variables

Inclusion of present land use in the rating system was based upon two assumptions: that it would be preferable to designate as protected agricultural areas those parcels that are already being used for agricultural purposes, and that agricultural parcels currently surrounded by forested parcels should receive less priority in the rating system (shades projected by near forest stands have a negative impact on agricultural parcels, like for example on the amount and quality of pastures). This meant considering not only the present land use of the parcel itself but also the land use of neighbouring parcels. In order to combine both, forest parcels were assigned the minimum rating (0), while parcels with other land uses were assigned a rating equal proportionate to the share of the perimeter *not* shared with forest parcels.

On the other hand, the purpose of including past use is to maintain or recover, to some extent, past land use. Rating of past land use was established at 100 for parcels used as arable land or pastures in 1956–57 for shrublands (areas covered by woody perennial plants, smaller than trees), and zero for forest parcels, based upon automatic classification of the aerial photographs from that year. This is based on evidence that farmers often transform former shrublands into farmed area (Corbelle-Rico et al. 2012).

Automatic Classification of Historical Photographs

For the automatic classification of the aerial photographs from 1957 (previously orthorectified and mosaicked), an object-oriented automatic approach was employed. Automatic classification of black and white (panchromatic) images is usually a difficult task given the low amount of spectral information they possess; object-oriented methods usually attain better results in terms of classification accuracy and consistency (Laliberte et al. 2004; Marignani et al. 2008). Such an approach fit well with this work, as land parcels can easily be used as "objects" to be classified. Classification of panchromatic images typically benefits from the use of textural measures (Cots Folch et al. 2007), but even with the use of this complementary information, the number of classes in the final classification is generally low – usually only two (Hutchinson et al. 2000; Laliberte et al. 2004; Pillai et al. 2005) or three classes, at most (Carmel and Kadmon 1998; Kadmon and Harari-Kremer 1999). Despite this result, global accuracy values are usually low, ranging between 60% (Carmel and Kadmon 1998) and 80% (Kadmon and Harari-Kremer 1999), though sometimes reaching 90% (Pillai et al. 2005).

The goal of the classification process was the identification of four land cover classes: arable land, pastures, shrubland, and forest. These classes were easily distinguishable separable by their mean grey levels, with arable having the highest values (very light grey in the original photographs) and forest the lowest (very dark

grey, almost black), while pastures had grey values in the middle. Only shrublands were not distinguishable, due to their heterogeneous nature – however, such a pattern can be easily detected via textural image processing (e.g., the use of local variance, which typically yields higher values for heterogeneous areas than for the rest of the classes). The use of textural features to improve image classification – proposed by Haralick et al. (1973) – has been demonstrated by previous studies (e.g., Cots Folch et al. 2007; Johansen et al. 2007).

A supervised decision tree was fitted to a sample of 100 parcels selected by quota sampling in order to obtain 25 of each class, which are then visually classified. The methodology makes use of the J48 algorithm available in the R package RWeka (R Development Core Team 2008; Hornik et al. 2009). Global accuracy of the resulting classification was assessed by means of a ten-fold cross-validation process[1] which estimated to result in approximately 88% global accuracy.

Independence of Variables

In any set of attributes or variables that are considered to be the input for MCDA, it is likely that variables could be correlated. If this was the case, some of the variables would be redundant, and the results of the analysis would be affected (Malczewski 2000). To test the independence of the seven variables included in this work, correlation was assessed using Spearman's rank correlation coefficient (a non-parametric measure of dependence between two variables). The seven variables were not demonstrated to be redundant (Table 3); correlation coefficients were between −0.25 and 0.25 in every case.

Table 3 Correlation between rated values of the seven variables used in the study

Variables	1	2	3	4	5	6	7
1. Soil productivity	1.00	0.20	−0.14	−0.03	0.00	0.10	0.04
2. Parcel slope		1.00	−0.04	−0.08	0.01	0.24	0.09
3. Parcel area			1.00	0.24	0.20	−0.12	−0.14
4. Parcel shape				1.00	0.04	0.00	−0.07
5. Parcel access					1.00	0.04	−0.03
6. Present land use						1.00	0.15
7. Past land use							1.00

Spearman's rank correlation coefficient

[1]Sample data is randomly divided into ten subsets. One subset is left out of model fitting process, and is used to assess classification accuracy; the process is then repeated nine more times until every sample observation has been used for accuracy assessment.

Allocation of Preserved Farmland

The proposal for the selection of parcels to form the protected agricultural space depends on estimation of agricultural land demand by active farms in the municipality. Current practice for planners in Galicia usually assumes demand to be equal to, or slightly greater than, current area of agricultural land, although other slightly more complex methods are occasionally used. When most farms in the municipality are specialized in dairy farming, a rough estimation of demand is sometimes derived from total existing headage (livestock units) and maximum stocking rate (livestock units/ha) allowed by European Common Agricultural Policy subsidies.

Results

Once the amount of desired protected land is established, the MCDA system accumulates the highest-scoring parcels until the desired total area is reached. Results of the MCDA system are summarized in Fig. 4. Figure 4a presents the relationship between parcel scores obtained using Weight Set 1 (x) and the area accumulated by parcels having received equal or higher scores (y). This graphic is intended to allow decision-makers (planners) to determine which parcels should be included in the protected area; for a desired amount of preserved farmland (y) (based on estimated demand of land by local farms) the curve indicates the minimum suitability score above which parcels should be selected. As the area intended for protection increases, parcels with lower scores must be selected to accumulate the total area. Obviously, the use of different weight sets would result in slightly different curves, but this influence is rather limited. Correlation is high between parcel scores resulting from the use of Weight Sets 1 and 2 (Fig. 4b), 2 and 3 (Fig. 4c), 1 and 3 (Fig. 4c), meaning that selection of weights has relatively little influence on the relative ranking of parcels. In other words, parcels with most suitable characteristics would appear among the highest-scoring parcels for all weight sets, therefore the final set of parcels included in the protected area would not change very much. Table 4 presents the Spearman's rank correlation coefficient for the three sets of scores: as shown in Figs. correlation is especially high in the case of the highest and lowest scores (Fig. 4b, c and d), as these correspond to parcels generally characterized by high/low ratings for all of the variables.

An example of application of the system (based upon the curve in Fig. 4a) is presented in Fig. 5. Three proposals for total protected agricultural area are presented – for hypothetical target areas of 86 km^2 (the total of current area occupied by agriculture in the municipality), 68 km^2 (−20%) and 105 km^2 (+20%). It can be seen in the figure how the protected area is formed by parcels with the highest suitability values (in orange), and how additional areas could be added (in blue and green) to increase the total amount of protected area. In comparing with Fig. 2, it is apparent that the selected parcels generally follow current land use, meaning that zoning would not impose severe changes in land use.

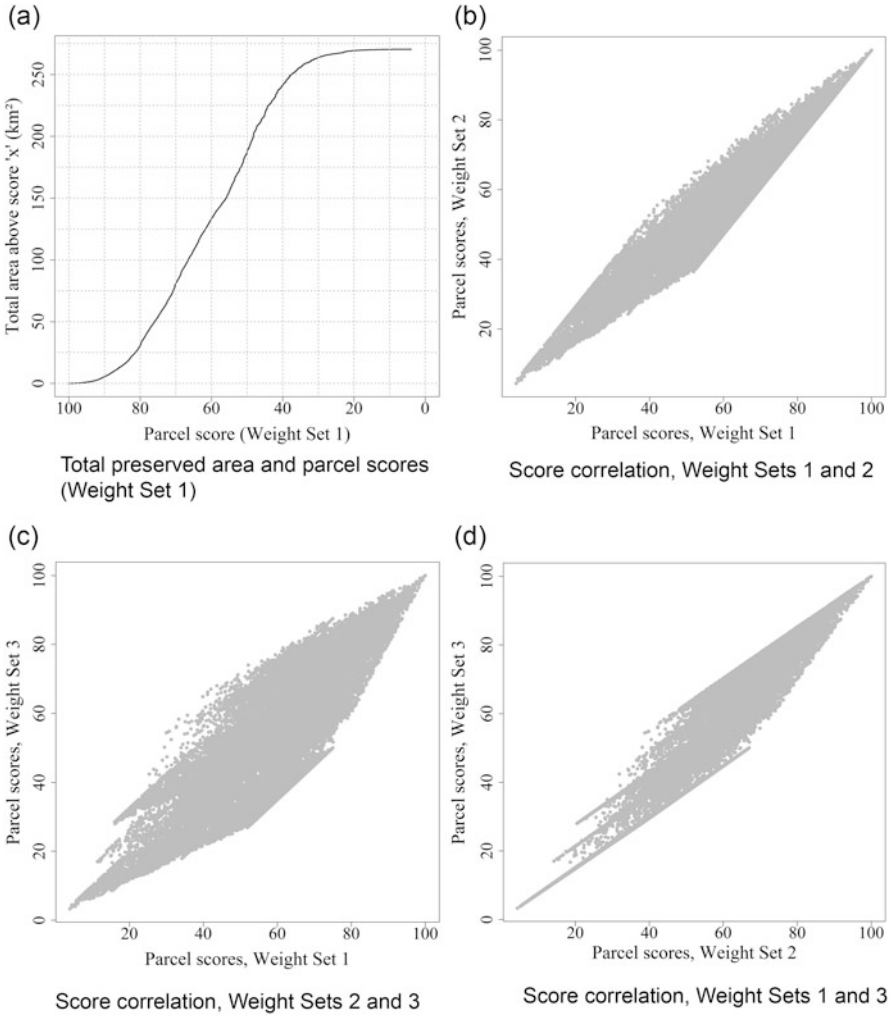

Fig. 4 Area of accumulated parcels above a given score (**a**) and correlation between parcel suitability scores resulting from different weight sets (**b**, **c**, **d**)

Table 4 Correlation between parcel scores resulting from different weight sets

	Weight set 2	Weight set 3
Weight set 1	0.96	0.88
Weight set 2	–	0.96

Spearman's rank correlation coefficient

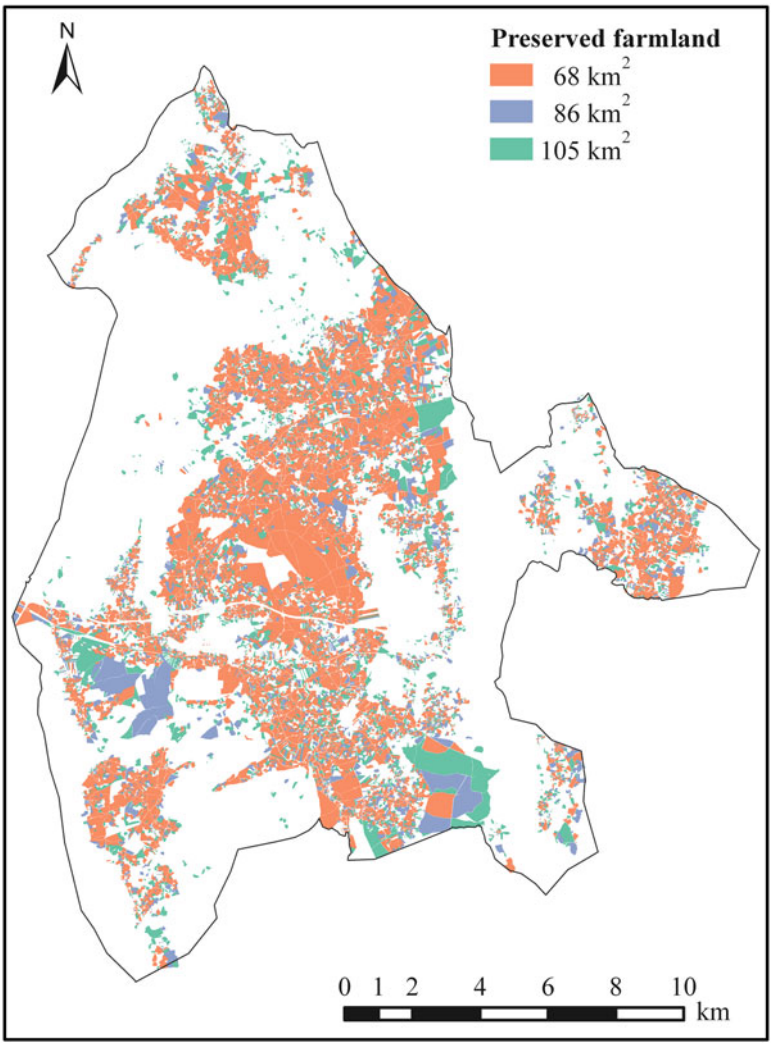

Fig. 5 Selected parcels for different total areas of protected agricultural land. (Spatial data source: Spanish Land Parcel Information System, http://sigpac.mapa.es/fega/visor)

Discussion

The results of the sensitivity analysis show that a high degree of correlation exists between scores calculated for each parcel using each of the three different weight sets. In practical terms, this means that when the desired amount of protected area is either proportionally small or large with respect to the total area of the municipality, maps resulting from the different weight sets would be quite similar – whereas when

the desired total protected area is close to one half of the total municipal area, the differences between the solutions from the various weight sets are likely to be at their maximum. In this particular case, desired total protected area is about one third of the total municipal area (around 86 km^2), which means inclusion of parcels over a score threshold around 70 (see Fig. 4a). Around this point, scatterplots show strong correlation between various weight set solutions, even if we consider that weight sets 1 and 3, for example, do not represent slightly different views – but rather, opposite approaches. In conclusion, the selection of the most suitable parcels for agricultural use may be considered relatively stable, regardless of the set of weights used in the MCDA.

An approach based upon the rating of individual parcels inherently enables the inclusion of characteristics of the parcel structure in the decision system. It may be argued that this would also be possible using a more conventional raster-based approach, but a parcel-based approach was also demonstrated to fit well with the automatic classification of past land use. In addition, the parcel-based approach also fits better the actual practice of planners, who tend to avoid, whenever possible, dividing parcels into several zoning categories. Regarding the classification of past land use, the success of the proposed methodology is clearly dependent upon the degree of change that has since affected the parcel, as well as on the availability of historic remotely-sensed data or data about past land uses.

In terms of implementation in real cases, our experience showed that the large number of zoning categories that exist under the law makes it impossible to simply transfer the results of the MCDA system to the final plan. At least six other categories of protected rural land may be included in the plan, for example "right-of-way preservation" for public utilities or infrastructures, "landscape preservation", or "preservation of waterways". Nevertheless, the results of the system were used as a guideline; after adjustments were made to include or exclude individual parcels – sometimes only fractions of parcels – in order to account for other categories of protected rural land, the results served to designate the protected agricultural area. While many of these additional categories of protected rural land could also be included in the proposed SDSS, interaction with planners revealed their preference for simple systems from which they could derive results to be used externally, in combination with other information in a desktop GIS. In this fashion, the SDSS proposed here served as only a component of the larger, personalized SDSS that planners employed.

Conclusions

Preservation of agricultural land is a current topic on the agenda of many States and regions of the European Union, for a variety of reasons. This work proposed the development of an objective system to define the boundaries of protected agricultural land adaptable to the estimated demand for agricultural land at the municipal or parish level. The system integrates information variously related to

the productivity and ecological importance of agricultural land, while requiring minimal input data sources. Nevertheless, it is far from our intention to present it as a model to be followed strictly. Other variables, different weighting applied to them, or different means of analysis could be successfully employed in other geographic areas.

It is remarkable that the influence of subjective decisions (the estimation of weights for the different groups of variables) upon the final result is relatively low, which suggests that the approach is robust enough for the system to be trustworthy both for planning technicians and local inhabitants. Two other innovative features are should be highlighted as well. First, the parcel-based nature of the process, as opposed to a raster-based approach is a rare feature that contributes to the applicability of the system in a practical context. Second, the combination of object-oriented classification with the use of textural features for the classification of historical black and white aerial photographs is also innovative, rendering very high accuracy values while offering easy applicability via the use of commonly-used, readily available software.

Acknowledgments We gratefully acknowledge the planning team formed by GAU and LaboraTe-USC technicians for their opinions and ideas which contributed to this work. We also thank the anonymous reviewer, who provided many constructive and helpful comments that helped to improve our manuscript.

References

Androkovich, R. A. (2013). British Columbia's agricultural land reserve: Economic, legal and political issues. *Land Use Policy, 30*(1), 365–372.

Baldock, D., Beaufoy, G., Brouwer, F., & Godeschalk, F. (1996). *Farming at the margins: Abandonment or redeployment of agricultural land in Europe*. London/The Hague: Institute for European Environmental Policy.

Bengston, D. N., Fletcher, J. O., & Nelson, K. C. (2004). Public policies for managing urban growth and protecting open space: Policy instruments and lessons learned in the United States. *Landscape and Urban Planning, 69*, 271–286.

Bojórquez-Tapia, L. A., Díaz-Mondragón, S., & Ezcurra, E. (2001). Gis-based approach for participatory decision making and land suitability assessment. *International Journal of Geographical Information Science, 15*(2), 129–151.

Bouhier, A. (1979). La Galice: essai geographique d'analyse et d'interpretation d'un vieux complexe agraire, Ph.D. Thesis, University of Poitiers.

Brabec, E., & Smith, C. (2002). Agricultural land fragmentation: The spatial effects of three land protection strategies in the eastern United States. *Landscape and Urban Planning, 58*(2), 255–268.

Bunce, M. (1998). Thirty years of farmland preservation in North America: Discourses and ideologies of a movement. *Journal of Rural Studies, 14*(2), 233–247.

Calvo, S., Méndez, G., & Díaz, R. A. (2011). Los paisajes culturales de agras en Galicia y su dinámica evolutiva. *Journal of Depopulation and Rural Studies, 10*, 7–38.

Carmel, Y., & Kadmon, R. (1998). Computerized classification of Mediterranean vegetation using panchromatic aerial photographs. *Journal of Vegetation Science, 9*, 445–454.

Ceballos-Silva, A., & López-Blanco, J. (2003). Delineation of suitable areas for crops using a multi-criteria evaluation approach and land use/cover mapping: A case study in central Mexico. *Agricultural Systems, 77*, 117–136.

Collantes Gutiérrez, F., & Pinilla Navarro, V. (2011). *Peaceful surrender: The depopulation of rural Spain in the twentieth century*. Newcastle: Cambridge Scholars Publishers.

Corbelle-Rico, E., Crecente-Maseda, R., & Santé-Riveira, I. (2012). Multi-scale assessment and spatial modelling of agricultural land abandonment in a European peripheral region: Galicia (Spain), 1956–2004. *Land Use Policy, 29*(3), 493–501.

Cots Folch, R., Aitkenhead, M. J., & Martínez Casanovas, J. A. (2007). Mapping land cover from detailed aerial photography data using textural and neural network analysis. *International Journal of Remote Sensing, 28*(7), 1625–1642.

Daniels, T. L., & Reed, D. E. (1988). Agricultural zoning in a metropolitan county: An evaluation of the Black Hawk County, Iowa, program. *Landscape and Urban Planning, 16*(4), 303–310.

Díaz-Fierros Viqueira, F., & Gil Sotres, F. (1984). *Capacidad productiva de los suelos de Galicia*. Madrid: Universidade de Santiago de Compostela.

DLG. (2005). *Land abandonment, biodiversity and the CAP, Dienst Landelijk Gebied*. Utrecht: Government Service for Land and Water Management of the Netherlands.

EEA. (2004). *High nature value farmland. Characteristics, trends and policy challenges*, *European Environment Agency. Report No 39 (1/2004)*. Luxembourg: EUR-OP.

EEA. (2006). *Land accounts for Europe 1990–2000. Towards integrated land and ecosystem accounting, European Environment Agency*. Report No 11/2006, Copenhagen.

FAO. (1976). *A framework for land evaluation, Food and Agriculture Organization of the United Nations*. Soils Bulletin 32.

Gellrich, M., Baur, P., Koch, B., & Zimmermann, N. E. (2006). Agricultural land abandonment and natural forest re-growth in the Swiss mountains: A spatially explicit economic analysis. *Agriculture, Ecosystems and Environment, 118*, 93–108.

Geneletti, D. (2007). An approach based on spatial multicriteria analysis to map the nature conservation value of agricultural land. *Journal of Environmental Management, 83*, 228–235.

Geneletti, D., & van Duren, I. (2008). Protected area zoning for conservation and use: A combination of spatial multicriteria and multiobjective evaluation. *Landscape and Urban Planning, 85*(2), 97–110.

Gorry, G. A., & Morton, M. S. (1971). *A framework for management information systems*. Cambridge: Massachusetts Institute of Technology.

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification'. *IEEE Transactions on Systems, Man and Cybernetics, 3*(6), 610–621.

Höchtl, F., Lehringer, S., & Konold, W. (2005). "Wilderness": What it means when it becomes a reality – a case study from the Southwestern Alps. *Landscape and Urban Planning, 70*, 85–95.

Hornik, K., Zeileis, A., Hothorn, T., & Buchta, C. (2009). *RWeka: An R interface to Weka*. R package version 0.3-15. URL: http://CRAN.R-project.org/

Hutchinson, C. F., Unruh, J. D., & Bahre, C. J. (2000). Land use vs. climate as causes of vegetation change: A study in SE Arizona. *Global Environmental Change, 10*, 47–55.

INE. (1963). *Primer Censo Agrario de España. Año 1962. Resultados provisionales. Segunda parte: datos municipales*. Madrid: Spanish National Statistics Institute (INE). 1964.

INE. (2010a). *Agricultural Census 1999, Spanish National Statistics Institute (INE)*. URL: http://www.ine.es

INE. (2010b). *Population Census, Spanish National Statistics Institute (INE)*. URL: http://www.ine.es

Johansen, K., Coops, N. C., Gergel, S. E., & Stange, Y. (2007). Application of high spatial resolution satellite imagery for riparian and forest ecosystem classification. *Remote Sensing of Environment, 110*, 29–44.

Kadmon, R., & Harari-Kremer, R. (1999). Studying long-term vegetation dynamics using digital processing of historical aerial photographs. *Remote Sensing of Environment, 68*, 164–176.

Koomen, E., Dekkers, J., & van Dijk, T. (2008). Open-space preservation in the Netherlands: Planning, practice and prospects. *Land Use Policy, 25*, 361–377.

Laliberte, A. S., Rango, A., Havstad, K. M., Paris, J. F., Beck, R. F., McNeely, R., & González, A. L. (2004). Object-oriented image analysis for mapping shrub encroachment from 1937 to 2003 in southern New Mexico. *Remote Sensing of Environment, 93*(2), 198–210.

Lichtenberg, E., & Ding, C. (2008). Assessing farmland protection policy in China. *Land Use Policy, 25*(1), 59–68.

MacDonald, D., Crabtree, J. R., Wiesinger, G., Dax, T., Stamou, N., Fleury, P., Gutiérrez Lazpita, J., & Gibon, A. (2000). Agricultural abandonment in mountain areas of Europe: Environmental consequences and policy response. *Journal of Environmental Management, 59*, 47–69.

MacGregor, R. (1980). The vanishing land. *Maclean's Magazine, 12*, 46–47.

Makowski, M., & Wierzbicki, A. P. (2000). Architecture of decision support systems. In A. P. Wierzbicki, M. Makowski, & J. Wessels (Eds.), *Model-based decision support methodology with environmental applications* (pp. 47–70). Dordrecht: Kluwer Academic Publishers.

Malczewski, J. (1999). *GIS and multicriteria decision analysis*. Toronto: John Wiley and Sons.

Malczewski, J. (2000). On the use of weighted linear combination method in GIS: Common and best practice approaches. *Transactions in GIS, 4*(1), 5–22.

Malczewski, J. (2006). Gis-based multicriteria decision analysis: A survey of the literature. *International Journal of Geographical Information Science, 20*(7), 703–726.

Marignani, M., Rocchini, D., Torri, D., Chiarucci, A., & Maccherini, S. (2008). Planning restoration in a cultural landscape in Italy using an object-based approach and historical analysis. *Landscape and Urban Planning, 84*(1), 28–37.

Matthews, K. B., Shwarz, G., Buchan, K., Rivington, M., & Miller, D. (2008). Wither agricultural DSS? *Computers and Electronics in Agriculture, 61*, 149–159.

McCown, R. L. (2002). Changing systems for supporting farmers' decisions: Problems, paradigms, and prospects. *Agricultural Systems, 74*, 179–220.

Millington, J. D. A. (2007). *Modelling land-use/cover change and wildfire regimes in a Mediterranean Landscape*, Ph.D. dissertation, King's College, London.

Moreira, F., Rego, F. C., & Ferreira, P. G. (2001). Temporal (1958–1995) pattern of change in a cultural landscape of Northwestern Portugal: Implications for fire occurrence. *Landscape Ecology, 16*, 557–567.

Mottet, A., Ladet, S., Coqué, N., & Gibon, A. (2006). Agricultural land-use change and its drivers in mountain landscapes: A case study in the Pyrenees. *Agriculture, Ecosystems and Environment, 114*, 296–310.

National Agricultural Lands Study. (1980). *Where Have all the Farmlands Gone*. Washington, DC: National Agricultural Lands Study.

Nellis, L., & Maca, M. N. (1986). The effectiveness of zoning for agricultural lands protection: A case study from Cache County, Utah. *Landscape and Urban Planning, 13*, 45–54.

Ontario Institute of Agrologists. (1975). *Foodland preservation or starvation: A statement on land use policy*. Toronto: Executive Committee of the Council of Ontario Institute of Agrologists.

Pillai, R. B., Weisberg, P. J., & Lingua, E. (2005). *Object-oriented classification of repeat aerial photography for quantifying woodland expansion in central Nevada*', 20th Biennial Workshop on Aerial Photography, Videography, and High Resolution Digital Imagery for Resource Assessment, October 4–6, Weslaco, Texas.

Pinto Correia, T., & Breman, B. (2008). Understanding marginalisation in the periphery of Europe: A multidimensional process. In F. Brouwer, T. van Rheenen, S. S. Dhilliion, & A. M. Elgersma (Eds.), *Sustainable land management. Strategies to cope with the marginalisation of agriculture*. Cheltenham: Edward Elgar Publishing Limited.

R Development Core Team. (2008). *R: A language and environment for statistical computing, R foundation for statistical computing*, Vienna, Austria. ISBN 3-900051-07-0. URL: http://www.R-project.org

Romero Calcerrada, R., & Perry, G. L. W. (2004). The role of land abandonment in landscape dynamics in the SPA Encinares del río Alberche y Cofío, Central Spain, 1984–1999. *Landscape and Urban Planning, 66*, 217–232.

Singer, M. J., Tanji, K. K., & Snyder, J. H. (1979). Planning uses of cultivated cropland and pastureland. In M. T. Beatty, G. W. Petersen, & L. D. Swindale (Eds.), *Planning the uses and management of land*, *Agronomy series, n. 21* (pp. 225–272). Madison: American Society of Agronomy.

Skinner, M. W., Kuhn, R. G., & Joseph, A. E. (2001). Agricultural land protection in China: A case study of local governance in Zhejiang province. *Land Use Policy, 18*(4), 329–340.

Soliva, R. (2006). Landscape stories: Using ideal type narratives as a heuristic device in rural studies. *Journal of Rural Studies, 23*, 62–74.

Steiner, F. R., Pease, J. R., & Coughlin, R. E. (Eds.). (1994). *A decade with LESA: The evolution of land evaluation and site assessment*. Ankeny: Soil and Water Conservation Society.

Suárez Seoane, S., Osborne, P. E., & Baudry, J. (2002). Responses of birds of different biogeographic origins and habitat requirements to agricultural land abandonment in northern Spain. *Biological Conservation, 105*, 333–344.

Tan, R., Beckmann, V., van den Berg, L., & Qu, F. (2009). Governing farmland conversion: Comparing China with the Netherlands and Germany. *Land Use Policy, 26*(4), 961–974.

Tulloch, D. L., Myers, J. R., Hasse, J. E., Parks, P. J., & Lathrop, R. G. (2003). Integrating GIS into farmland preservation policy and decision making. *Landscape and Urban Planning, 63*, 33–48.

UNESCO. (2008). *Operational guidelines for the implementation of the World Heritage Convention, United Nations Educational, Scientific and Cultural Organization*. Intergovernmental Committee for the Protection of the World Cultural and Natural Heritage. URL: http://whc.unesco.org/archive/opguide05-en.pdf

Uran, O., & Janssen, R. (2003). Why are spatial decision support systems not used? Some experiences from the Netherlands. *Computers, Environment and Urban Systems, 27*, 511–526.

Witlox, F. (2005). Expert systems in land-use planning: An overview. *Expert Systems with Applications, 29*, 437–445.

Aggregate and Disaggregate Dynamic Spatial Interaction Approaches to Modeling Coin Diffusion

Marion Le Texier and Geoffrey Caruso

Abstract With the 2002 introduction of the euro as a common currency in Europe, the possibility has emerged to assess international mobility using this new tracer, given that every coin bears a specific national side. Using a simple two-country framework, four dynamic modeling strategies were designed in order to simulate the diffusion of coins and to understand how this diffusion is affected by population size, mobility rates and coin exchange processes. Methodological implications are raised with respect to aggregation, synchronicity and stochasticity issues.

Although each model converges to an equilibrium, the time to reach this end stage and the level of coin mixing in each country strongly varies with the modeling strategy. Calibration is undertaken with French data, using mobility rates as adjustment variables. The experiment shows that convergence to a perfect mix of coins can only be obtained if reciprocal exchanges are modeled, with a time horizon around 2064 – while non-reciprocal models indicate an imperfect mix converging in the year 2020 at the latest.

Keywords Diffusion • Euro • Spatial interaction • Dynamic system • Individual agent-based model

Introduction

> The Euro is much more than a mere economic instrument. It is a symbol of the *grand mission* of European integration, a tool to ensure democracy and peace on a continent too often ravaged by war (Time, 30th April 2012).

M. Le Texier (✉)
University Paris Diderot, Paris, France

University Rouen Normandie, Mont-Saint-Aignan, France
e-mail: marion.le-texier@univ-rouen.fr

G. Caruso
University of Luxembourg, Luxembourg, Luxembourg

J.-C. Thill (ed.), *Spatial Analysis and Location Modeling in Urban and Regional Systems*, Advances in Geographic Information Science,
https://doi.org/10.1007/978-3-642-37896-6_9

This is the way Michael Schuman, an American journalist, reports on what Europeans say about the euro and its recent crisis. It seems that the meaning of the common currency for Europeans goes far beyond its economic utility, notably since it facilitates exchanges between residents of the Eurozone.

The euro was put into circulation on January 1, 2002 as a common currency for 11 countries – each euro coin bearing on one side a common EU symbol, and on the other side, a national symbol. At that precise point in time, all coins minted in a given country were located in their countries of origin. Euro coins then started to circulate, conveyed across space and borders by people and along multiple monetary transactions. Assumably, the euro constitutes a wonderful tracer of international mobility and can somehow reveal patterns of European integration. Does a rapid mixing of the different national coins, and a high level of foreign euro coins in a given country, mean stronger integration? Interestingly enough, after 10 years, French coins still represented 66% of total coins in France, while the *Banque Nationale de France* produces only 20% of the coins of the Eurozone (Grasland et al. 2012). Although European integration cannot be reduced to the pace of coins' mixing, analyzing the mechanisms of the euro diffusion process may shed light on particular interaction flows.

Empirical studies based upon surveys of money bags across the EU have revealed intricate patterns of diffusion that go well beyond the simple proximity-based contagion diffusion from borders. A series of models have been developed in the past few years, based upon physical or epidemic processes, to reflect both upon the spatial diffusion of money and the time at which a homogeneous mix of coins across national states can be reached. Models have used approaches as diverse as Brownian movement, Levy flights, or Markov chains.

In this research, we use a simple spatial interaction perspective to the modeling of euro coins since it is well grounded in geographical research to represent flows of people and goods [see Tobler (1970) or Fotheringham and O'Kelly (1989) to name but two]. This model explicitly takes into account the existence, size and characteristics of cities, in addition to distance effects. However, a specificity of coin diffusion is that flows are not directly observed. Rather, it is the mix of coins in different places that can be measured, i.e., the result of a series of consecutive flows and interpersonal interactions occurring between many possible origins and destinations. The question then arises as to whether this diffusion process can be modeled using aggregate behavioral assumptions at the country level, or whether each individual, wearing a money bag, must be modeled across his/her trips and transactions in order to obtain a realistic representation of the diffusion process. The money bags will be described by the share of coins of each particular country of minting that they contain.

This study first recalls basics of dynamic spatial interaction models and discusses existing models of coin diffusion and their assumptions. These modeling strategies raise fundamental issues related to the treatment of time, randomness and agents' aggregation. We then provide a comparative analysis of four models of coin circulation, which are simulated using a simplified geography of two countries. Our models differ in two aspects: (1) the scale of analysis, i.e., aggregate, (place-based)

versus individual-based approaches; (2) the time granularity, i.e., asynchrony versus synchrony of both mobility and exchange processes; and (3) the definition of the exchange process, i.e., the number of coins involved in each transaction. We analyze the time, existence, and characteristics of a steady state in the coins' mixing process through numerical simulation. We then conduct a sensitivity analysis of the time to convergence and proportion of the foreign coin mix to the spatial interaction parameters: population size and mobility rates (in which distance effects and the attractiveness of places are assumed to be embedded).

Methodology

Randomness, Time and Behavior in Spatial Interactions

The observation of population migration within the United Kingdom at the end of the nineteenth century led several authors to describe empirical spatial interaction laws (Ravenstein 1889), which were then formalized by analogy to the physical law of gravitation (Reilly 1929, 1931; Stewart 1947; Huff 1964; Tobler 1970). Hence the emergence of spatial interaction theory as a field of research examining the location and flow of social and economic processes, emphasizing the role of the attractiveness of places and their relative distances. Spatial interaction models have since evolved in different directions (Roy 2004), including three that are particularly important for this study: probabilistic treatment, cumulative dynamics, and explicit consideration of individual behavior. In the following section, we position our work with respect to these three evolutions.

Randomness and Stochasticity

Probabilistic treatment was introduced for spatial interaction models by Huff (1964) whose models aimed at identifying theoretical market areas within a set of central places included conditional probability distributions based upon place location. Probabilistic models were then popularized with the emergence of random utility theory and discrete choice models (Wegener 2000) and notably through the use of multinomial Logit models (Nijkamp and Reggiani 1987).

In spatial interaction models, randomness is mainly found within the destination choice stage. For coin diffusion this stage may not represent the main stochastic element. We hypothesize that the mixing of coins results from two sequential stages: (1) a mobility stage, i.e., flows of individuals who choose a destination; and (2) an exchange stage, i.e., flows of money between individuals in a given place. The latter is very likely to be stochastic compared to mobility, which is a reasoned choice made by individuals pursuing a specific objective. Even if we assume a simplified spatial interaction framework, the fact that there is a stochastic element at destination – and a cumulative effect through time due to repetitive movements and repetitive

exchanges – make coin diffusion a complex system. In the present work, we control for the spatial interaction stochasticity by leaving the destination choice aside and considering two cities only. We thus focus on the exchange stochasticity, making our approach closer to epidemic models which consider the probability of an individual to randomly have a contagious interaction at constant rate in time (Andersson and Britton 2000).

Complex Dynamics

Gravity models also evolved to account for complex system dynamics after the work of Allen and Sanglier (1979) and Wilson (1981). Again transferred from physical concepts, the integration of the 'arrow of time' (Prigogine and Stengers 1979; Haken 1977) in the analysis of cities lead to the emergence of principles of self-organization in geographical (Pumain 1998) or economic contexts (Arthur 1988), with particular focus on non-linear dynamics, irreversibilities and path-dependency (O'Sullivan 2004; Guermond 2008). Temporal and cumulative effects have been considered in various fields and at various geographical scales, including transportation and urban economics, urban geography, and interregional migrations (Anas 1983; White 1977, 1978; Allen and Sanglier 1979, 1981; Wilson 1981; Pumain et al. 1989; Weidlich and Haag 1988; Aziz Alaoui and Bertelle 2009).

How and whether or not the system can reach a steady state is analyzed in dynamic systems, and typically entropy is used as a measure of the degree of organization of the system (Wilson 1967, 1970). In the easiest cases, interactions between places lead to stable situations where the aggregate system descriptors do not change, or evolve only slowly. Yet, the resulting structure of the system, even if stable, is often found to be very sensitive to the values of the input parameters (Bretagnolle et al. 2003). In many cases, unstable situations are observed with either internal fluctuations within the system or external perturbations. Non-linear interactions can indeed lead to extremely complicated sets of bifurcations and trajectories, with no possibility to pre-determine any equilibrium point or other attractors (Wilson 1981, 2002; Allen and Sanglier 1981; Allen 1997; Bretagnolle et al. 2003). Moreover, in the case of the evolution of systems of cities through time, if one can find stable mathematical convergence properties, such as scaling laws (Batty 2008), this relative stability and regularity at the macro level might be coupled with a high volatility at the micro level (in terms of cities or firms, for example) (Portugali 2000; Pumain 2003; Pumain et al. 2009).

Furthermore, the causal relationships within a geographical system may create progressively growing contrasts between spatial units; pushing the system away from maximized entropy (Berry 1964). In econophysics, such a behavior has been observed in the distribution of wealth. Successive monetary transactions between agents in a closed system reach a thermodynamic equilibrium with Boltzmann-Gibbs distribution (Chakraborti et al. 2011; Dragulescu and Yakovenko 2001). In population or economic exchange systems, the speed toward convergence has been shown to depend on the degree of mobility (Rappaport 2004), the frequency of interactions (Busnel et al. 2008), as well as on the differences between places at the

original state, as illustrated in the case of economic convergence between countries (Barro and Sala i Martin 1992).

In the case of coin diffusion – which is a mixing process – in time, we can expect a homogenization across the system; however, we also hypothesize that the mixing process should reach an end, at which time the proportions of each coin type in each city match the overall proportions originally produced. Also, there is no apparent reason to think that the diffusion is reversible and would, for example, after reaching a certain mix, return to 100% national coins in each country. There are a few redistribution interventions from National Banks from time to time, but the system is largely bottom-up, driven by millions of individual exchanges. We therefore expect an irreversible process and some form of convergence. Whether some complex trajectories can also be observed because of path dependence is not known a priori. Similarly, some regions may have specific behavior given relative location, lower population ranking, or other regional factors; however, these considerations are beyond the scope of this initial research, in which simplistic geographical structure will be used.

Individual Choice and Aggregation

Aggregate spatial models directly describe the mechanisms of a system and its evolution using an ad hoc analytical form and a limited series of aggregated state variables (Weidlich 2003). Such a strategy provides indicators regarding system behavior and a set of interpretative clues without the need for simulations (Edwards et al. 2003). The main purpose of an aggregate model is then to understand the evolution of the phenomena overall, rather than its internal dynamics (Sanders 2007). Sometimes hypotheses regarding individual behavior can be integrated at an upper level, as is the case, for example, with the synergetic model of interregional migrations by Weidlich and Haag (1988).

Nevertheless, the form of aggregate models often requires simplification of the system under analysis; in spatial interaction models, this simplification is implemented via the hypothesis that agents are homogeneous and rational (Roy and Less 1981). Moreover the elementary interactions of individual components within a system may show collective dynamics that call into question the capability of equation-based modeling strategies to grasp the macroscopic behavior of many systems in many fields, including chemistry, biology, finance and social sciences (Shnerb et al. 2000; Gilbert and Troitzsch 2011). Interactions at the elementary levels are responsible for the production of intermediate or global processes and dynamics which may be important enough to contrast with macroscopic descriptions and predictions (as in the Lotka-Volterra model). The concept of emergence, relying on the idea that understanding a macro-scale phenomena necessitates the understanding of the most elementary components of the system (Solé et al. 1999), is therefore often preferred by modelers, despite difficulties in adaptation to real systems (Gil-Quijano et al. 2012). Compared to aggregate models, individual-based models (IBM) are also perceived as being easily applicable to empirical

issues because they are more often grounded in a heuristic approach to prediction (Bretagnolle et al. 1999).

An early attempt to integrate individuals in models was conducted by Orcutt (1957) who implemented operating characteristics at the individual and household levels to understand macroeconomic laws. In the particular context of spatial diffusion, the idea dates back to Hagerstrand's model where both a hierarchical model of the society and individuals within it, as well as spatial and temporal constraints upon individuals, are considered (Hägerstrand 1970, 1952). Many models actually account for mechanisms at different scales, notably in geography (Sanders 2007), in order to highlight impacts between scales (White et al. 1997; Quijano et al. 2007). Nevertheless, the complexity of those interrelations may make this analysis impossible; a solution is to consider the system at the most disagreggate level only. In transport research, the disaggregation of models to individuals or households offers a direct way to include economic decision-making explicitly in the choice of activities and trips, in contrast to the ad hoc "physics" of aggregate spatial interaction (Lerman 1979; Kitamura 1984a,b).

It is difficult to form a theory or discover facts about how people behave with coins and the many currency transactions they might have. Money bags can be surveyed at different time intervals, but each exchange of coins cannot be surveyed. The individual behavior which can be represented in a model is therefore simplistic, and randomness may play an important role. Even with this caveat, it is interesting to survey individuals. In this study, individuals have their own money bags, which they carry out when moving out, and whose content changes after a monetary transaction. As stated earlier, we consider only movement between two cities for simplification, therefore mobility choice at the individual level is not considered. Aggregate and disaggregate versions of the exchange stage were developed for the purpose of comparison.

Models of Coin Diffusion

One can identify two approaches to the analysis of coin diffusion in the literature. The first approach considers the circulation of coins to be a process of spatial migration between origins and destinations with the aim of understanding what type of persons and places contribute to circulating coins across borders, and whether barriers can be identified (Berroir et al. 2005). The second approach is epidemiological – the epidemic disease being represented by the presence of foreign coins in each country. Of course, coins are not strictly analogous to diseases or biological vectors since they cannot reproduce themselves; however, the circulation process shares similarities with epidemics in the sense that they both result from movements of people and individual contacts.

Currency circulation has been studied as a proxy of human mobility in many different scientific fields. Historians have been using coin distribution as a proxy for ancestral mobilities for decades already (Bursche 2002; Moisil 2002; Oberländer-

Törnoveanu 2002; Tsotselia 2002; Ujes 2002). Other disciplines began to consider this particular material only recently. Physicists and mathematicians have analyzed dollar bills (Brockmann and Hufnagel 2007; Brockmann 2008, 2010; Brockmann and Theis 2008) or euro coins (van Blokland et al. 2002; Stoyan et al. 2004; Seitz et al. 2009, 2012) using data publicly contributed over the Internet. Geographers have compared static potential models with observed distribution of dollar bills (Tobler 1981), or analyzed border and social effects in terms of "contamination" by foreign euro coins (Berroir et al. 2005; Grasland et al. 2002, 2005a,b, 2012; Grasland and Guérin-Pace 2003; Grasland 2009).

The case of euro coins has attracted interest because it is a unique example of a diffusion process for which the initial state is perfectly known (all coins minted were in their home countries on December 31, 2001). It offers the potential to determine whether the movement of currency follows a particular theoretical law, and whether we can predict the date upon which a perfect mix can be reached. The perfect mix is theoretically the final stage of the diffusion process where the share of each coin type is equal across countries and remains stable. The proportion corresponds to the proportions of the different coins in the total produced since the introduction of the Euro.

From a physical perspective, the diffusion between two places is expected to occur via dispersal from the place where the phenomena is the most concentrated toward the place where it is the least concentrated, as in heat or particle diffusion. Using Markov chains, van Blokland et al. (2002), Stoyan (2002), Stoyan et al. (2004), Seitz et al. (2009, 2012) designed models in which each country is characterized by the proportion of foreign coins, and entropy is measured through time. The proportions evolve according to constant but asymmetric transition probabilities. Those models are aggregated in the sense that they assume that the money bag of each person involved in a transaction is solely characterized by the proportion of foreign coins in his or her country. Using this modeling framework in the case of Germany, the perfect mix of coins was expected by 2012 (Stoyan 2002; Stoyan et al. 2004; Seitz et al. 2009); if losses (in terms of movement of the coins outside of the Eurozone, savings and losts) and new production of coins are accounted for, the perfect mix is expected to be delayed until 2050 (Seitz et al. 2012).

Three key assumptions underlie these models, the implications of which upon the evolution of the system will be tested. First, nation states are used as the relevant elementary entity, which raises the question of being less aggregate by considering individuals rather than territories. Second, the coins available in a given place depends only of the previous stage, meaning that the mobility and the exchange process are asynchronous. Third, the authors consider a uni-directional exchange at destination as only the 'movers' get money from the transactions. We consider the effects of these three assumptions by designing and comparing four different modeling strategies for the exchange process.

Another modeling approach has been taken by Grasland and Guérin-Pace (2003), which is not dynamic but includes both population and distance effects, as in gravity models. The authors analyze the effect of the distance decay parameter. Berroir

et al. (2005) use a similar strategy to identify barrier effects. Grasland et al. (2005a) employ a dynamic perspective which is close to the previous Markov models but considers distance between places. The probability for a coin to move toward a particular city is defined according to transition probabilities that decrease with distance from the border. The mix of coins is analyzed with respect to distance. The authors show that even a very active border area may only lead to a small level of penetration of foreign coins as observed at a given point in time since the coins are continuously going forward and backward from the border. When compared to the previous physical models, Grasland et al. (2005a) obtain a better fit to the observed mixing of coins in France and other countries.

In addition to testing the implications of the three assumptions underlying the Markov models discussed above, this model employs a spatial interaction (distance) effect, as in Grasland and Guérin-Pace (2003) and Grasland et al. (2005a). This is conducted in a simpler manner by analyzing sensitivity of model outputs to different cross-border mobility rates.

Model

Definitions and General Assumptions

Elements of the System

The model utilizes a simple geographical setting comprised of two places, and undifferentiated interaction flows that represent the movement of people between the two places. In the future, a more complex geographical system including several places with different population and attractiveness attributes (capital cities, tourist attractions,...) would enrich the approach and would facilitate the inclusion of different motives for and intensity of movement (travel, holidays, daily commute, weekend trips, etc.). In the present work, a simpler setting enables the focused examination of the impact of the various modeling assumptions stated above.

The geographical system consists of two places, a and b, endowed with total populations P_a and P_b. Within each population, a portion is mobile, M_a and M_b; the rest is immobile. The immobile agents are able to participate in transactions, although they do not move out of their home city. Mobility rates for each place are given by $m_a = \dfrac{M_a}{P_a}$ and $m_b = \dfrac{M_b}{P_b}$.

Each place produces a certain number of coins (one coin for each individual) with a specific symbol. A and B are the total number of coins from each place available in the system. A_a^t, B_a^t, A_b^t and B_b^t correspond to the total number of coins located in a city (a, b) at time t. The initial conditions are such that A and B are distributed only among their own total population at the beginning of the system: $A_a^{t_0} = A$; $A_b^{t_0} = 0$; $B_a^{t_0} = 0$; $B_b^{t_0} = B$, which depicts the Eurozone at the time euro currency entered circulation (January 1, 2002).

We define the stock of coins available in each place at time t according to their place of origin in relative terms: α_a (and, respectively, β_a, α_b, β_b) being the share of A (and, respectively, B) among the total number of coins located in a or b at time t (so that $\alpha_a^t + \beta_a^t = 1$ and equivalently for b):

$$\alpha_a^t = \frac{A_a^t}{A_a^t + B_a^t} \tag{1}$$

$$\beta_a^t = \frac{B_a^t}{A_a^t + B_a^t} \tag{2}$$

$$\alpha_b^t = \frac{A_b^t}{A_b^t + B_b^t} \tag{3}$$

$$\beta_b^t = \frac{B_b^t}{A_b^t + B_b^t} \tag{4}$$

We take particular interest into the value of α_b and β_a since they indicate the level of foreign coins in a given place. For example ten years after the Euro entered circulation, the share of foreign Euro coins was 35% in France (Grasland et al. 2012) and 75% in the Netherlands (Eurodiffusie 2011).

In this model, the system is closed. Each city produces one coin for each individual and there is no additional production nor loss of coins :

$$A_a^t + B_a^t = P_a \tag{5}$$

$$A_b^t + B_b^t = P_b \tag{6}$$

Last, we assume that every mobile person makes a trip to the other city at each time step. Every mobile person also makes a monetary transaction at destination. Thus, for a mobile agent, a time step consists in four successive actions : (1) a move, (2) a transaction, (3) a move, and (4) a transaction. For an immobile agent, a time step consists only in transactions.

Three additional assumptions underlie the mixing process. First, the direction of the transaction is either one-way or two-ways: in each transaction there is either a giver or a receiver, or both agents participate in the transaction. Such a transaction may represent a situation in which an individual purchases an item in a shop and receives change from the shopkeeper. Second, the exchange of money is either synchronous or asynchronous with the mobility of people: i.e. everybody moves and conducts transaction at the same time or sequentially. In the first case, it is considered that movers can only conduct a transaction with an immobile agent at destination; while in the second case, the people and money available in the city are defined after the moves of mobile agents. Third, each individual may interact with one or with several individuals, thus heterogeneity can be introduced in the quantity of transactions that each individual conducts.

Convergence Speed, Equilibrium Value and Perfect Mix

The methodology used in this study includes an analysis which examines how the mix of coins changes with time, and whether or not it converges to an equilibrium. The result of interest is the proportion of foreign coins within each city at equilibrium – i.e., β_a^* and α_b^*, which we define as: $\beta_a^* = lim_{t\to\infty}\beta_a(t)$ and $\alpha_b^* = lim_{t\to\infty}\alpha_b(t)$. We thus expect a static equilibrium, from which the properties of the system will remain unchanged over time. The equilibrium point of β_a is such that $\forall t \in N$, $\beta_a^{t+1} = \beta_a^t$ (similarly for α_b). We denote by t^* the moment at which these asymptotes are reached. Obviously, t^* is the same for both places.

Through simulation it is possible to analyze the existence and characteristics of this simplified system's equilibrium, as well as the time required to reach it. This process is presented in the following section. Prior to attempts to reach convergence, and independent of the equilibrium solution, the perfect mix was characterized as a mix in which $\alpha_a = \alpha_b$ and, accordingly, $\beta_a = \beta_b$. Since the shares of α and β are complementary, we can define the perfect mix using our two variables of interest: $\beta_a = 1 - \alpha_b$.

Set of Models

We design a set of three models, the last one being subdivided into two variants. Models *I* and *II* are aggregated, at a level similar to that of existing models in the literature – i.e., averaged money bags were calculated during the exchange process. Models *IIIa* and *IIIb* are disaggregated – each individual money bag is calculated in the exchange process.

For all models, numerical simulations are conducted (see section "Theoretical Simulations") in order to capture the diffusion process and cumulative effects. Some analytical properties can be derived from the first model using a discrete or continuous dynamic formalization (see Appendix). A thorough continuous analysis and comparison with numerical outputs is nevertheless outside the scope of the present work.

Model I

Model *I* implements an aggregated one-way exchange. This is in line with the models proposed by Stoyan (2002), Stoyan et al. (2004) and Seitz et al. (2009): the diffusion mix results from external trips only and is one-way. The one-way exchange applies hardly to real-life situation since no money is received by local population at destination. The money bag of immobile agents being unchanged at transaction time, the money available for a transaction is directly described by (A_i, B_i) at destination place for the receiving mobile agents. Mobility and exchange are asynchronous: at each time step t, mobile agents move then make a transaction

at destination. The mobile agents bring back a proportion of coins A and B that corresponds to the proportion in the destination place prior to their arrival.

The stock of money A (and, respectively, B) available in city a (and, respectively, b) at the end of a time step is therefore defined according to:

$$A_a^{t+1} = (1 - m_a)P_a\alpha_a^t + m_aP_a\alpha_b^t \tag{7}$$

$$B_a^{t+1} = (1 - m_a)P_a\beta_a^t + m_aP_a\beta_b^t \tag{8}$$

$$A_b^{t+1} = (1 - m_b)P_b\alpha_b^t + m_bP_b\alpha_a^t = A - A_a^{t+1} \tag{9}$$

$$B_b^{t+1} = (1 - m_b)P_b\beta_b^t + m_bP_b\beta_a^t = B - B_a^{t+1} \tag{10}$$

Each equation is made of two parts: the first refers to the unchanged money bags of the $(1 - m_a)P_a$ and $(1 - m_b)P_b$ immobile agents, and the second to the number of coins moving from a to b (or b to a), taken by the m_aP_a and m_bP_b mobile agents.

Model II

In daily life, both a buyer and a shopkeeper can receive money from a transaction. We therefore adapt Model *I* to realign with this hypothesis that coin transactions can be two-way. This kind of transactions may represent the money exchanged between a shop-keeper and a buyer, bills and tips, shared expenses, etc. If not every time, at least on average, mobile people and the people they meet can be both givers and receivers in turn. Model *II* therefore implements the idea that all people present in a given city and a given time have the same probability to get a coin from either a mobile agent visiting the city or from an immobile agent living in this same city. This two-way exchange process requires the total currency available in each place to be updated with the currency brought back by all agents, in order to constitute a stock before the next transactions take place.

Since Model *II* is aggregated and two-way, it automatically assumes that movements are synchronous – all mobile agents move together to contribute their coins to the stock, then each returns home with a money bag characterized by this averaged intermediate stock.[1] The averaged intermediate stock represents the average of the destination stock after the arrivals of mobile agents.

In Model *II*, we first define the intra-time update of the stock of coins available in cities after mobility has taken place:

[1] A two-ways asynchronous model would correspond to a situation where the total currency available in each place would be updated after the moves of agents. Mobile agent would nevertheless be able to conduct a transaction only with immobile agents at destination. Consequently, they would return home with the same money bag as in Model *I*. Besides, if the locals are the receivers, then they would receive the same money as if they were moving to the other city, again resembling Model *I*. A two-ways asynchronous model would thus add little sense at an aggregated level and is not analyzed in this paper.

$$\bar{A}_a^t = (1 - m_a)P_a\alpha_a^t + m_bP_b\alpha_b^t \tag{11}$$

$$\bar{B}_a^t = (1 - m_a)P_a\beta_a^t + m_bP_b\beta_b^t \tag{12}$$

$$\bar{A}_b^t = (1 - m_b)P_b\alpha_b^t + m_aP_a\alpha_a^t \tag{13}$$

$$\bar{B}_b^t = (1 - m_b)P_b\beta_b^t + m_aP_a\beta_a^t \tag{14}$$

Hence, the share of each coins available for transactions in city a at time t is denoted by:

$$\bar{\alpha}_a^t = \frac{\bar{A}_a^t}{\bar{A}_a^t + \bar{B}_a^t} \tag{15}$$

$$\bar{\beta}_a^t = \frac{\bar{B}_a^t}{\bar{A}_a^t + \bar{B}_a^t} = 1 - \bar{\alpha}_a^t \tag{16}$$

$$\bar{\alpha}_b^t = \frac{\bar{A}_b^t}{\bar{A}_b^t + \bar{B}_b^t} \tag{17}$$

$$\bar{\beta}_b^t = \frac{\bar{B}_b^t}{\bar{A}_b^t + \bar{B}_b^t} = 1 - \bar{\alpha}_b^t \tag{18}$$

with $\bar{\alpha}_a^t$ describing the share of money A in the city a after every mobile agent moved, and accordingly, $\bar{\alpha}_b^t$ describing the share of A in b, $\bar{\beta}_a^t$ the share of B in a and $\bar{\beta}_b^t$ the share of B in b.

Given these intermediate proportions (Eqs. 15, 16, 17, and 18), the stocks of coins in each place at the end of the time step (Eqs. 7, 8, 9, and 10) are redefined into:

$$A_a^{t+1} = (1 - m_a)P_a\bar{\alpha}_a{}^t + m_aP_a\bar{\alpha}_b{}^t \tag{19}$$

$$B_a^{t+1} = (1 - m_a)P_a\bar{\beta}_a{}^t + m_aP_a\bar{\beta}_b{}^t \tag{20}$$

$$A_b^{t+1} = (1 - m_b)P_b\bar{\alpha}_b{}^t + m_bP_b\bar{\alpha}_a{}^t \tag{21}$$

$$B_b^{t+1} = (1 - m_b)P_b\bar{\beta}_b{}^t + m_bP_b\bar{\beta}_a{}^t \tag{22}$$

Model III

Model *III* uses a disaggregate approach, with the exchange process defined at the level of individuals. Mobility is still implemented at the inter-city level, since the model includes only two cities and the share of mobile people is known and fixed. Models *IIIa* and *IIIb* are both two-way asynchronous models.

In contrast to Model *II*, there is no intermediate, aggregate 'pot of coins' but a sequence of iterative meetings between two individuals present at the same place. Exchanges are sequentially simulated after randomly pairing individuals in the list

of people present. However, in order to be comparable with the previous aggregate models in terms of quantity of exchanges and the mixing process, two additional assumptions were made:

First, for an individual, the mixing occurs at the destination, but also back home (as in Model *I* and *II*, in which the second stage was included in the update process). The equivalent in disaggregate modeling is to model interpersonal exchanges twice: agents conduct transactions at the destination, and then return home where the conduct transactions again. The list of individuals present at destination is drawn from the list of immobile people of the destination place and the mobile agents from the concurrent city. Back home, the list is made of the population of the city – the immobile agents plus the mobile agents (having returned home).

Second, the quantity of transactions is set similarly to aggregated models in which every exchanger is assumed to have one interaction. Therefore the number of transactions is fixed at half of the population present in a place, i.e., $E_a = [(1 - m_a)P_a + m_bP_b]/2$ and $E_b = [(1 - m_b)P_b + m_aP_a]/2$ for the first round at destination, and $E'_a = P_a/2$ and $E'_b = P_b/2$ for the second round at home.

The mix of coins in a given place after completion of a full time step is then measured when every mobile agent is back home:

$$\alpha_a^t = \sum_{i_a}^{P_a} \frac{\alpha_{i_a}^t}{P_a} \tag{23}$$

$$\beta_a^t = \sum_{i_a}^{P_a} \frac{\beta_{i_a}^t}{P_a} \tag{24}$$

$$\alpha_b^t = \sum_{i_b}^{P_b} \frac{\alpha_{i_b}^t}{P_b} \tag{25}$$

$$\beta_b^t = \sum_{i_b}^{P_b} \frac{\beta_{i_b}^t}{P_b} \tag{26}$$

where $(\alpha_{i_a}^t, \beta_{i_a}^t)$ and $(\alpha_{i_b}^t, \beta_{i_b}^t)$ represent the money bag mix for each individual i at time t after exchange.

The pairing of individuals necessitates a third assumption: E_a, E_b, E'_a and E'_b are sample pairs obtained from a random drawing with replacement. This reflects the ideas that: (1) all exchanges are sequential and the model is thus asynchronous; (2) each agent can make several (or no) transactions regardless of its origin; and (3) at destination, the probability to exchange with a local or a visitor is equal. The latter sounds reasonable, although we can imagine that visitors may concentrate in parts of the city and have more interactions among themselves, as it may be the case in souvenir shops, for instance.

Finally, the exchange between two people itself necessitates several assumptions. We know from Nuno et al. (2005) that individuals have different strategies across transactions, leading to different numbers of coins being exchanged. We also know that the value of coins and the characteristics of individuals influence the number of coins exchanged (Grasland 2009). It is therefore interesting to test the impact of how many coins are exchanged at each transaction onto the convergence level and rate of foreign coin diffusion.

We consider two possibilities. In model *IIIa*, the exchange consists of averaging the money bags of the paired individuals. After exchange, the two money bags are thus similar.

For a pair of individuals i = (1,2) and τ, referring to an intra-time exchange iteration, we have

$$\alpha_1^\tau = \alpha_2^\tau = \alpha_1^{\tau-1} + \alpha_2^{\tau-1}/2 \tag{27}$$

and similarly for β (which is also equal to $1 - \alpha$). Note that when $\tau = 1$ then $\alpha^{\tau-1} = \alpha^\tau$ and $\beta^{\tau-1} = \beta^\tau$.

In model *IIIb*, a different strategy is used that is expected to lead to the quickest mixing of coins. Instead of averaging the money bags, agents 1 and 2 simply swap money bags:

$$\alpha_1^\tau = \alpha_2^{\tau-1} \tag{28}$$

$$\alpha_2^\tau = \alpha_1^{\tau-1} \tag{29}$$

and similarly for β. This type of transaction could correspond, for instance, to a situation in which two individuals exchange plenty of coins of small values against a few coins of high values.

These models represent only two strategies out of many possibilities. There are intermediate situations in which only a part of each money bag is exchanged, or even more extreme situations where only one coin is exchanged during a transaction. Testing this variety of strategies in its entirety is not feasible for one study. With models *IIIa* and *IIIb*, we aim to show that a disaggregate model of exchange based upon individuals, rather than aggregated exchange, might in itself provide insight into the complexity of the diffusion process, even within a simplified geography.

Theoretical Simulations

In this section, numerical outputs from models *I*, *II*, *IIIa* and *IIIb* are analyzed and compared based upon the sensitivity of the equilibrium value (β_a^*) and convergence speed (t^*) to changes in population and mobility rates.

Exogenous Parameters

Sensitivity analysis is based upon a fixed $P_b = 1{,}000$ and $m_b = 0.10$ and an incrementally varying P_a from 100 to 2,000, and m_a from 0.01 to 0.20. The benchmark situation is such that both cities have equivalent exogenous parameters, i.e., $P_a = P_b = 1{,}000$ and $m_a = m_b = 0.10$. m_a and m_b can theoretically be any proportion, but 0.10 is used as a mid-range benchmark. As an indication, 1.1% of the French active population was a cross-border commuter in 2011 (Floch 2011). This would be our lower limit, knowing that other trip purposes contribute to the diffusion of coins.

Population is fixed at the values indicated. These figures were chosen in order to avoid both very small values (because of random effects), and very high values (out of concern for computational limits of this particular research effort). The median population in the Eurozone countries in 2011 was 9.5 million inhabitants. The first and third quartiles were, respectively, 1.9 million and 24 million inhabitants (Eurostat 2012). Sensitivity ranging from 100 to 2,000 was therefore considered to be reasonable.

Results

Benchmark: Equal Population and Mobility Ratios

The benchmark case uses same characteristics for both places, in order to measure the effects of the modeling strategy without external influence (i.e., without altering the values of exogenous parameter inputs P_a, P_b, m_a and m_b). Following Seitz et al. (2012), if there are no external influences, we expect convergence and perfect mix at equilibrium for the two aggregated models. Also, there is no a priori reason not to reach convergence in the two disaggregated cases, since apart from stochasticity, there is no additional interaction.

All of the models converge, and the convergence process is depicted in Fig. 1. Models *I* and *IIIb* demonstrate very similar behavior. The exchange process is actually very similar in both models, even though the elementary scale is different: in Model *IIIb*, people swap money bags, and they do not mix up their coins. In Model *I*, visitors change their money bags to mimic the locals' money bags. It is surprising, however, to find Model *IIIb* so close to the most basic aggregated model, and so different from Model *IIIa*. This definitely highlights the importance of carefully considering what an aggregate model assumes in terms of individual behavior.

Comparing Models *IIIa* and *IIIb* shows that when more coins are exchanged between individuals, convergence is accelerated. The same idea holds when comparing Model *I* and Model *II*, in which people at the destination also receive money, instead of being givers only.

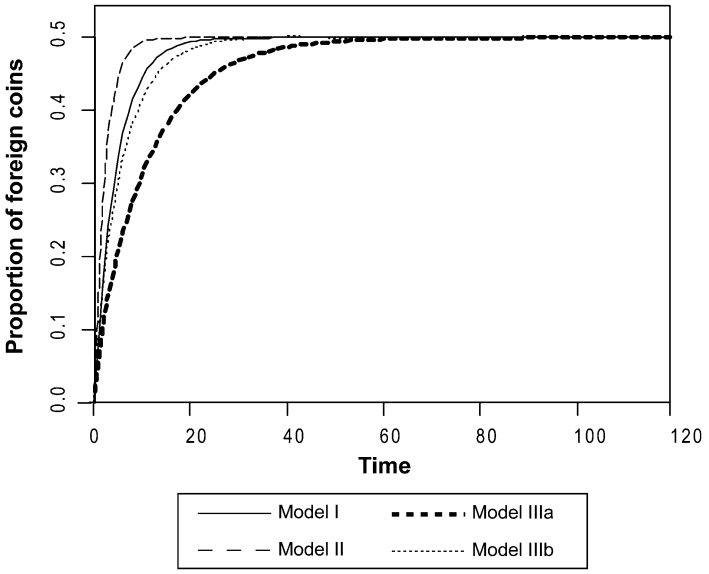

Fig. 1 Evolution of coin mixing (β_a) over time for the four models

Stochasticity in Models *IIIa* and *IIIb* may raise additional questions regarding these results. Therefore Fig. 2a, b report the uncertainty around the convergence state after a Monte Carlo process with 100 replications. The variability of the equilibrium is small in the case of Model *IIIb* and almost nonexistent in Model *IIIa*.

Sensitivity to Mobility and Population Differentials

We show in Fig. 3a–d the sensitivity of the equilibrium mix (β_a^*) to changes in mobility rates and population differentials between places. The grids represent incremental variations of the ratio of the two input parameters between the two cities. The darker the square, the higher the proportion of foreign coins in place a at equilibrium. The color gradient corresponds to different values for each model.

Figure 3a reveals a strong difference in the behavior of Model *I* compared to the other models. Results from Model *I* are quite counter-intuitive. In fact we expected the differentials of population between places to determine the level of the coins mixing at equilibrium as is the case in Models *II*, *IIIa* and *IIIb*. The proportion of foreign coins at equilibrium should correspond to the ratio of populations, since each city produces one coin per individual: the higher the population of b compared to a, the higher the share of foreign coins in a at equilibrium.

But in Model *I*, β_a^* is only dependent on mobility differentials. Indeed, a closer look at the model shows that the total number of coins is preserved, but not the global share of each coin type in the system – which changes with time in a complex

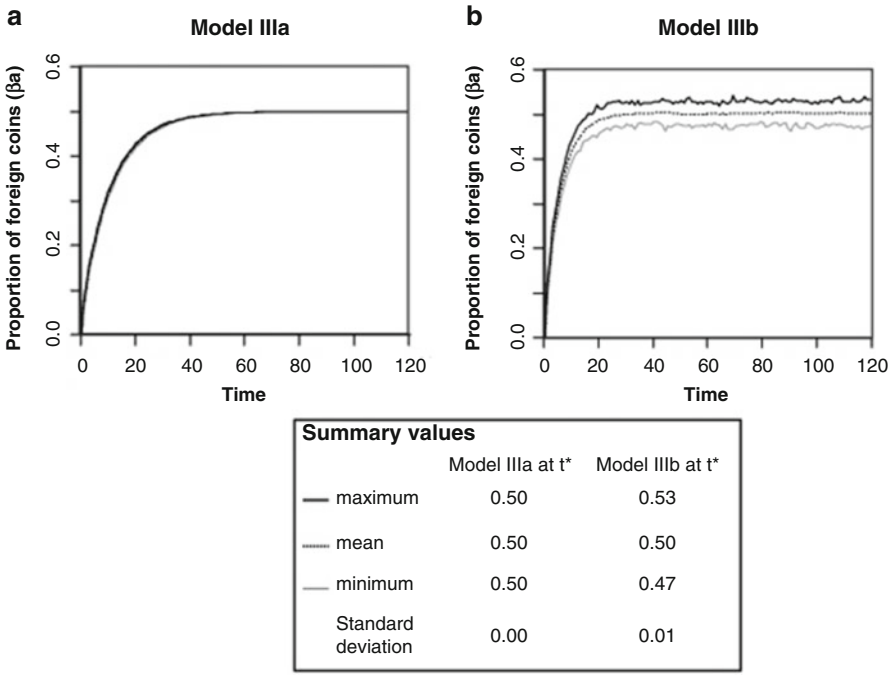

Fig. 2 Evolution of coin mixing (β_a) over time – Monte Carlo analysis for Models *IIIa* and *IIIb*

manner (see section "Calibration: France Versus the Eurozone" and "Appendix"). This transformation in the type of coins is of course unwanted. Model *I* is however very useful methodologically because it is the closest to Stoyan (2002), Stoyan et al. (2004), Seitz et al. (2009) models who postulate a constant number of coins, but do not control for it. Indeed, the dynamic is not simulated on the stocks but on the share of the origins of coins in Germany and the rest of the Eurozone. When simulating the dynamics of these stocks in Model *I*, we show that *A* or *B* coins are "created" or "deleted" symetrically revealing that the system is not properly closed.

Figure 4a–d show the sensitivity of the time to reach equilibrium, i.e., the sensitivity of t^* to changes in mobility rates and populations. The darker the square, the more time is needed to reach equilibrium. Again, the color gradient corresponds to different values for each model. Model *I* again shows a unique pattern. However, there are also strong differences for t^* between Models *II*, *IIIa* and *IIIb*, whereas these models had similar patterns for β_a^*.

The convergence speed in Model *I* is not influenced by changes in the population ratio, but is highly dependent upon mobility rates. Quite surprisingly, equivalent mobility rates for cities do not lead to the quickest mixing of coins. On the contrary, the higher is the differential of mobility in favor of city *a*, the quicker is the convergence speed to the equilibrium value of the share of foreign coins within the

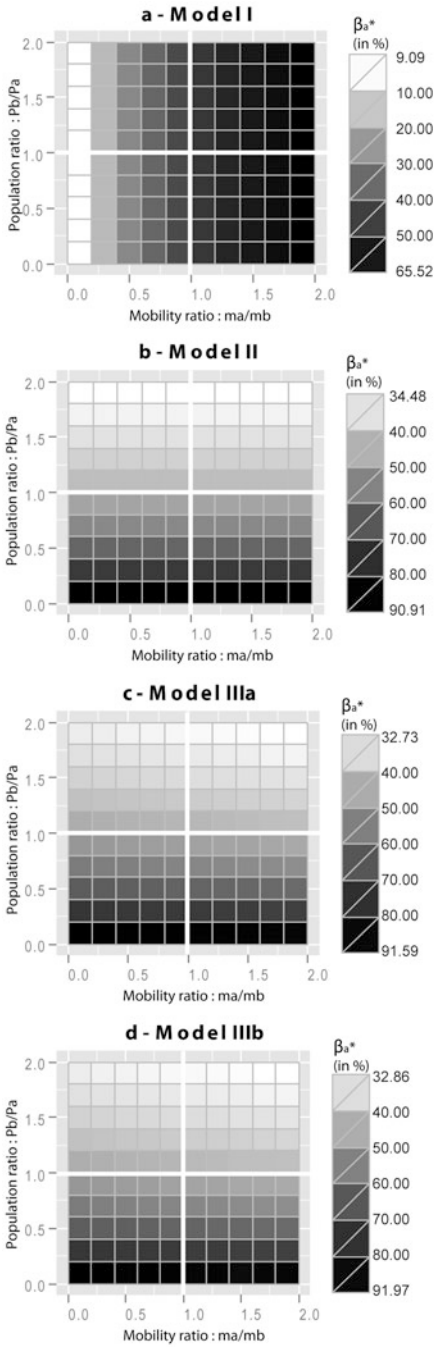

Fig. 3 Sensitivity of the equilibrium mix β_a^* to changes in the ratio of populations and mobility rates

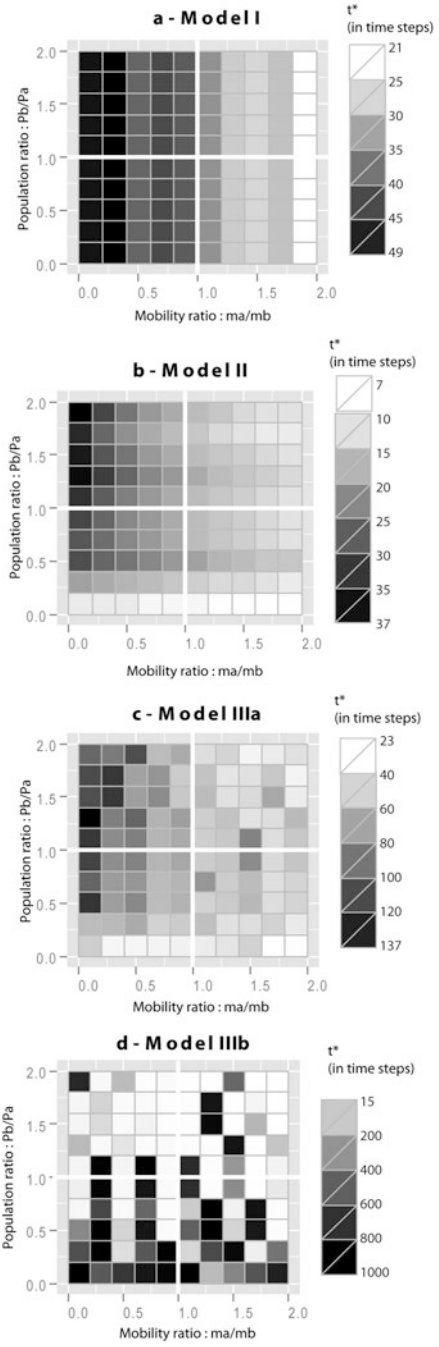

Fig. 4 Sensitivity of the time to convergence, t^*, to changes in the ratio of populations and mobility rates

city (i.e., β_a^*). Although the simulation shows that some combinations of mobility rates are associated with lower time to convergence than others, the resulting value of t^* is not strictly proportional to the ratio of mobility.

In comparing Fig. 4b–d, we observe a certain similarity of the impacts of mobility rates and population differentials in Models *II* and *IIIa*, with a positive effect of mobility superiority and a negative effect of population superiority upon the convergence speed. Model *IIIb* seems to be much sensitive to the interaction of both mobility rates and population differentials than model *IIIa*, which appears itself more sensitive than model *II*.

Calibration: France Versus the Eurozone

This section presents the calibration of our four models using data for France to determine whether the evolution of the mix can be predicted, and whether or when the perfect mix can be reached. In order to calibrate the model, the proportion of foreign coins in France is observed at 16 points in time via a representative survey of an average of 1,500 individuals and money bags per survey (Grasland et al. 2012). The error (*Err*) of each model is defined as the sum of the squared residuals between the estimated values and the 16 observations.

Before testing the four models, we first develop an empirical model. We test for various functional forms and find that a logistic function[2] best fits the data. Using maximum likelihood estimation we find:

$$\beta_a(t) = 0.3541 * (1 - e^{-t/43.72}) \tag{30}$$

The error amounts to 25.98. The curve is displayed in Fig. 5. The model results in an asymptotic value of the mix at $\beta_a = 0.3541$, to be obtained in April 2015. However β_a is far below the expected full mix level 0.75, i.e. the ratio of the two populations. The logistic curve is often used to model diffusion processes (Casetti 1969), although in this case, despite a good fit to empirical data, its usage remains ad hoc to represent mobility and diffusion. This small test emphasizes that some processes come into play which considerably reduce the mixing process – typically additional barrier effects, hoarding, external dissemination of coins, etc., which it is not possible to reflect within the framework of this analysis.

We now turn to fitting our aggregate and disaggregate models. This is an attempt to understand how the introduction of mobility and exchange behavior can affect our ability to forecast the diffusion process. In this calibration, populations are exogenously fixed, while mobility rates (m_a for France and m_b for the rest of the Eurozone) are adapted in order to minimize the error term. As explained earlier, our interest is indeed to use euro coin diffusion as a marker of international movements.

[2] Asymptotic regression with lower limit at 0 (2 parameters).

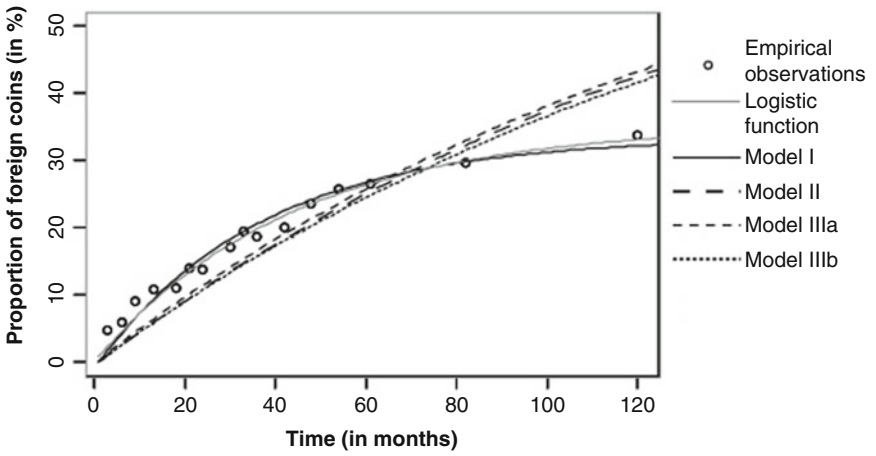

Fig. 5 Fitting of the estimated diffusion curves to empirical observations

Population parameters are taken according to Eurostat (Eurostat 2012). On January 1, 2012 the population of the Eurozone was 333 million inhabitants, as compared to 65 million in France alone (19.5% of the Eurozone total). In order to avoid computational burdens, the number of agents implemented is set at 654 (P_a) and 2,674 (P_b).

Again, the mobility rates can take any value ranging from 0 to 1. In the absence of global information about international mobility between France and the rest of the Eurozone, there is thus no preconceived starting point for the exploration. First, a crude calibration is conducted in order to narrow down the values of m_a and m_b by reducing *Err* by trial and error. Population being fixed, this exploration reveals monotonic behavior. Then, we vary m_a and m_b systematically from 0.001 to 0.030 by a 0.001 increment and run each model for this set of parameters.

The models also include a time increment. The unit was defined as a month, which of course impacts the results. However, for a given set of population and migration rates, varying the time increment is similar to changing the number of exchanges. Here one exchange per month was assumed for simplification. The measurement unit of the mobility rate is therefore to be interpreted with these two assumptions in mind.

Simulation results are displayed in Fig. 5 and Table 1.

Figure 5 shows that β_a from Model *I* follows a curve that is very close to the logistic estimate. Error is slightly increased but much lower than that of the other three models. Model *I* predicts an equilibrium occurring in May 2020. As with the logistic model, perfect mix is not reached: only 34% of coins in France are foreign coins at equilibrium. Upon closer look at the resulting equilibrium values, we find that they correspond to $\beta_a^* = m_b P_b^2$ for France ($\beta_b^* = m_a P_a^2$ for Eurozone) and $\alpha_b^* = m_a P_a^2$ for France ($\alpha_a^* = m_b P_b^2$ for Eurozone).

Table 1 Calibration of the four models to empirical values

	Logistic	Model *I*	Model *II*	Model *IIIa*	Model *IIIb*
P_a		654	654	654	654
P_b		2,674	2,674	2,674	2,674
m_a		0.009	0.001	0.017	0.003
m_b		0.018	0.001	0.001	0.002
Err	0.003	0.004	0.026	0.023	0.025
β_a^*	0.35	0.34	0.80	0.80	0.80
t^*	161	221	861	756	746
Mix date	04/2015	05/2020	09/2073	12/2064	02/2064

Consequently, the numbers of coins A^t and B^t differ from the initial amounts across time. After calibration, for example, Model *I* transforms the number of French coins from 654 to 1332 and in parallel reduces the number of Eurozone coins from 2674 to 1996. Total number of coins is constant, but this model does not preserve the initial share. This is problematic, but also striking, in that this is the best model to fit the data. We interpret this as the result of the production of new coins and losses (coins leaving the Eurozone or being hoarded). There are no reported numbers related to losses (European 2003), but we know, for example that new currency production has not being homogeneous in time and space. According to the European Central Bank, 56% of all French euro coins in existence in 2010 were minted prior to the end of 2002; for Luxembourg, this share is 31%.

Model *II*, Model *IIIa* and Model *IIIb* have the same equilibrium value. Indeed, models incorporating a reciprocal exchange process predict that the convergence value will be close to the ratio of population differences, leading to a longer time until equilibrium. Indeed, Model *II* expects an equilibrium in January 2073, i.e. more than 70 years after the introduction of the euro as a common currency. The design of the exchange process does not appear to have great influence on the convergence speed, since Model *IIIa* converges after December 2064 and Model *IIIb* after February 2064. Nevertheless, the fact that the individuals switch the entire contents of their money bags in Model *IIIb* results in an equilibrium value varying from an estimation of 80.1% to 80.7% of total euro coins estimated to be foreign euro coins in France.

Even if Models *II*, *IIIa* and *IIIb* give close *Err* or, equilibrium value and time to convergence, the estimated values of m_a and m_b are very different from one modeling strategy to another. The aggregate model (model *II*) uses equivalent mobility rates for the two places, and the mobility rates are very low (only 1% of the global population of place *a* and place *b*), signifying that the number of interactions between France and the rest of the Eurozone is assumed to be very low. Besides, the design of the exchange process impacts the values of the input parameters m_a and m_b; the lower the number of coins exchanged during a transaction, the higher the mobility of the residents of place *a* has to be to compensate for the differences of population. Indeed, model *IIIa* estimates that the mobility of French people (17%) is

much higher than the mobility of people of the rest of the Eurozone entering France (1%), while model *IIIb* allows for few differences between those mobility rates; the mobility in France being set much lower (3%).

Conclusion

This study aimed at illustrating the results of a particular modeling strategy for analysis using a dynamic spatial interaction model.

Model *I* has been found to be very unique among the models tested: the effects of the input parameters on the equilibrium value and convergence speed differ largely from results in Models *II*, *IIIa* and *IIIb*. The integration or non-integration of substantial reciprocity in the exchange process seems indeed to be the preeminent choice in the design of the models. Consequently, the calibration of Model *I* resulted in very different adjusted parameters and predictions compared to the three other models.

The mechanisms under investigation in Model *I* (in which external trips are the impetus for diffusion, and exchange is only one-way (van Blokland et al. 2002; Stoyan et al. 2004; Seitz et al. 2009, 2012)) and the logistic curve offer very similar calibration to empirical values and time predictions. Since Hägerstrand (1952) the logistic model has often been shown to fit diffusion processes over time, our result may question the benefit of aggregate models to provide modeling benefits within such a simple geographical space. Conversely the disaggregate approaches helped in revealing how aggregate patterns emerge over time from individual behavior.

The system being comprised of two countries, randomness in Models *IIIa* and *IIIb* is not defined at the stage of the destination choice like it has mainly been the case since Huff (1964), but rather is implemented at the moment of the exchange. Models *IIIa* and *IIIb* result in a low sensitivity of the system to the elementary interactions of individuals components. This is due to all mobile agents sharing identical mobility behavior. Low sensitivity to random pairing (equal probability for agents to participate in a transaction) could as well been deduced from the central limit theorem (Roy and Less 1981).

The calibration to empirical values has shown minor differences in the convergence speed of the three models, as expected from Hägerstrand (1952), Shnerb et al. (2000), Sanders (2007), and Quijano et al. (2007). These differences stand in the design of the elementary transactions: mean behavior (model *II*) vs mean behavior with stochasticity (model *IIIa*) and differences in the number of coins exchanged (model *IIIa* and model *IIIb*).

These results are based upon testing of four models, with particular assumptions. Expanded testing with additional models, using a broader range of assumptions and additional sensitivity analyses, will be undertaken in future research to follow up on these results.

In the study of this simple system, we have considered that individuals will participate in one exchange per time step on average, to simplify the comparison with the aggregate models. However, the question of the impact of the number of exchanges per time step is still unanswered, and will be the subject of further research. Controlling the range of variability within the different types of exchanges is expected to bring new understanding of the particular effects of mobility upon the diffusion process.

We should be prudent when using coin diffusion modeling to generalize about international movements. Coins are not equivalently involved in transactions according to the prices of the products, or the payment strategies adopted by individuals (Nuno et al. 2005), or even the characteristics of the people. Empirical surveys (Grasland et al. 2012) have shown for example that the content of money bags depends on individual characteristics: elderly women carry on average more coins than young men, etc. Also, cash represents a substantial parts of payments in Europe (de Meijer 2010), which might not be the case in other regions of the world. In the European context we assume that coin diffusion is a possible proxy for analyzing individual mobility and, subsequently, the internationalization of territories.

Finally, further research efforts will attempt to: (1) increase the realism/complexity of the geography of the system by introducing new places (in a 1-D and/or 2-D space) and by differentiating the attractiveness of the different places; and (2) articulate assumptions on macroscopic mobility with sociological parameters of individual behavior. These assumptions may be best investigated by preliminary social network analysis and methods based in time-geography in order to accurately model mobility decisions (scheduling of activities and destination choices).

Appendix

Mathematical resolutions of Model *I* are presented in the following subsections.

Discrete Formalism

From Eqs. (1) and (3) on the one hand and (5) and (6) on the other hand, within Eq. 7, we find that:

$$A_a^{t+1} = (1 - m_a)P_a\alpha_a^t + m_a P_a \alpha_b^t$$

$$A_a^{t+1} = (1 - m_a)P_a\frac{A_a^t}{P_a} + m_a P_a \frac{A_b^t}{P_b}$$

$$A_a^{t+1} = \frac{(1 - m_a)P_a A_a^t}{P_a} + \frac{m_a P_a A_b^t}{P_b}$$

$$A_a^{t+1} = (1 - m_a)A_a^t + \frac{m_a P_a}{P_b}(P_b - B_b^t)$$

$$A_a^{t+1} = (1 - m_a)A_a^t + \frac{m_a P_a}{P_b}(P_b - (P_b - B_a^t))$$

$$A_a^{t+1} = (1 - m_a)A_a^t + \frac{m_a P_a}{P_b}B_a^t$$

$$A_a^{t+1} = (1 - m_a)A_a^t + \frac{m_a P_a}{P_b}(P_a - A_a^t)$$

$$A_a^{t+1} = (1 - m_a)A_a^t + \frac{m_a P_a}{P_b}(P_a - A_a^t) \tag{31}$$

Knowing that for A_a, we can proceed similarly with others:

$$B_a^{t+1} = (1 - m_a)B_a^t + \frac{m_a P_a}{P_b}(P_b - B_a^t) \tag{32}$$

$$A_b^{t+1} = (1 - m_b)A_b^t + \frac{m_b P_b}{P_a}(P_a - A_b^t) \tag{33}$$

$$B_b^{t+1} = (1 - m_b)B_b^t + \frac{m_b P_b}{P_a}(P_b - B_b^t) \tag{34}$$

Knowing that $A_a^{t+1} = (1 - m_a)A_a^t + \frac{m_a P_a}{P_b}(P_a - A_a^t)$, if we define $x = (1 - ma)$, $y = \frac{m_a P_a}{P_b}$, $a = x - y$ and $b = yP_a$, we have:

$$aA_a^t + b \tag{35}$$

We know that arithmetico-geometric series may be solved by the following theorem: $\forall n \geq n_0, U_{n+1} = aU_n + b$, having $r = \frac{b}{1-a}$, thus $\forall n \geq n_0, U_n = a^{n-n_0}(U_{n0} - r) + r$, with U being the arithmetic series and r being the common difference. Consequently, the change in the stocks of coins of each city varies according to the populations in the two cities and mobility rates of the city where coins are observed.

Continuous Formalism

Trying to define the variation of A_a, A_b, B_a et B_b under the formalism of deterministic ordinary differential equations of first order, we find that:

$$A_a^{t+1} - A_a^t = -m_a\left(1 + \frac{P_a}{P_b}\right)A_a^t + \frac{m_a P_a^2}{P_b}$$

Consequently, with d being the derivative, the equilibrium point which verifies $\frac{dA_a}{dt} = 0$ is given by the following expression:

$$\frac{dA_a}{dt} = \frac{P_a^2}{P_b + P_a}$$

Similarly, the equilibrium points which verify $\frac{dB_a}{dt} = 0$, $\frac{dA_b}{dt} = 0$ and $\frac{dB_b}{dt} = 0$ are given by the expressions:

$$B_a = \frac{P_a P_b}{P_b + P_a}$$
$$A_b = \frac{P_a P_b}{P_b + P_a}$$
$$B_b = \frac{P_b^2}{P_b + P_a}$$

The variation between two consecutive states are dependent upon both the mobility rates in the place of minting and the populations differences between the two places. The value of the equilibrium points of the redistribution of coins between cities is only dependent upon the population shares, i.e., upon the number of coins originally produced.

Acknowledgements The authors are grateful to Nathalie Corson, Florent Le Néchet, Hélène Mathian, Romain Reuillon and Clara Schmitt for their help. This research benefited from funding by the *Fonds National de la Recherche* (FNR) in Luxembourg (AFR grant PHD-09-158).

References

Allen, P., & Sanglier, M. (1979). A dynamic model of growth in a central place system. *Geographical Analysis, 11*, 256–272.

Allen, P., & Sanglier, M. (1981). Urban evolution, self organisation and decision-making. *Environment and Planning, 13*, 168–183.

Allen, P. M. (1997). *Cities and regions as self-organizing systems: Models of complexity*. London: Taylor and Francis.

Anas, A. (1983). Discrete choice theory, information theory and the multinomial logit and gravity models. *Transportation Research B, 17*, 13–23.

Andersson, H., & Britton, T. (2000). *Stochastic epidemic models and their statistical analysis*. New York: Springer.

Arthur, W. B. (1988). Urban systems and historical path dependence. In J. Ausubel & R. Herman (Eds.), *Cities and their vital systems: Infrastructure, past, present and future* (pp. 85–97). Washington, DC: National Academy Press.

Aziz-Alaoui, M., & Bertelle, C. (Eds.) (2009). *From system complexity to emergent properties*. Berlin: Springer.

Barro, R. J., & Sala-i-Martin, X. X. (1992). Convergence. *Journal of Political Economy, 100*, 223–251.

Batty, M. (2008). The size, scale, and shape of cities. *Science, 319*(5864), 769–771.

Berry, B. J. L. (1964). Cities as systems within systems of cities. *Papers in Regional Science, 13*, 147–163.

Berroir, S., Grasland, C., Guérin-Pace, F., & Hamez, G. (2005). La diffusion spatiale des pièces euro en Belgique et en France. *Revue belge de Géographie (Belgeo), 4*, 345–358.

Bretagnolle, A., Daudé, E., & Pumain, D. (2003). From theory to modelling: Urban systems as complex systems. Cybergeo: European Journal of Geography, Dossiers, 13e Colloque Européen de Géographie Théorique et Quantitative, Lucca, Italie, 8–11 Septembre 2003. http://cybergeo.revues.org/2420

Bretagnolle, A., Mathian, H., Pumain, D., & Rozenblat, C., Dossiers, 11th European Colloqium on Quantitative and Theoritical Geography, Durham Castle, UK, September 3–7 1999. Long-term dynamics of European towns and cities: towards a spatial model of urban growth. Cybergeo: Revue européenne de géographie [on line], mis en ligne le 29 Mars 2000. http://cybergeo.revues.org/566

Brockmann, D. (2008). Money circulation science – fractional dynamics in human mobility. In K. Rainer, G. Radons, & I. M. Sokolov (Eds.), *Anomalous transport: Foundations and applications* (pp. 459–484). Weinheim: Wiley-VCH.

Brockmann, D. (2010). Following the money. *Physics World, 23*, 31–34.

Brockmann, D., & Hufnagel, L. (2007). The scaling law of human travel – a message from George. In: B. Blasius (Ed.), *Nonlinear modeling in ecology, epidemiology and genetics* (pp. 109–127). Hackensack: World Scientific.

Brockmann, D., & Theis, F. (2008). Money circulation, trackable items, and the emergence of universal human mobility patterns. *IEEE Pervasive Computing, 7*, 28–35.

Bursche, A. (2002). Circulation of Roman coinage in northern Europe in late antiquity. *Histoire & Mesure, XVII*(3), 121–141.

Busnel, Y., Bertier, M., & Kermarrec, A. M. (2008). On the impact of the mobility on convergence speed of population protocols. Research Report, Institute National de Recherche en Informatique et Application (INRIA), No 6580, July 2008.

Casetti, E. (1969). Why do diffusion processes conform to logistic trends? *Geographical Analysis, 1*(1), 101–105.

Chakraborti, A., Muni Toke, I., Patriarca, M., & Abergel, F. (2011). Econophysics review: II Agent-based models. *Quantitative Finance, 11*(7), 1013–1041.

de Meijer, C. R. W. (2010). The single euro cash area. *European Payments Council Newsletter, 8*(10), 1–9.

Dragulescu, A., & Yakovenko, V. M. (2001). Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States. *Physica A, 299*, 213–221.

Edwards, M., Huet, S., Goreaud, F., & Deffuant, G. (2003). Comparing an individual-based model of behaviour diffusion with its mean field aggregate approximation. *Journal of Artificial Societies and Social Simulation, 6*(4), http://jasss.soc.surrey.ac.uk/6/4/9.html.

Eurodiffusie. (2011). Volg de verspreiding van de euromunten door Europa! http://www.eurodiffusie.nl/. Accessed Feb 2011.

European, C. (2003). The introduction of euro banknotes and coins – one year after. *Official Journal of the European Union*. [COM(2002) 747 final - Official Journal C 36 of 15.02.2003].

Eurostat. (2012). Statistic database. http://epp.eurostat.ec.europa.eu/. Accessed Sept 2012.

Floch, J.-M. (2011). Vivre en deca de la frontière, travailler au-delà. *Insee Première, 1337*, 1–4.

Fotheringham, A., & O'Kelly, M. E. (1989). *Spatial interaction models: Formulations and applications*. Dordrecht: Kluwer.

Gil-Quijano, J., Louail, T., & Hutzler, G. (2012). From biological to urban cells: Lessons from three multilevel agent-based models. In N. Desai, A. Liu, & M. Winikoff (Eds.), *Principles and practice of multi-agent systems* (pp. 620–635). Berlin/Heidelberg: Springer.

Gilbert, N., Troitzsch, K. G. (2011). *Simulation for the social scientist*, 2nd ed. Buckingham: Open University Press.

Grasland, C. (2009). Spatial analysis of social facts. In: F. Bavaud, & C. Mager (Eds.), *Handbook of theoretical and quantitative geography* (pp. 117–174). Lausanne: FGSE.

Grasland, C., & Guérin-Pace, F., 5–9 septembre 2003. Euroluca: A simulation model of euro coins diffusion. In *Proceedings of the 13th European Colloquium on Theoretical and Quantitative Geography* (pp. 24), Lucca.

Grasland, C., Guérin-Pace, F., Le Texier, M., & Garnier, B. (2012). Diffusion of foreign euro coins in France, 2002–2012. *Population and Societies, 488*, 1–4.

Grasland, C., Guérin-Pace, F., & Nuno, J.-C. (2005a). Interaction spatiale et réseaux sociaux. modélisation des déterminants physiques, sociaux et géographiques de la diffusion des pièces euro entre pays de la zone euro. In: 2eme Colloque "Systèmes Complexes en SHS". Paris.

Grasland, C., Guérin-Pace, F., Terrier, C., 2005b. La diffusion spatiale, sociale et temporelle des pièces euros étrangères : un problème complexe. Actes des journées de Méthodologie Statistique.

Grasland, C., Guérin-Pace, F., & Tostain, A. (2002). La circulation des euros, reflet de la mobilité des hommes. *Population et Sociétés, 384*, 4.

Guermond, Y. (2008). From classic models to incremental models. In Y. Guermond (Ed.), *The modeling process in geography: From determinism to complexity* (GIS series, pp. 15–38). London: ISTE and WILEY.

Haken, H. (1977). *Synergetics, an introduction*. New York: Springer.

Hägerstrand, T. (1952). *The propagation of innovation waves. B(4)*. Lund: Royal University of Lund.

Hägerstrand, T. (1970). What about people in regional science? *Regional Science Association Papers, 24*, 7–21.

Huff, D. (1964). Defining and estimating a trading area. *Journal of Marketing, 28*, 34–38.

Kitamura, R. (1984a). A model of daily time allocation to discretionary out-of-home activities and trips. *Transportation Research B, 18*, 255–266.

Kitamura, R. (1984b). Incorporating trip chaining into analysis of destination choice. *Transportation Research B, 18*, 67–81.

Lerman, S. (1979). The use of disaggregate coice models in semi-Markov process models of trip chaining behaviour. *Transportation Science, 13*, 273–291.

Moisil, D. (2002). The Danube limes and the barbaricum (248–498 a.d.). A study in coin circulation. *Histoire & Mesure, 17*(3), 79–120.

Nijkamp, P., & Reggiani, A. (1987). Spatial interaction and discrete choice: Staties and dynamics. In J. Hauer, H. Timmermans, & N. Wrigley (Eds.), *Contemporary developments in quantitative geography*. Dordrecht: Reidel.

Nuno, J., Grasland, C., Blasco, F., Guérin-Pace, F., & Olarrea, J. (2005). How many coins are you carrying in your pocket? *Physica A, 354*, 432–436.

Oberländer-Törnoveanu, E. (2002). La monnaie byzantine des vi-viiieme siècles au-delà de la frontière du bas-danube. entre politique, économie et diffusion culturelle. *Histoire & Mesure, 17*(3), 155–196.

Orcutt, G. (1957). A new type of socio-economic system. *Review of Economics and Statistics, 58*, 773–797.

O'Sullivan, D. (2004). Complexity science and human geography. *Transactions of the Institute of British Geographers, 29*(3), 282–295.

Portugali, J. (2000). *Self-organization and the city*. Berlin: Springer.

Prigogine, I., & Stengers, I. (1979). *La nouvelle alliance*. Paris: Gallimard.

Pumain, D. (1998). Les modèles d'auto-organisation et le changement urbain. *Cahiers de Géographie de Québec, 42*(117), 349–366.

Pumain, D. (2003). Une approche de la complexité en géographie. *Géocarrefour, 78*(1), 25–31.

Pumain, D., Paulus, F., & Vacchiani-Marcuzzo, C. (2009). Innovation cycles and urban dynamics. In D. Lane, D. Pumain, S. Van der Leeuw, & G. West (Eds.), *Complexity perspectives on innovation and social change* (Methodos series, Vol. 7, pp. 237–260). Berlin: Springer.

Pumain, D., Sanders, L., & Saint-Julien, T. (1989). *Villes et auto-organisation*. Paris: Economica.

Quijano, J. G., Piron, M., & Drogoul, A. (2007). Vers une simulation multi-agent de groupes d'individus pour modéliser les mobilités résidentielles intra-urbaines. *Revue internationale de Géomatique/European Journal of GIS and Spatial Analysis, 17*(2), 161–181.

Rappaport, J. (2004). A simple model of city crowdedness. Research Working Paper RWP 04-12, Federal Reserve Bank of Kansas City.

Ravenstein, E. (1889). The laws of migration. *Journal of the Royal Statistical Society of London, 52*(2), 241–305.

Reilly, W. J. (1929). *Methods for the study of retail relationships* (No. 2944). Austin: University of Texas Bulletin.

Reilly, W. (1931). *The law of retail gravitation*. New York: Knickerbocker Press.

Roy, J. R. (2004). *Spatial interaction modelling: A regional science context*. Berlin/New-York: Springer.

Roy, J. R., & Less, P. (1981). On appropriate microstate descriptions in entropy modelling. *Transportation Research B, 15*, 85–96.

Sanders, L. (2007). Objets géographiques et simulation agent, entre thématique et méthodologie. *Revue internationale de Géomatique/European Journal of GIS and Spatial Analysis, 17*(2), 135–160.

Seitz, F., Stoyan, D., & Tödter, K.-H. (2009). *Coin migration within the euro area* (Discussion Paper Series 1: Economic Studies 27). Frankfurt: Deutsche Bundesbank.

Seitz, F., Stoyan, D., & Tödter, K.-H. (2012). Coin migration and seigniorage within the euro area. *Jahrbücher f. Nationalökonomie u. Statistik, 232*(1), 84–92. Deutsche Bundesbank.

Shnerb, N. M., Louzoun, Y., Bettelheim, E., & Solomon, S. (2000). The importance of being discrete: Life always wins on the surface. In *PNAS (Proceedings of the National Academy of Sciences of the United States of America)* (Vol. 97).

Solé, R. V., Gamarra, J. G. P., Ginovart, M., & Lopez, D. (1999). Controlling chaos in ecology: From deterministic to individual-based models. *Bulletin of Mathematical Biology, 61*, 1187–1207.

Stewart, J. (1947). Empirical mathematical rules concerning the distribution and equilibrium of population. *Geographical Review, 37*, 461–486.

Stoyan, D. (2002). Statistical analyses of euro coin mixing. *Mathematical Spectrum, 35*, 50–55.

Stoyan, D., Stoyan, H., & Dödge, G. (2004). Statistical analyses and modelling of the mixing process of euros coins in Germany and Europe. *Australian & New Zealand Journal of Statistics, 46*, 67–77.

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*, 234–240.

Tobler, W. R. (1981). A model of geographical movement. *Geographical Analysis, 13*(1), 1–20.

Tsotselia, M. (2002). Recent Sasanian coin findings on the territory of Georgia. *Histoire & Mesure, 17*(3), 143–153.

Ujes, D. (2002). Coins of the Macedonian kingdom in the interior of Balkans. Their inflow and use in the territory of the Scordisci. *Histoire & Mesure, 17*(3), 7–41.

van Blokland, P., Boot, L., Hiremath, K., Hochstenbach, M., Koole, G., Pop, S., Quant, M., & Wirosoetisno, D. (2002). The euro diffusion project. *Proceedings of the 42nd European Study Group with Industry*.

Wegener, M. (2000). Spatial models and GIS. In A. S. Fotheringham & M. Wegener (Eds.), *Spatial models and GIS: New potential and new models* (pp. 3–20). London: Taylor and Francis.

Weidlich, W. (2003). Socio-dynamics – a systematic approach to mathematical modelling in the social sciences. *Chaos, Solitons and Fractals, 18*, 431–437.

Weidlich, W., & Haag, G. (1988). The migratory equations of motion. In W. Weidlich & G. Haag (Eds.), *Interregional migration* (pp. 21–32). Berlin: Springer.

White, R., Engelen, G., & Uljee, I. (1997). The use of constrained cellular automata for high-resolution modelling of urban land use dynamics. *Environment and Planning B, 24*, 323–343.

White, R. W. (1977). Dynamical central place theory. *Geographical Analysis, 9*, 226–243.

White, R. W. (1978). The simulation of central place dynamics: Two sector systems and the rank size rule. *Geographical Analysis, 10*, 201–208.

Wilson, A. (1967). A statistical theory of spatial distribution models. *Transportation Research, 1*, 253–269.
Wilson, A. (1970). *Entropy in urban and regional modelling*. London: Pion.
Wilson, A. (1981). *Catastrophe theory and bifurcation: Application to urban and regional system*. London: Croom Helm.
Wilson, A. G. (2002). Complex spatial systems: Challenges for modellers. *Mathematical and Computer Modelling, 36*(3), 379–387.

Comprehensive Evaluation of the Regional Environmental and Economic Impacts of Adopting Advanced Technologies for the Treatment of Sewage Sludge in Beijing

Guofeng Zhang, Xiaojing Ma, Jingjing Yan, Jinghua Sha, and Yoshiro Higano

Abstract In recent years, more than one million tons of sewage sludge has been discharged annually in Beijing. The rate of sewage sludge treatment was less than 50% in 2010. Untreated sewage sludge has critically polluted the waterways. The purpose of this paper is to analyze the impact of sewage sludge treatment on the development of the regional economy and on the environment. In this report, we use Lingo software to simulate an economic model and an environmental model with an input-output table and perform a linear optimization of these models. The economic model describes the relationship between economic activities and the emission of water pollutants. The environmental model describes the change in the amount of water pollutants that are generated in the model. Beijing is divided into 11 sub-regions. A comprehensive environmental policy is coupled with the introduction of advanced technology to reduce water pollutants. Based upon the results of the simulation, we can provide detailed information about economic growth, water pollutant reduction, policy subsidies and the number of new sewage and sewage sludge plants needed in each sub-region.

G. Zhang (✉)
School of Economy and Trade, Hebei GEO University, Shijiazhuang, 050031, China

Graduate School of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 3058572, Japan
e-mail: zgfjjgl@hotmail.com

X. Ma
School of Resources, Hebei GEO University, Shijiazhuang 050031, China

J. Yan • J. Sha
School of Humanities and Economic Management, China University of Geosciences, No. 29 Xueyuan Rd, Beijing 100083, China

Y. Higano
School of Economy and Trade, Hebei GEO University, Shijiazhuang 050031, China

J.-C. Thill (ed.), *Spatial Analysis and Location Modeling in Urban and Regional Systems*, Advances in Geographic Information Science,
https://doi.org/10.1007/978-3-642-37896-6_10

Keywords Comprehensive evaluation • Economic model • Environmental model • Sewage sludge • Simulation • Sustainable development • Beijing • Linear optimization

Introduction

Due to its rapid economic and population growth, Beijing's municipal sewage emissions are increasing each year. In 2010, the city produced more than 1.4 billion tons of sewage emissions (Beijing Environmental Protection Bureau 2011). Many sewage treatment plants that have been constructed by the Beijing municipal government have adopted advanced technologies, and the sewage treatment rate increased to 80% in 2010 (Beijing Water Authority 2011). However, the amount of sewage sludge has also increased. Sewage sludge is the byproduct of sewage treatment, and these byproducts pollute the water environment. If this sewage sludge cannot be treated, 50% of the water pollutants removed by sewage treatment will return to the environment (Yang 2010). However, the need for sewage sludge treatment has not been addressed by the government. The rate of sewage sludge treatment was less than 50% in 2010. More than 20,000 tons of total nitrogen is emitted by untreated sewage sludge every year (Zhou 2011), which accounts for approximately 30% of the total net load of total nitrogen in Beijing.

Recently, the government has realized the importance of environmental protection. Accordingly, *The Twelfth Five-Year Plan of Economic and Social Development* (Beijing Municipal Development and Reform Commission 2011) requires that all sewage sludge be treated by 2015 and load of COD (chemical oxygen demand) be reduced by 8.7% in 2015 compared with 2010. Therefore, the government has adopted an integrated policy that promotes water conservation, forestation, reduction of working capital and the introduction of advanced sewage sludge treatment technologies. To determine the optimal development plan for Beijing, it is beneficial to use a simulation method to evaluate the regional environmental and economic impacts of adopting advanced technologies for the treatment of sewage sludge.

The purpose of this paper is to construct a comprehensive simulation model for optimized solutions, in order to evaluate the regional environmental and economic impacts of adopting advanced technologies for the treatment of sewage sludge, and to develop an optimal policy solution to realize the goal of environmentally sustainable economic development.

The spatial implications of an optimal economic and environmental allocation depend on the time span that is allowed for adjustments. In the short term, only a limited set of adjustments may be feasible. The location of labor and firms may be fixed in the short run, making abatement technology a possibility. Moreover, a regionally uniform policy has different impacts under the different conditions in each sub-region (Siebert 1985). In this study, the simulation duration is 10 years, and the government adopts a uniform water pollutant reduction policy for each

sub-region. Under these conditions, one hypothesis is that water pollutants will be transferred among the sub-regions, but the economy of each sub-region is closed. Another hypothesis is that for both the economic and environmental scope, regional differences should be expected by residents of each sub-region.

Methodology and Data

Methodology

Many studies have estimated the impacts of sewage sludge treatment on economic development and the environment. While some of these studies have performed economic evaluations, which focused on analyses of capital investment and operation cost of sewage sludge treatment (Kim and Parker 2008; Shi 2009a, b); those studies, however, ignore the environmental impacts of sewage sludge treatment. A life cycle assessment is one method that can be used to perform a comprehensive evaluation of the technical, economic and environmental aspects of sludge treatment (Murray et al. 2008; Hong et al. 2009; Enrica et al. 2011). However, this method cannot select the optimal sewage sludge treatment technology. The analytic hierarchy process (AHP) method has been used in China to evaluate the technological and economic impacts of sewage sludge treatment (Mao et al. 2010), but the use of expert scoring in this approach is subjective, as different experts may provide differing evaluations. The establishment of an optimization and comprehensive evaluation model via a computer simulation is a suitable method for evaluating water purification policies (Higano and Sawada (1997); Fumiaki and Yoshiro 2000; Mizunoya et al. 2007 and Yan 2010), but this approach has not yet been used in the evaluation of sewage sludge treatment.

In this study, we construct a comprehensive linear programming model that considers all factors influencing water pollutant emission, such as population growth, gross regional product (GRP) growth, types of land use and industry structure. This comprehensive model consists of one objective function (Maximize GRP) and two sub-models (a water pollutant model and an economic model). The economic model describes the relationship between economic activity and the emission of water pollutants. The water pollutant model describes changes in the level of water pollutants generated. The pollutants measured in this study are T-N (total nitrogen), T-P (total phosphorus) and COD (chemical oxygen demand). In the model, Beijing is divided into 11 sub-regions. The simulation duration is from 2011 to 2020. Simulation for this comprehensive model is performed via Lingo software, which is a fast, efficient and straightforward tool for solving linear and nonlinear optimization problems (Li and Higano 2007). By comparing the simulation results for different scenarios, we can estimate the regional environmental and economic impacts of sewage sludge treatment. Based upon the simulation results, we can provide detailed information about economic growth, water pollutant reduction, policy subsidies and the number of new plants needed in each sub-region.

Data

Three types of data are used in this study. One portion includes published data pertaining to population growth, investment, GRP, economic output values, and the amount of sewage, as well as sewage sludge discharge and treatment. The data of population growth, investment, GRP can be obtained from *"Beijing Statistical Yearbook 2011,"* (Beijing Municipal Bureau of Statistics 2012); the data of economic output values from *"Beijing input-output extension table 2010"* (Beijing Municipal Bureau of Statistics 2011); the data of amount of sewage from *"Beijing Water Resources Bulletin2011,"* (Beijing Water Authority 2012); the data of amount of sewage sludge discharge and treatment from *"Beijing Environment Bulletin 2011,"* (Beijing Environmental Protection Bureau 2012). Another portion is composed of survey data, which include detailed information about advanced sewage and sewage sludge treatment technologies. The technical data of sewage treatment and sewage sludge treatment are based on Gaopidian and Huairou sewage treatment plants located in Beijing, and two sewage sludge treatment plants of Dalian city and Jiaxing city which are the model plants for sewage sludge disposal, respectively. All these treatment plants employ advanced sewage or sludge treatment techniques that are commonly used worldwide and, have more effective on pollutant removal performance than any other existing plants in Beijing. The third portion is composed of calculated data based upon the published data, such as the coefficient of water pollutant emissions, the coefficient of discharged sewage pollutant emissions and others.

Model Description

Structure of the Simulation Model

In this study, we draw on the research of Higano and Sawada (1997), Higano and Yoneta (1999), Hirose and Higano (2000), Mizunoya et al. (2007) and Yan (2010). Based upon these previous studies, we constructed a comprehensive linear programming model. Moreover, this study demonstrates improvements over previous water pollutant models by introducing advanced sludge treatment technologies into the model.

The comprehensive model consists of one objective function (Maximize GRP) and two sub-models (a water pollutant model and an economic model) (Fig. 1). The economic model describes the relationship between economic activities and the emission of water pollutants. The environmental model describes changes in the generation of water pollutants.

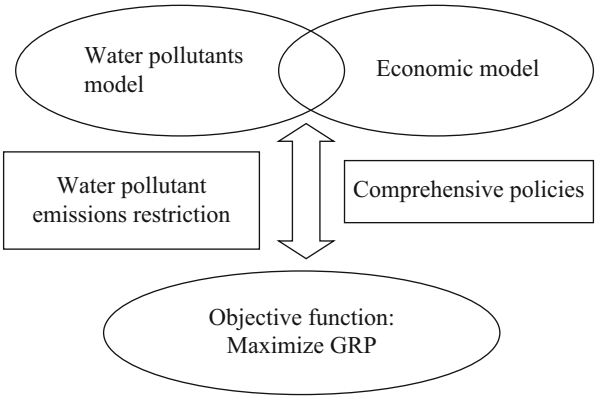

Fig. 1 Composition of the sub-models

We assumed that water pollutants for economic activity flow into the rivers in Beijing (Fig. 2). There are three sources of water pollutant discharge from economic activity: household, industry, and nonpoint sources. Herein, the pollutants contained in rainfall have been examined separately as one part of total water pollutant emissions in the City of Beijing regardless of its amount are very small (Hirose and Higano (2000); Mizunoya et al. 2007; Yan 2010). Consequently, the pollutants emitted via rainfall were excluded from nonpoint sources. A portion of sewage flows into sewage plants through pipes, and sewage sludge is produced during the process of sewage treatment. Finally, the water pollutants contained in the treated and untreated sewage and sewage sludge flow into the rivers. Since it is difficult to collect the water pollutants added by nonpoint sources, we assume that these flow into rivers directly.

In this study we propose an integrated policy to achieve sustainable development, in both the environmental and economic senses (Fig. 3). Currently, the policies used to reduce water pollutants are named "Forestation for water conservation" and "Reduction of working capital". "Forestation for water conservation" policy is that using government subsidies to encourage foresting on vacant land to enhance soil and water conservation. "Reduction of working capital" policy is that the exploitation of decrease of working capital for achieving water pollutant emission reduction by degrading the sectors with high emissions. In order to further reduce pollution of waterways, we propose the construction of new sewage and sewage sludge treatment plants to manage untreated sewage and sewage sludge in this study. Those are the policy of subsidy for reducing capital stock and subsidy for water conservation, which can decrease the pollution from industry and nonpoint source, respectively, and the policy of installation of new sewage and sewage sludge treatment plants, which can contribute to pollution reduction caused by industry and household.

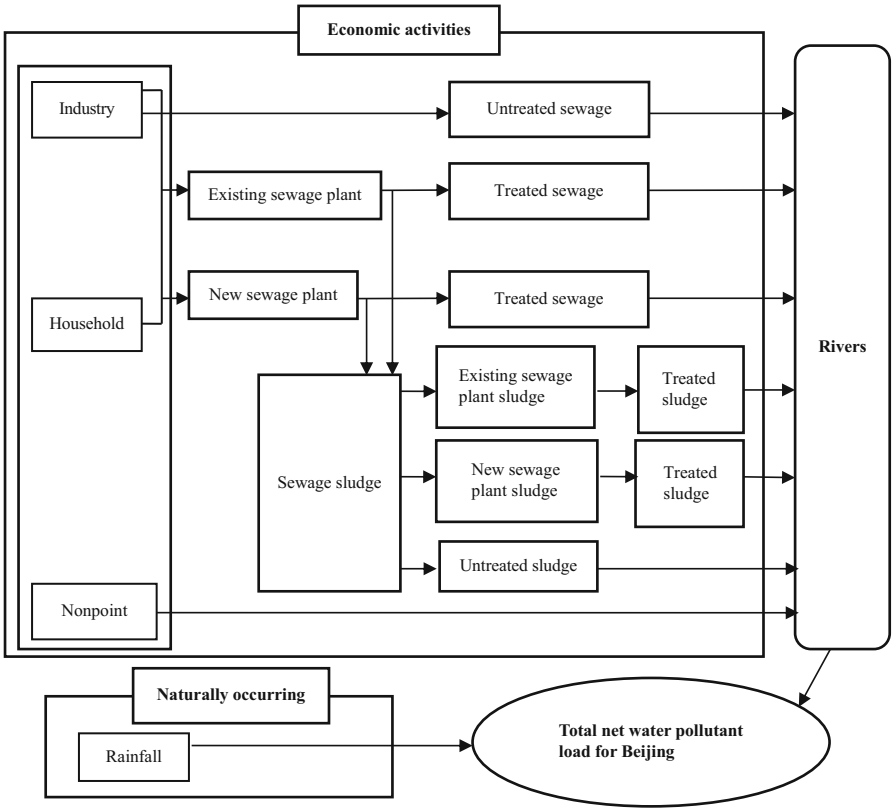

Fig. 2 Framework of the water pollutant model

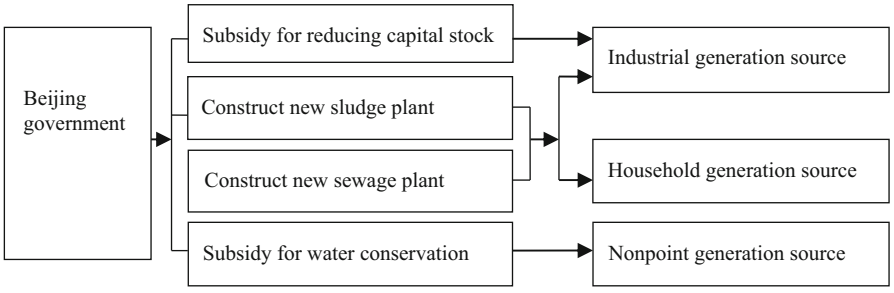

Fig. 3 The framework of the economic model

The policy variables are represented by coefficients in the simulation model, and the environmental and economic impact of different combinations of policies are discussed based upon the simulation results.

Classification of Water Pollutants

T-P, T-N and COD are commonly used in the literature to describe water pollution (Kyou et al. 1998; Chae and Shin 2007; Wang et al. 2008; Jing et al. 2009). Thus, the model in this study includes organic pollution parameters that are commonly used in the literature to describe water pollution. T-P, T-N and COD coded as water pollutant 1, 2 and 3, respectively, in this analysis.

Sub-Regions of Beijing

To facilitate the collection of data and the implementation of policies, in this analysis Beijing is divided into 11 regions based upon administrative divisions. The Central City sub-region of Beijing, which contains six districts, is regarded as one region because some districts, such as Dongcheng and Xicheng, share sewage and sewage sludge treatment facilities with other districts. The other ten districts each represent a separate sub-region (Table 1 and Fig. 4).

Classification of Water Pollutant Sources

The pollutant sources are divided into three categories: nonpoint, household and industry. Load of water pollutants via nonpoint source is decided by land area and the coefficient of water pollutant emissions for different land use. Land use is categorized into four types based upon the "*Beijing year book 2011*" (Beijing

Table 1 Subdivision of the study area

No.	Sub-region	No.	Sub-region
1	Dongcheng	4	Shunyi
1	Xicheng	5	Changping
1	Haidian	6	Daxing
1	Chaoyang	7	Mentougou
1	Shijingshan	8	Huairou
1	Fengtai	9	Pinggu
2	Fangshan	10	Miyun
3	Tongzhou	11	Yanqing

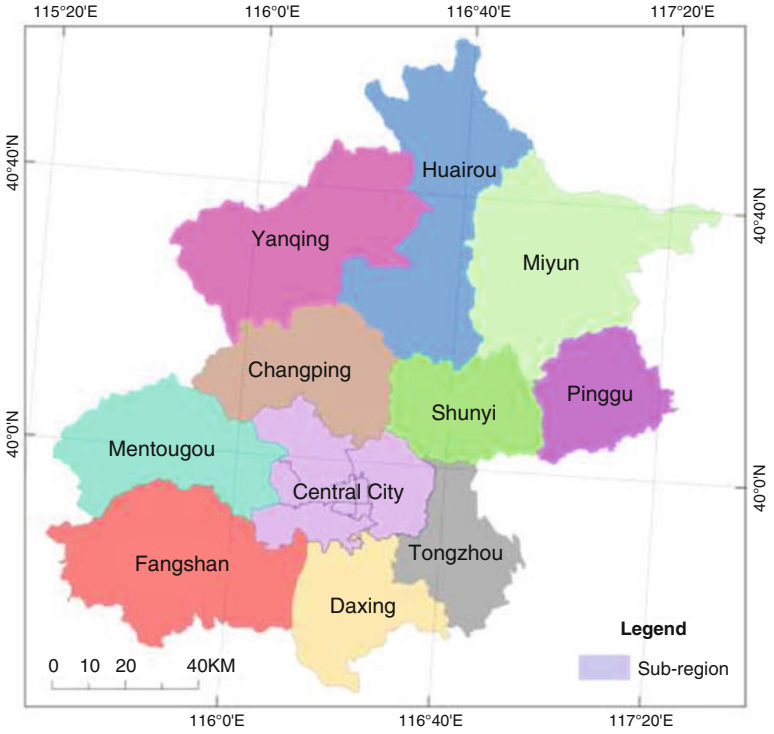

Fig. 4 The map of Beijing administrative division

Table 2 Coding of land use types

No.	Land use
1	Agricultural area
2	Forest
3	Urban area
4	Other land use

Municipal Bureau of Statistics 2012), (Table 2), and industry is divided into fifteen categories based upon *"Beijing input-output extension table 2010"* (Beijing Municipal Bureau of Statistics 2011), (Table 3).

New Technology

We introduce two types of new sewage treatment technology and two types of new sewage sludge treatment technology. Membrane Bio-Reactor (MBR) technology is a popular sewage treatment technology in China, while the High Turbidity Sewage Purification System (HTSPR) technology is a physical and chemical method of

Table 3 Industry classification and coding

No.	Industry
1	Agriculture
2	Forestry
3	Animal husbandry
4	Fishery
5	Minerals mining
6	Processing of petroleum, Coking and processing of nuclear fuel
7	Chemical industry
8	Metallurgical industry
9	Iron and steel industry
10	Manufacture of communication equipment, computers and other electronic equipment
11	Other manufacturing
12	Production and distribution of electric power and heat
13	Construction
14	Transportation, warehousing and postal service
15	Other services

Table 4 New sewage treatment technologies

Technology	Membrane bio-reactor (MBR)	High turbidity sewage purification system (HTSPR)
Investment for one plant (million CNY)	100	110
Operation cost (CNY/)	0.8	0.6
Sewage sludge treatment scale(tons/day)	500,00	10,000
Influent (mg/L)	T-P: 6	T-P: 7.5
	T-N: 65	T-N: 100
	COD: 450	COD:800
Effluent (mg/L)	T-P: 0.5	T-P: 1
	T-N: 15	T-N: 10
	COD: 30	COD: 30

sewage treatment technology. Anaerobic digestion-fluidized bed drying (A-F) is a German sewage treatment technology that has been used in Dalian, a city in Liaoning province. Finally, fluidized bed drying-combustion (F-C) technology, which uses combustion is a sewage treatment technology that has been used in Jiaxing, a city in Zhejiang province; this technology was developed in Japan. Detailed information about these technologies is presented in Tables 4 and 5.

Table 5 New sewage sludge treatment technologies

Technology	Anaerobic digestion-fluidized bed drying (A-F)	Fluidized bed drying-combustion technology (F-C)
Investment for per plant (million CNY)	150	290
Operation cost (CNY/ton)	893	998
Sewage sludge treatment scale(tons/day)	600	1500

Table 6 Scenario composition

Scenarios	Water pollutant reduction rate (in 2020 compared with 2010)	Advanced sewage technology	Advanced sewage sludge technology	Water conservation (forestation) subsidy	Reduction of working capital
Scenario 1	0%	×	×	○	○
Scenario 2	15%	×	×	○	○
Scenario 3	15%	×	○	○	○
Scenario 4	15%	○	○	○	○

Table 7a Maximum allowable amount of water pollutants for 2011–2020 in scenario 1 (in tons)

Years	Maximum allowable amount of T-P	Maximum allowable amount of T-N	Maximum allowable amount of COD
2011	5387	58,548	202,861
2012	5387	58,548	202,861
2013	5387	58,548	202,861
2014	5387	58,548	202,861
2015	5387	58,548	202,861
2016	5387	58,548	202,861
2017	5387	58,548	202,861
2018	5387	58,548	202,861
2019	5387	58,548	202,861
2020	5387	58,548	202,861

Case Setting

This study explores four scenarios (Table 6). The ○ symbol indicates that a policy was adopted, and the × symbol indicates that a policy was not adopted. We defined the water pollutant reduction rate as the percent decrease in the water pollutant emissions level in 2020 compared with 2010. Since government policy requires that the COD emission rate should be reduced by 8.7% in 2015 compared with 2010, we set a target of a 15% reduction in water pollutant emissions in 2020 compared with 2010. Maximum allowable amount of water pollutant from 2011 to 2020 for every scenario is calculated based upon the real data of 2010 (Tables 7a and 7b).

Scenarios 1 and 2 simulate the current situation, in which no advanced technology is adopted. Scenario 3 simulates the implementation of the integrated

Table 7b Maximum allowable amount of water pollutants for 2011–2020 in scenario 2, 3 and 4 (in tons)

Years	Maximum allowable amount of T-P	Maximum allowable amount of T-N	Maximum allowable amount of COD
2011	5301	57,607	199,596
2012	5216	56,679	196,383
2013	5132	55,767	193,221
2014	5049	54,869	190,110
2015	4968	53,985	187,049
2016	4888	53,116	184,038
2017	4810	52,261	181,075
2018	4732	51,420	178,160
2019	4656	50,592	175,291
2020	4581	49,777	172,469

policy approach, which addresses both economic and environmental aspect, with the introduction of advanced sewage sludge treatment technologies. Scenario 4 simulates the implementation of the integrated policy in conjunction with the introduction of both advanced sewage and sewage sludge treatment technologies.

Simulation Model Formulation

Objective Function

The maximization of the objective function places primary importance upon economic activities, are described by the equilibrium solutions of the following structural equation:

$$Max \sum_{t=1}^{10} \frac{1}{(1+\rho)^{(t-1)}} GRP(t) \tag{1}$$

ρ: social discount rate which is a measure used to help guide choices about the value of diverting funds to social projects (exogenous), $\rho = 0.05$;
$GRP(t)$: Gross regional product (endogenous);

The Water Pollutant Model

1. Total water pollutant load for Beijing

$$TQ^p(t) = \sum_{j=1}^{11} WP_j^p(t) \qquad (\mathrm{p} = 1, \mathrm{T-P}; \mathrm{p} = 2, \mathrm{T-N}; \mathrm{p} = 3, \mathrm{COD}) \tag{2}$$

$TQ^p(t)$: total net load of water pollutant p for Beijing at time t (endogenous);
$WP_j^p(t)$: load of water pollutant p in region j at time t (endogenous);

2. The constraints for the total water pollutant load for Beijing

$$TQ^p(t) \leq TQC^p(t) \tag{3}$$

$TQC^p(t)$: the maximum allowable amounts of water pollutant p at time t;

3. Water pollutant load of sub-region

$$WP_j^p(t) = QR_j^p(t) + RQ_j^p(t) \tag{4}$$

$QR_j^p(t)$: load of water pollutant p in rivers at time t (endogenous);
$RQ_j^p(t)$: load of water pollutant p from rainfall at time t (endogenous);

4. Water pollutant flow through rivers

$$QR_j^p(t) = (1 - \nu) \cdot SECQ_j^p(t) \tag{5}$$

ν: river self-purification rate (exogenous);
$SECQ_j^p(t)$: water pollutant p contributed by economic activities in the region j at time t (endogenous);

5. Water pollutant contributed by economic activities

$$SECQ_j^p(t) = HQ_j^p(t) + UIQ_j^p(t) + NQ_j^p(t) - SEQ_j^p(t) - SLQ_j^p(t) \tag{6}$$

$HQ_j^p(t)$: water pollutant p emitted by households in region j at time t (endogenous);
$UIQ_j^p(t)$: water pollutant p emitted by industry in region j at time t (endogenous);
$NQ_j^p(t)$: water pollutant p emitted by nonpoint sources in region j at time t (endogenous);
$SEQ_j^p(t)$: water pollutant p reduced by sewage plants in region j at time t (endogenous);
$SLQ_j^p(t)$: water pollutant p reduced by sewage sludge plants in region j at time t (endogenous);

6. Load of water pollutants from nonpoint sources

$$NQ_j^p(t) = \sum_{g=1}^{4} EL^g \cdot L_j^g(t) \tag{7}$$

EL^g: coefficient of water pollutant p emitted by land use g (exogenous);
$L_j^g(t)$: area of land use g in region j at time t (endogenous);

7. Water pollutants emitted by households

$$HQ_j^p(t) = Z_j(t) \cdot EH^p \tag{8}$$

$Z_j(t)$: number of households in region j at time t (endogenous);
EH^p: emission coefficient of water pollutant p per household (exogenous);

$$Z_j(t+1) = Z_j(t) \cdot (1+\mu) \tag{9}$$

μ: household growth rate (exogenous);

8. Water pollutants emitted by industry

Level of water pollutants for industry is dependent upon production. We describe relationship of production and emission of water pollutant via coefficient of water pollutants emissions of industry.

$$UIQ_j^p(t) = \sum_{m=1}^{15} x_j^m(t) \cdot EUI^m \tag{10}$$

$x_j^m(t)$: production of industry m in region j at time t (endogenous);
EUI^m: emission coefficient of water pollutant p of industry m (exogenous);

9. Water pollutant reduced by sewage plants

Water pollutant reduced by sewage plants has two portions. One portion is reduced by existing sewage plants. Nine types of technology have been use for these sewage plants. Another portion is reduced by new sewage plants, which will use two types of advanced technologies

$$SEQ_j^p(t) = SEQ_j^a(t) + SEQ_j^b(t) \tag{11}$$

$SEQ_j^a(t)$: load of water pollutant p reduced by the existing sewage plants using original technology a in region j at time t (endogenous);
$SEQ_j^b(t)$: load of water pollutant p reduced by the new sewage plants, which use advanced technology b in region j at time t (endogenous);

$$SEQ_j^a(t) = \sum_{a=1}^{9} QSE_j^a(t) \cdot \alpha^a \tag{12}$$

$QSE_j^a(t)$: the amount of sewage treated by existing plants, which use original technology a, in region j at time t (exogenous);

α^a: coefficient of reduction of pollutant p by original sewage technology a (exogenous);

$$SEQ_j^b(t) = \sum_{b=1}^{2} QSE_j^b(t) \cdot \alpha^b \tag{13}$$

$QSE_j^b(t)$: the amount of sewage treated by new sewage plants, which use advanced technology b, in region j at time t (endogenous);
α^b: coefficient of reduction of pollutant p by advanced sewage technology b (exogenous);

10. Water pollutant reduced by sewage sludge plants

Water pollutant reduced by sewage sludge plants has two portions. One portion is reduced by existing sewage sludge plants. Five types of technology have been use for these sewage plants. Another portion is reduced by new sewage sludge plants, which will use two types of advanced technologies.

$$SLQ_j^p(t) = SLQ_j^c(t) + SLQ_j^d(t) \tag{14}$$

$SLQ_j^c(t)$: load of water pollutant p reduced by the existing sewage sludge plants, which use original technology c, in region j at time t (endogenous);
$SLQ_j^d(t)$: load of water pollutant p reduced by the new sewage sludge plants, which use advanced technology d, in region j at time t (endogenous);

$$SLQ_j^c(t) = \sum_{c=1}^{5} ES^c \cdot QSL_j^c(t) \tag{15}$$

ES^c: coefficient of reduction of pollutant p by original sludge technology c (exogenous);
$QSL_j^c(t)$: amount of sewage sludge treated by existing plants, which use original technology c, in region j at time t (exogenous);

$$SLQ_j^d(t) = \sum_{n=1}^{2} ES^d \cdot QSL_j^d(t) \tag{16}$$

ES^d: coefficient of reduction of pollutant p by new sludge technology d (exogenous);
$QSL_j^d(t)$: amount of sewage sludge treated by new plants, which use advanced technology d in region j at time t (endogenous);

11. Load of water pollutants from rainfall

$$RQ_j^p(t) = ER^p(t) \cdot L_j \tag{17}$$

ER^p: emission coefficient of rainfall for pollutant p (exogenous), $ER^1 = 47$ kg/km^2-year, $ER^2 = 1124$ kg/km^2-year, $ER^3 = 2091$ kg/km^2-year (Hirose and Higano (2000); Yan 2010);
L_j: total area of region j (exogenous);

The Economic Model

Policy Treatment for Nonpoint Sources

The Beijing government converted land that was assigned to other land use into forest land, providing subsidies to improve the water quality.

1. Total land area

$$\overline{L}_j(t) = \sum_{g=1}^{4} L_j^g(t) \tag{18}$$

$\overline{L}_j(t)$: total land area in region j at time t (endogenous);
$L_j^g(t)$: land area comprised of land use g in region j at time t (endogenous);

2. Conservation of land to forest land

Beijing government converted the majority of other land use into forest land to improve the water conservation function.

$$L_j^g\,(t+1) = L_j^g(t) + \Delta L_j^g(t) \qquad (\mathrm{g} = 2) \tag{19}$$

$L_j^g\,(t+1)$: area of forest in region j at time $t+1$ (endogenous);
$\Delta L_j^g(t)$: increased area of forest that was converted from other land uses in region j at time t (endogenous);

$$\Delta L_j^g(t) = L_j^{42} \tag{20}$$

L_j^{42}: conversion from other land uses *(g = 4)* to forest *(g = 2)* in region j at time t (endogenous);

$$L_j^{42} \geq \lambda^4 \cdot S_j^{42}(t) \tag{21}$$

λ^4: reciprocal of the subsidy for one unit of conversion to forest (exogenous);
$S_j^{42}(t)$: subsidy for region j given by the Beijing government for conversion of other land use into forest (endogenous);

Policy Treatment for Production Generation Sources

This production function is derived from Harrod-Domar model through the relationship between capital accumulation and production. The production of industry m is restricted by working capital and subsidy for reducing working capital (Yan 2010). Capital accumulation is dependent upon the investment and depreciation of capital.

$$x_j^m(t) \leq \alpha^m \left\{k_j^m(t) - s_j^m(t)\right\} \qquad (m = 1, 2, \ldots, 15) \tag{22}$$

$$k_j^m(t+1) = k_j^m(t) + I_j^m(t+1) - f^m \cdot k_j^m(t) \tag{23}$$

$x_j^m(t)$: production of industry m in region j at time t (endogenous);
$k_j^m(t)$: working capital available for industry m in region j at time t (endogenous);
$s_j^m(t)$: subsidy given by the Beijing government for industry m at time t (endogenous);
α^m: ratio of capital to output in industry m (exogenous);
$I_j^m(t)$: investment in industry m in region j at time t (endogenous);
f^m: rate of capital depreciation of industry m (exogenous);

Policy Treatment for New Sewage Plant Construction

The investment and maintenance cost of new plants is covered by subsidies from the Beijing government.

1. Total amount of sewage discharge

The total amount of sewage is determined by the amount of industrial and household emissions.

$$TQSE_j(t) = \sum_{m=1}^{15} x_j^m(t) \cdot \eta^m + Z_j(t) \cdot \eta^z \tag{24}$$

$TQSE_j(t)$: amount of sewage discharge in region j at time t (endogenous);
$x_j^m(t)$: production of industry m in region j at time t (endogenous);
η^m: sewage emission coefficient for industry m (exogenous);
$Z_j(t)$: number of households in region j at time t (endogenous);
η^z: household sewage emission coefficient (exogenous);

2. Total amount of sewage treatment

The total amount of sewage treatment depends upon two quantities. The first amount is the amount of sewage treated by existing sewage plants, which use original technologies, and the second is the amount of sewage treated by new sewage plants, which use advanced technologies.

$$TQSE_j(t) \geq TQSET_j(t) \tag{25}$$

$$TQSET_j(t) = \sum_{a=1}^{9} QSE_j^a(t) + \sum_{b=1}^{2} QSE_j^b(t) \tag{26}$$

$TQSE_j(t)$: amount of sewage discharge in region j at time t (endogenous);
$TQSET_j(t)$: total amount of sewage treatment in region j at time t (endogenous);
$QSE_j^a(t)$: the amount of sewage treated by existing sewage plants, which use original technology a, in region j at time t (exogenous);
$QSE_j^b(t)$: the amount of sewage treated by new sewage plants, which use advanced technology b, in region j at time t (endogenous);

3. Increase in the amount of sewage treatment

The increase in the amount of sewage treatment depends upon the investment in new plant construction.

$$TQSET_j(t+1) = TQSET_j(t) + \Delta TQSET_j(t) \tag{27}$$

$TQSET_j(t+1)$: total amount of sewage treatment in region j at time $t+1$(endogenous);
$\Delta TQSET_j(t)$: increase in the quantity of sewage treatment in region j at time t (endogenous);

$$\Delta TQSET_j(t) = \sum_{b=1}^{2} \Delta QSE_j^b(t) \tag{28}$$

$\Delta QSE_j^b(t)$: quantity of sewage treatment increased by new sewage plants, which use advanced technology b, in region j at time t (endogenous);

$$QSE_j^b(t+1) = QSE_j^b(t) + \Delta QSE_j^b(t) \tag{29}$$

$$\Delta QSE_j^b(t) \leq \Phi \cdot I_j^b(t) \tag{30}$$

$QSE_j^b(t+1)$: the amount of sewage treated by new sewage plants, which use advanced technology b, in region j at time $t+1$ (endogenous);
$I_j^b(t)$: investment in new sewage plants construction, which use advanced technology b, in region j at time t (endogenous);
Φ: sewage treatment coefficient per unit of investment (exogenous);

4. Maintenance costs of new sewage plants

The maintenance costs of new sewage plants depend upon the amount of sewage treatment by the new plants.

$$MC_j^b(t) = \zeta_j^b \cdot QSE_j^b(t) \tag{31}$$

$MC_j^{\mathbf{b}}(t)$: Maintenance costs of new sewage plants, which use technology *b,* in region *j* at time *t* (endogenous);
ζ_j^b: maintenance costs per ton of sewage for new sewage plants, which use technology *b*, in region *j* (exogenous);

5. Financial and subsidies constraints

$$I_j^b(t) + MC_j^b(t) \leq S_j^b(t) \tag{32}$$

$S_j^b(t)$: subsidies for new sewage plants construction, which use advanced technology *b*, in region *j* by Beijing government at time *t* (endogenous);

Policy Treatment for Sewage Sludge Construction

The investment and maintenance costs of new sewage sludge plants are subsidized by the Beijing government.

1. Total amount of sewage sludge discharge and treatment

The total amount of sewage sludge is determined by the amount of sewage treatment and can be divided into two portions. One portion is the amount of sewage sludge treated by existing sewage sludge plants, which use original technologies, and the other portion is the amount of sewage sludge treated by new sewage sludge plants, which use advanced technologies.

$$TQSL_j(t) \leq TQSET_j(t) \cdot ESE \tag{33}$$

$TQSL_j(t)$: total amount of sewage sludge discharged in region *j* at time *t* (endogenous);
$TQSET_j(t)$: total amount of sewage treatment in region *j* at time *t* (endogenous);
ESE: coefficient of sewage sludge discharged by sewage treatment (exogenous);

$$TQSLT_j(t) \leq TQSL_j(t) \tag{34}$$

$$TQSLT_j(t) = \sum_{c=1}^{5} QSL_j^c(t) + \sum_{d=1}^{2} QSL_j^d(t) \tag{35}$$

$TQSLT_j(t)$: total amount of sewage sludge treatment in region j at time t (endogenous);
$QSL_j^c(t)$: amount of sewage sludge treated by existing sewage sludge plants, which use original technology c, in region j at time t (exogenous);
$QSL_j^d(t)$: amount of sewage sludge treated by new sewage sludge plants, which use advanced technology d, in region j at time t (endogenous);

2. Increase in the amount of sewage sludge treatment

The increase in the amount of sewage sludge treatment is determined by investment in new sewage sludge plant construction.

$$TQSLT_j(t+1) = TQSLT_j(t) + \Delta TQSLT_j(t) \tag{36}$$

$$\Delta TQSLT_j(t) = \sum_{d=1}^{2} \Delta QSL_j^d(t) \tag{37}$$

$TQSLT_j(t+1)$: total amount of sewage sludge treatment in region j at time $t+1$ (endogenous);
$\Delta TQSLT_j(t)$: increase in the amount of sewage sludge treatment in region j at time t (exogenous);
$\Delta QSL_j^d(t)$: increase in the amount of sewage sludge treatment by new sewage sludge plants, which use advanced technology d, in region j at time t (endogenous);

3. Investment in new sewage sludge plants

$$\Delta QSL_j^d(t) \le \lambda_j^d \cdot I_j^d(t) \tag{38}$$

λ_j^d: sewage sludge treatment coefficient per unit of investment for advanced technology d in region j (exogenous);
$I_j^d(t)$: investment in new sewage sludge plants construction, which use advanced technology d, in region j at time t (endogenous);

4. Maintenance costs of new sewage sludge plants

The maintenance cost of new sewage sludge plants depends upon the amount of sewage sludge treatment by these new plants.

$$MC_j^d(t) = \zeta_j^d \cdot QSL_j^d(t) \tag{39}$$

$MC_j^d(t)$: maintenance costs of new sewage sludge plants, which use advanced technology d, in region j at time t (endogenous);
ζ_j^d: maintenance costs per ton of sewage sludge treatment for new sewage sludge plants, which use advanced technology d, in region j (exogenous);

5. Financial and subsidies constraints

$$I_j^d(t) + MC_j^d(t) \leq S_j^d(t) \tag{40}$$

$S_j^d(t)$: subsidies for new sewage sludge plants construction, which use advanced technology d, in region j, from the Beijing government at time t (endogenous);

6. Subsidies constraints of the Beijing government

$$S(t) \geq \sum_{j=1}^{11}\sum_{m=1}^{15} S_j^m(t) + \sum_{j=1}^{11} S_j^{42}(t) + \sum_{j=1}^{11}\sum_{b=1}^{2} S_j^b(t) + \sum_{j=1}^{11}\sum_{d=1}^{2} S_j^d(t) \tag{41}$$

$S(t)$: total subsidy for water pollutant reduction from the Beijing government at time t (exogenous)

7. Transportation costs for sewage sludge

Based upon the technical characteristics and the amount of sewage sludge, we suggest that sewage sludge plants should be constructed in one sub-region. Sewage sludge from other sub-regions will be transported to this sub-region, and the selection of the optimal region for the construction of new plants is based upon the costs of this transportation. The transportation cost is determined by distance, the amount of sewage and the coefficient of transportation cost. Since, new sewage sludge plant will be constructed nearby the place of sewage plant. We ignore the intra-regional transport of sewage sludge in this analysis.

$$COST_i^j(t) = \sum_{j=1}^{11} DIS_i^j QSL_j^d(t) Et \quad \text{j} = 1, 2 \ldots 11; \text{i} = 1, 2 \ldots 11 \tag{42}$$

$$T\_COST_i^j = \sum_{t=1}^{10} COST_i^j(t) \tag{43}$$

$COST_i^j(t)$: total transportation cost of moving sewage sludge from region j to region i at time t (endogenous);
DIS_i^j: distance between sub-region i and sub-region j;
Et: coefficient of transportation cost, $Et = 93$, when distance of transportation equal or less than fifty kilometers; $Et = 163$, when distance of transportation more than fifty and less than one hundred kilometers (Zhang et al. 2006);
$T\_COST_i^j$: total transportation cost of moving sewage sludge from region j to region i over 10 years (endogenous);

Market Balance

The total production of each industry is determined by balances between supply and demand (Mizunoya et al. 2007). The production is dependent on the Leontief input-output coefficient matrix, consumption, investment and net export (Yan 2010). We added variables that were related to the investment in advanced technologies to describe the impact of new plant construction on production.

$$X^m(t) \geq A \cdot X^m(t) + C(t) + i^m(t) + \beta_m^b I^b(t) + \beta_m^d I^d(t) + e^t(t) \tag{44}$$

$$I^b(t) = \sum_{j=1}^{11} \sum_{b=1}^{2} I_j^b(t) \tag{45}$$

$$I^d(t) = \sum_{j=1}^{11} \sum_{d=1}^{2} I_j^d(t) \tag{46}$$

$$X^m(t) = \sum_{j=1}^{11} x_j^m(t) \tag{47}$$

$X(t)$: the column vector of the *mth* element is the total product of industry m in the target area at time t (endogenous);
A: input-output coefficient matrix (exogenous);
$C(t)$: total consumption at time t (endogenous);
$i^m(t)$: total investment in industry m at time t (exogenous);
β_m^b: the column vector of the *mth* coefficient is the production in industry m induced by new sewage plant construction (exogenous);
$I^b(t)$: total investment in new sewage plant construction at time t (exogenous);
β_m^d: the column vector of the *mth* coefficient is the production induced in industry m by new sewage sludge plant construction (exogenous);
$I^d(t)$: total investment in new sewage sludge plant construction at time t (exogenous);
$e^t(t)$: column vector of net export at time t (endogenous);

Gross Regional Product

$$GRP(t) = \sum_{m=1}^{15} v^m \cdot X^m(t) \tag{48}$$

v: the row vector of the *mth* element that is the rate of added value in the *mth* industry (exogenous);

$X(t)$: the column vector of the *mth* element is the total product of industry m in the target area at time t (endogenous);
$v^m \cdot X^m(t)$: economic add value of industry m at time t;

Simulation Results and Discussion

Economic Effects

Objective Function

In this study, we adopted incorporated integrated policies and introduced advanced technologies into analysis and simulation. A feasible solution was reached for every scenario (Fig. 5). The results of the simulation indicate that the introduction of advanced technologies in sewage and sewage sludge treatment is an effective strategy for increasing GRP growth. The highest value of the objective function is obtained in Scenario 4, which incorporated the policies of adopting both advanced sewage and sludge technologies, as well as subsidy for forestation water conservation and reduction of working capital. When only advanced sewage treatment technologies are adopted along with the two subsidies (Scenario 3), there is a decrease in GRP growth of more than 524 billion CNY compared to Scenario 4. Similarly, when advanced technologies are not adopted and only subsidies are implemented (Scenario 2), there is a decrease in GRP growth of more than 3651 billion CNY. Moreover, under Scenario 1, in which there is no reduction in the

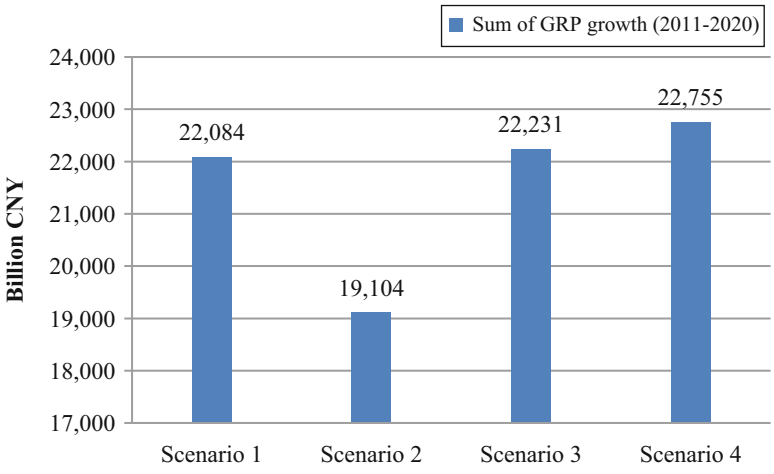

Fig. 5 Sum of GRP growth over 10 years

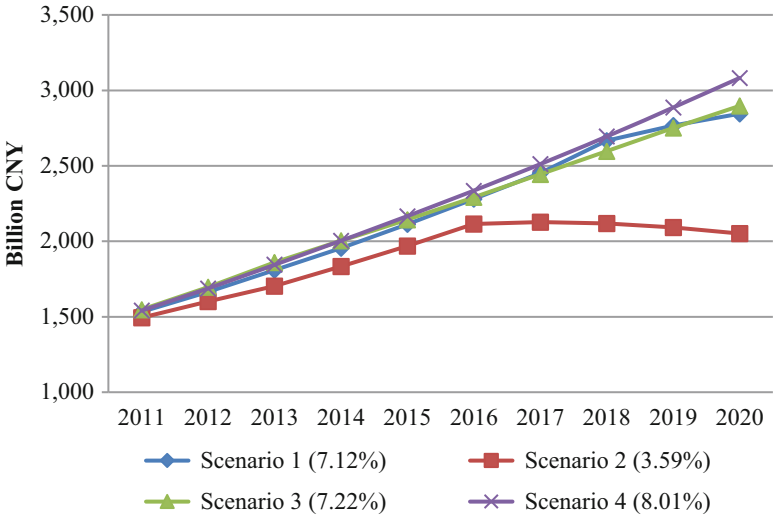

Fig. 6 Average GRP increase in Beijing, 2011–2020, for each simulation scenario

water pollution in 2020 compared with 2010, no adoption of advanced technologies and only subsidies implemented, there is a difference of more than 671 billion CNY.

Change in GRP for Each Scenario

In this simulation, the average rate of the increase in GRP of Beijing for 10 years is 7.12%, 3.59%, 7.22%, and 8.01% in Scenario 1, 2, 3, and 4, respectively (Fig. 6). A comparison of the results from Scenario 1 and Scenario 2 demonstrate that constraints resulting from high levels of water pollutants will limit economic development. Furthermore, the GRP will decrease after 2016 in Scenario 2. However, if we introduce advanced sewage treatment technologies, as in Scenario 3, the GRP growth rate is demonstrated by the simulation to reach a higher level. It should be noted here that the rates for Scenario 1 and 3 do not differ much, but the load of water pollutant in Scenario 1 is much more than it in Scenario 3 (Figs. 8, 9 and 10). However, the increase in the GRP is greatest when both advanced sewage and sewage sludge treatment technologies are adopted, as in Scenario 4. In this scenario, the GRP in 2020 is more than twice that of 2011.

Economic Value Added by Each Industry Sector

The simulation results for Scenario 3 and 4 indicate that the total combined economic value added of all the industry sectors is increased year by year. (Figs. 7a and 7b). However, we note that:

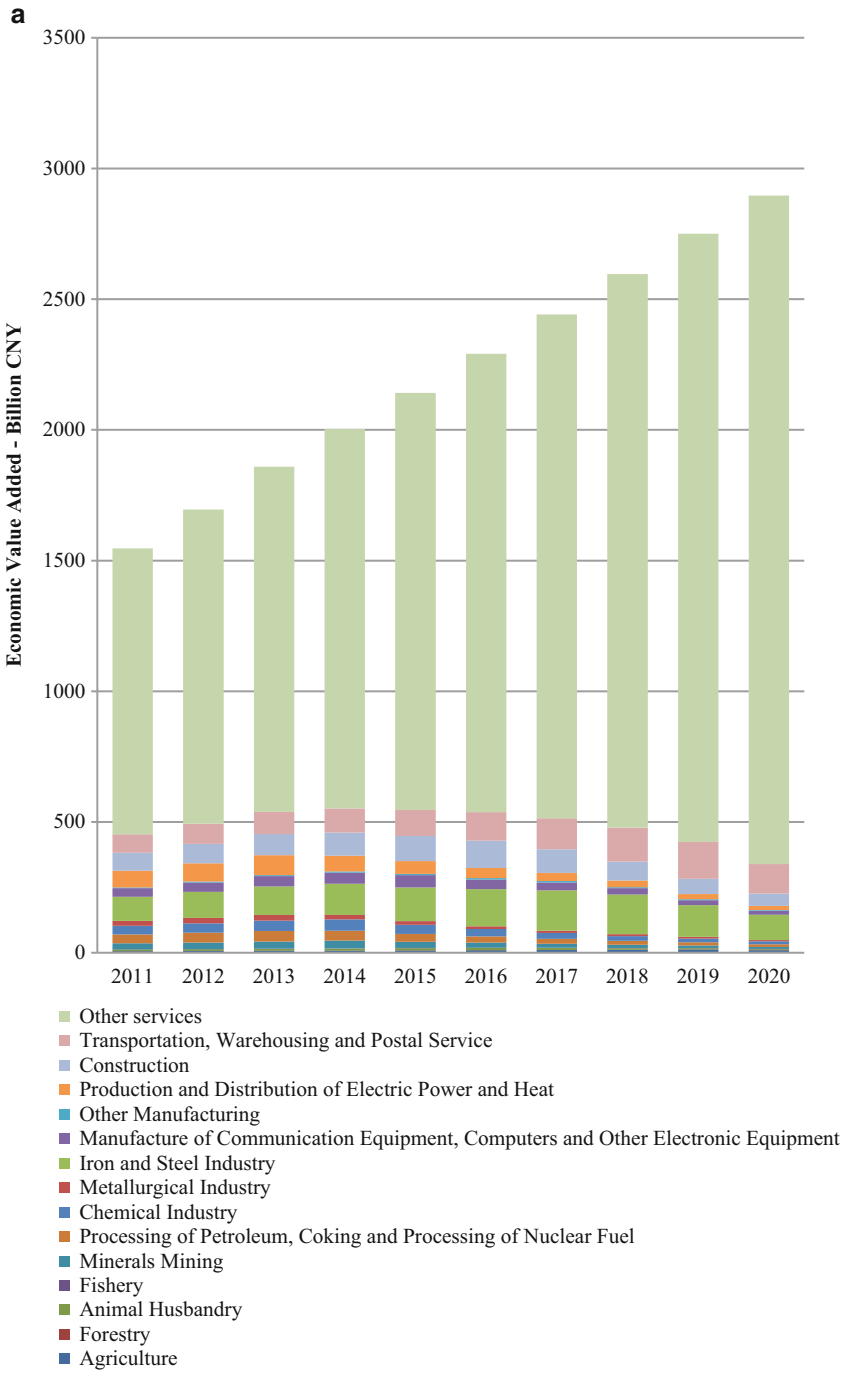

Fig. 7a Economic value added for each industry sector in Scenario 3, 2011–2020

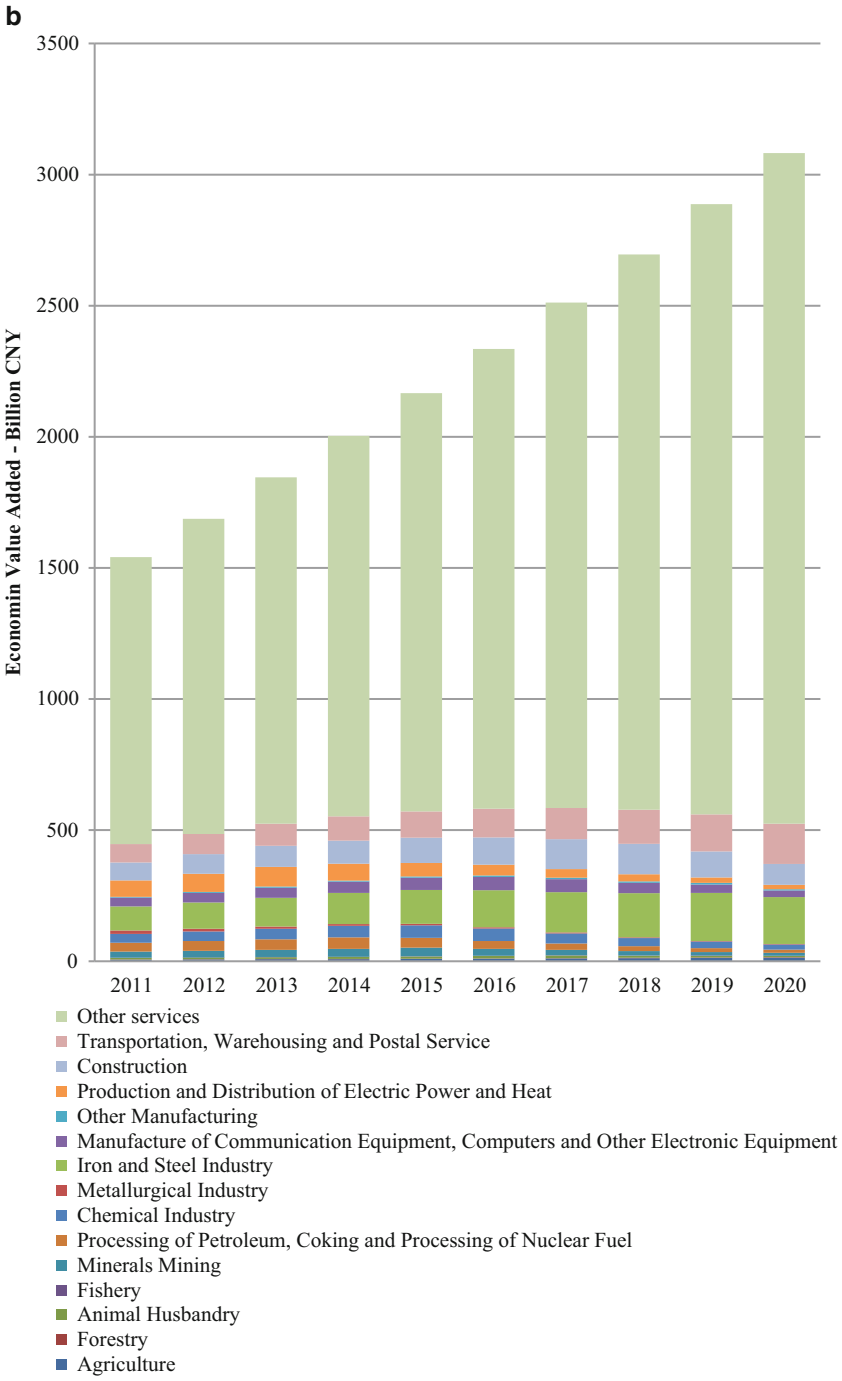

Fig. 7b Economic value added for each industry sector in Scenario 4, 2011–2020

Table 8 Comparison of simulation results with maximum allowable amounts of pollutants in 2020 (in tons)

Scenarios	T-P Maximum allowable amount	T-P Simulation result	T-N Maximum allowable amount	T-N Simulation result	COD Maximum allowable amount	COD Simulation result
Scenario 1	5388	5367	58,548	58,548	202,861	201,606
Scenario 2	4581	4533	49,777	49,777	172,469	142,455
Scenario 3	4581	4571	49,777	49,777	172,469	135,272
Scenario 4	4581	779	49,777	40,471	172,469	172,469

First, outside of the metallurgical industry and the electric and heating power industry, the economic value added of all other industries combined is greater for Scenario 4 than for Scenario 3 in every year. This result, taken in context with the water pollutant emission results presented in Table 8, demonstrates that the level of water pollutants is greatly reduced by the adoption of advanced sewage and sewage sludge treatment technologies, as in Scenario 4. This outcome makes more allowances for the industrial development to emit certain pollutants. However, the development of the metallurgical industry, as well as the electric and heat power industries, is limited in Scenario 4, due to the inefficiency in reduction of COD by sewage sludge treatment technology as compared to sewage treatment technology. The economic value added coefficient for these two industries is lower than that of other industries, while the COD emission coefficient for these two industries is greater than that of other industries.

Second, in Scenario 4, economic value added by the tertiary industry sectors (i.e, transportation and warehousing postal service and other services) is increased year by year, resulting from the low coefficient of water pollutant emission and the high coefficient of economic value added. In contrary, economic value added for the some secondary industry sectors (i.e, minerals mining, processing of petroleum, coking, processing of nuclear fuel, chemical industry, metallurgical industry and production and distribution of electric power and heat power) generally demonstrate a downward trend, due to the low coefficient of economic value added and high coefficient of water pollutant emissions. Yet the other of the secondary industry sectors (i.e, iron and steel industry, manufacture of communication equipment, computers and other electronic equipment and other manufacturing) is increased first and then decreased, owing to the high coefficient of economic value added and high coefficient of water pollutant emissions. The primary industry sectors do not change significantly.

Environmental Effects

Water Pollutant Restrictions and Simulation Results in 2020

The level of water pollutants allowed under the restrictions and the simulation results in 2020 are displayed in Table 8. In this table, it is apparent that, in Scenario 1, 2 and 3, the resulting amount of T-N is equivalent to the maximum level of pollutant allowable under the restrictions, and the resulting amounts of T-P and COD are less than the maximum allowable constrained amounts. However, in Scenario 4, the resulting amount of COD is equivalent to the maximum level of pollutant allowable under the restrictions, and the resulting amounts of T-P and T-N are less than maximum allowable under the restrictions. These results show that the T-N removal efficiencies of the original technologies and the advanced sewage treatment technologies are lower than the removal efficiencies for T-P and COD. In contrast, the COD removal efficiency of advanced sewage sludge treatment technology is lower than the removal efficiencies for T-P and T-N. These findings suggest that the reduction of COD should be considered as the most important factor when adopting both sewage and sewage sludge treatment technologies. The results of this simulation are different from those of previous studies, which did not consider both the economic and environmental impacts of sewage sludge treatment.

Load of Water Pollutant for 2011 to 2020

The resulting total amount of water pollutants for each scenario is presented in Figs. 8, 9 and 10. The largest amount of total water pollutants appears in Scenario 1, in which the water pollutant reduction rate is 0% in 2020 compared with 2010. In scenario 1, the total net loads of T-P, T-N and COD from 2011 to 2020 are 53,832 tons, 583,698 tons and 2,021,044 tons, respectively. In comparing Scenario 2, 3 and 4, the simulation results indicate substantially lower net loads of T-P and T-N form 2011–2020 when sewage sludge treatment technologies are adopted. In Scenario 4, over 10 years, 38,843 and 38,558 fewer tons of T-P is released than in Scenarios 3 and 2, respectively. In Scenario 4, over 10 years, 96,602 and 96,514 fewer tons of T-N is released than in Scenarios 3 and 2, respectively. However, the total net load of COD over 10 years is higher for Scenario 4, by more than 255,733 tons and 198,777 tons as compared to Scenario 3 and 2, respectively. Because the COD removal efficiency of the advanced sewage sludge treatment technology is lower than that of its removal efficiency for T-P and T-N, the removal of COD becomes the objective in Scenario 4.

Subsidy Structure

In this simulation, the Beijing government pays subsidies for water pollutant reduction. Figure 11 presents the subsidy structure. In Scenarios 1 and 2, all

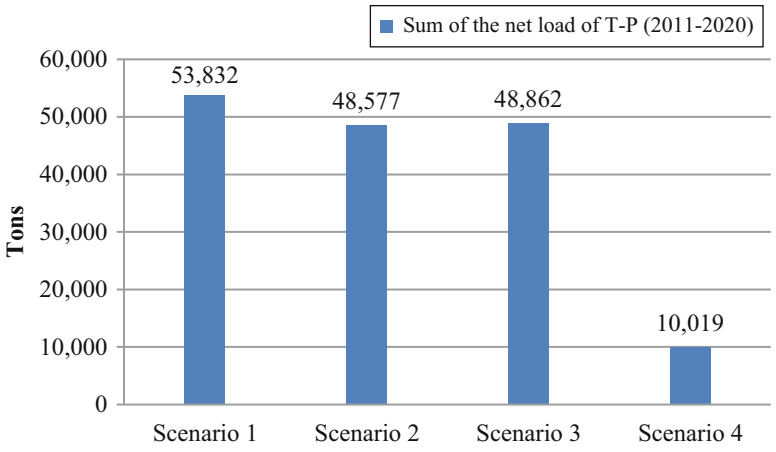

Fig. 8 Total net load of T-P from 2011 to 2020 for each scenario

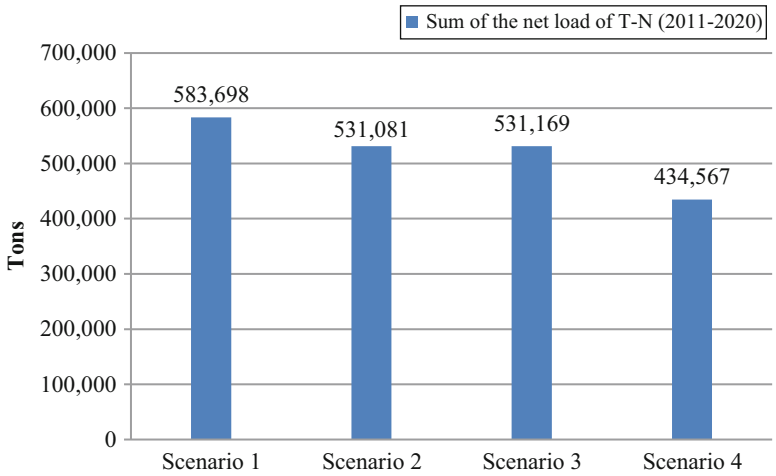

Fig. 9 Total net load of T-N from 2011 to 2020 for each scenario

subsidies are allocated to the policy of reduction of the working capital. The results demonstrate that the water pollutant reduction efficiency of this policy is higher than that of the water conservation by forestation policy. However, the policy of reduction of the working capital is a double-edged sword. As the amount of subsidy granted under this policy increases, GRP decreases. In Scenario 3, in which advanced sewage technology is adopted, the subsidies for water conservation by forestation and the construction of new sewage plants account for 16% and 30% of total subsidies, respectively. Subsidy for the policy of reduction of working capital is 54%, which is 46 percentage points lower in Scenario 3 compared with Scenario 2.

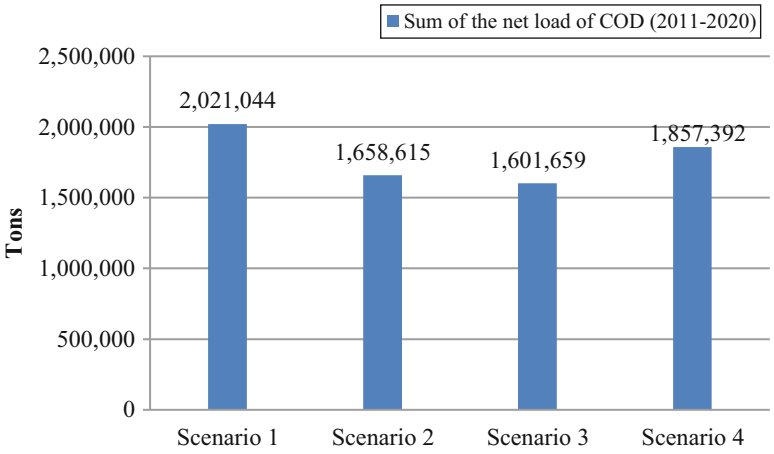

Fig. 10 Total net load of COD from 2011 to 2020 for each scenario

When both advanced sewage and sewage sludge treatment technologies are adopted in Scenario 4, there is no subsidy for the policies of reduction of the working capital and water conservation by forestation. The subsidies for the construction of new sewage and sewage sludge plants account for 24% and 76% of the total subsidy, respectively. This simulation result demonstrates that the construction of new treatment plants is the most efficient tool for water pollutant reduction.

Regional Analysis

To comprehensively consider the impacts upon economic development and the environment, we select Scenario 4 as the optimal scenario. In this section, we will analyze the economic and environmental impacts in each sub-region under the conditions of scenario 4. The model assumes that uniform water pollutant reduction polices are applied throughout the 11 sub-regions by the Beijing government. However, uniform policies do not always lead to identical sub-regional impacts (Siebert 1985).

Sub-regional Economic Trends

Figures 12 and, 13 show the economic trends for each sub-region. Figure 12 demonstrates that the GRP increases annually in each sub-region. In terms of the GRP size of each region, the GRP of the Central City sub-region accounts for more than 70% of total GRP in Beijing each year. This denotes that the Central City sub-region plays an important role in the economic development of Beijing. Figure 13

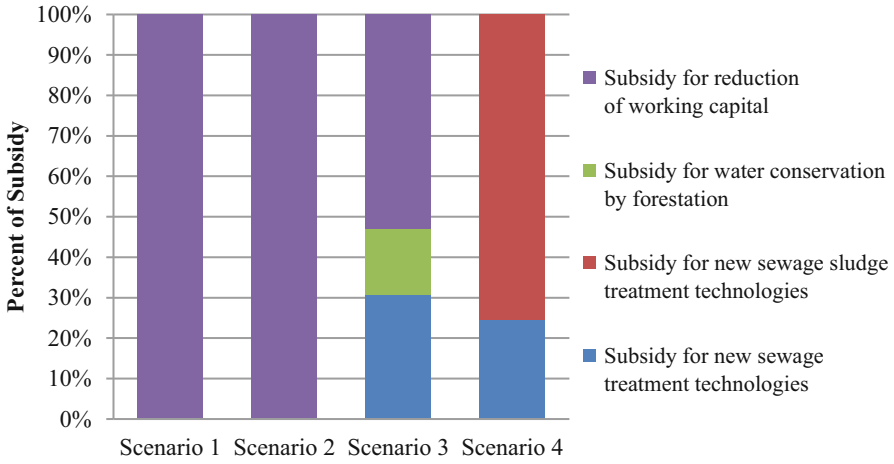

Fig. 11 Allocation of Beijing government subsidy to various water pollutant reduction policies in four scenarios

demonstrates that the rate of economic growth differs for each sub-region. In five of the sub-regions, the annual average GRP grows faster than the average rate in Beijing from 2011 to 2020 (8.01%). In particular, the annual average GRP growth rate of Changping is the highest, at 9.31%.

Environment Intensity Trend

Based upon the analysis above, the COD constraint becomes the target of interest for scenario 4. Therefore, in this section, we select COD intensity as a proxy for environmental impact (Table 9); this metric is an important indicator of environmentally sustainable regional economic development.

This simulation result indicates that COD levels decrease from year to year in each sub-region. Simultaneously, the capacity for sustainable development in each sub-region increases each year. The average COD intensity is related to the efficiency of environmental protection and economic development of each sub-region. However, compared to the average COD intensity of Beijing (82 tons/billion CNY of GRP), only Changping, Central City and Shunyi sub-regions have more efficient environmental protection and economic development. This result also demonstrates that the implementation of a uniform policy does not result in a uniform capacity for sustainable development. The imbalances in capacity can be explained by differences in the initial capital accumulation, population and industry structure and the limitation of labor and firms migration.

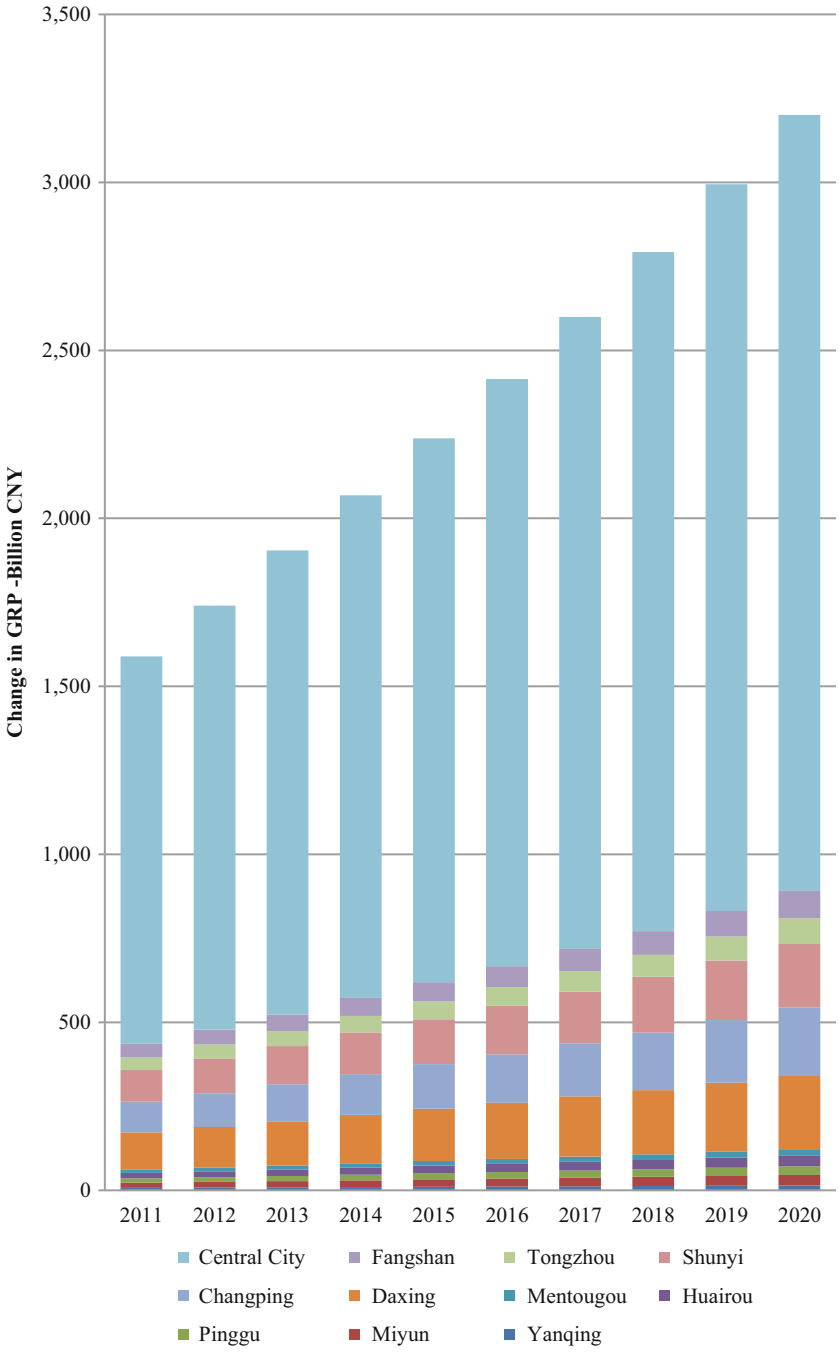

Fig. 12 Change in GRP for each sub-region

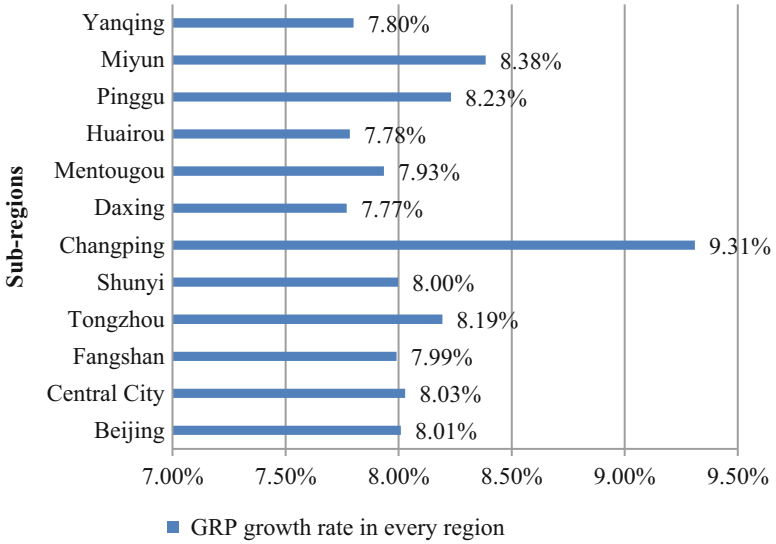

Fig. 13 Rate of GRP growth for each sub-region

Table 9 COD intensity trend from 2011 to 2020 (in tons/billion CNY)

Sub-regions	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Average intensity	Ranking (low to high)
Changping	111	107	103	99	51	47	44	39	36	33	67	1
Central City	107	91	89	81	76	69	65	59	55	51	74	2
Shunyi	174	171	66	62	59	54	50	47	43	40	77	3
Fangshan	170	168	111	104	99	91	85	77	72	67	104	4
Daxing	144	143	138	136	129	121	85	78	62	57	109	5
Huairou	179	176	172	170	160	148	138	127	119	93	148	6
Mentougou	236	228	220	141	135	122	113	105	100	92	149	7
Tongzhou	261	251	242	151	143	131	124	116	111	97	163	8
Pinggu	219	212	206	200	190	147	136	127	120	109	167	9
Yanqing	480	452	198	182	172	160	149	138	129	120	218	10
Miyun	369	349	332	284	262	242	227	209	193	180	265	11

Optimal Regions for New Sewage Sludge Plant Construction

We selected the optimal regions for the construction of new sewage sludge plants based upon transportation costs (Fig. 14). Since new sewage sludge plant will construct in the place of sewage plants, we do not consider intra-regional transport of sludge.

In this simulation, the transportation costs associated with of building new sewage sludge plants in Central City sub-regions is 2965 million CNY, which is lower than those associated with construction of plants in other sub-regions in

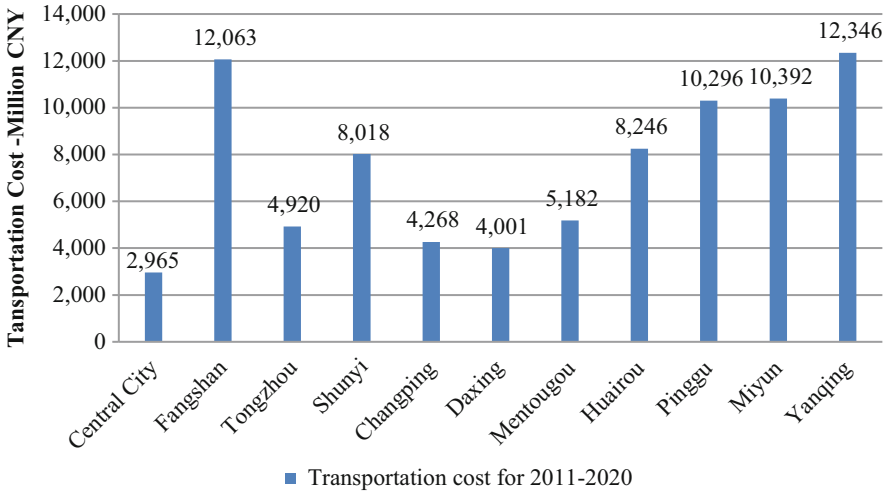

Fig. 14 Transportation costs associated with proposed Central City sub-region sludge plant from the sub-regions of Beijing

Beijing. Therefore, the optimal location for new sewage sludge plants is Central City sub-region.

The Amount of Sewage Sludge Transportation in Sub-regions

Sewage sludge is the source of the water pollutants. Based on the simulation results, sewage sludge discharge in other regions should be treated Central City sub-region. Table 10 show the amount of sewage sludge transportation from the other sub-regions to Central City sub-region. The simulation results predict that approximately 8,084,288 tons of sewage sludge will be transported to Central City sub-region from the other sub-regions over 10 years.

Compensation for Environmental Costs

However, the "polluter pays" principle requires a region to bear the environmental cost that it causes in another region (Siebert 1985). This principle also plays an important role in China. The environmental cost is the damage caused by water pollutants. The compensation for Central City sub-region equals the sum of sewage sludge treatment cost (including investment and maintenance cost) of all sub-regions (Table 11). This simulation result predicts that from 2011 to 2020 approximate 7569 million CNY would be compensated to Central City sub-region by the other sub-regions.

Table 10 Amount of sewage sludge transported to Central City from other sub-regions (in tons)

Sub-regions	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Total
Central city	287,297	349,976	374,820	403,756	421,236	442,543	442,543	453,459	453,459	453,459	4,082,548
Fangshan	50,589	50,589	59,215	60,866	60,866	60,866	60,866	60,866	60,911	61,335	586,969
Tongzhou	50,589	50,589	50,589	63,011	63,011	63,600	63,600	63,600	63,600	65,436	597,625
Shunyi	39,749	39,749	77,465	81,681	81,713	81,713	81,713	81,713	81,713	81,713	728,922
Changping	47,874	47,874	47,874	47,874	65,643	65,643	65,643	65,643	65,643	66,166	585,877
Daxing	50,589	50,589	59,215	60,866	60,866	60,866	60,866	60,866	60,911	50,589	576,223
Mentougou	14,454	14,454	14,454	17,415	17,415	17,572	17,572	17,572	17,572	17,572	166,052
Huairou	20,790	20,790	20,790	20,790	20,790	20,790	20,790	20,790	20,790	20,790	207,900
Pinggu	19,998	19,998	19,998	19,998	19,998	21,937	22,039	22,076	22,076	22,325	210,443
Miyun	16,261	16,261	16,261	18,265	18,454	18,454	18,454	18,454	18,454	18,590	177,908
Yanqing	10,841	10,841	17,427	17,786	17,821	17,821	17,821	17,821	17,821	17,821	163,821
Total	609,031	671,710	758,108	812,308	847,813	871,805	871,907	882,860	882,950	875,796	8,084,288

Table 11 Compensation to Central City sub-region (in million CNY)

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Total
Central city	434	356	352	381	389	410	396	413	406	406	3943
Fangshan	78	45	59	56	54	54	54	54	54	55	563
Tongzhou	78	45	45	65	56	57	57	57	57	60	577
Shunyi	61	36	95	76	73	73	73	73	73	73	706
Changping	73	43	43	43	71	59	59	59	59	60	569
Daxing	33	21	21	21	21	21	46	35	47	42	308
Mentougou	22	13	13	18	16	16	16	16	16	16	162
Huairou	32	19	19	19	19	19	19	19	19	21	205
Pinggu	31	18	18	18	18	21	20	20	20	20	204
Miyun	24	15	15	18	17	17	17	17	17	17	174
Yanqing	16	10	20	16	16	16	16	16	16	16	158
Total	882	621	700	731	750	763	773	779	784	786	7569

Table 12 Subsidy scheme for new sewage plant construction (in million CNY)

Sub-regions	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Total
Central city	82	376	135	163	145	168	116	149	122	122	1577
Fangshan	0	0	26	10	6	6	6	6	6	8	75
Tongzhou	0	0	0	38	8	9	8	8	8	13	92
Shunyi	0	0	115	36	26	25	25	25	25	25	303
Changping	0	0	0	0	54	11	11	11	11	12	110
Daxing	37	5	5	5	5	5	54	14	37	19	184
Mentougou	0	0	0	9	2	2	2	2	2	2	21
Huairou	0	0	0	0	0	0	0	0	0	6	6
Pinggu	0	0	0	0	0	6	1	1	1	2	12
Miyun	0	0	0	6	2	1	1	1	1	2	15
Yanqing	0	0	16	1	0	0	0	0	0	0	17
Total	119	381	297	268	248	233	224	217	213	211	2412

Subsidy Trend for New Sewage Plant Construction

Table 12 presents the optimal subsidy scheme for new sewage plant construction. The subsidy consists of investments in new plant construction and maintenance costs. This simulation result assumes that the subsidy for the Central City sub-region is in the greatest among all the sub-regions; this subsidy accounts for 65% of the total subsidies for sewage plant construction in Beijing. The population and GRP of the Central City sub-regions accounts for 60% and 70% of the total population and GRP of Beijing, respectively, in 2010. Although the rate of sewage treatment reached 95% in 2010, a large amount of new sewage is discharged as the population and GRP continue to grow.

Table 13 The number of new plants to be constructed in each region

Sub-regions	Sewage plants (MBR)	Sewage plants (HTSPR)	Sewage sludge plants (A-F)	Sewage sludge plants (F-C)	Total
Central city	2	3	2	1	8
Fangshan	0	3	–	–	3
Tongzhou	0	12	–	–	12
Shunyi	0	5	–	–	5
Changping	0	7	–	–	7
Daxing	1	1	–	–	2
Mentougou	0	1	–	–	1
Huairou	0	2	–	–	2
Pinggu	0	7	–	–	7
Miyun	0	6	–	–	6
Yanqing	0	2	–	–	2
Total	3	49	2	1	55

The Number of New Sewage and Sewage Sludge Plants to be Constructed in Each Region

The number of new sewage and sewage sludge plants that are proposed to be constructed over 10 years is shown in Table 13. MBR and SPR represent the advanced technologies that will be used in the new sewage plants, while A-F and F-C represent the advanced technologies that will be used in the new sewage sludge plants. In this simulation, it is necessary to construct three new sewage plants featuring the MBR technology and forty-nine plants featuring the HTSPR technology in Beijing. Two sewage sludge plants with the A-F technology and one sewage sludge plant with the F-C technology will need to be constructed in Central City sub-region.

Conclusion

In this study, we create an integrated (economic-environment) model and use a computer simulation to analyze the regional environmental and economic impacts of adopting advanced technologies for the treatment of sewage sludge. Moreover, we discuss the issues of economic development, environmental conservation, environmental allocation and the compensation of each sub-region. The following conclusions are based upon the simulation results.

1. When both advanced sewage and sewage sludge technology are adopted, this policy becomes a very effective tool for reducing environmental pollutants and improving economic development. In the optimal scenario (Scenario 4), the total GRP for 2011–2020 is 22,755 billion CNY. This figure is 3651 billion CNY

more than the total GRP for 2011–2020 under Scenario 2, in which advanced sewage sludge treatment technology is not adopted. In the optimal scenario, the average rate of economic growth from 2011 to 2020 is 8.01%, which is 4 percentage points higher than that of Scenario 2 (3.59%). Moreover, the total net loads of T-P and T-N are 10,019 tons, and 434,567 tons, respectively, in Scenario 4. These numbers are 79%, 18%% lower, respectively, than the net loads for Scenario 2. The total net load of COD is 1,857,392 tons which is 12% higher than in Scenario 2.

2. The optimal budget expenditures for the policy are 2.4 billion CNY for new sewage plant construction and 7.6 billion CNY for new sewage sludge plant construction. The optimal location for new sewage sludge plants is Central City sub-region. Approximately 8,084,288 tons of sewage sludge is transported to Central City sub-region from other sub-regions over 10 years. Thus, the other sub-regions should compensate Central City sub-region approximately 7569 million CNY for 10 years.
3. The pollutant removal efficiency of advanced sewage and sewage sludge technology are different. The COD removal efficiency of advanced sewage technology is higher than that of sewage sludge technology. Thus, COD should be selected as a limiting factor when advanced sewage sludge technology is adopted. This finding differs from those of other studies, which only considered sewage treatment.
4. In the short run, since the location of labor and firms fixed, the adoption of a uniform water pollutant reduction policy by the Beijing government has different impacts on sub-regional economic development and environmental conservation. For example, the capacity for sustainable development in Changing sub-region, Central City sub-region and Shunyi sub-region are much higher than in the other sub-regions. However, in the long run, with labor and firms migration in sub-regions, economic and environmental development of each sub-region will in the same level.
5. In this study, an assumption of the analysis was that the Beijing economic system function as a closed sub-national region, we conducted the model with a short-run 10 years prediction. Our ultimate goal is to reconstruct the model as an open system using a long-run perspective, extending the study area to encompass north China, and eventually a national or global study area. In that case we would consider both the pollution flow and the economic relations among sub-regions.
6. Sewage sludge recycling is an important future extension of our research. Using current technology, power and biogas can be generated during the process of sewage sludge treatment. In further research, this model will be expanded to four sub-models: (1) the energy balance model; (2) the greenhouse gas (GHG) balance model; (3) the water pollutants model; and, (4) the socioeconomic model.
7. This study focuses only on organic pollution measures (T-P, T-N and COD) commonly used in the literature to describe water pollution. As it is important for public health, other toxic waste, such as heavy metal, should be investigated further, though this is outside of the research scope. Additionally, about the problem of the inefficiency in reduction of COD lower than which of T-P and T-N in Scenario 4, introduction of new techniques and application of a mixed policy leading to higher COD emission reduction are two assignments remain in the future work.

Acknowledgements This work was financially supported by The Key Laboratory of Carrying Capacity Assessment for Resource and Environment, Ministry of Land and Resources (Chinese Academy of Land and Resource Economics, China University of Geosciences Beijing) (CCA2012.11), National Natural Science Foundation of China (41101559), The Fundamental Research Funds for the Central Universities.

References

Beijing Environmental Protection Bureau. (2011). *Beijing Environment Bulletin 2010* [EB/OL]: http://www.bjepb.gov.cn/portal0/tab181/. (in Chinese).

Beijing Environmental Protection Bureau. (2012). *Beijing Environment Bulletin 2011* [EB/OL]: http://www.bjepb.gov.cn/portal0/tab181/. (in Chinese).

Beijing Municipal Bureau of Statistics. (2011). *Beijing input-output extension table 2010*. Beijing: BMBS.

Beijing Municipal Bureau of Statistics. (2012). *Beijing statistical yearbook 2011*. Beijing: BMBS.

Beijing Municipal Development and Reform Commission. (2011). The twelfth five-year plan for the national economic and social development of Beijing [EB/OL]: http://www.bjpc.gov.cn/fzgh_1/guihua/12_5/Picture_12_F_Y_P/.

Beijing Water Authority. (2011). *Beijing water resources Bulletin 2010* [EB/OL]: http://www.bjwater.gov.cn/tabid/207/Default.aspx. (in Chinese).

Beijing Water Authority. (2012). *Beijing water resources Bulletin 2011* [EB/OL]: http://www.bjwater.gov.cn/tabid/207/Default.aspx. (in Chinese).

Chae, S. R., & Shin, H. S. (2007). Effect of condensate of food waste (CFW) on nutrient removal and behaviours of intercellular materials in a vertical submerged membrane bioreactor (VSMBR). *Bioresource Technology, 98*(2), 373–379.

Enrica, U., Ivet, F., Jordi, M., & Joan, G. (2011). Technical, economic and environmental assessment of sludge treatment wetlands. *Water Research, 45*, 73–582.

Fumiaki, H., & Yoshiro, H. (2000). A simulation analysis to reduce pollutants from the catchment area of Lake Kasumigaura. *Studies in Regional Science, 30*(1), 47–63.

Higano, Y., & Sawada, T. (1997). The dynamic policy to improve the water quality of Lake Kasumigaura. *Studies in Regional Science, 26*(1), 75–86. (in Japanese).

Higano, Y., & Yoneta, A. (1999). Economic policies to relieve contamination of Lake Kasumigaura. *Studies in Regional Science, 29*(3), 205–218. (in Japanese).

Hirose, F., & Higano, Y. (2000). A simulation analysis to reduce pollutants from the catchment area of Lake Kasumigaura. *Studies in Regional Science, 30*(1), 47–63. (in Japanese).

Hong, J. L., Hong, J. M., Otaki, M., & Jolleit, O. (2009). Environmental and economic life cycle assessment for sewage sludge treatment processes in Japan. *Waste Management, 29*, 696–703.

Jing, D. B., et al. (2009). COD, TN and TP removal of typha wetland vegetation of different structures. *Polish Journal of Environmental Studies, 18*(2), 183–190.

Kim, Y., & Parker, W. (2008). A technical and economic evaluation of the pyrolysis of sewage sludge for the production of bio-oil. *Bioresource Technology, 99*, 1409–1416.

Kyou, H. L., Jong, H., & Tae, J. P. (1998). Simultaneous organic and nutrient removal from municipal wastewater by BSACNR process. *Korean Journal of Chemical Engineering, 15*(1), 9–14.

Li, B., & Higano, Y. (2007). An environmental socioeconomic framework model for adapting to climate change in China. In R. Cooper, K. Donaghy, & G. Hewings (Eds.), *Globalization and regional economic modeling in advances in spatial science* (pp. 327–349). New York: Springer.

Mao, H., Xu, D. Q., & Wang, W. J. (2010). The evaluation model and application research of sludge disposal method in sewage plant. *Environmental science and management, 35*(1), 181–184. (in Chinese).

Mizunoya, T., Sakurai, k., Kobayashi, S., Piao, S. H., & Higano, Y. (2007). A simulation analysis of synthetic environment policy: Effective utilization of biomass resources and reduction of environmental burdens in Kasumigaura basin. *Studies in Regional Science, 36*(2), 355–374.

Murray, A., Horvath, A., & Nelson, K. (2008). Hybrid life-cycle environmental and cost inventory of sewage sludge treatment and end-use scenarios a case study from China. *Environmental Science & Technology, 42*(9), 163–3169.

Shi, J. (2009a). Cost-effectiveness analysis and evaluation on the municipal wastewater sludge treatment and disposal system (I). *Water and Wastewater Engineering, 35*(8), 32–35. (in Chinese).

Shi, J. (2009b). Cost-effectiveness analysis and evaluation on the municipal wastewater sludge treatment and disposal system (II). *Water and Wastewater Engineering, 35*(9), 56–59. (in Chinese).

Siebert, H. (1985). Spatial aspects of environmental economics. In A. V. Kneese & J. L. Sweeney (Eds.), *Handbook of natural resource and energy economics* (pp. 125–164). Amsterdam: Elsevier.

Wang, Y. H., Inamori, R., Kong, H. N., Xu, K. Q., Inamori, Y., Kondoc, T., & Zhang, J. X. (2008). Influence of plant species and wastewater strength on constructed wetland methane emissions and associated microbial populations. *Ecological Engineering, 32*, 22–29.

Yan, J. J. (2010). *Comprehensive evaluation of effective biomass resource utilization and optimal environmental policies in rural areas around Big Cities in China: Case study of Miyun country in Beijing*. Doctoral dissertation for Sustainable Environmental Studies, Graduate School of Life and Environmental Sciences, University of Tsukuba.

Yang, X. P. (2010). Status and ideas of the Beijing municipal sludge disposal. *Water & Wastewater Information, 8*, 17–18. (in Chinese).

Zhang, Y. A., Gao, D., Chen, T. B., Zheng, G. D., & LI, Y. X. (2006). Economical evaluation of different techniques to treatment and dispose sewage sludge in Beijing. *Ecology and Environment, 15*(2), 234–238. (in Chinese).

Zhou, J. (2011). The sludge status and the treatment strategy of Beijing. *The Water Industry Market, 4*, 30–32. (in Chinese).

Part III
Spatial Systems Optimization

Locational Modeling in Spatial Analysis: Development and Maturity of Concepts

M. E. O'Kelly

Abstract This paper provides a selective review of the field of location modeling in spatial analysis. The review is written on the basis of 30 years of experience in the area of spatial analysis, as well as an awareness of ideas in location theory emerging from recent conferences such as ISOLDE, RSAI, and INFORMS. The work synthesizes opinions and impressions from conference programs, informal conversations with colleagues at various stages of their careers, and general awareness of disciplinary norms. The paper points to several characteristics of a field in maturity. It also suggests that current ideas are, in many cases, quite well foreshadowed in the basic foundational literature; yet, at the same time, there are two kinds of emerging advances. One is a number of mathematical breakthroughs that allow an improved approach to previously difficult problems, and the other is a set of new problems that were unanticipated in the previous literature. A key attribute of maturity is that techniques from the field have resulted in real world locational decisions. The paper is illustrated with examples from location theory and applied mathematical modeling. This is both a backward- and forward-looking appreciation for some ideas that have matured and are ready, in some cases, for new mathematical and computational tools.

Keywords Location theory • Spatial analysis • Development of locational analysis

Introduction and Prior Work

Smith et al. (2009) have recently reviewed the *growth to maturity* of locational analysis, so it may seem too soon to revisit some of the issues in the field. Nevertheless, this chapter discusses the maturity of location theory with a special emphasis on

M.E. O'Kelly (✉)
Department of Geography, The Ohio State University, Columbus, OH, USA
e-mail: okelly.1@osu.edu

J.-C. Thill (ed.), *Spatial Analysis and Location Modeling in Urban and Regional Systems*, Advances in Geographic Information Science,
https://doi.org/10.1007/978-3-642-37896-6_11

its role in urban and regional analysis.[1] Generally, in this paper, location theory will be used to refer to a broad range of concepts and ideas relating to the location of activities or facilities. Location models are the mathematical constructs used in many cases to encode these abstract ideas. While by no means an uncomplicated use of terms, locational analysis is used as a rubric for ideas that overarch theory and models (see Webber 1984 on explanation, prediction, and planning). The classical concerns of location theory are often associated with the canonical works of Losch and Christaller on central places, von Thunen on agricultural land use allocation, and Weber (and later Hoover) on industrial location (see Griffith et al. 2013). As of the mid- 1980s the tendency to ground locational analysis in these classics declined a little, in favor of comparatively sterile optimization with less of the rich economic reasoning that underpinned the early works. What, if anything, does the mature and more practical field need to pull from the past? The reason this question is relevant is that there is a generation of operational and mathematical modelers in the field of locational analysis that has very little connection to the themes of the early theorists. There is also a sense that maturity connotes a field past its prime, and hence in the waning phase. That view is particularly strong among those who have moved away from mathematical models. In this paper, some arguments will be brought forth to suggest that the field is in fact poised to add a new level of integration between ideas from theory and practice, and to do so with more sophisticated tools. While the scope of this introductory remark is very broad, the paper focuses on a range of specific examples of how the field can build on its foundations. The paper draws on 30-, 40-, and even 50-year retrospectives.

Retrospective

With this background in mind, note that location theory was already well established in the last half of the twentieth century and was clearly a major component of what Scott (2000) called a great half-century of progress in spatial analysis. Lea (1973) provided a comprehensive annotated bibliography of the state of the art at that time, and given the rapid development of the field, it is quite difficult to imagine a similarly comprehensive overview today. Rather, there are some specialized slices of this literature, notably the useful review paper on spatial optimization by Tong and Murray (2012). Mulligan et al. (2012) have also reviewed the re-emergence of locational ideas arising from a core in central place theory in a review for the *Annals of Regional Science*. Fujita (2010) has given a very laudatory review of von Thunen's many contributions to economics and O'Kelly and Bryan (1996) provide a geographic assessment of Thunen's fundamental spatial insights. These are fields that are not quite the same as the ones reviewed in Smith et al. (2009), though

[1] Another related paper reports on the 51st ERSA Conference in Barcelona. The presence of location analysis in regional science is noted (Royuela 2012).

there are some obvious overlaps, and indeed part of the point of this paper is to stress the role of locational analysis in a wider array of fields, including urban and regional modeling. Several core concepts pervade the field and these themes reappear throughout this paper. The paper argues that the spatial economic view contributes rich insights to locational analysis, and offers promising ideas that can steer operational and mathematical tools in improved directions.

A great influence in this march to maturity was the kind of careful synthesis and panorama provided by scholars in urban modeling and planning. As ably illustrated by Batty (2008) in *Fifty Years of Urban Modeling: Macro-Statics to Micro-Dynamics,*[2] there have been enormous advances in spatial analysis, and there are now connections between urban land use models and many aspects of transport and movement. Taking the added step to incorporate land use planning and optimization in the same scope of work, these fields are seen as highly relevant to each other. Analysts such as Lew Hopkins and Brit Harris for example were keenly aware of the ideas that are at the foundation of current problems in spatial optimization. Many of these models have since assimilated the refinements that come with advances in applied mathematical programming. A typical trajectory takes us from enumeration, to standard optimization, to customized tools. Several introductory texts on Integer Programming use the plant location problem to illustrate, for example, lagrangean relaxation and dual ascent (Nemhauser and Wolsey 1988). The point of this illustration is that someone coming to the field today should consider the pathway to the current state-of-the-art built from earlier stages of contributions. This perspective can also provide a promising pathway to areas of future growth linking urban and regional modeling and location optimization. Newly emerging scholars should appreciate the foundations for some of the third or even fourth generation advances.

Current Status

In locational analysis, there have been many in-depth topical reviews that successfully parse the field by drawing distinctions between planar and discrete models, deterministic vs. stochastic models, and so on. The end members of these various continua have also been thoroughly treated in textbooks, such as Daskin (1995) on facility location with a particular emphasis on logistics and supply chains. Chen (2010) comes close to the theme of this paper by looking at a broad array of spatial modeling applications. There are also books and papers which have chosen to explore one end of a particular continuum (network location: Handler and Mirchandani 1979).

[2]*Centre for Advanced Spatial Analysis (CASA), University College London, UK; "presents chronological and conceptual history of urban land use-transportation models movement in the context of current developments."*

From these texts it is clear that locational analysis has contributed some important breakthroughs. As stated above, there are sustained rounds of work on some core problems, with notable achievements as a result. The people providing such insights are justifiably lauded as pioneers for having provided unusually deep insights – including Goldman, Hakimi, Koopmans and Beckman, and such scholars as Drezner and Wesolowsky for their flair in applied mathematical models (for examples and references see Drezner 1995; Drezner and Hamacher 2002). Richard Church also has recently been recognized at the North American Meetings of the Regional Science Association International (RSAI) with a well-received Fellow's Lecture (RSAI, Miami, 2011) that described the emergence of powerful locational insights into the set cover problem, many of which derived from a spatial analysis perspective. O'Kelly's plenary at the 2013 *Spatial Analysis and Methods* group at the Association of American Geographers (AAG) meeting provided a guide to hub location models. (This book itself provides current views of selected applied locational models.)

Selected Attributes of a Mature Field

As a mature field, location theory has the following attributes (among others):

1. There is an emerging sense of the *stages of growth* and development of the field, devoting persistent attention to some key topics, and supported by retrospective reviews. The field has emerged from earlier approaches and demonstrates increasingly refined attacks on some core problems.
2. There exists a system of *assessing new contributions*. A good review system is very important: i.e., how new material is screened, selected, generated, and accepted.
3. The field has had an impact on technology and on practical *location and network decisions*.
4. The field demonstrates fruitful and *active collaborations* with aspatial disciplines, such as applied mathematics and computer science.

These ideas are discussed in the following subsections.

A Brief Review of the Stages of Location Models

Many methodological fields move through a developmental cycle from youth – where new ideas are adopted and new problems emerge from a practical or theoretical concern – through maturity and potentially old age. This metaphor for cyclic change is well worn, but in this paper, it is argued that the notion of maturity – indicating ideas that have been developed and tested – is indeed apposite for the field of location theory. In this field, applied mathematical ideas are tested and expanded

upon and provide the drive for further expansion of the state of the art, and of the realm of possible practical applications. Computational and theoretical advances make larger practical instances attainable. The following sections, discuss both the retrospective view and the encouraging signs of current activity.

It should be clear from the introductory remarks that maturity is viewed as a positive attribute and not one that connotes a field that is past its prime. This is especially clear when one examines the increasing depth of field that has come from persistent attention to key topics. It is quite revealing to look for problems that have been around for some time. The *Papers in Regional Science,* for example, provided a CD with contents from 1955–2004. A brief content analysis for papers with the word "location" in their title shows a persistent concern with this topic. The early years of that journal show the presence of topics which have become rather classical concerns for the field and still garner a good bit of attention. Beckmann (1955) offered "Some Reflections on Losch's Theory of Location," and the first few years of publication includes papers by Tiebout, Henderson, Goldman, Stevens, Kain, and Lowry, to name just a few who have become central figures in all aspects of location theory. Over time, there was a notable increase in papers of this type beginning in 1991 (to an average of 4 per year). By the early 2000s, the rate of papers had not abated, and themes now show an increasing concern with agglomeration, innovations, and hubs, among other topics. Similar stock-taking has occurred for other journals (Geographical Analysis, see Griffith et al. 2013; and Journal of Regional Science, see Duranton 2010).

Certainly there have been facets and aspects of locational analysis that have been left behind, as the field has moved on, though Mulligan et al. (2012) argue that some of these ideas are ripe for reemergence. For example, the graphical packing ideas of Christaller and Losch are now better accommodated in Geographic Information Systems (GIS), and in such systems, there is no need to confine the analysis to an *isotropic plain* that was a feature of early models. In fact, non-uniform spatial distributions are the *raison d'etre* for locational analysis, and overcoming spatial separation between resource sites is a fundamental task for economic geography. All this is to say that the earliest forms of graphical spatial models have been eclipsed and are replaced by a rejuvenated set of tools. The result may be a new cycle starting from remnants of the older stages. (See Sui et al. 2009 for a detailed treatment of geospatial technology in location.)

It is clear from consultation with colleagues at conferences that there is a need on the part of some experienced researchers to reaffirm the macro contributions of the field of location theory, especially in the context of urban and regional analysis. There is also a tendency to accept core concepts as either so central that their origins can be elided, or so 'obvious' that normal inquisitive exploration has all but stopped. This myopia can lead to certain ideas being reinvented.[3]

[3]I am especially grateful to Gordon Mulligan for discussions on this point.

Assessing New Contributions Via Journals, Grants, and Patents: Different Paths to Innovation

One would expect a mature field to have a fairly elaborate system of gatekeepers and peer review. It takes a certain kind of imaginative foresight to gauge the potential worth of an idea through the panel review of a proposal; the peer review stage of the journal; or the assessment of a patent application. Each of these three cases, although seemingly unrelated, presents a slightly different problem.

(A) A research proposal is written prospectively for an activity to be carried out that may or may not produce transformative results. There is in fact a fairly vigorous debate as to whether a research proposal should be gauged on the basis of its probable success, but in the light of pressure to make wise use of public funds, agencies such as the National Science Foundation and the European Science Foundation have well documented criteria designed to gauge "broader impact." In some cases, this could be seen as placing a weight on the likelihood of a research project delivering practical or applied results. Location theory performs well on this dimension.

(B) The journal paper review comes from a slightly different point of view. An article being reviewed is the result of an intensive period of research and reflects the results of much research activity. The paper is judged during peer review on the basis of its contribution to existing literature.

(C) Yet a third juncture at which the location theorist might run into a form of assessment is in the area of patent application. A brief detour into patent terminology might be instructive here: the patent is centrally defined by its claims, and the US Patents and Trademarks Office (USPTO) examiner reviewing those claims looks for the work to be non-obvious and novel. The review expects to the patent to disclose some distinctive new methods or a new system, in contrast to the existing body of knowledge (the so-called prior art). Developing an acceptable version of the patent is referred to as patent prosecution and the *prosecution history* might read something like the kind of back and forth from a journal editor or a review panel to an investigator. The examiner (from the USPTO perspective) requires that the patent applicant hone the ideas to a set of allowable claims. Usually, these are refined in the process to be a bit narrower than the original claims, as the USPTO has a strong interest in placing tight boundaries around the granted patents which become the basis for a grant of a short-term monopoly on these ideas.

Impact on Real System Technology and Practical Tools

Research accomplishments detailed in journals and books, patents, and a number of demonstrable location theories have resulted from core scientific funding. Patented

technology is fundamental to many current businesses in the area of location-based services. One recent example (Bailey and Huber 2012) makes extensive reference to GIS and location and transport concepts in connection with the determination of trade areas. The references at the beginning of the patent contain numerous citations to well-known geographic and location authorities. Many businesses are built on a number of key technology areas, including spatial data handling, tiling and relational structures in spatial databases, and the locational awareness of devices such as cell phones. While a very narrow conception of location theory might miss this fact, these creative endeavors use fundamental spatial and locational ideas in very broad ways – from packing and arranging data, to arranging the organization of spatial data bases, to the use of spatial (database) keys for the purpose of rapid index of locations. In the broad sense, the spatial sciences community has contributed to the mature technology that is at the foundation of location-based services. It is important to note that the ideas have come from a range of aspatial fields, including computer science and telecommunications.

ReVelle (1993) demonstrated that there are integer-friendly location problems and there has been a major emphasis in the literature on developing transformations of new problems to the underlying or fundamental (canonical) versions of some core location models. One would be the plant location problem (Erlenkotter 1978) and another, the Generalized Assignment Problem (Ross and Soland 1975). More recent mathematical innovations are rapidly absorbed and deployed in the area of locational analysis (decomposition, ordered medians, a difference between two convex functions, and the entire gamut of numerical exploration tools).

Another motivation has been the emergence of an evolved form of GIS in location.. As noted by Church (2002), the field of GIS "*has evolved into a mature research and application area involving a number of academic fields including Geography, Civil Engineering, Computer Science, Land Use Planning, and Environmental Science. GIS can support a wide range of spatial queries that can be used to support location studies. GIS will play a significant role in future location model development and application.*" Murray (2010) echoes this view and goes further to explain that GIS is contributing important advances in this field. Specific examples include transport (Thill 2001), multi-criteria decision making (Thill 2000), and spatial covering (O'Kelly and Murray 2004).

A partial list of the attributes of applied problems is that they tend to be for a real client, at a larger scale, and requiring realistic parameters. Because the client frequently modifies the constraints or goals, the techniques used for solution need to be robust to allow modification. Efficient, high-speed, and flexible exploration of a large number of solutions is also a primary concern. The output might contain maps and locational recommendations, and the ability to tie the visualization of the solutions to GIS. On the other hand, more mathematical models are ideally suited as a test bed for mathematical techniques (e.g. integer programming), are sometimes illustrated first with small test cases, and then larger randomly generated data sets. Sometimes these random data sets succeed in establishing the maximal size of the problem, but they may lack some convenient exploitable spatial structure. The main

concern in such models is the development of high quality solutions (measured for example by lower bounds), and perhaps gaining insight into problem structure. The output from such efforts is often an elegant characterization of the problem, with definite ideas on the suitability of the algorithm to problems of various size and characterization.

One area with which this author is very familiar, and where both applied and algorithmic advances have been notable is in the domain of hub networks. Space does not permit a complete review here (see Campbell and O'Kelly 2012), but one example from each side will make the point. The French postal system (La Poste) has been the subject of substantial facility (hub) location and network analysis and optimization. Winkenbach et al. 2012, and a presentation at the *European Working Group on Locational Analysis* (EWGLA) by one of the co-authors of that report (Lemarié 2011) summarized the work. Typically, the problem demonstrates all the points above: a very large number of variables and parameters, and a need for adaptable solutions. In the other direction, algorithms for the hub problem have emphasized particular protocols, the relative difficulties in solving special cases, and the desire to characterize the solution space in fairly precise ways. An exemplary line of work demonstrating this approach is by Contreras et al. (2012).

Spatial and Aspatial Scientific Collaboration

Mathematical and algorithmic methods are also powerful tools within the field of location theory (Tong and Murray 2012). The spatial view helps to develop extensions and modification of algorithms. Some geographers imagine or devise a novel network approach to a problem without worrying too much about practical solvability, or maybe even realizing that this is a non-standard case. There may be merit in taking a problem up as a spatial challenge, rather than starting from the prescribed canonical methods. Instead of tools in search of a problem (e.g. linear programming), geographers are perhaps in the advantageous position of having *spatial* ideas first and then seeking a means to tackle the problem.

Creative model formulations do not always suggest immediate insights on how to solve the problem. Many times the initial model formulation has been attacked by rather direct enumeration, which may later be superseded by tighter model formulations. Naïve model formulations provide fertile ground for collaborators to improve the picture. This has happened in numerous ways in the history of geographers working in location theory (e.g. Scott on plant location; Holmes et al. 1972 on medians; and, undoubtedly, in many other areas where fast computations have eclipsed initial techniques). New techniques also come into play in that a problem that is initially difficult becomes more amenable to solution with the evolution of improved methods. It would be foolish to think that the current methodological toolkit is the final chapter.

Geography is quite central in the location analysis framework, along with collaborators and other colleagues from a variety of disciplines (notably Operations

Research, Regional Science, Management Science, etc.). It is also clear from collaboration with computer science and applied math that geographers can bring special insight to network problems.[4]

Activity in a Mature Field

As mentioned above, a field that is mature will already have had several attempts to take stock and synthesize. As well as the many excellent overall literature reviews, this author has recently participated in two papers that examine aspects of the field that are certainly relevant to the claim of maturity. Indeed, writing parts of those papers is partly what prompted this re-examination of the issue (Campbell and O'Kelly 2012; Griffith et al. 2013). As will be illustrated in various ways in the following sections, location theory is also a very active area. Three rather holistic measures of this activity level are given next: (i) active clusters of work presented at conferences; (ii) technical breakthroughs, typically developed as primary research; and, (iii) standard test data sets.

Conference Activity

Participants from multiple disciplines meet regularly and interact at clusters at conferences and in special sub-sections or study groups. The *International Symposium on Locational Decisions* (ISOLDE) and the *Triennial Symposium on Transportation Analysis* (TRISTAN) are examples of specialized conferences. The *Section On Location Analysis* (SOLA) in INFORMS is a study group, and the track of locations papers at the *Regional Science Association International* (RSAI) is a specialized sub-track within a conference. Indeed, it is now relatively easy to pursue specialized interests at INFORMS and RSAI in location theory, with this being somewhat more hit-and-miss at the *Association of American Geographers* (AAG), and even less central in very large general purpose conferences such as the *Transportation Research Board* (TRB). One other remarkable trend is the cloning of special interest groups around a certain regional cluster or thematic cluster; for example the European Working Group on Location Analysis (EWGLA) has successfully hosted a number of conferences with very high caliber interaction and attendance in the past few years.

During the period 2011–2012, partially in preparation for this paper, the author attended and observed locational sessions at EWGLA, RSAI, INFORMS, TRB, and the AAG. These focused observations are in addition to the normal awareness one

[4]There is also a lot of good material in the classics: I especially recommend Haggett et al. (1977) and the theoretical geography of Bunge (1966).

would have from perusing programs and conference proceedings and special issues (Table 1 summarizes the main features of conferences with a location theme). There are positive signs that there are active systems in place to nurture new participation and regeneration. There are changing ways in which academic conferences are organized. There is a large amount of material at conferences, and this requires "break out" sessions with specialized tracks or meetings. For example, there is the RSAI track in location models and the AAG special interest groups in spatial analysis. Such special sessions tend to be open across a range of career stages to nurture new doctoral candidates and graduates.

Technical Breakthroughs

An added indication of activity in the field is the emergence of new and previously unanticipated technical tools. For example, in the 1980s, one would be expected to be aware of Lagrangean relaxation, decompositions, and so on, as exemplified in Nemhauser and Wolsey (1988). Techniques that may have seemed abstract or at least difficult to adapt to some problems, were not adopted as quickly. However, with the emergence of some excellent work and exemplary pedagogic and applied cases, these tools are now being adapted in many location problems. For example, ordered medians use a structured redefinition of the data to dispose of certain sub-problems rather quickly. The ideas of pre-processing and exploitation of data structures in spatial location has a very long lineage in geography. Rushton (1989) and Hillsman (1980) developed techniques to sort distance matrices and to retain values below a threshold; such data structures were used in location software at the University of Iowa. Recently, Church (2002) has developed several special data structures (including vector assignment and spatially based search algorithms such as BEAMER, etc.).

Standard Problems and Data Sets

A final indicator of activity is the existence of standard data sets, test problems, and benchmarks special cases. (See the large number of location test cases in OR-library http://people.brunel.ac.uk/mastjjb/jeb/info.html which is a collection of test datasets for various OR problems.) These core ideas then lead to second and third generation attacks on the basic problems with the good results alluded to above and discussed further below. For example, a standard test data set developed from the 1970 air passenger data set (the CAB data) has been widely used to test hub location problems.

A detailed review of development in the case of hubs has been published recently and will not be repeated here (Campbell and O'Kelly 2012). Suffice it to say that the hub model trajectory shows all the markers of a developed and mature model in

Table 1 Some key features and highlights of selected conferences in 2011–2012

Conference	Type of sessions	Frequency/Location in 2011–2012	Major themes (number of sessions)
ISOLDE	Specialized group	Tri-annual (postponed in 2011; held in 2012)	Facility location models (5), applications (4), hub location, railway facility location, location models under uncertainty, continuous facility location (3), emergency facility location, network design (2)
EWGLA	Specialized sub-group	Annual (Nantes, October 12–14, 2011)	Discrete location and supply chain (2), hub location (3), applications, location under uncertainty (2), logistics and transportation, continuous location (2), covering models, competitive location
RSAI	Special track	Annual (Miami, November 10–12, 2011)	Organized track in location and spatial modeling (7) and other papers with a location focus
INFORMS	Cluster	Annual (Charlotte, November 13–18, 2011)	Cluster in location analysis: Community-based OR; locational decisions, continuous location, new directions, facility location under uncertainty (2), location models (2), location analysis, facility location, energy facility location issues

Sources
ISOLDE Program
EWGLA Program
RSAI: http://www.narsc.org/newsite/wp-content/uploads/2011/10/entire_program10212011.htm
INFORMS: https://informs.emeetingsonline.com/emeetings/formbuilder/clustersessionlist.asp?clnno=2726&mmnno=206

location. Moreover, it developed with awareness that the hub model gives rise to a similar source of non-convexity that is found in the interactive land use plan design model (Snickars 1978). This also helped to place the model correctly in the context of work by both regional scientists and planners (Brotchie 1969; Lundqvist 1973 on land use plan design as a quadratic program). In addition, the hub model matured with NSF support and a major breakthrough came as a result of a competition run by NSF for geo-scientists to collaborate with the mathematical community. The NSF's support allowed geographers to work with the operational research community (O'Kelly et al. 1995).

Theoretically Motivated Examples

Despite some remarkably clear indications of the centrality of transport in location theory, the role of transport as a source of agglomeration and growth has been underplayed. Some might trace this to the fact that transport services are themselves simply a means to overcome spatial separation and are best thought of as a drain on the value of the transported commodities (i.e. the iceberg model, Krugman 1992). It should be clear that the iceberg idea is simply a way to avoid modeling transport as a separate sector of the economy and does a great disservice to the representation and importance of transportation. Further ideas on transport, agglomeration, and urban growth are discussed very briefly in the following sections.

Impacts of Transport on Agglomeration and Growth

The failure to treat transport as endogenous can be traced back to the Lowry model in regional science. That model emphasized the difference between basic and non-basic employment and treated basic employment as essentially export manufacturing (e.g. steel mills) and the service sector as the sector that provides the employees of the manufacturing industry with retail and other services. In fact, the transport sector, to the extent that there are transshipment and other material handling activities, should itself be able to generate and support many jobs. Such is the theme, for example, of the McMaster Institute for Transport and Logistics which studies the critical role of logistics and supply chains. The employment and agglomeration effect of hubs drives home the fact that the organization and location of facilities to achieve these efficiencies can themselves drive the employment of workers in freight, airport services, distribution systems, and many other ancillary sectors. These sectors have very large impacts on location; for example, note the employment strength of air freight hub cities like Memphis, TN, and Louisville, KY (on the positive side), and the weakness of Wilmington, OH, where the loss of the DHL hub has had profound employment effects.

Location of Urban Growth

Following success in urban growth simulation models, the literature is now moving well beyond the initial and second generation models (Landis and Zhang 1998a, b). For example the SLEUTH model is a viable platform for exploring many aspects of urban development. Implicit in the model is a layer showing the suitability of the surface for development, or perhaps with slope as a surrogate. So, in lieu of early *isotropic* models, Clarke and Gaydos (1998) have established a way to estimate and evaluate the development potential of the land use and the implied locational patterns that go along with these trends. Many in the location specialization are probably unaware of the allied effort in this field and probably do not see them as directly locational; but these models treat broad ideas that hinge on location and accessibility. These ideas evolved into agent based model (ABM) approaches, which are grounded in theory and computation, and tend to have behavioral roots in line with the traditions of location theory and location modeling. For example, CSISS specialty meetings and SPACE workshops promoted the connection between cellular automata and land use models. There are also important branches – for example a distinction between optimal land use planning (land use allocation), and integrated land use and transportation modeling. Such themes are active areas of research in the journal *Computers, Environment, and Urban Systems*.

Conclusions

Is the challenge of solving bigger applied problems sufficiently interesting to keep theoretically inclined scholars invested in the field? I know from experience that even the most homogenous and coherently organized conference will typically have algorithmic papers that fail to excite some spatial modelers as well as more descriptive or applied work that has little to offer the pure mathematical audience. At least from recent experience, these sub-groups seem to be willing to hold together under the rubric of *location*, and to (politely) absorb each other's work. It is also important to note that some well-trained applied mathematicians are being drawn to the field. All this is to say that the field has had a good deal of cumulative and on-going development. Some of this can be traced to the core foundations of the French and German schools of location theory, surveyed in Ponsard (1983). Those who doubt or are unaware of the deep history of our sub-field would do well to browse some of these pages.

Work can sometimes be critiqued for being neither sufficiently realistic to provide real guidance nor sufficiently mathematical to be capable of offering a real breakthrough in modeling terms. Arguably, though, some of the most capable and influential work occurs when the skilled mathematical modeler attacks a really challenging practical problem and is forced to focus on the solvability and realism of the answer.

The continuum of models can be examined from the perspective of Webber (1984), who proposed that the role for models could fall in a kind of uneasy three-way contrast: explanatory, scientific, and applied. This paper has primarily looked at the scientific and applied foci. The reason is that the requirement for a good explanatory model, as suggested by Webber, is a theory of how social systems operate, and most location models are more likely to be concerned with other goals. The chances of having a causal model for the extrapolation of the exogenous parameters are also quite slim in the presence of open social systems, as explained by Webber (1984).

Scientific models align quite well with mathematical formulations of core location problems. As such, they lean towards careful characterization of the solutions space, the feasible region, the objective function, and the tightness of the formulation. The applied focus on the other hand is much more closely aligned with the notion of actually doing something with the location model in terms of resulting in a decision about a location of a facility or spatial activity of some kind. However, a moment's reflection by anyone who has more than a cursory awareness of the field will surely suggest that this is a false dichotomy. No applied modeler can be completely unconcerned with the intrinsic theoretical aspects of the model that they are using, and most modelers would be motivated to use the best available practical method to solve their problems. Such a solution, of course, hinges in many cases on deep and satisfying mathematical and scientific analysis of the model. On the other hand, few theoreticians are so far from reality that they do not have at least some interest in testing their ideas on a numerical problem of some size. Nonetheless, their interest sometimes stops with a large test case (for example, a randomly generated database) reporting run times and the power of their algorithm vs. other contenders. The applied case can take advantage of these ideas, but is also likely to use realistic data and to report the location and allocation results ahead of the run time. If the applied model is so computationally intensive that the determination of an actual answer is prohibitive, the applied analyst makes a careful exploration of the literature to seek better formulations, bounds and short cuts, or other techniques that will render the problem tractable.

With this possibly porous distinction between scientific and applied models in place, there are signs of maturity along both ends of the continuum. Scientific models have become more carefully formatted with the same problem set up in different ways with increasing attention paid to the polyhedral properties (facets, cuts, valid inequalities). The practical angle, in contrast, is more likely to add features to an existing problem, fine tuning it to address issues of capacity, stochastic demand, and so on.

In conclusion, it is important to recap the idea that the maturity of the field is a mark of the advanced development and rich tradition of past efforts. In order for maturity to lead to further positive developments, it is suggested that the current generation learn from the past, draw on it, and potentially shape new forms of discourse. The active member of SOLA and of RSAI for example might usefully ask these questions: What has been gained by location analysis in linking theory and applied work, and where does this lead? Is there a trend towards mainstream

relevance for location theory, or is it past its prime and headed for old age and decline? It is perhaps presumptuous to make any declarations on the definite answer here, but the debate can start from the issues raised here. A lot depends here on precisely an issue brought up in this paper: i.e. that to remain a mainstream activity, to develop future new breakthroughs, and to continue to develop this two-way link between theory and applied work, the current generation of workers needs to absorb the ideas from a generation of academic scientists who are about to reach a critical career phase. For the topic to persist and develop, there needs to be an active development strategy to assure the lasting impact of the ideas outlined in this paper.

Further development of interdisciplinary collaboration via scientific conferences would be a great aid to the field. An interesting question: which form of conference forum is most productive? What might meetings or collaborations look like in the future? Where is the best place to cluster specialists for a scientific meeting? Would conferences be stronger if they were less specialized and encouraged more collaborative meetings between applied planners and mathematical theorists?[5] There are a number of fundamental principles that are shared by many of the strands of literature alluded to in this paper. Among these are notions of constrained optimality, least-cost efficient solutions, and applied versions of academic optimization problems based on mathematical methods. But there is a fruitful role for optimal location ideas in planning even if there are so many imperfections in real world markets that make the orderly models of optimal location somewhat unsuited. The sets of questions intersect, but the core advances in the exclusive positions take different pathways. It is hoped that this paper has contributed to a fundamental awareness of the relationships between optimal location and physical land use/activity planning among both seasoned professionals and the new generation of academic scholars.

Acknowledgement I would like to thank Gordon Mulligan for initial conversations and feedback on this topic, and Mark Partridge for some helpful guidance on missing links. Kevin Cox reminded me of some important links at the University of Toronto. I also received helpful comments from my colleagues: Yongha Park and Meng Guo. Thanks also to Julia Elmer for constructive comments. The referees and the Editor, Jean-Claude Thill, provided detailed and constructive comments on a preliminary version. The author is grateful to The National Science Foundation (BCS-1125840) for support of current research on hub location models.

References

Bailey, W. G., & Huber, W. A. (2012) *Methods and systems for optimizing network travel costs*. United States Patent Office (USPTO 8332247).

Batty, M. (2008). Fifty years of urban modeling: Macro-statics to micro-dynamics (S. Albeverio, D. Andrey, P. Giordano & Alberto Vancheri, Eds.) (pp. 1–20). Mendrisio: Physica-Verlag HD.

Beckmann, M. (1955) Some reflections of Losch's theory of location. Papers in Regional Science, 1(1), 139–148.

[5]Thanks to Julia Elmer for suggesting this question.

Brotchie, J. F. (1969). A general planning model. *Management Science, 16*, 265.
Bunge, W. (1966). *Theoretical geography* (Vol. 1). Gleerup: Royal University of Lund, Department of Geography.
Campbell, J. F., & O'Kelly, M. E. (2012). Twenty-five years of hub location research. *Transportation Science, 46*(2), 153–169.
Chen, Y. (2010). Urban systems: Contemporary approaches to modelling (C. S. Bertuglia, G. Leonardi, S. Occelli, G. A. Rabino, R. Tadei, & A. G. Wilson Eds.). *Geographical Analysis, 24*(3), 281–282.
Church, R. L. (2002). Geographical information systems and location science. Computers & Operations Research, 29(6), 541–562.
Clarke, K., & Gaydos, L. (1998). Loose-coupling a cellular automaton model and GIS: Long-term urban growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographical Information Science, 12*(7), 699–714.
Contreras, I., Cordeau, J. F., & Laporte, G. (2012). Exact solution of large-scale hub location problems with multiple capacity levels. Transportation Science, *46*(4), 439–459.
Daskin, M. (1995). *Network and discrete location models, algorithms, and applications*. New York: Wiley.
Drezner, Z. (1995). *Facility location: A survey of applications and methods*. New York: Springer.
Drezner, Z., & Hamacher, H. (2002). *Facility location: Applications and theory*. Berlin: Springer.
Duranton, G. (2010). Introduction: The Journal of regional Science at 50: Looking forward to the next 50 years. *Journal of Regional Science, 50*(1), 1–3.
Erlenkotter, D. (1978). A dual-based procedure for uncapacitated facility location Operations Research 26(6), 992–1009.
Fujita, M. (2010). The evolution of spatial economics: From Thünen to the new economic geography. *Japanese Economic Review, 61*(1), 1–32.
Griffith, D. A., Chun, Y., O'Kelly, M. E., Berry, B. J. L., Haining, R. P., & Kwan, M. P. (2013). Geographical analysis: The first forty years, Geographical Analysis, 45, 1–27.
Haggett, P., Cliff, A. D., & Frey, A. (1977). *Locational models*. New York: Wiley.
Handler, G. Y., & Mirchandani, P. B. (1979). *Location on networks: Theory and algorithms*. Cambridge, MA: MIT Press.
Hillsman, E. L. (1980). *Heuristic solutions to location-allocation problems: a user's guide to ALLOC IV, V and VI*. Iowa City: Department of Geography, University of Iowa.
Holmes, J., Williams, F. B., & Brown, L. A. (1972). Facility location under a maximum travel restriction: An example using day care facilities. *Geographical Analysis, 4*(3), 258–266.
Krugman, P. R. (1992). *A dynamic spatial model*. Cambridge, MA: National Bureau of Economic Research.
Landis, J., & Zhang, M. (1998a). The second generation of the California urban futures model. Part 1: Model logic and theory. *Environment and Planning B: Planning and Design, 25*(5), 657–666.
Landis, J., & Zhang, M. (1998b). The second generation of the California urban futures model. Part 2: Specification and calibration results of the land-use change submodel. *Environment and Planning B: Planning and Design, 25*(6), 795–824.
Lea, A. C. (1973). *Location-allocation models: A review*. Toronto: University of Toronto.
Lundqvist, L. (1973). Integrated location-transportation analysis: A decomposition approach. Regional and Urban Economics, 3, 233–262.
Mulligan, G. F., Partridge, M. D., & Carruthers, J. I. (2012). Central place theory and its reemergence in regional science. *The Annals of Regional Science, 48*(2), 405–431.
Murray, A. T. (2010). Advances in location modeling: GIS linkages and contributions. Journal of Geographical Systems, 12(3), 335–354.
Nemhauser, G. L., & Wolsey, L. A. (1988). *Integer and combinatorial optimization*. New York: Wiley.
O'Kelly, M. E., & Bryan, D. (1996). Agricultural location theory: von Thünen's contribution to economic geography. *Progress in Human Geography, 20*(4), 457–475.

O'Kelly, M. E., & Murray, A. T. (2004). A lattice covering model for evaluating existing service facilities. *Papers in Regional Science, 83*(3), 565–580.

O'Kelly, M. E., Skorin-Kapov, D., & Skorin-Kapov, J. (1995). Lower bounds for the hub location problem. *Management Science, 41*(4), 713–721.

Ponsard, C. (1983). *History of spatial economic theory*. Berlin/New York: Springer.

ReVelle, C. (1993). Facility siting and integer-friendly programming. *European Journal of Operational Research, 65*(2), 147–158.

Ross, G. T., & Soland, R. M. (1975). A branch and bound algorithm for the generalized assignment problem. Mathematical Programming, 8(1), 91–103.

Royuela, V. (2012). Regional science trends through the analysis of the main facts of the 51st ERSA conference. *Investigaciones Regionales, 24*, 13–39.

Rushton, G. (1989). Applications of location models. *Annals of Operations Research, 18*(1), 25–42.

Scott, A. J. (1970). Location-allocation systems: A review. *Geographical Analysis, 2*(2), 95–119.

Scott, A. J. (2000). Economic geography: The great half-century. *Cambridge Journal of Economics, 24*(4), 483–504.

Smith, H. K., Laporte, G., & Harper, P. R. 2009. Locational analysis: Highlights of growth to maturity. Journal of the Operational Research Society, 60(4), 140–148.

Snickars, F. (1978). Convexity and duality properties of a quadratic intraregional location model. *Regional Science Urban Economics, 8*(1), 5–20.

Sui, D. Z., et al. (2009). *Geospatial technology and the role of location in science*. Dordrecht: Springer Netherlands.

Thill, J.-C. (2000). Geographic information systems for transportation in perspective. *Transportation Research Part C: Emerging Technologies, 8*(1–6), 3–12.

Thill, J.-C. (2001). Spatial multicriteria decision making and analysis: A geographic information sciences approach. *International Planning Studies, 6*(4), 427–435.

Tong, D., & Murray, A. T. (2012). Spatial optimization in geography. *Annals of the Association of American Geographers, 102*(6), 1290–1309.

Webber, M. J. (1984). *Explanation, prediction and planning: The Lowry model*. London: Pion.

Winkenbach, M., Kleindorfer, P. R., Lemarié, B., Levêque, C., Roset, A., & Spinler, S. (2012). A mixed integer linear programming model for solving large-scale integrated location-routing problems for urban logistics applications at Groupe La Poste [On line: http://idei.fr/doc/conf/pos/papers_2012/winkenbach.pdf].

Spatial Uncertainty Challenges in Location Modeling with Dispersion Requirements

Ran Wei and Alan T. Murray

Abstract Geographic information systems provide the capacity to digitally create, store, manipulate, analyze and display various types of spatial information. While these functionalities enable handling of spatial data in a much more rapid and precise way than traditional paper-based approaches, uncertainties nevertheless remain in digital information and are not likely to ever be completely eliminated. Location modeling, as an important spatial analytical approach, must therefore confront the various uncertainties and errors in spatial data. In this chapter we detail multi-objective models structured to account for data uncertainty in support of location siting when spatial dispersion is a prerequisite. These models explicitly incorporate facets of data uncertainty and can be applied in a manner that enables evaluation of uncertainty impacts with statistical confidence. Solution approaches are discussed, along with the case study setting, to demonstrate how these models address spatial uncertainty in a context supporting facility location planning.

Keywords Spatial uncertainty • Spatial dispersion • Location modeling

Introduction

Locational decisions are critical for both public and private sectors. A good locational choice can translate to numerous economic and social benefits, while poor locational decisions may result in lost income, excessive costs or increased risk to people and property. A location problem involves determining the best place to site people, goods, services and/or activities which interact across space, while maintaining certain constraints or conditions. A location model typically can be

R. Wei (✉)
School of Public Policy, University of California at Riverside, Riverside, CA 32521, USA
e-mail: ran.wei@ucr.edu

A.T. Murray
Department of Geography, University of California at Santa Barbara, Santa Barbara, CA 93106, USA
e-mail: amurray@ucsb.edu

J.-C. Thill (ed.), *Spatial Analysis and Location Modeling in Urban and Regional Systems*, Advances in Geographic Information Science,
https://doi.org/10.1007/978-3-642-37896-6_12

structured using a mathematical model, consisting of decision variables, one or more objectives and constraints (Murray 2010; Tong and Murray 2012). As an important spatial analytical method, a wide variety of location models have been developed to assist locational decision making. Most models aim to enhance accessibility to facilities or services, such as fire stations, warehouses and public parks, to name but a few. However, there are some types of facilities which are necessary for society, but undesirable to locate in close proximity to people or certain activities. Typical examples of such facilities include nuclear power plants, chemical processing laboratories, addiction treatment centers, halfway houses and waste disposal sites. In order to limit localized impacts and achieve an equitable distribution of these facilities, dispersion requirements are often enacted implicitly or explicitly. Related location models, known as dispersion models, have been developed to account for dispersion requirements. Detailed reviews of dispersion models can be found in Church and Garfinkel (1978), Kuby (1987), Erkut and Neuman (1989), Murray et al. (1998), Curtin and Church (2006) and Church and Murray (2009).

Dispersion models generally encapsulate a number of spatially-based features, including facility and demand locations, distance or travel time, adjacent areas, etc. These spatial features are model inputs, requiring creation and processing using geographical information systems (GIS). It is widely acknowledged that uncertainty exists in digital spatial information and is unlikely to ever be completely eliminated (Goodchild 1993; Heuvelink 1998; Longley et al. 2015). One can therefore conclude that the spatial information associated with dispersion modeling is also uncertain in various ways. Dispersion models are typically deterministic, assuming model inputs to be precise and accurate. This assumption obviously deviates from the truth, which may result in biased or unreliable analysis and interpretation. In general, little is known about the impacts of data error and uncertainty in spatial analysis (Hunter and Goodchild 1996; Murray 2003; Wei and Murray 2012; Yao and Murray 2013). Of note in this chapter is the work of Wei and Murray (2012), which demonstrates that imperfections in spatial data could have significant impacts on dispersion model findings. This means that there is a need to account for spatial uncertainty in dispersion models and associated analyses. Other spatial analytical methods have concluded this as well (Heuvelink et al. 1989; Aerts et al. 2003; Goovaerts 2006; Griffith et al. 2007; Malizia 2013).

In this chapter, we detail probabilistic approaches to assess the impacts of spatial uncertainty in dispersion modeling. Uncertainty in spatial restrictions is explicitly represented by a probability term derived from the positional errors of spatial features. The probability term combined with multi-objective optimization accounts for potential spatial uncertainty impacts with a level of statistical confidence. The next section details a basic dispersion problem and the model most commonly used to address such problems. Following this, an approach to derive a probability associated with dispersion restrictions is detailed. Two multi-objective models are then reviewed to explicitly account for spatial data uncertainty. A case study illustrates how uncertainty can be addressed, as well as implications for planning and decision making. Finally, we conclude with a discussion of the results, commentary and suggestions for further investigation of the effects of policy change.

Dispersion Modeling

A commonly encountered planning problem involves siting one or more facilities within a region where geographic concentration is to be avoided. As noted previously, such facilities could be waste processing or recycling centers, franchise establishments, medical centers, etc. The anti-covering location problem (ACLP) is one of the most widely applied dispersion models (Church and Murray 2009). The ACLP seeks a configuration of sites that maximizes benefits associated with sited facilities, while maintaining a prespecified spatial separation between facilities. It has been utilized in forest harvest scheduling (Barahona et al. 1992; Hochbaum and Pathria 1997; Murray 1999; Goycoolea et al. 2005), retail planning (Zeller et al. 1980), nature reserve design (Downs et al. 2008), and planning and policy evaluation (Murray et al. 1998; Grubesic and Murray 2008; Wei and Murray 2012). Consider the following notation:

$$i = \textit{index of spatial planning units}$$

$$\beta_i = \textit{benefit associated with selecting unit } i$$

$$\Phi_i = \textit{set of units that spatially conflict with unit } i$$

$$X_i = \begin{cases} 1, & \textit{if unit } i \textit{ is selected for facility placement}, \\ 0, & \textit{otherwise}. \end{cases}$$

Units in this context could, for example, refer to commercial parcels for waste recycling centers or other types of undesirable service facilities. Φ_i represents the set of units that are too close to, or spatially conflict with, unit i. With this notation, the ACLP can be formulated as follows (Church and Murray 2009):

$$\text{Maximize} \qquad \sum\nolimits_{\mathrm{i}} \beta_{\mathrm{i}} X_{\mathrm{i}} \tag{1}$$

$$\textit{Subject to} \qquad X_i + X_j \leq 1 \quad \forall i, j \in \Phi_i \tag{2}$$

$$X_i = \{0, 1\} \quad \forall i \tag{3}$$

The objective (Eq. 1) is to maximize the benefits associated with the selection of facility sites. Constraints (2) ensure that no two conflicting facility locations can be simultaneously selected. Binary integer requirements are stipulated in (3).

The defining feature of the ACLP is the dispersion orientation in siting facilities. The separation constraints (Eq. 2) reflect this, and are assumed to be known with certainty. However, given that spatial data is fraught with uncertainty and error (Goodchild and Gopal 1989), the spatial dispersion restrictions in the ACLP may potentially become ambiguous. The dispersion model is applied in a case study and then its associated uncertainty issues are introduced.

Case Study

The case study involves alcohol beverage control and the societal impact of changes in associated public policy. The state of Pennsylvania is considering a move away from controlling alcohol outlets, leaving only state laws mandating that alcohol outlets not be closer than 200 ft. from another outlet nor within 300 ft. of a school, church, hospitals, parks and playgrounds. These measurements are required to be made in Euclidean distance by the Pennsylvania Liquor Control Board, regardless of any obstacles (PLCB 2008). As the state of Pennsylvania currently restricts the sale of alcoholic beverages for off-premises consumption to specific amounts/configurations within particular types of outlets (e.g. state-run stores and beer distributors, of which there are a limited number), a substantial increase in the number of outlets is expected under privatization, improving consumer access. The issue, then, is where these new outlets would be located. This concern arises because there is a notable correlation between alcohol sales and crime rates (Grubesic and Pridemore 2011). To explore this issue, Grubesic et al. (2012) employed the ACLP to assess the likely regional and local impacts of such public policy change.

In this case study, six neighborhoods of Philadelphia are considered here: Poplar, Hartranft, North Central, Fairmout, Fishtown and West Kensington. These neighborhoods were chosen specifically to represent areas in which violent crime is a major problem. Thus, an increase in alcohol outlets would likely exacerbate already bad social conditions. Figure 1 depicts this region, which is located in the north of Center City, Philadelphia. This illustration shows the locations of 183 existing outlets and 934 commercial parcels where an alcohol outlet could conceivably be sited under privatization.

The dispersion model is structured using Python, and subsequently solved using a commercial optimization package, Gurobi (2017). Computational processing was carried out on a MS Windows-based, Intel Xeon (2.53 GHz) computer with 6 GB of RAM.

Application of the ACLP indicates that the number of alcohol outlets locating in these neighborhoods could increase to 270, while still maintaining the 200 ft. separation restriction between outlets required by state law. The spatial configuration associated with this analysis is displayed in Fig. 2. However, this analysis did not account for spatial data uncertainty, which is problematic as the parcel boundary data utilized in the analysis are only accurate to within 25 ft. The issue then is what are the implications of data uncertainty in this case, both in terms of total number

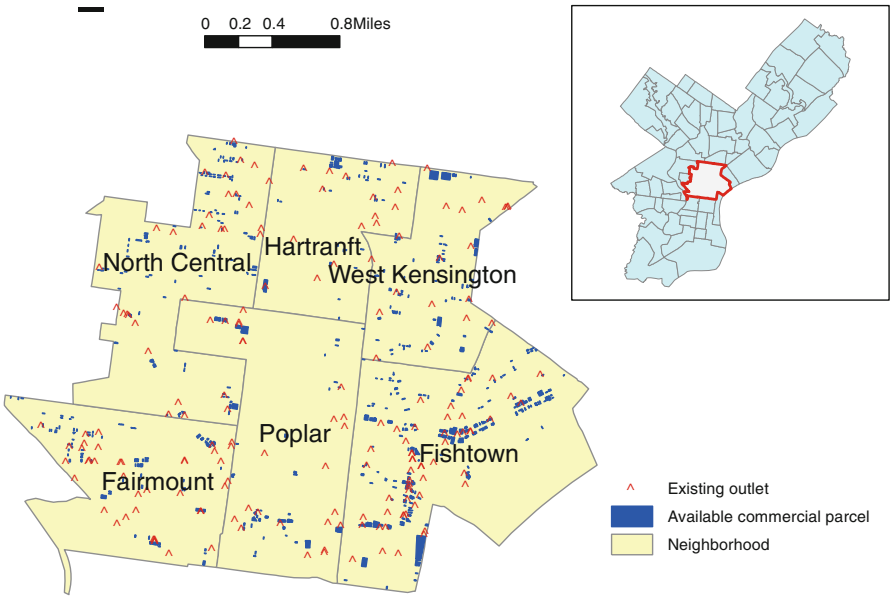

Fig. 1 Studied neighborhoods in the city of Philadelphia

as well as spatial configurations of potential alcohol outlets under privatization. The next section will present a probabilistic approach for characterizing uncertainty impacts in dispersion modeling.

Uncertainty Implications

The modeling of uncertainty and error in spatial data dates back to classical measurement theory, in which the positional errors of points are represented as a circular normal distribution (Greenwalt and Shultz 1962). Even today, this approach is the most commonly used and accepted point error model (Goodchild 1993). Many map accuracy standards, such as the National Map Accuracy Standard (NMAS), the American Society of Photogrammetry and Remote Sensing (ASPRS) Accuracy Standard and the National Standard for Spatial Data Accuracy (NSSDA), are based upon this point error model. Leung and Yan (1998) extended the circular normal model to other types of spatial features such as line and polygon in which the errors of all points along the line or polygon boundary are independent and conform to the same circular normal distribution. While such assumptions may not hold in reality, this general error model represents an approach to describe the accuracy of an entire geographic dataset when the spatial structure of uncertainty information is not available, which is the case for most geographic datasets (Goodchild et al. 1999).

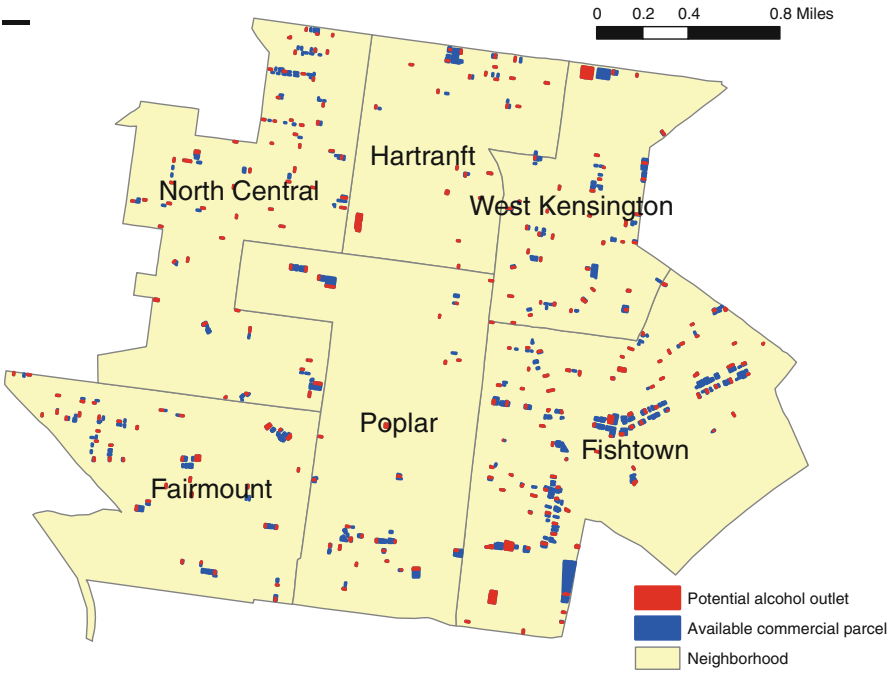

Fig. 2 Spatial configuration of potential alcohol outlets derived using the ACLP (270 outlets)

Utilizing the error model developed in Leung and Yan (1998), a probability term for distance-based separation constraints can be derived. Let U_i and U_j be two units, l the minimum allowable distance between U_i and U_j if spatial certainty is assumed, V_k a vertex of U_i, and V_t a vertex of U_j. Suppose that the distance, $d(V_k, V_t)$, between V_k having coordinates (x_k, y_k) and V_t having coordinates (x_t, y_t) is the minimum distance l. Accounting for spatial error/uncertainty, the distance may not be completely certain, but rather variable depending on associated error. When positional errors in both coordinates, ε_x and ε_y, are independent and represented by the same normal distribution, consistent with the general error model, the minimum distance can be viewed as a quadratic form of normally distributed variables:

$$\varepsilon_x \sim N\left(0, \sigma^2\right), \quad \varepsilon_y \sim N\left(0, \sigma^2\right) \tag{4}$$

$$X_k - X_t \sim N\left(x_k - x_t, 2\sigma^2\right), \quad Y_k - Y_t \sim N\left(y_k - y_t, 2\sigma^2\right) \tag{5}$$

$$L^2 = (X_k - X_t)^2 + (Y_k - Y_t)^2 \tag{6}$$

There is much discussion in the statistical literature regarding the distribution of the quadratic form of normal variables (Mathai and Provost 1992). It is straightforward to conclude that the minimum distance between U_1 and U_2 in Eq. (6) conforms to a noncentral chi-square distribution:

$$\frac{L^2}{2\sigma^2} \sim noncentral\ \chi^2\left(k = 2, \lambda = \frac{l^2}{2\sigma^2}\right) \tag{7}$$

As a result, given a prespecified separation distance d reflecting the condition that two units are too close to each other, the probability p_{ij} that the minimum distance between U_i and U_j is not larger than d can be expressed as:

$$p_{ij} = P\left(L \leq d\right) = F_{nc-\chi^2}\left(\frac{d^2}{2\sigma^2}\right) \tag{8}$$

This represents the probability that a required dispersion condition is violated if both units are simultaneously selected, given corresponding assumptions of how error is distributed.

Interpreting this in the context of the ACLP means that constraints (2) become ambiguous when considering spatial error and the conflict violation probability p_{ij} indicates whether the dispersion constraint between unit i and j should be actually imposed given a confidence level $1-\alpha$. Specifically, the proximity condition could be relaxed if $p_{ij} < 1-\alpha$; otherwise the condition must be imposed. A framework within which dispersion restriction can be structured while explicitly addressing data uncertainty therefore emerges.

Model Extensions

There is uncertainty in spatial data, creating ambiguity in the proximity constraints of the ACLP. The challenge is how to account for this in the context of the ACLP, and spatial analysis more generally. Wei and Murray (2012) developed a multi-objective model to integrate spatial data uncertainty in the ACLP. Some separation constraints in the ACLP become uncertain and may or may not need to be imposed, and others may have been overlooked but need to be imposed. In order to incorporate the known and potential separation constraints into the ACLP, two conflict sets are introduced to represent certain and uncertain conditions for unit i as follows:

$$\Omega_i = set\ of\ units\ that\ are\ a\ known\ conflict\ with\ unit\ i$$

$$\Psi_i = set\ of\ units\ that\ potentially\ conflict\ with\ unit\ i$$

In addition to the two conflict sets, new decision variables are also necessary:

$$Y_{ij} = \begin{cases} 1, & \textit{if the separation restriction between units i and j is relaxed,} \\ 0, & \textit{otherwise.} \end{cases}$$

These variables track whether an uncertain proximity condition is imposed or not. Employing this additional notation, an extension of the ACLP can be structured (Wei and Murray 2012):Error-anti-covering location problem I (E-ACLP I)

$$\textit{Maximize} \qquad \sum\nolimits_i \beta_i X_i \tag{9}$$

$$\textit{Minimize} \qquad \sum\nolimits_i \sum\nolimits_{j \in \Psi_i} p_{ij} Y_{ij} \tag{10}$$

$$\textit{Subject to} \qquad X_i + X_j \leq 1 \quad \forall i, j \in \Omega_i \tag{11}$$

$$X_i + X_j - Y_{ij} \leq 1 \quad \forall i, j \in \Psi_i \tag{12}$$

$$X_i = \{0, 1\} \quad \forall i \tag{13}$$

$$Y_{ij} = \{0, 1\} \quad \forall i, j \in \Psi_i$$

The first objective, (Eq. 9), remains to maximize the benefits associated with selecting units for facility siting, as structured in the ACLP. The second objective, (Eq. 10), is to minimize the total probability of violating separation constraints that are spatially uncertain. The two objectives are in conflict in the sense that gains in benefits can be achieved only by relaxing uncertain proximity constraints. Constraints (11) ensure that no two selected facility sites conflict among the known restrictions. Constraints (12) track separation of those facilities that might possibly be in conflict. These are the constraints that could be acceptable to relax or ignore, depending upon the level of uncertainty. When Y_{ij} equals one, both X_i and X_j could take the value of one in constraint (12), indicating that both units could be selected for siting facilities; otherwise, at most one of them can be selected. Constraints (13) impose binary integer restrictions on decision variables.

By design, the ACLP and E-ACLP I are closely related. In fact, the ACLP is equivalent to the E-ACLP I when $\Omega_i = \Phi_i$ and $\Psi_i = \varnothing$. In practice, however, it would be expected that $\Omega_i \subset \Phi_i$ and $(\Omega_i \cup \Psi_i) \supset \Phi_i$, indicating that the total facility siting benefits would increase when uncertain proximity constraints are relaxed, as fewer overall spatial restrictions would be imposed. Alternatively, total benefits would decrease when uncertain restrictions are imposed, because more restrictions would be imposed relative to the ACLP.

The E-ACLP I explicitly accounts for spatial data uncertainty by introducing the certain and uncertain conflict sets, Ω_i and Ψ_i. Wei and Murray (2012) utilized a rather simplistic approach to determine known and potential proximity conflicts. If the minimum distance between two units is less than the specified separation distance minus positional accuracy, $(d-\varepsilon)$, the conflict was deemed known (certain); if the distance is between $(d-\varepsilon)$ and $(d+\varepsilon)$, the conflict is thought to be possible (uncertain). Such classification lacks a strict statistical interpretation. However, it can be enhanced by integrating the conflict violation probability approach detailed previously. Specifically, given a specified confidence level $1-\alpha$, unit j belongs to the known (certain) conflict set Ω_i if $p_{ij} \geq 1-\alpha$; otherwise, it belongs to the potential (uncertain) conflict set Ψ_i if $0 < p_{ij} < 1-\alpha$. Using this approach, each known and potential conflict set is associated with a specific confidence level. It should not be a surprise that the E-ACLP I is not the only way to account for spatial uncertainty. Wei and Murray (2015) presents an alternative approach, where relaxations of uncertain constraints associated with unit i are allowed, or they are not, instead of relaxing each pair of potential dispersion restrictions individually. Consider the following additional notation:

$$\widetilde{p}_i = \textit{confilct violation probability associated with unit } i$$

$$Z_i = \begin{cases} 1, & \textit{if unit } i \textit{ is relaxed}, \\ 0, & \textit{otherwise}. \end{cases}$$

The probability, $\widetilde{p}_i$, is not related to each constraint relaxation but rather to each spatial unit. Relaxing unit i reflects that all potential separation constraints associated with unit i could be relaxed. Specification of this unit conflict violation probability, $\widetilde{p}_i$, could take the following form based on the previous discussion:

$$\widetilde{p}_i = \sum\nolimits_j p_{ij} \tag{14}$$

New decision variables, Z_i, are then needed to track whether unit constraints are to be relaxed or not. This alternative model can therefore be structured as follows (Wei and Murray 2015):

Error-anti-covering location problem II (E-ACLP II)

$$\textit{Maximize} \qquad \sum\nolimits_i \beta_i X_i \tag{15}$$

$$\textit{Minimize} \qquad \sum\nolimits_i \widetilde{p}_i Z_i \tag{16}$$

$$\textit{Subject to} \qquad X_i + X_j \leq 1 \quad \forall i,j \in \Omega_i \tag{17}$$

$$X_i + X_j - Z_i - Z_j \leq 1 \quad \forall i, j \in \Psi_i \tag{18}$$

$$X_i = \{0, 1\} \quad \forall i \tag{19}$$

$$Z_i = \{0, 1\} \quad \forall i$$

The first objective, (Eq. 15), remains to maximize the total benefits of selecting units for siting facilities and the second objective, (Eq. 16), is to minimize the total facility conflict violation probabilities. Constraints (17) ensure the certain separation requirements are satisfied between known conflicting cases. Constraints (18) link the relaxation decision of unit i to all of the units in the set Ψ_i. When Z_i equals one, all units in Ψ_i could be selected concurrently for facility placement. Constraints (19) impose binary integer restrictions on decision variables.

When comparing the formulation of E-ACLP I and II, their differences in dealing with potential/uncertain proximity restrictions are obvious. In the E-ACLP I each potential separation restriction pair is independent of other potential restrictions. Alternatively, all potential restrictions involving one unit are bundled together in the E-ACLP II. The differences between these two models will be illustrated in the case study.

Research has shown the ACLP to be challenging to solve (Moon and Chaudhry 1984; Erkut et al. 1996; Murray and Church 1997; Murray and Kim 2008; Gamarnik and Goldberg 2010). Thus, extensions of the ACLP, such as the multi-objective E-ACLP I and II, are also expected to be computationally demanding. Further, one must also contend with the complications of multiple objectives.

Impacts of Uncertainty

After introducing the issue of spatial uncertainty and reviewing models to explicitly account for this, an investigation of the impact of uncertainty in alcohol beverage control is detailed. The positional accuracy of the parcels is estimated to be 25 ft. at the 90% confidence level, resulting in the standard deviation of positional error in both coordinates, σ, being 11.65 ft. (see FGDC 1998). The E-ACLP models detailed above can therefore be applied to evaluate the implications of spatial data uncertainty. The concern is that data uncertainty could influence the number and spatial distribution of permitted alcohol outlets which are facilitated under privatization.

Both E-ACLP I and E-ACLP II are utilized to assess the impacts of data uncertainty. They are structured using Python and solved using Gurobi. The constraint method is used to identify all trade-off solutions (see Cohon 1978). The probability of violating potential proximity restrictions, p_{ij}, is derived based upon

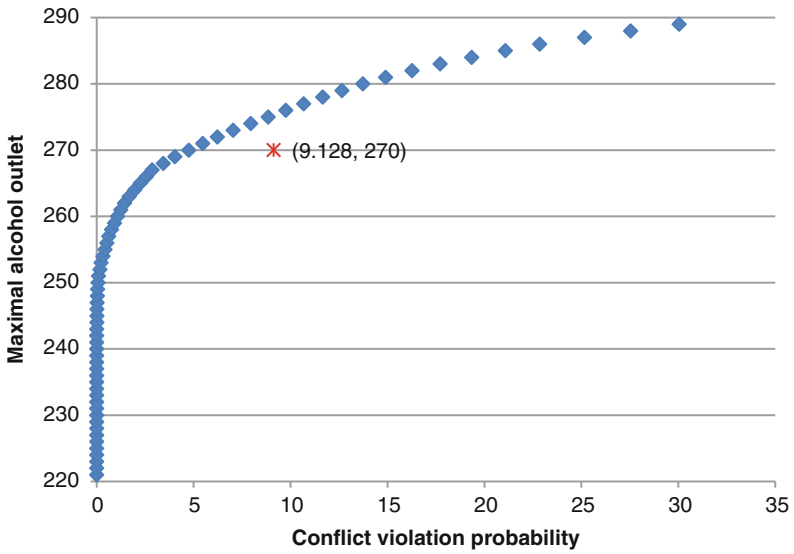

Fig. 3 Trade-off curve for the E-ACLP I

the minimum required separation distance and positional accuracy. The confidence level is set as 95%, indicating that the separation constraints could be relaxed only if the conflict violation probability is less than 0.95. Encountered models required less than 0.1 second to solve using Gurobi, with the constraint method used to identify all trade-off solutions associated with the two objectives.

The impacts of spatial uncertainty using the E-ACLP I are summarized in Fig. 3. Shown in Fig. 3 are trade-off solutions. The x-axis ("Conflict violation probability") corresponds to objective (10), which equals the total probability of relaxed potential proximity constraints. The y-axis ("Maximal alcohol outlets") corresponds to objective (9), indicating the total number of alcohol outlets that could be located in these neighborhoods. Accounting for spatial uncertainty suggests that the expected number of alcohol outlets could be as few as 221 but as high as 289 at the 95% confidence level. This is in contrast to the 270 identified using the ACLP, assuming spatial data certainty. This is an 18.15% reduction in total alcohol outlets in the lowest case and a 7.04% increase in highest case. The Pareto optimal curve shows that an increase in alcohol outlets means that the total violation probability increases, or that more potential proximity constraints are ignored. In the case of 221 outlets (Fig. 3), all potential restrictions are imposed (zero total violation probability). On the other extreme, the 289-outlet case is achieved by relaxing all potential restrictions. Remaining solutions summarized in Fig. 3 reflect a trade-off ranging between these extremes (221 to 289 total alcohol outlets) with variation in the total violation probability ranging from 0 to 30.05, respectively. Beyond the total number of potential alcohol outlets, each solution is associated with a different spatial pattern. Figure 4 shows the spatial configuration of the 221-outlet case, while

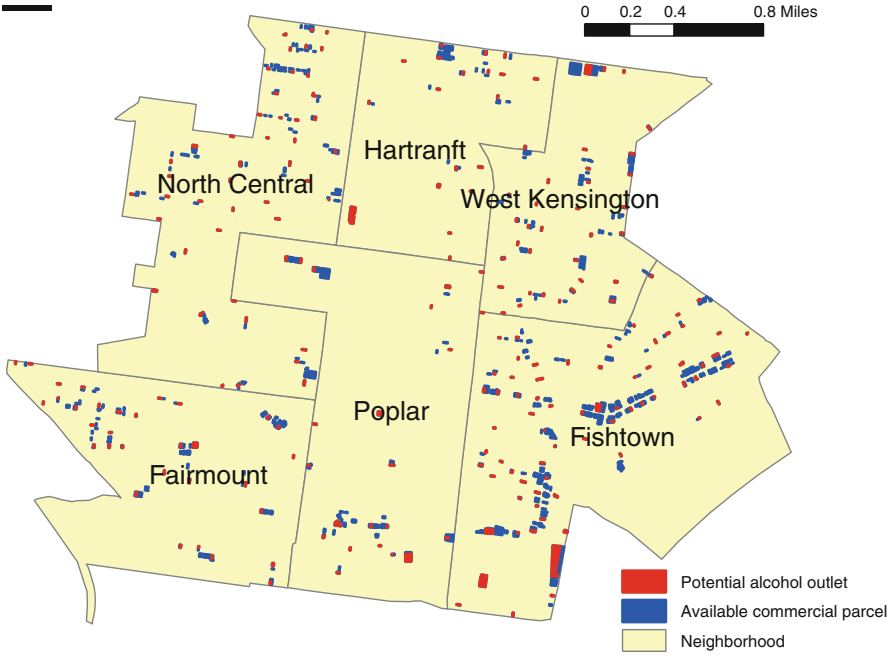

Fig. 4 Spatial configuration of potential alcohol outlets derived using the E-ACLP I (221 outlets)

that of the 289-outlet case is displayed in Fig. 5. A trade-off solution with 270 outlets (4.04 total violation probability) is also given in Fig. 6. There are clear spatial pattern differences in Figs. 4, 5, and 6, which would be expected to have varying regional and local impacts.

The impacts of spatial uncertainty using the E-ACLP II are summarized in Fig. 7. The x-axis corresponds to objective (16), representing the sum of violation probability associated with each unit ("Unit conflict probability"). This ranges from 0 to 411.41. The y-axis indicates the total number of possible alcohol outlets ("Maximal alcohol outlets"), as shown in Fig. 3 previously. Since the two extremes in Fig. 7 characterize the scenarios of relaxing none through relaxing all potential restrictions, the range of the x-axis is the same as Fig. 3, with the spatial configurations of the two extreme cases being the same obtained using the E-ACLP I (Figs. 4 and 5). However, the other tradeoff solutions exhibit differing spatial patterns. Figure 8 shows the spatial configuration of 270 outlets (a 132.39 total conflict probability) for the E-ACLP II; this can be contrasted with the spatial configuration depicted in Fig. 6 obtained using the E-ACLP I. we conclude that trade-off solutions may differ depending upon which of the two models is used and the associated spatial patterns may differ as well.

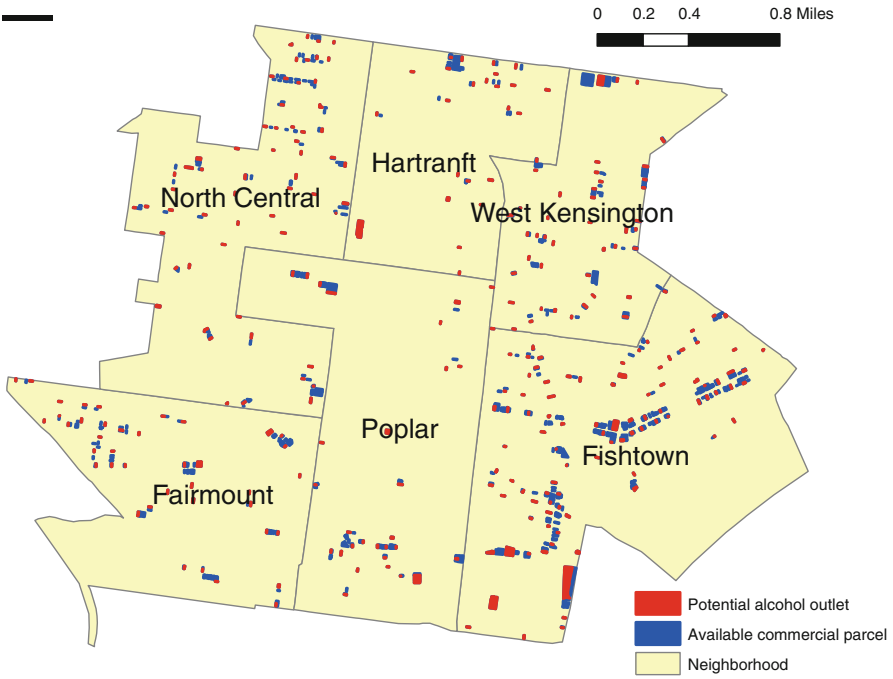

Fig. 5 Spatial configuration of potential alcohol outlets derived using the E-ACLP I (289 outlets)

Discussion and Conclusions

The results presented illustrate that spatial data uncertainty has real and significant impacts on analyses. If uncertainty/error in spatial information is ignored, findings would be erroneous, biased or misguided. For the study area analyzed in this research, the total number of alcohol outlets ranged widely from 221 to 289 at the 95% confidence level when the impacts of data uncertainty were assessed, in contrast to 270 alcohol outlets suggested under an assumption of positional accuracy. This implies that alcohol sales privatization might lead to a 20.77% growth in alcohol outlets in the most conservative case, or increase as high as 57.92% when accounting for spatial data uncertainty. Given the high crime rate of these neighborhoods, even the 20.77% increase could be cause for concern in the community.

Though not mentioned in the case study section, comparisons were also undertaken to contrast the probabilistic approach developed here with the more simplistic strategy employed in Wei and Murray (2012). The analysis suggests that the total number of alcohol outlets varies from 252 to 286, a smaller range than the 221 to 289 outlet range resulting from the probabilistic approach. These results differ because

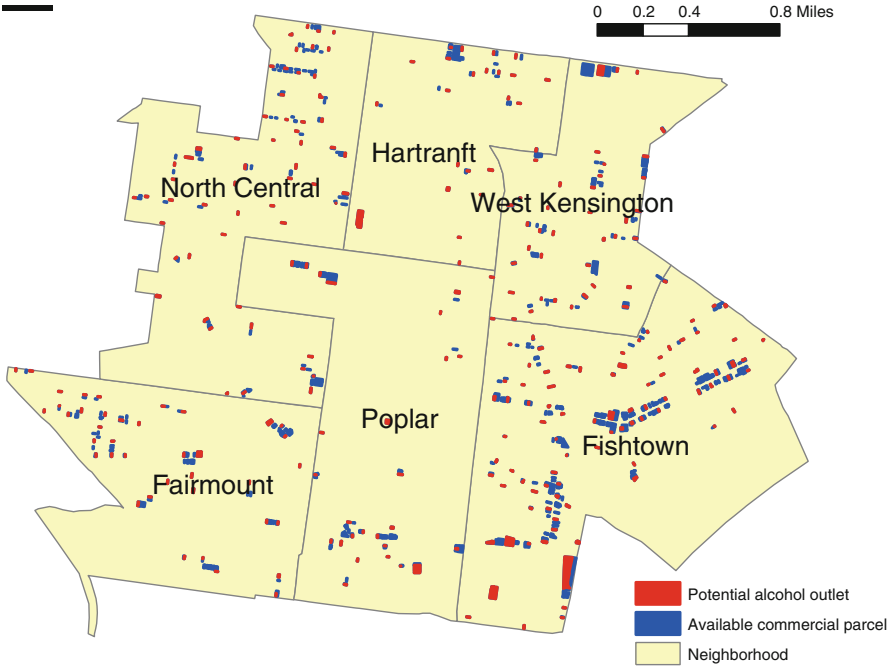

Fig. 6 Spatial configuration of potential alcohol outlets derived using the E-ACLP I (270 outlets)

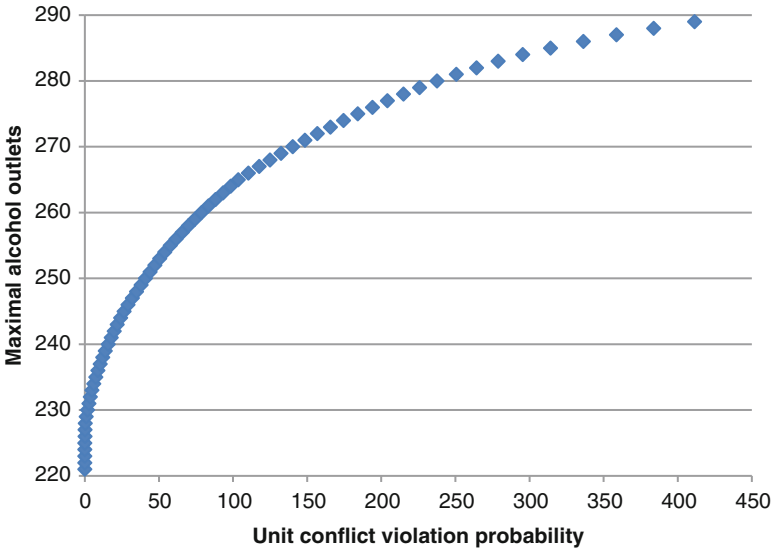

Fig. 7 Tradeoff curve for the E-ACLP II

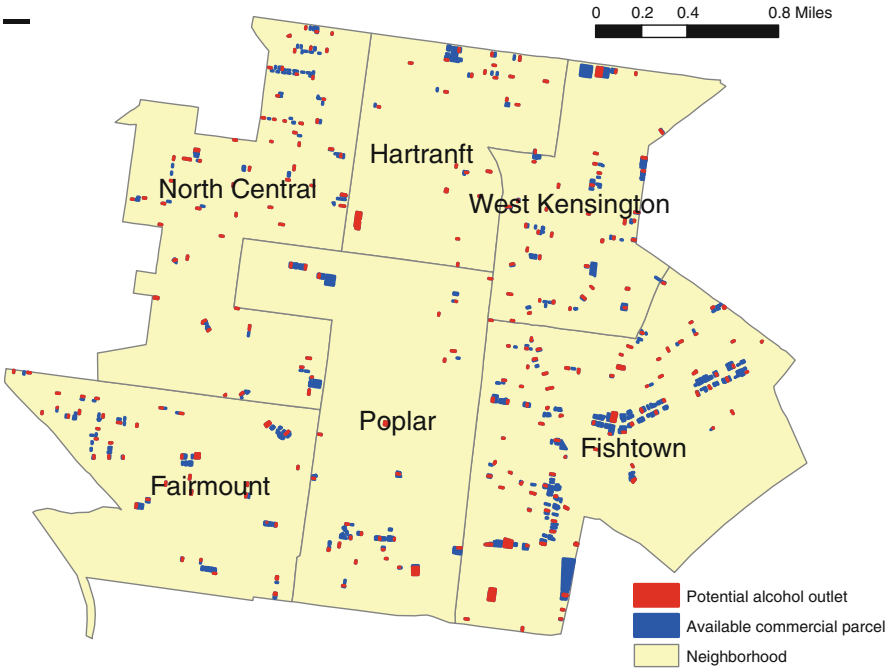

Fig. 8 Spatial configuration of potential alcohol outlets derived using the E-ACLP II (270 outlets)

different numbers of certain and uncertain conflicts are identified, changing the constraining conditions. For example, 3452 known and 820 potential conflicts are derived using the simplistic approach, but 3408 known and 1601 potential conflicts exist using the probabilistic approach. More conflicts are considered uncertain when conflict violation probabilities are utilized. The implications are significant, as the incorporation of a richer interpretation of error through probabilistic means results in a broader range of variability in the number of facilities and the associated spatial configurations that emerge.

This chapter detailed approaches to address spatial uncertainty in dispersion models by developing the conflict violation probability for use in the E-ACLP I and II models. The application results demonstrated that spatial data uncertainty significantly affects dispersion modeling results. The integration of spatial uncertainty makes analysis more complicated, given that a range of trade-off outcomes are possible; but they remain finite and can be evaluated to assess socio-economic implications. In the context of alcohol beverage control, the potential increase of alcohol outlets under privatization must be considered in the context of a known correlation with higher crime rates. The increase in outlets should be investigated further to determine the potential for exacerbating crime in neighborhoods already plagued by its effects. For example, increasing the mandated separation distance between outlets could reduce alcohol sales per area; though, again, further investigation is

needed to determine whether this strategy may be effective in avoiding possible increases in crime rates. The models detailed here enable the effects of policies to be illustrated while explicitly taking into account spatial data uncertainty, and provide a mechanism for initial investigation into associated issues.

Acknowledgements The first author would like to acknowledge support through the 2012-13 Benjamin H. Stevens Graduate Fellowship in Regional Science as well as an Arizona State University Graduate College Completion Fellowship. This material is also based upon work supported by the National Science Foundation under grants 0924001 and 0922737. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Aerts, J. C. J. H., Goodchild, M. F., & Heuvelink, G. (2003). Accounting for spatial uncertainty in optimization with spatial decision support systems. *Transactions in GIS, 7*(2), 211–230.

Barahona, F., Weintraub, A., & Epstein, R. (1992). Habitat dispersion in forest planning and the stable set problem. *Operations Research, 40*, 14–21.

Church, R. L., & Garfinkel, R. S. (1978). Locating an obnoxious facility on a network. *Transportation Science, 12*(2), 107–118.

Church, R. L., & Murray, A. T. (2009). *Business site selection, location analysis, and GIS*. New York: Wiley.

Curtin, K. M., & Church, R. L. (2006). A family of location models for multiple-type discrete dispersion. *Geographical Analysis, 38*(3), 248–270.

Downs, J. A., Gates, R. J., & Murray, A. T. (2008). Estimating carrying capacity for sandhill cranes using habitat suitability and spatial optimization models. *Ecological Modelling, 214*(2-4), 284–292.

Erkut, E., & Neuman, S. (1989). Analytical models for locating undesirable facilities. *European Journal of Operational Research, 40*(3), 275–291.

Erkut, E., ReVelle, C., & Ülküsal, Y. (1996). Integer-friendly formulations for the r-separation problem. *European Journal of Operational Research, 92*(2), 342–351.

FGDC. (1998) *Geospatial positioning accuracy standards*. National standard for spatial data accuracy. Available at http://www.fgdc.gov/standards/projects/FGDC-standards-projects/accuracy

Gamarnik, D., & Goldberg, D. A. (2010). Randomized greedy algorithms for independent sets and matchings in regular graphs: Exact results and finite girth corrections. *Combinatorics, Probability and Computing, 19*(01), 61–85.

Goodchild, M. F. (1993). Data models and data quality: Problems and prospects. In M. F. Goodchild, B. O. Parks, & L. T. Steyaert (Eds.), *Visualization in geographical information systems* (pp. 141–149). New York: Wiley.

Goodchild, M. F., & Gopal, S. (1989). *The accuracy of spatial databases*. London: Taylor&Francis.

Goodchild, M. F., Shortridge, A. M., & Fohl, P. (1999). Encapsulating simulation models with geospatial data sets. In K. Lowell & A. Jaton (Eds.), *Spatial accuracy assessment: Land information uncertainty in natural resources* (pp. 123–129). Ann Arbor: Ann Arbor Press.

Goovaerts, P. (2006). Geostatistical analysis of disease data: Visualization and propagation of spatial uncertainty in cancer mortality risk using Poisson kriging and p-field simulation. *International Journal of Health Geographics, 5*(1), 7.

Goycoolea, M., Murray, A. T., Barahona, F., Epstein, R., & Weintraub, A. (2005). Harvest scheduling subject to maximum area restrictions: Exploring exact approaches. *Operations Research, 53*(3), 490–500.

Greenwalt, C.R., & Shultz, M.E. (1962). *Principles of error theory and cartographic applications*. ACIC Technical Report (Vol. 96), St. Louis, MO.

Griffith, D. A., Millones, M., Vincent, M., Johnson, D. L., & Hunt, A. (2007). Impacts of positional error on spatial regression analysis: A case study of address locations in Syracuse, New York. *Transactions in GIS, 11*(5), 655–679.

Grubesic, T., & Murray, A. (2008). Sex offender residency and spatial equity. *Applied Spatial Analysis and Policy, 1*(3), 175–192.

Grubesic, T. H., & Pridemore, W. A. (2011). Alcohol outlets and clusters of violence. *International Journal of Health Geographics, 10*(1), 30.

Grubesic, T. H., Murray, A. T., Pridemore, W. A., Philip-Tabb, L., Liu, Y., & Wei, R. (2012). Alcohol beverage control, privatization and the geographic distribution of alcohol outlets. *BMC Public Health, 12*, 1015.

Gurobi. (2017). http://www.gurobi.com

Heuvelink, G. B. M. (1998). *Error propagation in environmental modelling with GIS*. London: Taylor and Francis.

Heuvelink, G. B. M., Peter, A. B., & Stein, A. (1989). Propagation of errors in spatial modelling with GIS. *International Journal of Geographical Information System, 3*(4), 303–322.

Hochbaum, D. S., & Pathria, A. (1997). Forest harvesting and minimum cuts: A new approach to handling spatial constraints. *Forest Science, 43*(4), 544–554.

Hunter, G. J., & Goodchild, M. F. (1996). Communicating uncertainty in spatial databases. *Transactions in GIS, 1*(1), 13–24.

Kuby, M. J. (1987). Programming models for facility dispersion: The p-dispersion and maximum dispersion problems. *Geographical Analysis, 19*(4), 315–329.

Leung, Y., & Yan, J. (1998). A locational error model for spatial features. *International Journal of Geographical Information Science, 12*(6), 607–620.

Longley, P. A., Goodchild, M. F., Maguire, D., & Rhind, D. (2015). *Geographic information science and systems* (4th ed.). London: Wiley.

Malizia, N. (2013). The effect of data inaccuracy on tests of space-time interaction. *Transactions in GIS, 17*(3), 426–451.

Mathai, A. M., & Provost, S. B. (1992). *Quadratic forms in random variables: Theory and applications*. New York: M. Dekker.

Moon, I. D., & Chaudhry, S. S. (1984). An analysis of network location problems with distance constraints. *Management Science, 30*, 290–307.

Murray, A. T. (1999). Spatial restrictions in harvest scheduling. *Forest Science, 45*(1), 45–52.

Murray, A. T. (2003). Site placement uncertainty in location analysis. *Computers, Environment and Urban Systems, 27*(2), 205–221.

Murray, A. T. (2010). Advances in location modeling: GIS linkages and contributions. *Journal of Geographical Systems, 12*(3), 335–354.

Murray, A. T., & Church, R. L. (1997). Facets for node packing. *European Journal of Operational Research, 101*(3), 598–608.

Murray, A. T., & Kim, H. (2008). Efficient identification of geographic restriction conditions in anti-covering location models using GIS. *Letters in Spatial and Resource Sciences, 1*(2–3), 159–169.

Murray, A. T., Church, R. L., Gerrard, R. A., & Tsui, W. S. (1998). Impact models for siting undesirable facilities. *Papers in Regional Science, 77*(1), 19–36.

Pennsylvania Liquor Control Board (PLCB). (2008). Information booklet for retail license, LCB-119.

Tong, D., & Murray, A. T. (2012). Spatial optimization in geography. *Annals of the Association of American Geographers, 102*(6), 1290–1309.

Wei, R., & Murray, A. T. (2012). An integrated approach for addressing geographic uncertainty in spatial optimization. *International Journal of Geographical Information Science, 26*(7), 1231–1249.

Wei, R., & Murray, A. T. (2015). Spatial uncertainty in harvest scheduling. *Annals of Operations Research, 232*(1), 275–289.

Yao, J., & Murray, A. T. (2013). Continuous surface representation and approximation: Spatial analytical implications. *International Journal of Geographical Information Science, 27*(5), 883–897.

Zeller, R. E., Achabal, D. D., & Brown, L. A. (1980). Market penetration and locational conflict in franchise systems. *Decision Sciences, 11*(1), 58–80.

The Nearest Neighbor Ant Colony System: A Spatially-Explicit Algorithm for the Traveling Salesman Problem

Jean-Claude Thill and Yu-Cheng Kuo

Abstract Inspired from the behavior of real ants, the ant colony algorithm has provided a new approach for solving discrete optimization problems, such as the traveling salesman problem. However, traditional ant colony algorithms may consume an inordinate amount of computing time to converge to a solution. In this chapter, a new heuristic algorithm called the nearest neighbor ant colony system (NNAC) is proposed in order to reduce computing time, without sacrificing on the optimality properties of the solutions. The NNAC is an intelligent form of the original ant colony system that follows a spatial strategy. Thanks to a search strategy that eliminates a large number of inefficient solutions up front on the basis of proximity-based neighborhoods, the NNAC is able to find the best solution in a fraction of the time that the conventional ant colony system consumes. The paper summarizes the principles of heuristics based on ant colony systems and highlighted some of their limitations. The proposed NNAC algorithm is presented in detail. The NNAC is tested on five different data sets and compared to a traditional ant colony system heuristic.

Keywords TSP • Traveling Salesman Problem • Ant colony system • Spatially explicit modeling

Introduction

The Traveling salesman problem (TSP) is one of the most notable operational models of spatial planning. It involves finding the shortest way of visiting each of a given set of locations (also known as cities) exactly once and returning to the

J.-C. Thill (✉)
Department of Geography and Earth Sciences, University of North Carolina, Charlotte, NC 28223, USA
e-mail: jean-claude.Thill@uncc.edu

Y.-C. Kuo
Department of Geography, University at Buffalo – The State University of New York, Buffalo, NY, USA

J.-C. Thill (ed.), *Spatial Analysis and Location Modeling in Urban and Regional Systems*, Advances in Geographic Information Science,
https://doi.org/10.1007/978-3-642-37896-6_13

starting point at the end of the tour. If travel cost is substituted for the distance between each pair of locations, the TSP is then to find the least-cost way of visiting all the locations.

The TSP has a number of real-world applications such as bus routing (Spada et al. 2005), delivery and repair vehicle routing (Weigel and Cao 1999), newspaper delivery (Song et al. 2002), business logistics (Exnar and Machac 2011), and individual trip planning (Thill and Thomas 1991). Because it is such a prominent problem of spatial planning and since many problems can be structured as a TSP on networks, the TSP is fairly widely implemented in Geographic Information Systems (Curtin et al. 2014). There are also a number of applications beyond these conventional applications. In industrial engineering and in sciences, the TSP is used to schedule a robot to drill holes in a circuit board (Ball and Magazine 1988), for genome sequencing (Johnson and Liu 2006), to design the layout of a satellite module (Sun and Teng 2003), optimal production sequencing (Jeong et al. 1997), as well as many other novel uses. Perhaps as important is that the TSP has become a fundamental platform for the study and assessment of general methods that can be applied to solve discrete optimization problems in different fields of research. There are several reasons for the TSP to play such an important role in combinational optimization. First, the TSP is a conceptually simple problem that turns out to be hard to solve because of being an NP-hard problem. Second, the TSP has no other additional constraints that are usually difficult to control in practice (Hoos and Stützle 2005). Finally, the TSP is the substructure of many other problems arising in real-life practical situations (Christofides 1979).

Many approaches have been developed to solve the TSP (Rego et al. 2011). Recently, ant colony algorithms have been proposed as a new heuristic approach for solving discrete optimization problems in general, and the TSP in particular. These algorithms simulate the collective intelligence of ants living in the same colony, particularly their foraging behavior. Although the Ant Colony System (ACS) and its variants have been successfully applied to various optimization fields, it does come with major drawbacks. Most significantly, when the number of cities to visit becomes large, huge computation time and failure to find an optimal solution at convergence become problematic. This chapter aims at improving the performance of the ACS with respect to these two key considerations by proposing enhancements that take advantage of the known spatial structure of the set of cities to be visited. An overview of the TSP is provided in the second section, while the principles of the Ant Colony Algorithm are presented in the third section. The enhanced ACS algorithm, dubbed the Nearest Neighbor Ant Colony (NNAC) algorithm, is presented next. The implementation of the NNAC algorithm is then pursued on Oliver-30, a dataset commonly used in testing and benchmarking described in Whitley et al. (1989), as well as four other derivative datasets. This provides the test bed for comparing the original ACS and the new NNAC algorithm. Conclusions are drawn at the end of the chapter.

The Traveling Salesman Problem

The traveling salesman problem (TSP) was first treated mathematically in the special case of the so-called Hamiltonian circuit in the 1800s by the Irish mathematician William R. Hamilton and by the British mathematician Thomas P. Kirkman. The general form of the TSP was first discussed in a series of publications by K. Menger in the late 1920s and the computational complexity of solutions to the TSP was addressed by H. Whitney in a 1934 seminar talk at Princeton University (Flood 1956; Schrijver 2005).

The TSP is known to be a NP hard problem for which no known efficient algorithm exists. The computational time to solve NP hard problems increases exponentially with problem size. Compared with any polynomial time problem whose computational time increases as N^2, where N is the problem size, the NP hard problem increases as 2^N instead. The TSP is very difficult to solve optimally due to its combinatorial complexity.

Since Dantzig et al. (1954) introduced the technique of plane cutting in integer programming, several modern and high-performance techniques have been developed to find the exact TSP solution. These include the branch and cut method (Padberg and Rinaldi 1991; Applegate et al. 2006), the branch and bound method (Little et al. 1963; Held and Karp 1970) and a few others. With enhanced linear programming-based techniques and greater computing power, problems of increasingly large sizes have been solved optimally, from Dantzig et al.'s 49 German cities in 1954 to Applegate et al.'s 85,900-city instance solved by the CONCORDE branch-and-cut algorithm (Applegate et al. 2006).

Given the large computing time of exact algorithms, practical TSP solution procedures are necessarily heuristic (Curtin 2007). Various approximation methods, whose rate of growth of computation time is a low order polynomial in n, have been experimentally observed to perform well (Christofides 1979; Curtin et al. 2014). The nearest neighbor heuristic (Hoos and Stützle 2005) is a form of greedy algorithm. Other approaches include simulated annealing (Meer 2007), Tabu search (Gendreau et al. 1994), and the Lin-Kernihan (LK) heuristic (Lin and Kernihan, Lin and Kernighan 1973). The LK algorithm is a particularly prominent TSP solution method. It is a tour improvement method that attempts to improve a given graph by exchanging a set of edges in order to obtain an alternative graph of lower cost (Applegate et al. 2006). The basic concept of the LK algorithm is based on the 2-opt moves method (Flood 1956).

Several heuristic methods based on a biological metaphor hold great promises for solving combinatorial optimization problems of the complexity exhibited by the TSP. The main approaches here are the genetic algorithm (GA), Neural Networks (NN), Particle Swarm Optimization (PSO), and Ant Colony Optimization. Some solution strategies resort to hybrid approaches where multiple heuristics are integrated (see for instance Marinakis et al. 2010; He and Mo 2011). The well-known approximate search technique of genetic algorithm (GA) simulates evolutional

mechanisms of genetic material such as selection, mutation and crossover to determine the better solution from its prior generation (Shengwu and Chengjun 2002). The GA has a better chance that some other algorithms to avoid getting trapped in the local minimum because the mutation operator alters one or more gene values in a chromosome and can result in entirely new genes being added to the gene pool; hence the algorithm has the opportunity to escape from the local minimum. Neural networks find an optimal TSP solution through learning by simulating the human nervous system. Unsupervised forms of neural networks such as Kohonen's self-organizing map or adaptive resonance theory (ART) (Goldstein 1990; Vishwanathan and Wunsch 2001; Mulder and Wunch 2003) iteratively fit a prototype tour to the set of nodes to the visited in the solution space. PSO has a population of candidate solutions (particles) that move in the search-space, guided by their own best known position as well as the entire swarm's best known position in this space (Cunkas and Ozsaglam 2009). Ant colony optimization is described in more detail in the following section.

Ant Colony Optimization

Ants are known to have developed evolved social strategies to efficiency find food supply and identify short paths between their nest and food sources. Inspired from the behavior of real ant colonies, Dorigo's (1992) Ant System (AS) is a heuristic method that simulates individual and collective behaviors of an ant colony. It is often regarded as one of the most advanced techniques for approximate optimization across diverse domains of application (Blum 2005). Since the early 1990s, the ant colony algorithm has been used to solve discrete optimization problems, such as the traveling salesman problem (Colorni et al. 1992; Dorigo et al. 1996; Dorigo and Gambardella 1997), vehicle routing problems (Reimann et al. 2004; Abousleiman et al. 2017), packet-switched communication network problems (Schoonderwoerd et al. 1997; Di Caro and Dorigo 1998), the network design problem (Poorzahedy and Abulghasemi 2005), land cover zoning and planning (Li et al. 2011), and others.

Rather simple principles of behavior can explain not only the ability of ants to find the shortest path but also the ability to adjust to changes in their decision environment. Each ant has an inclination for certain chemical compounds called pheronomes, deposited on trails by other ants. When ants arrive at decision nodes where they have to decide what direction to follow, they tend to follow the path that has been discovered by most ants, which means that stronger pheromone has been deposited on this path. A strong pheromone on the shorter path will obviously be created much faster than on longer paths. This will prompt an increasing number of ants to choose the shorter path until most ants have found the shortest path.

As a form of swarm optimization approach, ant colony heuristics assign a large number of virtual ant agents to the task of exploring many possible paths between

cities. As it builds a tour, each ant selects the next city to visit through a stochastic rule that features the amount of virtual pheromone deposited on the edges. The ants explore untraversed edges and in doing so, they deposit pheromone on edges until their tour is completed (local trail updating). Once all the ants have completed the shortest tour, virtual pheromone is deposited on the complete route (global trail updating). The amount of pheromone is inversely proportional to the tour length.

The AS has been used in different optimization applications and results have shown this algorithm to be capable of finding optimal solutions. In addition, AS is easy to understand, to program, and to combine with other heuristic algorithms. However, is suffers from serious drawbacks. Most notably, it consumes large computation time and gets stuck on local minima rather easily. The algorithm routinely takes at least 1–2000 iterations of all virtual ants through a solution search to find an optimal result and, even then, the algorithm could still have reached stagnation behavior (local minimum) without finding a global optimum.

Different strategies have been proposed to remedy this deficiency. For instance, Dorigo et al. (1996) introduced the Elitist Ant System (EAS) in which improvement is achieved by providing additional reinforcement to the edges belonging to the best tour found since the start of the algorithm. Stützle and Hoos (1997) developed a Max-Min ant system which imposes constraints on the updating of pheromone laid on the trail in order to solve the stagnation situation: ants can only deposit pheromone on the edges which belong to the tour that outperforms others in or up to the current iteration. Wu et al. (Wu et al. 1999) applied mutation features to the ant colony algorithm in order to prevent the early stagnation behavior and to decrease the computational time. In the Rank-based Ant System (Bullnheimer et al. 1999) the amount of pheromone deposited on the edges of a tour depends on how well this tour ranks against others. Zhu and Yang (2004) proposed a dynamic pheromone updating approach, which ensures that every ant's contribution can be efficiently adopted during the search phase. Dorigo and Gambardella's (1997) Ant Colony System (ACS) differs from the AS by granting virtual ants more intelligent ability to not only exploit learned knowledge but also of exploring tours other than good tours identified hitherto in the algorithm.

The search strategy of ACS and other ant colony algorithms involves the maintenance of a candidate list of all nodes that remain to be visited. Virtual ants search every possible node on this list for their next position, whether this is likely to be an efficient move or not. This strategy causes computation time to be wasted on searching inefficient candidates. Inefficient candidates are those nodes selected by ant agents at their current position that will, by no means, become optimal solutions. Furthermore, because pheromone is also deposited when inefficient nodes are visited, the probability that other ants choose the same route is raised, which may induce the algorithm to step into a stagnation situation without finding the best route, or lead to search results that are randomly scattered because ant agents are misled by the pheromone.

The Nearest Neighbor Ant Colony Algorithm

The Principle

As mentioned previously, the ant colony algorithms search every possible node listed in the candidate list. Thus, for a 30-node problem, 30! possible combinations are searched in order to find an optimal solution. In fact, most of the computation time involved in finding a solution may be wasted on node combinations that are not efficient. The Nearest Neighbor Ant Colony (NNAC) algorithm is proposed to improve computation time that is wasted on searching the inefficient combinations and to avoid early stagnation.

Careful observation of the Traveling Salesman Problem and of its solutions in the Euclidean plane indicates that optimal solutions commonly share a few fundamental properties. These properties are as follows:

1. Every node is linked by a solution edge to one of its nearest neighbors.
2. Outlier nodes, which are relatively far away from any others in the solution space, are not placed at the end of the visiting node sequence.
3. No two edges on a solution tour intersect, thus creating a planar solution.

These fundamental properties are exploited here in order to enhance the efficiency of ant colony algorithms. The proposed Nearest Neighbor Ant Colony (NNAC) algorithm incorporates a new searching strategy for the TSP that takes into account the spatial arrangement of cities across the study region. The algorithm responds to the spatial clumpiness of cities in the search for better paths traversing and connecting cities. It modifies the ant colony system to exhibit the properties described above so as to converge faster and thus reduce computing time. In addition, the NNAC implements a mutation function (Wu et al. Wu et al. 1999) that is intended to minimize the chance of stagnation on a local optimum. Some similar considerations drawn from local measurement of nearest neighbors were introduced by earlier researchers, in particular Doerner et al. (2002), Wetcharaporn et al. (2006), and Qi (2007). The novel structures of the proposed algorithm are discussed in more detail in the rest of this section.

The Nearest Neighbor Searching Strategy

The ACS algorithms sequentially search through many possible solutions (although not necessarily all of them); however, most of these solutions can readily be dismissed for being obviously suboptimal. In line with the first solution principle introduced above, the proposed NNAC algorithm focuses on nodes within a certain distance threshold r from current node i, instead of searching all nodes in the candidate list as a TSP tour is being built. The nearest candidate list, NNAC-candi,

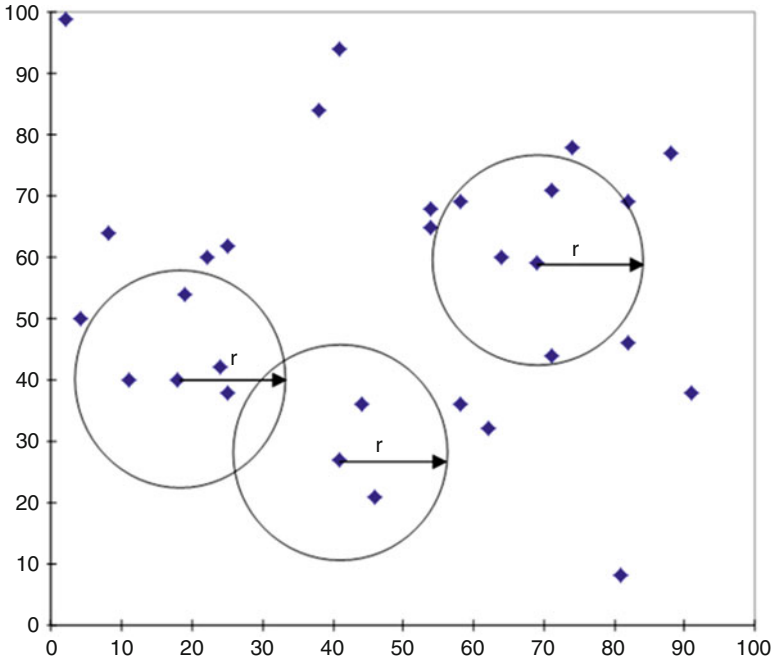

Fig. 1 The NNAC searching strategy

formed of nodes within the preset threshold eliminates most inferior nodes, which can reduce computation time to a great degree.

In contrast to the ACS algorithms that save every node that has never been visited to the candidate list, the NNAC-candi list of the Nearest Neighbor Ant Colony (NNAC) only saves a limited number of potential nodes which are located within a certain distance r to the ant's current location, as depicted in Fig. 1. It is a dynamic concept as it needs to be recomposed after each move on the tour by identifying all nodes encompassed by the moving circular window (buffer) of radius r. Practically, a sorted list of neighbors for all nodes can be computed in a pre-processing step and only those that are not tabu must be considered and placed in the NNAC-candi list. This search strategy takes into consideration the spatial process that generated the nodes to be visited. As a result, the number of nodes on an NNAC-candi list depends on the spatial structure of nodes in the vicinity of the current node: a local cluster produces a long list, while a sparse layout of competing and mutually avoiding nodes yields a short list. The search radius r is a critical parameter whose value directly impacts on the computation time as well as the optimality of the solution. The conventional candidate list is a limit case of the NNAC-candi list. The probability of choosing next node j from the NNAC-candi list, given the current node, follows the original ACS approach. It should be noted that parameter r must

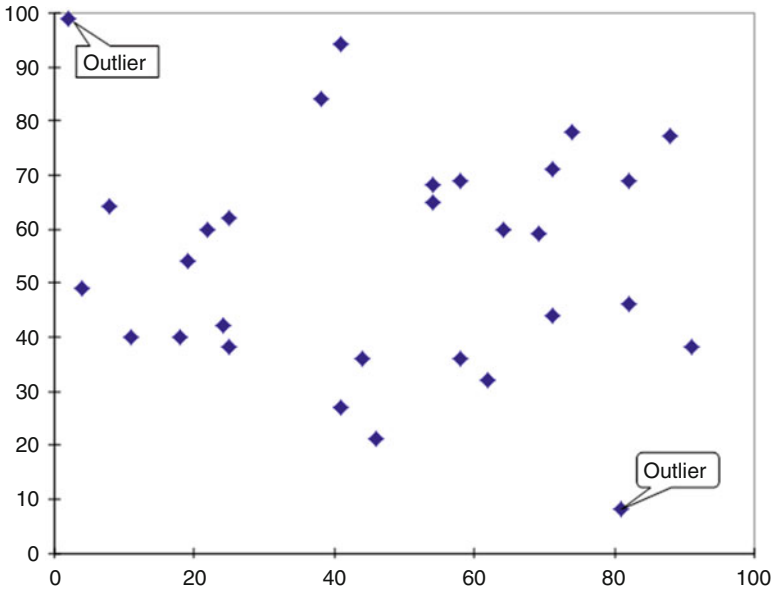

Fig. 2 Outlier nodes

be carefully selected to avoid: too small a value may exclude an optimal or near-optimal solution as the NNAC-candi list may be very short (and possibly empty), while a large r results in no computational savings.

The Problem of Outlier Nodes

Let us consider the nodes that are relatively far away from any others. The probability of selecting any of these outliers (Fig. 2) until close to the end of a tour is rather low since the distance between pairs of nodes is a controlling parameter of the probability function that drives the ACS algorithm. Therefore, in solutions produced by an ACS algorithm, outliers are usually listed at the end of the visiting list, which is at variance with known optimal solutions (second principle). Hence, whether an outlier is correctly placed in the visiting nodes sequence or not during the searching procedure is a critical criterion for finding an optimal solution. In order to enhance the proposed algorithm so that outliers are not ignored, an outlier-first searching strategy is deployed. Instead of randomly selecting a start node as in ACS, outliers are here given higher priority of being processed.

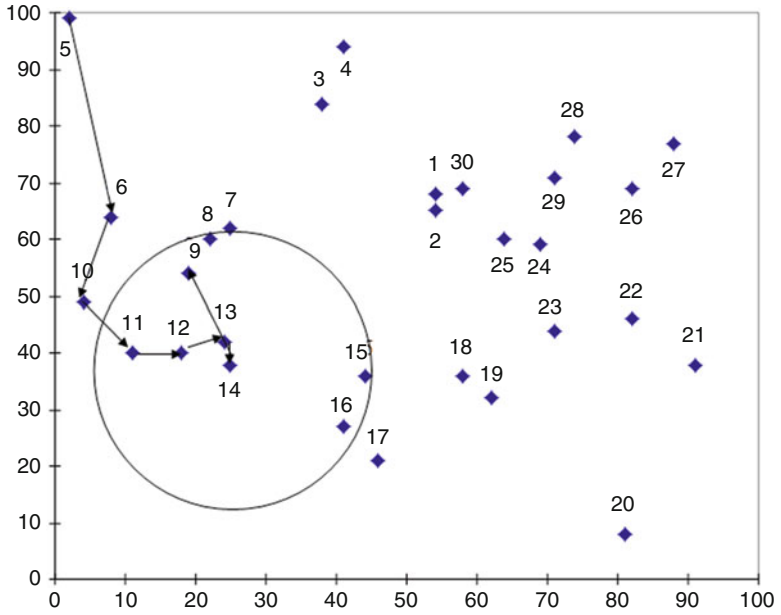

Fig. 3 Planar searching strategy

Planar Searching Strategy

Let us consider the scenario depicted by Fig. 3, where an ant agent starts from node 5. According to the NNAC, this ant chooses the sequence of nodes 6, 10, 11, 12, 13, 14, and so on. Up to this point, the searching sequence has followed a nearest neighbor searching strategy, according to which each new edge connects a node to its nearest neighbor. Let us now look at the NNAC-candi list of an agent arrived at node 14; this list includes nodes 9, 15, 16, and 8. According to the probability function controlling edge formation, the ant agent chooses node 9 as its next destination. This selection satisfies the searching strategy rule that the node is within searching distance r. However, the final result will not be the optimal route because it violates a fundamental property of TSP solutions, namely that segments $\overline{12,13}$ and $\overline{9,14}$ intersect. In order to reduce the likelihood that this property is violated, an improved pheromone updating rule is proposed.

Two pheromone updating strategies are applied in the proposed algorithm to represent the reinforcement mechanism that happens when agents explore routes and find good solutions. One is a local updating rule, while the other is an updating rule that takes into account the best tour found by an agent during each iteration. The latter replaces the global updating rule used in the ACS algorithm. Both rules

contribute to reinforcing the ant agents' learning from their prior action and to speeding the convergence of the algorithm, while preserving the desired properties of TSP solutions.

Local Updating Rule

In the NNAC algorithm, the pheromone density value on each segment is updated right after an ant agent passes through. However, instead of adding a constant amount of pheromone on each segment as in the ACS, the amount decreases according to how far the ant agent has traveled so far. It is consistent with the expectation that the significance of information for making an optimal decision does not remain constant as an agent progresses along the route: information is more important at an early stage than in later phases. The route choices made early in the search are more important than those made later because each selection has a compounding impact on subsequent selections. A correct decision at an early stage will lead to a higher chance that the ant will make correct decisions at a later stage. When a selection is made at a more advanced stage of the tour, the remaining choices are so limited that its impact on the entire optimization becomes relatively minor. Meanwhile, the amount of pheromone also depends on whether the ant's searching strategy violates optimal route properties. For instance, if an ant agent picks its next node in a way that violates the properties of route planarity, the amount of pheromone deposited on this segment is discounted.

The local pheromone updating rule for any segment $\overline{ij}$is mathematically expressed as:

$$\tau_{ij}(t) = (1-\rho_1) \quad \tau_{ij}\,(t-1) + \Delta\tau_{ij}(t) \tag{1}$$

where t is the time index, such that one and only one edge is traversed per time unit, and $0 < \rho_1 < 1$ is a user-defined coefficient representing the rate at which pheromone evaporates between consecutive time periods t-1 and t. The term $\Delta\tau_{ij}(t)$ controls the deposit of pheromone during time period t. It lets ant agents learn the best action to perform in each possible state; it is defined as

$$\begin{cases} \Delta\tau_{ij}(t) = \frac{\sum \tau_{kl}(t)}{L} & \text{if segment } \overline{ij} \quad \in \textit{planar} \\ \Delta\tau_{ij}(t) = \frac{\sum \tau_{kl}(t)}{L} \quad \delta & \text{otherwise,} \end{cases} \tag{2}$$

where L is the cumulative distance from start node s to current node i, the $\sum\tau_{ij}$ is the accumulated pheromone density from start node s to current node i, and $0 < \delta < 1$ is the scalar penalty assigned for violating the optimal route properties.

Iteration-Best Updating Rule

The iteration-best path is the shortest route of m runs (m ants) on any given iteration. The iteration-best updating rule is performed after all ants have completed their tour in the current iteration so as to reinforce the pheromone density on the segments belonging to the iteration-best route. By laying pheromone on the shortest route only, the NNAC iteration-best updating rule embodies the principle of "winner take all", which is intended to provide advantage to those segments which are identified as being part of the best combination.

The iteration-best rule for updating the pheromone value on segment $\overline{\text{ij}}$ is expressed mathematically as follows:

$$\tau_{ij}\,(t = n) = (1 - \rho_2)\ \ \tau_{ij}\,(t = 1) + \Delta\tau_{ij} \tag{3}$$

where t is the time index, $0 < \rho_2 < 1$ is a user-defined coefficient representing the rate at which pheromone evaporates between the start (t = 1) and the end (t = n) of the tour construction. The term $\Delta\tau_{ij}$ controls the deposit of pheromone on the best tout of the iteration, in inverse proportion of its length. It is defined as

$$\Delta\tau_{ij} = \begin{cases} (L_{iteration-best})^{-1} & \text{if } \overline{\text{ij}} \in \text{iteration best tour} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where $L_{iteration-best}$ is the length of the iteration's shortest tour.

Mutation

Like most heuristic methods, ant colony optimization has the disadvantage of sticking on local minima; as a result, it may not be able to find the globally optimal route. This situation can be avoided by randomly disturbing the pheromone density of certain segments by resetting or increasing their values. This behavior is similar to the mutation function in genetic algorithms (Lin et al. 1993). During the iteration-best pheromone updating phase, the mutation function randomly switches the order of two nodes on the iteration-best route, which will cause an incorrect update of the pheromone density and also give the NNAC algorithm the opportunity to find the real global shortest path, if it happens to have erred towards a local optimum.

Procedures of NNAC and Comparison with ACS

The NNAC heuristic involves the iteration through a number of steps, which parallel the original ACS algorithm. For sake of clarity, a run of the simulation is defined

as a complete route performed by a single ant agent. An iteration of an ant-based algorithm encompasses all the routing decisions made by all m ants. Therefore, in an m-ant simulation, an iteration has m runs and an ant run consists of n edges connecting nodes to form an ant's tour. A detailed description of ACS is available in Dorigo and Stützle (2004).

1. Choosing a start node

Instead of randomly selecting a start node from the set of nodes to be visited on a run as the original ACS does, the NNAC gives outliers more chances to be selected as a start node. For each odd run number, the NNAC randomly selects a start node from the outlier node pool. Conversely, on even runs, the start node is randomly selected from dataset as the traditional ACS does.

2. List of nodes that remain to be visited: NNAC-candi versus candidate list

For each ant k at node i, a NNAC-candi list is created. Only nodes within distance r from node i can be listed:

$$NNAC - candi_k = (j_1, j_2, j_3 \ldots j_n)\,,$$

where j_n is a node within distance r of node i that has not yet been visited. In the ACS algorithm, all nodes which have not been visited are listed in the candidate list:

$candi_k = (j_1, j_2, j_3 \ \ldots \ j_n)$, where j_n is a node that has not yet been visited.

3. Choosing next node j

The NNAC algorithm applies the same choice rule as the ACS for selecting the node to which an ant moves at a certain run, given the list of nodes that are yet to be visited. This rule is given by

$$j = \begin{cases} \arg\max\limits_{\substack{s \in candi_k(t) \\ s \in NNAC - candi_k(t)}} \left\{[\tau_{is}(t)]\,[\eta_{is}]^{\beta}\right\} & \text{if } q \le q_0 \\ S & \text{otherwise} \end{cases} \tag{5}$$

where q is a uniformly distributed random number defined on [0, 1], q_0 is a predefined parameter ($0 \le q_0 \le 1$), and S is the node that achieves the highest value on the following stochastic function:

$$p_{ij}^{k}(t) = \begin{cases} \dfrac{[\tau_{ij}(t)]^{\alpha}[\eta_{ij}]^{\beta}}{\sum\limits_{\substack{u \in candi_k(t) \\ u \in NNAC - tabu_k(t)}} [\tau_{iu}(t)]^{\alpha}[\eta_{iu}]^{\beta}} \\ \text{if } j \in candi_k(t) \text{ for ACS}, j \in NNAC - candi_k(t) \text{ for NNAC} \\ 0 \qquad \text{otherwise} \end{cases} \tag{6}$$

where $\tau_{ij}(t)$ is the density of pheromone on edge (i,j) during time period t, η_{ij} is the inverse value of distance between node i and node j, and α , β are user-defined parameters which represent the importance of pheromone and distance, respectively, in the selection of the next node to visit. The latter equation is in fact the node selection rule of the original AS heuristic.

4. Local pheromone updating rule

In the ACS and NNAC algorithms, the pheromone density on each segment is updated right after an ant agent passes through. However, while a constant amount of pheromone is added on each segment in the ACS, the amount decreases according to how far the ant agent has traveled so far. The NNAC local updating rule is given by Eqs. (1) and (2). In contrast, in the standard ACS heuristic, the latter is replaced by

$$\Delta\tau_{ij}^{k} = \begin{cases} \frac{Q}{L_k} & \text{if ant k uses path (i, j) in its tour} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where Q is a constant and L_k is the length of the tour of the k^{th} ant starting from its start node i, visiting all nodes, and returning to start node i.

5. Iteration-Best and global pheromone updating rule

In addition to the local updating rule, the NNAC performs an iteration-best pheromone updating rule at the end of each iteration so as to reinforce the pheromone density on the segments belonging to the iteration-best route. This rule is given by (3)–(4). It supersedes the ACS global pheromone updating rule, which is triggered at the end of each iteration when an ant k has identified the shortest tour during the searching period. The global pheromone updating rule is given by following equations:

$$\tau_{ij}(t = n) = (1 - \rho) \quad \tau_{ij}(t = 1) + \Delta\tau_{ij} \tag{8}$$

where

$$\Delta\tau_{ij} = \begin{cases} \left(L_{gb}\right)^{-1} & \text{if (i, j)} \in \text{global best tour} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

and ρ is the pheromone decay parameter and $L_{gb}$is the length of the best of all the tours produced by all m agents since the beginning of the iteration.

6. Mutation

If the iteration-best route returns the same value for a number X of consecutive iterations, two nodes in the iteration-best visiting list are randomly switched and the pheromone values are updated accordingly. Number X is a user-defined constant that controls the rate of mutation by setting the permissible number of identical consecutive solutions. No mutation is allowed in the ACS heuristic.

7. Additional iterations

Steps 1 through 6 are repeated for a predetermined number of iterations and the shortest routes obtained in each iteration are then compared to produce the TSP solution.

Experimentation Results

In this section, we implement and test the NNAC heuristic on Oliver-30 (Whitley et al. 1989), a 30-city test problem commonly used in the literature for validation and benchmarking purposes. The NNAC heuristic is also tested on four other randomly generated 30-node datasets for consistency analysis. The experimentation results and the analysis of consistency of applying the NNAC algorithm on different datasets are reported below.

NNAC Performance on Five Datasets

Ant population size is set at 6 for all of the following experiments.[1] The outlier nodes are identified by visually examining the different datasets. In addition, good values of parameters α and β are determined by testing multiple combinations, the best of which is reported below for each data set. The other user-defined parameters are set as follows:

$$\text{pheromone decay } (\rho) = 0.5$$

$$\text{segment intersected penalty } (\delta) = 0.7$$

$$q_0 = 0.5.$$

Figure 4 depicts the results of one out of 50 trials with 100 iterations conducted on each test dataset. The left-hand side of each panel shows a plot of the length of the best route found by ant agents in each iteration of the depicted trial; the right-hand side maps the route that has the shortest length on all 100 iterations. As the results on the five 30-node datasets indicate, the NNAC generates TSP routes that are reasonable. Furthermore, the algorithm rather quickly converges to minimum-length routes that dominate the solutions after 50 iterations or so.

[1] Our experiments reveal that smaller population sizes degrade the performance of the solution algorithm. Conversely, large sizes increase computational time with no performance benefits.

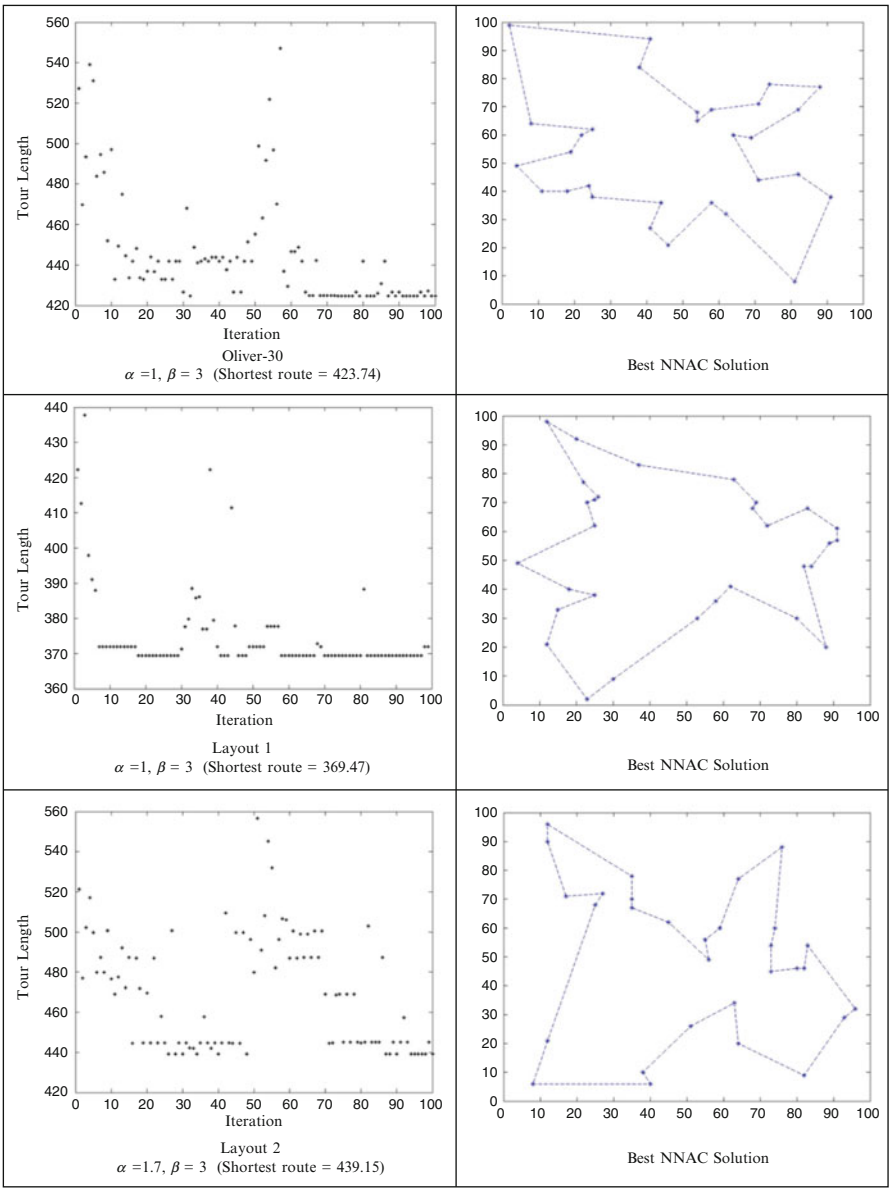

Fig. 4 NNAC results for five different city layouts

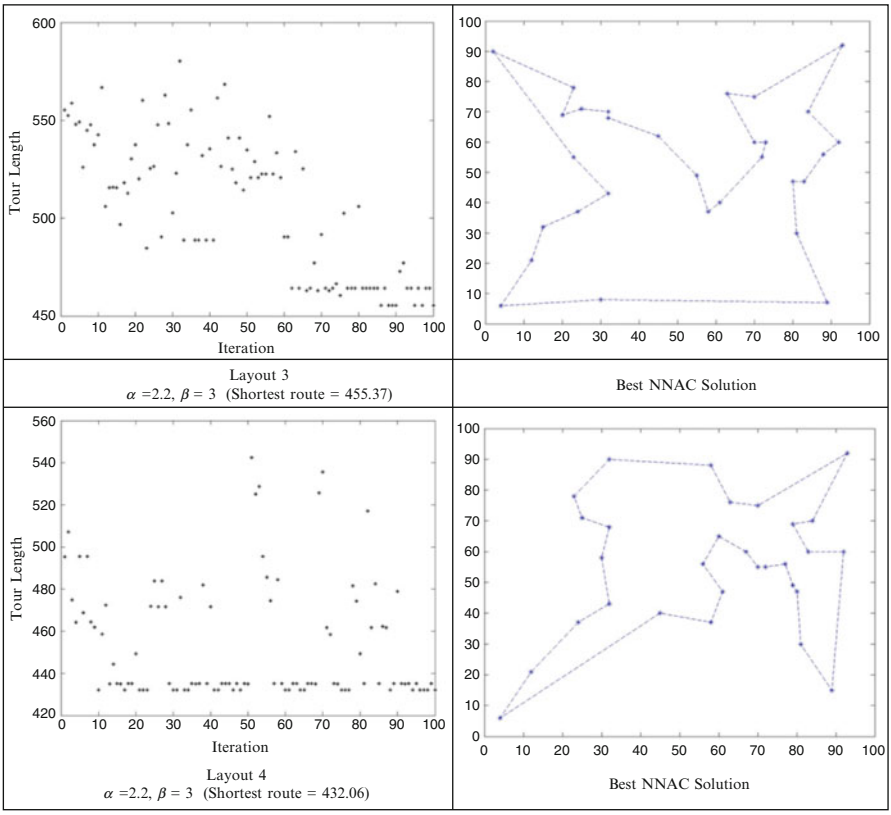

Fig. 4 (continued)

Consistency and Validation of NNAC Solutions

The Oliver-30 and four different 30-node layouts are also used to test the consistency of NNAC solutions. Results are obtained by running the NNAC algorithm on 50 trials; each trial includes 100 iterations with six ant agents. Table 1 reports the mean, standard deviation, and minimum of the best routes obtained over 50 trials. The average error is computed as the ratio between the mean of 50 best tour lengths and the length of the shortest tour. As shown in Table 1, the average error and the standard deviation are consistently small, which indicates that the performance of the NNAC algorithm is very consistent across all five test datasets.

While the NNAC heuristic is found to produce consistent TSP solutions, it is a worthy solution method only if it can be validated against known solutions. The best NNAC solution on Oliver-30 depicted in Fig. 4 (with a length of 423.74) is also found to be the best known solution of this test dataset (Whitley et al. 1989). No similar statement can be made for the other four test layouts as there are no known

Table 1 Summary statistics of NNAC best tour lengths for five test data sets

	Best tour length			
	Mean	Standard deviation	Minimum (best solution)	Average error (%)
Oliver-30	425.46	2.44	423.74	0.4
Layout 1	369.47	0.00	369.47	0.0
Layout 2	444.97	5.77	439.15	1.3
Layout 3	459.54	4.50	455.37	0.9
Layout 4	438.32	7.31	432.06	1.5

best solutions for these data sets. However, the best solutions attained in our tests can be benchmarked against known fundamental properties of optimal solutions reported in section "The Principle". All best solutions generated in our experiments meet the optimal properties mentioned which are:

1. Every node is linked by a solution edge to one of its nearest neighbors;
2. Outliers are correctly selected: the to-nodes of outliers are limited to their nearer neighbors;
3. No two link segments intersect on the route.

Furthermore, the NNAC heuristic is found to converge to a best solution exhibiting these desirable properties in 100 iterations or less (See Fig. 4). This represents a significant improvement in efficiency in solving the TSP with ant colony algorithms. In fact, the NNAC heuristic outranks the efficiency of the ACS algorithms by a large margin, since the latter typically need thousands of iterations to find the best solution. It has already been noted earlier that the standard deviations and average error percentages yielded by the NNAC algorithm on various data sets are small, which indicates that the solutions generated by the NNAC algorithm tightly cluster around the best value. These two considerations lead us to conclude that the NNAC heuristic is a reliable and robust solution approach to the TSP.

Experiment Comparison

The performance of the NNAC algorithm can also be assessed by comparing its solutions to those of the ACS on the same dataset. Dorigo and Gambardella (1997) conducted Ant Colony System (ACS) experiments with three different specifications of the local pheromone updating rule (1) on the Oliver-30 dataset. Their TSP solutions are reported here. The three ACS variants differ by the term $\Delta\tau_{ij}$ that controls the deposit of new pheromone during each time period. Specifically, in the simple ACS, $\Delta\tau_{ij} = \tau_0$, where τ_0 is a constant; the ant-Q updating rule uses $\Delta\tau_{ij} = \gamma \max \tau_{ij}$, where γ is a user-defined parameter; finally $\Delta\tau_{ij} = 0$ is assumed in a naïve ACS heuristic.

Table 2 Experiment comparison on Oliver-30

Heuristic	Tour length		
	Average	Standard deviation	Minimum
NNAC	**425.46**	**1.44**	**423.74**
Simple ACS	424.74	2.83	423.74
Ant-Q	424.70	2.00	423.74
Naïve ACS with $\Delta\tau(r,s)=0$	427.52	5.21	423.74

The ACS solutions with the three different pheromone updating rules serve to benchmark the NNAC solutions. The local updating and global updating rules in the NNAC are described by Eqs. (1–2) and (3–4), respectively. Table 2 shows the results obtained from Dorigo and Gambardella (1997) and NNAC.

According to the results shown in Table 2, the NNAC heuristic and the ACS with all three alternate specifications of $\Delta\tau_{ij}$can find the best TSP solution with a route length of 423.74. Although the simple ACS and Ant-Q approaches have better performance in term of average tour length, the lower standard deviation of the NNAC indicates that the NNAC algorithm more consistently finds a close to optimal solution than the variants of the ACS algorithm.

It should be pointed out that experiments with NNAC involve 25,100-iteration trials with 6 ants. On the other hand, results reported for the three ACS variants were generated by 25 2500-iteration trials with 10 ants. Therefore, the total number of ant runs in the latter experiments is 625,000, which contrasts with a mere 15,000 runs of the NNAC. The NNAC uses 2.4% of the ant runs of more traditional ACS algorithms. Remarkably, not only does the NNAC generate TSP solutions with less variance than the ACS algorithm, it also does so in a fraction of the computing time.

Conclusions and Future Enhancements

While ant colony heuristics have been successfully applied to various optimization problems such as the TSP, they suffer from huge computation time and from a tendency to settle on local optima at convergence. In this chapter, we have proposed several enhancements to the conventional ACS algorithms. The proposed NNAC algorithm capitalizes on the optimal route properties to reduce computing time without sacrificing on the optimality properties of the solutions. The performance of this algorithm was shown to be better than that of the ACS algorithm on computing time by eliminating inefficient routes in advance. Instead of searching every possible node, the NNAC searching strategy exploits the spatial structure existing among cities to be visited by focusing on nodes within a certain radius from the current node. Thus, the time saving comes from narrowing down the possible solutions from 30! to less than $5^{25} * 4!$ in a 30-node problem.

The NNAC also brings to bear prior knowledge on the spatial distribution of nodes in handling nodes which are relatively far away from the other nodes. These spatial outliers are largely ignored during the searching phase of ACS, which causes pheromone density to be inadequately updated and computing time to be wasted. In the NNAC algorithm, the randomly selected start node strategy is replaced by a conditional outlier-first strategy. The advantages of the conditional outlier-first strategy are:

1. An increase of the reliability of pheromone update rules;
2. A Reduction of the information noise generated by ignored outliers;
3. A reduction of the computing time.

Two enhanced pheromone updating rules are implemented in the NNAC. The local pheromone updating rule follows the traditional ACS approach but gives more intelligence while updating the pheromone density. In lieu of increasing the pheromone density uniformly, the NNAC local updating rule adds an amount according to the distance that an ant has traveled, which reinforces the importance of information received in the early searching stage. The NNAC local updating rule also assigns a penalty to non-planar link segments. The iteration-best updating rule, on the other hand, reinforces the information on the most efficient solution in each iteration, which reduces the probability of inefficient nodes being selected in the remaining iterations.

A mutation function is also implemented in the NNAC algorithm. The mutation function prevents the algorithm from being limited to a local minimum and falling in convergence without finding the global minimum. The mutation function in the NNAC is executed in the global pheromone update phase. By doing so, it gives the NNAC the chance to escape from a possible local minimum and to find the real global minimum value.

Currently the NNAC algorithm focuses on problem of small to moderate sizes. The NNAC features several solution strategies designed to consistently generate optimal solution properties. It provides an approach for eliminating inefficient node combinations in advance so as to reduce the computing time on the TSP problem. Thorough testing of the heuristic is in order to identifying best performing combination of parameters. Further parameter sensitivity analysis is needed to establish rules of thumb for the selection of appropriate parametric specifications. It is also anticipated that further enhancement of the efficient of the NNAC solution approach can be achieved by capitalizing to a greater extent on the spatial structure of nodes. Particularly, knowledge on the presence of local clusters of nodes can be exploited to avoid searching along dominated solution paths. It is our contention that the articulation of techniques of spatial point pattern analysis with ant colony optimization principles will lead to solution approaches suitable for large-scale TSP problems. Also, the algorithm should be tested and customized to problems on networks, which have become rather standard in an era where much geospatial data are available on transportation infrastructure and their operational properties, such as speed and capacity.

References

Abousleiman, R., Rawashdeh, O., & Boimer, R. (2017). Electric vehicles energy efficient routing using ant Colony optimization. *SAE International Journal of Alternative Powertrains, 6*(1).

Applegate, D. L., Bixby, R. E., Chvatal, V., & Cook, W. J. (2006). *The traveling salesman problem: A computational study*. Princeton: Princeton University Press.

Ball, M. O., & Magazine, M. J. (1988). Sequencing of insertions in printed circuit board assembly. *Operations Research, 36*, 192–201.

Blum, C. (2005). Ant Colony optimization: Introduction and recent trends. *Physics of Life Reviews, 2*, 353–373.

Bullnheimer, B., Hartl, R. F., & Strauss, C. (1999). A new rank based version of the ant system: A computational study. *Central European Journal of Operations Research, 7*, 25–38.

Christofides, N. (1979). The traveling salesman problem. In N. Christophides, A. Mingozzi, P. Toth, & C. Sandi (Eds.), *Combinatorial Optimization* (pp. 131–149). New York: Wiley.

Colorni, A., Dorigo, M., & Maniezzo, V. (1992). An investigation of some properties of an ant algorithm. In *Parallel problem solving from nature conference (PPSN 92)* (pp. 509–520). New York.

Cunkas, M., & Ozsaglam, M. Y. (2009). A comparative study on particle swarm optimization and genetic algorithms for traveling salesman problems. *Cybernetics and Systems, 40*, 490–507.

Curtin, K. (2007). Network analysis in geographic information science: Review, assessment, and projections. *Cartography and Geographic Information Science, 34*, 103–111.

Curtin, K., Voicu, G., Rice, M. T., & Stefanides, A. (2014). A comparative analysis of traveling salesman solutions from geographic information systems. *Transactions in GIS, 18*, 286–301.

Dantzig, G., Fulkerson, R., & Johnson, S. (1954). Solution of a large-scale traveling salesman problem. *Operations Research, 2*, 393–410.

Doerner, K. F., Gronalt, M., Hartl, R. F., Reimann, M., Strauss, C., & Stummer, M. (2002). Savings ants for the vehicle routing problem. In S. Cagnoni, J. Gottlieb, E. Hart, M. Middendorf, & G. R. Raidl (Eds.), *Applications of evolutionary computing* (pp. 11–20). Berlin/Heidelberg: Springer LNCS 2279.

Di Caro, G., & Dorigo, M. (1998). AntNet: Distributed stigmergetic control for communication networks. *Journal of Artificial Intelligence Research, 9*, 317–365.

Dorigo, M. (1992). *Optimization, learning and natural algorithms*. Unpublished Ph.D. dissertation, Dipartimento di Elettronica, Politecnico di Milano, Milan, Italy.

Dorigo, M., & Gambardella, L. (1997). Ant colony system: A cooperative learning approach to the travelling salesman problem. *IEEE Transactions on Evolutionary Computation, 1*, 53–66.

Dorigo, M., Maniezzo, V., & Colorni, A. (1996). Ant system: Optimization by a Colony of cooperating agents. *IEEE Transactions on System, Man, and Cybernetics—Part B: Cybernetics, 26*, 29–41.

Dorigo, M., & Stützle, T. (2004). *Ant colony optimization*. Cambridge, MA: MIT Press.

Exnar, P., & Machac, O. (2011). The travelling salesman problem and its application in logistic. *WEAS Transactions on Business and Economics, 18*, 163–173.

Flood, M. M. (1956). The traveling salesman problem. *Operations Research, 4*, 61–75.

Gendreau, M., Hertz, A., & Laporte, G. (1994). A Tabu search heuristic for the vehicle routing problem. *Management Science, 40*, 1276–1290.

Goldstein M (1990) Self-organizing feature maps for the multiple traveling salesman problem (MTSP). In: *Proceedings IEEE international conference on neural networks*, Paris, pp. 258–261.

He, X. B., & Mo, Y. W. (2011). Solving the TSP by simulated annealing genetic algorithm based on Google maps JavaScript API. *Advanced Materials Research, 201–203*, 733–737.

Held, M., & Karp, R. M. (1970). The traveling salesman problem and minimum spanning trees. *Operations Research, 18*, 1138–1162.

Hoos, H. H., & Stützle, T. (2005). *Stochastic local search: Foundations and applications*. San Francisco: Morgan Kaufmann.

Jeong, E. Y., SC, O., Yeo, Y. K., Chang, K. S., Chang, J. Y., & Kim, K. S. (1997). Application of traveling salesman problem (TSP) for decision of optimal production sequence. *Korean Journal of Chemical Engineering, 14*, 416–421.

Johnson, O., & Liu, J. (2006). A traveling salesman approach for predicting protein functions. *Source Code for Biology and Medicine, 1*, 3.

Li, X., Lao, C. H., Liu, X. P., & Chen, Y. M. (2011). Coupling urban cellular automata with ant Colony optimization for zoning protected natural areas under a changing landscape. *International Journal of Geographical Information Science, 25*, 575–593.

Lin, S., & Kernighan, B. W. (1973). An effective heuristic algorithm for the traveling-salesman problem. *Operations Research, 21*, 498–516.

Lin, F. T., Kao, C. Y., & Hsu, C. C. (1993). Applying the genetic approach to simulated annealing in solving some NP-hard problems. *IEEE Transactions on Systems, Man and Cybernetics, 23*, 1752–1767.

Little, J. D. C., Murty, K. G., Sweeney, D. W., & Karel, C. (1963). An algorithm for the traveling salesman problem. *Operations Research, 11*, 972–989.

Marinakis, Y., Marinaki, M., & Dounias, G. (2010). A hybrid particle smarm optimization algorithm for the vehicle routing problem. *Engineering Applications of Artificial Intelligence, 23*, 463–472.

Meer, K. (2007). Simulated annealing versus metropolis for a TSP instance. *Information Processing Letters, 104*, 216–219.

Mulder, S. A., & Wunch, D. C. (2003). Million city traveling salesman problem solution by divide and conquer clustering with adaptive resonance neural networks. *Neural Networks, 16*, 827–832.

Padberg, M., & Rinaldi, G. (1991). A branch-and-cut algorithm for the resolution of large-scale symmetric traveling salesman problems. *SIAM Review, 33*, 60–100.

Poorzahedy, H., & Abulghasemi, F. (2005). Application of ant system to network design problem. *Transportation, 32*, 251–273.

Qi, C. (2007). An ant colony system hybridized with randomized algorithm for TSP. In: *Eighth ACIS international conference on software engineering, artificial intelligence, networking, and parallel/distributed computing* (SNPD 2007), pp. 461–465.

Rego, C., Gamboa, D., Glover, F., & Osterman, C. (2011). Traveling salesman problem heuristics: Leading methods, implementations and latest advances. *European Journal of Operational Research, 211*, 427–441.

Reimann, M., Doerner, K., & Hartl, R. F. (2004). D-ants: Savings based ants divide and conquer the vehicle routing problems. *Computers & Operations Research, 31*, 563–591.

Schoonderwoerd, R., Holland, O., Bruten, J., & Rothkrantz, L. (1997). Ant-based load balancing in telecommunications networks. *Adaptive Behavior, 5*, 169–207.

Schrijver, A. (2005). On the history of combinatorial optimization (till 1960). In K. Aardal, G. L. Nemhauser, & R. Weismantel (Eds.), *Handbook of discrete optimization* (pp. 1–68). Amsterdam: Elsevier.

Shengwu, X., & Chengjun, L. (2002). A distributed genetic algorithm to TSP. In: *Proceedings of the 4th world congress on intelligent control and automation* (Vol. 3, pp. 1827–1830).

Song, S. H., Lee, K. S., & Kim, G. S. (2002). A practical approach to solving a newspaper logistics problem using a digital map. *Computers and Industrial Engineering, 43*, 315–330.

Spada, M., Bierlaire, M., & Liebling, T. M. (2005). Decision-aiding methodology for the school bus routing and scheduling problem. *Transportation Science, 39*, 477–490.

Sun, Z. G., & Teng, H. F. (2003). Optimal layout design of a Satellite Module. *Engineering Optimization, 35*, 513–529.

Stützle, T., & Hoos, H. H. (1997). MAX-MIN ant system and local search for the traveling salesman problem. In: *IEEE Int'l conference on evolutionary computation* (pp. 309–314). IEEE Press.

Thill, J. C., & Thomas, I. (1991). Towards conceptualizing trip-chaining behavior: A review. *Geographical Analysis, 19*, 1–17.

Vishwanathan, N., & Wunsch, D. C. II. (2001). ART/SOFM: A hybrid approach to the TSP. In: *Proceedings of neural networks* (Vol. 4, IJCNN '01, pp. 2554–2557). International Joint Conference, Washington, DC.

Wetcharaporn, W., Chaiyaratana, N., & Tongsima, S. (2006). DNA fragment assembly: An ant colony system approach. In F. Rothlauf et al. (Eds.), *Applications of evolutionary computing. EvoWorkshops 2006. Lecture notes in computer science* (Vol. 3907, pp. 231–242). Berlin/Heidelberg: Springer.

Weigel, D., & Cao, B. (1999). Applying GIS and OR techniques to solve Sears technician-dispatching and home delivery problems. *Interfaces, 29*, 112–130.

Whitley, D., Starkweather, T., & Fuquay, D. (1989). Scheduling problem and traveling salesman: The genetic edge recombination operator. In: *Proceedings of the third international conference on genetic algorithm*, Morgan Kaufmann, Palo Alto, CA, pp. 133–140.

Wu, Q. H., Zhang, J. H., & XH, X. (1999). An ant colony algorithm with mutation features. *Journal of Computer Research and Development, 36*, 1240–1245.

Zhu, Q. B., & Yang, Z. J. (2004). An ant colony optimization algorithm based on mutation and dynamic pheromone updating. *Journal of Software, 5*, 185–192.

Large-Scale Energy Infrastructure Optimization: Breakthroughs and Challenges of CO_2 Capture and Storage (CCS) Modeling

Jordan K. Eccles and Richard S. Middleton

Abstract Secure, sustainable, and cost-effective energy development will be one of the greatest global challenges in coming decades. This development will include an extensive range of energy resources including coal, conventional and unconventional oil and natural gas, wind, solar, biofuels, geothermal, and nuclear. CO_2 capture and storage (CCS) infrastructure is a key example; meaningful CCS in the US could involve capturing CO_2 from hundreds of CO_2 sources, including coal-fired and natural gas power plants, and transporting a volume of CO_2 greater than US oil consumption. Here, we highlight breakthroughs and future challenges for CCS infrastructure optimization and modeling. We start with the evolution of CCS infrastructure modeling from early attempts to represent the capture (sources), transport (network), and storage (sinks) of CO_2, through to the integration of more advanced spatial optimization (or location-allocation) approaches including mixed integer-linear programming. We then highlight key future challenges and opportunities, including the representation of significant uncertainties throughout the CCS supply chain and the ability to represent policy and business decisions into CCS infrastructure optimization. Finally, we examine the role that next-generation CCS infrastructure modeling can have in wider massive-scale energy network investments.

Keywords CO_2 capture and storage (CCS) • Spatial optimization • Uncertainty • Mixed integer-linear programs (MIP)

J.K. Eccles
Senior Analyst at Leidos, 11955 Freedom Drive Reston, VA 20190

R.S. Middleton (✉)
Los Alamos National Laboratory, Earth and Environmental Sciences, Los Alamos, NM 87545, USA
e-mail: rsm@lanl.gov

J.-C. Thill (ed.), *Spatial Analysis and Location Modeling in Urban and Regional Systems*, Advances in Geographic Information Science,
https://doi.org/10.1007/978-3-642-37896-6_14

Introduction

Secure, sustainable, and cost-effective energy development will be one of the greatest global challenges in coming decades. This development will include an extensive range of energy resources including coal, conventional and unconventional oil and natural gas, wind, solar, biofuels, geothermal, and nuclear. To meet global energy demands with these resources, energy infrastructure will have to be constructed on an unprecedented scale. In the US alone, this will include construction of perhaps hundreds of power plants, including wind and solar farms and several hundred thousand kilometers of pipelines and transmission lines. Integrating this diverse range of energy sources into an integrated, cost-effective system will take careful and comprehensive planning. Location-allocation modeling, or spatial optimization, should and will be a critical tool in this planning.

CO_2 capture and storage (CCS) is a climate change mitigation technology that reduces CO_2 emissions by capturing CO_2 at large stationary sources (e.g., coal-fired power plants), transporting the CO_2 in dedicated pipelines, and injecting and storing the CO_2 in geologic reservoirs (e.g., deep saline aquifers or depleted oil and gas reservoirs) (Middleton et al. 2012a; Stauffer et al. 2011). To have a meaningful impact, the US alone would have to install CO_2 capture infrastructure on hundreds of coal-fired and gas power plants as well as build a CO_2 pipeline network capable of transporting a volume of CO_2 greater than US oil consumption (Middleton et al. 2012b). Optimizing CCS infrastructure on this scale is large and complex problem, a problem well-suited to spatial optimization. For example, commercial-scale CCS involves capturing CO_2 from point sources, transporting the CO_2 over a network, and delivering the CO_2 to specific sinks. That is, this is a classic source-network-sink optimization problem.

Here, we explore and discuss the evolution of CCS infrastructure modeling, from early developments introduced by CCS researchers through to the current state-of-the-art led by experts in infrastructure modeling and optimization. The early researchers used their CCS technology expertise to identify that there was an important source-network-sink problem to solve, but did not necessarily have the required tools to adequately solve CCS problems. Later, researchers with operations research and optimization backgrounds were brought into the equation, establishing classic optimization approaches to the problems. And more recently, CCS infrastructure modeling has begun to mature to the stage where advanced concepts—such as system uncertainty, fluctuating CO_2 flows, and real-world policy—have been integrated into models. Specifically, we identify how spatial optimization has solved many research gaps in CCS modeling, but also discuss how CCS presents a unique set of challenges particularly in scale and uncertainty and that these issues must be resolved going forward. We also highlight how many of the breakthroughs and challenges developed by CCS infrastructure modeling can be solved by a flexible modeling approach and how the solutions developed for CCS modeling can be applied to energy infrastructure in general.

Early CCS Development

The IPCC's Special Report on CCS, published in 2005, illustrates the early emphasis on CO_2 capture technology and economics, with much less emphasis on CO_2 transport and storage (Metz et al. 2005), let alone any mention of integrated infrastructure modeling. This is of no surprise. For instance, the cost to capture CO_2 is still perhaps as much as an order of magnitude more expensive than either the transport or storage of CO_2, hence the early focus on CO_2 capture technology and economics. Arguably, this capture-focus remains today since capture costs still have the greatest flexibility for cost reduction. Similarly, without any large-scale CCS source-network-sink infrastructure in place (largely still true today), system CCS infrastructure modeling and optimization received scant attention. Although geological storage had been proposed in some detail over more than a decade before the IPCC report (Koide et al. 1992), later including storage below the ocean seafloor (Koide et al. 1997), very little advancement in economic characterization of geological storage had been made by the time the IPCC report was compiled and published. At the time, geological storage in deep saline aquifers was widely regarded as a sequestration option with the largest physical potential, a belief borne out by subsequent refinements of storage characterization, most notably the continuing efforts of the US Department of Energy to map out storage potential in North America (USDOE 2010).

The IPCC report devotes some attention to transport costs and very little to storage costs. Without widespread adoption of CCS technology at the time, and indeed with only perhaps a single large commercial CCS project using a pipeline of any length—Weyburn (Petroleum Technology Research Centre 2011)—little attention was paid to CO_2 transportation. Conversely, even without CCS in place, millions of tonnes of natural (i.e., non-anthropogenic) CO_2 have been transported large distances for several decades, taking CO_2 from natural geologic formations (such as salt domes) and injecting the CO_2 into depleted oil fields for enhanced oil recovery (EOR). Currently, EOR annually stores between 50 and 70 million tonnes (50–70 $MtCO_2$/year) in the US. However, transportation and integrated infrastructure modeling were not a high priority. In short, the IPCC report heavily concentrated on CO_2 capture technology, devoted some attention to geologic storage engineering and costs, but with almost no emphasis on CO_2 transportation and integrated CCS modeling.

The IPCC report relied on just three reports for CO_2 injection and storage costs: Bock (2002), Allinson et al. (2003a), and Hendriks and Bock (1993), a covering a broad geographic range (Australia, the United States, and Europe, respectively). In the US, Bock's report for the Electric Power Research Institute (with the US DOE) found ranges between \$0.40/tCO_2 and \$4.50/tCO_2. In Australia, Allinson et al.'s article in the APPEA Journal found a similar \$0.20 to \$5.10 per tonne (Allinson et al. 2003b). Hendriks et al. (2004), in a report for Ecofys with the Netherland's TNO, found a range between \$1.90 and \$6.20 for Europe. Offshore costs were slightly higher and had a wider range, and oil and gas fields were roughly the same (Metz et al. 2005).

Perhaps because the ranges were relatively narrow or because their magnitude was so small compared to the estimated cost of capture, most large models seemed to focus on variations in capture parameters as controlling the degree of penetration of CCS. It is in fact rare to find any kind of documentation of transport or storage costs in the stabilization scenario modeling that accompanies many IPCC reports (Bernstein et al. 2007) or in, for example, the US Energy Information Administration's National Energy Modeling system at the time (EIA 2007). Highly sophisticated techno-economic models for power plants focused their attention on capture (Rubin et al. 2005, 2007) since the characteristics of power plant/capture systems were easier to model with little computational effort.

The preferred method for incorporating transport and storage cost estimation into large scale modeling is well-illustrated by Bock's analysis of transport and storage cost for deep saline aquifers in the United States (Bock 2002). By 2000, the Bureau of Economic Geology at the University of Texas at Austin had compiled and digitized maps for 20 aquifers with information on over a dozen geological parameters, all of which might impact the storage capacity of the aquifer, the cost of injecting CO_2, or both (BEG 2000). Bock used this information to estimate the injection rate of CO_2 through a single well, which is an important value because it determines the number of wells that a sequestration site would require and the cost of drilling those wells is a major component of the overall cost of storage (Bock 2002). Although this analysis was almost certainly flawed—for example, it appeared to overestimate the injection rate by a considerable amount (Eccles et al. 2009)—it is certainly a good initial framework for estimating storage costs. The framework presented related total costs and injection rate, which continues to be a good way of representing the cost of CO_2 abatement in cost per tonne. Beyond this, however, the variation in geology and its effect on cost and capacity vanished when this component of the cost model was coupled to the transport component.

Similarly, the cost of transport was simplified drastically. Literature on transport costs at the time revealed that there might be a dramatic variation in the cost of transport depending on the scale at which CO_2 was transported (Kuby et al. 2011b) and the distance the pipeline had to cover (Chandel et al. 2010). Indeed, data on natural gas pipelines showed fairly clearly that although there was a great deal of variation in cost for short pipelines, cost of long pipelines was largely dependent on these two factors (Parker 2004). The transport component of models, then, could simply focus on arriving at two summary statistics for the pipeline network: the average length of pipeline and the average size of a source (or cluster of sources).

Bock's report is an excellent example of this method, in which the author estimates the cost of transport by determining an average distance and power plant size and determining the cost of transport from them (Bock 2002). This philosophy also appears in the IPCC report, where transport costs tend to be summarized by these two variables. Several figures demonstrate how far the understanding of source-sink matching has come just by the scale of their axes—plotting transport costs as a function of distance, the distance variable stretches up into the many thousands of kilometers, which now seems quite unlikely (Metz et al. 2005).

This approach very clearly discards the important fact that although one might summarize any pipeline network by the average size and length of pipelines, the development of that network will rarely include any pipelines that are actually at the average size and length. That is, most sources would face a cost to link to a sink that is different than the average, which might in turn affect their decision to link to that sink or to even capture in the first place. Critical in determining whether it is appropriate to distill the cost of transport (and storage) to average values is understanding the distribution of costs associated with transport. If the cost is not normally distributed or has a high variance, very little certainty is gained by reporting or using average figures in integrated modeling. The averaging approach was clearly meant to simplify first-order calculations in a complex problem, but could not survive very long without investigation as to the nature of variation in actual transport or storage costs.

Later investigations revealed variations in transport cost ranging over tens of dollars per tonne of CO_2, which is clearly large enough to impact decisions from pipeline routing and storage location all the way to the decision to capture itself, generally dominated by the cost of capture. Using an average length completely discards critical decision information and renders the results of these analyses suspect at best. Early transport cost figures did not include the variation in transport costs that would enable valuable conclusions to be drawn from CCS models.

This variation was clearly understood by the authors of the early studies, but they especially lacked a good framework for estimating how this variation would interact with the geographic and geologic variation in storage costs. This led to a combination of approaches in which each component was analyzed separately and a summarizing statistic or two was used to couple the components together. Thus, instead of interacting, the components could simply be added together to get the total cost of a CCS system.

Until recently, this was how major CCS modeling by the EPA (Dooley et al. 2008) and the EIA (2007) was conducted—the best understanding of a combined average cost of transport and storage was added to better-understood economics of capture and used to estimate the deployment or potential of CCS. This was not limited to government agencies or research organizations; various private estimates also took this approach (e.g., BCG 2008; McKinsey Climate Change Initiative 2008), since it had major advantages in terms of complexity, computational requirements, and data availability.

It was fairly clear, however, that this was not a permanent solution, even to the authors of the IPCC special report, who note that "[t]he full range of costs is acknowledged to be larger than shown" (Metz et al. 2005). While large-scale efforts to improve characterization of storage capacity focused on physical capacity (Ciferno et al. 2010), research into the impact of geological factors on storage cost by Eccles et al. found storage costs that ranged over several orders of magnitude just from the basic cost of injection, ranging from the sub-$\$1/tCO_2$ found by Bock (2002) all the way up to thousands of dollars per tonne (Eccles et al. 2009).

Such research paralleled decades-old conventional wisdom in the oil and gas industry as well as any other extractive industry: resources are rarely utilized at their

average cost, and the economic ease of extraction plays a major role in the resource utilization (Craft et al. 1991). Leaders in the CCS field called for more advanced characterization of storage sites, including the use of techno-economic models to evaluate realistic or economic storage potential (Bachu et al. 2007).

At the same time, modelers began to acknowledge (or advocate) that storage and transport costs could not be taken separately and just added together (Dooley et al. 2008). This was especially clear in the special case of geological storage with enhanced oil recovery. Although deep saline aquifers could (incorrectly) be imagined to be large, homogeneous bodies with similarly homogeneous characteristics, it was entirely obviously that oil fields are quite different from one another, and that not only would the costs of injecting CO_2 in oil fields vary based on the geology, the offsetting revenue from recovered oil would also vary (Nemeth et al. 2011).

Geographic Optimization and CCS

The problem, as presented starkly by EOR assessments, was an excellent opportunity for simple optimization: given various options for storage, each with a capacity and cost, and various agents who might utilize this storage, what configuration of source-sink matching would minimize total costs for the storage and transport system (Middleton et al. 2011)? This early optimization was pioneered by Dooley et al. (2004), who developed an optimization framework that could solve relatively simple versions of the source-sink matching problem. The approach represented a considerable advance over the uncoupled summary statistic one and represented an important move toward geographic optimization.

However, this approach still suffered from a number of drawbacks, possibly stemming from the research questions it may have been meant to solve. The authors devoted time and attention on enhanced oil recovery and geological characterization of the oil fields, but treated saline aquifers as economically homogeneous bodies (Dahowski and Dooley 2004). Although the documentation of the model is sparse, it appears that the cost of storage in saline aquifers for evaluations that are relatively massive in scale (i.e. all of North America and later China) is exactly the same over thousands of square kilometers of these aquifer's surface footprint (Dahowski et al. 2009; Dooley et al. 2004). The saline aquifers, moreover, completely dominate the source-sink matching, in which hundreds of sources in their analysis optimize simply the distance to the closest saline formation and nothing else, which renders the results largely unusable for policy planning.

At the same time, other modelers focused more heavily on understanding or integrating realistic costs of storage and, where possible, doing integrated source-sink matching with transport optimization (McCoy and Rubin 2008; Middleton and Bielicki 2009; Middleton et al. 2012d). After Dooley et al.'s evaluation of North America, but before Dahowski et al.'s evaluation of China, the EPA (2008) published a detailed estimate of roughly 80 cost categories for geological storage, many of which would depend on geological characteristics or project requirements

that stem from the characteristics such as the number of wells required. This allowed researchers to develop advanced semi-analytical models for geological storage, such the one underlying *SimCCS* (Middleton and Bielicki 2009). As more geological data became available, numerical simulations or models based on them, such as *CO_2-PENS* (Stauffer et al. 2009; Viswanathan et al. 2008), also became more prolific, allowing a much better characterization of geological potential.

At the same time, pipeline optimization and the methods associated with location optimization began to appear in CCS literature. An MIT project developed a cost weighting surface (Herzog et al. 2007), a necessary prerequisite for the cost distance/backlink calculations that form the basis of pipeline routing algorithms. *SimCCS* uses a similar but more detailed surface (Middleton and Bielicki 2009, 2013; Middleton et al. 2012d). Since then, a proliferation of transport models that utilize geographic optimization methods have emerged including Han and Lee (2011), Mendelevitch et al. (2010), Morbee et al. (2012), Kemp and Kasim (2010), and van den Broek et al. (van den Broek et al. 2010).

At this point, the development of transport and storage modeling is now mature enough to have completely integrated capture, transport, and storage components that consider all elements of the CCS system in combination (Keating et al. 2010). The results from integrated models have provided interesting perspectives on the potential of carbon capture and storage.

First, large-scale characterizations of geological storage potential indicate that there is plenty of storage capacity (USDOE 2010). The total technical potential may reach tens of billions of tonnes of storage space in North America alone, which is more than enough to sequester decades or centuries of emissions. This abundance, however, is not necessarily evenly distributed geographically. Just as with oil and gas extraction, it is not possible to utilize a storage resource anywhere. Although sedimentary formations are common across the United States, Eccles et al. (2011) find that the vast majority of cheap, abundant storage potential is concentrated in the Michigan and Ohio Basins, in Cretaceous sediments along the East Coast, and in Texas, especially along the Gulf Coast. Fortunately, these areas are easily accessible for many large CO_2 sources, but in terms of physical surface area, they make up a relatively small portion of sedimentary basins in the United States in general (Eccles et al. 2011).

These findings indicate that at least at first, carbon storage might be utilized for *below* the "average" cost of storage because these large, cheap reservoirs could be easily exploited. From the wrong perspective, the variation in carbon storage potential might lead to the conclusion that because many geological formations are unsuitable for cheap, large-scale storage, the technology does not have much potential for abating climate change. However, an integrated examination of geographic variation in geology indicates that cheap, large-scale storage sites might be deployed first, and thus the initial cost of utilization might be less than a summary statistic might indicate.

Second, transport of CO_2 over long distances can be relatively cheap, with appropriate coordination and scale, enabling the use of large, low-cost reservoirs that may not be proximate to sources. This may be apparent even when simply

examining techno-economic models (McCoy and Rubin 2008) or summaries of transport costs for pipelines (Chandel et al. 2010)—transporting CO_2 at scale (greater than perhaps 10 MT of CO_2 per pipeline route per year) decreases the cost per tonne of transport by a significant amount. Chandel et al. indicate that cost for transport (including labor and materials, both for pipelines and pumping stations) can drop from $0.018 per tonne per km for low volume transport (roughly 1 MT per year) to under $0.010 per tonne per km for 10 MT per year; even larger pipelines might see cost drop as low as $0.006 per tonne per km (Chandel et al. 2010). These numbers put the levelized cost of transport for 250 km of pipeline and 10–20 MT of CO_2 per year at a relatively low $2/t. This does not include the cost of rights of way, but is obviously smaller in scale than the variation in the cost of geological storage and the total cost of capture.

In earlier versions of geographic optimization, we see the argument that transport and storage will be low-cost because most sources are close to some kind of sink (Dooley et al. 2004; Wildenborg et al. 2004). Geological variations and their impact on cost appear to contradict this, because not every storage site is going to a feasible sink economically. However, we can see that geographic optimization indicates that large-scale transport is relatively low cost and national-scale geological data indicates that there are at least a few huge, low-cost storage options; whether sources and sinks are close to each other may be largely irrelevant.

Large-scale optimization brings these elements together and can more or less easily determine what the nature of an integrated, planned network of pipelines would be and how much it would cost. Integrated optimization is fairly standard for transport and storage modeling. As compared to a decade ago, there are many different models competing for attention in academic literature, where geographers have made a substantial impact on improving and refining our understanding of the physical potential for CO_2 storage and the nature of the infrastructure that would support that storage system.

Issues with Optimization

Advances brought by traditional geographic optimization have not been able to overcome some interesting challenges posed by geological storage and the CCS system in general.

The first most obvious of these is geological uncertainty. CO_2 storage pilot projects have until recently been demonstration-scale, with a few large commercial-scale endeavors (Michael et al. 2010). Of these, Sleipner Vest is probably the most well-known of non-EOR projects. From a geological perspective, the Utsira sand, into which the Sleipner project injects about a million tonnes of CO_2 per year, is an ideal environment. It is high-permeability, unconsolidated sand, and even though it is almost 200 m thick, the horizontal well-bore is perforated along only 100 m and can still sustain injection rates of nearly 3000 tonnes per day (Eccles et al. 2009).

The Snovhit project was expected to have similar performance to Sleipner Vest, but rapid increases in reservoir pressure early in its operation indicated that the Snovhit geology was not exactly as expected. Consequently, engineers at Statoil had to intervene to increase injectivity (Eiken et al. 2011). From a technical perspective this is a success story since, although the site did not operate as expected, Statoil was able to adapt to the situation. However, from a planning or operations perspective, the project is considered a minor disaster. In a large-scale storage system, Snovhit would not have been a one-off project, it would have been one of an interconnected series of CO_2 storage sites selected for their physical and economic characteristics as better than other sites that would have been more costly to utilize. The unexpected characteristics of Snovhit would have increased its cost and made it less reliable as a CO_2 sink than other sites, but the source-sink matching algorithm would obviously not have been able to take that into account.

Geological uncertainty is especially pernicious because Snovhit was supposedly well characterized (Eiken et al. 2011). Uncertain geologic characteristics can be challenging to incorporate into linear optimization, but Monte Carlo frameworks and other techniques could be used to create hybrid optimization models that would incorporate uncertain characteristics; uncertainty in storage and cost estimates are somewhat linked (Middleton et al. 2012c).

First, optimization models will have to be constrained to require the inclusion of reserve capacity. Snovhit isn't the only storage project with uncertain operating parameters. Even the very small (48 t/day) Nagaoka project experienced wide variations in injection performance, and systems that manage 100-megaton-scale emissions must be able to handle such intermittent performance (Research Institute of Innovative Technology for the, 2007). The cost of possible intermittency must be traded off against the cost of emissions, putting CCS in the same reliability boat as renewables like wind and solar.

Second, models must be able to appropriately digest uncertain operating parameters and provide useful results the incorporate the known uncertainty in geology. This is distinct from the problem of including reserve capacity in that reserve capacity is generally meant to accommodate short-term issues with performance. In electrical production and transmission, this might be the failure of a generator but not necessarily the complete removal of that generator from the grid. With geological storage, there can be considerable uncertainty as to the performance of the sequestration site on a long-term basis, as with Snovhit. Although various sensing techniques can mitigate this uncertainty, it is entirely possible that in addition to seeing daily fluctuations in the capacity of an injection well, a storage site operator might encounter systemic flaws in geology which permanently reduce the operating capacity of the site (Middleton et al. 2012c).

This might not even be the result of geological uncertainty. Underground injection has long-lived effects on the pressure environment, which reduces the injectivity of wells and requires compensation through either increased injection pressure or active reservoir management or simply toleration of the reduced performance (Eccles et al. 2012). In some cases, active reservoir management may be too expensive (because of fluid disposal costs) or increasing pressure may not be

feasible (because the pressure may be near the fracture limit), which would mean that tolerance for decreased performance might be the only option. If wells are not correctly spaced, this decreased performance could be dramatic (Eccles et al. 2012). Perhaps most importantly, the decreased performance is more or less directly related to the permeability of the formation, the parameter at issue in the Snovhit problems.

The optimization framework must therefore not only account for short-term reserve capacity, but also the cost of building out the system to account for possible long-term disruptions. The degree and nature of the long-term disruptions is in many ways quite uncertain, more or less in direct relation to the degree to which geological conditions are well-characterized.

In addition to accounting for geological uncertainty, spatial optimization must take into consideration complex decision variables in which different components of the CCS system might interact in unexpected ways. This somewhat ambiguous statement is best illustrated in the consideration of peaking plants and the dispatch of power (Middleton and Eccles 2013). Traditionally, CCS is modeled as a monolithic technology with constant operation that matches the capacity factor of the source to which it is attached. For baseload coal plants, this is not necessarily incorrect, but for plants with high ramp rates that operate as peaking plants or load-following plants, there is an interesting problem: what size should the capture facility be? It could be sized anywhere from zero to the maximum CO_2 output of the plant – depending on the plant's dispatch characteristics, though, it could be underutilized to a considerable degree. If the plant is a peaking natural gas plant, it might sit unused for most of the day, which obviously affects the economics of its operation.

The key element of the problem is the cost of CO_2 disposal vs. the cost of emissions (Middleton and Eccles 2013); with a high enough carbon price (or, critically, a low enough disposal cost) the plant would oversize its capture equipment. The carbon price is a matter of policy and is essentially out of the hands of the modeler (Kuby et al. 2011a), but the disposal cost can be a decision parameter. Whether the plant sends its emissions to a local EOR producer vs. sending them to a big, centralized CCS facility a few states over is critical to the initial decision of installing capture equipment, a feedback loop that is challenging to incorporate into the structure of geographic optimization models.

Broadly speaking, this decision comes down to incorporating the option to underutilize infrastructure. There are other reasons to underutilize infrastructure (or the equivalent, to oversize its initial construction), which includes resources coming online at different times and the storage capacity or reserve capacity of underutilized infrastructure. But these decision options need to be integrated very carefully with the simple but powerful source-sink matching optimization that currently exists. Even something as unobtrusive as fault tolerance (what happens when a pipeline goes down?) needs consideration in planning, optimization, and cost estimation.

Finally, slow movement on climate change policy has thrown into sharp relief the unusual question of the nature of the objective function. Lacking a financial incentive to avoid emissions, proponents of CCS have turned to carbon utilization

as a motivation for CCS deployment. Carbon dioxide is currently used most widely in EOR, where each tonne of injected CO_2 might extract 3 (using industry standards) or more (using reservoir modeling) barrels of oil. At the current price of oil, it is not difficult to see that the financial incentives for capture are much different than we might expect under a carbon price, which has led to renewed interest in enhanced oil recovery using CCS as the source for the CO_2.

In this environment, it is not necessarily appropriate to minimize costs: the actors in the system would rather want to maximize revenue, especially the oil producers. It was never particularly clear what entities would be operating the pipeline transport or CO_2 storage portions of the CCS system, but in the current environment, it seems it is almost certainly not going to be the same entity that operates the CO_2 source, so the different motivations must be integrated into optimization modeling. Additionally, echoing earlier issues of uncertainty, this environment might dramatically change over the planning lifetime of a CCS project. It is certainly not out of the question that even the somewhat recalcitrant United States might have a carbon price in the next two decades; what impact would this have halfway through the planning and/or deployment of a large-scale carbon storage system? How will the policy treat stranded capital or ongoing investments? Perhaps more importantly, will the policy itself be based on the lessons learned or results from optimization modeling and deployment of CCS systems?

Beyond these questions of uncertainty and motivation, it is clear that deployment of CCS on the scale that would impact CO_2 emissions dramatically is an application of geographic optimization that has never been seen before. The CO_2 emissions from the US electric power sector alone would produce the same volume of (compressed) CO_2 as the entire would consumes in oil. We have certainly used geographic optimization for pipeline planning and transport optimization, but never at such a large scale. The idea of designing a system that would move a billion tonnes of CO_2 from the ground up is mind-boggling, and there are almost certainly challenges of working at that scale that have not even been considered, much less integrated into modeling.

Flexible Modeling Approach

To deal with these issues, a flexible modeling approach that draws from the best traditions of spatial optimization is necessary. As the name implies, flexible modeling could involve many different techniques to modify basic network-sink optimization. Some of these variations might be innovative applications of linear programming, and some might involve extending linear optimization to mixed-integer programming, using creative constraints or decision variables, or even shifting to hybrid approaches, especially involving unconventional optimization modeling.

Mixed-integer programming is the easiest technique to incorporate, as is clear from the fact that it is already used in several optimization models (Middleton and Bielicki 2009; Morbee et al. 2012). One reason to use mixed-integer programming is the prevalence of fixed pipeline sizes in industry and modeling; pipelines have to be scaled to accommodate the optimum CO_2 flux but have to use one of the fixed pipeline sizes, so MIP has to be used (Middleton 2013). MIP techniques are common in transport optimization, and thus it may not be particularly revolutionary to say that they should be used here. At the same time, however, MIP solutions take more computation to find, so MIP models have to be constructed or deployed carefully in complex or very large-scale optimization problems.

MIP's limitations are especially important when the issues of geological uncertainty and peaking plant's decision options (i.e. how much capture capacity to install) are considered. At first glance, this may not be apparent. Reserve requirements that can mitigate daily fluctuations in transport or sequestration capacity may simply require constraints in a MIP model that would overbuild capacity given probabilities, expected values, and so on. Even longer-term issues might also be solved with clever MIP applications such as, for example, multiple time periods for capacity to come online to deal with reduced performance with possible acceleration or the equivalent of spinning reserves in electricity transmission applied to CCS systems (Middleton et al. 2012e). This dramatically increases the complexity of the MIP problems to be solved, but that is not necessarily prohibitive.

However, modelers may want to consider a more flexible solution, which is to employ a variety of techniques at different scales specifically designed to incorporate uncertainty and probability. At local scales for individual projects, this might not be much different than the techniques we see today. But at larger scales with more uncertainty, modelers may wish to take advantage of Monte Carlo techniques. To deal with large-scale geological uncertainty, Monte Carlo models might solve many variations on the same problem to arrive at a variety of outcomes and distributions of their costs; this has previously been done with CCS and MIPs (Middleton et al. 2012c). In transportation routing, this is of course very challenging to interpret, because it is in some ways only the beginning of the decision process that leads to a build-out on the ground. Nonetheless, it can be paired with the more-detailed local routing to give industry stakeholders and policymakers a multi-scale tool that could correctly incorporate uncertainty and risk at some levels but provide useful, detailed routing and cost information at other levels.

Beyond iterating network-sink problems in a Monte Carlo framework, evolutionary or agent-based algorithms deployed at the right scale could also narrow decision options and also provide a "fuzzier" picture of optimal decisions. These types of algorithms, in some ways like Monte Carlo models, can give multiple decision options that are close to the optimum, which might be especially useful in accounting for geological uncertainty or determining how to plan simulations are a more detailed scale. They have another advantage, which is that they tend to incorporate behavior by the agents involved (like power plant, pipeline, and storage

site operators) that is easier to tweak or to limit so as to match the conditions or behavior that are observed in the real world. These agent models might be especially effective if the agents solve simple transport optimization problems as part of the model structure.

A variety of models using different techniques at different scales might thus incorporate varying degrees of complexity in transport optimization yet in their entirety provide more value than a rigid approach to multi-scale modeling using linear or mixed integer programming alone. The flexible suite of models allows evaluation of CCS resources, decisions, and policy and many levels and with many degrees of confidence, from the detailed and somewhat traditional pipeline routing at a local or project level used to actually plan a CCS installation to the plethora of techniques that can be applied at national or international scales to assess the mitigation potential of the resource.

Regardless of the structure of these models in the future, it is fairly clear that spatial optimization has dramatically improved the assessment of CCS at all scales already. The flexible approaches of the future will continue to refine our understanding of CCS potential in the energy system. Perhaps most importantly, the unique challenges presented in CCS modeling may provide valuable insight into hybrid or flexible modeling techniques in other arenas, generalizing the issues and solutions to spatial optimization as a field.

The scalable, flexible approach is almost certainly applicable to energy infrastructure modeling as a whole. Spatial optimization is critical in understanding the deployment of future energy projects, from transmission for wind projects (e.g., Phillips and Middleton 2012) to the CCS pipeline infrastructure issues described above. Other resources are likely to have the same or more uncertainty as we find in CCS. Intermittency in renewable generation is another critical issue that will impact its deployment and the deployment of infrastructure that supports it, another clear application of spatial optimization with uncertain parameters that might require a flexible approach. Integrating plugin hybrid electric vehicles as an energy storage system in the smart grid of the future likewise requires different approaches to deal with uncertainty at different scales.

The techniques that will be developed to deal with the issues in CCS optimization will almost certainly have value in these and other arenas, making them not simply useful to the CCS research community, but to the larger energy infrastructure research community and to the field of spatial optimization. Flexible modeling approaches can help solve critical challenges in energy production in the future, in a world constrained by climate change and resource allocation as well as increasing demand. Spatial distribution of resources will be a feature of the energy infrastructure of the future for decades to come, and location allocation will thus be an invaluable tool in understanding and planning for the energy needs of civilization in years to come (Figs. 1 and 2).

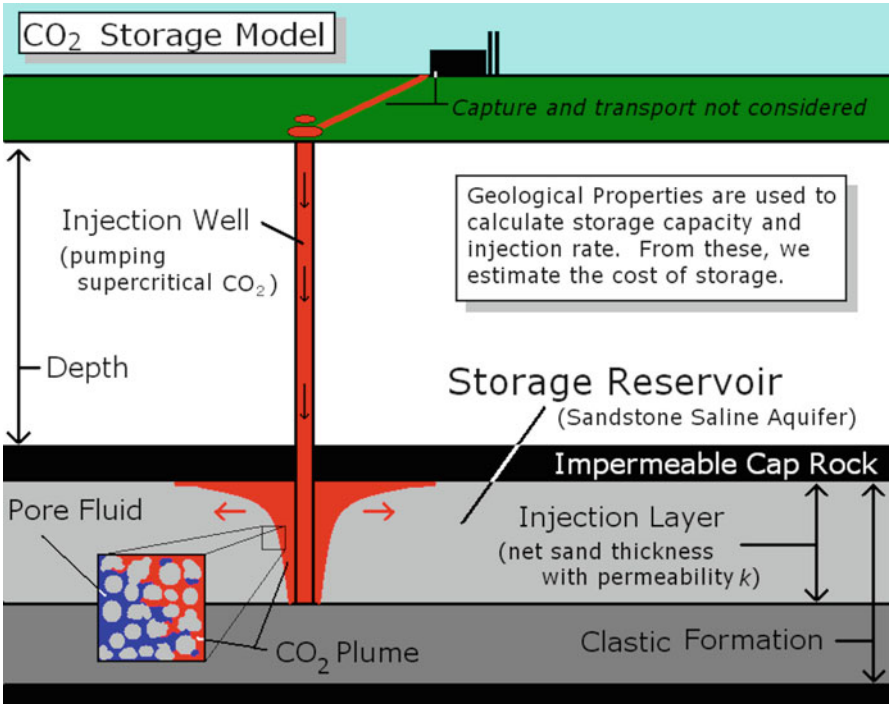

Fig. 1 Schematic of the CO_2 capture and injection/storage process; CO_2 transport is not depicted

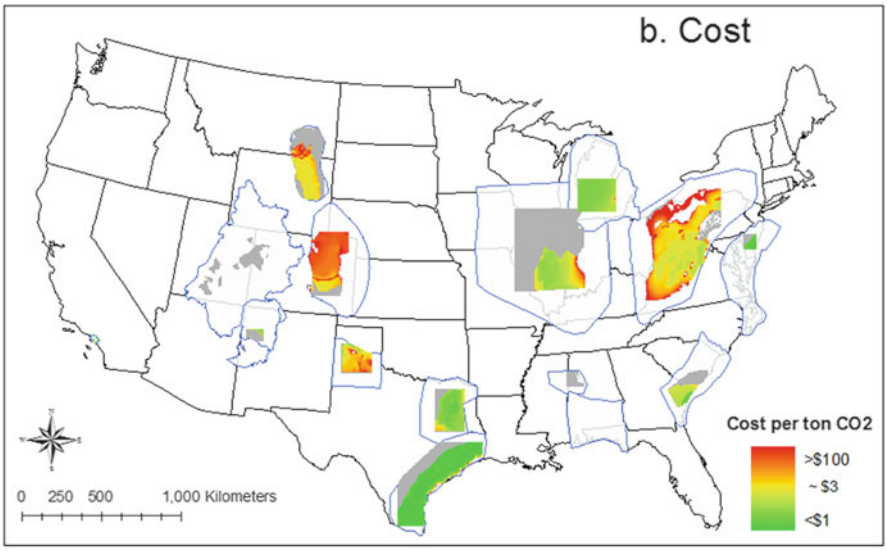

Fig. 2 Variability of CO_2 injection and storage costs for major deep saline aquifers in the United States (Eccles et al. (2011))

References

Allinson, W. G., Nguyen, D. N., & Bradshaw, J. (2003a). Dealing with carbon dioxide–Threat or opportunity-The economics of geological storage of CO2 in Australia. *APPEA Journal-Australian Petroleum Production and Exploration Association, 43*, 623–636.

Allinson, W. G., Nguyen, D. N., & Bradshaw, J. (2003b). The economics of geological storage of CO2 in Australia. *APPEA Journal, 623*.

Bachu, S., Bonijoly, D., Bradshaw, J., Burruss, R., Holloway, S., Christensen, N. P., & Mathiassen, O. M. (2007). CO2 storage capacity estimation: Methodology and gaps. *International Journal of Greenhouse Gas Control, 1*, 430–443. http://www.sciencedirect.com/science/article/B483WP-434PB475MY-431/432/433c432dd437f435d420d413dd5377fa87713e87710b87716f.

BCG. (2008). *Carbon capture and storage: A solution to the problem of carbon emissions* (p. 11). Boston: Boston Consulting Group.

BEG. (2000). *Carbon-dioxide sequestration*. http://www.beg.utexas.edu/environqlty/co2seq/co2data.htm. Carbon Dioxide Sequestration – Study Areas.

Bernstein, L., Bosch, P., Canziani, O., Chen, Z., Christ, R., Davidson, O., Hare, W., Huq, S., Karoly, D., Kattsov, V. et al., 2007. *Climate change 2007: Synthesis report. An Assessment of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.

Bock, B. (2002). *Economic evaluation of CO2 storage and sink enhancement options* (p. 476). Palo Alto/Muscle Shoals/Washington, DC: EPRI, TVA, U.S. DOE.

Chandel, M. K., Pratson, L. F., & Williams, E. (2010). Potential economies of scale in CO(2) transport through use of a trunk pipeline. *Energy Conversion and Management, 51*, 2825–2834.

Ciferno, J. P., Litynski, J. L., & Plasynski, S. I. (2010). *DOE/NETL carbon dioxide capture and storage RD&D roadmap* (p. 78). National Energy Technology Laboratory.

Craft, B. C., Hawkins, M., & Terry, R. E. (1991). *Applied petroleum reservoir engineering*. Prentice Hall, 0130398845%7 2.

Dahowski, R. T., & Dooley, J. J. (2004). Carbon management strategies for US electricity generation capacity: A vintage-based approach. *Energy 29*, 1589–1598. http://www.sciencedirect.com/science/article/B1586V1582S-1584CC1587RP1583-J/1582/f1075aa1583d1581ec1597f1762e90278dd90227e90277cbf.

Dahowski, R. T., Li, X., Davidson, C. L., Wei, N., Dooley, J. J., & Gentile, R. H. (2009). A preliminary cost curve assessment of carbon dioxide capture and storage potential in China. *Energy Procedia 1*, 2849–2856. http://www.sciencedirect.com/science/article/B2984K-2844W2840SFYG-F2844/2842/a391538b391197ae311453f391538c0299724bf0299720a

Dooley, J. J., Dahowski, R. T., Davidson, C. L., Bachu, S., Gupta, N., & Gale, J. (2004). A CO2-storage supply curve for North America and its implications for the deployment of carbon dioxide capture and storage systems. In *Proceedings of the seventh international conference on greenhouse gas control technologies*. Amsterdam: Elsevier.

Dooley, J. J., Dahowski, R. T., & Davidson, C. L. (2008). *On the long-term average cost of CO2 transport and storage* (p. 6). Richland: Pacific Northwest National Laboratory.

Eccles, J. K., Pratson, L., Newell, R. G., & Jackson, R. B. (2009). Physical and economic potential of geological CO2 storage in saline aquifers. *Environmental Science & Technology, 43*, 1962–1969.

Eccles, J. K., Pratson, L., Newell, R. G., & Jackson, R. B. (2011). The impact of geologic variability on capacity and cost estimates for storing CO2 in deep-saline aquifers. http://www.sciencedirect.com/science/article/pii/S0140988311002891. Energy Economics.

Eccles, J. K., Pratson, L., & Chandel, M. K. (2012). Effects of well spacing on geological storage site distribution costs and surface footprint. *Environmental Science & Technology*.

EIA. (2007). *The electricity market module of the national energy modeling system*. Washington, DC: US EIA.

Eiken, O., Ringrose, P., Hermanrud, C., Nazarian, B., Torp, T. A., & Høier, L. (2011). Lessons learned from 14 years of CCS operations: Sleipner, In Salah and Snøhvit. *Energy Procedia, 4*, 5541–5548. http://www.sciencedirect.com/science/article/pii/S1876610211008204.

EPA. (2008). *Geological CO2 Sequestration Technology and Cost Analysis* (p. 61). Washington, DC: EPA Office of Water.

Han, J.-H., & Lee, I.-B. (2011). Development of a scalable and comprehensive infrastructure model for carbon dioxide utilization and disposal. *Industrial & Engineering Chemistry Research, 50*, 6297–6315.

Hendriks, C. A., & Blok, K. (1993). Underground storage of carbon dioxide. *Energy Conversion and Management, 34*, 949–957.

Hendriks, C., Graus, W., & van Bergen, F. (2004). *Global carbon dioxide storage potential and costs*. Utrecht: Ecofys.

Herzog, H., Li, W., Zhang, H., Diao, M., Singleton, G., & Bohm, M. (2007). *West coast regional carbon sequestration partnership: Source sink characterization and geographic information system based matching* (p. 82). Cambridge, MA: Public Interest Energy Research (PIER) Program: California Energy Commission.

Keating, G. N., Middleton, R. S., Stauffer, P. H., Viswanathan, H. S., Letellier, B. C., Pasqualini, D., Pawar, R. J., & Wolfsberg, A. V. (2010). Mesoscale carbon sequestration site screening and CCS infrastructure analysis†. *Environmental Science and Technology, 45*, 215–222. http://dx.doi.org/210.1021/es101470m.

Kemp, A. G., & Kasim, A. S. (2010). A futuristic least-cost optimisation model of CO2 transportation and storage in the UK/UK continental shelf. *Energy Policy, 38*, 3652–3667.

Koide, H., Tazaki, Y., Noguchi, Y., Nakayama, S., Iijima, M., Ito, K., & Shindo, Y. (1992). Subterranean containment and long-term storage of carbon dioxide in unused aquifers and in depleted natural gas reservoirs. *Energy Conversion and Management, 33*, 619–626. http://www.sciencedirect.com/science/article/B616V612P-497B619ND-MS/612/612e615def618cd597dc5859b5850b5859c5855e5852c5912.

Koide, H., Shindo, Y., Tazaki, Y., Iijima, M., Ito, K., Kimura, N., & Omata, K. (1997). Deep subseabed disposal of CO2 – The most protective storage. *Energy Conversion and Management, 38*, S253–S258. http://www.sciencedirect.com/science/article/B256V252P-254DS259V240-251H/252/7303021b7303057a7303029b7303024fa5196636979789c5196636979754.

Kuby, M. J., Bielicki, J. M., & Middleton, R. S. (2011a). Optimal spatial deployment of CO_2 capture and storage given a price on carbon. *International Regional Science Review, 34*, 285–305.

Kuby, M. J., Middleton, R. S., Bielicki, J. M. (2011b). Analysis of cost savings from networking pipelines in CCS infrastructure systems. In: J. Gale, C. Hendriks, W. Turkenberg (Eds.), 10th international conference on greenhouse gas control technologies, pp. 2808–2815.

McCoy, S. T., & Rubin, E. S. (2008). An engineering-economic model of pipeline transport of CO2 with application to carbon capture and storage. *International Journal of Greenhouse Gas Control, 2*, 219–229. http://www.sciencedirect.com/science/article/B283WP-214R215G218M216-214/212/9979191aebc9979106d9941927bf9979147ee9979195ff9979743e.

McKinsey Climate Change Initiative. (2008). *Carbon capture & storage: Assessing the economics*. McKinsey & Company.

Mendelevitch, R., Herold, J., Oei, P.-Y., & Tissen, A. (2010). *CO2 highways for Europe: Modelling a carbon capture, transport and storage infrastructure for Europe* (CEPS Working Document No. 340).

Metz, B., Davidson, O., De Coninck, H., Loos, M., & Meyer, L. (2005). *IPCC special report on carbon dioxide capture and storage*. Working Group III of the Intergovernmental Panel on Climate Change, Cambridge.

Michael, K., Golab, A., Shulakova, V., Ennis-King, J., Allinson, G., Sharma, S., & Aiken, T. (2010). Geological storage of CO2 in saline aquifers-A review of the experience from existing storage operations. *International Journal of Greenhouse Gas Control, 4*, 659–667.

Middleton, R. S. (2013). A new optimization approach to energy network modeling: Anthropogenic CO2 capture coupled with enhanced oil recovery. *International Journal of Energy Research*. https://doi.org/10.1002/er.2993.

Middleton, R. S., & Bielicki, J. M. (2009). A scalable infrastructure model for carbon capture and storage: SimCCS. *Energy Policy, 37*, 1052–1060.

Middleton, R. S., & Brandt, A. R. (2013). Using infrastructure optimization to reduce greenhouse gas emissions from oil sands extraction and processing. *Environmental Science and Technology*. https://doi.org/10.1021/es3035895.

Middleton, R. S., & Eccles, J. K. (2013). The complex future of CO2 capture and storage: Variable electricity generation and fossil fuel power. *Applied Energy, 108*, 66–73.

Middleton, R. S., Bielicki, J. M., Keating, G. N., & Pawar, R. J. (2011). Jumpstarting CCS using refinery CO2 for enhanced oil recovery. In: J. Gale, C. Hendriks, W. Turkenberg (Eds.), 10th international conference on greenhouse gas control technologies, pp. 2185–2191.

Middleton, R. S., Keating, G. N., Stauffer, P. H., Jordan, A. B., Viswanathan, H. S., Kang, Q. J., Carey, J. W., Mulkey, M. L., Sullivan, E. J., Chu, S. P., Esposito, R., & Meckel, T. A. (2012a). The cross-scale science of CO2 capture and storage: from pore scale to regional scale. *Energy & Environmental Science, 5*, 7328–7345.

Middleton, R. S., Kuby, M. J., & Bielicki, J. M. (2012b). Generating candidate networks for optimization: The CO2 capture and storage optimization problem. *Computers, Environment and Urban Systems, 36*, 18–29.

Middleton, R. S., Kuby, M. J., Wei, R., Keating, G. N., & Pawar, R. J. (2012c). A dynamic model for optimally phasing in CO2 capture and storage infrastructure. *Environmental Modelling & Software, 37*, 193–205.

Middleton, R. S., Keating, G. N., Viswanathan, H. S., Stauffer, P. H., & Pawar, R. J. (2012d). Effects of geologic reservoir uncertainty on CO2 transport and storage infrastructure. *International Journal of Greenhouse Gas Control, 8*, 132–142.

Middleton, R. S., Keating, G. N., Viswanathan, H. S., Stauffer, P. H., & Pawar, R. J. (2012e). Effects of geologic reservoir uncertainty on CO_2 transport and storage infrastructure. *International Journal of Greenhouse Gas Control, 8*, 132–142. 1750–5836.

Morbee, J., Serpa, J., & Tzimas, E. (2012). Optimised deployment of a European CO2 transport network. *International Journal of Greenhouse Gas Control, 7*, 48–61.

Nemeth, K., Hill, G., Sams, K., Hovorka, S. D., Ripepi, N., Pashin, J., Hills, D., Rhudy, R., Trautz, R. C., Esposito, P. R., Gandee, J. E., Locke, C. D., Lindner, J., Han, F.-X., Luthe, J. C., Conrad, M., & McCollum, C. (2011). *Southeast regional carbon sequestration partnership (SECARB) Phase II final report* (p. 622). Norcross: Southern States Energy Board.

Parker, N. (2004). Using natural gas transmission pipeline costs to estimate hydrogen pipeline costs. UC Davis: Institute of Transportation Studies. Retrieved from: http://www.escholarship.org/uc/item/9m40m75r

Petroleum Technology Research Centre. (2011). *Weyburn-Midale research project*.

Phillips, B. R., & Middleton, R. S. (2012). SimWIND: A geospatial infrastructure model for optimizing wind power generation and transmission. *Energy Policy, 43*, 291–302.

Research Institute of Innovative Technology for the, E. (2007). *Nagaoka project: Demonstration test and monitoring at the Iwanohara test site*. http://www.rite.or.jp/English/lab/geological/demonstration.html

Rubin, E. S., Rao, A. B., & Chen, C. (2005). Comparative assessments of fossil fuel power plants with CO_2 capture and storage. In E. S. Rubin, D. W. Keith & C. F. Gilboy (Eds.), 7th international conference on greenhouse gas control technologies (pp. 285–294). Elsevier Science.

Rubin, E. S., Chen, C., & Rao, A. B. (2007). Cost and performance of fossil fuel power plants with CO_2 capture and storage. *Energy Policy, 35*, 4444–4454.

Stauffer, P. H., Viswanathan, H. S., Pawar, R. J., & Guthrie, G. D. (2009). A system model for geologic sequestration of carbon dioxide. *Environmental Science and Technology, 43*, 565–570.

Stauffer, P. H., Keating, G. N., Middleton, R. S., Viswanathan, H. S., Berchtold, K. A., Singh, R. P., Pawar, R. J., & Mancino, A. (2011). Greening coal: Breakthroughs and challenges in carbon capture and storage. *Environmental Science & Technology, 45*, 8597–8604.

USDOE. (2010). *Carbon sequestration atlas of the United States and Canada Online (III)*. Morgantown: National Energy Technology Laboratory.

van den Broek, M., Brederode, E., Ramirez, A., Kramers, L., van der Kuip, M., Wildenborg, T., Turkenburg, W., & Faaij, A. (2010). Designing a cost-effective CO_2 storage infrastructure using a GIS based linear optimization energy model. *Environmental Modelling & Software, 25*, 1754–1768.

Viswanathan, H. S., Pawar, R. J., Stauffer, P. H., Kaszuba, J. P., Carey, J. W., Olsen, S. C., Keating, G. N., Kavetski, D., & Guthrie, G. D. (2008). Development of a hybrid process and system model for the assessment of wellbore leakage at a geologic CO_2 sequestration site. *Environmental Science & Technology, 42*, 7280–7286.

Wildenborg, T., Gale, J., Hendriks, C., Holloway, S., Brandsma, R., Kreft, E., & Lokhorst, A. (2004, September). Cost curves for CO 2 storage: European sector. In *Proceedings of the 7th international conference on greenhouse gas control technologies* (GHGT-7), pp. 5–9.

The Recharging Infrastructure Needs for Long Distance Travel by Electric Vehicles: A Comparison of Battery-Switching and Quick-Charging Stations

Linda Christensen, Sigal Kaplan, Thomas C. Jensen, Stefan Røpke, and Allan Olsen

Abstract On-road electric vehicle recharging infrastructure is essential in the transformation of electric vehicles into a practical transportation option. This study focuses upon assessing the need for recharging infrastructure for long distance travel for a large market share of electric vehicles, finding the optimal infrastructure deployment, and understanding the economic, social and environmental costs and benefits associated with the optimal infrastructure deployment. The analysis considers quick-charging and battery-switching as plausible recharging technologies. Results show: (i) the promotion of electric vehicles is beneficial when considering economic costs and benefits for operators and users, tax redistribution, and environmental externalities, even with a relatively modest market share; (ii) the number of required recharging stations for satisfaction of the travel demand is at the magnitude of 1–2% of the current gasoline infrastructure, under the assumption of wide availability of off-road recharging at home and the workplace; (iii) the optimal deployment of the recharging stations is along the main national highways outside of urban conurbations, under the assumption of wide availability of home recharging; (iv) the battery-switching technology is far more attractive to the consumer than the quick-charging technology for long-distance travel requiring more than one recharging visit.

Keywords Electric vehicles • Recharging stations • Location optimization • Socio-economic analysis • Battery-switching • Quick-charging • Spatial-optimization • EVs

L. Christensen • T.C. Jensen • S. Røpke • A. Olsen
Department of Management Engineering, Technical University of Denmark, Bygningstorvet 116, B 2800, Kongens Lyngby, Denmark
e-mail: lich@dtu.dk; tcje@dtu.dk; ropke@dtu.dk

S. Kaplan (✉)
The Geography Department, The Hebrew University, Mt. Scopus, Jerusalem 9190501, Israel
e-mail: Sigal.kaplan@mail.huji.ac.il

J.-C. Thill (ed.), *Spatial Analysis and Location Modeling in Urban and Regional Systems*, Advances in Geographic Information Science,
https://doi.org/10.1007/978-3-642-37896-6_15

Introduction

The mass utilization of electric vehicles (EVs) was proposed at the end of the 1970s as a remedy for an automobile market under the constraints of an oil shortage, high oil prices, and depletion of environmental resources (Blair 1978; Charlesworth and Baker 1978). Research efforts during more than three decades have focused on technological improvements, market penetration, and impact assessment of EVs (Blair 1978; Carmody and Haraden (1982); De Luchi et al. 1989; Giese et al. 1983; Hamilton 1980; Kurani et al. 1994). However, the market penetration of EVs has been negligible so far because of high unit costs, limited driving range, and lack of recharging infrastructure (Chéron and Zins 1997; Dagsvik et al. 2002; Pearre et al. 2011). Nevertheless, recent years have witnessed growing expectations for a rapid EV growth in the future, following battery technology innovations and governmental commitment to EV promotion through legislation, investments, and taxation policies (Andersen et al. 2009; Brady and O'Mahony 2011; Brown et al. 2010; Dagsvik et al. 2002; Hidrue et al. 2011).

Adequate EV recharging infrastructure has a key role in transforming EVs into a viable transportation option for large-scale adoption (Andersen et al. 2009; Dagsvik et al. 2002; Hidrue et al. 2011; Wang and Lin 2009; Christensen 2011). Accordingly, a significant body of research has been dedicated to developing recharging technologies (Calasanzio et al. 1993; Fernández and Trinidad 1997; Yang and Liaw 2001) and promoting recharging standardization (Brown et al. 2010). In parallel, research has been devoted to impact assessment of EV recharging on the electric power system. Themes of interest are the minimization of the burden induced by vehicle recharging on the power grid (Hartmann and Özdemir 2011; Mullan et al. 2011; Perujo and Ciuffo 2010); the dual role of EVs as both consumption and storage devices (Andersen et al. 2009; Kristoffersen et al. 2011); and the development of intelligent grid management systems that allow power load optimization for example by allowing flexible recharging rates according to the power consumption rate (Ahn et al. 2011; Amoroso and Cappuccino 2011; Van den Bossche 2010).

The optimal deployment of EV recharging stations on the basis of consumer behavior is scarcely explored, although a handful of studies sheds light on the infrastructural needs and optimal deployment of refueling stations for hydrogen and natural gas (Frick et al. 2007; Kuby and Lim 2005; Kuby et al. 2009; Nicholas et al. 2004). Differences exist in terms of objective functions, as Nicholas et al. (2004) proposed minimizing the travel time to the refueling stations, Kuby and Lim (2005) and Kuby et al. (2009) focused on maximizing the refueled traffic volumes without changes in the selected routes, and Frick et al. (2007) considered a multiple-objective function including the minimization of travel distance to population conurbations, the locations of pipelines, and commercial opportunities. Recently, Kim and Kuby (2012) proposed a model that allows vehicles to deviate from shortest paths in order to refuel. Differences exist also in terms of scale, as Kuby and Lim (2005) and Kim and Kuby (2012) considered a synthetic network, Nicholas et al. (2004) referred to metropolitan regions, while Frick et al. (2007) considered a

national model for Switzerland. Kuby et al. (2009) explored both a metropolitan scale model for Orlando and a state scale model for Florida.

The unique features of EVs impede the direct application of the aforementioned methodologies to the optimal location of EV recharging stations. First, EV recharging can be done anywhere (i.e., at home or at activity location), provided adequate connection to the electricity grid, and hence the assumption that people prefer to recharge in proximity to their origin or destination is not necessarily substentiated in the case of EV recharging behavior. Second, the location of EV recharging infrastructure is highly flexible since it is independent of current or planned networks for the deployment of natural gas or hydrogen pipelines.

Only a few recent studies examined the optimal deployment of EV recharging infrastructure revealing that this line of research is still at its nascent stage. Wang and Lin (2009) proposed a model that selects a set of stations to locate, minimizing total cost, such that all vehicle paths given as input to the model are feasible. The model is used for locating EV quick-charging stations for intercity travel along a coastal road in Taiwan. Wang and Wang (2010) extended the previous model by including a dual-objective function of minimum location cost and maximum population coverage. Li et al. (2010) suggested using demand-responsive portable recharging units and analyzed a synthetic network assuming 100 EVs.

The current study joins the line of location optimization research for EV recharging stations by assessing the need for on-road EV recharging stations based upon travel patterns. The purpose of this analysis is to provide an efficient spatial distribution of EV recharging stations at the national level, with a focus upon understanding its benefits and costs while considering economic, social and environmental goals. The importance of this issue is threefold. First, although evidence suggests that consumers would prefer overnight recharging at home over on-road recharging stations (Skippon and Garwood 2011), the deployment of publicly accessible recharging infrastructure remains important as a pre-condition for large-scale EV market penetration in urban and rural areas (Andersen et al. 2009; Dagsvik et al. 2002; Hidrue et al. 2011; Wang and Lin 2009). Second, understanding the costs and benefits of quick-charging versus battery-switching is of interest since these technologies are currently underway in several world regions including Australia, Asia, Europe, and the United States. Finally, EV recharging infrastructure requires significant private and public investments for infrastructure development (Andersen et al. 2009; Li and Ouyang 2011). Consequently, efficient infrastructure deployment is essential, in particular in the initial stages of EV market penetration.

The contribution of the current study to the locational optimization literature regarding EV recharging infrastructure is fourfold. First, while existing studies on EV recharging stations have focused on small-scale local networks, the current study analyzes a nationwide case including thousands of potential EV chanrging station locations. Second, while previous studies regarding the location of EV recharging stations have represented the consumer perspective, the current study proposes a comprehensive socio-economic analysis including social and environmental considerations. Third, while previous studies on the location of EV stations have ignored possible travel pattern changes due to recharging, the current study -in line with

the recent study of Kim and Kuby (2012) on hydrogen refueling stations - accounts for detours from the originally intended route, due to the need to reach the EV recharging stations. Finally, the current study considers home-based EV recharging opportunities.

The remainder of the paper is organized as follows. Section "Case study: Denmark" describes the background of the case study. Section "Methodology" presents the methodology applied in the current study. The results of the empirical analysis are introduced in the "Results" section. Finally, section "Discussion and conclusions" presents a discussion of these results; conclusion are also drawn and further research is proposed in this last section.

Case Study: Denmark

Background

Denmark, renowned as a world leader in clean energy production and high energy efficiency, is seeking to decrease the fossil fuel dependency in the transport sector. Consequently, in addition to its high taxation on the purchase of private vehicles, the Danish government allocated a large budget for research regarding EVs between the years 2008 and 2011 and currently grants a tax exemption for the purchase of new EVs until 2015. Furthermore, as part of the energy strategy for 2050, the Danish government intends to establish a specific fund for kick-starting EV recharging infrastructure, encouraging EV standardization, promoting research and development efforts for renewable energy in the transport sector, and acting to tighten European Union standards on vehicle energy efficiency and CO_2 emissions. As yet another part of this energy strategy, the Danish government has an ambitious goal of 18% reduction in fossil fuel consumption in the energy and transport by 2020, and an even more ambitious goal of 100% renewable energy in 2050.

Although the current EV market share is negligible (approximately 500 registered EVs nationwide in 2011), several studies predicted its potential in Denmark on the basis of different market conditions. Eskebæk and Holst (2009) predicted an EV market share of 13% by 2020 on the basis of semi-structured interviews with key consultants in the EV and energy sectors, media articles and car industry statistics. Better Place, an international company developing EV battery-switching and charging devices, was more optimistic, expecting EVs to achieve 20% of Danish market share by 2020 (Rosted et al. 2009). Similar expectations were embraced by Kristoffersen et al. (2011). Based upon a stated preference survey among 1593 new car-buyers and the assumptions of a 150-kilometer EV range and a vehicle purchase price of about EUR 25000 Euros, Mabit and Fosgerau (2011) predict a much higher potential EV market share. Thus, the market share of EVs is expected to be significant, necessitating efficient infrastructure deployment.

Road Network and Candidate Locations

The current road network in Denmark totals 73,197 km of paved roads, including 1111 km of motorways. Along the road infrastructure there are about 2200 gasoline stations, of which 87% are considered as possible candidates for EV recharging stations. The candidate facility locations are distributed across Denmark's regions with about 13% in the capital region of Copenhagen. Regarding the distribution of the candidate locations according to the road hierarchy, 27% are located near national motorways, 43% near arterial roads, 30% near regional roads, and only 0.3% near local roads.

Technology Scenarios

The year 2020 serves as the target year for scenario development. EV off-road recharging infrastructure is assumed to be widely available to the general public in Denmark by 2020 in agreement with the goals of the Danish government for the reduction of fossil fuel dependency and the development of EV infrastructure. Specifically, two types of off-road recharging facilities are currently considered: normal plugs and recharging poles. Normal plugs facilitate overnight recharging at home with the connection to the existing electricity grid. The price of the plug is assumed to be an integral part of the EV purchase transaction and therefore its price is internalized in the EV purchase price. The recharging speed of poles depends on the number of phases and the electric current. For example, a single-phase 16 ampere recharging pole can charge a medium-size EV in 7 h, while a three-phase electric power recharging pole reduces the time to less than 3 h. As a result of off-road charging infrastructure availability at home, at workplaces and shopping facilities, the main charging demand for on-road recharging stations will comprise long-distance travelers with a daily kilometrage of over 100 km.

The representation of the road network for the target year was conducted by using the road infrastructure development for 2020 embedded in the Danish National Transport Model. The network comprises 31,533 links and contains information about the road hierarchy, directionality and number of lanes, length, and speed limit. In order to simulate realistic traffic flow conditions, average daily traffic volumes were assigned to the network and congested travel speed was calculated. Average daily traffic volumes were preferred over morning peak hour volumes since long-distance travel is distributed across daily periods.

The current study assumes a driving range of 150 km and a practical driving range of 120 km for a medium-size EV with maximum speed of 110 km/h. Currently, the new generation of EVs with Li-Ion batteries have a driving range of 120–180 km before recharging is necessary. However, these driving ranges are only obtained if cars are driving at a speed of 80 km/h, and are significantly shorter when the speed exceeds 110 km/h (Christensen 2011). Notably, the choice of vehicle

with a specific driving range by Danish drivers derives not only from technological limitations, but also from driving needs. Since over 90% of the Danish travelers have a daily kilometrage of up to 100 km, it is reasonable to assume that the main market demand would be for low-cost medium range EVs.

Two technology scenarios are evaluated on the basis of existing recharging technology: differentiating between quick-charging and battery-switching. Quick-charging stations have sufficiently high voltage to recharge 80% of the battery in approximately 20 min. Battery-switching stations replace the EV battery pack with a fully charged battery in approximately 5 min. Both scenarios assume service times according to information currently available from recharging suppliers, and no waiting times according to the assumption of sufficient capacity to provide immediate recharging services. Other than for their features, the two technologies differ in terms of their construction costs, which are expected to be about EUR 34,000 (DKK 250,000) for quick-charging stations and about EUR 400,000 (DKK 3000,000) for battery-switching stations.

In terms of externalities, the current study accounts for emissions from the usage life-cycle phase, namely emissions resulting from electricity production and vehicle tailpipe emissions. The current study assumes zero CO_2 emissions from EVs, while CO_2 emissions from fossil-fueled vehicles are assumed to decrease with the improvement of European Union Standards. The argument for zero emission from EVs is reasonable due to the Danish transition to wind and bio-mass energy, and assuming that power production is covered by the European carbon trading system Danish Government (2011). The costs of other tailpipe pollutant emissions from EVs are assumed to be roughly 40% of the tailpipe emission costs of other vehicles. The costs of EV noise emissions also comprise roughly 40% of the noise emissions from fossil-fueled vehicles. In the absence of data regarding the impact of EVs on road safety, the current study assumes the same accident costs per kilometer for EVs and fossil-fueled vehicles.

Methodology

The research methodology included four steps: (i) evaluating the need for EV recharging stations on the basis of travel and activity patterns; (ii) analyzing the induced EV market share on the basis of the optimal infrastructure deployment; (iii) examining the optimal deployment of on-road EV recharging stations to satisfy the travel demand under land-use constraints on possible recharging sites; (iv) analyzing the costs and benefits associated with the optimal deployment of infrastructure.

Identification of Travel Patterns That Necessitate Recharging

Evaluating the need for recharging stations was conducted by means of agent-based recharging heuristics on the basis of the Danish National Travel Survey (NTS), while considering expected trends regarding EV driving range and a prominent scenario regarding the deployment of EV recharging infrastructure in Denmark (see Christensen 2011).

The agent-based recharging heuristics account for daily driving distance, available time windows for recharging on the basis of the drivers' activity patterns, urban versus interurban driving cycle, season, availability and type of recharging infrastructure at activity locations, and EV capabilities, such as range and speed. Car manufacturers' data provide the input to the heuristics in terms of size, driving range, maximum speed, battery and engine capacity, and estimated market prices of EVs.

The data regarding the travel and activity patterns are extracted from the NTS dataset for the period 2006–2010 from a representative sample of 47,848 car-using respondents. The survey data consist of respondents' 24-h travel diaries detailing individual trips and activities, as well as socio-economic characteristics. Although the NTS contains data related to fossil-fueled cars, it is currently the most suitable source in Denmark for evaluating the needs of EVs derived from travel and activity patterns. The NTS contains detailed information regarding the travel patterns of one adult in each household, rather than of all household members.

The current study overcomes this limitation for households with more than one licensed driver by employing hot-deck imputation (as detailed in Andridge and Little 2010) that generates matching household members on the basis of relevant criteria for car use, such as region, urban area type, travel weekday, family type, age and gender. Following the imputation procedure, car travel patterns are generated on the basis of the number of cars and complementary car use across household members and daily periods. The sample representativeness is maintained by adjusting the weights of respondents who have a dual role as both individuals and matching household members.

Assessment of EV Market Share

The market share analysis is based on 2976 observations obtained via a stated preference (SP) survey of 372 respondents (Jensen et al. 2012). The recharging technology for on-road stations in the survey was assumed to be a generic technology with 5–10 min recharging time (Jensen et al. 2012).

A logit model was estimated for the purchase propensity of EVs as a function of vehicle characteristics and infrastructure deployment as follows (Train 2002):

Table 1 Logit model for EV market shares. Elasticities are calculated with EVs representing 1% of the car market

Variable	Unit	Range	Elasticity
Purchase price	1000 DKK	19–998[a]	−2.02
Fuel costs (EV)	DKK/km	0.14–0.52	−0.84
Fuel cost (conventional)	DKK/km	0.26–1.39	0.61
Driving range (EV)	Km	112–208	1.22
Driving range (conventional)	Km	420–910	−0.36[b]
Carbon emission (EV)	g/km	34–127	−0.95
Carbon emission (conventional)	g/km	70–234	0.74
Top speed (EV)	km/h	94–173	1.73
Top speed (conventional)	km/h	111–230	−0.98
Battery lifetime (EV)	1000 km	100–250	0.69
Number of battery stations	Amount	0–30	0.31
Charging at work place	Dummy	0/1	0.25
Charging in city centers and at larger train stations	Dummy	0/1	0.33
Charging in city centers	Dummy	0/1	0.28
Charging at larger train stations	Dummy	0/1	0.21

Source: Jensen et al. (2012)

[a]The lower bound for the purchase price reflects that in some cases the respondents chose a used reference car

[b]The parameter is non-significant at the 5% confidence level (i.e. the elasticity could be 0)

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j=1}^{J} e^{V_{nj}}} = \frac{e^{\beta' x_{ni}}}{\sum_{j=1}^{J} e^{\beta' x_{nj}}} \tag{1}$$

where P_{ni} is the probability of individual n to choose alternative i given J alternatives ($j = 1, \ldots J$), x_{ni} is the vector of alternative attributes for alternative i and individual n, and β is the vector of parameters to be estimated.

Results presented in Table 1 illustrate that consumers evaluate EVs versus fossil-fueled vehicles by considering purchase price and operating costs, vehicle capabilities, and environmental aspects. Availability of on-road recharging stations, recharging facilities at the workplace, and possibility to recharge close to home or at public parking lots only play a minor role in the acceptability of EVs.

Since the model is constructed in a hypothetical setting, it is necessary to calibrate the model to reproduce recently observed EV sales by adjusting the constant term for EVs (ASC_{EV}). This is a non-trivial task since EV sales are presently low and possibly reflect the initial state of a market penetration curve rather than an equilibrium state. In order to account for the uncertainty about the initial state of the demand curve, the model is calibrated for two alternative base-year scenarios. In the first scenario, the model is calibrated for the actual number of sold cars retrieved by using the recent sales figures from the Danish Car Importers Association. The figures show that the annual sales will probably reach 500 EVs in

2012. Nevertheless, the EVs that are currently on the market in Denmark only cover around 50% of the market because of the unavailability of the electric versions of medium sized cars, multi-purpose vehicle versions of small family cars and sport utility vehicles. Therefore, in the second scenario the model is calibrated while considering the recent market penetration of an electric version of a typical five door family cars which represents half of the market share in Denmark. According to this scenario, the EV car sales in 2013 are expected to rise to 900 cars.

Edison Model for Optimal Location of EV Recharging Stations

This study optimally locates quick-charging stations by applying an improved version of the methodology firstly developed as part of the Edison project (Electric vehicles in a Distributed and Integrated market using Sustainable energy and Open Networks, see Olsen and Nørrelund 2012). The input for the optimization procedure includes: (i) the road network and link flow speeds according to the traffic conditions; (ii) daily tours as trip sequences between origins and destinations; (iii) assumed off-road recharging infrastructure; (iv) the characteristics of the trips that necessitate recharging according to the travel pattern analysis (i.e., origin, destination and trip schedule); (v) the desired number of EV recharging stations; and (vi) candidate locations for EV recharging stations. The output of the optimization procedure consists of: (i) the optimal spatial deployment of EV recharging stations according to the defined objective function; (ii) the number of recharging incidents per station and per daily journey; and (iii) the route detour per trip due to recharging in terms of time and distance traveled. Hence, the output provides valuable insights regarding the efficiency of the infrastructure deployment from both the private investors' and the consumers' perspectives.

The model structure consists of the three stages illustrated in Fig. 1.

At the first stage, the potential recharging points per tour are identified on the basis of the chosen travel routes. The need to recharge is identified as a function of travel distance, speed and driver's assumed risk aversion. The chosen travel route is selected according to the shortest path algorithm, while allowing users to define the objective function for minimization on the basis of a linear combination of travel distance and congested travel times. The potential recharging demand points are estimated by a simulation procedure considering: a constant battery discharge rate; a homogenous risk aversion buffer of 20 km remaining driving range demarcating the decision point at which the search for a recharging location begins; a driving range of 120 km; and a recharging rate of 80% for quick-charging and 100% for battery-switching. Notably, the model assumes that recharging is possible at intermediate stops where other activities are performed, provided that the activities are of sufficient duration.

Figure 2 illustratesthis concept, as a single tour from A to B is presented at the top of the figure. The tour represents an actual tour extracted from the Danish National Travel Survey. At the top, the intersections and cumulative distance (in kilometers)

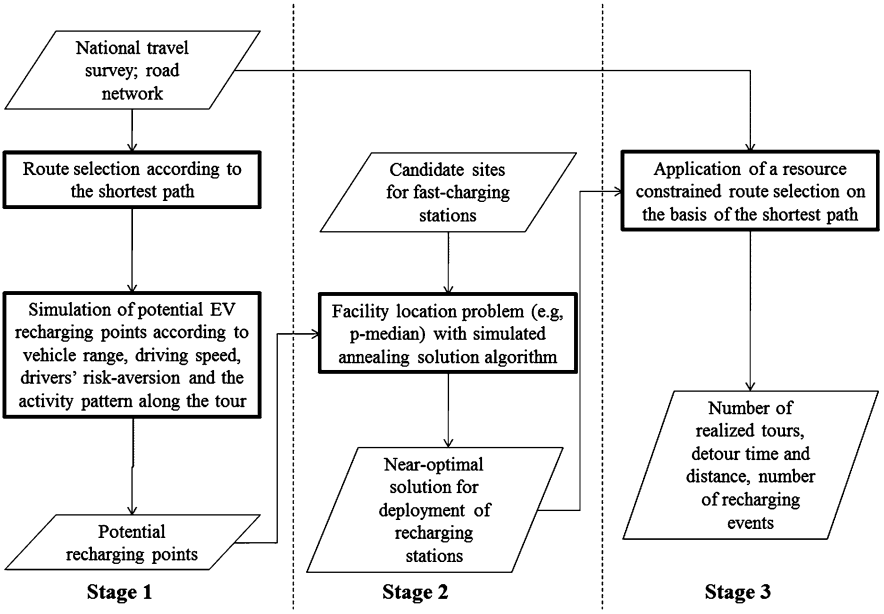

Fig. 1 Edison model structure for the optimal location of recharging stations

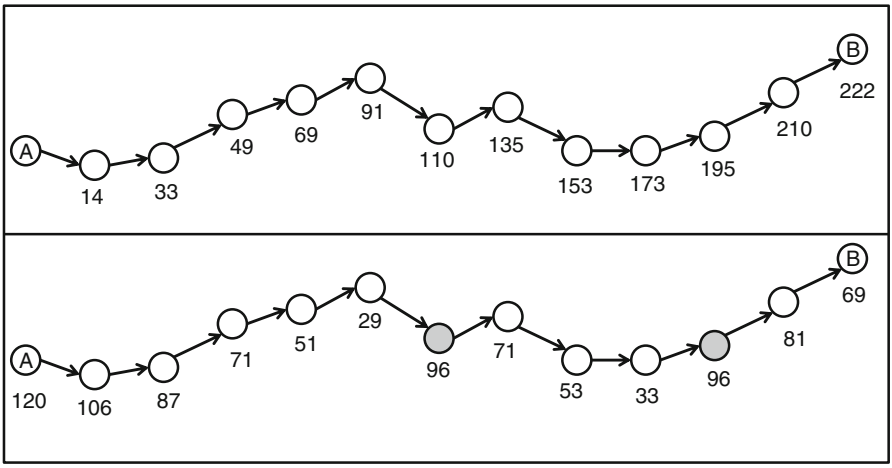

Fig. 2 Calculating potential recharging points. *Top*: A tour from node A to node B extracted from the Danish National Travel Survey; cumulative distance from node A is depicted. *Bottom*: The same tour is illustrated with remaining driving range and potential recharging points

along the tour are depicted. At the bottom, the potential recharging points that result from this tour are presented (assuming that the quick-charging technology is used) with the remaining driving range (in kilometers) at each node according to the

discharge rate of the battery. While traveling from the sixth to the seventh node heading from A to B, the vehicle exceeds the 20 km threshold and the seventh node is consequently selected as a potential recharging point. The range of the vehicle is reset to 96 km (80% of the battery range of 120 km), assuming that the vehicle is quick-charged at this point. Between the tenth and eleventh node the remaining range drops below 20 km again, and node 11 is selected as the second potential recharging point. The potential recharging points for all tours are used as input (EV demand points) to the facility location model.

At the second stage, the model seeks the optimal deployment of recharging stations on the basis of the distribution of the potential recharging points and the candidate sites for EV recharging stations. The current study employs a *p*-median facility location problem (as in Tansel et al. 1983) to find the near-optimal deployment of *p* recharging stations by minimizing the total driving distance/travel time from EV demand points to the recharging stations. Because the number of facilities is too large to be solved to proven optimality, a simulated annealing meta-heuristic is used to find a near-optimal solution of the *p*-median facility location problem (for details, see Olsen and Nørrelund 2012).

At the third stage, given the optimal deployment of the recharging stations, travel routes between origins and destinations along the tour as a trip sequence are re-selected on the basis of a resource constrained shortest path algorithm (Irnich and Desaulniers 2005). Hence, trips that require en-route recharging are re-routed in order to perform the recharging task at recharging stations, resulting in detours. Trips that cannot be served by the recharging infrastructure due to the length of the detour are aborted, thus simulating the share of trips that would not be performed due to the lack of infrastructure. The number of recharging incidents per tour and per station is also calculated at the third stage. Further details regarding the algorithm employed in stage three can be found in Røpke et al. (2012).

Socio-economic Evaluation

The socio-economic analysis accommodates the calculation of user benefits and detour costs, investments and operational costs, externalities (e.g., noise emissions, greenhouse gases and other pollutant emissions), and tax distortion.

The current study employs a utilitarian approach for estimating the road user benefits. In particular, the consumer surplus due to the change in the number of on-road recharging stations is calculated on the basis of the logit model for the purchase of EVs as follows (Train 2002):

$$\Delta E\left(CS_n\right) = \frac{1}{\alpha_n}\left[\ln\left(\sum_{j=1}^{J^1} e^{V_{nj}^1}\right) - \ln\left(\sum_{j=1}^{J^0} e^{V_{nj}^0}\right)\right] \tag{2}$$

where α_n is the estimated car price parameter, and the indices zero and one refer, respectively, to the base case and the change in the attribute. This expression accounts for the perceived costs and benefits during the hypothetical vehicle purchase in the SP experiment, and can be interpreted as the percentage change in the consumer's willingness to pay as a result of a change in one EV attribute. The total user benefits are obtained by multiplying the consumer surplus in the total new car sales (conventional cars and EV) after the change in the attributes.

The detour costs are calculated independently and added as a cost, with the argument that the daily detour costs could not be foreseen by the respondents to the SP survey. The survey only provided information about an ideal recharging time of 5–10 min per visit; it did not provide any information regarding the daily number of recharging stops required or the need to make a detour in order to reach recharging stations while on long distance trips. The detour costs are calculated on the basis of the value of time DKK 80 per hour used in transport feasibility studies in Denmark. A penalty for infeasible trips resulting from lack of recharging infrastructure along the route is assumed to amount to three extra traveling hours. The detour externalities are included as costs.

The costs associated with the recharging infrastructure are the operation and maintenance costs of the recharging stations and charging poles in the cities. The total building cost of the battery-switching stations is approximately DKK 3 million; operating cost is assumed to be 10% of annual investment costs.

The externalities cover climate change, air pollution and noise and are based upon the assumptions in the Danish Transport Economic Unit Prices (Danish Ministry of Transport 2010). CO_2 emissions from EVs are set to zero within the European Emission Trading System. The cap on the total emissions from heavy industry and power production means that extra emissions due to a larger EV fleet are offset by reductions elsewhere in Europe. It is assumed that, apart from recharging detours, the total amount of traffic remains the same - i.e. the annual kilometrage of both EVs and conventional cars remains 18,000 kilometers.

The tax distortion is set to 20% of the net revenue loss in accordance with the guidelines from the Danish Ministry of Finance (1999).

All of the costs and benefits over the years 2012–2030 are calculated as the net present value in 2012 with a 5% interest rate.

Results

Demand for EVs

The model presented in Table 1 is used to predict the market share of new EV sales. The prediction is based on the assumption that the market price of EVs excluding the battery will decrease to the level of conventional cars in 2020, and that the battery price would decrease by approximately 40% (Danish Energy Authority 2011). In addition, the tax exemption for EVs will be in place until 2015, but the registration

tax for EVs would still be lower than the registration tax of conventional cars due to their higher energy efficiency.

According to the demand model predictions, considering the alternative base-year sales scenarios of 500 EVs (low-demand) and 900 EVs (high-demand), respectively, the EV fleet in 2020 will increase to 8600 for the first scenario and 15,200 vehicles for the second scenario (comprising a maximum of 0.6% of the total vehicle fleet) in the absence of on-road recharging infrastructure. The effect of the additional recharging stations can only be calculated for the battery-switching technology because the SP survey considered only an ideal recharging time of 5–10 min. The additional demand for EVs in 2020 that is generated by the deployment of on-road battery-switching infrastructure is predicted to be 3000–5000 vehicles with 15 battery-switching stations; 7000–12,000 with 30 stations; and 14,000–26,000 with 50 stations, under the conditions of the low-demand and high-demand scenarios, respectively.

Travel Patterns and Infrastructure Deployment

Table 2 presents the average number of recharging visits per daily tour, the average detour time per recharging visit, and the average detour time and distance per daily tour. The distribution of the number of recharging visits per daily tour is a decaying exponential function with most travelers recharging only once or twice daily. While the detour distance is significantly reduced with the increase in the number of recharging stations, the detour time does not significantly decrease, indicating that the detour time is mostly a result of the recharging time rather than the detour travel time.

A small share of the travelers cannot reach a recharging station if only 15 recharging stations are deployed. This share is negligible when the number of facilities is increased to 50 quick-charging stations or 30 battery-switching stations.

Notably, the recharging time per daily tour for the quick-charging technology is more than triple the recharging time of the battery-switching technology. Under these conditions, it can be assumed that the quick-charging technology would not be an inducement to purchase a new EV for the purpose of long-distance travel.

Figure 3 shows the optimal locations of EV recharging stations and the number of recharging visits at each station. For the quick-charging technology, the number of visits is based upon a total EV fleet of 15,200 vehicles in 2020. Notably, while it is not assumed that this technology would be an encouraging factor in the purchase of EVs for long-distance travel, it is assumed that consumers who already bought EVs would use them also for the purpose of long-distance travel. For the battery-switching technology, Fig. 3 depicts the high-demand scenario. As expected, most of the stations - as well as the busiest stations - are located along the main national highways outside urban conurbations.

For two reasons the location of recharging stations is different for the two technologies. The first reason is that the range of an EV after a quick charge is lower

Table 2 Results from the recharging infrastructure deployment model for 15, 30 and 50 quick-charging stations and battery-switching stations

	Quick-charging stations			Battery-switching stations		
	15	30	50	15	30	50
Average number of recharging visits per daily tour (number of visits)	2.35	2.10	2.06	1.91	1.84	1.81
Average detour and recharging time per daily tour (minutes)	67.03	52.63	48.94	21.18	18.64	16.52
Average detour travel time per daily tour (minutes)	19.96	10.55	7.69	11.64	9.46	7.48
Average detour distance per day (kilometers)	21.45	8.66	5.52	11.09	8.62	5.76
Average detour and recharging time per recharging visit (minutes)	28.48	25.02	23.73	11.10	10.15	9.14
Average detour distance per recharging visit (kilometers)	9.11	4.12	2.67	5.81	4.69	3.18
Share of cars that need to charge which are unable to charge (percent)	5.7	3.9	1.0	3.5	0.8	0.3

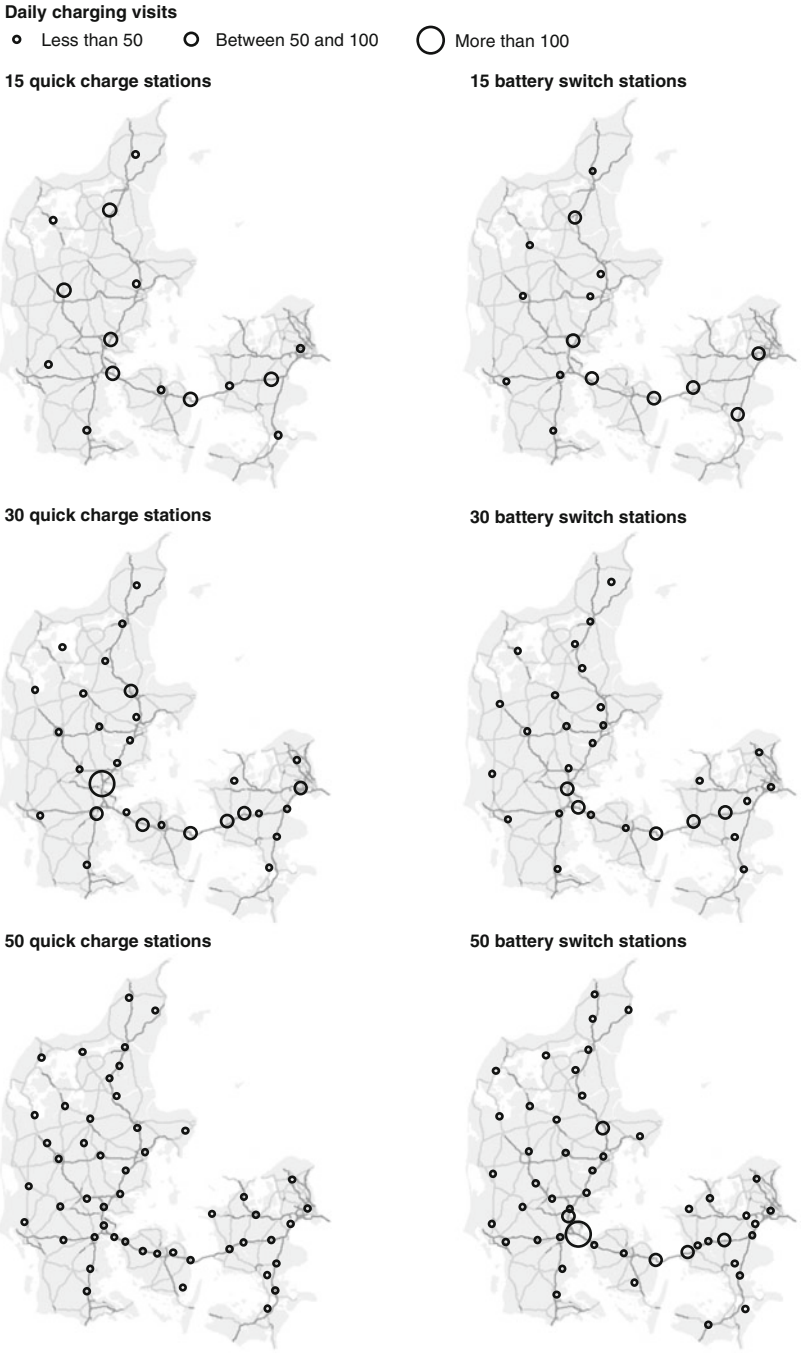

Fig. 3 The locations of 15, 30 and 50 quick-charging and battery-switching-stations

Table 3 Results of the socio-economic analysis under the high-demand scenario (900 EV sales in 2013)

Monetary benefits (Millions DKK[a])	15 stations	30 stations	50 stations
Consumer surplus	1366	3215	6713
Tax distortion	−328	−767	−1579
Investment and operation	−114	−228	−379
Externalities (excluding CO_2)	134	312	638
Detour costs	−58	−103	−173
Net benefits (million DKK)	1000	2429	5220
CO_2 reduction (kiloton)	159	372	770

[a]1 EUR = 7.5 DKK

than the range of an EV after a battery change. The second reason is that the set of locations is the output of a stochastic search procedure, which is a near-optimal solution.

Socio-economic Analysis

According to Table 3, in the case of the high-demand base-year scenario, the benefit of deploying 15 battery-switching stations and equipping all EVs with a switchable battery is assessed to be DKK 1.0 billion. In the case of the low-demand scenario, the benefit is halved.

The results are highly dominated by a road user benefit of more than DKK 1.366 billion calculated as the consumer surplus. Notably, the benefit of the reduced externalities is only 10% of the consumer surplus. The most important cost is a tax distortion, which consists of government revenue loss on purchase tax, energy taxes, etc. Investment and operation of the recharging infrastructure, and the detour costs, both have a lesser effect.

The socio-economic benefit gained by the addition of one recharging station increases with the number of stations, from DKK 67 million for 15 stations to DKK 162 million for 30 stations, and DKK 261 million for 50 stations. This increase is due to the perceived consumer benefit due to the additional recharging opportunities, as well as to the increase in the EV fleet as a result of the increase in the number of stations.

Discussion and Conclusions

The results show that wide-scale market penetration of EVs is correlated with infrastructure deployment. Deployment. This finding is in accordance with previous studies (Christensen et al. 2010; Hidrue et al. 2011; Stathopoulos and Marcucci

2012). Furthermore, the results show the importance of efficiency in deployment of EV recharging stations, as only 15 battery-switching stations or 30 quick-charging stations (1–2% of the current infrastructure) are sufficient for satisfying the recharging needs of 96% of the EV stock. Notably, the results obtained are under the assumption that by 2020 in Denmark, EV off-road recharging infrastructure will be widely available at home, at workplaces and at shopping centers for use by the general public. Therefore, the main demand for on-road recharging stations will be comprised of long-distance travelers with a daily kilometrage of over 100 km.

The results indicate that, under the assumption of optimal infrastructure deployment, the main reason for time and production losses as a result of recharging detours is related to the recharging time rather than to the detour travel time or detour distance. The results show that for long-distance travel the average detour time for a long-distance daily tour including recharging with quick-charging technology is about 50–60 min. Considering that long recharging time is among the three main concerns of consumers along with range anxiety and purchase price (Hidrue et al. 2011), spending 50–60 min per day at quick-charging stations may be a severe barrier to EV market penetration. This barrier is largely alleviated if battery-switching is considered since the recharging time reduces to only 15–20 min for a daily long-distance tour, which is nearly equivalent to re-filling a gasoline tank several times. Moreover, battery-switching would be associated with lower driving range anxiety due to the higher recharging capacity allowing fewer daily recharging incidents. Thus, according to the results of the current study, battery-switching seems a better solution in terms of alleviating the barriers for wide-scale EV adoption.

The current study does not explicitly incorporate capacity constraints. However, the results indicate that some recharging stations will serve over 50 cars daily. Therefore, stations should be designed to accommodate the daily distribution of recharging visits with adequate capacity in order to avoid aditional waiting time.

The analysis shows that a reduction of 160–770 Kilotons in CO_2 is feasible for the target year of 2020, assuming a relatively modest share of EVs comprising 0.7–1.5% of the total vehicle stock. This reduction is feasible without major policy changes apart from full availability of off-road recharging options at home and at activity locations, and the efficient deployment of EV recharging infrastructure. Interestingly, this result is in agreement with the assessment of the Rotterdam Climate Initiative (RCI) that such a reduction is possible by introducing green vehicles and fuels (Geerlings 2012).

The socio-economic analysis shows the positive net benefit of providing battery-switching station infrastructure. Results show that of the benefits of EVs, a large part is related to the willingness to pay, which is estimated on the basis of an SP survey. Notably, SP survey are associated with a high degree of uncertainty; in particular, SP surveys are susceptible to compatibility bias and strategic response bias. The former bias occurs when respondents are not responsible for the consequences of their selection, while the latter bias occurs when respondents anticipate that their responses would influence product design. Compatibility bias could result in overestimation of the consumer willingness to pay, while Strategic response bias

could result is overestimation of the required EV features, for example speed and range. Nevertheless, SP surveys are the best tools for investigating technologies with little or no market penetration. Bearing these limitations in mind, the current study shows that even with a relatively modest market share, promoting the EV could be beneficial.

The current study is the first analysis of the deployment of EV recharging stations from a comprehensive socio-economic perspective. However, the study is not without limitations, and as such it helps to uncover several interesting issues for further research regarding optimal location of EV recharging infrastructure. First, the current study is based upon relatively conservative assumptions regarding off-road recharging infrastructure, fuel prices, EV market share and driving range. Other, less conservative, scenarios could be considered for further research, in particular with respect to fuel prices. Second, this study is conducted under the assumption that the travel patterns and route selection are rational and known. A beneficial future line of research would be to incorporate uncertainty as well as bounded rationality into the model. Third, the current study is conducted under the assumption of population homogeneity - for example with respect to risk aversion. However, it would be beneficial to incorporate population heterogeneity within the decision-making processes related to recharging. Fourth, the current study assumes that the recharging stations do not have capacity constraints and that travelers do not learn from their previous recharging experience. Hence, it would be beneficial to incorporate both capacity constraints and learning experience by allowing feedback across decision models. In conclusion, the current study is based upon a single-technology demand function. However, the results indicate that it would be beneficial to explore data collection regarding travelers' preferences underlying the choice between competing recharging technologies - namely recharging time, the number of daily charging visits and charging costs.

Acknowledgments The current study comprises a portion of the "Greening European Transportation Infrastructure for Electric Vehicles" project funded by the Trans-European Transport Network (TEN-T) program, as well as a portion of the Edison project funded by the Danish ForskEl program. The authors wish to acknowledge the contribution of Stefan Mabit and Anders Fjendbo Jensen to the market share model, and the contribution of Min Wen, Gilbert Laporte, Oli Madsen, and Anders Nørrelund to the development of the Edison model.

References

Ahn, C., Li, C. T., & Peng, H. (2011). Optimal decentralized recharging control algorithm for electrified vehicles connected to smart grid. *Journal of Power Sources, 196*, 10,369–10,379.

Amoroso FA, Cappuccino G (2011) Impact of charging efficiency variations on the effectiveness of variable-rate-based recharging strategies for electric vehicles. Journal of Power Sources 196: 9574–9578.

Andersen, P. H., Mathews, J. A., & Rask, M. (2009). Integrating private transport into renewable energy policy: The strategy of creating intelligent recharging grids for electric vehicles. *Energ Policy, 37*, 2481–2486.

Andridge, R. R., & Little, R. J. A. (2010). A review of hot deck imputation for survey non-response. *International Statistical Review, 78*, 40–64.

Blair PD (1978) Modeling energy and power requirements of electric vehicles. Energy Conversion 18:127–134.

Brady, J., & O'Mahony, M. (2011). Travel to work in Dublin: The potential impacts of electric vehicles on climate change and urban air quality. *Transportation Research Part D-Transport and Enivornment, 16*, 188–193.

Brown, S., Pyke, D., & Steenhof, P. (2010). Electric vehicles: The role and importance of standards in an emerging market. *Energ Policy, 38*, 3797–3806.

Calasanzio, D., Maja, M., & Spinelli, P. (1993). Fast recharging of lead/acid batteries. *Journal of Power Sources, 46*, 375–381.

Carmody, W. J., & Haraden, J. (1992). An electric grid for transportation in Los Angeles. *Energy, 17*, 761–767.

Charlesworth, G., & Baker, T. M. (1978). Transport fuels for the post-oil era. *Energ Policy, 6*, 21–35.

Chéron, E., & Zins, M. (1997). Electric vehicle purchasing intentions: the concern over battery charge duration. *Transportation Research Part A-Policy and Practice, 31*, 235–243.

Christensen, L. (2011). Electric vehicles and the costumers. Report WP 1.3 version 0.1, EDISON project. http://www.edison-net.dk/Dissemination/Reports/Report_011.aspx

Christensen, L., Kveiborg, O., & Mabit, S. L. (2010). The market for electric vehicles – What do potential users want? In: *Proceedings of the 12th world conference on transport research*, Lisbon, Portugal.

Dagsvik, J. K., Wennemo, T., Wetterwald, D. G., & Aaberge, R. (2002). Potential demand for alternative fuel vehicles. *Transportation Research Part B: Methodological, 36*, 361–384.

Danish Energy Authority. (2011). Report on the conditions for the deployment of charging stations for EVs (in Danish). http://www.ens.dk/da-DK/Info/Nyheder/Nyhedsarkiv/2011/Documents/Redeg_ladestandere_elbiler_jan2011_final.pdf

Danish Government (2011) *Energy strategy 2050 – From coal, oil and gas to green energy*. The Danish Ministry of Climate and Energy.

Danish Ministry of Finance. (1999). Guide to the preparation of socioeconomic evaluations (in Danish). http://www.fm.dk/publikationer/1999/vejledning-i-udarbejdelse-af-samfundsoekonomiske-konsekvensvurderinger/

Danish Ministry of Transport. (2010). Transport economic unit prices 2010 (in Danish). www.dtu.dk/centre/Modelcenter/Samfunds%c3%b8konomi/Transport%c3%b8konomiske%20Enhedspriser.aspx

De Luchi, M., Wang, Q., & Sperling, D. (1989). Electric vehicles: Performance, life-cycle costs, emissions, and recharging requirements. *Transportation Research Part A-Policy and Practise, 23*, 255–278.

Eskebæk, L., & Holst, J. (2009). *Electric vehicles on the Danish market in 2020 and the economic value of the company Better Place*. Master Thesis, Copenhagen Business School.

Fernández, M., & Trinidad, F. (1997). Recharging strategies for valve-regulated lead/acid batteries in electric-vehicle applications. *Journal of Power Sources, 67*, 125–133.

Frick, M., Axhausen, K. W., Carle, G., & Wokaun, A. (2007). Optimization of the distribution of compressed natural gas (CNG) refueling stations: Swiss case studies. *Transportation Research Part D-Transport and Environment, 12*, 10–22.

Geerlings, H. (2012). The governance of the introduction of electric vehicles in Rotterdam. Presented at the NECTAR Cluster 2 Workshop on Urban and Regional Transport, Regulation Sustainability and E-Mobility, Dresden, Germany.

Giese, R., Jones, P. C., & Kroetch, B. G. (1983). Electric vehicles: Market penetration and positive externalities. *Technological Forecasting and Social Change, 24*, 137–152.

Hamilton, W. (1980). Energy use of electric vehicles. *Transportation Research Part A-Policy and Practise, 14*, 415–421.

Hartmann, N., & Özdemir, E. D. (2011). Impact of different utilization scenarios of electric vehicles on the German grid in 2030. *Journal of Power Sources, 196*, 2311–2318.

Hidrue, M. K., Parsons, G. R., Kempton, W., & Gardner, M. P. (2011). Willingness to pay for electric vehicles and their attributes. *Resource and Energy Economics, 33*, 686–705.

Irnich, S., & Desaulniers, G. (2005). Shortest path problems with resource constraints. In G. Desaulniers, J. Desrosiers, & M. M. Solomon (Eds.), *Column generation* (pp. 33–65). New York: Springer.

Jensen, A. F., Cherchi, E., & Mabit, S. L. (2012). On the stability of preferences and attitudes before and after experiencing an electric vehicle. In: *13th international conference on travel behaviour research*, Toronto, Canada.

Kim, J. G., & Kuby, M. (2012). The deviation-flow refueling location model of optimizing a network of refueling stations. *International Journal of Hydrogen Energy, 37*, 5406–5420.

Kristoffersen, T. K., Capion, K., & Meibom, P. (2011). Optimal recharging of electric drive vehicles in a market environment. *Applied Energy, 88*, 1940–1948.

Kuby, M., & Lim, S. (2005). The flow-refueling location problem for alternative-fuel vehicles. *Socio-Economic Planning Sciences, 39*, 125–145.

Kuby, M., Lines, L., Schultz, R., Xie, Z., Kim, J. G., & Lim, S. (2009). Optimization of hydrogen stations in Florida using the flow-refueling location model. *International Journal of Hydrogen Energy, 34*, 6045–6064.

Kurani KS, Turrentine T, Sperling D (1994) Demand for electric vehicles in hybrid households: An exploratory analysis. Transport Policy 1:244–256.

Li, Z., & Ouyang, M. (2011). The pricing of recharging for electric vehicles in China: Dilemma and solution. *Energy, 36*, 1–14.

Li, Z., Sahinoglu, Z., Tao, Z., & Teo, K. H. (2010). Electric vehicles network with nomadic portable recharging stations. In: *Proceeding of the IEEE 72nd vehicular technology conference* (pp. 1–5).

Mabit SL, Fosgerau M (2011). Demand for alternative-fuel vehicles when registration taxes are high. Transportation Research Part D-Transport and Environment 16:225–231.

Mullan, J., Harries, D., Bräunl, T., & Whitely, S. (2011). Modelling the impacts of electric vehicle recharging on the Western Australian electricity supply system. *Energ Policy, 39*, 4349–4359.

Nicholas, M. A., Handy, S. L., & Sperling, D. (2004). Using geographic information systems to evaluate siting and networks of hydrogen stations. *Transportation Research Record, 1880*, 126–134.

Olsen, A., & Nørrelund, A. V. (2012). Location of future fast charge stations. Appendix 1 in Christensen, L., (2011) *Electric vehicles and the costumers.* Report WP 1.3 version 0.1, EDISON project. http://www.edison-net.dk/Dissemination/Reports/Report_011.aspx

Pearre, N. S., Kempton, W., Guensler, R. L., & Elango, V. V. (2011). Electric vehicles: How much range is required for a day's driving? *Transportation Research Part C-Emergency, 19*, 1171–1184.

Perujo, A., & Ciuffo, B. (2010). The introduction of electric vehicles in the private fleet: Potential impact on the electric supply system and on the environment. A case study for the province of Milan, Italy. *Energ Policy, 38*, 4549–4561.

Ropke, S., Christensen, L., Kaplan, S., Madsen, O. B. G, & Olsen, A. (2012). *Evaluating the placement of charging stations for electrical vehicles: an application of the resource constrained shortest path problem*. DTU Transport working paper.

Rosted, J., Kjeldsen, C., Bisgaard, T., & Napier, G. (2009). *New nature of innovation*. Report submitted to the OECD Committee for Industry, Innovation, and Entrepreneurship. http://www.newnatureofinnovation.org/full_report.pdf

Skippon S, Garwood M (2011) Responses to battery electric vehicles: UK consumer attitudes and attributions of symbolic meaning following direct experience to reduce psychological distance. Transportation Research Part D-Transport and Environment 16:525–531.

Stathopoulos, A., & Marcucci, E. (2012). Modelling the purchase of alternative fuel vehicles: intangible determinants? Presented at the NECTAR Cluster 2 Workshop on Urban and Regional Transport, Regulation Sustainability and E-Mobility, Dresden, Germany.

Tansel, B. C., Francis, R. L., & Lowe, T. J. (1983). Location on networks: A survey. Part I: the p-center and p-median problems. *Management Science, 29*, 482–497.

Train, K. (2002). *Discrete choice methods with simulation*. Cambridge, MA: Cambridge University Press.

Van den Bossche, P. (2010). Electric vehicle recharging infrastructure. In: G. Pistoia (Ed.), *Electric and hybrid vehicles* (pp. 517–543). Elsevier.

Wang, Y. W., & Lin, C. C. (2009). Locating road-vehicle refueling stations. *Transportation Research Part E-Logistics & Transportation Review, 45*, 821–829.

Wang, Y. W., & Wang, C. R. (2010). Locating passenger refueling stations. *Transportation Research Part E: Logistics and Transportation Reviews, 46*(5), 791–801.

Wen, M., Laporte, G., Madsen, O. B. G., Nørrelund, A. V., & Olsen, A. (2012). *Locating replenishment stations for electric vehicles*. Manuscript submitted for publication.

Yang, X. G., & Liaw, B. Y. (2001). Fast-charging nickel- metal hybrid traction batteries. *Journal of Power Sources, 101*, 158–166.

Transport Strategies in Reverse Logistics for Establishing a Sound Material-Cycle Society

JongJin Yoon and Shigeru Morichi

Abstract Improvement of transport efficiency in reverse logistics contributes not only to direct efficiency gains but also to promotion of waste recycling. This paper first examines the existing situation of waste management and related reverse logistics systems in Japan. Some examples of efficient reverse logistics systems in the United States are also discussed. In considering the case of the Tokyo Prefecture area, alternative transport systems for waste management reverse logistics are evaluated using a vehicle routing model based upon a genetic algorithm. The modeling results show that the milk-run type cooperative truck transport in combination with bulk-hauling railway transport is the most cost-effective option with significant benefits in terms of lower CO_2 emissions.

Keywords Reverse logistics • Reverse chains • Recycling • Cooperative transport

Introduction

Because of the limits of final disposal sites and finite natural resources, a Sound Material-Cycle Society is needed, which is defined by the Japanese Ministry of the Environment (2010) as:

> . . . a society in which the amount of new resource extraction is minimized at all stages of social and economic activities, from resource extraction through production, distribution, consumption and disposal, through a range of measures such as reduction of waste generation and use of circulative resources, thereby minimizing environmental loads. (p. 136)

J. Yoon (✉)
CTI Engineering Co., Ltd, Tokyo, Japan
e-mail: yoon.jongjin@gmail.com

S. Morichi
National Graduate Institute for Policy Studies, Tokyo, Japan

J.-C. Thill (ed.), *Spatial Analysis and Location Modeling in Urban and Regional Systems*, Advances in Geographic Information Science,
https://doi.org/10.1007/978-3-642-37896-6_16

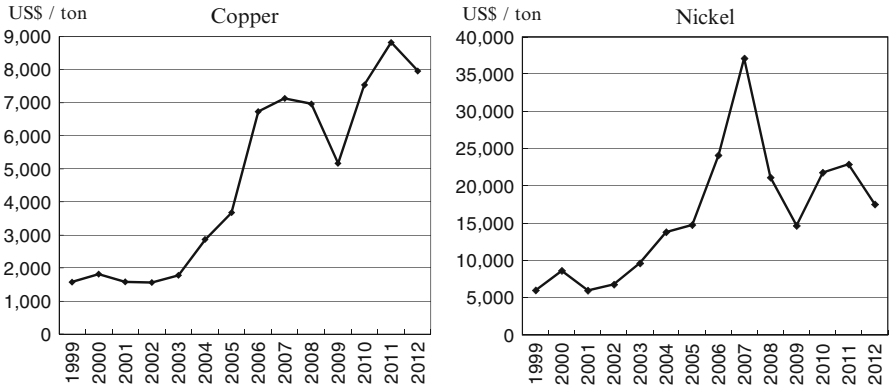

Fig. 1 Trends in metals pricing, 1999–2012 (Source: Elaborated from International Monetary Fund data (2013))

In Japan, about 35 million tons of waste goes to landfills for final disposal annually (Yoon, 2007a) and, according to the Ministry of the Environment, the existing capacity of final disposal sites is sufficient only for the next 10 years. A shortage of final disposal sites for waste is expected to become a serious problem owing to the limited availability of land in Japan. Moreover, recent surges in the price of metal commodities have emphasized the importance of resource conservation through recycling and reuse. The price of copper has increased fivefold, and that of nickel increased threefold in comparison with 1999 pricing (Fig. 1). Therefore, it may be financially beneficial to recycle waste as raw materials, though further investigation is needed to determine whether the costs of the recycling process provide savings over the cost of the metals and other commodities at current pricing. Clearly, the recycling of waste can serve not only to reduce waste disposed of in final landfill sites but also to make sustainable use of finite and costly natural resources.

However, the lack of efficient reverse logistics processes in Japan, such as integrated waste collection and intermodal waste transport systems, is the principal barrier to promoting efficient recycling or waste disposal management. For example, in Japan, waste is transported mainly by trucks in small lots because the scale of firms involved in waste collection and transportation is generally very small. As a result, although transport cost accounts for about 3% of costs across all industries, on average, transport cost in industries such as waste treatment and recycling accounts for about 20–30% of all costs (Yoon 2007b).

The waste stream consists of materials and resources of very low or negative marginal value. For these, the cost of logistics can often be the crucial factor in determining whether a material stream is considered a waste or a resource, and therefore whether it enters the disposal stream or the reprocessing stream. Such logistics costs may include storage, inter-change, handling, transportation, and administration. Low-cost transport could turn low value non-hazardous waste into a viable resource (Alker and Ravetz 2006). It can therefore be said that

without significant improvement in reverse logistics, we cannot achieve the needed efficiency in waste management and recycling operations.

Reverse logistics encompasses the logistics management skills and activities involved in reducing, managing, and disposing of wastes (Kopicki et al. 1993). Reverse logistics is defined by Stock (1998) as:

> . . . the term refers to the role of logistics in product returns, source reduction, recycling, materials substitution, reuse of materials, waste disposal, and refurbishing, repair and remanufacturing. (p. 20)

Additionally, Rogers and Tibben-Lembke (1999) define reverse logistics as:

> . . . the process of planning, implementing, and controlling the efficient, cost-effective flow of raw materials, in-process inventory, finished goods, and related information from the point of consumption to the point of origin for the purpose of recapturing value or proper disposal. (p. 2)

Significant research related to reverse logistics has been accomplished as follows. Fleischmann (2001) analyzes mainly quantitative models for reverse logistics, and Prahinski and Kocabasoglu (2006) review the literature in reverse supply chain management (RSCM) including materials recovery, product upgrade and so on. Dowlatshahi (2000) categorizes literature on reverse logistics by topic into five groups, and analyzes it. Lee et al. (2002) discuss an effective and efficient reverse logistics program for OEMs (original equipment manufacturers). Kopicki et al. (1993), Johnson (1998), Spicer and Johnson (2004), and Krumwiede and Sheu (2002) discuss the role of third-party logistics (3PL) providers in reverse logistics. Alker and Ravetz (2006) insist that developing appropriate locations of facilities such as large recycling sites, Energy for Waste (EFW) facilities, incinerators and landfills which have an input to the strategic transport network for wastes as part of an integrated and efficient intermodal waste transport system will help to shift the balance from waste disposal towards resource management.

Until the 1980s, the scope of reverse logistics was limited to the movement of material against the primary flow, from the customer toward the producer (Rogers and Tibben-Lembke 2001). Due to the following economic and environmental aspects of reverse logistics (Krumwiede and Sheu 2002; Prahinski and Kocabasoglu 2006), the importance of reverse logistics has increased and the definition has been expanded to also include the following:

- Environmental aspects: the site- and capacity-related limits of final disposal sites; the finite nature of natural resources; and increased consumer environmental awareness
- Economic aspects: the need for management of product returns and increased sales opportunities in secondary and global markets

Though the importance of studying reverse logistics and engaging in reverse logistics activities has increased due to these economic and environmental aspects, limited data has restricted research on the subject in Japan.

In light of this background, this paper analyzes the current state of reverse logistics in Japan, mainly considering industrial waste and private collection companies undertaking reverse logistics. Utilizing data from governmental and

private sector sources, including interviews, key policy issues are explored and quantitative analysis is undertaken in order to assess policy measures to respond to the critical issues. Finally, policy suggestions are made based upon the results of the analysis.

The Current State of Reverse Logistics in Japan

Supply chain management (SCM) can be defined by Simchi-Levi et al. (2000) as:

> . . . a set of approaches utilized to efficiently integrate suppliers, manufacturers, warehouses, and stores, so that merchandise is produced and distributed at the right quantities, to the right locations, and at the right time, in order to minimize system-wide costs while satisfying service level requirements. (p. 1)

Most SCM research concentrates on the forward movement and transformation of the materials from the suppliers to the end-user/consumer. However, the reverse flow of products from consumers to upstream businesses has not received much study (Prahinski and Kocabasoglu 2006).

Therefore, we examine the current status and problems of reverse logistics from the following two viewpoints: (a) improvement of transport efficiency, and (b) supply chain integration.

Improvement of Transport Efficiency

In comparison with forward logistics, Yoon (2007a) listed characteristic features of reverse logistics in Japan as "rarely time sensitive: the converse of just-in-time logistics" (p. 335), further characterizing reverse logistics as a small lots transportation endeavor converging upon waste management facilities. When such characteristics are considered, it appears that the efficiency of logistics systems may potentially be improved—perhaps by changing the reverse logistics transportation method to cooperative transportation and "milk-run"-type routing. Milk-run routing refers to the means of transportation where a single truck cycles around multiple suppliers to collect or deliver freights (Kitamura and Okamoto 2012). In addition, efficiency may potentially be improved by shifting the transport mode from truck to bulk capacity railway or coastal shipping.

However, in Japan, waste is transported mainly by trucks in small lots and "cooperative transport," in which a number of shippers or carriers jointly operate freight vehicles to reduce their costs for collecting and delivering goods and provide higher levels of service to their customers, is for the most part not practiced in reverse logistics systems. As the scale of firms involved in waste collection and transportation is generally very small, it has not been feasible to construct a network of waste generators, which is necessary in milk-run type transportation routing.

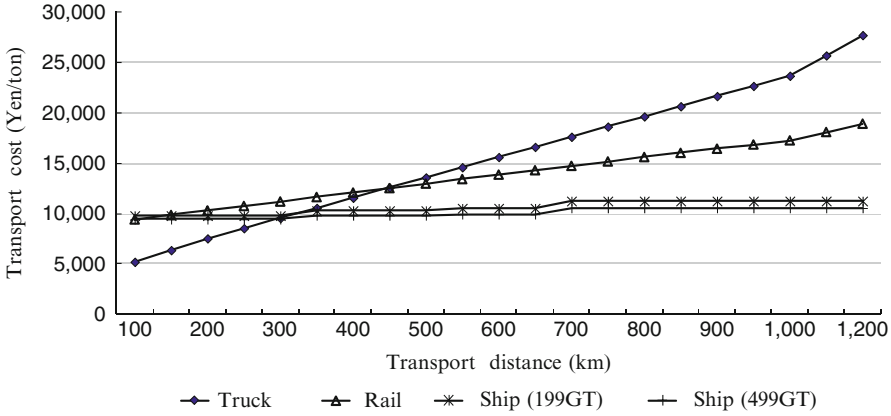

Fig. 2 Transport cost in Japan by transport distance and transport mode (Source: Elaborated from Koutsuunihon Inc. data (2004) and Japan Maritime Center data (2000))

It is notable that in forward logistics, according to the 7th National Cargo Net Flow Report (logistics census) for the year 2000 (Ministry of Land, Infrastructure and Transport of Japan 2002), trucking accounts for 83% of goods transport, while railway and shipping account for 13%. However, if ferrous scrap is excluded, 99% of reverse logistics contents are transported by truck, which indicates that the potential for efficiency improvements through modal shift has perhaps not yet been explored.

As illustrated in Fig. 2, if transport cost is plotted by transport distance and examined by mode, for a transport distance up to 300 km, trucking has a cost advantage. For distances over 300 km, maritime shipping is most cost effective. Moreover, if the distance exceeds 450 km, rail is less costly than trucking, but more costly than maritime shipping.

For this reason, as transport distance increases, the modal share of rail and maritime shipping increases in forward logistics (Fig. 3). Yet, irrespective of transport distance, the share of trucking is currently dominant in reverse logistics (Fig. 3). This indicates the importance of exploring the use of rail transport and coastal shipping, as well as other essential processes and practices of reverse logistics (Yoon 2007b).

Chain Integration

In forward logistics, minimization of system-wide costs is achieved through supply chain management. However, for reverse logistics, there is a lack of so-called chain strategy which can efficiently manage reverse logistics via integrated networks.

The principal factors behind this are: (1) unlike the case of production and sales activities in forward logistics, there is no incentive for integration of reverse chains;

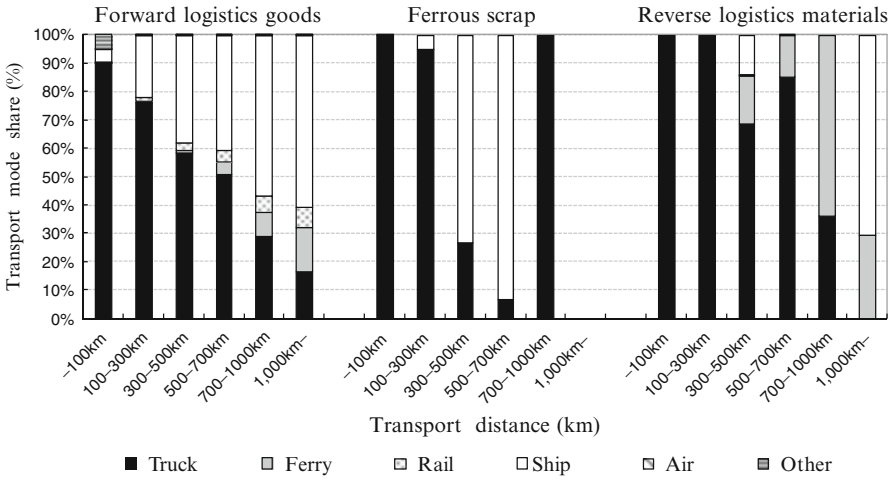

Fig. 3 Transport modal choice in Japanese reverse logistics transport by transporting distance and contents transported (Source: Elaborated 3-day survey of 7th National Cargo Net Flow Report (logistics census) for year 2000 (Ministry of Land, Infrastructure and Transport of Japan 2002)

(2) there is a lack of a principal coordinator for structuring reverse chains; and (3) in Japan, firms involved in waste collection, transportation, and disposal operate at a smaller scale.

About 50% of the top 300 waste-related firms in Japan have less than 50 employees (Yano Research Institute Ltd. 2005). Also, according to the National Federation of Industrial Waste Management Associations (2000) in Japan, industrial waste collection and conveyance contractor firms with less than 20 employees comprise approximately 40% of the total; among these firms, collection and transportation firms with fewer than five employees comprise approximately 40%. Regarding storage activities to transport waste efficiently, 77% of industrial waste collection and conveyance contractors do not have such storage because of their smaller scale, as mentioned above.

The Market for Reverse Logistics in Japan

As previously mentioned, reverse logistics lacks a so-called chain strategy that involves efficiency improvements in transport and integration of the supply chain. In addition, higher transportation costs for lower-value reverse logistics materials, as compared to that of forward logistics goods, calls for a significant improvement in the efficiency of reverse logistics. For the reasons mentioned above, recycled products have a decreased ability to maintain price competitiveness in the market.

To improve the supply chain management for reverse logistics, the concept of Third-Party Service Provider (3PSP) has been increasingly adopted in the

United States. Collaborating with metal dealers, 3PSPs integrate reverse chains and provide better services related to them. One representative example is the Waste Management Company (WMC), which operates throughout the United States with 413 waste-collecting vehicle stations, 33 railway-based transfer stations, and 131 recycling depots. WMC reaps benefits from economies of scale through integration and management of reverse chains. Its recent annual sales top 1.2 trillion yen with a work force of over 44,000 (Waste Management Inc. 2011). It has improved efficiency and reduced costs through effectual structuring of reverse logistics and integration of reverse chains. The WMC has accomplished this through the provision of rail-based waste transfer stations, collection of waste at transfer stations, and transportation by railway to reduce transport cost. For example, at the rail-based waste transfer station at Annapolis, Maryland, waste is first collected at the transfer station and then the railway is utilized to transport the bulk volume of waste to the final disposal site located 145 km away. As one railway trip transports the equivalent load of 100 trucks, significant efficiency gains accompany decreased transport cost. Moreover, in the US, in addition to private companies, local governments are also involved in improving the efficiency of reverse logistics. For example, in Montgomery County, Maryland, waste is first collected at the transfer station by trucks, then transported 30 km to the incineration plant by railway, in order to achieve transport efficiency.

There is a sizable domestic market for reverse logistics in Japan. In 1997, the reverse logistics market accounted for 2.5 trillion yen and is expected to expand further in the future (Hosoda 2000). Moreover, there is an increasing trend of moving of waste and recycling materials over longer distances.

Considering these long-term trends in the market, it appears to be even more imperative to improve the efficiency of reverse logistics in Japan. By appropriately structuring reverse logistics chains, better service provision to waste dispatchers can be achieved. In concert with cooperation and coordination among the country's various regions and sectors, these combined efforts can help Japan to reap the benefits of economies of scale in reverse logistics.

Previous Studies Modeling Efficiency Improvement in Reverse Chains

As previously mentioned, the cost of logistics can often be the crucial factor in determining whether a material stream is considered a waste or a resource. As such, improvement of transport efficiency in reverse logistics has received some attention in recent years. As vehicle operating cost is a significant component in calculating transportation cost, the Vehicle Routing Problem (VRP) has been an important model, used extensively to solve reverse logistics problems.

Kim et al. (2009) presented a vehicle routing approach to minimize the transport distance of end-of-life consumer electronic goods collected by local authorities and

major manufacturers for routing to four regional recycling centers. They applied a tabu search heuristic to solve the vehicle routing problem. Mar-Ortiz et al. (2011) also optimized the design of reverse logistics networks for the collection of electric and electronic equipment waste. In this research, the associated facility location and collection routing problems were mathematically formulated as integer programming models.

Buhrkal et al. (2012) considered the waste collection vehicle routing problem with time windows to find cost-optimal routes for garbage trucks. They presented a mathematical modeling formulation including lunch and rest breaks, which has been solved using the meta-heuristic adaptive large neighborhood search. Bonomo et al. (2012) proposed a method to optimize waste collection using graph theory and mathematical programming tools. Kanchanabhan et al. (2011) proposed a model to plan the allocation of waste bins and a route network using GIS.

Sasikumar et al. (2010) explored the optimal routes which minimize total distance traveled and corresponding transportation costs for a third-party reverse logistics provider. They proposed a hybrid approach of combining the Sweep Algorithm and Clarke-Wright Savings Algorithm with the simulated annealing algorithm. Anbuudayasankar et al. (2010) developed three unified heuristics—based upon an extended branch-and-bound heuristic, a genetic algorithm and simulated annealing—to solve a simultaneous delivery and collection problem with constrained capacity.

A Model to Evaluate Policy Measures

In this section, a proposal is presented to improve the efficiency of reverse logistics in Japan based upon reverse chains. In particular, bulk transport via milk-run type cooperative transport and modal shift options are examined in terms of their effectiveness. For modal shift, as discussed in the US examples above, transport by railway is examined.

The model is applied to the case of the 23 wards of Tokyo Prefecture, the core of the Tokyo metropolitan area with a population of approximately eight million and an area of 606 km^2. There are a large number of waste dispatchers and waste processing facilities in the 23 wards of Tokyo. Therefore, some simplification is required in operationalizing the concepts underlying this research in order to achieve clarity. For example, we assume that there is only one waste dispatcher in each ward and only one waste processing facility which serves the 23-ward metropolitan area. In the case of transporting waste to an exurban area, it is assumed that only one waste transfer station exists in Tokyo Prefecture, for the purpose of later transporting waste by rail to a processing facility in Yamaguchi Prefecture.

Also, we consider only plastic waste for this case study because plastic is one of the most widely-used products in Japan, and plastic waste may be routed either to the disposal stream or the reprocessing stream depending upon the transportation cost.

The next section presents a basic description of the study area and the cases analyzed. This is followed by an explanation of the optimization model developed for application in this case study.

Description of the Test Case for Model Application

In the process of this research into the case of plastic waste, different schemes of collecting and transporting waste to the processing site were compared in terms of total cost of collection, as well as transportation and environmental impacts. The comparison was between the direct and independent type of vehicle routing and the milk-run type cooperative transport that is used for transporting the waste from collection nodes to waste processing facilities.

In this research, cooperative transport indicated waste dispatchers working cooperatively to move the waste of the prefecture to reduce their costs, including use of intermodal transport. Intermodal transport in this research included the usage of truck and rail. In the case of cooperative transport, waste was collected at the Tokyo Prefecture transfer station, from where the bulk load of waste was transported to the processing facility in Yamaguchi Prefecture by railway.

Also, two different methods of plastic waste processing were considered, namely incineration and thermal recycling. For the processing facility in Tokyo Prefecture, incineration was assumed and the cost figure (33,000 yen/ton) was obtained from a research report by Yano Research Institute (2005). Likewise, in the case of the processing facility in Yamaguchi Prefecture (1062 km from central Tokyo), waste was recycled via thermal recycling and the cost figure (12,000 yen/ton) was obtained by interview with the waste processing company. The loading capacity of trucks was assumed to be 6 tons, considering road conditions in Tokyo Prefecture. It was assumed that a truck did not use toll roads while transporting waste.

As mentioned above, some simplification was necessary in order to clearly show the contrast between different collection schemes. We supposed that there was only one waste dispatcher in each ward and only one waste processing facility serving the 23 wards of Tokyo Prefecture. The amount of waste from each waste dispatcher was assumed to be 1.5 tons, which was the average waste loading amount per truck, except in the case of ferrous scrap (Ministry of Land, Infrastructure and Transport of Japan 2002). The collection time at each collection node (or dispatcher) was 30 min (t_j); operating hours of trucks were fixed at 8 h (T_e). Travel times (t_{ij}^v) were calculated using the traffic assignment method. The information on basic units of vehicle operating cost according to travel speed (β[yen/km]) and money value of time ($\alpha = 87.44$ yen) were obtained from a cost-benefit analysis manual (Ministry of Land, Infrastructure and Transport 2003). Transport costs (d_{ij}) from a dispatcher to another dispatcher or to the transfer station were calculated using travel time (t_{ij}^v), travel distance (L_{ij}) and parameters such as basic unit of vehicle operating cost (β) and money value of time (α).

Table 1 Alternatives for waste collection, transport and processing

	Transport method	Processing site	Processing method
Case 1	Direct and independent truck trips	Tokyo prefecture	Incineration
Case 2	Direct and independent truck trips	Yamaguchi prefecture	Thermal recycling
Case 3	Cooperative truck transport & rail	Yamaguchi prefecture	Thermal recycling

The current waste transport pattern consists of transporting wastes from each dispatcher to the processing facility (located in Tokyo Prefecture) using direct and independent truck trips, with each trip comprised of a run directly from a dispatcher to the processing facility. For this model, the current waste transport pattern was assumed, in consideration of the present status of the very small-scale firms involved in waste collection and transportation. This assumption is fully characterized in Case 1—and similarly characterized in Case 2—in terms of longer-distance direct and independent transport to the Yamaguchi Prefecture processing facility. It was assumed that each truck takes the shortest route from a dispatcher to the processing facility (Case 1 and Case 2), as identified by Dijkstra's algorithm.

On the other hand, in the case of milk-run type cooperative transport (Case 3), the model that is presented below is applied to find the optimized cooperative truck transport route. Also, it is assumed that each truck uses the shortest route from a dispatcher to another dispatcher or the transfer station—again, as identified by Dijkstra's algorithm.

As shown in Table 1, three alternatives were considered for managing the waste of Tokyo Prefecture. Case 1, expressing the current waste transport pattern, involved transporting waste from dispatchers to the Tokyo Prefecture processing facility using direct and independent truck trips (each trip ran directly from a dispatcher to the processing facility). Case 2 also used a similar transporting method (direct and independent), but the processing facility was located outside the metropolitan area (Yamaguchi Prefecture, 1062 km from central Tokyo), therefore each trip ran directly from a dispatcher to the Yamaguchi Prefecture processing facility by truck. Finally, in Case 3, a milk-run type cooperative truck transport system was used to collect and transport the waste to the Tokyo Prefecture transfer station, from where the bulk load of waste was transported to the processing facility in Yamaguchi Prefecture by railway.

Model Structure

The model is formulated as an optimization problem with an objective to minimize the total cost of cooperative transport, as shown in Eq. (1). This model was developed under the assumption of a fixed demand without fluctuation. The complete model structure for Case 3 is as given below.

Minimize

$$TC = \sum_{i=0}^{n} \sum_{j=0}^{n} \sum_{v=1}^{m} \left(d_{ij} + c_j\right) x_{ij}^{v} \tag{1}$$

Subject to

$$\sum_{j=0}^{n} \sum_{v=1}^{m} x_{ij}^{v} = 1 \quad \forall_i \in N \tag{2}$$

$$\sum_{j=0}^{n} x_{ij}^{v} - \sum_{j=0}^{n} x_{ji}^{v} = 0 \quad \forall_i \in N, v \in V \tag{3}$$

$$\sum_{j=1}^{n} \sum_{v=1}^{m} x_{0\,j}^{v} \leq V \tag{4}$$

$$\sum_{i=0}^{n} \sum_{j=0}^{n} \left(t_{ij}^{v} + t_j\right) x_{ij}^{v} \leq T_e \quad v \in V \tag{5}$$

$$\sum_{i=0}^{n} \sum_{j=0}^{n} q_j x_{ij}^{v} \leq Q_v \quad v \in V \tag{6}$$

$$x_{ij}^{v} \in \{0, 1\} \, \forall_{(i,j)} \in A, \quad v \in V \tag{7}$$

$$N = \{0, 1, 2, \cdots, n\} \tag{8}$$

$$V = \{1, 2, 3, \cdots, m\} \tag{9}$$

$$A = \{(i, j) \, | i, j \in N\} \tag{10}$$

Where

TC: total cost; d_{ij}: transport cost from node i to j;
c_j: waste collection cost at node$j$; x_{ij}^{v}: if truck v uses (i,j) path: 1, if not: 0;
t_{ij}^{v}: transport time from node i to j; t_j: time required for waste collection at node j;
T_e: maximum time of operation by vehicle v; q_j: amount of waste at node j;
Q_v: loading capacity of truck.

Here, Eq. (1) is the objective function, in which total cost (TC) includes waste collection cost (c_j) and waste transport cost (d_{ij}). The transport cost also includes both vehicle operating cost (r_{ij}) and travel time cost (w_{ij}). The waste collection cost, however, depends upon the hours of operation for the waste collection operation and amount of collected waste. Time is converted into monetary value. Also, transport cost (d_{ij}) and waste collection cost (c_j) are as given below.

$$d_{ij} = r_{ij} + w_{ij} \tag{11}$$

$$r_{ij} = L_{ij} \times \beta \tag{12}$$

$$w_{ij} = t_{ij}^{v} \times \alpha \tag{13}$$

$$c_j = t_j \times \alpha \tag{14}$$

Where

r_{ij}: vehicle operating cost; w_{ij}: vehicle travel time cost;
L_{ij}: travel distance from node i to j; α: money value of time;
β: basic unit of vehicle operating cost according to travel speed.

The remaining equations represent the model's various constraints. Eq. (2) requires that only one truck visits one collection node once. Eq. (3) requires that the number of arriving trucks is equal to the number of departing trucks at the collection node. Eq. (4) ensures that the number of trucks departing from a depot is less than the total number of trucks available. Eq. (5) specifies truck return within operating hours, and Eq. (6) ensures that the volume of waste loaded into trucks is less than the vehicle's capacity. Equation (7) is the binary constraint. Equation (8) is the set of nodes, and Eq. (9) is the set of vehicles. Finally, Eq. (10) is the set of arcs.

A genetic algorithm has been successfully used to solve optimization problems of the vehicle routing model (Anbuudayasankar et al. 2010; Jula et al. 2005; Baker and Ayechew 2003). In this research, in view of the efficacy of the genetic algorithm, the genetic algorithm is utilized a for the purpose of optimization (minimization of total cost) to obtain the optimal result for the milk-run type cooperative transport case (Case 3).

In the application of the partially matched crossover genetic operator, we apply elitist roulette selection and scramble mutation. The number of individuals and the number of generations are set to 200 and 5000, and the maximum number of generations is a termination criterion. The crossover and mutation rates were set to 0.6 and 0.03 respectively. These rates were chosen because of their successful implementation for solving optimization problems in previous research by Sakawa and Tanaka (1996) and Iba (2005).

For comparative evaluation of each case in terms of environment, CO_2 emissions are calculated using CO_2 Basic Emission Units per km by Japan Federation of Freight Industries (2007).

Results

Computing the total cost for each alternative case can be performed directly for Case 1 and Case 2, as they use direct and independent transport and do not involve optimization modeling. However, Case 3 uses cooperative transport and requires utilization of the model to compute the total cost.

The layout of waste collection nodes (dispatchers) and the waste processing facility in the case of direct and independent transport (for case 1) is shown in Fig. 4.

As previously mentioned, Case 2 also uses a similar direct and independent transport from waste collection dispatchers, but the processing facility is located outside the metropolitan area (Yamaguchi Prefecture, 1062 km from central Tokyo), therefore each trip runs directly from a dispatcher to the Yamaguchi Prefecture processing facility by truck. The location of the Yamaguchi Prefecture processing facility is shown in Fig. 5.

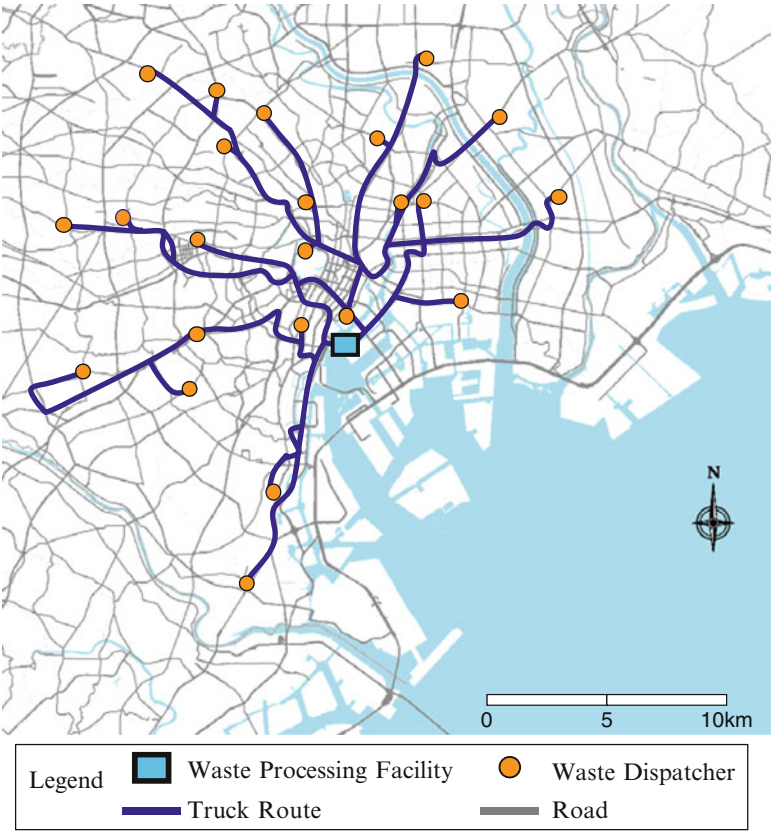

Fig. 4 Case 1 direct and independent routes for waste transportation (Tokyo Prefecture)

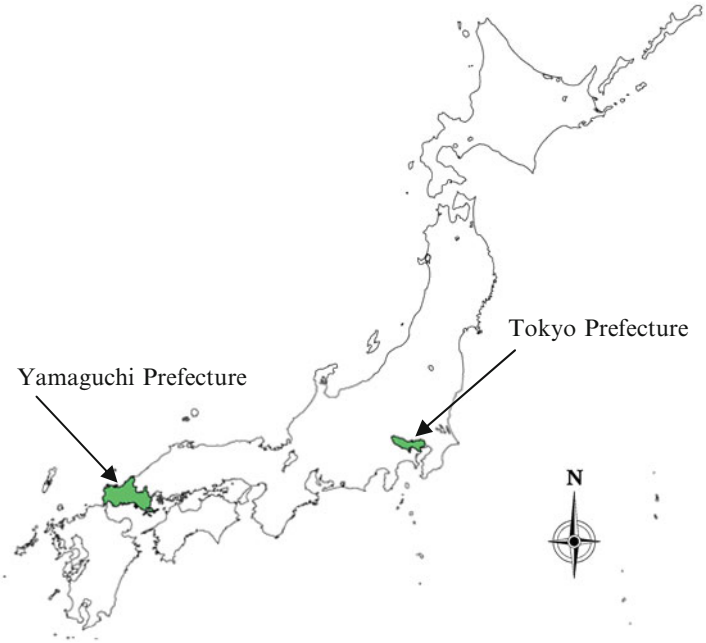

Fig. 5 The location of Yamaguchi Prefecture processing facility (Case 2 and Case 3)

The optimized cooperative truck transport route for Case 3 is shown in Fig. 6. Waste collected to the Tokyo Prefecture transfer station, from where the bulk load of waste is transported to the processing facility in Yamaguchi Prefecture by railway.

Total cost computation for each case is displayed in Fig. 7.

As expected, for Case 1, the transport cost is low since total transport distance is short compared with the other cases, however, the processing cost is large since the processing facility is located within the Tokyo metropolitan area where land price is very high.

By relocating the processing facility outside the metropolitan area (in Yamaguchi Prefecture), the processing cost can be significantly reduced—but very high transport cost is incurred—if the direct and independent truck transport is used, resulting in increased total cost (Case 2).

On the other hand, if cooperative truck transport is used to collect the waste at the Tokyo Prefecture transfer station and then railway is used for transporting the bulk load of waste to the processing facility outside of the Tokyo metropolitan area (in Yamaguchi prefecture), the transport cost is significantly reduced (Case 3) compared with Case 2. Case 3 results in the minimum total cost, thus demonstrating the most efficient system of waste management.

In addition, for each alternative case of waste management assumed for Tokyo Prefecture, the environmental cost in terms of CO_2 emissions is computed. The results are presented in Table 2. If CO_2 emissions for Case 1 stand as a reference,

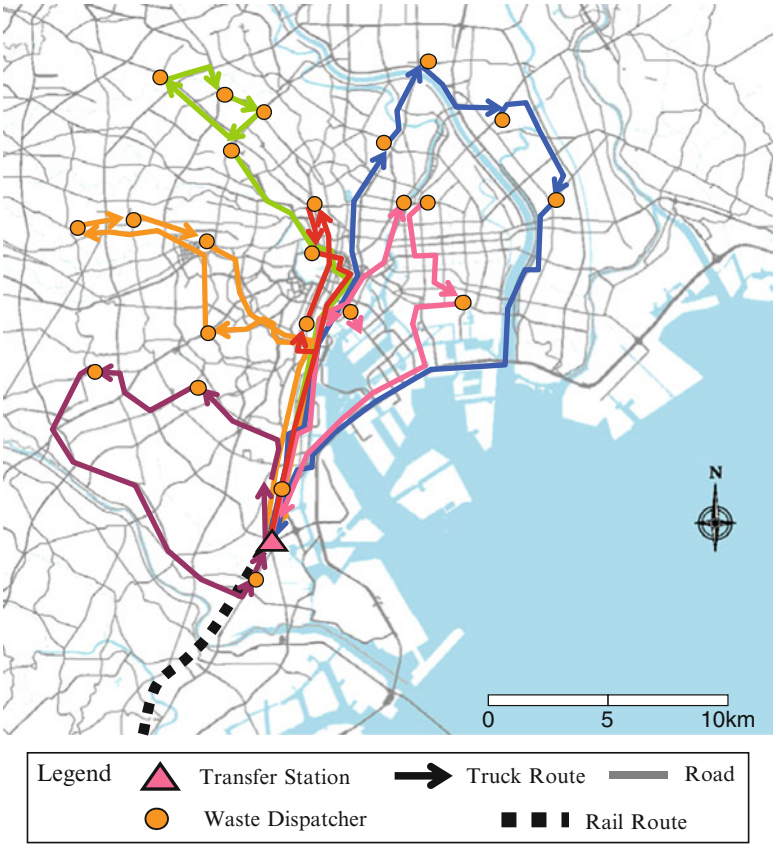

Fig. 6 Cooperative transport routes for waste transportation (Tokyo Prefecture)

Case 2 increases CO_2 emissions by 116%—apparently due to longer hauling distance by an inefficient mode of transport. Here also Case 3 shows a significant advantage by reducing CO_2 emissions by 22%.

The above results also have important implications for recycling policies. The current policy approach places heavy emphasis on regulation and awareness to promote recycling of waste for resource conservation and environmental management.

However, because of the relatively higher cost for waste recycling, due in particular to transport cost, only a small portion of waste is recycled. If greater emphasis is placed on building efficient reverse logistics chains, the total cost of waste recycling can be significantly reduced, as demonstrated by this analysis.

Such reduction in the cost of waste recycling would significantly expand the market for the recycling business, as shown in Fig. 8. If the cost for waste recycling can be reduced from C_0 to C_1, the supply curve of recycled material shifts down from S_0 to S_1. As a result, the equilibrium quantity of recycled material increases.

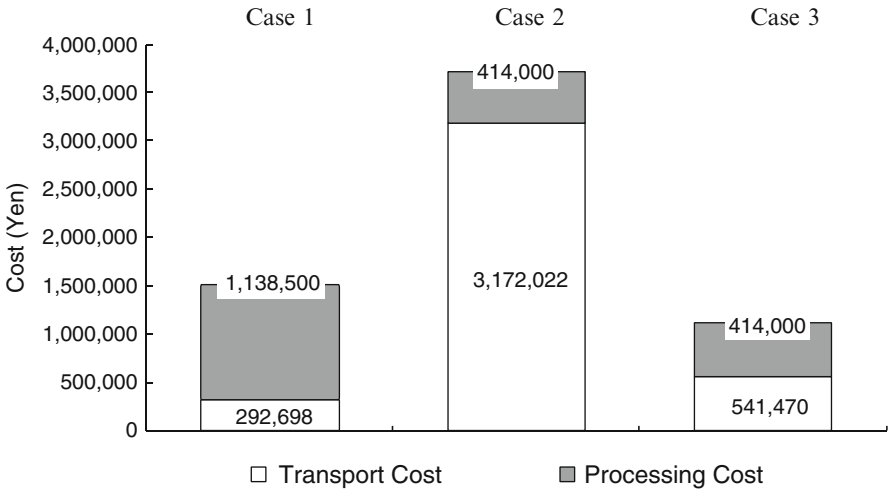

Fig. 7 Comparison of total cost

Table 2 Comparison of environmental impact

	Case 1	Case 2	Case 3	
	Truck	Truck	Truck	Rail
Transport distance (km)	564	28,083	294	1062
Amount of waste (t)	34.5	34.5	34.5	
CO_2 (g/ton·km)	153	153	153	21
Total CO_2 emission (g)	2,978,130	6,445,049	2,321,205	

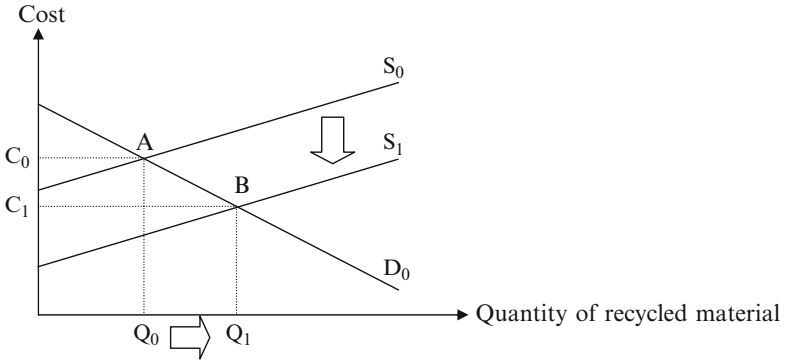

Fig. 8 Effect of transport cost on recycling market

Conclusions

Unlike forward logistics, reverse logistics have not yet fully evolved to provide efficient services to associated sectors. The key bottleneck in improving reverse logistics is the transport component. Therefore improvement of transport efficiency in reverse logistics contributes not only to direct efficiency gains but also to the promotion of waste recycling.

This paper first examined the existing situation of waste management and related reverse logistics systems in Japan. In particular, the inefficiency of reverse logistics is highlighted as a barrier to promoting recycling and efficient waste disposal. Some examples of efficient reverse logistics systems in the US were also discussed. In a case study of the Tokyo Prefecture area, alternative transport systems for waste management reverse logistics were evaluated via an optimization model utilizing a genetic algorithm. Under this model's particular assumptions, the modeling results showed that milk-run type cooperative truck transport to a Tokyo Prefecture transfer station, in combination with bulk-hauling railway transport to a waste processing facility in Yamaguchi Prefecture, is the most cost-effective of those studied. This option (Case 3) demonstrates significant benefits in terms of lower CO_2 emissions. It is recommended that efficient policy measures such as intermodal transport in reverse logistics should be further studied to determine the potential for reducing the cost of waste recycling and promoting and expanding the market for recycling materials.

In many cases, the construction of new waste facilities depletes open land, and often faces city land-use restrictions and opposition from residents. For these reasons, new waste facilities are typically constructed in the urban periphery or in exurban areas, which necessitates long-distance transportation. Moreover, the existence of economies of scale in this field requires long-distance transportation; there is clearly an increasing trend of moving waste and recycling materials over longer distances. Hence, it is necessary to improve transport efficiency in reverse logistics by structuring reverse chains to overcome the smaller scale of waste-related firms in Japan, in order to make it possible to reap the benefits of economies of scale.

References

Alker, S., & Ravetz, J. (2006). *Spatial planning for integrated waste management*. Glasgow: Enviro Centre Ltd.

Anbuudayasankar, S. P., Ganesh, K., Lenny Koh, S. C., & Mohandas, K. (2010). Unified heuristics to solve routing problem of reverse logistics in sustainable supply chain. *International Journal of Systems Science, 41*, 337–351.

Baker, B. M., & Ayechew, M. A. (2003). A genetic algorithm for the vehicle routing problem. *Computers & Operations Research, 30*, 787–800.

Bonomo, B., Duran, G., Larumbe, F., & Marenco, J. (2012). A method for optimizing waste collection using mathematical programming: A Buenos Aires case study. *Waste Management & Research, 30*, 311–324.

Buhrkal, K., Larsen, A., & Ropke, S. (2012). The waste collection vehicle routing problem with time windows in a city logistics context. *Procedia – Social and Behavioral Sciences, 39*, 241–254.

Dowlatshahi, S. (2000). Developing a theory of reverse logistics. *Interfaces, 30*(3), 143–155.

Fleischmann, M. (2001). *Quantitative models for reverse logistics*. Berlin/Heidelberg: Springer.

Hosoda, E. (2000). Economic aspects of recycling of construction waste. *Waste Management Research, 11*(2), 105–116.

Iba, H. (2005). *Foundations of genetic algorithms*. Tokyo: Ohmsha.

Inc, K. (2004). *Data book on freight rates in Japan 2004*. Tokyo: Koutsuunihon Inc.

International Monetary Fund. (2013). http://www.imf.org/external/data.htm. Accessed 5 Mar 2013.

Japan Federation of Freight Industries. (2007). *Suji de Miru Butsuryu* (Logistics in Figures). Japan Federation of Freight Industries, Tokyo.

Japan Maritime Center. (2000). *Report on cost analysis of Japanese domestic maritime shipping*. Tokyo: Japan Maritime Center.

Japanese Ministry of the Environment, Government of Japan. (2010). *2010 Annual report on the environment, the sound material-cycle society and the biodiversity in Japan*. http://www.env.go.jp/en/wpaper/2010/fulltext.pdf. Accessed 12 Aug 2011.

Johnson, P. F. (1998). Managing value in reverse logistics systems. *Transportation Research Part E, 34*(3), 217–227.

Jula, H., Dessouky, M., Ioannou, P., & Chassiakos, A. (2005). Container movement by trucks in metropolitan network: Modeling and optimization. *Transportation Research Part E, 41*, 235–259.

Kanchanabhan, T. E., Abbas Mohaideen, J., Srinivasan, S., & Lenin Kalyana Sundaram, V. (2011). Optimum municipal solid waste collection using geographical information system (GIS) and vehicle tracking for Pallavapuram municipality. *Waste Management & Research, 29*, 323–339.

Kim, H., Yang, J., & Lee, K. (2009). Vehicle routing in reverse logistics for recycling end-of-life consumer electronic goods in South Korea. *Transportation Research Part D, 14*, 291–299.

Kitamura, T., & Okamoto, K. (2012). Automated route planning for milk-run transport logistics using model checking. In *Proceeding of 2012 third international conference on networking and Computing*, pp. 240–246.

Kopicki, R., Berg, M. J., Legg, L., Dasappa, V., & Maggioni, C. (1993). *Reuse and recycling: Reverse logistics opportunities*. Oak Brook: Council of Logistics Management.

Krumwiede, D. W., & Sheu, C. (2002). A model for reverse logistics entry by third-party providers. *Omega, 30*, 325–333.

Lee, J., McShane, H., & Kozlowski, W. (2002) Critical issues in establishing a viable supply chain/reverse logistic management program. In *Proceedings of IEEE International Symposium on Electronics and the Environment*, pp. 150–156.

Mar-Ortiz, J., Adenso-Diaz, B., & Gonzalez-Velarde, J. L. (2011). Design of a recovery network for WEEE collection: The case of Galicia, Spain. *Journal of the Operational Research Society, 62*, 1471–1484.

Ministry of Land, Infrastructure and Transport of Japan. (2002). *The 7th national cargo net flow report for year 2000*. http://www.mlit.go.jp/seisakutokatsu/census/houkoku/rep_all.pdf Accessed 11 Mar 2013.

Ministry of Land, Infrastructure and Transport of Japan. (2003). *A cost-benefit analysis manual*. http://www.mlit.go.jp/road/ir/ir-council/hyouka-syuhou/1pdf/s1-1.pdf Accessed 13 Mar 2013.

National Federation of Industrial Waste Management Associations. (2000). *Report of factual investigation on industrial waste collectors and transporters*. Tokyo: National Federation of Industrial Waste Management Associations.

Prahinski, C., & Kocabasoglu, C. (2006). Empirical research opportunities in reverse supply chains. *Omega, 34*, 519–532.

Rogers, D. S., & Tibben-Lembke, R. (1999). *Going backwards: Reverse logistics trends and practices*. Pittsburg: Reverse Logistics Executive Council.

Rogers, D. S., & Tibben-Lembke, R. (2001). An examination of reverse logistics practices. *Journal of Business Logistics, 22*(2), 129–148.

Sakawa, M., & Tanaka, M. (1996). *Genetic algorithms*. Tokyo: Asakura.

Sasikumar, P., Noorul Haq, A., & Baskar, P. (2010) A hybrid algorithm for the vehicle routing problem to third party reverse logistics provider. In *Proceeding of 2010 8th international conference on supply chain management and information systems* (SCMIS 2010).

Simchi-Levi, D., Kaminsky, P., & Simchi-Levi, E. (2000). *Designing and managing the supply chain-concepts, strategies, and case studies*. Boston: McGraw-Hill.

Spicer, A. J., & Johnson, M. R. (2004). Third-party demanufacturing as a solution for extended producer responsibility. *Journal of Cleaner Production, 12*, 37–45.

Stock, J. R. (1998). *Development and implementation of reverse logistics programs*. Lombard: Council of Logistics Management.

Waste Management Inc. (2011) *2011 annual report*. http://investors.wm.com/phoenix.zhtml?c=119743&p=irol-reportsannual. Accessed 15 Mar 2013.

Yano Research Institute Ltd. (2005). *The market prospects of the waste disposal treatment business and recycling business strategy*. Tokyo: Yano Research Institute Ltd..

Yoon, J. (2007a). Analyzing current reverse logistics and system-wise approach towards efficiency improvement. *Journal of the Japan Society of Civil Engineers G, 63*(4), 332–344.

Yoon, J. (2007b). A study on reverse logistics for establishing a sound material-cycle society. *Transport Policy Studies Review, 10*(1), 67–73.

Index

A
Abatement, 224, 326
Acceptability, 348
Accessibility, 60, 65, 107, 108, 153, 180, 277, 284
Adaptive resonance theory (ART), 304
Advertisements, 67, 69
Africa, 66, 68, 73, 74, 87–93
Agglomeration economies, 12, 26, 29
Algorithm, 11, 38, 43, 46, 71, 153, 184, 272, 278, 301–319, 331, 349, 351, 370, 372, 374, 379
Analytic hierarchy process (AHP), 225
Amenities, 98, 100, 105, 106
Annual discharge, 226
Ant colony optimization, 303–305, 311
Areal interpolation, 10, 12, 13
Autocorrelation, 2, 14, 18, 20, 24, 98, 100, 103, 104, 108, 112, 113

B
Bandwidth, 48, 102, 104
Battery-switching stations, 346, 352–357
Bayesian interpolation, 9–29
Bifurcations, 196
Biofuels, 164, 324
Biogas, 259
Bioreactor, 230, 231
Bounded rationality, 358
Branding, 64–66, 83, 85, 86
Bulkhauling, 379
Business district, 34, 35, 55, 107

C
Caribbean, 68, 73, 82–83, 89, 91, 92
Cellphone, 49
Cellular automata, 277
Central places, 26, 266
Chow-Lin method, 13
City identify, 65
City images, 63–93
Climate change, 327, 329, 332, 335, 352
Clusters, 36, 46, 47, 51–54, 104, 114, 136, 273, 275, 279, 319
Coal-fired power plants, 324
CO_2 capture, 323–336
Cohesion policy, 123–139
Coin diffusion, 193–218
Coin flows, 194
Cokriging, 99
Coldspots, 3, 113, 136
Combinatorics, 303
Common agricultural policy (CAP), 174, 185
Communication, 59, 64, 66–71, 85, 86, 99, 115, 131, 138, 248, 304
Commuting, 11, 20, 29, 34, 59, 60, 158
Commuting flows, 11
Competitive advantage, 64
Competitiveness, 85, 124, 126, 134, 135, 138, 139, 368
Components, 5, 36, 64, 98, 103, 109, 129, 152, 153, 159, 176, 197, 215, 327, 329, 332
Computational, 5, 103, 109, 207, 213, 269, 286, 303, 305, 308, 314, 326, 327
Conditional autoregressive regression (CAR), 14, 15

J.-C. Thill (ed.), *Spatial Analysis and Location Modeling in Urban and Regional Systems*, Advances in Geographic Information Science,
https://doi.org/10.1007/978-3-642-37896-6

Confirmatory spatial data analysis (CSDA), 4
Conservation, 153, 178, 224, 227, 232, 237, 250, 251, 258, 259, 364, 377
Contagious diffusion, 196
Contiguity, 101, 124, 131, 138
Convergence, 4, 124–128, 135, 136, 138, 139, 195–197, 202, 206–209, 211, 212, 214, 215, 302, 310, 318, 319
Cost-benefit analysis, 165, 371
CO_2 storage, 323–336
Crossscalar, 5
Crossvalidation, 47, 50, 111, 184

D
Data cube, 5
Decomposition, 271, 274
Demographic data, 159
Denmark, 12, 88, 344–352, 357
Dimensional extraction, 69
Dimension reduction, 65
Diseconomies, 12
Distance decay, 102, 199

E
Economic data, 13
Elasticities, 348
Electric vehicles, 335, 341–358
Emission, 225–227, 232, 235, 237, 238, 248, 259, 346, 352, 374, 378
Employment, 10, 23, 26, 176, 276
Energy infrastructure, 323–336
Entropy, 196, 199
Environmental, 98, 100, 107, 108, 115, 146, 149, 150, 152, 154, 159, 164, 174, 175, 223–260, 342, 343, 348, 363, 365, 371, 376–378
Equilibrium, 196, 202, 206–210, 213–215, 218, 233, 348, 377
Equity, 3, 124–126, 134, 136, 138, 139
Euclidean distance, 108, 286
Europe, 4, 11, 66, 68, 73, 87–93, 123, 124, 126, 132, 134–138, 152, 164, 174, 175, 216, 325, 343, 352
European Regional Development Fund (ERDF), 125, 131
European regional policy, 125–128, 138
European Union, 4, 12, 123–139, 174, 178, 188, 344, 346
Evolutionary algorithms, 303
Exchanges, 85, 130, 194–198, 204, 205, 213, 216
Experimentation, 314
Expert scoring, 225
Exploratory spatial data analysis (ESDA), 4
Ex-post evaluation, 146, 163
Externalities, 26, 346, 351, 352, 356
Extrapolation techniques, 115, 151, 278

F
Forestation, 224, 227, 232, 244, 250, 251
France, 87, 90, 91, 194, 200, 201, 209, 212–215
Functional areas, 47
Functional intensity, 47–48, 54–56
Functional relationships, 4
Functional zones, 33–61

G
Generalization, 4
Generalized cross-entropy (GCE), 13
GeoCells, 124, 127–132, 134, 138
Geocode, 105
Geographically weighted regression, 99–102, 116
Geographic information science, 2, 3
Geographic information systems (GIS), 1–5, 38, 59, 99, 101, 105, 107, 108, 117, 119, 152, 176, 180, 182, 188, 269, 271, 284, 370
Geography, 1–3, 73, 194, 196, 198, 206, 216, 269, 271, 272, 274
Georeferenced, 97
Geotagged, 35
Geotweets, 35
Globalization, 68, 69, 138, 149
GLS estimators, 13
Google Search, 68
Granularity, 5, 195
Gravitation, 195
Gravity model, 196
Greenhouse gas (GHG), 259, 351
G statistic, 108

H
Hedonic, 4, 97–120
Heuristics, 303, 304, 318, 347, 370
High performance computing, 303
Hotspots, 3, 113, 136
Housing rent, 98
Human movements, 33–61

I
Incineration, 369, 371, 372
Incinerators, 365
Income disparities, 12
Individual behavior, 195, 197, 198, 215, 216
Industrial location, 3, 266
Infrastructural refueling, 342
Infrastructure, 5, 36, 66, 106, 126, 138, 146, 147, 152, 158, 165, 188, 319, 323–336, 341–358, 367, 368, 371
Inpout-output model, 139
Input-output table, 226
Integer programming, 267, 271, 333–335
Interlinkages, 132
Intermodal, 364, 365, 371, 379
Internal heterogeneity, 10
Investment, 64, 70, 124, 126, 128, 139, 225, 226, 231, 232, 238–241, 243, 255, 257, 333, 342, 343, 351, 352, 356

J
Journey-to-work, 11
Just-in-time, 366

K
Kernel density, 47
Kohonen neural network, 304

L
Labor catchment, 11
Lagrangean relaxation, 267, 274
Landfills, 364, 365
Land use scanner, 151–155, 159, 161, 166
Land use simulation model, 148, 150
Latent dirichlet allocation (LDA), 42, 43, 45, 46, 50–52
Latin America, 66
Learning, 3, 304, 310, 358
Lemmatization, 69
Linear programming, 225, 303
Lingo, 225
Lingua franca, 68
Linguistic, 86
Lisbon strategy, 135, 138
Local indicators of spatial association (LISA), 99, 104
Local labor market areas (LLM), 10, 11, 18, 23, 24, 26, 28
Locational analysis, 266–269, 271, 272
Location-allocation, 324
Logistics, 5, 267, 275, 276, 302, 363–379
Logit, 153, 195, 347, 348, 351

M
Machine learning, 3
Macroeconomic, 127, 128, 164, 198
Mapping, 38, 40, 86, 99, 100, 115
Map segmentation, 37–40, 50
Markov property, 14
Measurement, 2, 102, 213, 286, 287, 306
Metadata, 36, 43, 45
Metaheuristic, 370
Metropolitan areas, 26, 105, 147, 370, 375, 376
Middle East, 66, 68, 73, 76–77, 87–92
Migration, 133, 149, 164, 195–198, 213, 252, 259
Mixed-integer programming, 334
Model coupling, 117
Monte Carlo framework, 331, 334
Moran's I, 103, 104, 108, 109, 112–114, 135
Multicollinearity, 100, 108, 109
Multicriteria decision making, 271
Multidimensional scaling, 4, 63–93
Multiobjective optimization, 284
Mutation, 304–306, 311, 313, 319, 374

N
Nearest neighbor ant colony system, 301–319
Neopositivism, 3
Netherlands, 87, 92, 145–166, 174, 201
Network science, 3
Neural network, 303, 304
Newspapers, 67
Nitrogen, 224, 225
Nonpoint emission, 227
Nonstationarity, 98, 104
Nonstationary process, 101
NP-hard, 302
NUTS II regions, 12, 17, 20, 28

O
Oceania, 68, 73, 77–78, 87, 88, 90–92
Opensource, 108, 109
Operations research, 2, 3
Optimality, 279, 307, 318, 351
Optimization, 3, 5, 225, 266, 267, 272, 279, 284, 286, 302–305, 310, 311, 318, 319, 323–336, 342, 343, 349, 371, 372, 374, 375, 379

Ordinary least squares (OLS), 13, 98–101, 103, 104, 109–111, 113–116
Ordination, 71

P
Parsing, 69
Particle swarm optimization, 303
Pathdependency, 197
Periphery, 158, 379
Pheronomes, 304
Pipelines, 324–327, 329, 330, 332–335, 342, 343
PlotGoogleMaps, 109, 115
P-median, 351
Points of interest, 33–61
Policy implications, 377
Principal components analysis (PCA), 109
Privatization, 105, 286, 287, 292, 295, 297
Productivity, 10, 12, 26, 180–182, 184, 189

Q
Quick-charging stations, 341–358

R
Railways, 107, 108, 275, 366, 367, 369–372, 376, 379
Randomization, 113
Real estate, 98, 100, 105, 107, 159
Recapturing, 365
Recharging infrastructure, 341–358
Recycling, 259, 285, 364, 365, 369–372, 377–379
Redevelopment, 164–166
Refilling station, 357
Refurbishing, 365
Regional growth, 124, 128, 132, 139
Regional growth diffusion, 123–139
Regional integration, 128
Regionalization, 11, 12, 149
Regional science, 1, 3, 4, 60, 86, 266, 268, 269, 273, 276, 279
Regions, 3, 10, 34, 46, 55, 68, 123, 132, 147, 174, 197, 225, 254, 342, 369
Regression kriging (RK), 99, 100, 102–103, 116
Relative location, 27, 29, 197
Remanufacturing, 365
Reverse logistics, 363–379
Robust estimator, 20

S
Scale, 13, 14, 61, 69, 159, 175, 178, 181, 207, 231, 232, 271, 324, 326, 328–330, 333, 342, 343, 364, 366, 368, 369, 379
Scoring, 185
Self-organization, 196
Serbia, 87, 105, 107
Sewage sludge, 223–260
Shipping, 366, 367
Similarity, 64, 65, 68, 71, 79, 84, 113, 212
Simulated annealing, 303, 351, 370
Simulation, 4, 123–139, 148–153, 158–164, 166, 167, 195, 197, 202, 206, 212, 213, 224–229, 233–249, 251, 252, 254, 257, 258, 277, 311, 312, 329, 334, 349
Single act, 125, 131, 135
SLEUTH, 277
Social media, 64, 67, 86
Source-network-sink optimization framework, 324, 325
Spain, 9–29, 87, 90, 93, 124, 175, 178
Spatial
  analysis, 1–5, 101, 150, 151, 265–279, 284, 289
  concentration, 136
  data, 3, 4, 38, 98, 99, 108, 119, 175, 177, 179, 187, 271, 284, 286, 287, 289, 291, 293, 295, 297, 298
  disaggregation, 12, 20, 29
  dynamic, 124, 131, 193–218
  error, 98, 100, 109, 110, 288, 289
  interaction, 138, 193–218
  lag, 98, 100, 101, 109
  optimization, 5, 266, 324, 332, 333, 335
  planning, 126, 146–148, 165, 166, 301, 302
Spatially autoregressive regression (SAR), 13
Spatial science, 3
Spatial weights matrix, 14
Spatiotemporal model, 119
Spillovers, 26
Stemming, 69, 328
Subsidies, 70, 185, 225, 227, 237, 238, 240, 242, 244, 245, 249–251, 257
Sustainability, 324
Sustainable development, 227, 252, 259

T
Tabu search, 303, 370
Taxation, 342, 344
Taxicabs, 36, 40, 49, 59
Taxi trajectories, 34, 47, 49, 59
Taxonomy, 45

Telecommunications, 59, 271
Territorial fragmentation, 124
Textual analysis, 71
Tokenization, 69
Topic discovery, 40–46
Topic model, 36, 37, 42, 44–46, 48–50
Tourism, 61, 67
Transmission, 324, 331, 334, 335
Transportation, 2, 5, 26, 34, 45, 49, 52, 60, 98, 174, 196, 231, 242, 248, 254, 255, 273, 275–277, 319, 325, 334, 342, 364, 366, 368–372, 375, 377, 379
Transport cost, 327, 364, 367, 369, 373, 374, 376–378
Transshipment, 276
Traveling salesman problem, 301–319
Trending, 77
Turbidity, 230, 231

U
Uncertainty, 208, 275, 283–298, 324, 330–335, 348, 357, 358
Unemployment, 107, 149
United States, 4, 11, 83, 84, 87–93, 174, 178, 325, 326, 329, 333, 336, 343, 369
Unit of analysis, 10–13, 29
Untreated sewage, 224, 227
Urban computing, 59
Urban core, 105, 106
Urban density, 145–166
Urban development, 146, 149, 151, 152, 154–161, 163–165, 277
Urban extensification, 165, 166
Urban growth boundary, 277
Urbanization, 146–149, 155, 158, 161, 162, 174
Urban planning, 35, 48, 58–61, 146, 163
Urban system, 4, 27, 138, 277
URLs, 68, 69

V
Validation, 50, 314, 316–317

W
Waste recycling, 377, 379
Water pollutants, 224–227, 229, 232–236, 245, 248, 249, 255, 259
Waterways, 188, 227
Webcrawling, 68, 85
Welfare, 3, 13, 26, 149, 166
Wireless technologies, 66
World cities, 4