

Contributions to Statistics

Olga Valenzuela · Fernando Rojas ·
Luis Javier Herrera · Héctor Pomares ·
Ignacio Rojas *Editors*

Theory and Applications of Time Series Analysis

Selected Contributions from ITISE 2022

 Springer

Contributions to Statistics

The series **Contributions to Statistics** features edited and conference volumes in theoretical and applied statistics. Composed of refereed selected contributions, the volumes present the latest developments in all the exciting areas of contemporary statistical research.


Olga Valenzuela · Fernando Rojas ·
Luis Javier Herrera · Héctor Pomares ·
Ignacio Rojas
Editors


Theory and Applications of Time Series Analysis

Selected Contributions from ITISE 2022


 Springer


Editors

Olga Valenzuela 
Faculty of Sciences
University of Granada
Granada, Spain

Luis Javier Herrera 
ETSIIT, CITIC-UGR
University of Granada
Granada, Spain

Ignacio Rojas 
ETSIIT, CITIC-UGR
University of Granada
Granada, Spain

Fernando Rojas 
ETSIIT, CITIC-UGR
University of Granada
Granada, Spain

Héctor Pomares 
ETSIIT, CITIC-UGR
University of Granada
Granada, Spain

ISSN 1431-1968

Contributions to Statistics

ISBN 978-3-031-40208-1

ISBN 978-3-031-40209-8 (eBook)

<https://doi.org/10.1007/978-3-031-40209-8>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Preface

This book gathers the extended versions of a selection of the best contributions presented in the eighth edition of the International Conference on Time Series and Forecasting, ITISE 2022, held in Gran Canaria (Spain) in June 2022.

Fortunately, this conference was characterized by a return to normality, something that had not been achieved at ITISE since 2019. Let's remember that the 2020 edition was cancelled due to the outbreak of the COVID-19 coronavirus and that the 2021 edition was held under very special conditions, with hybrid face-to-face/online sessions. We can now happily remember the old saying "nothings lasts forever", which in its more popular version in Spanish we usually say "no hay mal que dure 100 años" (literally translated: "there is no evil that lasts a hundred years"). Let's cross our fingers and hope that the following editions maintain this new normality.

As is well known, this congress aims to provide a friendly discussion forum for scientists, engineers, educators, and students to discuss the latest ideas and achievements in the fundamentals, theory, models, and applications in the field of time series analysis and forecasting. More specifically, the main topics of the last edition of the Congress were

1. Time series analysis and forecasting

- Nonparametric and functional methods
- Vector processes
- Probabilistic approaches to modeling macroeconomic uncertainties
- Uncertainties in forecasting processes
- Nonstationarity
- Forecasting with Many Models. Model integration
- Forecasting theory and adjustment
- Ensemble forecasting
- Forecasting performance evaluation
- Interval forecasting

- Data preprocessing methods: Data decomposition, seasonal adjustment, singular
- Spectrum analysis, detrending methods, etc.

2. **Econometrics and forecasting**

- Econometric models
- Economic and econometric forecasting
- Real macroeconomic monitoring and forecasting
- Advanced econometric methods

3. **Advanced methods and on-line learning in time series**

- Adaptivity for stochastic models
- On-line machine learning for forecasting
- Aggregation of predictors
- Hierarchical forecasting
- Forecasting with computational intelligence
- Time series analysis with computational intelligence
- Integration of system dynamics and forecasting models

4. **High dimension and complex/big data**

- Local versus global forecasts
- Dimension reduction techniques
- Multiscaling
- Forecasting Complex/Big data

5. **Forecasting in real problems**

- Health forecasting
- Atmospheric science forecasting
- Telecommunication forecasting
- Hydrological forecasting
- Traffic forecasting
- Tourism forecasting
- Marketing forecasting
- Modelling and forecasting in power markets
- Energy forecasting
- Climate forecasting
- Financial forecasting and risk analysis
- Forecasting electricity load and prices
- Forecasting and planning systems

High-quality candidate papers from the Conference ITISE 2022 (15 contributions) were invited to submit an extended version of their conference paper to be considered for this special publication in the book series of Springer: Contributions to Statistics. For the selection procedure, the information/evaluation of the chair of every session, in conjunction with the review comments and the summary of reviews, was taken into account.

So, now we are pleased to have reached the end of the whole process and present the readers with these final contributions that we hope will provide a clear overview of the thematic areas covered by the ITISE 2022 conference. For the sake of readability, the contributions presented in this book have been classified into different chapters according to their content. Some chapters of the book contain pure theoretical contributions. On the other hand, there are chapters with more practical contributions with the intention of providing the readers with a more real-world view of the field. As is common in these editions, a specific chapter of the book has been dedicated to Econometrics, one of the most prominent applications of time series modeling and forecasting. In the following, we will make a short summary of what the reader may find in every chapter of the book:

- *Theoretical Aspects of Time Series*. Although in the field of time series it is difficult to separate the theoretical aspects from the practical ones, since the presentation of many of the theoretical developments usually ends with practical examples where these developments could be applied, the papers in this first part have been selected for being mainly theoretical. In contrast to previous editions, in which this chapter was the one that included the largest number of contributions, in this edition only two papers have been selected for this part. The first one deals with the study of several online parameter estimation methods for irregular autoregressive models. The second one introduces a new class of stochastic processes that the authors call “Costationary Whitenoise Processes”, and propose a theoretical framework to estimate their underlying parameters.
- *Econometrics*. This part aims at presenting some recent developments of time series research applied to analysis and forecasting methods in Econometrics. Four contributions have been selected. The first of these analyzes data from Airbnb in Milan to find out which variables have the greatest influence on what apartments are more likely to be rented. In the second one, the authors test the ability of a semi-Markov chain model to reproduce some typical facts, such as the persistence of volatility, of many financial time series during the COVID-19 pandemic. The third, very original in our view, makes use of machine learning approaches to address the forecasting of stock price direction by analyzing news sentiment. Finally, the last contribution of this part tries to analyze the published recommendations of stockbrokers, their accuracy, and their real impact on portfolio trades that include stocks of companies in the construction sector.

- *Time Series Analysis Applications*. The third part of the book is dedicated to applications of time series analysis different from the ones related to Econometrics. The idea is to state explicitly that applications of time series reach practically any scientific discipline imaginable. Three contributions were selected for this part. The first one presents an application of an automatic entropy-based clustering algorithm for seasonal time series. In the second one, the authors examine the impact of climate variables on the transmissibility of coronavirus and predict the dynamics of confirmed COVID-19 cases using time series models and climate data. This part ends with a paper focused on time series analysis of Interferometric Synthetic Aperture Radar data for monitoring the displacement of the Earth’s surface. The authors also describe the time series analysis tools in the least-squares wavelet software that can be used for processing non-stationary time series which may not be evenly sampled.
- *Time Series Forecasting*. The fourth part of the book is dedicated to some applications of time series prediction. The first contribution explores one of the most widely used models that explicitly deal with mixed-frequency datasets, the MIDAS (MIXed DATA Sampling) models. The authors compare the nowcasting and forecasting performance of different variants of these models when predicting the GDP growth of the four largest Euro Area economies between 2011Q4 and 2020Q3. In the next contribution, a new theoretical and practical scheme for automatic forecasting model selection using R programming is developed, which objectively can quantify which model has the “best” forecasting capacity with dimensionless measures. Finally, in the last paper of this part, recurrent neural networks with different types of gated cells for forecasting time series with multiple seasonalities are compared.
- *Time Series Applications in Energy*. The last part of the book is dedicated to the applications of time series analysis and forecasting methods in one of the areas of greatest impact in recent years: energy. In particular, we have selected contributions focused on power grids; wind energy production prediction; and daily short-term gas, coal, oil, and carbon emission futures.

Last but not least, we would like to point out that this edition of ITISE was organized by the University of Granada (UGR), Spain, together with the Spanish Chapter of the IEEE Computational Intelligence Society. The Guest Editors would also like to express their gratitude to all the people who supported them in the compilation of the book, especially to the contributing authors for their submissions, the chairs of the different sessions, and to the anonymous reviewers for their comments and useful suggestions in order to improve the quality of the papers.

We wish to thank our main sponsors as well: the Department of Computer Engineering, Automation and Robotics of the University of Granada and the Research Center for Information and Communications Technologies (CITIC-UGR) for their

support and grants. Finally, we wish also to thank Dr. Veronika Rosteck, Springer Editor, for their interest in editing a book series of Springer based on the best papers of ITISE 2022.

We hope the readers of this book find these contributions interesting and helpful.

Granada, Spain
May 2023

Olga Valenzuela
Héctor Pomares
Luis Javier Herrera
Fernando Rojas
Ignacio Rojas

Contents

Theoretical Aspects of Time Series

Online Estimation Methods for Irregular Autoregressive Models	3
Felipe Elorrieta, Lucas Osses, Matias Cáceres, Susana Eyheramendy, and Wilfredo Palma	

Costationary Whitenoise Processes and Local Stationarity Testing	19
Alessandro Cardinali	

Econometrics

The Effects of Healthy and Sustainable Transportation, Commerce, and Spillover on Airbnb Performance	31
Jorge Chica-Olmo and Ruggero Sainaghi	

A Semi-Markov Approach to Financial Modelling During the COVID-19 Pandemic	47
Riccardo De Blasis	

From News to Sentiments and Stock Price Directions	59
Fennee Chong and Bharanidharan Shanmugam	

Recommendations of Stockbrokers Versus Fuzzy Portfolio Approach in Construction Sector	71
Anna Łyczkowska-Hanćkowiak and Aleksandra Wójcicka-Wójtowicz	

Time Series Analysis Applications

Automatic Clustering for Seasonal Time Series Based on Entropy	89
Miguel Ángel Ruiz Reina	

Modelling and Predicting the Dynamics of Confirmed COVID-19 Cases Based on Climate Data	105
Yuzhi Cai, Fangzhou Huang, and Jiao Song	

Least-Squares Wavelet Analysis of Rainfalls and Landslide Displacement Time Series Derived by PS-InSAR 117
 Ebrahim Ghaderpour, Claudia Masciulli, Marta Zocchi, Roberta Marini, Giandomenico Mastrantoni, Francesca Reame, Gianmarco Pantozzi, Niccoló Belcecchi, Gabriele Scarascia Mugnozza, and Paolo Mazzanti

Time Series Forecasting

Macroeconomic Forecasting Evaluation of MIDAS Models 135
 Nicolas Bonino-Gayoso and Alfredo Garcia-Hiernaux

Relative Measures of Forecasting: Lambda-Family-Measures 155
 Miguel Ángel Ruiz Reina

Recurrent Neural Networks for Forecasting Time Series with Multiple Seasonality: A Comparative Study 171
 Grzegorz Dudek, Slawek Smyl, and Paweł Pełka

Time Series Applications in Energy

Markov Processes for the Management of a Microgrid 189
 Salvatore Vergine, César Álvarez-Arroyo, Guglielmo D’Amico, Juan Manuel Escaño, and Lázaro Alvarado-Barrios

Deep Learning Applied to Wind Power Forecasting: A Spatio-Temporal Approach 207
 Rubén del Campo, Eloy Anguiano, Álvaro Romero, and José R. Dorronsoro

Forecasting Natural Gas Prices with Spatio-Temporal Copula-Based Time Series Models 221
 Sven Pappert and Antonia Arsova

Theoretical Aspects of Time Series

Online Estimation Methods for Irregular Autoregressive Models



Felipe Elorrieta , Lucas Osses, Matias Cáceres, Susana Eyheramendy , and Wilfredo Palma

Abstract In the last decades, due to the huge technological growth observed, it has become increasingly common that a collection of temporal data rapidly accumulates in vast amounts. This provides an opportunity for extracting valuable information through the estimation of increasingly precise models. But at the same time, it imposes the challenge of continuously updating the models as new data become available. Currently available methods for addressing this problem, the so-called online learning methods, use current parameter estimations and novel data to update the estimators. These approaches avoid using the full raw data and speeding up the computations. In this work, we consider three online learning algorithms for parameters estimation in the context of time series models. In particular, the methods implemented are gradient descent, Newton-step and Kalman filter recursions. These algorithms are applied to the recently developed irregularly observed autoregressive (iAR) model. The estimation accuracy of the proposed methods is assessed by means of Monte Carlo experiments. The results obtained show that the proposed online estimation methods allow for a precise estimation of the parameters that generate the data both for the regularly and irregularly observed time series. These online approaches are numerically efficient, allowing substantial computational time savings. Moreover, we show that the proposed methods are able to adapt the parameter estimates quickly when the time series behavior changes, unlike batch estimation methods.

Keywords Autoregressive model · Online estimation · Streaming data · Gradient descent

Supported by ANID Fondecyt Iniciacion 11200590 grant and the ANID Millennium Science Initiative ICN12_009, awarded to the Millennium Institute of Astrophysics.

F. Elorrieta (✉) · L. Osses · M. Cáceres

Department of Mathematics and Computer Science, Universidad de Santiago, Santiago, Chile
e-mail: felipe.elorrieta@usach.cl

S. Eyheramendy

Faculty of Engineering and Sciences, Universidad Adolfo Ibañez, Santiago, Chile

F. Elorrieta · S. Eyheramendy · W. Palma

Millennium Institute of Astrophysics, Santiago, Chile

1 Introduction

Several natural or social phenomena can be measured sequentially over time. There are a wide range of tools to explain the temporal behavior of these phenomena. Among the most popular methods are the autoregressive moving average (ARMA) models. This model can be useful, for example, to explain the time dependence of a stationary time series and to forecast its future behavior. However, these forecasts become very uncertain over a large prediction horizon. This is an issue, since in many cases temporal data are still being collected, which makes the time series models useless if they are not updated recurrently.

This updating can be done in two ways: in batch or online setting. In the online setting, each training instance is processed once “on arrival” without the need for storage and reprocessing. In contrast, for batch learning, the whole dataset is required to recalculate the current learning.

Avoiding storing and reprocessing previous information can be very useful to optimize the computation time of the models if we need to fit large time series or if we intend to model many time series simultaneously. The second case is a very common challenge in time series classification based on features that are extracted from the temporal behavior of each time series [19]. For example, in astronomy, astronomical objects can be classified by the temporal behavior of their brightness [5, 6, 23, 25]. However, currently astronomical data is being processed in real time [2, 16] so that constant updating of time series models is a challenge which is addressed in this work.

Several works have already addressed the problem of online estimation in regularly observed time series [1, 18]. However, in some cases temporal data may not be obtained regularly. This problem occurs frequently in astronomy [11], climatology [20] or high-frequency finance [14].

Some models have been proposed in the literature to fit irregularly observed time series. These models can be separated into two groups. The first one includes models that assume continuous times, such as the Continuous Autoregressive Moving Average (CARMA [17]) or Continuous-Time Fractionally Integrated ARMA Process (CARFIMA [27]) models. These models are the solution of differential stochastic equations and assume that there are small time gaps between observations.

On the other hand, the second group considers discrete times representation of irregularly observed time series. Some models that follow this approach are the irregular autoregressive (iAR [10]), complex irregular autoregressive (CiAR [7]), bivariate irregular autoregressive (BiAR [8]) and the irregular autoregressive moving average (iARMA [22]) models.

In this work, we are particularly interested in proposing online estimation methods for the irregularly observed autoregressive (iAR) model. For this purpose, we follow two different approaches. First, following the work of Anava et al. [1], we use numerical methods such as Gradient Descent and Newton–Raphson to derive an update equation for the parameter of the iAR model. In the second approach, we

obtain a recursive estimate of the posterior distribution of the model parameter, under a Bayesian structure [26].

The structure of this paper is as follows. In Sect. 2, we describe the irregularly observed autoregressive model and the three online estimation methods proposed in this work. In Sect. 3, we perform Monte Carlo experiments to evaluate the proposed online estimation methods in different simulated scenarios. The application to real-life data is shown in Sect. 4. Finally, the main findings of this work are discussed in Sect. 5.

2 Methods

2.1 Irregular Autoregressive Model

The irregular Autoregressive (iAR) model was introduced by Eyheramendy et al. [10] and it is defined as

$$y_{t_j} = \phi^{t_j - t_{j-1}} y_{t_{j-1}} + \sigma \sqrt{1 - \phi^{2(t_j - t_{j-1})}} \epsilon_{t_j}, \quad (1)$$

where ϵ_{t_j} is a white noise sequence with zero mean and unit variance, σ is the standard deviation of y_{t_j} , and t_j are the observational times such that $j = 1, \dots, n$.

The minus log-likelihood function of the process, under Gaussianity, is given by

$$l(\theta) = \frac{n}{2} \log(2\pi) + \frac{1}{2} \sum_{i=1}^n \log v_{t_i} + \frac{1}{2} \sum_{i=1}^n \frac{e_{t_i}^2}{v_{t_i}}, \quad (2)$$

where the initial values are defined by $e_{t_1} = y_{t_1}$, $v_{t_1} = \sigma^2 + \delta_{t_1}^2$. In addition, the one-step predictor is defined by $\hat{y}_{t_j} = \phi^{t_j - t_{j-1}} y_{t_{j-1}}$, while the innovation is defined as $e_{t_j} = y_{t_j} - \hat{y}_{t_j}$. The innovation variance is $v_{t_j} = \sigma^2(1 - \phi^{2(t_j - t_{j-1})}) + \delta_{t_j}^2$, where $\delta_{t_j}^2$ is the known variance of the measurement errors. Note that the parameter ϕ describes the autocorrelation function of the process. Furthermore, y_{t_j} is a weakly stationary process under $0 < \phi < 1$.

Generally, this model is fitted in an irregularly observed time series using all the information available, i.e., in a batch setting. As mentioned above, in this work, we propose online estimation procedures for the iAR model. With these methods, the estimation of the parameter ϕ can be updated when new observations arrive without the need to reprocess the previous information. In this work, we propose three online estimation methods, which are introduced in the following subsection.

2.2 Online Estimation Algorithms

A first approach to implement the online estimation algorithms for the iAR model is based on the first-order optimization of a given loss function [1]. We propose two estimation algorithms that follow this idea. The first one is called the Online Newton Step (ONS) method. The ONS method was introduced by Hazan et al. [15] and is based on the Newton–Raphson method. The main goal of this method is to exploit the information to accelerate the convergence of the optimization. On the other hand, we propose to use the Online Gradient Descent (OGD) method [3, 28], which aims to optimize the parameter estimation by computing the gradient of a loss function.

A second approach is based on a Bayesian estimation method for the linear regression model. The algorithm that was built following this idea is the Online Bayesian Regression (OBR), which assumes a prior distribution on the parameter of the model and combines it with the likelihood of the observations to obtain the posterior distribution. The parameters of the resulting distribution contain the estimation results for the parameter of the iAR model.

iAR Online Newton Step (iAR-ONS) The ONS algorithm uses the second derivative of the loss function. The adaptation of this algorithm for online estimation of the parameter (ϕ) of the iAR process is as follows:

Input: ϕ_1 initial value; learning rate η ; $A_1 = \eta$.

for $j = 1$ **to** $n-1$ **do**

 Get $\hat{y}_{t_j} = \phi_j^{t_j - t_{j-1}} y_{t_{j-1}}$

 Observe y_{t_j} and obtain the loss function $l_j(\phi_j)$;

 Let $\nabla_j = \nabla l_j(\phi_j)$; Update $A_{j+1} = A_j + \nabla_j \nabla_j$;

 Parameter update $\phi_{j+1} = (\phi_j - \frac{1}{\eta} A_{j+1}^{-1} \nabla_j)$

end

Algorithm 1: iAR-ONS

In our implementation, we assume that $l_j(\phi_j)$ is the quadratic loss function defined by $(y_{t_j} - \hat{y}_{t_j})^2$. Later, the gradient is defined by $\nabla_j = -2(y_{t_j} - \hat{y}_{t_j})y_{t_{j-1}}(t_j - t_{j-1})\phi_j^{t_j - t_{j-1} - 1}$.

iAR Online Gradient Descent (iAR-OGD) The second online estimation method that we propose in this work is the iAR-OGD algorithm, which aims to optimize the parameter estimation of the iAR process using the gradient descent method. The steps to implement this algorithm are as follows:

Input: ϕ_1 initial value; learning rate η .

```

for  $j = 1$  to  $n-1$  do
  Get  $\hat{y}_{t_j} = \phi_j^{t_j - t_{j-1}} y_{t_{j-1}}$ ;
  Observe  $y_{t_j}$  and obtain the loss function  $l_j(\phi_j)$ ;
  Let  $\nabla_j = \nabla l_j(\phi_j)$ ;
  Parameter update  $\phi_{j+1} = (\phi_j - \frac{1}{\eta} \nabla_j)$ 
end

```

Algorithm 2: iAR-OGD

Here the loss function $l_j(\phi_j)$ and the gradient ∇_j are defined as in the iAR-ONS algorithm.

iAR Online Bayesian Regression (iAR-OBR) The OBR algorithm assumes a Gaussian prior distribution on the parameter ϕ of the iAR model, and combines it with the likelihood of the current observation being analyzed, which is also Gaussian. In addition, the parameter ϕ is defined by a random walk between the observations [26]. This information can be combined to obtain the posterior distribution of the parameter of the iAR model given the first j observations, in which its mean and its variance are the estimated parameters ϕ_j and P_j respectively. The parameters of the posterior distribution can be obtained using a modified version of the Kalman recursions, which are defined as follows:

Input: ϕ_1, P_1 initial values.

```

for  $j = 2$  to  $n$  do
  Observe  $y_{t_{j-1}}$  and  $y_{t_j}$ ;
  Let  $S_j = y_{t_{j-1}} P_{j-1} y_{t_{j-1}} + \sigma^2(1 - \phi_{j-1}^{2(t_j - t_{j-1})})$ ;
  and  $K_j = P_{j-1} y_{t_{j-1}} S_j^{-1}$ ;
  Update  $\phi_j = \phi_{j-1} + K_j [y_{t_j} - y_{t_{j-1}} \phi_{j-1}^{(t_j - t_{j-1})}]$ ;
  and  $P_j = P_{j-1} - K_j S_j K_j$ ;
end

```

Algorithm 3: iAR-OBR

3 Simulation Experiments

In this section, we assess the parameter estimation accuracy of the three online estimation methods for the iAR model proposed in this work. For the following experiments, we generate each iAR process with 400 observations with random observational times coming from the following four distributions:

- Regular time with constant gaps.
- Time gaps following a Uniform distribution with parameters $a = 0.5$ and $b = 1.5$.
- Time gaps following a Gamma distribution with parameters $\alpha = 3$ and $\beta = 3$.

- Time gaps following a mixture of two exponential distributions with means $\lambda_1 = 15$ and $\lambda_2 = 2$ and weights $w_1 = 0.15$ and $w_2 = 0.85$, respectively, such that

$$f(\nabla_j | \lambda_1, \lambda_2, \omega_1, \omega_2) = \omega_1 g(\nabla_j | \lambda_1) + \omega_2 g(\nabla_j | \lambda_2) \quad \forall j = 2, \dots, n$$

where $g(\cdot)$ is the exponential distribution and $\nabla_j = t_j - t_{j-1}$ is the j -th time gap.

We chose these distributions in order to assess whether the size of the gaps in which the observations are taken can affect the results of the estimation methods. Both for the generation of the iAR process and the observational times, we use functions available in the iAR package of R [9].

For each time distribution, we implemented the three following scenarios. First, we perform a simple sanity check, in which we generate the iAR model with a parameter ϕ and assess whether the proposed methods estimate the parameter accurately. The second experiment consists of an abrupt change in the parameter of the model, i.e., the parameter of the model remains constant for the first half of each simulated iAR process and for the second half it changes to another value. Finally, the last scenario is the constant change, in which the parameter of the model changes slowly in time. The abrupt change and the constant change experiments allow to evaluate whether the online estimation can adapt to changes in the process structure.

Following the Monte Carlo method, each experiment was repeated 100 times. In order to assess the proposed estimation methods we use three evaluation measures. The first of them is the parameter estimation accuracy. Second, we evaluate the goodness of fit by computing the Mean Squared Error (MSE) of the fitted values. Finally, we evaluate the computation time of the proposed algorithms.

For the sanity check and constant change scenarios, we assumed that the first 50% of the measurements of the time series were observed, while for the abrupt change scenario we assumed that the initial 62.5% of the time series was observed. From the observed values, we estimate the iAR model parameter in a batch setting. This batch estimate was used as the initial value of the parameter for the online estimation methods performed for the remaining observations. Finally, in each scenario, we also compare the online estimation methods with the batch estimation which uses all the values of each time series. The batch estimation of the iAR model is obtained from the `IARloglik` function of the iAR package. From now we will refer to it as iAR-MLE.

Table 1 shows a summary of the Monte Carlo experiments for the sanity check scenario where the parameter of the simulated model is $\phi = 0.5$ for the four distributions used to generate the observational times. In this table we present the last value estimated by each estimation method, denoted by $\hat{\phi}_{400}$. Note that the three online estimation methods achieve a fairly accurate estimate at the end of the time series. In the comparison between the online estimation methods, the iAR-ONS method stands out with a closer estimation to the true value of the parameter for three of the four time distributions.

In order to assess to the goodness of fit of each estimation method, we present the MSE obtained from the fitted values estimated from the beginning of the online

Table 1 Summary of the Monte Carlo experiments for the sanity check scenario performed using a parameter $\phi = 0.5$ in the iAR model for the four distributions used to generate the observational times. Columns 2–5 of this table show the last value estimated for each method. The last four columns show the mean squared error (MSE) computed for the fitted values

Obs. time	$\hat{\phi}_{400}$ MLE	$\hat{\phi}_{400}$ OBR	$\hat{\phi}_{400}$ OGD	$\hat{\phi}_{400}$ ONS	MSE MLE	MSE OBR	MSE OGD	MSE ONS
Regular	0.502	0.48	0.49	0.504	0.74	0.7	0.74	0.78
Unif (0.5, 1.5)	0.5	0.475	0.479	0.496	0.74	0.7	0.74	0.77
Gamma (3, 3)	0.502	0.473	0.484	0.494	0.66	0.62	0.66	0.69
Exp. M(15, 2, 0.15, 0.85)	0.503	0.481	0.481	0.533	0.76	0.71	0.73	0.8

estimation until the end of each simulated series. From the MSE we observe that the batch estimation method has higher values than the iAR-OBR method, indicating that the use of this online estimation method gives better results for the fitted values in this setting.

Furthermore, the complete trajectory of the parameters estimated using the proposed methods are presented in Fig. 1. Note from Fig. 1a–c that the online estimation methods converge quickly to the true parameter, particularly when the time gaps are small.

Table 2 shows the results of the Monte Carlo simulations for the abrupt change scenario in which the parameter of each simulated process of size $n = 400$ starts with a parameter of $\phi = 0.7$ for the first 200 observations and changes to $\phi = 0.3$ for the last 200 observations. As can be noticed, the proposed online estimation methods can adapt to the structural change produced in the time series giving a last parameter estimation closer to the true parameter of the second half of time series, unlike the batch estimation which obtains an approximated last estimated parameter of 0.5 for each observational time. The fact that the batch estimation fails to estimate well the parameters used to generate each time series impacts on the time series fit, as can be seen from the mean squared error value. As in the sanity check scenario, the online method that obtains the best fit to the series is the iAR-OBR method, which achieves the lowest mean squared error for the four distributions used to generate the observational times. Figure 1d–f shows that the iAR-OGD method converges slower to the true value of the parameter than the remaining online estimation methods.

The last simulation experiment that we perform is the constant change scenario. In this example, the parameter of the iAR model decreases constantly from an initial value of $\phi = 0.8$ to a final value of $\phi = 0.4$. Table 3 shows that, as in the abrupt change scenario, the batch estimation of the iAR model parameter is very distant from the true parameter. In particular, the last value estimated by the batch method is close to the average of the values used to generate the simulated time series. In contrast, the online estimation methods achieve values closer to the last parameter. Here the iAR-OBR method stands out with a final parameter estimate closer to 0.4. Regarding the goodness of fit, the iAR-OBR method again has the best fitted values

Table 2 Summary of the Monte Carlo experiments for the abrupt change scenario performed using a parameter $\phi = 0.7$ for the first half and $\phi = 0.3$ for the second half of each simulated time series for the four distributions used to generate the observational times. Columns 2–5 of this table show the last value estimated for each method. The last four columns show the mean squared error computed for the fitted values

Obs. time	$\hat{\phi}_{400}$ MLE	$\hat{\phi}_{400}$ OBR	$\hat{\phi}_{400}$ OGD	$\hat{\phi}_{400}$ ONS	MSE MLE	MSE OBR	MSE OGD	MSE ONS
Regular	0.502	0.294	0.309	0.319	0.93	0.84	0.94	0.96
Unif (0.5, 1.5)	0.494	0.299	0.316	0.31	0.92	0.82	0.91	0.92
Gamma (3, 3)	0.497	0.282	0.327	0.3	0.85	0.74	0.82	0.82
Exp. M(15, 2, 0.15, 0.85)	0.492	0.311	0.369	0.399	0.85	0.8	0.86	0.91

Table 3 Summary of the Monte Carlo experiments for the constant change scenario performed using an initial parameter of $\phi = 0.8$ which decreases to $\phi = 0.4$ for the four distributions used to generate the observational times. Columns 2–5 of this table show the last value estimated for each method. The last four columns show the mean squared error computed for the fitted values

Obs. time	$\hat{\phi}_{400}$ MLE	$\hat{\phi}_{400}$ OBR	$\hat{\phi}_{400}$ OGD	$\hat{\phi}_{400}$ ONS	MSE MLE	MSE OBR	MSE OGD	MSE ONS
Regular	0.604	0.409	0.416	0.479	0.75	0.64	0.69	0.71
Unif (0.5, 1.5)	0.597	0.427	0.429	0.483	0.75	0.65	0.7	0.72
Gamma (3, 3)	0.597	0.431	0.453	0.483	0.68	0.57	0.6	0.63
Exp. M(15, 2, 0.15, 0.85)	0.599	0.443	0.465	0.541	0.76	0.71	0.74	0.78

according to the mean squared error. Finally, Fig. 1g–i shows a very slow decay of the batch estimation, while the online estimations adapt quickly to the true parameter, particularly the iAR-OBR and iAR-ONS methods.

In order to evaluate the efficiency of the estimation methods proposed in this work, we obtained the computation times of each estimation method for different sample sizes of the simulated times series, from $n = 50$ to $n = 600$. For each sample size, the computation time was obtained 100 times. Later, we calculated the mean of the computation times obtained in each repetition. In Fig. 2, it can be observed that the online estimation methods are faster than the batch estimation for all the sample sizes evaluated. Particularly, for time series of 100 observations, batch estimation (0.0131 s) takes on average 12 times longer than iAR-OGD and iAR-ONS estimation (0.0011 s) and 4 times longer than iAR-OBR estimation (0.0036 s). Furthermore, this difference increases as more observations are required to be estimated.

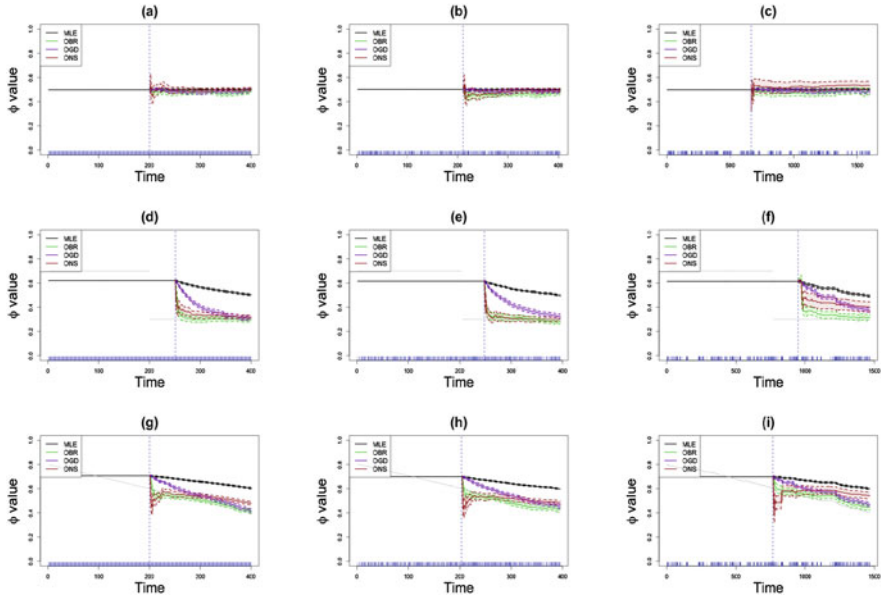


Fig. 1 Mean and confidence intervals of the online estimates for simulated irregular autoregressive processes. The three rows represent (from top to bottom) the sanity check, abrupt change, and constant change experiments. The three columns represent (from left to right) the examples for regular, Gamma (3, 3) and Exp. M(15, 2, 0.15, 0.85) times. The black line represents the batch estimation of the iAR model. The brown, purple, and green lines represent the parameter estimation using the iAR-ONS, iAR-OGD, and the iAR-OBR methods, respectively. Finally, the gray line is the true parameter with which the time series were simulated

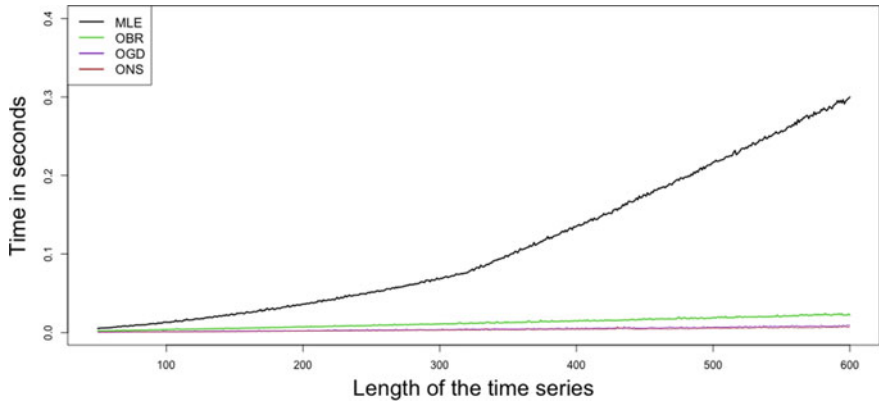


Fig. 2 Mean of the computation times of the estimation methods for the iAR model for simulated processes with sample sizes from $n = 50$ to $n = 600$. The black line represents the batch estimation of the iAR model. The brown, purple, and green lines represent the parameter estimation using the iAR-ONS, iAR-OGD, and the iAR-OBR methods, respectively

4 Application to Real-Life Data

4.1 Flow of the River Nile

For our first application, we use the Nile dataset available in R. In this time series, the observations correspond to the measurement of the annual flow of the river Nile below the Aswan Dam, in 10^8 m^3 . This dataset contains 100 observations regularly observed between the years 1871 and 1970. The Nile time series has been analyzed in several previous studies because it has an apparent change point in the year 1898, which was first detected by Cobb [4]. Figure 3a shows the Nile time series. Before implementing the estimation methods proposed in this work, we preprocessed the data to ensure a constant mean and variance equal to one. For this purpose, we first estimate the trend of the time series using the lowess method. Later, we remove the trend from the series. Finally, we standardize the de-trended time series.

Figure 3b shows that the online estimation methods proposed in this work detect this change point, exhibiting a break in the increasing trend of the parameter estimates observed before 1898. After this break, the parameter estimates slowly decrease until reaching an estimate close to the batch estimate (0.258) at the end of the time series. Here, the iAR-OBR method has its last estimate (0.261) closest to the Batch estimate of the parameter of the iAR model.

4.2 Infant Heart Rate

The second real dataset that we use in this work is the time series of the record of a 66-day-old infant heart rate (in beats per minute). This time series is available in the R package wavethresh [13] under the name BabyECG. The time series contains

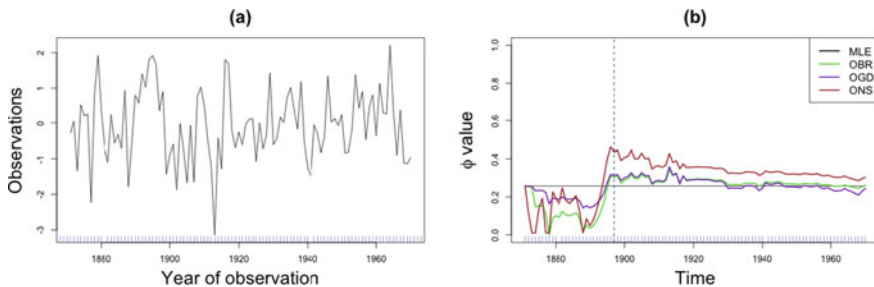


Fig. 3 Online estimation of the flow of the river Nile time series. **a** Shows the time series of the flow of the river Nile. **b** Shows the parameter estimation at each point of the time series. The black line represents the batch estimation of the iAR model. The brown, purple, and green lines represent the parameter estimation using the iAR-ONS, iAR-OGD, and the iAR-OBR methods, respectively. The vertical line represents the year in which the change point occurs

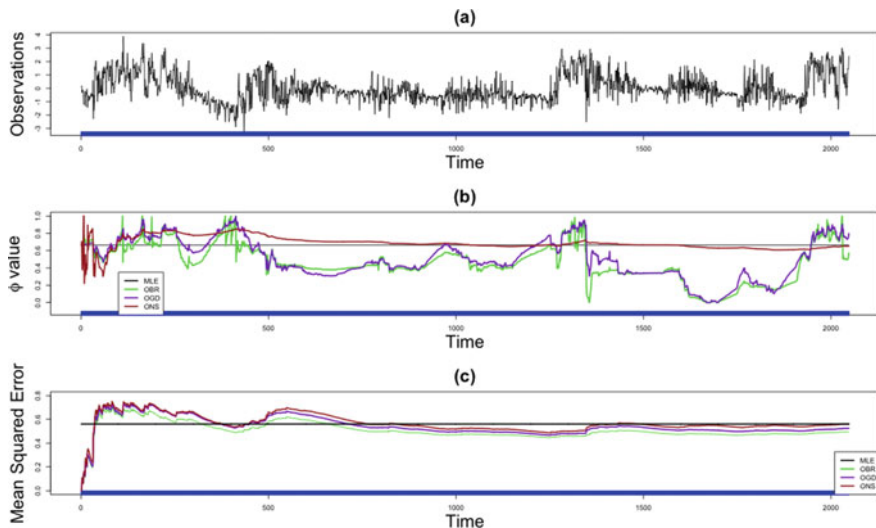


Fig. 4 Online estimation of the infant heart rate time series. **a** Shows the time series of the infant heart rate. **b** Shows the parameter estimation at each point of the time series. **c** Shows the mean squared error estimated at each point of the time series. The black line represents the batch estimation of the iAR model. The brown, purple, and green lines represent the parameter estimation using the iAR-ONS, iAR-OGD, and the iAR-OBR methods, respectively

2048 observations sampled regularly every 16 seconds between the 21:17:59 of a day and the 06:27:18 of the following day. This data was introduced by Nason et al. [21] as an example of a locally stationary time series due to the variations in its behavior over time, as can be seen in Fig. 4a. Considering these variations over time, the BabyECG time series is an interesting example to assess the ability of online estimation methods to adapt to structural changes in the time series. As in the previous example, we remove the trend from the time series and then standardize it before applying the estimation methods.

Figure 4 shows high variability in the parameter estimates obtained from the online estimation methods, particularly for the iAR-OGD and iAR-OBR. This result indicates that the online estimation methods are able to detect the structural changes in the time series. Furthermore, this can be validated by the goodness of fit obtained for this time series from the online estimations. Note in Fig. 4c that the mean squared error obtained by the online estimation methods is consistently lower than those obtained from the batch estimation for more than half of the observations in the time series. By calculating the mean of the estimated mean squared errors for each observation, we note that the method with the best fit of this time series is the iAR-OBR with an average MSE of 0.51. This value is lower than the MSE obtained from the batch estimation (0.56).

4.3 *Brightness of Astronomical Object*

For the last application to real data, we use the time series of the brightness of an astronomical object. Unlike the time series used in the previous examples, this time series is irregularly observed. The astronomical object we consider in this example corresponds to an Active Galactic Nuclei (AGN) observed in the ZTF survey [2] and coded as “ZTF20aagiimu”. The data processing was done by the broker ALerCE [12]. This object has been studied by Sanchez-Saez et al. [24], who have proposed it as a candidate for a Changing-state AGN, i.e., an object that changes state in time, so it is interesting to see if the estimation methods proposed in this work are able to detect these behavioral changes. This time series contains 177 observations that were measured in more than 3 years. This astronomical time series is presented in Fig. 5a. Before implementing the estimation methods, we did the preprocessing of the time series explained in the example in Sect. 4.1.

Figure 5b shows that the three online estimation methods obtain an abrupt increase in the estimated parameter during the second half of the time series, which may be explained by a change of the state of this astronomical object. Furthermore, if we observe the mean squared error obtained for each estimation method we notice that again these values are lower for the iAR-OBR method (Fig. 5c), which reaches an average MSE of 0.6, while the Batch estimation reaches an MSE of 0.7. According to this result, a parameter estimate which is able to be adapted to the new observations obtains better fitted values for this time series than one that uses all the observed data.

5 Discussion

In this work, we presented three online estimation methods for the irregular autoregressive (iAR) model. These methods are the first approach in the literature to the online estimation of irregularly observed autoregressive processes, since so far the online estimation methods applied to time series have been implemented on regular-time ARMA processes.

In the experiments carried out throughout this paper, it is observed that the proposed methods show at least two advantages over batch estimation. First, the online estimation methods have proven to be capable of adapting to structural changes in the processes. In addition, we show that the proposed methods have shorter computation times than the batch maximum likelihood estimator.

Among the proposed estimation methods, the iAR-OBR method achieves the best goodness-of-fit indicators in the time series analyzed. In addition, this method showed a faster adaptation to the structural changes of the time series. On the other hand, the iAR-ONS and iAR-OGD methods are less computationally expensive.

Based on the results obtained, we consider that online estimation methods can make an efficient estimation of streaming data, where new data arrive frequently and

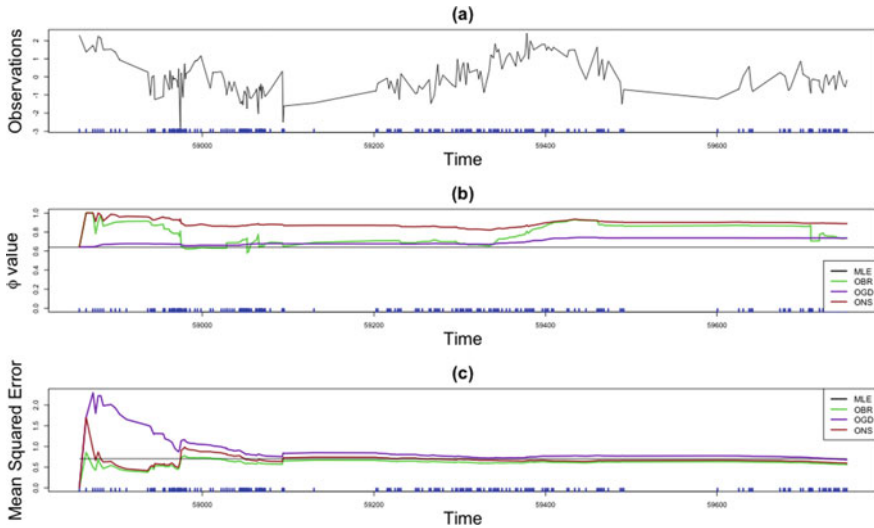


Fig. 5 Online estimation of the brightness of an astronomical object. **a** Shows the time series of the brightness of a changing-state AGN candidate, **b** Shows the parameter estimation at each point of the time series. **c** Shows the mean squared error estimated at each point of the time series. The black line represents the batch estimation of the iAR model. The brown, purple, and green lines represent the parameter estimation using the iAR-ONS, iAR-OGD, and the iAR-OBR methods, respectively

models must be updated to avoid becoming obsolete. This efficiency is reached in terms of estimation accuracy and computation time.

In future works, we aim to extend these methods to perform online estimations of other time series models for irregularly observed data (such as the complex irregularly observed autoregressive model (CiAR [7]) and the bivariate irregularly observed autoregressive model (BiAR [8]) models).

References

1. Anava, O., Hazan, E., Mannor, S., Shamir, O.: Online learning for time series prediction. In: Shalev-Shwartz, S., Steinwart, I. (eds.) Proceedings of the 26th Annual Conference on Learning Theory. Proceedings of Machine Learning Research, vol. 30, pp. 172–184. PMLR, Princeton, NJ, USA (2013). <http://proceedings.mlr.press/v30/Anava13.html>
2. Bellm, E.C., Kulkarni, S.R., Graham, M.J., Dekany, R., Smith, R.M., Riddle, R., Masci, F.J., Helou, G., Prince, T.A., Adams, S.M., Barbarino, C., Barlow, T., Bauer, J., Beck, R., Belicki, J., Biswas, R., Blagorodnova, N., Bodewits, D., Bolin, B., Brinell, V., Brooke, T., Bue, B., Bulla, M., Burruss, R., Cenko, S.B., Chang, C.K., Connolly, A., Coughlin, M., Cromer, J., Cunningham, V., De, K., Delacroix, A., Desai, V., Duev, D.A., Eadie, G., Farnham, T.L., Feeney, M., Feindt, U., Flynn, D., Franckowiak, A., Frederick, S., Fremling, C., Gal-Yam, A., Gezari, S., Giomi, M., Goldstein, D.A., Golkhou, V.Z., Goobar, A., Groom, S., Hacopians, E., Hale, D., Henning, J., Ho, A.Y.Q., Hover, D., Howell, J., Hung, T., Huppenkothen, D., Imel, D., Ip, W.H., Ivezić, Ž., Jackson, E., Jones, L., Juric, M., Kasliwal, M.M., Kaspi, S., Kaye,

- S., Kelley, M.S.P., Kowalski, M., Kramer, E., Kupfer, T., Landry, W., Laher, R.R., Lee, C.D., Lin, H.W., Lin, Z.Y., Lunnan, R., Giomi, M., Mahabal, A., Mao, P., Miller, A.A., Monkevitz, S., Murphy, P., Ngeow, C.C., Nordin, J., Nugent, P., Ofek, E., Patterson, M.T., Penprase, B., Porter, M., Rauch, L., Rebbapragada, U., Reiley, D., Rigault, M., Rodriguez, H., van Roestel, J., Rusholme, B., van Santen, J., Schulze, S., Shupe, D.L., Singer, L.P., Soumagnac, M.T., Stein, R., Surace, J., Sollerman, J., Szkody, P., Taddia, F., Terek, S., Sistine, A.V., van Velzen, S., Vestrand, W.T., Walters, R., Ward, C., Ye, Q.Z., Yu, P.C., Yan, L., Zolkower, J.: The zwicky transient facility: system overview, performance, and first results. *Publ. Astron. Soc. Pacific* **131**(995), 018002 (2018). <https://doi.org/10.1088%2F1538-3873%2Faaecbe>
3. Cesa-Bianchi, N., Long, P.M., Warmuth, M.K.: Worst-case quadratic loss bounds for prediction using linear functions and gradient descent. *IEEE Trans. Neural Netw.* **7**(3), 604–619 (1996). <https://doi.org/10.1109/72.501719>
 4. Cobb, G.W.: The problem of the Nile: conditional solution to a changepoint problem. *Biometrika* **65**(2), 243–251 (1978), <http://www.jstor.org/stable/2335202>
 5. Debosscher, J., Sarro, L.M., Aerts, C., Cuypers, J., Vandenbussche, B., Garrido, R., Solano, E.: Automated supervised classification of variable stars. I. *Methodology* **475**, 1159–1183 (2007)
 6. Elorrieta, F., Eyheramendy, S., Jordán, A., Dékány, I., Catelan, M., Angeloni, R., Alonso-García, J., Contreras-Ramos, R., Gran, F., Hajdu, G., Espinoza, N., Saito, R.K., Minniti, D.: A machine learned classifier for RR Lyrae in the VVV survey **595**, A82 (2016)
 7. Elorrieta, F., Eyheramendy, S., Palma, W.: Discrete-time autoregressive model for unequally spaced time-series observations. *A&A* **627**, A120 (2019). <https://doi.org/10.1051/0004-6361/201935560>
 8. Elorrieta, F., Eyheramendy, S., Palma, W., Ojeda, C.: A novel bivariate autoregressive model for predicting and forecasting irregularly observed time series. *Mon. Not. R. Astron. Soc.* **505**(1), 1105–1116 (2021). <https://doi.org/10.1093/mnras/stab1216>
 9. Elorrieta, F., Eyheramendy Susana, P.W.: iAR: irregularly observed autoregressive models (2022). <https://CRAN.R-project.org/package=iAR>, r package version 1.1.9
 10. Eyheramendy, S., Elorrieta, F., Palma, W.: An irregular discrete time series model to identify residuals with autocorrelation in astronomical light curves. *Mon. Not. R. Astron. Soc.* **481**(4), 4311–4322 (2018). <https://doi.org/10.1093/mnras/sty2487>
 11. Feigelson, E.D., Babu, G.J., Caceres, G.A.: Autoregressive times series methods for time domain astronomy. *Front. Phys.* **6**, 80 (2018). <https://www.frontiersin.org/article/10.3389/fphys.2018.00080>
 12. Förster, F., Cabrera-Vives, G., Castillo-Navarrete, E., Estévez, P.A., Sánchez-Sáez, P., Arredondo, J., Bauer, F.E., Carrasco-Davis, R., Catelan, M., Elorrieta, F., Eyheramendy, S., Huijse, P., Pignata, G., Reyes, E., Reyes, I., Rodríguez-Mancini, D., Ruz-Mieres, D., Valenzuela, C., Álvarez-Maldonado, I., Astorga, N., Borissova, J., Clocchiatti, A., Cicco, D.D., Donoso-Oliva, C., Hernández-García, L., Graham, M.J., Jordán, A., Kurtev, R., Mahabal, A., Maureira, J.C., Muñoz-Arancibia, A., Molina-Ferreiro, R., Moya, A., Palma, W., Pérez-Carrasco, M., Protopapas, P., Romero, M., Sabatini-Gacitua, L., Sánchez, A., Martín, J.S., Sepúlveda-Cobo, C., Vera, E., Vergara, J.R.: The automatic learning for the rapid classification of events (ALeRCE) alert broker. *Astron. J.* **161**(5), 242 (2021). <https://doi.org/10.3847/1538-3881/abe9bc>
 13. Guy, N.: Wavethresh: wavelets statistics and transforms (2022). <https://CRAN.R-project.org/package=wavethresh>, r package version 4.7.0
 14. Hautsch, N.: *Econometrics of Financial High-Frequency Data*. Springer, Berlin (2012)
 15. Hazan, E., Agarwal, A., Kale, S.: Logarithmic regret algorithms for online convex optimization. *Mach. Learn.* **69**(2–3), 169–192 (2007)
 16. Ivezić, Ž., Kahn, S.M., Tyson, J.A., Abel, B., Acosta, E., Allsman, R., Alonso, D., AlSayyad, Y., Anderson, S.F., Andrew, J., et al.: LSST: from science drivers to reference design and anticipated data products. **873**, 111 (2019)
 17. Jones, R.H.: Fitting a continuous time autoregression to discrete data. In: *Applied Time Series Analysis*, pp. 651–682 (1981). <http://ci.nii.ac.jp/naid/10030486300/en/>
 18. Liu, C., Hoi, S.C.H., Zhao, P., Sun, J.: Online arima algorithms for time series prediction. In: *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pp. 1867–1873. AAAI'16, AAAI Press (2016)

19. Maharaj, E.A., D'Urso, P., Caiado, J.: Time Series Clustering and Classification, first edn. Chapman Hall/CRC Press (2019)
20. Mudelsee, M.: Climate time series analysis: classical statistical and bootstrap methods. In: Atmospheric and Oceanographic Sciences Library, second edn., vol. 51. Springer International Publishing (2014)
21. Nason, G.P., Von Sachs, R., Kroisandt, G.: Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Stat. Soc.: Ser. B (Stat. Methodol.)* **62**(2), 271–292 (2000). <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00231>
22. Ojeda, C., Palma, W., Eyheramendy, S., Elorrieta, F.: A novel first-order autoregressive moving average model to analyze discrete-time series irregularly observed. In: Valenzuela, O., Rojas, F., Herrera, L.J., Pomares, H., Rojas, I. (eds.) *Theory and Applications of Time Series Analysis and Forecasting. ITISE 2021. Contributions to Statistics*. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-14197-3_7
23. Richards, J.W., Starr, D.L., Butler, N.R., Bloom, J.S., Brewer, J.M., Crellin-Quick, A., Higgins, J., Kennedy, R., Rischard, M.: On machine-learned classification of variable stars with sparse and noisy time-series data. **733**, 10 (2011)
24. Sánchez-Sáez, P., Lira, H., Martí, L., Sánchez-Pi, N., Arredondo, J., Bauer, F.E., Bayo, A., Cabrera-Vives, G., Donoso-Oliva, C., Estévez, P.A., Eyheramendy, S., Förster, F., Hernández-García, L., Arancibia, A.M.M., Pérez-Carrasco, M., Sepúlveda, M., Vergara, J.R.: Searching for changing-state AGNs in massive data sets. i. applying deep learning and anomaly-detection techniques to find AGNs with anomalous variability behaviors. *Astron. J.* **162**(5), 206 (2021). <https://doi.org/10.3847/1538-3881/ac1426>
25. Sánchez-Sáez, P., Reyes, I., Valenzuela, C., Förster, F., Eyheramendy, S., Elorrieta, F., Bauer, F.E., Cabrera-Vives, G., Estévez, P.A., Catelan, M., Pignata, G., Huijse, P., Cicco, D.D., Arévalo, P., Carrasco-Davis, R., Abril, J., Kurtev, R., Borissova, J., Arredondo, J., Castillo-Navarrete, E., Rodríguez, D., Ruz-Mieres, D., Moya, A., Sabatini-Gacitúa, L., Sepúlveda-Cobo, C., Camacho-Iñiguez, E.: Alert classification for the ALeRCE broker system: the light curve classifier. *Astron. J.* **161**(3), 141 (2021). <https://doi.org/10.3847/1538-3881/abd5c1>
26. Särkkä, S.: Batch and recursive Bayesian estimation, pp. 27–50. *Institute of Mathematical Statistics Textbooks*, Cambridge University Press (2013)
27. Tsai, H.: On continuous-time autoregressive fractionally integrated moving average processes. *Bernoulli* **15**(1), 178–194 (2009). <https://doi.org/10.3150/08-BEJ143>
28. Zinkevich, M.: Online convex programming and generalized infinitesimal gradient ascent. In: *International Conference on Machine Learning*, pp. 928–935 (2003)

Costationary Whitenoise Processes and Local Stationarity Testing



Alessandro Cardinali

Abstract This paper provides three main contributions to the study of multiscale locally stationary processes. We introduce a new class of stochastic processes that we call *Costationary Whitenoise* (CWN) processes, and propose a theoretical framework to estimate their underlying parameters. We use this setup to devise a non-parametric approximation method for the unknown distribution function of unobservable innovations. We address this task by using CWN order statistics to approximate the quantiles of unobservable innovations. Finally, we use the above frameworks to derive a non-parametric bootstrap stationarity test for multiscale locally stationary processes. The finite sample performances of this test are assessed through simulations showing that our method successfully controls rejection rates for stationary and locally stationary processes with both Gaussian and Student- t distributed innovations. Finally, by applying our test to equity returns, we are able to associate the presence of non-stationarities to the occurrence of economic shocks.

Keywords Local stationarity · Costationarity · Whitenoise · Bootstrap test

1 Introduction

For a stationary time series the statistical properties remain constant over time. For a locally stationary (LS) time series the statistical properties can change slowly over time. As a consequence, such series can appear stationary when examined over short intervals, but appear non-stationary when examined on larger scales. Priestley [12, 13] provides a comprehensive review of locally stationary processes and their history. A more recent review is provided by Dahlhaus [4]. The methods described in this article can be applied to locally stationary time series that are a triangular stochastic arrays defined in the rescaled time t/T , where T represents the sample size. Based on this setup, Dahlhaus [3] proposed *locally stationary Fourier* (LSF) processes whose underlying pseudo-spectral structure is defined in terms of Fourier basis. The *locally*

A. Cardinali (✉)

School of Computing, Electronics and Mathematics, University of Plymouth, Plymouth, England
e-mail: alessandro.cardinali@plymouth.ac.uk

stationary wavelet (LSW) model due to Nason et al. [9], instead, decomposes the local structure of the process among different scales through a set of non-decimated wavelets used as basis functions. In the following, we will consider the latter family of processes defined as

$$X_{t;T} = \sum_{j=1}^{\infty} \sum_{k=-\infty}^{\infty} W_j \left(\frac{k}{T} \right) \psi_j(t-k) \epsilon_{j,k}, \quad (1)$$

where $\{\psi_j(t-k)\}_j$ is a family of discrete non-decimated wavelet filters whose local support includes the neighborhood of t , and is spanned by the index k . The parameter j is integer valued and represents the scale of the corresponding wavelet. The function $W_j(k/T)$ is a time localized amplitude of bounded variation, referring to dyadic scales indexed by j . Finally, $\epsilon_{j,k}$ is a sequence of doubly indexed i.i.d. standardized random variables. This setup allows the definition of a time-varying generalization of the classical spectra having a well-defined limit in the rescaled time $z \in [0, 1)$ defined as

$$S_j(z) = \lim_{T \rightarrow \infty} \left| W_j \left(\frac{[zT]}{T} \right) \right|^2, \quad (2)$$

where we have set $k = [zT]$, and $[x]$ is the integer part of x .

Assumption 1 For all $j = 1, 2, \dots, \infty$ and for $z \in [0, 1)$, the wavelet spectra satisfies $0 < \inf_z S_j(z) \leq \sup_z S_j(z) < \infty$. Moreover $\sum_j \sup_z S_j(z) < \infty$.

This multiscale LS framework has proven useful to estimate time-varying unconditional statistical properties of non-stationary time series. The estimation of local spectra is relevant in a wide range of disciplines such as climatology, neuroscience, and economics, where the underlying phenomena are characterized by regime changes that cannot be appropriately taken into account by classical stationary models. A biased estimate of the multiscale spectra can be obtained by the squared non-decimated wavelet coefficients $I_{l,k} = |\sum_t \psi_l(t-k) X_{t;T}|^2$, whereas an asymptotically unbiased estimate can be obtained as in Nason et al. [9] by

$$L_{j,k} = \sum_l A_{j,l}^{-1} I_{l,k}, \quad (3)$$

where the $A_{j,l}$ are inner products of two autocorrelation wavelets defined at scales j and $l = 1, 2, \dots, \infty$, see Nason et al. [9] for further details.

Assumption 2 For each $j = 1, 2, \dots, \infty$ and for $k = 0, 1, \dots, T-1$, the wavelet periodogram sequences $\{L_{j,k}\}_k$ have the L_2 NED property, as stated in Definition 17.2 by Davidson [6].

In this paper, we use this framework to refine the study of time-varying linear combinations of LS processes which are (co)stationary, see Cardinali and Nason [1]. In particular, we devise a methodology to discover linear combinations which are

not only stationary but also uncorrelated, and we call these *Costationary Whitenoise* (CWN) processes. Furthermore, we use these new combinations to devise a bootstrap non-parametric test for local stationarity. The contribution of this paper is therefore threefold. We first propose a new consistent estimation methodology for the leading coefficients of CWN processes. We then use CWN sequences to estimate the CDF of (unobservable) locally stationary innovations. Finally, we incorporate this methodology into a bootstrap local stationarity test. The article is structured as follows. Section 2 reviews the costationarity framework and the new estimation approach for parameters of CWN processes. Section 3 illustrates the CDF estimation of innovations based on CWN sequences, and Sect. 4 introduces the bootstrap non-parametric local stationarity test. Section 5 shows the empirical size and local power of the test by means of simulations. Section 6 illustrates an application of the test to equity returns.

2 Costationary Whitenoise Processes

In this section we introduce a new class of stochastic processes and propose a consistent estimator for their underlying parameters. Given $X_{t;T}$ and $Y_{t;T}$ (possibly) locally stationary processes, let

$$Z_t = \alpha_t X_{t;T} + \beta_t Y_{t;T}, \tag{4}$$

where α_t and β_t are deterministic functions of bounded variation. In the following discussion we will focus on piecewise constant functions α_t, β_t , measurable on a disjoint sequence of half-opened dyadic intervals. In this paper, intervals of dyadic length have been considered for computational convenience; however, in principle, the theory we present will apply to intervals of arbitrary length. For a discussion on segmentation issues and regularity conditions concerning costationary solutions, we refer the interested reader to Cardinali and Nason [1]. Therefore, define $\tilde{L}_{j,k}^Z$ and \bar{L}_j^Z , respectively, as the smoothed unbiased multiscale periodogram and average periodogram for the Z_t sequence. In the following, the periodogram will be smoothed using the wavelet shrinkage approach as outlined in Nason et al. [9]. Cardinali and Nason [1] showed that when the functions α_t and β_t are estimated as a result of the optimization

$$(\hat{\alpha}_t, \hat{\beta}_t) = \underset{\alpha_t, \beta_t}{\operatorname{argmin}} \frac{1}{J_T T} \sum_{j,k} (\tilde{L}_{j,k}^Z - \bar{L}_j^Z)^2, \tag{5}$$

the resulting combinations are second-order (co)stationary sequences. As in Cardinali and Nason [1], we will consider a_t and b_t piecewise constant deterministic functions with C equally spaced changepoints each.

Remark 1 In Cardinali and Nason [1], the functions α_t and β_t were expressed in terms of orthogonal Haar wavelet expansions depending upon C coefficients each, thus allowing for an efficient optimization process. We use the same approach here

for the functions a_t and b_t . Therefore, the optimization will be conducted targeting the underlying Haar wavelet coefficients $\mathbf{a} = a_1, \dots, a_C$ and $\mathbf{b} = b_1, \dots, b_C$ rather than the piecewise constant functions a_t and b_t . The optimization will start for $C = 1$ and continue for larger values, until successful convergence is achieved.

Therefore, in this section, we consider time-varying linear combinations

$$W_t = a_t X_{t:T} + b_t Y_{t:T}, \quad (6)$$

which are not only stationary but also uncorrelated. We call these combinations *Costationary Whitenoise* (CWN) processes. Define the parameters $\theta = (\mathbf{a}, \mathbf{b})$ and let $\hat{\theta}$ be their estimate. Then, consistent estimates of θ can be obtained as a result of the optimization

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} \frac{1}{T^2} \sum_{j=1}^{J_T} \left(\sum_{k=0}^{T-1} |\tilde{L}_{j,k}^\theta - 2^{-j}| \right)^2, \quad (7)$$

where $\tilde{L}_{j,k}^\theta$ is the smoothed version of the unbiased multiscale periodogram for the W_t sequence. The loss function appearing in (7) is based on the results from Fryzlewicz et al. [7], where it was shown that the multiscale spectra for a whitenoise process is given by $S_j(z) = 2^{-j}$ for $z \in [0, 1)$, and $j = 1, 2, \dots$. Therefore, the metric from (7) imposes stronger conditions on the spectra of the resulting processes than those from (5). In fact its minimization requires the spectra of W_t not only to be time invariant for each scale $j = 1, 2, \dots$ (implying stationarity as for processes Z_t), but also imposes a specific rate of decay across scales. This latter additional condition also implies absence of autocorrelations for the sequences W_t . The construction of CWN sequences from real data requires the sequences \hat{a}_t and \hat{b}_t to be consistent estimators of the functions a_t and b_t , respectively. The following result establishes this property.

Proposition 1 *Let a_t and b_t be piecewise constant functions with C equally spaced changepoints each. For $t = 0, 1, \dots, T - 1$, let W_t be defined as in (6) and let $C \leq \log_2(T)/2$. Define \hat{a}_t and \hat{b}_t as the result of the optimization defined in (7). Then $\hat{a}_t \xrightarrow{P} a_t$ and $\hat{b}_t \xrightarrow{P} b_t$ as $T \rightarrow \infty$.*

Proof (Proposition 1) The dependence of the CWN smoothed periodogram and spectra with respect the piecewise constant functions a_t and b_t and their underlying Haar wavelet coefficients $\theta = (\mathbf{a}, \mathbf{b})$ will be denoted as $\tilde{L}_{j,k}^\theta$ and $S_j^\theta(z)$, respectively. The same notation will apply to sequences, vectors, and matrices containing multiscale periodograms and spectra of CWN processes. The proof is based on the fact that the r.h.s. of Eq. (7) can be rewritten as a quadratic form

$$\frac{1}{T^2} \sum_j^{J_T} \left(\sum_{k=0}^{T-1} |\tilde{L}_{j,k}^\theta - 2^{-j}| \right)^2 = \hat{\mathbf{g}}_T(\theta)' W \hat{\mathbf{g}}_T(\theta), \quad (8)$$

where $\hat{\mathbf{g}}_T(\theta)$ is a J_T -dimensional vector and W is a square identity matrix. For $j = 1, \dots, J_T$ the elements of the vector $\hat{\mathbf{g}}_T(\theta)$ are $\hat{g}_{j:T}(\theta) = T^{-1} \sum_k |\tilde{L}_{j,k}^\theta - 2^{-j}|$, where the dependence of $L_{j,k}^\theta$ with respect to the data and the functions a_t and b_t is illustrated by Eqs. (6) and (3). Moreover, W is the square identity matrix of dimension(s) $\log_2(T)$. [1] showed that any combination Z_t (therefore also combinations W_t) is an LS process with well-defined unbiased multiscale periodogram and spectra so that $\lim_{T \rightarrow \infty} \mathbb{E} L_{j,k}^\theta = S_j^\theta(z) + O(T^{-1})$. This setup can lead to an *identified* one-step Generalized Method of Moments (GMM) estimator (Hansen, 1982), when the number of moments ($\log_2(T)$) is larger than the number of parameters ($2C$). The consistency of this estimator can be proved by verifying that the following conditions are satisfied. These are the rescaled time equivalent of the conditions in Theorem 2.5 by Newey and McFadden [10], up to a remainder of order $O(T^{-1})$.

Lemma 1 *Let $\hat{\mathbf{g}}_T(\theta)$ be a vector-valued function, continuous with respect to θ , such that $\mathbf{g}(\theta) = \lim_{T \rightarrow \infty} \mathbb{E}[\hat{\mathbf{g}}_T(\theta)] = O(T^{-1})$ if and only if $\theta = \theta_0 \in \Theta$, for some compact set Θ . In addition let $\hat{\mathbf{g}}_T(\theta)$ and W be defined as in Eq. (8). Then Proposition 1 holds if the following conditions are fulfilled:*

1. $\hat{\mathbf{g}}_T(\theta) \xrightarrow{P} \mathbf{g}(\theta)$;
2. W converges in probability to a positive definite matrix;
3. $W\mathbf{g}(\theta) = O(T^{-1})$ if and only if $\theta = \theta_0$;
4. $\theta \in \Theta$ which is compact;
5. $\hat{\mathbf{g}}_T(\theta)$ is continuous with respect to θ with probability 1;
6. $\lim_{T \rightarrow \infty} \mathbb{E}[\sup_\theta \|\hat{\mathbf{g}}_T(\theta)\|] < \infty$.

Proof (Lemma 1) For all $j = 1, \dots, J_T$, we have

1. Since $\tilde{L}_{j,k}^\theta$ is an L_2 NED sequence, so is the absolute value of their difference from the deterministic term 2^{-j} . Therefore, Theorem 20.19 by Davidson [6] implies $\hat{g}_{j:T}(\theta) \xrightarrow{P} g_j(\theta) + O(T^{-1})$ where $g_j(\theta) = \int |S_j^\theta(z) - 2^{-j}| dz$. This determines the convergence in probability of the sample moments in $\hat{\mathbf{g}}_T(\theta)$.
2. $W = I$ is a deterministic identity matrix, which is positive definite.
3. $\lim_{T \rightarrow \infty} \mathbb{E} \hat{g}_{j:T}(\theta) = O(T^{-1})$ if and only if $S_j^\theta(z) = 2^{-j}$ for all $z \in [0, 1)$.
4. the parameter space is $\Theta = \mathbb{R}^{2C}$, which is a compact set.
5. $\hat{g}_{j:T}(\theta)$ is continuous for all $\theta \in \Theta$ since CWN processes are special cases of Locally Stationary processes, thus having well-defined multiscale periodogram and spectra. The condition follows by noting that for each k in $\hat{g}_{j:T}(\theta)$ we have $\lim_{T \rightarrow \infty} \mathbb{E}|L_{j,k}^\theta - 2^{-j}| = \pm(S_j^\theta(z) - 2^{-j}) + O(T^{-1})$ if $2^{-j} \leq S_j^\theta(z)$, respectively. We conclude by noting that the absolute value is continuous everywhere, including at the point $S_j^\theta(z) = 2^{-j}$.
6. $\lim_{T \rightarrow \infty} \mathbb{E} \sup_{\theta \in \Theta} \|\hat{g}_{j:T}(\theta)\| < \infty$ by Assumption 1 and the application of the continuous mapping theorem.

Lemma 1 establishes that $\hat{\theta}$ is a consistent estimator for the parameters θ , and therefore for the functions a_t, b_t that are obtained by the inverse Haar wavelet transform applied to the estimated parameters. This concludes the proof of Proposition 1.

We therefore define as realized Costationary Whitenoise processes the sequences

$$\widehat{W}_t = \widehat{a}_t X_{t:T} + \widehat{b}_t Y_{t:T}. \quad (9)$$

Since realized CWN processes are obtained as a result of the optimization (7), Proposition 1 implies that \widehat{W}_t will converge weakly to a zero-mean stationary and uncorrelated sequence.

3 Estimation of CDF for Innovations

Among the many possible specifications of W_t , we will mainly be interested in those where $X_{t:T}$ is a locally stationary (LS) process (for which data are available) and $Y_{t:T}$ is an independent Gaussian Whitenoise (GWN) process that we can simulate. Let $Y_{t:T} \sim GWN(0, \sigma^2)$ and W_t be defined as in Eq. (6), then some convenient statistical properties characterize the resulting CWN processes. In fact, higher order cumulants for W_t have a simplified form since all the Gaussian higher order cumulants of $Y_{t:T}$ are equal to zero. This property, along with stationarity and absence of autocorrelation, makes the order statistics of sequences \widehat{W}_t good instruments to estimate the quantiles of the unobservable innovations ε . In order to obtain a smoothed finite sample estimate of the innovations quantiles, we also benefit from the availability of multiple CWN sequences, attainable by simulating multiple GWN processes $Y_{t:T}$. For $i = 1, 2, \dots, M$, let $W_t^{(i)}$ be a CWN sequence resulting from performing the optimization described in (7) using a simulated GWN process $Y_{t:T}^{(i)}$, and let $W_{(t)}^{(i)}$ be the corresponding order statistics. Then a bootstrap estimator for the quantiles of innovations ε is given by

$$\hat{\varepsilon}_{(t)} = \frac{1}{M} \sum_{i=1}^M \hat{\lambda}_i W_{(t)}^{(i)}, \quad (10)$$

where $\hat{\lambda}_i = 1/\hat{\sigma}_{W_t}$ is the inverse of the estimated standard deviation for the stationary sequence $W_t^{(i)}$. The empirical CDF for the innovations can be obtained from these smoothed quantiles and will be denoted as $\widehat{\mathcal{F}}$.

4 A Non-parametric Bootstrap Stationarity Test

The variety of models proposed to represent locally stationary processes led to a variety of stationarity tests that have been proposed over the last decades. Priestley and Subba Rao [11] proposed a two-way ANOVA test to detect significant

spectral variation with respect to time and frequency, whereas Starica and Granger [14] proposed a test evaluating departures of the periodogram from a parametric spectrum over different intervals. Several tests have been proposed more recently, see Dahlhaus [4] and references therein for a review. Cardinali and Nason [1] proposed a parametric multiscale bootstrap testing approach based on the representation from Eq. (1), which however was restricted to Gaussian innovations. In this section, we generalize this latter methodology by proposing a bootstrap algorithm which is based on the non-parametric CDF estimation illustrated in the previous section. Our stationarity test is based on the following hypotheses:

H_o (stationarity) : $S_j(z)$ is a constant function of z , for all j ,

and

H_a (local stationarity) : $S_j(z)$ is not a constant function of z , for some j ,

and for $z \in [0, 1)$. In order to assess these hypotheses, we will consider the test statistic

$$\hat{\rho}_T = T^{-1} \sum_j^{J_T} \sum_k (\tilde{L}_{j,k} - \hat{L}_j)^2, \quad (11)$$

where \hat{L}_j is the average smoothed periodogram at scale j . This statistics was also used in Cardinali and Nason [1] along with a parametric bootstrap resampling scheme. We now propose its use in association with a non-parametric resampling scheme, and the whole algorithm for the testing procedure is as follows:

1. Estimate the multiscale spectra and the average spectra with the smoothed unbiased periodogram $\{\tilde{L}_{j,k}\}_{j,k}$ and its time averages $\{\hat{L}_j\}_j$, respectively. Calculate the test statistic $\hat{\rho}_T$ as defined in Eq. (11).
2. Use CWN processes to approximate the CDF of innovations by means of $\hat{\mathcal{F}}$, as illustrated in Sect. 3.
3. Simulate innovations $\varepsilon^* \sim \hat{\mathcal{F}}$ (see previous section). Then, for $n = 1, 2, \dots, N$, simulate stationary sequences $X_{t,n}^*$ of pseudodata from Eq. (1), where for each scale j we replace the amplitudes $W_j(k/T)$ with the square root of the average periodogram \hat{L}_j .
4. Calculate the smoothed periodogram $\tilde{L}_{j,k}^*$ from the series $X_{t,n}^*$, then calculate $\tilde{L}_{j,k}^{**} = \nu \tilde{L}_{j,k}^*$, where $\nu = \text{sd}(\tilde{L}_{j,k})/\text{sd}(\tilde{L}_{j,k}^*)$.
5. For each simulated stationary sequence $X_{t,n}^*$, calculate the test statistics $\tilde{\rho}_{T,n}^{**} = T^{-1} \sum_j^{J_T} \sum_k (\tilde{L}_{j,k}^{**} - \hat{L}_j^{**})^2$, where $\tilde{L}_{j,k}^{**}$ was defined in the previous step, and \hat{L}_j^{**} is its average for scale j .
6. Use the simulated statistics to calculate the empirical p-values

$$p_N = N^{-1} \sum_{n=1}^N \mathbb{I}(\tilde{\rho}_{T,n}^{**} > \hat{\rho}_T).$$

7. Reject the null hypothesis if $p_N < \alpha$, for a chosen significance level $\alpha\%$.

Remark 2 The periodogram obtained from simulated series has typically larger variance than the periodogram from original data. This variance bias needs to be tackled in order to compare the two types of periodograms and calculate meaningful empirical p-values. This is addressed in step 4 above by rescaling the simulated periodogram by ν , the ratio of standard deviations for the periodograms from data and simulations, respectively.

5 Simulation Studies

In this section, we conduct a simulation study to assess the finite sample properties of our local stationarity test. Computations of our test are conducted by using the *least asymmetric* wavelet with support $\mathcal{L}_1 = 8$, see Daubechies [5] for a discussion of their mathematical foundations. In order to estimate \mathcal{F} , we simulate 10 realizations of CWN processes by estimating functions a_t and b_t with $C = 3$ changepoints. As in Cardinali and Nason [2], the representation of these functions is in terms of C Haar coefficients. In this and the next section, we consider dyadic sample sizes to maximize computational efficiency. In order to evaluate the empirical p-values, for each test we compute 200 bootstrap samples under the null hypothesis. We evaluate both empirical size and power by simulating one stationary and one locally stationary process of size $T = 64, 128, 256, 512$, then look at the empirical rejection rates at 5% significance levels. For each sample size, we simulate 500 realizations and for each proposed model we consider both Gaussian and Student- t distributed innovations with five degrees of freedom. We initially evaluate the empirical rejection rate under the null hypothesis by simulating the stationary process:

- Model 1. AR(2): $X_t = 0.75X_{t-1} - 0.4X_{t-2} + \epsilon_t$.

We then evaluate the local power by applying our test to simulations of a locally stationary process:

- Model 2. Autoregressive model with innovations having smooth time-varying volatility:

$$X_{t:T} = 0.8X_{t-1} + \sigma(t/T)\epsilon_t;$$

with $\epsilon_t \sim i.i.d.(0, 1)$, where $\sigma(t/T) \rightarrow \sigma(z)$ and

$$\sigma(z) = \frac{3}{2} + \sin(2\pi z) + 0.3 \cos(2\pi z).$$

The results of our simulations are displayed in Table 1, where we reported results for both Gaussian (top section) and Student- t (bottom section) innovations. Looking at the results for Gaussian simulations, overall our test shows a well-controlled empirical size for all sample sizes (Model 1). The empirical rejection rate for the

Table 1 Empirical rejection rates for stationarity tests: the top section refers to Gaussian innovations. The bottom section refers to Student- t innovations (5 *d.o.f.*)

$d\mathcal{F}$	T	Model 1	Model 2
$N(0, 1)$	64	0.01	0.86
	128	0.00	0.95
	256	0.00	0.97
	512	0.00	1.00
$t(5)$	64	0.01	0.71
	128	0.00	0.88
	256	0.00	0.95
	512	0.00	0.99

local alternative represented by Model 2 is also very large, even for moderate sample sizes, and rapidly growing toward 1. Looking at the results for Student- t simulations, overall our test still shows a well-controlled empirical size for all sample sizes (Model 1). The rejection rate for Model 2 is still very large for sample sizes equal or larger than 128. As in the Gaussian case, the rejection rate grows rapidly for larger samples. Looking at the overall results, we have evidence that the test is very effective for both distributions considered in our experiment.

6 Stationarity Testing of Equity Returns

We applied our test to a time series of log-returns for the ProShares UltraShort Bloomberg Natural Gas Exchange Traded Fund. This Fund typically invests in Natural Gas Future Contracts but can switch to other types of derivatives (e.g., swaps) in case of market disruptions. We considered a series of 512 returns covering the period August 20, 2020–September 7, 2022. In order to locate economic shocks, we conducted our test on rolling windows of size 128. We choose this length since simulations showed that our test has very good power for this and larger sample sizes. The results of our sequential test are displayed in Fig. 1. The red circles correspond to rejections of the null hypothesis and are located at the center of the corresponding window. Their occurrence characterizes the presence of shocks causing significant changes in the underlying market. We mainly noted two of such periods, a shorter one during spring 2021 and a longer one corresponding to summer 2021. Both periods seem to be caused by COVID and subsequent financial and logistic disruptions. The non-stationarities found during summer 2021 seem to mark the end of a relatively low volatility state and the beginning of a high volatility state that lasts to these days. This change can be associated to the sharp increase of gas prices occurred in the

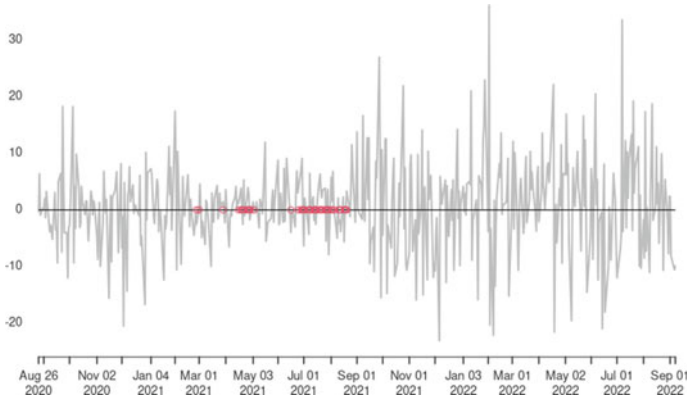


Fig. 1 Sequential stationarity testing using moving windows of size 128. The vertical left axis refers to returns value for quotes of the ProShares UltraShort Bloomberg Natural Gas Exchange Traded Fund. The red circles represent rejection of the stationarity hypothesis over the corresponding window

second half of summer 2021. The underlying reasons that determined the beginning of this turbulence could be the insufficient supply of natural gas, perhaps due to logistic troubles following the pandemic, or due to a maximal level of supply that was already reached with the infrastructures in place at that time.

References

1. Cardinali, A., Nason, G.P.: Costationarity of locally stationary time series. *J. Time Ser. Econom.* **2**(1), Article 1 (2010)
2. Cardinali, A., Nason, G.P.: Costationarity of locally stationary time series using costat. *J. Stat. Softw.* **55**, 1 (2013)
3. Dahlhaus, R.: Fitting time series models to nonstationary processes. *Ann. Stat.* **25**, 1–37 (1997)
4. Dahlhaus, R.: Locally stationary processes. In: *Handbook of Statistics*, vol. 30, pp. 351–413 (2012)
5. Daubechies, I.: *Ten Lectures on Wavelets*. Philadelphia, SIAM (1992)
6. Davidson, A.: *Stochastic Limit Theory*. Oxford University Press, Oxford (1994)
7. Fryzlewicz, P., Van Belleghem, S., von Sachs, R.: Forecasting non-stationary time series by wavelet process modelling. *Ann. Inst. Statist. Math.* **55**, 737–764 (2003)
8. Hansen, L.P.: Large sample properties of generalized methods of moments estimators. *Econometrica* **50**, 1029–1054 (1982)
9. Nason, G., von Sachs, R., Kroisandt, G.: Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Stat. Soc. B* **62**, 271–292 (2000)
10. Newey, W., McFadden, D.: Large sample estimation and hypothesis testing. In: *Handbook of Econometrics*, pp. 2111–2245. Elsevier Science (1994)
11. Priestley, M., Subba Rao, T.: A test for non-stationarity of time-series. *J. R. Stat. Soc. B* **31**, 140–149 (1969)
12. Priestley, M.: *Spectral Analysis and Time Series*. Academic Press (1983)
13. Priestley, M.: *Non-linear and Non-stationary Time Series Analysis*. Academic Press (1988)
14. Starica, C., Granger, C.: Nonstationarities in stock returns. *Rev. Econ. Stat.* **87**, 503–522 (2005)

Econometrics

The Effects of Healthy and Sustainable Transportation, Commerce, and Spillover on Airbnb Performance



Jorge Chica-Olmo  and Ruggero Sainaghi 

Abstract This study explores the effects of location variables on Airbnb listing performance. In particular, three groups of antecedents are explored: healthy and sustainable transportation systems (metro, tourist bus and bike-sharing), commerce (food and non-food establishments and shop centers) and spillover (nearby Airbnb listings). The study includes five groups of control variables, articulated in no-location variables (size, contractual terms, rules, host and guest) and one widely used location antecedent (distance to the city center). Three hypotheses are tested, all supposing a positive effect on listing revenue generated by the three additional groups of location variables (transport, commerce and spillover). The empirical study is carried out in the city of Milan (Italy), a complex destination able to attract (before the COVID-19 outbreak) different targets with an important focus on fashion and design. Four models are proposed, and empirical findings support the hypotheses. Interestingly, food shops are not significant, while the non-food and commercial centers positively affect the revenues.

Keywords Airbnb · Listing performance · RevPAR · Healthy and sustainable transportation · Commerce · Spillover effect

1 Introduction

The start-up of Airbnb and more generally of other commercial peer-to-peer accommodation platforms (P2P APs) has progressively changed the tourism industry. A plethora of studies were recently published, and some literature reviews have identified many areas of inquires [28, 32, 33].

J. Chica-Olmo
University of Granada, 18011 Granada, Spain
e-mail: jchica@ugr.es

R. Sainaghi (✉)
IULM University, 20143 Milan, Italy
e-mail: ruggero.sainaghi@iulm.it

The worldwide diffusion of peer-to-peer accommodation platforms, and especially of listings, attracted the interest of researchers to understand the determinants of host price and revenue [26]. This is not surprising, considering that the main motivation to rent an apartment in a peer-to-peer accommodation platform is to earn an additional income, followed by the so-called social features [14]. This research stream focuses on measuring the performance determinants, using the price, revenue or more than one performance indicator as a dependent variable. The current results clearly suggest that the listing antecedents are significantly different from the hotel determinants, considering the deep differences in the business models.

In this work, we consider that the revenue of Airbnb accommodation depends on its characteristics. A first wave of listing performance papers has identified some promising determinants, focusing on the listing characteristics, such as listing type, amenities, contractual terms and size. Other promising antecedents include the host, guest reviews, location patterns and destination characteristics. A second wave of studies has shown some more complex patterns, where some variables are able to moderate the results [45], and some relationships change their effects in different research contexts [40]. Furthermore, some counter-intuitive findings—as the effect generated by professional hosts and by the number of guests' reviews on performance—show different signs moving from price to occupancy rate or revenue per available room (RevPAR) [22]. The large majority of papers is based on ordinary least squares (OLS) hedonic pricing models [27]. However, some new methodologies—such as spatial econometrics models [9], geographical weighted regression [49], quantile regression method [19], multilevel model [10] and machine learning frameworks [5]—appear to be more efficient. The second generation of studies, paradoxically, has opened many new questions and has shown the paucity of current knowledge, asking for new theoretical and empirical studies. Location, in particular, appears a promising determinants and new models are necessary to evaluate the complexity of listing locational patterns. This is the gap that this paper contributes to fill and is epitomized in the following research questions (later discussed in the Literature Review).

Research question 1. What is the effect generated by location on listing performance?

Research question 2. What are the effects generated by three locational variables (healthy and sustainable transportation systems; commerce; spillover) on Airbnb listing performance?

Research question 3. How can the listing performance be measured?

2 Literature Review

As an old adage goes, there are three main variables that influence hospitality: location, location and location. However, researchers analyzing the performance of Airbnb listings have used less intensive location patterns than in the hospitality

field. In this study, the following locational variables are proposed: accessibility, transportation, commerce and spillover.

Previous studies have focused more on the so-called *accessibility* of points of transportation or to other local attractions, mainly measured by the relative distance from the city center [10, 21, 24, 40, 49]. These variables usually show a negative correlation with listing performance. Therefore, the higher the distance to the local attractors, the lower the performance of the listing. Inside this wide group of antecedents, distance from the city center is usually the most significant variable and was therefore used in this study [21].

Considering the focus of the second research question on healthy and sustainable transportation systems, commerce, and spillover, following the effects of these antecedents are succinctly discussed. The variable of *transportation* is usually measured by the distance from a destination hub [37] and, in particular, metro areas [10]. The relationships between transportation variables and Airbnb listing results are usually negative, suggesting that the higher the distance, the lower the listing performance. In this connection, a study revealed that the ability to be located near a transportation hub increased Airbnb listing results [7]. By contrast, however, another study showed the absence of a significant correlation [21]. In this study, a positive correlation was expected between healthy and sustainable transportation and Airbnb performance. Airbnb guests, thanks to the transportation hub, can obtain rapid access to a destination's attractions. Another study confirms that public transport is the most frequently used mode of transportation for tourists who visit small areas, and the authors have found that the tourists' level of education and the price of public transportation are significant variables relating to the choice of transportation mode [18]. While transportation is usually measured by focusing on metro stations, in this study two additional variables are considered. The first is tourist buses. Many guests use this service to get an overview of a destination's attractions [21]. The second is represented by bike-sharing systems. This mode of transportation combines several advantages: it is a sustainable way to get around in a city and can be cheaper than other alternatives. Unsurprisingly, 73% of tourists claim to like or be very favorable to the use of bike-sharing [13]. Therefore, this transportation system is of great interest to a large percentage of tourists who are likely to be interested in selecting accommodation close to this service. Consequently, the second research question explores the ability of healthy and sustainable transportation systems to influence Airbnb listing performance.

Commerce is the second location variable considered. Shopping is an important activity involving many guests in all destinations around the world [11]. Examples of shopping activities include the purchase of fashion products, gifts, food and souvenirs [16, 17]. This study explores two different variables related to commerce. The first focuses on food and beverages. Hotel guests do not have kitchen facilities, but Airbnb listings are usually equipped with a kitchen. Previous studies revealed that the presence of a kitchen does not improve Airbnb listing performance. Guests can simply buy food and beverages and choose apartments located closer to centers of commerce. The second sub-variable related to commerce includes non-food and beverage products. This variable is expected to be relevant for many tourist destinations, and, in

particular, for the city of Milan, which is considered as an icon for design products and fashion goods worldwide. Usually, non-food shops are clustered in blocks, as in specific streets famous for design, artisanal products and similar goods. There are only a handful of previous studies exploring the non-food and beverage variables, and their results are partially contradictory [3, 24, 46].

The third locational factor is *spillover*. Focusing on listing price (the object of this paper), the spillover effect is the ability of neighboring apartments to influence the rate of another listing. This agrees with the idea that near-spatial data are related [39]. Some authors have used the word “contagion” to describe this process [9]. An updated literature review analyzed the tourism studies that have explored the spillover effect [4]. The articles cited by these authors are mainly focused on tourism demand. In this paper, the attention is on spatial spillover that can be defined in terms of economic externalities “which produce non-compensated or indirect impacts for a receiver situated nearby” [50]. As anticipated, previous studies are mainly focused on tourism demand and explore the spatial spillover effects among nearby regions [47]. The spillover effect can be positive or negative. The positive effect of spillover identifies complementary interactions among neighbors, while the negative effect suggests the presence of direct competition. Moving from the general field of tourism demand to Airbnb listings, the spatial spillover effect has rarely been explored. Some authors have found a positive effect on listing prices generated by their neighbors (Airbnb apartments) [9]. Another study investigated the agglomeration effect in New York City [43]. The study shows a positive agglomeration effect on Airbnb listing revenue, reinforced by host tenure and mitigated by host capacity.

Based on the extensive analysis of previous literature, we have classified the relevant variables into two groups: location and non-location variables. Within location variables, we are interested in transport, commercial and spillover effects on RevPAR. The current study tests the following hypotheses, focusing on location variables:

- H1. The accessibility of sustainable and healthy transportation positively affects the revenue per available room of Airbnb listings.
- H2. The spatial distribution of commercial shops positively affects the revenue per available room of Airbnb listings.
- H3. The spillover effect positively influences the revenue per available room of Airbnb listings.

3 Materials

3.1 Study Area

This study assumes the city of Milan as a relevant destination for testing the proposed hypotheses. Two main reasons support this decision. First, Milan is located in Italy, an important worldwide country for tourism, well epitomized by the World Tourism Organization (UNWTO) that ranks this country (before the COVID-19 outbreak)

fifth in terms of arrivals and sixth for receipts. Second, Milan is a complex destination where different targets of guests are attracted and mixed. While other Italian cities such as Rome, Venice and Florence center around leisure guests, Milan is more of a business and a meeting, incentive, congress, exhibition (MICE) destination [30]. Milan is the economic capital of Italy, where the stock market is located. Milan is also the second largest trade-fair center in Europe (Sainaghi R., Mauri, Ivanov, & d'Angella, 2018). At the same time, Milan embraces an extraordinary cultural heritage, including, among others, the Duomo, Leonardo da Vinci's *Last Supper* and many famous museums (Brera, Pinacoteca Ambrosiana, Museo del Novecento, etc.). The brand and image of the city were strengthened by the Milan Expo, hosted in 2015. This event was able to improve hotel performance and reduce tourism seasonality [29, 34].

3.2 Data and Variables

This study is mainly based on AirDNA data. AirDNA uses information taken from the web related to Airbnb listings, reporting a rich set of information for each rental (such as host and guest information, size of listings and other contractual characteristics). AirDNA data has been used in many previous studies. In this study, pre-pandemic data was used. During the COVID-19 outbreak, Milanese Airbnb reservations reduced considerably. A second level of data is represented by locational factors that were measured using open data published by the Municipality of Milan.

Focusing on locational variables, a geographic information system (GIS) was downloaded. Spatial density ratios were calculated based on GIS data using the approaches developed in previous papers [36]. In particular, the density ratio was used for bike-share parking and commercial shops (food and beverage commerce as well as non-food shops). The search radius per unit area (km^2) of 200 m was used to obtain indexes for shared parking. A search radius of 500 m per km^2 was used for commercial establishments (food and non-food). In addition, to determine the spatial effect of big commercial establishments, we have considered the number of shopping centers in each district. The number of apartments used was 10,922. Table 1 shows the statistics of variables and their descriptions.

In this paper, we try to explain the RevPAR. Figure 1 shows the spatial distribution of the apartments and their RevPAR. The highest values are in the city center ("Centro Storico"). To control variables and location, this study has used two groups of variables: location and non-location. Within non-location variables, we have considered five groups of determinants: size, contractual terms, rules, host and guest. Among the location variables are spatial trend, transportation, commerce and spillover. Each group is succinctly discussed. With respect to non-location characteristics, size is operationalized using three variables, the number of bedrooms (Bedrooms), bathrooms (Bathrooms) and the maximum number of guests (MaxGuests). Previous studies show a positive and strong impact of size variables on price [6]. Generally speaking, the sample is mainly composed of small listings,

Table 1 Descriptive statistics and description of variables—N = 10,922 apartments

Variable	Mean	Min	Max	Std. Dev.	Description
RevPAR	60.283	10.028	570.372	36.806	Revenue per Available Room (€)
<i>No-location factors</i>					
<i>Size</i>					
Bedrooms	1.302	1	10	0.596	Number of bedrooms
Bathrooms	1.160	0	8	0.426	Number of bathrooms
MaxGuests	3.525	1	16	1.508	Number of maximum guests
<i>Contractual terms</i>					
CancePolic	0.556	0	1	0.497	1 if apt. has cancelation policy, 0 otherwise
Security	1.016	0	45	2.379	Security deposit (in hundreds of €)
Cleaning	39.197	0	116	28.704	Cleaning fee (€)
ExtPeopFee	0.133	0	2.84	0.172	Extra people fee (in hundreds of €)
Checkin	0.0001	0	1	0.010	1 if check-in is after 19:00, 0 otherwise
Checkout	0.424	0	1	0.494	1 if check-out is 11:00, 0 otherwise
MinimumStay	2.836	1	365	10.390	Minimum stay (days)
<i>Rules</i>					
Pets	0.207	0	1	0.405	1 if pets are allowed, 0 otherwise
<i>Host</i>					
Superhost	0.222	0	1	0.416	1 if the apt. is superhost, 0 otherwise
ResponR	91.667	0	100	21.597	Response rate
ResponT	0.625	0	1	0.484	1 if response time is within an hour
NPhotos	21.039	1	200	12.246	Number of photos
Experience	2.249	0	9.469	1.810	Number of years of the apartment in database
<i>Guest</i>					
Nreviews	0.374	0	7.9	0.629	Hundreds of reviews
OverRat	4.653	1	5	0.432	Overall rating
<i>Location factors</i>					
<i>Spatial trend</i>					
DtCityCenter	4.056	0.099	14.133	2.107	Distance from Duomo (km)
<i>Transport</i>					
DtMetro	0.813	0.009	5.584	0.632	Distance from nearest metro station (km)
DtTuribus	1.721	0.024	7.128	1.160	Distance from nearest tourist bus stop (km)

(continued)

Table 1 (continued)

Variable	Mean	Min	Max	Std. Dev.	Description
DBikePark	2.800	0.000	11.459	2.332	Share-bike parking density (index)
<i>Commerce</i>					
DFood	0.0002	0	0.0021	0.0002	Medium-large commercial food establishments density (index)
DNon-food	0.001	0	0.026	0.002	Medium-large commercial non-food establishments density (index)
Shop Centers	2.530	1	5	1.441	Number of shopping centers into district

considering that the average number of bedrooms is 1.3. Contractual terms include a group of antecedents regulating some technical conditions. Strict cancellation policy (CancePolic), security deposit fee (Security), cleaning fee (Cleaning) and extra people fee (ExtPeopFee) usually affect the rates positively [20, 42]. The only variable listed under the third block (rules) considers whether pets are allowed or not (Pets). Roughly 20% of samples offer this service. In another study [9], this variable has a negative effect on price. The host group focuses on five relevant variables. Approximately, 22% of hosts are classified as superhosts (Superhost). The response rate of superhosts (ResponR) is very high (91.6%), and 62.5% respond within an hour (ResponT). The average number of published photos (NPhotos) is roughly 21. The average experience or membership (Experience) is 820 days (2.2 years). Previous studies have shown positive effects generated by these antecedents on dependent variables [2, 31, 40]. The guest group focuses on the number of reviews (Nreviews) and overall ratings (OverRat). While this latter antecedent positively affects the rates [3], the number of reviews primarily has a negative effect on price [42], but a positive effect on occupancy [44] and revenue [10] and RevPAR [43]. In this work, all of the above variables are control variables.

Regarding location factors, the distance to the city center has traditionally been considered [21], and this variable represents the spatial drift or trend [8]. Here, this variable is considered as a control variable. However, transport, commerce and revenue of neighboring apartments have not been widely considered in the literature, and for this they are considered as interest factors. Transport mode of choice is an important aspect of tourism research [18], and Airbnb guests are likely to be willing to pay more money for better accessibility to places of interest. Hence, apartment prices will be higher in locations close to transportation hubs. To contrast the hypothesis (H1) that RevPAR depends on the accessibility of sustainable and healthy transportation, we have considered three variables: the Euclidean distance from the nearest metro station (DtMetro), the Euclidean distance from the nearest tourist bus stop (DtTuribus) and bike-share parking density (DdBikePark). The distance to the nearest subway station from a rental has been included in the hedonic hotel price models [48], but not much in the Airbnb hedonic models. Accessibility to tourist bus routes can be an antecedent of Airbnb listing results and is a key locational

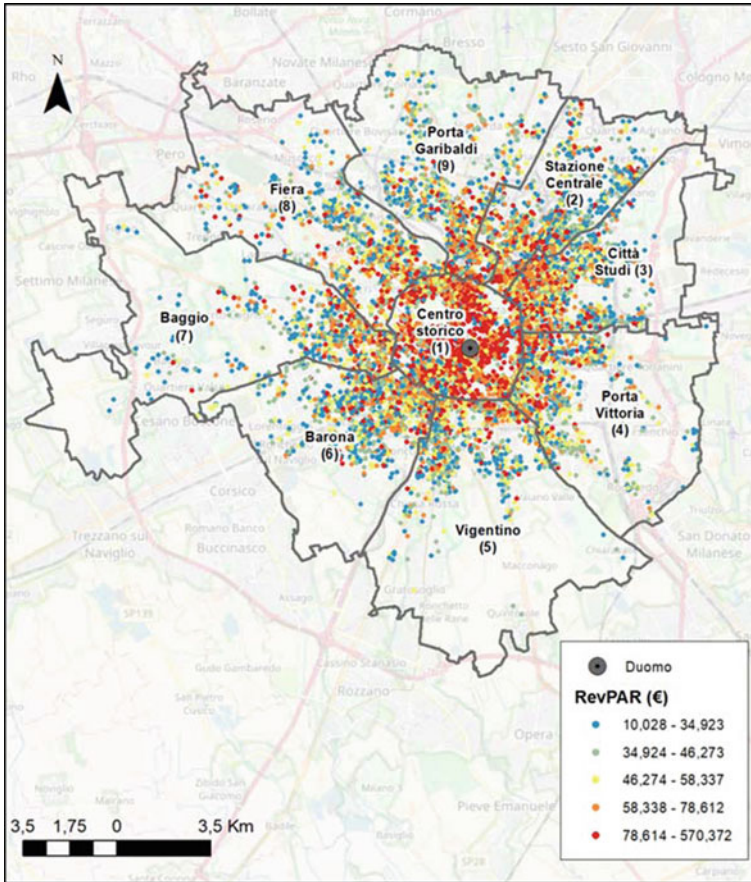


Fig. 1 Milan’s districts, city center and Revenue per Available Room (RevPAR) in euros

asset to hosts and guests [41]. Bike-sharing is also a potential determinant [13], and therefore the concentration of bike-share parking can influence the revenue of Airbnb apartments. A second map (not reported in this study because of space limitations) shows the spatial distribution of transport variables. The map suggests a strong concentration of the three analyzed transport variables in the city center, while the connections with the peripheral Milanese blocks are mainly covered by metro lines. Transport depicts a clear center/peripheral distribution. We assume that the sign of the coefficients of the distance variables ($Dt_{CityCenter}$, Dt_{Metro} and $Dt_{Turibus}$) is negative because the revenue is inversely related to the distance of these points of interest, in order to reduce the friction cost. On the other hand, we expect that the areas where there is a greater abundance (density) of bike-sharing stations will have higher revenue because the tourist demand will be greater.

The commerce factors include three independent variables. The first two measure the density of medium to large commercial food and non-food establishments. The third measures the number of shopping centers in given districts. As previously explained in the literature reviews, these determinants were used less in previous studies. We expect a positive relationship to emerge between the dependent variable for non-food shops and shopping centers.

4 Methods

This paper adopts a semi-log hedonic model to explain the dependent variable, represented by the log of RevPAR [46]. In this case, the coefficients of continuous explanatory variables multiplied by 100 represent the percentage impact of these variables on RevPAR. However, when the explanatory variable is binary, this impact is obtained with $[\exp(b) - 1] * 100$ [12]. Ordinary least squares (OLS) is the classic model used in similar papers. Some authors, however, suggest caution, given the possible presence of perturbances and autocorrelations [1]. The inclusion in this study of spillover requires the integration of the OLS with some spatial models. In particular, a first spatial autoregressive (SAR) model was considered (see the following formula 1), and a second spatial error model (SEM) with spatial nuisance dependence was adopted (see the second following formula):

$$SAR : y = rWy + xb + e \quad (1)$$

$$SEM : y = Xb + u, \text{ with } : u = / Wu + e \quad (2)$$

The variable y is the dependent variable and, as anticipated, is the natural logarithm of RevPAR (lnRevPAR). X includes the independent variables (those listed in Table 1), while β are the associated parameters. The neighboring listings are summarized in the W variable, which includes a row-normalized spatial weights matrix. Wy ($W\ln\text{RevPAR}$) and r represent the spatially lagged dependent variable and global spatial spillover effect, respectively; Wu and $/$ represent the spatially lagged disturbances and their associated coefficient, respectively, and e represents a normal iid vector of disturbances.

To evaluate these models, the maximum likelihood (ML) is employed as a consistent method [1]. In this study, however, the perturbances show heteroscedasticity. In this case, the two-stage least squares (2SLS) is adequate [15]. Moran's I test was used to quantify the errors, considering the possible presence of spatial autocorrelation in the disturbances. Finally, to choose between the appropriateness of the two spatial models (SAR versus SEM), the Lagrange multiplier (LM) tests (LM-error and LM-lag) and the robust versions (RLM-error and RLM-lag) of the OLS models were calculated.

Focusing on the SAR model, the coefficients cannot be directly interpretable. Because of this, the direct and indirect effects of the explanatory variables were calculated. In particular, the indirect effects represent the spatial spillover effects.

5 Findings

Table 2 depicts four nested classical models estimated using OLS and two SAR models estimated with 2SLS. The first model (Mod1.ols) includes non-location traditional independent variables largely used in previous papers. The second model (Mod2.ols) annexes the location classical variable (distance to city center). The third model (Mod3.ols) adds the transport variables and the fourth model (Mod4.ols) appends commerce variables. The spatial models Mod4.2sls with W1 and Mod4.2sls with W2 include the spatial spillover effects with two different weight matrices, inverse distance and k-nearest, respectively.

In Table 2, Models adjusted R-squared for a nested model with each group of variables (\bar{R}_j^2) is shown. The Size group of variables is shown as the one with the greatest explanatory power. The rest of the non-location variables increase the explanatory capacity of the model by 33% $((0.294-0.221)/0.221)$, while the location variables, with respect to the non-location ones, allow to increase in the explanatory capacity by 43% $((0.421-0.294)/0.294)$. The distance to the city center is, without a doubt, the location variable with the greatest explanatory capacity. It is possible that this happens because the main tourist attraction is concentrated in the center of the city of Milan. This polarizing effect of the city center over the whole city is named spatial trend. However, this effect does not consider changes within each area caused by other location variables, such as accessibility to means of transport and commerce, that could cause a modulation in the revenue of the apartments.

The results shown in Table 2 allow us to contrast the established hypotheses. Thus, all coefficients of the transport variables (DtMetro, DtTuribus and DBikePark) are significant and have the expected sign. Therefore, Airbnb apartments that are located in areas close to sustainable means of transport, such as metro stations and tourist bus stops, or where there is a high density of shared bike parks (which is a healthy means of transport), have high revenue from available rooms (H1). This may be an indicator that there is a greater demand for apartments in these areas due to their proximity to these means of transport.

On the other hand, H2 has been contrasted since, as shown in Table 2, a positive and significant relationship has been detected both between RevPAR and the number of shopping centers in the district where the apartment is located (ShopCenters), as with the density of commercial non-food establishments in the area (DNon-Food). This may be due to the fact that tourists prefer to stay near commercial establishments related to shopping (fashion, gift, etc.). However, the proximity of food stores has not been detected to have a significant effect on RevPAR.

Table 2 OLS and 2SLS models. Dependent variable: lnRevPAR

Variable	Mod1		Mod2		Mod3		Mod4		Mod4.2sls (W1)		Mod4.2sls (W2)	
	Direct	Indir	Direct	Indir	Direct	Indir	Direct	Indir	Direct	Indir	Direct	Indir
Constant	2.743***	-	3.103***	-	2.985***	-	2.897***	-	-	-	-	-
<i>No-location factors</i>												
(1) Size							$\bar{R}_1^2 = 0.221$					
Bedrooms	0.110***		0.120***		0.112***		0.119***	0.006	0.118***	0.000	0.119***	0.065***
Bathrooms	0.239***		0.198***		0.197***		0.190***	0.010	0.189***	0.010	0.186***	0.102***
MaxGuests	0.057***		0.063***		0.062***		0.062***	0.003	0.061***	0.003	0.061***	0.033***
<i>Contractual terms</i>												
							$\bar{R}_2^2 = 0.257$					
CancePolic	0.034***		0.029***		0.029***		0.032***	0.002	0.031***	0.000	0.031***	0.017***
Security	0.009***		0.006***		0.007***		0.007***	0.000	0.006***	0.000	0.006***	0.003***
Cleaning	0.002***		0.001***		0.001***		0.001***	0.00001	0.001***	0.00001	0.001***	0.001***
ExtPeopFee	-0.016***		-0.075***		-0.078***		-0.083***	-0.004	-0.082***	-0.004	-0.079***	-0.043***
Checkin	1.995***		2.013***		2.009***		2.022***	0.104	2.031***	0.104	2.059***	1.130***
Checkoutb	0.106***		0.078***		0.073***		0.069***	0.004	0.069***	0.004	0.067***	0.037***
MinimuStay	0.001***		0.001***		0.001***		0.001***	0.00001	0.001***	0.00001	0.001***	0.001***
<i>Rules</i>												
							$\bar{R}_3^2 = 0.258$					
Pets	-0.0037		0.002		0.006		0.006	0.0003	0.006	0.0003	0.007	0.004
<i>Host</i>												
							$\bar{R}_4^2 = 0.285$					
Superhost	0.057***		0.066***		0.066***		0.066***	0.003	0.066***	0.003	0.066***	0.036***
ResponR	0.002***		0.001***		0.001***		0.001***	0.0001	0.001***	0.0001	0.001***	0.001***
ResponT	0.037***		0.036***		0.032***		0.031***	0.002	0.031***	0.002	0.031***	0.017***

(continued)

Table 2 (continued)

Variable	Mod1		Mod2		Mod3		Mod4		Mod4.2sls (W1)		Mod4.2sls (W2)	
	Direct	Indir	Direct	Indir	Direct	Indir	Direct	Indir	Direct	Indir	Direct	Indir
NPhotos	0.003***	0.002***	0.002***	0.002***	0.002***	0.0001	0.002***	0.0001	0.002***	0.0001	0.002***	0.001***
Experience	-0.015***	-0.014***	-0.013***	-0.013***	-0.013***	-0.001	-0.014***	-0.001	-0.013***	-0.001	-0.013***	-0.007***
<i>Guest</i>	$\bar{R}_5^2 = 0.294$											
Nreviews	0.084***	0.061***	0.055***	0.055***	0.055***	0.003	0.052***	0.003	0.052***	0.003	0.051***	0.028***
OverRat	0.045***	0.056***	0.058***	0.058***	0.058***	0.003	0.057***	0.003	0.057***	0.003	0.058***	0.031***
<i>Location factors</i>												
<i>Spatial trend</i>	$\bar{R}_6^2 = 0.396$											
DtCityCenter	-	-0.080***	-0.046***	-0.046***	-0.046***	-0.002	-0.032***	-0.002	-0.032***	-0.002	-0.021***	-0.012***
<i>Transport</i>	$\bar{R}_7^2 = 0.411$											
DtMetro	-	-	-0.063***	-0.063***	-0.063***	-0.003	-0.053***	-0.003	-0.053***	-0.003	-0.037***	-0.020***
DtTuribus	-	-	-0.022***	-0.022***	-0.022***	-0.002	-0.030***	-0.002	-0.030***	-0.002	-0.022***	-0.012***
DBikePark	-	-	0.025***	0.025***	0.025***	0.001	0.015***	0.001	0.015***	0.001	0.009***	0.005***
<i>Commerce</i>	$\bar{R}_8^2 = 0.421$											
DFood	-	-	-	-	-	0.708	-13.836	0.708	-13.836	0.708	2.041	1.120
DNonFood	-	-	-	-	-	0.762	14.896***	0.762	14.896***	0.762	8.379***	4.598***
ShopCenters	-	-	-	-	-	0.001	0.027***	0.001	0.027***	0.001	0.015***	0.008***
<i>Spillover effect</i>												
WlnRevPAR	-	-	-	-	-	-	0.050*	-	0.050*	-	0.356***	-

The H3 has also been contrasted, that is, the presence of spatial spillover effect, both globally and for the explanatory variables. These effects have been estimated considering two types of neighborhood specifications. The results (Table 2) indicate that both, global effect and the explanatory variables (indirect effects), are more significant if the k -nearest ($W2$) is considered versus the inverse of the distance ($W1$). Part of the presence of the global spillover effect could be due to the presence of spatial contagion, that is, as in real estate, apartment owners tend to imitate the prices of their neighbors, and therefore revenue will also be similar.

6 Discussion and Conclusion

This study basically confirms the relevancy of control variables (non-location antecedents) used in previous studies (presented and discussed in the literature review). Size is the most relevant determinant in terms of adjusted R^2 , while the other groups (contractual terms, host and guest), despite being highly significant and showing the expected signs, play a more marginal explanation power. Moving to locational variables, this study introduces three additional variables other than the classical distance to the city center. The three tested hypotheses are confirmed: healthy and sustainable transport systems, commerce and spillover effects significantly and positively affect the listing RevPAR.

Airbnb may be indirectly causing an increase in the carbon footprint due to the process of displacement of residents and gentrification that it could produce [10]. However, Airbnb could cause the contrary effect, this is, a reduction in pollution because guests opt for public transport due to it being cheap and accessible, which can have an impact on a reduction in the level of pollution. The rental of Airbnb apartments can also increase the demand for sustainable and healthy means of transport, which could not only have a positive effect on the environment, but could also affect the health of the tourists who use them. Moreover, this study sheds light on the relationships between Airbnb guests and city commercial offers. Thus, although accessibility to the nearest shopping center and medium-large non-food establishments has a significant and positive effect on RevPAR, this does not occur with food establishments. This is in line with the Airbnb guest profile, more interested in discovering local and authentic products [23]. The measurement of spillover effects contributes to filling an important gap in the Airbnb performance literature about competition and collaboration among hosts and more generally agglomeration and contagion effects. The two spatial econometrics models developed clearly illustrate the positive impact of spatial spillover effect on RevPAR. Finally, while previous studies were more centered on rates, this paper, in line with some recent academic articles (see Section Listing performance), suggested the higher relevance of RevPAR, able to consider the listing sales.

The new variables used in this study can orient the location choices and the communication strategies implemented by Airbnb listings with their guests. The proximity to healthy and sustainable transportation systems (metro, tourist bus and

bike sharing) can support the rise of RevPAR, as well as the proximity to non-food stores and more generally shop centers. Airbnb hosts can signal these services on their web pages. The spillover effect is supportive of agglomeration strategies. Nearby listings exercise a positive effect on RevPAR. This finding can change the perception of neighbors, moving from a competitive to a more collaborative approach. Finally, the dependent variable is more in line with hosts' focus on revenue rather than on rates. An increase in price, without a parallel rise in occupancy rate, does not improve the listing performance.

This is an explorative study based on a single case study. Despite the sample size being robust (more than 10 thousand listings), the findings must be confirmed in other different empirical settings.

References

1. Anselin, L.: *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht (1988)
2. Benítez-Aurioles, B.: Why are flexible booking policies priced negatively? *Tour. Manage.* **67**, 312–325 (2018)
3. Cai, Y., Zhou, Y., Scott, N.: Price determinants of Airbnb listings: evidence from Hong Kong. *Tour. Anal.* **24**(2), 227–242 (2019)
4. Chang, C.L., Hsu, S.H., McAleer, M.: Risk Spillovers in Returns for Chinese and International Tourists to Taiwan. *J. Travel Res.* **59**(2), 335–351 (2020)
5. Chattopadhyay, M., Mitra, S.K.: Do airbnb host listing attributes influence room pricing homogenously? *Int. J. Hosp. Manag.* **81**, 54–64 (2019)
6. Chen, Y., Xie, K.: Consumer valuation of Airbnb listings: a hedonic pricing approach. *International Journal of Contemporary Hospitality Management* **29**(9), 2405–2424 (2019).
7. Cheng, M., Jin, X.: What do Airbnb users care about? An analysis of online review comments. *Int. J. Hosp. Manag.* **76**, 58–70 (2019)
8. Chica-Olmo, J., Cano-Guervos, R., Chica-Olmo, M.: A coregionalized model to predict housing prices. *Urban Geogr.* **34**(3), 395–412 (2013)
9. Chica-Olmo, J., González-Morales, J.G., Zafra-Gómez, J.L.: Effects of location on Airbnb apartment pricing in Málaga. *Tour. Manage.* **77**, 103981 (2020)
10. Deboosere, R., Kerrigan, D.J., Wachsmuth, D., El-Geneidy, A.: Location, location and professionalization: a multilevel hedonic analysis of Airbnb listing prices and revenue. *Reg. Stud. Reg. Sci.* **6**(1), 143–156 (2019)
11. García, M.N., Muñoz-Gallego, P.A., Viglia, G., Gonzalez-Benito, O.: Be social! The impact of self-presentation on peer-to-peer accommodation revenue. *J. Travel Res.* **59**(7), 1268–1281 (2019)
12. Halvorsen, R., Palmquist, R.: The interpretation of dummy variables in semi-logarithmic equations. *American Economic Review* **70**(3), 474–475 (1980)
13. Kaplan, S., Manca, F., Nielsen, T.A., Prato, C.G.: Intentions to use bike-sharing for holiday cycling: An application of the Theory of Planned Behavior. *Tour. Manage.* **47**, 34–46 (2015)
14. Karlsson, L., Dolnicar, S.: Someone's been sleeping in my bed. *Ann. Tour. Res.* **58**, 159–162 (2016)
15. Kelejian, H.H., Prucha, I.R.: A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics* **17**(1), 99–121 (1998)
16. Keown, C., Jacobs, L., Worthley, R.: American tourists' perception of retail stores in 12 selected countries. *J. Travel Res.* **22**(3), 26–30 (1984)

17. Lehto, X.Y., Chen, S.Y., Silkes, C.: Tourist shopping style preferences. *J. Vacat. Mark.* **20**(1), 3–15 (2014)
18. Le-Klähn, D.-T., Roosen, J., Gerike, R., Hall, C.M.: Factors affecting tourists' public transport use and areas visited at destinations. *Tour. Geogr.* **17**(5), 738–757 (2015)
19. Lorde, T., Jacob, J., Weekes, Q.: Price-setting behavior in a tourism sharing economy accommodation market: a hedonic price analysis of Airbnb hosts in the Caribbean. *Tourism Management Perspectives* **30**, 251–261 (2019)
20. Moreno Izquierdo, L., Egorova, G., Peretó Rovira, A., Más Ferrando, A.: Exploring the use of artificial intelligence in price maximisation in the tourism sector: its application in the case of Airbnb in the Valencian Community. *Journal of Regional Research* **42**, 113–128 (2018)
21. Önder, I., Weismayer, C., Gunter, U.: Spatial price dependencies between the traditional accommodation sector and the sharing economy. *Tour. Econ.* **25**(8), 1150–1166 (2019)
22. Oskam, J., van der Rest, J.P., Telkamp, B.: What's mine is yours—but at what price? Dynamic pricing behavior as an indicator of Airbnb host professionalization. *Journal of Revenue and Pricing Management* **17**(5), 311–328 (2018)
23. Paulauskaite, D., Powell, R., Coca-Stefaniak, J.A., Morrison, A.M.: Living like a local: Authentic tourism experiences and the sharing economy. *Int. J. Tour. Res.* **19**(6), 619–628 (2017)
24. Perez-Sanchez, V.R., Serrano-Estrada, L., Marti, P., Mora-Garcia, R.T.: The what, where, and why of Airbnb price determinants. *Sustainability* **10**(12), 4596 (2018)
25. Ribes, J.F., Izquierdo, L., Rodríguez, A., Such Devesa, M.J.: The Rental Prices of the Apartments under the New Tourist Environment: A Hedonic Price Model Applied to the Spanish Sun-and-Beach Destinations. *Economies* **6**(23), 1–9 (2018)
26. Sainaghi, R.: The current state of academic research into peer-to-peer accommodation platforms. *Int. J. Hosp. Manag.* **89**, 102555 (2020)
27. Sainaghi, R.: Determinants of price and revenue for peer-to-peer hosts. The state of the art. *International Journal of Contemporary Hospitality Management* **33**(2), 557–586 (2020b).
28. Sainaghi, R., Baggio, R.: Clusters of topics and research designs in peer-to-peer accommodation platforms. *International Journal of Hospitality Management* **102393** (2019a).
29. Sainaghi, R., Baggio, R.: Destination Events, Stability, and Turning Points of Development. *J. Travel Res.* **60**(1), 172–183 (2019)
30. Sainaghi, R., Mauri, A.: The Milan World Expo 2015: hospitality operating performance and seasonality effects. *Int. J. Hosp. Manag.* **72**, 32–46 (2018)
31. Sainaghi, R., Abrate, G., Mauri, A.: Price and RevPAR determinants of airbnb listings: convergent and divergent evidence. *Int. J. Hosp. Manag.* **92**, 102709 (2021)
32. Sainaghi, R., Köseoglu, M.A., Mehraliyev, F.: The intellectual structure of the sharing economy. *Tour. Econ.* **27**(5), 1137–1156 (2020)
33. Sainaghi, R., Köseoglu, M.A., d'Angella, F., Mehraliyev, F.: Sharing economy: a co-citation analysis. *Curr. Issue Tour.* **23**(8), 929–937 (2019)
34. Sainaghi, R., Mauri, A., d'Angella, F.: Decomposing seasonality in an urban destination: the case of Milan. *Curr. Issue Tour.* **22**(16), 1919–1924 (2018)
35. Sainaghi, R., Mauri, A., Ivanov, S., d'Angella, F.: Mega events and seasonality: The case of the Milan World Expo 2015. *Int. J. Contemp. Hosp. Manag.* **31**(1), 61–86 (2019)
36. Silverman, B. W.: Density estimation for statistics and data analysis. Routledge (2018).
37. Soler, I.P., Gemar, G.: Hedonic price models with geographically weighted regression: An application to hospitality. *Journal of Destination Marketing Management* **9**, 126–137 (2018)
38. Tang, L.R., Kim, J., Wang, X.: Estimating spatial effects on peer-to-peer accommodation prices: Towards an innovative hedonic model approach. *Int. J. Hosp. Manag.* **81**, 43–53 (2019)
39. Tobler, W.R.: A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* **46**(1), 234–240 (1970)
40. Tong, B., Gunter, U.: Hedonic pricing and the sharing economy: how profile characteristics affect Airbnb accommodation prices in Barcelona, Madrid, and Seville. *Current Issues in Tourism* <https://doi.org/10.1080/13683500.2020.1718619> (2020)

41. Visser, G., Erasmus, I., Miller, M.: Airbnb: The emergence of a new accommodation type in Cape Town. South Africa. *Tourism Review International* **21**(2), 151–168 (2017)
42. Wang, D., Nicolau, J. L.: Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb. com. *International Journal of Hospitality Management* **62**, 120–131 (2017)
43. Xie, K.L., Kwok, L., Heo, C.Y.: Are Neighbors Friends or Foes? Assessing Airbnb Listings' Agglomeration Effect in New York City. *Cornell Hospitality Quarterly* **61**(2), 128–141 (2019)
44. Xie, K., Mao, Z.: The impacts of quality and quantity attributes of Airbnb hosts on listing performance. *Int. J. Contemp. Hosp. Manag.* **29**(9), 2240–2260 (2017)
45. Xie, K., Mao, Z.E., Wu, J.: Learning from peers: The effect of sales history disclosure on peer-to-peer short-term rental purchases. *Int. J. Hosp. Manag.* **76**, 173–183 (2019)
46. Yang, Y., Mao, Z.: Location advantages of lodging properties: A comparison between hotels and Airbnb units in an urban environment. *Ann. Tour. Res.* **81**, 102861 (2020)
47. Yang, Y., Fik, T.J., Zhang, H.L.: Designing a tourism spillover index based on multidestination travel: A two-stage distance-based modeling approach. *J. Travel Res.* **56**(3), 317–333 (2017)
48. Yang, Y., Wong, K.K., Wang, T.: How do hotels choose their location? Evidence from hotels in Beijing. *Int. J. Hosp. Manag.* **31**(3), 675–685 (2012)
49. Zhang, Z., Chen, R.J., Han, L.D., Yang, L.: Key factors affecting the price of Airbnb listings: A geographically weighted approach. *Sustainability* **9**(9), 1635 (2017)
50. Zhou, B., Wen, Z., Sutherland, I., Lee, S.K.: The spatial heterogeneity and dynamics of tourism-flow spillover effect: The role of high-speed train in China. *Tour. Econ.* **28**(2), 300–324 (2020)

A Semi-Markov Approach to Financial Modelling During the COVID-19 Pandemic



Riccardo De Blasis 

Abstract We assess the capabilities of the weighted-indexed semi-Markov chain (WISMC) model applied to high-frequency financial data during the COVID-19 pandemic which was characterised by periods of extreme volatility. In particular, we test the ability of the WISMC model to reproduce the typical stylised facts of the financial time series, such as the persistence of volatility. For a general analysis, we apply the model to three major indexes of the financial markets, i.e. the Standard & Poor 500 (SPX), the Dow Jones Industrial Average (DJI) and the Financial Times Stock Exchange 100 (FTSE) over a period that covers the first year of the COVID-19 pandemic, from January 2020 to December 2020. Moreover, we compare the results with the standard GARCH model. A Monte Carlo simulation shows that the WISMC model is able to reproduce the persistence of volatility and clearly outperforms the GARCH model in three different specifications.

Keywords Semi-Markov · WISMC · COVID-19 · Financial markets

1 Introduction

In this manuscript, we assess the capabilities of the weighted-indexed semi-Markov chain (WISMC) model by D'Amico and Petroni [7] to reproduce some of the stylised facts of the financial time series, such as the persistence of volatility, during the first year of the COVID-19 pandemic. The WISMC model belongs to a class of models that are part of the micro- to macro-approach, which focusses on the observable quantities and exploits the point process [12]. On the contrary, the macro- to micro-approach, which bases its analysis on the application of the standard econometric tools, is the general approach employed in the literature where the observed price is considered as a noisy representation of an unobserved price.

D'Amico and Petroni introduced the WISMC model as a generalisation of the semi-Markov model applied to financial data, which was initially exploited in [6].

R. De Blasis (✉)
Marche Polytechnic University, Ancona, Italy
e-mail: r.deblasis@univpm.it

Also, they provided some extensions of the model, in particular, in the multivariate setting [8, 9]. The main difference between the WISMC and the standard semi-Markov model is that while the latter employs two random variables, i.e. the observed price returns and the time between each price change, the former introduces a third variable that allows us to track the history of the process, thus allowing for better reproduction of the observed quantities. In the financial setting, the added variable is constructed using a function of the squared returns, e.g. the moving average or the exponentially weighted moving average (EWMA). This addition enables us to create a dependence of the returns' dynamic on the volatility level, which is an important feature of the financial time series.

As reported by D'Amico and Petroni, the application of the WISMC model to financial time series requires some calibrations to perform, among which is the discretisation of the continuous values of both the returns and the volatility index into discrete state spaces. This conversion can be performed manually by looking at the distribution of the processes or it can be automated. Some of the automatic algorithms have been proposed in [4, 11]. The latter will be the reference for the analysis of this manuscript by employing two machine learning algorithms, i.e. the Gaussian mixture model (GMM) and the k-means.

As the WISMC model has only been tested on financial markets during periods of stability, we assess its validity by considering the latest financial disruptions caused by the COVID-19 pandemic, in which the markets experienced extremely high volatility, especially around April 2020. To strengthen our results, we compare the WISMC model with some GARCH specifications, thus comparing the micro- to macro-approach and the macro- to micro-approach. Results from a Monte Carlo simulation show that the WISMC model is able to reproduce the persistence of volatility and clearly outperform the GARCH model in three different specifications.

The manuscript is organised as follows. Section 2 describes the semi-Markov and WISMC model while Sect. 3 covers the application of the model to financial data during the COVID-19 pandemic along with the comparisons with the GARCH model. Section 4 concludes the manuscript.

2 The Model

Before introducing the details of the weighted-indexed semi-Markov model, we describe the semi-Markov process which is the base of the WISMC. Initially, Levy [14] and Smith [19] introduced the semi-Markov processes independently. Other authors then studied them in [2, 17, 18]. After their introduction, the theory has been further implemented and expanded and the processes were used in many applications, from industrial to financial markets, see, e.g. [16, 20, 21]. For further analysis, the reader can refer to [1, 13].

In general, semi-Markov processes are considered a generalisation of the renewal processes and the Markov chain. Let us consider a probability space (Ω, \mathcal{F}, P) and two random variables

$$X_n : \Omega \rightarrow E \quad T_n : \Omega \rightarrow \mathbb{R}_+,$$

where $E = \{1, \dots, k\}$ is a finite state space, $n \in \mathbb{N}$ and $0 = T_0 < T_1 < T_2 < \dots$.

A process is called a *Markov renewal process* (X, T) with state space $E \times \mathbb{R}_+$ if

$$\begin{aligned} & \mathbb{P}(X_{n+1} = j, T_{n+1} - T_n \leq t | X_0, \dots, X_n; T_0, \dots, T_n) \\ & = \mathbb{P}(X_{n+1} = j, T_{n+1} - T_n \leq t | X_n) \quad a.s., \quad \forall n \in \mathbb{N}, j \in E, t \in \mathbb{R}_+. \end{aligned} \quad (1)$$

Assuming the time homogeneity of the process, the probabilities are independent of the time and can be written as

$$\mathbb{P}(X_{n+1} = j, T_{n+1} - T_n \leq t | X_n = i) = Q(i, j, t), \quad (2)$$

where Q is the *semi-Markov kernel*, with $Q(i, j, 0) = 0, \forall i, j \in E$.

Taking the limit of the kernel for $t \rightarrow +\infty$, we obtain the transition probabilities of the Markov chain $\{X_n\}_{n \in \mathbb{N}}$, with state space E . Thus, for each pair (i, j) ,

$$\lim_{t \rightarrow +\infty} Q(i, j, t) = P(i, j), \quad (3)$$

with $P(i, j) \geq 0$ and $\sum_{j \in E} P(i, j) = 1, i, j \in E$.

Moreover, dividing the semi-Markov kernel in (2) by the transition probabilities in (3), with the convention that $G(i, j, t) = 1$ if $P(i, j) = 0$,

$$G(i, j, t) = \frac{Q(i, j, t)}{P(i, j)}, \quad (4)$$

we can compute the conditional waiting time distribution function defined as

$$G(i, j, t) = \mathbb{P}(T_{n+1} - T_n \leq t | X_n = i, X_{n+1} = j). \quad (5)$$

Also, [2] reports that the increments $T_{n+1} - T_n$ are conditionally independent given the Markov chain X_n . In particular, when the state space E reduces to a single point, the increments are independent and identically distributed non-negative random variables, thus obtaining a *renewal process*.

Finally, the *semi-Markov process* with state space E and transition kernel $Q(i, j, t)$ is a continuous-time parameter process defined as

$$Y_t = X_n \text{ for } t \in [T_n, T_{n+1}). \quad (6)$$

The process can be seen as a system in a certain time t moving from one state to another with random sojourn times in between [2]. The sojourn intervals $[T_n, T_{n+1})$ are a random variable themselves. Their distribution depends on the state being visited X_n and the next state to be visited X_{n+1} . When their distribution is exponential, the semi-Markov process becomes a continuous-time Markov chain. Instead, if we ignore

the time variables, we obtain a discrete-time Markov chain. Therefore, as the semi-Markov follows the Markov property only at the jumps instants and because it is not a memoryless process, it cannot be considered a Markovian process. Hence, the name semi-Markov.

If we extend the semi-Markov model with the inclusion of the memory of the process, we can obtain a high-order semi-Markov process, see, e.g. [10, 15]. However, the estimation of the such process required a lot of effort when we increase the dimension of the state space and the memory. To overcome this limitation, D'Amico and Petroni [3] extended the semi-Markov process with the inclusion of a third variable which collects the history of the process. The authors also provided applications to financial markets with some extensions [5, 7].

We consider a stochastic process U_n with values in \mathbb{R} representing an index process with historical information of the semi-Markov process. D'Amico and Petroni [9] defined it as

$$U_n(\boldsymbol{\theta}) = \sum_{k=0}^{n-1} \sum_{a=T_{n-1-k}}^{T_n-k-1} f(X_{n-1-k}, T_n, a, \boldsymbol{\theta}) + f(X_n, T_n, T_n, \boldsymbol{\theta}), \quad (7)$$

where $f : E \times \mathbb{N} \times \mathbb{R} \rightarrow \mathbb{R}$ is a Borel measurable bounded function and U_0 is known and non-random. The size of the vector of parameters, $\boldsymbol{\theta}$, depends on the chosen function f .

We can define the process Y_t as a *weighted-indexed semi-Markov chain* (WISMC) if, $\forall n \in \mathbb{N}$, the following equality holds true:

$$\begin{aligned} \mathbb{P}(X_{n+1} = j, T_{n+1} - T_n \leq t | X_0, \dots, X_n; T_0, \dots, T_n; U_0, \dots, U_n) \\ = \mathbb{P}(X_{n+1} = j, T_{n+1} - T_n \leq t | X_n = i, U_n = v) := Q(i, j, t, v), \end{aligned} \quad (8)$$

with Q being the *indexed semi-Markov kernel*.

The condition affirms that we only need the knowledge of the last state i and the last value of the index process U_n to define the probability of the next state. Therefore, to identify the system at any jump time T_n we only need the triple of processes $\{X_n, T_n, U_n\}$. Also, we note that if the indexed semi-Markov kernel is constant in v , then it falls back to a semi-Markov kernel.

Moreover, for each pair (i, j) and each value of the index, we have that $Q(i, j, 0, v) = 0$ and the transition probabilities of the Markov chain, $\{X_n\}_{n \in \mathbb{N}}$, with state space E , are defined as

$$P(i, j, v) = \mathbb{P}(X_{n+1} = j | X_n = i, U_n = v). \quad (9)$$

Contrary to the probabilities in (3), they include the dependence on the index level.

Finally, the conditional waiting time distribution function also includes the dependence on the index level

$$G(i, j, t, v) = \mathbb{P}(T_{n+1} - T_n \leq t | X_n = i, X_{n+1} = j, U_n = v). \quad (10)$$

So far, we have not defined the functional form of the index, $U_n(\theta)$, even though it is necessary to apply the model to real data. The simplest and intuitive approach to financial data is to consider a moving average or the exponentially weighted moving average (EWMA) of the squared process, $(X_n)^2$ as in [5, 7], respectively. Choosing the dependence on the squared process, the authors introduce a dependence on the volatility and, in particular, on its clustering dynamic, which is an observed stylised fact in financial markets. Specifically, using the EWMA definition, the function becomes

$$f(X_{n-1-k}, T_n, a, \lambda) = \frac{\lambda^{T_n-a} X_{n-1-k}^2}{\sum_{k=0}^{n-1} \sum_{a=T_{n-1-k}}^{T_n-k-1} \lambda^{T_n-a}}. \quad (11)$$

3 Application to Financial Data

We assess the capabilities of the WISMC model during the first year of the COVID-19 pandemic by considering three major indexes of the financial markets, i.e. the Standard & Poor 500 (SPX), the Dow Jones Industrial Average (DJI) and the Financial Times Stock Exchange 100 (FTSE). We consider only the first year of the pandemic as it is the one in which the financial markets suffered the highest volatility. Considering the opening and closing times of the stock exchanges within the analysed time range, which is from January to December 2020, the price returns data, collected from the Thomson Reuters Datascope, are sampled at 5-second intervals to guarantee enough observations to estimate the model's parameters.

Figure 1 shows the three series of the percentage log-returns over the selected horizon along with their histogram. From the figure, it is clear that all series present a high volatility period of about 2 months just after the outburst of the COVID-19 pandemic in February 2020. Moreover, the histograms depict a high concentration of observations around zero returns in all series.

The summary statistics of the log-returns in percentage are reported in Table 1. From the figure, we observe that the average returns are approximately zero for all three series and that all standard deviations are very low, as expected from the visual analysis of the histograms. All series present autoregressive conditional heteroscedasticity as Engle's test at 1, 6 and 12 lags on the residuals of the GARCH models are all statistically significant at the 1% level. This means that a GARCH specification should correctly capture the time-varying structure of the variance.

The WISMC model requires the calibration of some of its parameters before proceeding with the estimation of the transition probability matrices $P(i, j, v)$ and the conditional waiting time distribution $G(i, j, t, v)$. Specifically, we have to convert the returns process into a discrete state space, as well as the volatility index which has values in \mathbb{R} . D'Amico and Petroni [7] based both discretisations on the visual analysis of their histograms. In a more recent paper [4], they employ a change point approach to overcome this subjectivity in the partitioning of the state space. Finally, De Blasis [11] proposed the use of some machine learning algorithm to automate

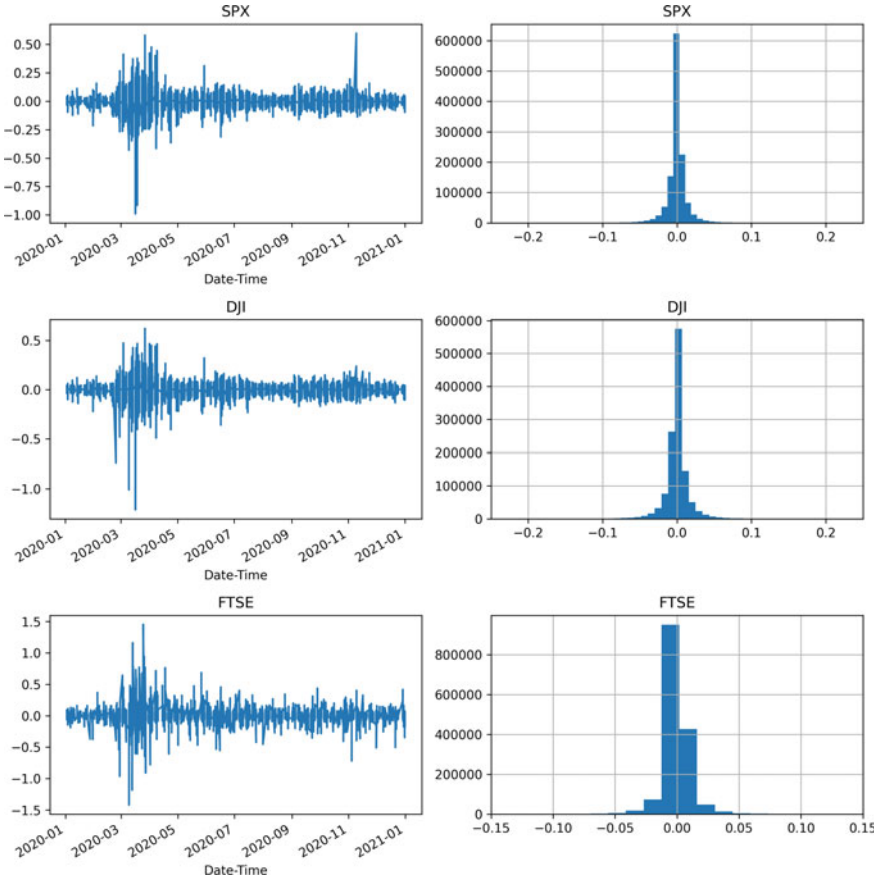


Fig. 1 Time series and histogram of the 5-second log-returns in percentage of the three major indexes, i.e. standard & Poor 500 (SPX), Dow Jones industrial average (DJI) and financial times stock exchange 100 (FTSE)

both procedures, specifically the Gaussian Mixture Model (GMM) and the k-means. In particular, the GMM algorithm proved to be very useful in the high-frequency setting when discretising the price returns. In our current analysis, we employ the GMM algorithm to discretise the price returns into a 3-state space, i.e. negative, zero and positive returns, and the k-means algorithm to discretise the volatility index into a 5-state space. We included five states for the volatility to extend the range of its possible values. With the application of the GMM algorithm to the discretisation of the returns, we filter the continuous values for the three series based on the bins defined in Table 2.

Further calibration would be needed to set the smoothing parameter of the EWMA function of the volatility index. However, empirical analysis suggests that values around 0.97 obtained the best results in applications to financial markets, see, e.g.

Table 1 Summary statistics of the 5-second log-returns in percentage for the three major indexes, i.e. standard & poor 500 (SPX), Dow Jones industrial average (DJI) and financial times stock exchange 100 (FTSE). The time range is from January to December 2020. ARCH refers to the statistic of Engle’s test for autoregressive conditional heteroscedasticity at 1, 6 and 12 lags

	SPX	DJI	FTSE
Obs.	1240676	1240683	1546345
Mean	0.0000	0.0000	0.0000
Median	0.0000	0.0000	0.0000
Max	0.5992	0.6208	1.453
Min	-0.9906	-1.2129	-1.4207
Std. dev.	0.0152	0.0181	0.0134
Skewness	-0.1059	-0.2976	-1.3887
Excess Kurtosis	61.8876	58.3234	618.3075
Jarque-Bera	197997658***	175865281***	24632753326***
ARCH(1)	36846.4***	49224.4***	164658.2***
ARCH(6)	73774.8***	85371.3***	210398.6***
ARCH(12)	89276.3***	101593.9***	216061.2***

*** Indicate the rejection of the null hypothesis of the tests at the 1% levels of significance

Table 2 Discretisation bins of the price returns employing the GMM algorithm

	State -1	State 0	State 1
SPX	[-0.991,0.010)	[0.010,0.039)	[0.039,0.599]
DJI	[-1.213,0.012)	[0.012,0.046)	[0.046,0.621]
FTSE	[-1.421,0.009)	[0.009,0.045)	[0.045,1.453]

[7, 8]. Therefore, in the remainder of the analysis, we will apply the same smoothing value, $\lambda = 0.97$.

After the model’s calibration, we need to perform the estimation of the transition probability matrices and conditional waiting time distributions. To this extent, we follow the estimation procedure outlined in Appendix B in [8].

Finally, to assess the validity of the WISMC model we perform a Monte Carlo simulation following the algorithm presented in [7] that we report here:

1. set $n = 0, X_0 = i, T_0 = 0, U_0 = v$, horizon time = T ;
2. given X_n and U_n , sample X from $P(i, j, v)$ and set X_{n+1} ;
3. given X_n and X_{n+1} , sample W from $G(i, j, t, v)$ and set $T_{n+1} = T_n + W$;
4. set U_{n+1} using (7) and (11);
5. if $T_{n+1} \geq T$ stop, else set $n = n + 1$ and go to 2.

With the simulations, we verify the ability of the WISMC model to reproduce the long-range serial correlation of the squared returns which is a typical stylised fact of the financial returns series. The autocorrelation function of the squared returns used for the analysis is as follows:

Table 3 Root mean square error (RMSE) and mean absolute error (MAE) between real and synthetic autocorrelation functions reported for WISMC model and compared with different GARCH models

		SPX	DJI	FTSE
RMSE	WISMC	0.013	0.012	0.014
	GARCH(1,1)	0.090	0.132	0.184
	GARCH(1,2)	0.152	0.169	0.160
	GARCH(2,1)	0.113	0.131	0.141
MAE	WISMC	0.011	0.010	0.011
	GARCH(1,1)	0.070	0.120	0.122
	GARCH(1,2)	0.111	0.146	0.140
	GARCH(2,1)	0.069	0.103	0.108

$$\Sigma(\tau) = \frac{\text{Cov}(Y^2(t + \tau), Y^2(t))}{\text{Var}(Y^2(t))}, \quad (12)$$

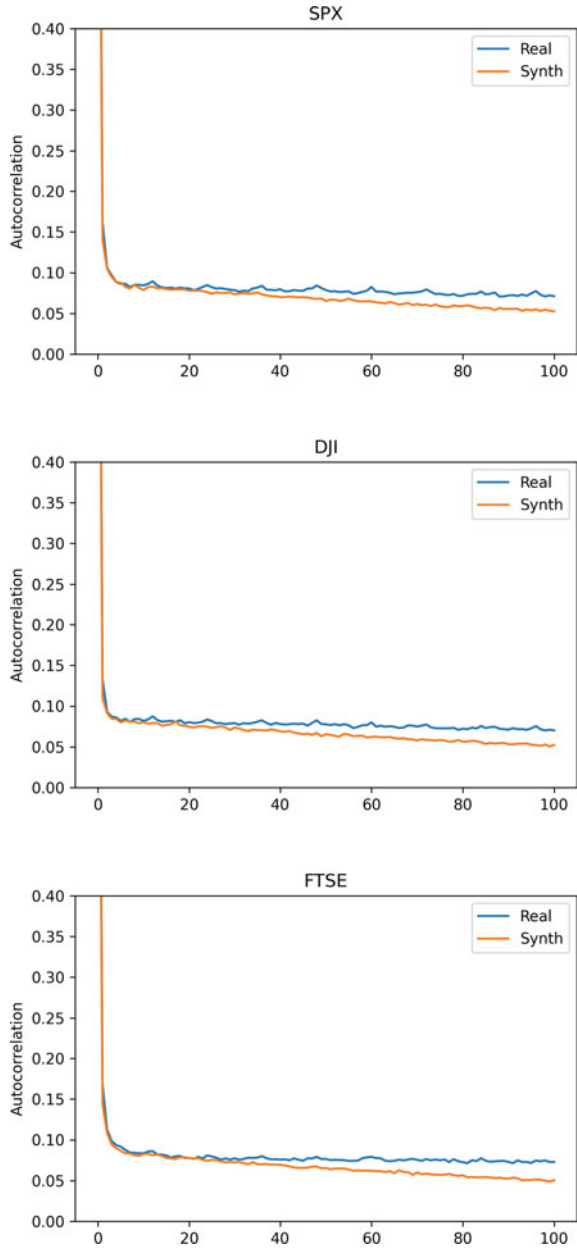
where Y is the process of the returns and τ is the time lag. We estimate the autocorrelation for both real and simulated data and compute the root mean square error (RMSE) and mean absolute error (MAE) for comparison. Additionally, as Engle's test of the squared returns suggested that the GARCH specification should suffice to model the varying volatility, we assess both WISMC and GARCH models' ability to reproduce the long-range serial correlation by comparing their RMSE values.

Figure 2 shows the autocorrelation functions of the squared returns for both real and synthetic time series using the WISMC model. In all three time series, i.e. SPX, DJI and FTSE, the model shows a good ability to reproduce the serial correlation of the real data. Specifically, the values of the autocorrelation from the simulated data at small lags are directly comparable with the values from the real data. Some differences between the real and synthetic values are evident when considering a higher number of lags.

However, to perform a proper comparison, we report the RMSE and MAE values for the WISMC model along with three different GARCH specifications in Table 3. In all cases, the WISMC model outperforms the GARCH specifications in reproducing the long-range serial correlation of the squared returns.

In addition, we test the WISMC model for the ability to reproduce the probability mass function of the price returns. To this extent, we employ the Kolmogorov–Smirnov (KS) hypothesis test to compare the distributions of both real and simulated data. Figure 3 reports the comparison of the distributions for the three indexes.

Fig. 2 Autocorrelation functions of the squared returns for real data and synthetic time series using the WISMV model



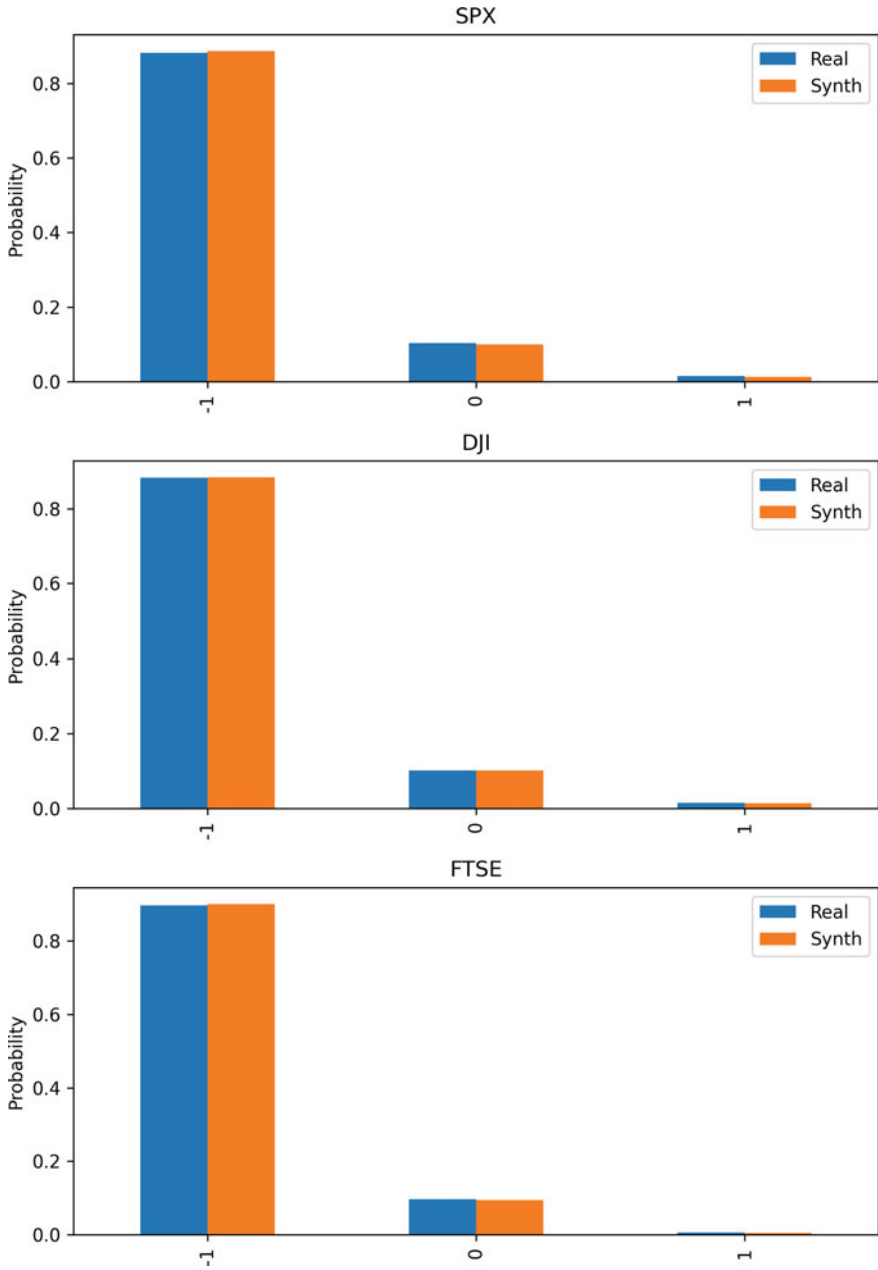


Fig. 3 Comparison of the probability mass function for real and simulated data for each index

4 Conclusion

In this manuscript, we tested the capability of the weighted-indexed semi-Markov chain (WISMC) model by D'Amico and Petroni [7] to reproduce the typical stylised facts of the financial time series, such as the persistence of volatility. In particular, we tested the model during a period of extreme volatility, like the COVID-19 pandemic. We applied the model to three major indexes of the financial markets, i.e. the Standard & Poor 500 (SPX), the Dow Jones Industrial Average (DJI) and the Financial Times Stock Exchange 100 (FTSE) in the first year of the COVID-19 pandemic, from January 2020 to December 2020 and compare the results with the general GARCH specification.

Results from the Monte Carlo simulation clearly show that the WISMC model is able to reproduce the long-range serial correlation of the squared returns of the selected time series. Moreover, it outperforms the GARCH model in three different specifications.

References

1. Barbu, V.S., Limnios, N.: Semi-Markov Chains and Hidden Semi-Markov Models toward Applications: their Use in Reliability and DNA Analysis. Lecture Notes in Statistics. Springer, New York, NY (2009). <https://link.springer.com/book/10.1007/978-0-387-73173-5>
2. Çinlar, E.: Markov renewal theory: a survey. *Manag. Sci.* **21**(7), 727–752 (1975). <https://doi.org/10.1287/mnsc.21.7.727>
3. D'Amico, G.: Age-usage semi-Markov models. *Appl. Math. Modell.* **35**(9), 4354–4366 (2011). <https://doi.org/10.1016/j.apm.2011.03.006>
4. D'Amico, G., Lika, A., Petroni, F.: Change point dynamics for financial data: an indexed Markov chain approach. *Ann. Finance* **15**(2), 247–266 (2019). <https://doi.org/10.1007/s10436-018-0337-0>
5. D'Amico, G., Petroni, F.: A semi-Markov model with memory for price changes. *J. Stat. Mech.: Theory Exp.* **2011**(12), P12009 (2011). <https://doi.org/10.1088/1742-5468/2011/12/P12009>
6. D'Amico, G., Petroni, F.: A semi-Markov model for price returns. *Phys. A: Stat. Mech. Appl.* **391**(20), 4867–4876 (2012). <https://doi.org/10.1016/j.physa.2012.05.040>
7. D'Amico, G., Petroni, F.: Weighted-indexed semi-Markov models for modeling financial returns. *J. Stat. Mech.: Theory Exp.* **2012**(07), P07015 (2012). <https://doi.org/10.1088/1742-5468/2012/07/P07015>
8. D'Amico, G., Petroni, F.: Copula based multivariate semi-Markov models with applications in high-frequency finance. *Euro. J. Oper. Res.* **267**(2), 765–777 (2018). <https://doi.org/10.1016/j.ejor.2017.12.016>
9. D'Amico, G., Petroni, F.: A micro-to-macro approach to returns, volumes and waiting times. *Appl. Stoch. Models Bus. Ind.* **37**(4), 767–789 (2021). <https://doi.org/10.1002/asmb.2622>
10. D'Amico, G., Petroni, F., Praticco, F.: First and second order semi-Markov chains for wind speed modeling. *Phys. A: Stat. Mech. Appl.* **392**(5), 1194–1201 (2013). <https://doi.org/10.1016/j.physa.2012.11.022>
11. De Blasis, R.: The weighted-indexed semi-Markov model: Calibration and application to financial modelling (2022). <https://papers.ssrn.com/abstract=4212391>
12. Fodra, P., Pham, H.: Semi-Markov model for market microstructure. *Appl. Math. Finance* **22**(3), 261–295 (2015). <https://doi.org/10.1080/1350486X.2015.1037963>

13. Janssen, J., Manca, R.: Applied Semi-Markov Processes. Springer New York, NY, first edn. (2006). <https://link.springer.com/book/10.1007/0-387-29548-8>
14. Levy, P.: Processus semi-Markoviens. In: Proceedings of the International Congress of Mathematicians. vol. III, pp. 416–426. North-Holland Publishing Co., Amsterdam, 1956, Amsterdam (1954)
15. Limnios, N., Opri, G.: Ch. 14. An introduction to semi-Markov processes with application to reliability. In: Handbook of Statistics, Stochastic Processes: Modelling and Simulation, vol. 21, pp. 515–556. Elsevier (2003). [https://doi.org/10.1016/S0169-7161\(03\)21016-6](https://doi.org/10.1016/S0169-7161(03)21016-6)
16. Pasricha, P., Selvamuthu, D., D'Amico, G., Manca, R.: Portfolio optimization of credit risky bonds: a semi-Markov process approach. *Financ. Innov.* **6**(1), 25 (2020). <https://doi.org/10.1186/s40854-020-00186-1>
17. Pyke, R.: Markov renewal processes: definitions and preliminary properties. *Ann. Math. Stat.* **32**(4), 1231–1242 (1961). <https://doi.org/10.1214/aoms/1177704863>
18. Pyke, R.: Markov Renewal Processes with Finitely Many States. *Ann. Math. Stat.* **32**(4), 1243–1259 (1961). <https://doi.org/10.1214/aoms/1177704864>
19. Smith, W.L.: Regenerative stochastic processes. In: *Proc. R. Soc. London. Ser. A, Math. Phys. Sci.* **232**(1188), 6–31 (1955). <https://doi.org/10.1098/rspa.1955.0198>
20. Swishchuk, A., Hofmeister, T., Cera, K., Schmidt, J.: General semi-Markov model for limit order books. *Int. J. Theor. Appl. Financ.* **20**(03), 1750019 (2017). <https://doi.org/10.1142/S0219024917500194>
21. Vasileiou, A., Vassiliou, P.C.G.: An inhomogeneous semi-Markov model for the term structure of credit risk spreads. *Adv. Appl. Probab.* **38**(1), 171–198 (2006). <https://doi.org/10.1239/aap/1143936146>

From News to Sentiments and Stock Price Directions



Fennee Chong and Bharanidharan Shanmugam

Abstract Forecasting stock price direction using news sentiment has gained momentum as trading activities are information driven in this digital era. News fuel pessimism or optimism in the investors mindset which are often translated by market sentiment indicators that influence stock price direction. The type of information that affects short-term market reactions are political news, business news and technical indicators. Of late, machine learning technique has become one of the popular approaches being applied in stock market predictions. The merit of this approach in predicting stock price directions using news sentiments as inputs has been confirmed by a case study included in this chapter.

Keywords News · Sentiments · Stock price directions · Machine learning

1 Introduction

Classical financial theory argues that market is efficient as investors are rational and trade is based on unbiased cognitive evaluation and maximisation of expected utility, hence investor sentiment is irrelevant. This assertion is supported by the Capital Pricing Model and the Efficient Market Hypothesis. Nonetheless, the occurrences of anomalies such as stock market crisis and housing market crisis have challenged this view, implying that the idiosyncrasies in human behaviour vary and this leads to the fluctuation in asset prices. Consequently, behavioural finance was introduced as an area of study on how psychological influences can drive market to behave irrationally. The influence of psychological biases is one of the key aspects of behavioural finance studies. Shefrin [1] wrote that

F. Chong (✉) · B. Shanmugam
Charles Darwin University, Casuarina, Australia
e-mail: fen.chong@cdu.edu.au

B. Shanmugam
e-mail: bharani.shanmugam@cdu.edu.au

... the market has a psychology, more specifically it has a character. It has thoughts, beliefs, moods, and sometimes stormy emotions (2005, pp. 203).

Many subsequent studies have provided evidence to support this claim. De Long et al. [2] posited that asset prices would be determined in part by noise traders. And noise traders are investors who make decisions based on sentiment, irrational beliefs or biases instead of fundamental analysis. This was supported by Kumar and Lee [3] who contended that feedback from social interactions generates emotions and sentiments that led to trade in concert rather than diversified trade in the asset market. Positive sentiment induces investors to be more assured about investing in certain assets while negative sentiment involves a greater level of uncertainty and therefore has the opposite effects [4].

The advancement of information, communication and technology sector has facilitated high-speed dissemination of news covering large demographics within a very short time frame, and this further expedites the formation of emotions and sentiment swings in the market. Therefore, in order to better understand and response to how the market behave, it is crucial for investors, regulators and analyst to incorporate sentiment analysis in market evaluation and price prediction.

2 What Contribute to Sentiment and Share Prices?

Investors access information from various sources to generate buy and sell decisions. Therefore, information availability becomes the key that determines the behaviour of investors on whether to buy or sell and this behaviour moves prices. Investors can be affected by news made available through formal channel such as company announcements, market news or via non-formal channel such as blogs, tweeter or other social platforms. Information will generate emotions that drive sentiments and eventually decision formation. The type of news that drives changes in market sentiment can be

- Political news: This can range from information released to the public from the central or state government, ministries, government departments or international news related to political changes of certain country.
- Business news: Referring to company announcements, reports or developments concerning changes in the industry, new product launches or change in managements.
- Rumours and speculation: These are often stories or report of uncertain or doubtful truth that are being circulated in the market with no concrete source.
- Other factors: Recent price movements, price projections and market dynamics of other assets available in the market.

All these news generate trading signals and fuel pessimism or optimism in the investors mindset which are often translated by market sentiment indicators that would influence stock price directions eventually.

3 Analysing Stock Price Trends

Predicting stock price trends or directions can be challenging but useful in aiding decision-making on the timing to buy, hold or sell certain stock. The common ways that have been used are

- **Fundamental analysis:** Asses the stock's intrinsic value to predict potential direction of stock prices. This method utilises both the qualitative and quantitative information included in the company's reports and financial statements, conditions within the industry and data of the overall state of the economy including but not limited to employment rates, gross domestic product, interest rates and others. The main objective is to examine whether a stock is undervalued or overvalued.
- **Technical analysis:** This method of stock analysis is considered in contrast to fundamental analysis as it bypasses the company's underlying fundamentals and focussing on the statistical trend of charts of a stock in predicting future direction of a stock. Information used in technical analysis includes historical price movements and trading volume. Technical analysis is more suitable to be used to formulate short-term trading strategy.
- **Sentimental analysis:** It is often referred to as opinion mining which is performed to evaluate subjective expressions. This method is an effective way to evaluate written or spoken language and classify the expression as positive, negative or neutral to give useful indication of emotions and predict market sentiment and responses based on the predicted emotions.
- **Machine learning techniques:** This is a powerful technique that uses algorithm to predict stock price trends. Input variables can be from one or a combination of the above techniques. The advantage of this method is its suitability for handling big data and the variety of data that may be used for analysis. Nonetheless, it is also prone to biases including sample biases, measurement bias, exclusion bias, recall bias and confirmation bias [5].

Investors and researchers can use one of the methods above or use one method in conjunction with the others in determining trends and directions of stock prices.

For this chapter, a case study is presented below to answer the following questions:

1. Do news sentiments impact stock price trends in the short run?
2. What is the predictive accuracy of news sentiment and technical analysis on stock trend using selected machine learning algorithms?

4 A Case Study on How News Sentiment Affects Stock Price Direction Using Machine Learning Model

To provide an example on whether news sentiment affect stock price direction using machine learning algorithms, the paper entitled "Do News Sentiments Predict Stock Price Direction?" presented in the 2022 ITSE will be discussed [6].

4.1 *Background of the Study*

As indicated earlier, behavioural finance argues that trading emotions such as fear and greed determine the overall sentiment of the financial market. On the other hand, news and information represent the primary factors that influence emotions and hence determine market sentiment [7, 8]. Types of news that are available in the financial market are historical prices which indicate trends and used in technical analysis, political news, business news as well as insider information. Most of the time, investors are conditioned to follow sentiment of the market. Therefore, a quantitative approach for measuring sentiment in text by deriving scores for positive or negative sentiments can gauge investors' decision-making process, and hence the stock price and market trend. Sentimental analysis emerged as a subfield of text mining in information technology and become popular about two decades ago [9, 10]. Numerous studies have provided evidence that machine learning algorithms can be effective in predicting stock prices [11, 12]. However, applying the machine learning method to determine stock prices directions is relatively new.

Bustos and Pomares-Quimbaya conducted a systematic review on the stock market forecast and found that variables such as technical indicators and raw market information are commonly used for prediction while recent studies have included social network news, and this addition has improved prediction accuracy [11]. The authors also indicated that frontier and emerging countries show more accurate stock price prediction using machine learning method compared to developed countries due to the higher level of randomness and complexity of the developed markets. On average, machine learning algorithms can predict at an accuracy of 0.7–0.8 level. In terms of machine learning algorithms used, they indicated that Support Vector Machine (SVM) remains the most popular technique to date.

Khan et al. predicted stock market trends using public sentiment from Twitter and political events by classifying them either a positive or negative news. They posited that the impact of public sentiment only produced 0–3% increase in accuracy while political events produced a significantly higher (10–20%) improvement of prediction accuracy [13]. Song et al. introduced a news-based stock portfolio construction approach using financial news sentiment scores and historical returns as ranking features to develop stock selection rules. This study achieves profound results whereby the selected portfolio using learning-to-rank algorithms outperformed both the S&P500 index and the hedge fund industry average performance during the same period [12]. In a related study, Kinyua et al. examined how market sentiment as a result of President Trump's tweets have on DJIA and S&P500 using two datasets: the tweets dataset and the stock market indices dataset. With machine learning regressions including linear regression, decision tree regression and random forest regression, they found that market reacted significantly to this information. In addition, except for the extreme tweet sentiment categories, generally, all other tweet sentiments led to a decline in the trend of the two market indices included in this study [14].

Chu et al. posited that financial news have persuasive impact on market performance. They introduced sentiment-of-topic to improve accuracy of prediction for traders. Incorporating financial news as a proxy for market sentiment is important as financial time-series data often display challenging behaviour which can compromise the accuracy for commonly used financial forecasting techniques. Finding showed double-digit improvement in portfolio return after including sentiment generated from financial news [15]. On the other hand, Renault contended that the size of the dataset used in a study has profound impact on results. Complex algorithms would not help to improve the accuracy of the results without the presence of good dataset and text mining skills. To provide reliable result, a dataset of approximately 100,000 to 250,000 labelled messages is considered sufficient [16].

A similar study by Weng et al. employed three machine learning algorithms including decision trees, neural networks and SVM to investigate whether time-series data, technical indicators and online news are capable of predicting next day's stock directions. In this experiment, they managed to achieve an 85% predictive accuracy which provided support to the value of online information in predicting stock trends [17]. These results were endorsed in a more recent study by Moulkalled et al. which examined similar set of variables using recurrent neural network (RNN), feedforward neural network (FFNN) and support vector regression (SVR). The authors also asserted that SVM provided the highest level of accuracy in directional prediction [18].

4.2 Machine Learning Models Used in This Case Study

Generally, there are two machine learning models: supervised and unsupervised. Many researchers have applied either single or a combination of machine learning approaches including SVM, Decision trees, Logic Regression and Random Forest to predict stock prices. For this case study, supervised machine learning approaches, Support Vector Machine (SVM) and Extreme Gradient Boosting (XGB) are used. Data collected from April 2015 to April 2021 are used for analysis. The types of news used in this study are business news and political news collected from Reuters and Bloomberg for stocks listed on the S&P500 index. A total of 304,000 news articles are gathered for analysis and this is within the range of similar studies [19–22]. Loughran & McDonald Financial Dictionary is used for sentiment classification while Valence Aware Dictionary and Sentiment Reasoner (VADER) is employed to detect sentiments expressed in social media [21, 23, 24]. News sentiments are classified as positive, negative or neutral.

Other input variables used are the historical prices and technical indicators. Seven technical indicators are included in the analysis including moving averages (MA), momentum (MOM), moving average convergence divergence (MACD), rate of change (ROC), relative strength index (RSI), commodity channel index (CCI) and Bollinger bands (BB) [21, 25]. Four models listed below are tested in this study while Model 1 serves as the base model.

Model 1:

$$\text{Stock Price Direction} = f(\text{Historical Price, Technical Indicators})$$

Model 2:

$$\text{Stock Price Direction} = f(\text{Historical Price, Technical Indicators, Political News Sentiment})$$

Model 3:

$$\text{Stock Price Direction} = f(\text{Historical Price, Technical Indicators, Business News Sentiment})$$

Model 4:

$$\text{Stock Price Direction} = f(\text{Historical Price, Technical Indicators, Political News Sentiment, Business News Sentiment}).$$

4.2.1 Support Vector Machine (SVM)

Over time, many prediction models have focussed on linear statistical time-series models for stock market prediction. However, the variance underlying the movement of stocks and other assets makes linear techniques suboptimal, and non-linear models to have lower predictive errors. SVM algorithm was developed by Vapnik [26]. It is based on statistical learning theory and has been applied to stock market prediction for many years. Binary classifiers are one of the best approaches that could be applied for prediction and support vector machines are one of the best binary classifiers. They create a decision boundary such that most points in one category fall on one side of the boundary while other points in the other category fall on the other side of the boundary. Radial kernel approach was applied in this case study as it is one of the most popular kernel functions and is defined as follows:

$$K(x_i, x_k) = \exp\left(-\frac{1}{\delta^2} \sum_{j=1}^n (x_{ij} - x_{kj})^2\right) \quad (1)$$

where δ is known as the bandwidth of the kernel function [27]. The primary advantage of this function is it can handle diverse input sets, as the condition on the geometry is very few for the input data. Additionally, it classifies test examples based on the example's Euclidean distance to the training points, and weights closer training points more heavily. Another primary advantage is the classification is based heavily on the most similar training examples and takes advantage of patterns in the data. This works exactly as what is needed for time-series data such as stock prices that display trends over a period of time.

4.2.2 Extreme Gradient Boosting (XGB)

Gradient boosting is a type of classifier used to merge different slow learners to produce a robust forecasted model [28]. The critical advantages of XGB are the flexibility, usage of parallel processing, handling the missing data at ease and works well with small-to-medium datasets. The following principle describes gradient boosting: initially, develop an error function, and that function is optimised at the time of model building. Following that, iteratively develop weak model for forecasting and finally combine all the weak models to create a more robust model. The Extreme Gradient Boost model is an early experimental model [29] and is an improved version of the gradient boosting algorithm, it is more scalable and efficient. This model supports the regularisation technique to avoid overfitting and has the capability to learn from non-linear datasets. Moreover, the parallelisation feature allows for it to train with multiple CPU cores [30]. It is one of the tree-based ensemble additive models that are composed of multiple base learners. In general, the XGBoost can be represented as

$$\begin{aligned}
 F &= \{ m_1, m_2, m_3, m_4 \dots m_n \} \text{ set of base learners,} \\
 \text{Predictive model} = \hat{y}_i &= \sum n_t = 1m_t(x_i)
 \end{aligned}
 \tag{2}$$

where \hat{y}_i is the final predictive model, which is the combination of all weak learners, and x is the input feature for each weak learner, i.e. m [30].

4.3 Experiment Design and Evaluation

This case study utilised the following framework (Fig. 1).

In this experiment, the output of the experiment will produce a value ranging from 0 to 1. For example, a value of 0.56 indicates a 56% of uptrend on the next day for a particular stock under study. For model evaluation, this case study adopted the train/test split and K-fold cross validation [31]. The validation process was as shown in Fig. 2. The train/test approach splits data into test data and trained data while K-Fold cross validation separates dataset into k subsets (folds) to be trained on $(k-1)$ random subset and test the outstanding subset in an iterative manner.

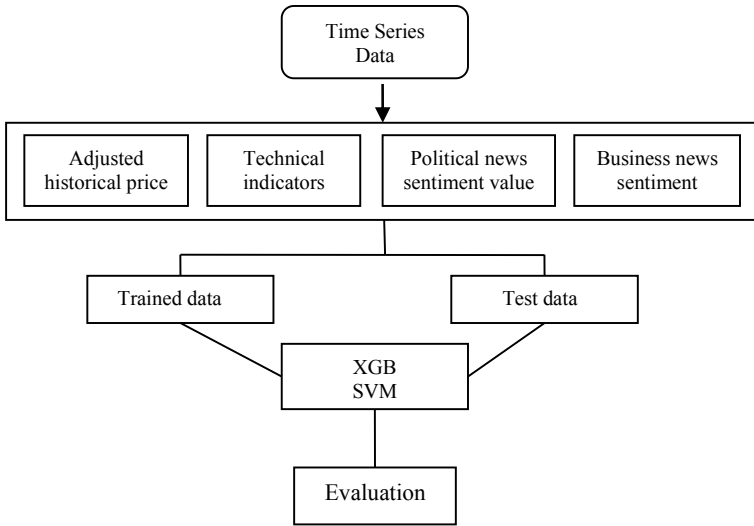


Fig. 1 Experiment framework

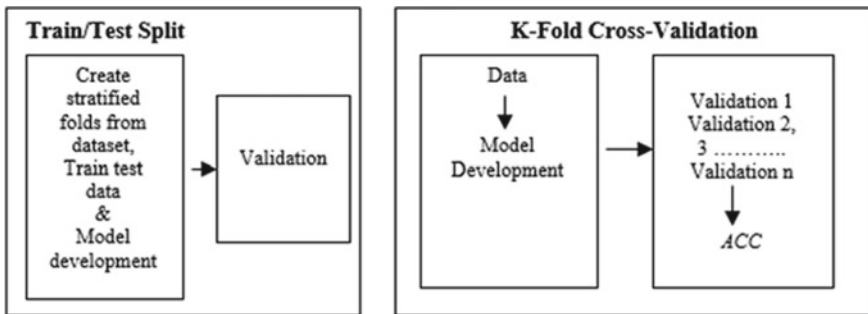


Fig. 2 Algorithm validation diagram

4.4 Analysis and Discussion of Results

This section reports and discusses results from the experiments. The accuracy metric was used as the evaluation criterion to examine the predictive power of sentimental proxies on stock price directions using the SVM and XGB algorithms in conjunction with technical analysis indicators and adjusted historical price. Table 1 exhibits the results of the experiment.

From the results displayed, the predictive value of daily political and business news sentiments on stock price trend was confirmed. Comparing the results across the four models analysed, it was found that the inclusion of both political and business news sentiments was able to further improve accuracy comparing to the base

Table 1 Test results

Evaluation	ML algorithms	Model 1	Model 2	Model 3	Model 4
Accuracy	SVM	0.521	0.530	0.521	0.527
(ACC)	XGB	0.818	0.826	0.818	0.826

Note Variables used: Model 1: Historical price and technical indicators, Model 2: Historical price, technical indicators and political news sentiment, Model 3: Historical price, technical indicators and business news sentiment, Model 4: Historical price, technical indicators, political news sentiment and business news sentiment

model (Model 1). Between political and business news, political news sentiment was proved to be able to predict stock price trend more accurately than its counterpart. However, the differences in terms of performance between models are mild. This implies that historical price movements and technical analysis indicators remain valid in predicting stock price directions. It was also observed that SVM’s prediction performance is relatively low at around 52% due to its inefficiency in kernel function and burden of memory complexity. Nonetheless, results from this study are in line with Schumaker and Chen’s study which produces 57% of directional accuracy using SMV while testing on a S&P500 dataset ranging from 26 October to 28 November 2005 [32].

Findings from this study provided support that accuracy evaluations are in the range of 50 to 88 percent; implying that the predictive values obtained from news sentiments are strong enough for market participants to make stock directional prediction using machine learning algorithms despite the variation detected in algorithm performances [13, 15].

Table 2 shows more details of the experiment. From the results, XGB has consistently outperformed SVM, showing overall accuracy of slightly more than 80 percent for all models. Generally, the performance of SVM and XGB confirmed the findings from previous studies [11, 13, 33]. While comparing the results of all models, it was obvious that Model 4 consisting of all variables (historical price, technical indicators, political news sentiment, business news sentiment) had outperformed other models which further recognises the predictive power of political and business news sentiment in stock price directions. The contribution of news sentiment to stock trend prediction in this study is also consistent to those of the results of Khan et al. which asserted that political news is having a more significant impact than other public news with 10–20% and 1–3% of increase in accuracy, respectively [13].

Table 2 Accuracy of results by machine learning algorithms and prediction models

Evaluation	Algorithms	Model 1	Model 2	Model 3	Model 4
Threefold Cross Validation	SVM	0.515	0.570	0.601	0.535
	XGB	0.815	0.826	0.816	0.822
Fivefold Cross Validation	SVM	0.576	0.528	0.488	0.547
	XGB	0.818	0.823	0.818	0.823
15-fold Cross Validation	SVM	0.562	0.523	0.518	0.507
	XGB	0.823	0.829	0.822	0.827
Data-split 70–30	SVM	0.431	0.505	0.478	0.518
	XGB	0.818	0.828	0.814	0.830
Data-split 80–20	SVM	0.584	0.503	0.506	0.498
	XGB	0.822	0.826	0.819	0.829

5 Conclusion

Forecasting stock price movement can be challenging; however, accurate predictions are beneficial to retail and institution investors, investment advisors and fund managers. In order to facilitate more accurate investment decisions, researchers and analysts have utilised predictive attributes including fundamental analysis, technical analysis and sentimental analysis in conjunction with powerful statistical platform. More recently, using sentimental indicators with reference to social network has gained momentum. News represents one of the most powerful influencers on the mood and the sentiments of decision-makers in the stock market. The recent advances in machine learning models, an artificial intelligence technique, have allowed application of software to predict more accurate outcomes. Using political news, business news, technical indicators and historical prices, Vo, Chong and Shanmugam conducted a study to predict S&P 500 next day's stock directions using the SVM and XGB machine learning algorithm [6]. Findings from this study confirms that news sentiments are valid predictors of stock price directions. Furthermore, political news produces more prominent result than business news. Both SVM and XGB can provide price direction prediction to a high level of accuracy. Among these two machine algorithms, XGB can produce prediction accuracy up to 83 percent, this result confirmed findings from previous studies [11, 13, 32]. Consequently, there are reasonable grounds to believe that the addition of political and business news sentiment to technical indicators and historical price information can produce promising model for traders to predict, and hence facilitate more accurate investment decisions.

References

1. Shefrin, H.: A behavioural approach to asset pricing. Elsevier Academic Press, USA (2005)

2. De Long, J., Shleifer, A., Summers, L., Waldman, R.: Noise trader risk in financial markets. *J. Polit. Econ.* **98**(4), 703–738 (1990)
3. Kumar, A., Lee, C.: Retail Investor Sentiment and Return Comovements. **61**(5), 2451–2486 (2006)
4. Kuhnen, C., Knutson, B.: The influence of effect on beliefs, preferences, and financial decisions **46**(3), 605–626 (2011)
5. Lim, H.: 7 types of data bias in machine learning. <https://lionbridge.ai/articles/7-types-of-data-bias-in-machine-learning/> (2020), last accessed 2022/8/26
6. Vo, R., Chong, F., Shanmugam, B.: Do News sentiments predict stock price direction? Presented in the ITISE 2022, Gran Canaria 27th–30th June 2022 (2022).
7. Shiller, R.: *Irrational Exuberance*. Princeton University Press, USA (2000)
8. Chong, F.: Disposition effect and flippers in the Bursa Malaysia. *The Journal of Behavioural Finance* **10**(3), 152–157 (2009)
9. Pang, B., Lee, L.: Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*. 2(1–2):1–135. <https://doi.org/10.1561/1500000011>
10. Liu, B., Zhang, L.: A survey of opinion mining and sentiment analysis. In *Mining text data*, 415–463. Springer (2012)
11. Bustos, O., Pomares-Quimbaya, A.: Stock market movement forecast: A systematic review. *Expert Syst. Appl.* **156**, 113464 (2020). <https://doi.org/10.1016/j.eswa.2020.113464>
12. Song, Q., Liu, A., Yang, S.Y.: Stock portfolio selection using learning-to-rank algorithms with news sentiment. *Neurocomputing* **264**, 20–28 (2017)
13. Khan, W., Malik, U., Ghazanfar, M.A., Azam, M. A., Alyoubi, K. H., Alfakeeh, A.S.: Predicting stock market trends using machine learning algorithms via public sentiment and political situation analysis. *Soft Computing*, 24, 11019–11043 (2020)
14. Kinyua, D., Mutigwe, C., Cushing, D.J., Poggi, M.: An analysis of the impact of President Trump’s tweets on the DJIA and S&P 500 using machine learning and sentiment analysis. *J. Behav. Exp. Financ.* **29**(2), 100447 (2021)
15. Chu, W.C., Wong, R.K., Chen, F., Ho, I., Lee, J.: Enhancing portfolio return based on sentiment-of-topic. *Data Knowl. Eng.* (2017). <https://doi.org/10.1016/j.datak.2017.07.004>
16. Renault, T.: Sentiment analysis and machine learning in finance: a comparison of methods and models on one million messages. *Digital Finance* **2**(1), 1–13 (2020)
17. Weng, B., Ahmed, M.A., Megahed, F.M.: Stock market on-day ahead movement prediction using disparate data sources. *Expert Syst. Appl.* **79**, 153–163 (2017)
18. Moulalled, M., El-Hajj, W., Jaber, M.: Automated stock price prediction using machine learning. In *Proceedings of the second financial narrative processing workshop (FNP 2019)*, Turku Finland No. 165, 16–24. Linköping University Electronic Press (2019)
19. Deng, S., Zhang, N., Zhang, W., Chen, J., Pan, J. Z., Chen, H.: Knowledge-driven stock trend prediction and explanation via temporal convolutional network. *Companion Proceedings of the 2019 World Wide Web Conference* (2019)
20. Ding, X., Zhang, Y., Liu, T., Duan, J.: Deep learning for event-driven stock prediction. *Proceedings of the Twenty-fourth International Joint Conference on Artificial Intelligence*, (2015)
21. Li, X., Wu, P., Wang, W.: Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. *Inf. Process. Manage.* **57**(5), 102212 (2020)
22. Turner, Z., Labille, K., Gauch, S.: Lexicon-based sentiment analysis for stock movement prediction. *World Academy of Science, Engineering and Technology: International Journal of Mechanical and Industrial Engineering*.14 (5) (2020)
23. Hutto, C., Gilbert, E.: Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and social media*, (2014)
24. Sohagir, S., Petty, N., Wang, D.: Financial sentiment lexicon analysis. 2018 IEEE 12th International Conference on Semantic Computing (ICSC) (2018)
25. Yıldırım, D.C., Toroslu, I.H., Fiore, U.: Forecasting directional movement of Forex data using LSTM with technical and macroeconomic indicators. *Financial Innovation* **7**(1), 1–36 (2021)

26. Cortes, C., Vapnik, V. Support Vector Networks, *Machine Learning*, 20, 273–297, <https://doi.org/10.1007/BF00994018>, (1995)
27. Halls-Moore, M.: Support vector machines: A guide for beginners, <http://www.quantstart.com/articles/Support-Vector-Machines-A-Guide-for-Beginners> (2014), last accessed 2022/8/29.
28. Friedman, J.H.: Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29**(5), 1189–1232 (2001)
29. Yan, D., Zhou, Q., Wang, J., Zhang, N.: Bayesian regularisation neural network based on artificial intelligence optimisation. *International Journal Production Research* **55**, 2266–2287 (2017)
30. Padhi, D.K., Padhy, N., Bhoi, A.K., Shafi, J., Ijaz, M.F.: A fusion framework for forecasting financial market direction using enhanced ensemble models and technical indicators. *Mathematics* **9**(21), 2646 (2021). <https://doi.org/10.3390/math9212646>
31. Shen, S., Jiang, H., Zhang, T.: Stock market forecasting using machine learning algorithms. <https://cs229.stanford.edu/proj2012/ShenJiang> Zhang-Stock Market Forecasting using MachineLearningAlgorithms.pdf (2012)
32. Schumaker, R.P., Chen, H.: Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZF in text system. *ACM Transactions on Information Systems* **27**, 1–19 (2009). <https://doi.org/10.1145/1462198.1462204>
33. Ali, M. R.: Prediction Accuracy of financial data – Applying several resampling techniques. Master thesis. North Dakota state University (2020)

Recommendations of Stockbrokers Versus Fuzzy Portfolio Approach in Construction Sector



Anna Łyczkowska-Hanćkowiak  and Aleksandra Wójcicka-Wójtowicz 

Abstract In portfolio management, we can use oriented fuzzy numbers (OFNs) which include information uncertainty and imprecision related to financial market. In such a case, we also utilise an expected fuzzy discount factor which leads to an imprecise present value (PV). Basing on those factors we can obtain recommendations (buy, sell, accumulate, reduce) for individual stocks included in the portfolio. Similarly, agents of stock-brokerage regularly issue analogous recommendation basing on their knowledge, experience and quantitative analysis. The main purpose of the paper is to analyse the published recommendations of stockbrokers, their accuracy and actual impact on the operations on portfolio including stocks of companies in construction sector. The created portfolio includes securities identified by a present value assessed with the use of trapezoidal-oriented fuzzy number (TrOFNs) in times of issuance of the recommendation. All theoretical considerations are illustrated by an empirical case study. The elements of portfolio are positively or negatively oriented.

Keywords Fuzzy numbers · Imprecision · Portfolio analysis · Recommendations

1 Introduction

Individual investors before making an investment decision frequently search for reliable information on particular company or sector. However, though they seek professional support in decision-making, often they have limited experience with advanced numerical methods. Therefore, they prefer the linguistic information, which is more comprehensible than numerical one, despite the fact that numerical methods can be more efficient. The required linguistic information can take various forms.

A. Łyczkowska-Hanćkowiak (✉)

WSB University in Poznan, ul. Powstańców Wielkopolskich 5, 61-895 Poznań, Poland
e-mail: anna.lyczkowska-hanckowiak@wsb.poznan.pl

A. Wójcicka-Wójtowicz

Poznań University of Economics and Business, Al. Niepodległości 10, 61-875 Poznań, Poland
e-mail: aleksandra.wojcicka-wojtowicz@ue.poznan.pl

An example of such information is the recommendation of stockbrokers who are regarded as professionals with broad knowledge and experience in this field [13, 15]. However, both knowledge and experience usually are classified as imprecise and inaccurate foundations which, in turn, can make the recommendation slightly imprecise as imprecision is one of the main characteristics of any financial market information.

Considering the main assumption of the uncertainty that any unknown future state of affairs is uncertain [4] we can *ex ante* analyse the efficiency of the obtained recommendation as the given recommendation might or might not realise in the future. This way we can assess whether the information was correct or not. As uncertainty is derived from the ignorance of future states of affairs, it can be modelled with a certain probability. This can be achieved if we can point out a specific time in which the analysed state of affairs will be known to the observer. It allows us to use the trapezoidal-oriented fuzzy numbers (TrOFNs) in making portfolio investment decisions [12, 14]. In our paper, we investigate individual entities from construction sector included in a portfolio with PVs described by trapezoidal-oriented fuzzy numbers. Furthermore, the analysis based on fuzzy discount factor is less difficult than the one based on a return rate.

Any security gives its owner a right to obtain future financial income that is payable at a specified maturity. The value of this gain can be presented as an anticipated future value of capital. However, as the present value of future cash flows is an approximated value, it can be expressed by trapezoidal-oriented fuzzy numbers which are simpler in case of operations.

In our research, a financial portfolio is understood as an arbitrary, finite set of assets. We consider both, a multi-asset portfolio and a single asset portfolio, as an individual security can be assumed a portfolio. Any analysed asset is regarded as a fixed security in a long position. The considered portfolio elements are positively or negatively oriented. Their fuzzy discount factor is tested. Then, a weighted sum of positively oriented discount factors and a weighted sum of negatively oriented factors is calculated (due to the fact that adding oriented fuzzy numbers is not associative). Consequently, a portfolio discount factor is obtained as a weighted addition of both sums. In our procedure, we use both entropy and energy measures for the estimation of the imprecision risk of obtained investment portfolios.

The paper is organised as follows. Introduction is followed by Sect. 2 which presents trapezoidal fuzzy numbers, along with their main properties. In Sect. 3, the evaluation of imprecision for oriented fuzzy numbers is shown. Section 4 includes the notion of imprecisely estimated PVs and expected discount factor. Section 5 compares the accurateness of brokers' recommendations and the recommendations resulting from the TrOFNs analysis. The research is illustrated by a case study presenting two strategies: one strategy follows the recommendations of stockbrokers and the other the recommendations of a portfolio analysis basing on TrOFN approach. Section 6 concludes the findings of the research.

2 Fuzzy Numbers—Overview

Fuzzy Numbers (FNs), which were initially introduced by Dubois and Prade [1], constitute the broadest set of numbers in fuzzy theory. However, they can be divided into smaller subgroups depending on their characteristics [10].

2.1 Oriented Fuzzy Numbers

Ordered fuzzy numbers were distinguished by Kosiński [2, 3]. The author introduced them intuitively as a fuzzy number model. Ordered fuzzy number S was determined for any sequence of numbers (a, b, c, d) as an ordered pair of continuous real functions (f_S, g_S) defined on the interval $[0, 1]$, i.e. $f_S : [0, 1] \rightarrow UP_S$ and $g_S : [0, 1] \rightarrow DOWN_S$, where $UP_S = [a, b]$ and $DOWN_S = [c, d]$. Functions f_S and g_S are, respectively, called rising (UP) and falling (DOWN), which satisfy the conditions: $f_S(0) = a, f_S(1) = b, g_S(1) = c, g_S(0) = d$. The numbers that are defined as above described are called Kosiński's numbers.

Further research into fuzzy numbers and ordered fuzzy numbers revealed a significant drawback of Kosiński's theory, i.e. that there exist such ordered fuzzy numbers which, in fact, are not fuzzy numbers. Therefore, the notion of ordered fuzzy numbers must have been revised by Piasecki [8]. Ultimately, the drawbacks of Kosiński's theory led to narrowing the notion to Oriented Fuzzy Numbers (OFNs).

Definition 1 [8] For any monotonic sequence $(a, b, c, d) \in \mathbb{R}$, Oriented Fuzzy Number—OFN $\overleftrightarrow{\mathcal{L}} = \overleftarrow{\mathcal{L}}(a, b, c, d, S_L, E_L)$ is a pair composed of an orientation $\overrightarrow{a, d} = (a, d)$ and a fuzzy number defined by its membership function $\mu_{\mathcal{L}}(\cdot | a, b, c, d, S_L, E_L) \in [0, 1]^{\mathbb{R}}$

$$\mu_{\mathcal{L}}(x) = \mu_{\mathcal{L}}(x | a, b, c, d) = \begin{cases} 0, & x \notin [a, d] \equiv [d, a], \\ S_L, & x \in [a, b] \equiv]a, b[, \\ 1, & x \in [b, c] \equiv [c, b], \\ E_L, & x \in]c, d[\equiv [c, d], \end{cases} \tag{1}$$

where the starting function $S_L \in [0, 1]^{[a, b]}$ and the ending function $E_L \in [0, 1]^{[c, d]}$ are semi-continuous from above, monotone function satisfying the condition $[\mathcal{L}]_{0+} = [a, d]$.

2.2 Trapezoidal-Oriented Fuzzy Numbers

Following the notion of Oriented Fuzzy Numbers, we can distinguish a subgroup of Trapezoidal-Oriented Fuzzy Numbers (TrOFNs) which are a special presentation of

OFNs. TrOFNs, determined by (a, b, c, d) , are divided into positively and negatively oriented. In case of a positive orientation, we expect a rise in the value of an estimated TrOFN and vice versa, in case of a negative orientation, we expect a drop in the value of an estimated TrOFN which can be presented as follows:

$$a < d \rightarrow (\overrightarrow{a, d}), \tag{2}$$

$$a > d \rightarrow (\overleftarrow{a, d}). \tag{3}$$

Definition of TrOFNs proposed by Piasecki in [8] is as follows:

Definition 2 For any monotonic sequence $(a, b, c, d) \subset \mathbb{R}$, TrOFN $\overleftrightarrow{Tr}(a, b, c, d) = \overleftrightarrow{\tau}$ is OFN $\overleftrightarrow{\tau} \in \mathbb{K}_{Tr}$ is determined explicitly by its membership functions $\mu_T \in [0, 1]^{\mathbb{R}}$

$$\mu_T(x) = \mu_{Tr}(x|a, b, c, d) = \begin{cases} 0, & x \notin [a, d] \equiv [d, a], \\ \frac{x-a}{b-a}, & x \in [a, b[\equiv]a, b], \\ 1, & x \in [b, c] \equiv [c, b], \\ \frac{x-d}{c-d}, & x \in]c, d] \equiv [c, d[. \end{cases} \tag{4}$$

Arithmetic operations on fuzzy numbers such as an extended sum \boxplus and dot product \boxtimes are not straightforward and must be defined as follows:

$$\overleftrightarrow{Tr}(a, b, c, d) \boxplus \overleftrightarrow{Tr}(p - a, q - b, r - c, s - d) = \begin{cases} \overleftrightarrow{Tr}(\min\{p, q\}, q, r, \max\{r, s\}), & (q < r) \vee (q = r \wedge p \leq s), \\ \overleftrightarrow{Tr}(\max\{p, q\}, q, r, \min\{r, s\}), & (q > r) \vee (q = r \wedge p > s). \end{cases} \tag{5}$$

$$\beta \boxtimes \overleftrightarrow{Tr}(a, b, c, d) = \overleftrightarrow{Tr}(\beta \cdot a, \beta \cdot b, \beta \cdot c, \beta \cdot d). \tag{6}$$

2.3 Energy and Entropy Measures

Energy measure can be used to evaluate the level of ambiguity risk which results from unequivocal recommendations. The higher the level of the ambiguity, the higher the risk of choosing the wrong alternative decision among the available ones. Entropy measure can be used to evaluate the level of indistinctness resulting from the inability to distinguish between the recommended and non-recommended decisions.

Both of these measures are important indicators in decision-making approach because the imprecision risk is estimated as a sum of ambiguity and indistinctness levels. In case of TrOFN (a, b, c, d) , the entropy and energy measures are determined in a following manner, respectively:

$$d(TrOFN(a, b, c, d)) = \frac{1}{2} \cdot |d + c - b - a|, \tag{7}$$

$$e(\text{TrOFN}(a, b, c, d)) = \frac{1}{4} \cdot |d - c + b - a|. \tag{8}$$

In the paper, the results are presented by an imprecise assets benefit (also regarded as a relative profit) defined in a portfolio approach by a function $\varpi : (\mathbb{K}_{Tr})^2 \times [0, 1] \rightarrow \mathbb{K}_{Tr}$ as follows:

$$(K, L, \lambda) = (\lambda \boxplus K) \boxplus ((1 - \lambda) \boxplus L). \tag{9}$$

Theorem 1 For any real number $\lambda \in [0, 1]$, we have

– for any pair $(\overleftarrow{K}, \overleftarrow{L}) \in (\mathbb{K}_{Tr}^- \times \mathbb{K}_{Tr}^-) \cup (\mathbb{K}_{Tr}^+ \times \mathbb{R} \mathbb{K}_{Tr}^-) \times (\mathbb{K}_{Tr}^+ \cup \mathbb{R})$

$$d(\varpi(\overleftarrow{K}, \overleftarrow{L}, \lambda)) = \lambda \cdot d(\overleftarrow{K}) + (1 - \lambda) \cdot d(\overleftarrow{L}), \tag{10}$$

$$e(\varpi(\overleftarrow{K}, \overleftarrow{L}, \lambda)) = \lambda \cdot e(\overleftarrow{K}) + (1 - \lambda) \cdot e(\overleftarrow{L}), \tag{11}$$

– for any pair $(\overleftarrow{K}, \overleftarrow{L}) \in ((\mathbb{K}_{Tr}^+ \cup \mathbb{R}) \times \mathbb{K}_{Tr}^-)$

$$d(\varpi(\overleftarrow{K}, \overleftarrow{L}, \lambda)) \leq \begin{cases} \lambda \cdot d(\overleftarrow{K}) - (1 - \lambda) \cdot d(\text{Core}(\overleftarrow{L})), & \varpi(\overleftarrow{K}, \overleftarrow{L}, \lambda) \in \mathbb{K}_{Tr}^+ \cup \mathbb{R}, \\ (1 - \lambda) \cdot d(\overleftarrow{L}) - \lambda \cdot d(\text{Core}(\overleftarrow{K})), & \varpi(\overleftarrow{K}, \overleftarrow{L}, \lambda) \in \mathbb{K}_{Tr}^- \cup \mathbb{R}, \end{cases} \tag{12}$$

where $\text{Core}(A) = \{x : \mu_A(x) = 1\}$, $A \in F(\mathbb{R})$,

– for any pair $(\overleftarrow{K}, \overleftarrow{L}) \in ((\mathbb{K}_{Tr}^+ \cup \mathbb{R}) \times \mathbb{K}_{Tr}^-) \cup (\mathbb{K}_{Tr}^- \times (\mathbb{K}_{Tr}^+ \cup \mathbb{R}))$

$$e(\varpi(\overleftarrow{K}, \overleftarrow{L}, \lambda)) \leq \min \left\{ \lambda e(\overleftarrow{K}), (1 - \lambda) e(\overleftarrow{L}) \right\}. \tag{13}$$

For more information about TrOFNs, see [5, 6, 9, 11].

3 Methods

3.1 Oriented Present Value

In case of portfolio analysis, it is worth stressing that OFNs are more convenient than FNs. Moreover, in [6] the fuzzy present value was replaced by the closest approximation of changes in market prices given by the quoted levels of the analysed security prices. Such a replacement was possible due to the redefinition of the PV determined as a present equivalent of a pay-off in a given time in the future. Taking all the above into account, we can assume that imprecise PV may be defined by an OFN [5, 6]. Such PV will be called an Oriented PV (OPV) represented by a

monotonic sequence (V_s, V_f, V_l, V_e) given by the formula:

$$OPV = TrOFN(V_s, V_f, V_l, V_e), \quad (14)$$

where the monotonic sequence $(V_s, V_f, \check{P}, V_l, V_e)$ is determined by \check{P} —quoted price, $[V_s, V_e] \subset \mathbb{R}^+$ —interval of all PV possible values, $[V_f, V_l] \subset [V_s, V_e]$ —interval of all prices that do not significantly differ from the quoted price \check{P} .

If we expect an increase in price, the present value is determined by a positively oriented TrOFN. Japanese candle models are examples of trapezoidal OPVs [9]. In turn, if we expect a decrease in price, the present value is determined by a negatively oriented TrOFN. Moreover, if we assume a time horizon $t > 0$ of a fixed investment, the asset is described by an anticipated future value and an assessed present value. The profit from owning a given asset is a basic return rate given by the identity:

$$r_t = \frac{V_t - V_0}{V_0} = \frac{V_t}{V_0} - 1. \quad (15)$$

Assuming that future value (FV) is a random variable $\check{V}_t : \Omega \rightarrow \mathbb{R}^+$ where ω is a set of elementary states ω of the financial market and present value is identified by a given quoted price \check{P} , then the return rate is a random variable determined by

$$r_t(\omega) = \frac{\check{V}_t(\omega) - \check{P}}{\check{P}}. \quad (16)$$

Practitioners of financial markets usually try to determine the lack of sufficient knowledge (uncertainty risk) by the probability distribution of a return rate given by a function of a cumulative distribution $F_r(\cdot|\bar{r}) : \mathbb{R} \rightarrow [0, 1]$.

3.2 Expected Discount Factor

If we assume that the expected value \bar{r} of the distribution and the expected discount factor \bar{v} exist then we can denote the expected discount factor (EDF) as

$$\bar{v} = (1 + \bar{r})^{-1}. \quad (17)$$

Then, following this direction of reasoning and taking together (15) and (16) we get the formula for the return rate:

$$r_t = r_t(V_0, \omega) = \frac{\check{P} \cdot (1 + \bar{r}(\omega))}{V_0} - 1 \quad (18)$$

and the imprecise EDF $V : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ given as a unary operator transforming PV [7]:

$$V(V_0) = \left(\frac{\check{P} \cdot (1 + \bar{r})}{V_0} \right)^{-1} = \frac{\bar{v}}{\check{P}} \cdot V_0. \tag{19}$$

As the imprecise PV, given by Eq. (14), is estimated with the use of TrOFN, then the imprecise EDF $\overleftrightarrow{V}(\overleftrightarrow{PV})$ also is defined as follows:

$$\overleftrightarrow{V}(\overleftrightarrow{PV}) = \overleftrightarrow{Tr} \left(\frac{V_s \cdot \bar{v}}{\check{P}}, \frac{V_f \cdot \bar{v}}{\check{P}}, \frac{V_l \cdot \bar{v}}{\check{P}}, \frac{V_e \cdot \bar{v}}{\check{P}} \right) = \overleftrightarrow{V} \left(\overleftrightarrow{Tr} (V_s, V_f, V_l, V_e) \right). \tag{20}$$

Consequently, the entropy and energy measures derived from imprecise EDF $\overleftrightarrow{V}(\hat{S})$ can be determined by

$$d(\overleftrightarrow{V}(\hat{S})) = \left| \frac{(V_e + V_l - V_f - V_e) \cdot \bar{v}}{2\check{P}} \right|, \tag{21}$$

$$e(\overleftrightarrow{V}(\hat{S})) = \left| \frac{(V_e - V_l + V_f - V_e) \cdot \bar{v}}{4\check{P}} \right|. \tag{22}$$

3.3 Portfolio Approach

In the research, the portfolio is regarded as an arbitrary set of assets which is finite. An asset (also referred to as share or stock) is, in turn, comprehended as a fixed security in a long-term investment. Each security is characterised by three elements:

- its price $\check{P}_i \in \mathbb{R}^+$,
- its imprecise present value $\overleftrightarrow{PV}_i = \overleftrightarrow{Tr} (V_s^{(i)}, V_f^{(i)}, V_l^{(i)}, V_e^{(i)})$,
- its EDF \bar{v}_i defined by (17).

As the PV of a whole portfolio π^* is the sum of its elements determined by TrOFNs, i.e. the sums of PVs of rising securities π^+ and falling securities π^- we can determine that

$$\begin{aligned} \overleftrightarrow{PV}^* &= \overleftrightarrow{PV}^+ \boxplus \overleftrightarrow{PV}^- \\ &= \overleftrightarrow{Tr} (V_s^{(+)}, V_l^{(+)}, V_f^{(+)}, V_e^{(+)}) \boxplus \overleftrightarrow{Tr} (V_s^{(-)}, V_l^{(-)}, V_f^{(-)}, V_e^{(-)}). \end{aligned} \tag{23}$$

Finally, we can determine the imprecise EDFs for a given portfolio (π^+ and π^*) as follows:

$$\overleftarrow{\mathcal{V}}^+ = \bar{r} \left(\sum_{Y_i \in \pi^+} \frac{\bar{v}^+ \cdot p_i^{(+)}}{v_i} \cdot D_s^{(i)}, \sum_{Y_i \in \pi^+} \frac{\bar{v}^+ \cdot p_i^{(+)}}{v_i} \cdot D_f^{(i)}, \sum_{Y_i \in \pi^+} \frac{\bar{v}^+ \cdot p_i^{(+)}}{v_i} \cdot D_l^{(i)}, \sum_{Y_i \in \pi^+} \frac{\bar{v}^+ \cdot p_i^{(+)}}{v_i} \cdot D_e^{(i)} \right) \quad (24)$$

$$\overleftarrow{\mathcal{V}}^* = \overleftarrow{Tr} \left(D_s^*, D_f^*, D_l^*, D_e^* \right) = \left(\frac{\bar{v}^* \cdot q^+}{\bar{v}^+} \square \overleftarrow{\mathcal{V}}^+ \right) \boxplus \left(\frac{\bar{v}^* \cdot q^-}{\bar{v}^-} \square \overleftarrow{\mathcal{V}}^- \right), \quad (25)$$

where

- $M^+ = \sum_{Y_i \in \pi^+} \check{P}_i, M^- = \sum_{Y_i \in \pi^-} \check{P}_i, M^* = M^+ + M^-$ are the values of the portfolios π^+, π^- and π^* ;
- $p_i^{+/-} = \frac{\check{p}_i}{M^{+/-}}$ represent the shares of a given asset $Y_i \in \pi^{+/-}$, respectively, in an individual portfolio ($\pi^{+/-}$);
- $q_i^{+/-} = \frac{M^{(+/-)}}{M^*}$ represent the share $q^{+/-}$ of an individual portfolio $\pi^{+/-}$ in a portfolio π^* ;
- $\bar{v}^{+/-} = \left(\sum_{Y_i \in \pi^{+/-}} \frac{p_i^{+/-}}{v_i} \right)^{-1}$, $\bar{v}^* = \left(\frac{q^+}{\bar{v}^+} + \frac{q^-}{\bar{v}^-} \right)^{-1}$ are, respectively, the EDFs $\bar{v}^+/\bar{v}^-/\bar{v}^*$ of the portfolios π^+, π^- and π^* .

The formula for an imprecise EDF for a portfolio π^- is analogous to the formula of an imprecise EDF for a portfolio π^+ .

Eventually, the formulas for energy and entropy are defined as follows:

$$d \left(\overleftarrow{\mathcal{V}}^{+/-} \right) = \sum_{Y_i \in \pi^{+/-}} \frac{\bar{v}^{+/-} \cdot q_i^{(+/-)}}{\bar{v}_i} \cdot d \left(\overleftarrow{\mathcal{V}}(Y_i) \right), \quad (26)$$

$$e \left(\overleftarrow{\mathcal{V}}^{+/-} \right) = \sum_{Y_i \in \pi^{+/-}} \frac{\bar{v}^{+/-} \cdot q_i^{(+/-)}}{\bar{v}_i} \cdot e \left(\overleftarrow{\mathcal{V}}(Y_i) \right), \quad (27)$$

where the energy and entropy measures must meet the conditions, respectively:

$$d \left(\overleftarrow{\mathcal{V}}^* \right) = \leq \begin{cases} \frac{\bar{v}^* \cdot q^+}{\bar{v}^+} \cdot d \left(\overleftarrow{\mathcal{V}}^+ \right) - \frac{\bar{v}^* \cdot q^-}{\bar{v}^-} \cdot d \left(\text{Core} \left(\overleftarrow{\mathcal{V}}^* \right) \right), & \overleftarrow{\mathcal{V}}^* \in \mathbb{K}_{Tr}^+ \cup \mathbb{R}, \\ \frac{\bar{v}^* \cdot q^-}{\bar{v}^-} \cdot d \left(\overleftarrow{\mathcal{V}}^- \right) - \frac{\bar{v}^* \cdot q^+}{\bar{v}^+} \cdot d \left(\text{Core} \left(\overleftarrow{\mathcal{V}}^* \right) \right), & \overleftarrow{\mathcal{V}}^* \in \mathbb{K}_{Tr}^- \cup \mathbb{R} \end{cases} \quad (28)$$

and

$$e \left(\overleftarrow{\mathcal{V}}^* \right) = \leq \min \left\{ \frac{\bar{v}^* \cdot q^+}{\bar{v}^+} \cdot e \left(\overleftarrow{\mathcal{V}}^+ \right), \frac{\bar{v}^* \cdot q^-}{\bar{v}^-} \cdot e \left(\overleftarrow{\mathcal{V}}^- \right) \right\}. \quad (29)$$

4 Findings

In the conducted research, we analyse a portfolio consisting of the assets of five companies listed on the Warsaw Stock Exchange (WSE). Those companies belong to the construction sector and are included in WIG (main index on the Warsaw Stock Exchange). The portfolio will be analysed basing on two different approaches and therefore we will distinguish Portfolio π_1 (basing solely on TrOFNs) and Portfolio π_2 (basing on brokers' recommendations).

The detailed composition of the analysed Portfolio π^* along with the number of individual securities is presented in Table 1.

Basing on the closing of WSE session on 8 September 2021, for each analysed security, its PV equal to TrOFN PV_i (describing its Japanese candles) is determined.

A quoted price \hat{P}_Y of each single element of the observed portfolio becomes the initial price on the subsequent following day (9 September 2021).

The obtained results of Portfolio π^* —TrOFN PV, energy and entropy measures, quoted price—are shown in Table 2.

The first two companies (variables Y_{COM} and Y_{MST}) are characterised by positively oriented TrOFNs, hence the investor might expect the increase of their value in the future. That would suggest that in the future the investor will have an opportunity to sell the securities with the profit. The other three variables (Y_{PEK} , Y_{SEL} and Y_{SNK}) are described by negatively oriented TrOFNs. In this case, the investor can expect a drop in their value in the future. That would suggest a future opportunity to buy the shares of given companies in the future at a lower price. For each variable, the

Table 1 Composition of initial portfolio π^*

Variable	Name	Tick	No. of stocks
Y_{COM}	COMPREMIUM	COM	1600
Y_{MST}	MOSTOSTALZAB	MST	950
Y_{PEK}	PEKABEX	PEK	550
Y_{SEL}	SELENAFM	SEL	1005
Y_{SNK}	ŚNIEŻKA	SNK	120

Table 2 Results of portfolio π^*

Variable	Present value	Quoted price	Energy measure	Entropy measure
Y_{COM}	$\overleftrightarrow{Tr}(4.86, 4.95, 5.04, 5.06)$	5.02	0.0305	0.0038
Y_{MST}	$\overleftrightarrow{Tr}(1.85, 1.89, 2.03, 2.14)$	2.04	0.1007	0.0176
Y_{PEK}	$\overleftrightarrow{Tr}(23.50, 23.10, 22.90, 22.90)$	22.70	0.0168	0.0042
Y_{SEL}	$\overleftrightarrow{Tr}(29.00, 29.00, 27.90, 27.20)$	28.00	0.0495	0.0060
Y_{SNK}	$\overleftrightarrow{Tr}(84.20, 84.20, 82.00, 81.80)$	82.00	0.0268	0.0006

Table 3 Characteristics of Portfolio π^+ , π^- and π^*

Portfolio	EDF \bar{v}	Imprecise EDF \overleftrightarrow{V}
π^+	0.9854	\overleftrightarrow{Tr} (0.9134, 0.9263, 0.9580, 0.9716)
π^-	0.9675	\overleftrightarrow{Tr} (0.9882, 0.9839, 0.9562, 0.9428)
π^*	0.9704	\overleftrightarrow{Tr} (0.9760, 0.9745, 0.9565, 0.9475)

above-mentioned initial price for the following day is given. Also, energy and entropy measures are calculated for each individual security. The linear portfolio analysis is not possible for considered portfolio π^* .

The next step included computing the portfolios π^+ , π^- and π^* determined by TrOFN (\overleftrightarrow{PV}).

$$\overleftrightarrow{PV}^+ = \overleftrightarrow{Tr} (9533.50, 9667.50, 9992.50, 10129.00), \tag{30}$$

$$\overleftrightarrow{PV}^- = \overleftrightarrow{Tr} (52174.00, 51954.0, 50474.50, 49747.00), \tag{31}$$

$$\overleftrightarrow{PV}^* = \overleftrightarrow{Tr} (61707.50, 61621.50, 60467.00, 59876.00). \tag{32}$$

The $\overleftrightarrow{PV}^*$ is defined by a negatively oriented TrOFNs, and therefore we conclude that we are to expect a decrease in the value of the portfolio in the future. EDFs of the given portfolios π^+ , π^- and π^* as well as the values of imprecise EDF are given in Table 3.

The obtained data from Table 3 enabled the computation of the energy and entropy measures of EDF \overleftrightarrow{V}^* of Portfolio π^* which, respectively, reached the levels of $d(\overleftrightarrow{V}^*) = 0.0233$ and $e(\overleftrightarrow{V}^*) = 0.0026$.

Basing on the above premises from the TrOFNs analysis we would advise the investor to buy additional shares of Y_{COM} and Y_{MST} to sell them in the future at a higher price and to sell some shares of Y_{PEK} , Y_{SEL} and Y_{SNK} to buy them cheaper in some time.

4.1 Case Study—Portfolio π_1

Portfolio π_1 is the case in which the investors follow the above-presented recommendations of the TrOFNs (the recommendations of TrOFNs are dominant). Therefore, the investors decide to

1. sell 20 shares of SNK,
2. sell 50 shares of PEK,
3. sell 5 shares of SEL,

Table 4 Composition of Portfolio π_1^*

Variable	Name	Tick	No. of stocks
Y_{COM}	COMPREMIUM	COM	2000
Y_{MST}	MOSTOSTALZAB	MST	1000
Y_{PEK}	PEKABEX	PEK	500
Y_{SEL}	SELENAFM	SEL	1000
Y_{SNK}	ŚNIEŻKA	SNK	100

- 4. buy 50 shares of MST,
- 5. buy 400 shares of COM.

Basing on the closing of WSE session on 10 September 2021, for each analysed security, its PV equal to TrOFN $\overleftrightarrow{P\hat{V}}_i$ (describing its Japanese candles) is determined.

A quoted price \hat{P}_Y of each single element of the observed portfolio becomes the initial price on the subsequent following day (11 September 2021).

For each variable, the above-mentioned initial price for the following day is given. Also, energy and entropy measures are calculated for each individual security. The linear portfolio analysis is not possible for considered portfolio π_1^* (Table 4).

The next step included computing the portfolios π_1^+ , π_1^- and π_1^* determined by TrOFN ($\overleftrightarrow{P\hat{V}}$).

$$\overleftrightarrow{P\hat{V}}_1^+ = \overleftrightarrow{Tr}(29210.00, 29310.00, 29920.00, 30200.00), \tag{33}$$

$$\overleftrightarrow{P\hat{V}}_1^- = \overleftrightarrow{Tr}(30370.00, 30330.00, 29830.00, 29310.00), \tag{34}$$

$$\overleftrightarrow{P\hat{V}}_1^* = \overleftrightarrow{Tr}(59580.00, 59640.00, 59750.0059750.00). \tag{35}$$

The $\overleftrightarrow{P\hat{V}}_1^*$ is defined by a positively oriented TrOFNs, and therefore we expect a further increase in the value of the portfolio in the future. EDFs of the given portfolios π_1^+ , π_1^- and π_1^* , as well as the values of imprecise EDF, are given in Table 5.

The obtained data from Table 5 enabled the computation of the energy and entropy measures of EDF $\overleftrightarrow{V}_1^*$ of Portfolio π_1^* which, respectively, reached the levels of $d(\overleftrightarrow{V}_1^*) = 0.02538$ and $e(\overleftrightarrow{V}_1^*) = 0.000248$.

Table 5 Characteristics of Portfolio π_1^+ , π_1^- and π_1^*

Portfolio	EDF \bar{v}_1	Imprecise EDF \overleftrightarrow{V}_1
π_1^+	0.9523	$\overleftrightarrow{Tr}(0.9314, 0.9350, 0.9549, 0.9642)$
π_1^-	0.9945	$\overleftrightarrow{Tr}(0.9784, 0.9771, 0.9608, 0.9441)$
π_1^*	0.9733	$\overleftrightarrow{Tr}(0.9547, 0.9559, 0.9579, 0.9579)$

4.2 Case Study—Portfolio π_2

Portfolio π_2 is the case in which the investors decided to seek additional support. This support comes in the form of brokers' recommendations as brokerage houses specialise in stock investments. In this case study, the recommendations of brokers are dominant. The official recommendations of brokers issued for analysed companies in the given time are presented in Table 6.

Comparing the recommendations in Strategy 1 (Portfolio π_1) and in Strategy 2 (Portfolio π_2) we can notice that the recommendations of brokers for analysed companies in the given time differ insignificantly from the recommendations reached on the basis of TrOFNs. Therefore, in case that the recommendations in both strategies are identical, the investor takes the same decision regarding individual company. However, in this strategy (if the recommendations differ), the investor chooses the recommendations of brokers.

Therefore, the investors decides to

1. sell 20 shares of SNK,
2. buy 50 shares of MST,
3. buy 5 shares of SEL,

while the number of shares of the remaining two companies stays the same. Finally, we obtain the following Portfolio π_2^* presented in Table 7.

Basing on the closing of WSE session on 10 September 2021, for each analysed security, its PV equal to TrOFN PV_i (describing its Japanese candles) is determined.

A quoted price \check{P}_Y of each single element of the observed portfolio becomes the initial price on the subsequent following day (11 September 2021).

Table 6 Recommendations of brokers for analysed companies

Company	Recommendation	Date of recommendation
ŚNIEŻKA	SELL	2021-09-08
MOSTALZAB	BUY	2021-09-07
PEKABEX	HOLD	2021-09-07
SELENAFM	BUY	2021-09-07
COMPREMUM	HOLD	2021-09-06

Table 7 Composition of portfolio π_2^*

Variable	Name	Tick	No. of stocks
Y_{COM}	COMPREMIUM	COM	1600
Y_{MST}	MOSTOSTALZAB	MST	1000
Y_{PEK}	PEKABEX	PEK	550
Y_{SEL}	SELENAFM	SEL	1010
Y_{SNK}	ŚNIEŻKA	SNK	100

Table 8 Results of portfolio π_2^*

Variable	Present value	Quoted price	Energy measure	Entropy measure
Y_{COM}	$\overleftrightarrow{Tr}(5.06, 5.06, 5.16, 5.20)$	5.18	0.0221	0.0018
Y_{MST}	$\overleftrightarrow{Tr}(2.27, 2.23, 2.13, 2.11)$	2.08	0.0597	0.00069
Y_{PEK}	$\overleftrightarrow{Tr}(21.90, 22.10, 22.80, 23.20)$	22.80	0.0419	0.0063
Y_{SEL}	$\overleftrightarrow{Tr}(28.10, 28.10, 27.70, 27.20)$	27.60	0.0225	0.0043
Y_{SNK}	$\overleftrightarrow{Tr}(81.40, 81.40, 82.00, 82.00)$	82.00	0.0070	0.0000

Furthermore, we calculate the portfolios π_2^+ , π_2^- and π_2^* determined by TrOFN (\overleftrightarrow{PV}).

$$\overleftrightarrow{PV}_2^+ = \overleftrightarrow{Tr}(28281.00, 28391.00, 28996, 29280.00), \tag{36}$$

$$\overleftrightarrow{PV}_2^- = \overleftrightarrow{Tr}(30651.00, 30611.00, 30107.00, 29582.00), \tag{37}$$

$$\overleftrightarrow{PV}_2^* = \overleftrightarrow{Tr}(58932.00, 59002.00, 59103.00, 59103.00). \tag{38}$$

The $\overleftrightarrow{PV}_2^*$ is defined by a positively oriented TrOFNs, and therefore we expect a further increase in the value of the portfolio in the future.

The obtained data enabled the computation of the energy and entropy measures for individual securities which are presented in Table 8.

The entropy and energy measures of EDF $\overleftrightarrow{V}_2^*$ of Portfolio π_1^* which, respectively, reached the levels of $d(\overleftrightarrow{V}_2^*)$ and $e(\overleftrightarrow{V}_2^*) = 0.000292$.

5 Conclusions

Portfolio analysis and management frequently use oriented fuzzy numbers (OFNs) which include information uncertainty and imprecision related to financial market. That enables the utilisation of an expected fuzzy discount factor which leads to an imprecise present value (PV). Finally, we obtain recommendations (buy, sell, accumulate, reduce) for individual stocks included in the portfolio. Similarly, agents of stock-brokerage regularly issue analogous recommendation basing on their knowledge, experience and quantitative analysis which also are imprecise.

The main purpose of the paper is to analyse the published recommendations of stockbrokers, their accurateness and actual impact on the operations on portfolio including stocks of companies in construction sector. The created portfolio includes securities identified by a present value assessed with the use of trapezoidal-oriented fuzzy number (TrOFNs) in times of issuance of the recommendation. All theoretical considerations are illustrated by an empirical case study. The elements of portfolio

are positively or negatively oriented. The main objective of this paper was to compare the usefulness of the recommendations made by stockbrokers and suggested by the fuzzy portfolio analysis which was conducted with the use of fuzzy discount factor.

The results show that the fuzzy portfolio analysis was more accurate for the companies in construction sector assessed by a forecast of the closest change in prices than the recommendations of the brokers.

Using OFNs an initial decrease in the value of a portfolio was anticipated. The decrease occurred in both of analysed strategies; however, in case of the second strategy, the decrease was of a smaller value.

The case study shows that recommendations obtained with the use of OFNs are better if we consider energy and entropy measures (their values are smaller). It can also result from the fact that OFN recommendation, unlike most of brokers' recommendations, are short-term recommendations.

The results based on OFNs are better because they are less imprecise.

References

1. Dubois, D., Prade, H.: Operations on fuzzy numbers. *Int. J. Syst. Sci.* **9**, 613–629 (1978)
2. Kosiński, W., Prokopowicz, P., Ślęzak D.: Fuzzy numbers with algebraic operations: algorithmic approach. In: Kłopotek, M., Wierzchoń, S.-T., Michalewicz, M., (eds.) *Proceedings of the IIS'2002*, pp. 311–320. Physica Verlag, Heidelberg (2002)
3. Kosiński, W.: On fuzzy number calculus. *Int. J. Appl. Math. Comput. Sci.* **16**(1), 51–57 (2006)
4. Łyczkowska-Hanćkowiak, A.: On application of oriented fuzzy numbers for imprecise investment recommendations. *Symmetry* **12**(10) (2020). <https://doi.org/10.3390/sym12101672>
5. Łyczkowska-Hanćkowiak, A.: The use of trapezoidal oriented fuzzy numbers in portfolio analysis. *Symmetry* **13**(9), 1722 (2021). <https://doi.org/10.3390/sym13091722>
6. Łyczkowska-Hanćkowiak, A., Piasecki, K.: The present value of a portfolio of assets with present values determined by trapezoidal ordered fuzzy number. *Oper. Res. Decis.* **28**(2), 41–56 (2018). <https://doi.org/10.5277/ord180203>
7. Łyczkowska-Hanćkowiak, A., Piasecki, K.: Portfolio discount factor evaluated by oriented fuzzy numbers, In: Fejfar, J., Hruška, M., (eds.) *Proceedings of the 39th International Conference Mathematical Methods in Economics MME 2021*, pp. 299–304. Czech University of Life Sciences Prague, Prague, Czech Republic (2021)
8. Piasecki, K.: Revision of the Kosiński's theory of ordered fuzzy numbers. *Axioms* **7**(1) (2018). <https://doi.org/10.3390/axioms7010016>
9. Piasecki, K., Łyczkowska-Hanćkowiak, A.: Representation of Japanese candlesticks by oriented fuzzy numbers. *Econometrics* **8**(1) (2019). <https://doi.org/10.3390/econometrics8010001>
10. Piasecki, K., Łyczkowska-Hanćkowiak, A.: Oriented fuzzy numbers versus fuzzy numbers. *Mathematics* **9**(3) (2021). <https://doi.org/10.3390/math9050523>
11. Piasecki, K., Łyczkowska-Hanćkowiak, A.: Imprecision indexes of oriented fuzzy numbers. In: Atanassov, K.-T., et al. (eds.) *International Workshop on Intuitionistic Fuzzy Sets and General Nets—IWIFSGN 2019, Artificial Intelligence and Soft Computing*, vol. 1308, pp. 1–13 (2021). https://doi.org/10.1007/978-3-030-77716-6_10
12. Piasecki, K., Wójcicka-Wójtowicz, A.: Application of the oriented fuzzy numbers in credit risk assessment. *Mathematics* **9**(5) (2021). <https://doi.org/10.3390/math9050535>
13. Wójcicka-Wójtowicz, A.: Influence of imprecision on credit risk assessment—Case study. In: *Hradec Economic Days 2020, Proceedings of the International Scientific Conference Hradec Economic Days 2020*, pp. 871–880 (2020). <https://doi.org/10.36689/uhk/hed/2020-01-098>

14. Wójcicka-Wójtowicz, A., Piasecki, K.: A scale of credit risk evaluations assessed by ordered fuzzy numbers, In: SSRN Electron. J. 1–7 (2019). <https://doi.org/10.2139/ssrn.3459822>
15. Wójcicka-Wójtowicz, A.: Can experts' knowledge in eNS inspire efficient classification of potential debtors?. In: Proceedings of the 38th INTERNATIONAL Conference on Mathematical Methods in Economics, pp. 650–655. Brno, Czech Republic (2020)

Time Series Analysis Applications

Automatic Clustering for Seasonal Time Series Based on Entropy



Miguel Ángel Ruiz Reina 

Abstract Automatic clustering for seasonal time series based on entropy is a tool developed to understand decision-making behaviours for economic agents. An unsupervised learning system reduces information and is a powerful statistical learning tool. This method is a multiple-choice classification solution under uncertain environments. The empirical application is in the tourist accommodation market, where international tourists must choose various accommodation options (hotels, tourist apartments, campsites and rural apartments). Seasonal uncertainty for offers can solve information gaps in understanding human behaviour. The three-dimensional information of spatial extension, spatial location and temporal extension is offered for the Spanish tourist market of foreigners who visit the Spanish Autonomous Communities from January 2001 to June 2022. The results have revealed similarities and dissimilarities among the analysed Spanish regions depending on the seasonal period. In addition, the internal verification criteria have allowed us to quantify similarities in intragroup behaviour as an added value to this study.

Keywords Time series · Clustering · Shannon entropy

1 Introduction

The economy has been connected to physics; many economic elements and people are connected, generating a high probability of chaos in each system. Understanding human behaviour patterns reduces uncertainty and reduces clutter in complex social systems. The ultimate goal of managing complex realities is to control the systems, optimise them, and reduce entropy. In this sense, unsupervised learning clustering of time series with seasonality provides tools to economic agents to quantify the decision-making process. However, modelling and clustering techniques vary depending on the problem or data set to be analysed; these techniques can affect

M. Á. R. Reina (✉)

Department of Applied Economics (Statistics and Econometrics), University of Malaga, Malaga, Spain

e-mail: ruizreina@uma.es

the results and the process of understanding decision-making [1]. In this order, this research proposes a novel unsupervised clustering system for time series with cyclical fluctuations in contexts of uncertainty for multiple choice. The proposed application aims to understand a disordered time-series system by making unconnected decisions and creating an order based on the seasonal automatic clustering system for spatio-temporal data [2].

Classification algorithms are not new in uncertain environments; the research directions are varied, highlighting: the granulation of systematic procedures, extensive data processing and adequate selection of characteristics and methods to find similar groups. Information theory allows machine learning algorithms to classify based on entropy and even find outliers. The problem can be divided roughly into pattern detection techniques; supervised machine learning, unsupervised learning and finding potential anomaly issues. The challenge of our work is to develop an unsupervised model according to similarities among groups and differences, although without prior knowledge of the categories, combining pre-processing, seasonal entropy calculations, time-series techniques and machine learning analysis. Shannon's concept of entropy (1948) measures uncertainty, and classification algorithms allow knowing spatio-temporal information for decision-making [4]. However, we will provide the scientific literature with automatic classification and internal verification techniques, ensuring that data partitions show within-group similarities and out-of-group differences.

In the contextualisation of our modelling, the main idea is to provide a nonparametric estimation of space–time information based on the measurement of seasonal entropy for the decision-making of economic agents. We establish an automatic clustering model based on specific stochastic processes' degree of uncertainty and randomness. In this way, without prior knowledge, we will find economic agents' behaviour patterns and habits. The classification system allows time-series researchers to reduce the information to smaller dimensions for data exploitation.

The empirical work will deal with how foreign tourists visiting Spain decide on tourist accommodation; we can indicate that uncertainty is created multifactorial, as we will see in the methodological section. The lack of knowledge presents the degree of anticipation before making the accommodation decision among hotels, tourist apartments, campsites and rural apartments for the regions of Spain (Autonomous Communities). The strategic and economic relevance of the Spanish tourism industry is represented by 12% of the Gross Domestic Product (GDP) in 2019, its loss of weight in the Spanish economy due to the crisis caused by COVID-19 and its resurgence after the crisis period. Measuring this uncertainty and its spatio-temporal grouping will help economic agents make efficient decisions. In this way, it will try to recover entry levels of foreign tourists in Spain similar to those of 2019, when it represented the second largest number of people in the world (82.7 million) of the arrival of people only behind France (89.4 million). Ahead of the world giant USA (79.6 million), in fourth position China with 62.9 million, in fifth Italy, with 62.1 and sixth position Turkey with 45.7 million people received [5]. COVID-19 has harmed tourism activity in 2020 and 2021. So much so that the sector's contribution went

from 12% of GDP in 2019 to 7.4% in 2021. It is expected that in 2022 the situation of this industry will improve and, with it, the contribution to national wealth [6].

The book chapter is structured as follows: Sect. 2 deals with the literature review, studying state of the art in depth, allowing us to understand and propose the new methodology; Sect. 3 completely describes the method for unsupervised entropy clustering for seasonal data. The novel clustering system contributes to the dimensional reduction of time series and its subsequent empirical exploitation; the data and empirical application section is the fourth section, and it describes the natural results of the study. Data collection empirically supports the decision-making process for theoretical researchers and practitioners; Sect. 5 ends with the revealing conclusions of the study.

2 Literature Review

Clustering metric methods are crucial for many real-world applications, and distance metrics provide learning models with better performance than is generally achieved [7]. Unsupervised learning automatically generates categorisation problems into new categories for the same attributes [8]. Initially, cluster analysis was investigated with cross-sectional data; in the 1960s, issues of the number of clusters, better clustering algorithms and answering questions about outliers were solved [4]. Subsequently, the development of data Extraction, Transformation and Loading (ETL) techniques led to an excellent availability that was unsuspected at the beginning of the investigation [9]. The dimensional reduction applied to time series seasonal allows to improve the following processes: understand the business problem, understand the data, prepare the data, build a model, evaluate and finally deploy.

In recent years, the increasing availability of time series, classification and reduction techniques has attracted increasing interest. Initially, a more significant contribution of time-series univariate clustering techniques was found, but research has allowed the development of multivariate clustering techniques to remain challenging [10]. In this sense, the dimensionality of the series is related to time; additionally, the analysis can be made up of several values that change in the same time scale; in this case, it would be called multivariate time series. Clustering techniques are variants for time series reducing dimensionality; they constitute a wide area of research [10]. The techniques that have been recently developed are based on algorithms: connectivity-based clustering [11], centroid-based clustering [12], distribution-based clustering [13] and density-based clustering [14]. The techniques used are broad and relevant, where the primary classification of clustering algorithms could be cited as follows: time-series clustering by features; the clustering approach is based on stochastic time-series models, density function and forecast density, functional method, splines, copulas or tail dependence and series transformations [4].

The spatio-temporal analysis depends on the context of the data set, the objective and the available data's taxonomy. The spatio-temporal dataset clustering methods

can be as follows: Descriptive and generative model-based clustering, Distance-based clustering methods, Density-based methods and the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) family, visual-aided approaches, micro clustering methods, and flocks and convoy, important places and borderline cases: patterns [1].

In particular, with the help of spatial statistical tools based on entropy and information theory, the analysis and interpretation fill the literature gaps not included in the previous paragraphs. Entropy-based uncertainty for space–time analysis has been used in geosciences [15], categorical variables of interest on a dataset linked to climate change and evolution of land cover types in Nordic areas [16], for non-Gaussian spatial random fields, are explored for spatial field reconstruction in environmental and sensor network monitoring [17] o analyse traffic predictions in smart urban areas [18]. For space–time data with seasonal cycles, the grouping method called Neighbourhoods Seasonal Entropy based on Median Seasonal Clustering Entropy (*MdSCENsrjt*) was developed, unsupervised modelling process that starts from complete information gaps and uses entropy as a measure of uncertainty to classify information [4]. This learning method is a sequential process that analyses the binary process’s decision-making. However, the number of clusters is open to the nature of the data set. Consequently, after the literature review, our proposal could improve the different clustering methods based on automatic groups for spatio-temporal data with cyclical fluctuations.

After reviewing the literature, in our work, we consider a gap in the literature for the concept of multichoice of Shannon’s entropy (1948) to make classifications and reductions according to the recognition of seasonal patterns. In particular, we propose a new method that classifies time series according to repetitive cyclical flows based on the distance to a reference series in contexts of knowledge gaps before analysis. We call it Automatic Clustering Entropy Seasonal (*AutoCEs*), and it is developed in a high triple dimension of spatial extension, spatial location and temporal extension to achieve information reduction. This contribution allows knowing the differences among the dynamic behaviour of an aggregate series and the different series that compose it.

3 Methods

This study proposes an automatic unsupervised clustering procedure for time series with cyclic fluctuations based on multi-optional entropy called *AutoCEs*. This assembled algorithm allows information to be grouped in contexts of prior global uncertainty about the space–time data set. This learning statistics technique allows for obtaining cyclical information based on the concept of Shannon’s Information Theory (1948). An overview of this proposal is included in Fig. 1, presenting the classification phases for High-Dimensional Time Series (HDTS): (1) spatio-temporal data pre-processing; for Reduction-Dimensional Time Series (RDTS): (2) Entropy modelling

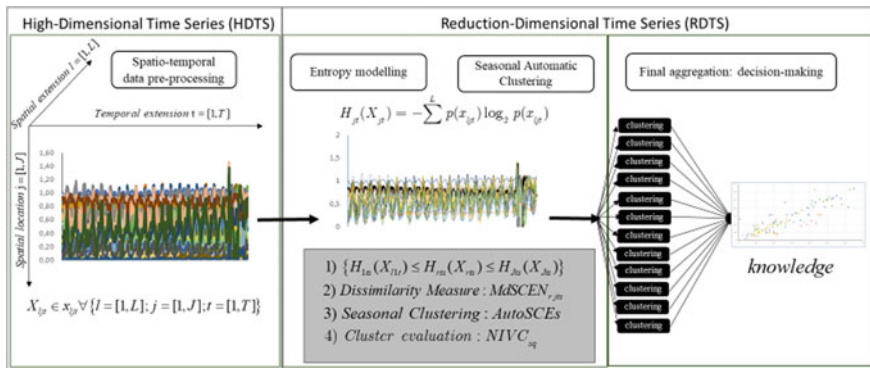


Fig. 1 AutoCEs: phases for High-Dimensional Time Series (HDTs) to Reduction-Dimensional Time Series (RDTS). Own elaboration

and Seasonal Automatic Clustering and (3) final aggregation: decision-making. Finally, we get the aggregation, and the knowledge is used in decision-making.

The *AutoCEs* process with seasonal fluctuations is based on the concept of Shannon entropy as a measure of uncertainty. We will divide our analysis process into two large blocks: HDTs and RDTS. Initially HDTs, the spatial information is disordered and contains hidden theoretical details on the original series to be analysed. This first process consists of ETL data. The second block of the analysis, and more important in this work, is RDTS, where the series is transformed into knowledge for decision-making. The series is grouped, and the uncertainty is quantified with Information Theory, mainly working with the original concept of Shannon entropy. Once the entropy has been quantified, we develop automatic clustering modelling for machine learning techniques. In the last step, clusters classified according to seasonal, cyclical fluctuations are obtained to make decisions efficiently for the field of application. In the following subsections, the points of the automatic decision algorithm will be detailed.

3.1 High-Dimensional Time Series (HDTs): Spatio-Temporal Data Pre-processing

Formally the set of $x_{ljt} \in X_{ljt} \forall (l = 1, 2, \dots, L; j = 1, 2, \dots, J; t = 1, 2, \dots, T)$ where l represents the options of the variables x in a space j for a temporal analysis period of each t in the data set (temporal space t is divided into seasonal flows s). We assume that X_{ljt} , as a collection of available data, the length will depend on the data engineering. Note that we can store data for variables with different samples, but for homogeneity in the study, we will assume the same length for each x_{ljt} . For a deep understanding, it is recommended to know the nature of the data, so prior

visualisation and analysis tools are necessary to understand the process [19]. This work models *AutoCEs* in the following subsections to reduce RDTs data, supporting our study compared with several unsupervised dissimilarity clustering methods such as Median Seasonal Clustering Entropy (*MdSCENsrjt*) [4].

Reduction-Dimensional Time Series (RDTs): Uncertainty, Clustering and Knowledge. Depending on the selected data set, a dimensional reduction can add value whatever the objective or scheme adopted for RDTs. In particular, spatio-temporal analysis is a large container, including several kinds of data that exhibit properties and opportunities for knowledge extraction [1]. Although clustering algorithms vary according to the technique [20], we provide an automatic unsupervised cluster number technique in this work. We will define the following intermediate steps: entropy modelling and automatic clustering.

Entropy modelling: the initial objective is to classify the temporal data set X_{ljt} , where we can define entropy measurement based on the portions of space–time. Shannon’s entropy measures the degree of disorder in spatio-temporal decision-making in a multifactorial context, represented as follows:

$$H_{jt}(X_{jt}) = - \sum_{l=1}^L p(x_{ljt}) \log_2 p(x_{ljt}) \quad (1)$$

The proportions of each series are obtained $p(x_{ljt}) = 0 \leq (x_{ljt}) / \left(\sum_{l=1}^L x_{ljt} \right) \leq 1$

and verified $\sum_{l=1}^L p(x_{ljt}) = 1$ for measurement in discrete times. The entropy vectors

would be organised for each temporary space $\overrightarrow{H_{jt}(X_{jt})}$ for $j \in \mathbb{R}^j$.

Once uncertainty has been defined as a pattern before grouping, it is convenient to indicate the group by uncertainties where the time series $H_{jt}(X_{jt}) = 0$ will represent a series with little variability in decision-making, and the maximum uncertainty will be found with values far from zero $\arg \max H_r(x_{rt})$. We will organise the information based on a reference series complying $0 \leq H_r(x_{rt}) \leq \arg \max H_r(x_{rt})$. *AutoCEs* will allow us to arrange the groups unsupervised around the proximity to the reference value $H_r(x_{rt})$ and dimensional reduction related to a seasonal pattern s .

Automatic clustering—*AutoCEs*: The seasonal sequential automatic cluster selection process begins after the preliminary phases of spatio-temporal data pre-processing and entropy measurement s . We will divide this last phase into the necessary sequences to obtain the final knowledge: sequence 1, the reference series and seasonal grouping will determine the centroid of the clusters; sequence 2, measures of dissimilarity; sequence 3, automatic clustering and clustering check. After this sequential automated work, we would enter the final phase of aggregation and efficient decision-making in the fields of application [1].

Part 1: having reduced the high-dimensional matrix X_{ljt} in a reduced entropy vector $\overrightarrow{H_{jt}(X_{jt})}$, it is convenient to work with the latter and find a time reference series $H_r(x_{rt})$ to be the centroid of the automatic clustering. In our paper, we will

work with monthly cyclical fluctuations, being $s \in [1, 12]$. However, the seasonal space could vary for quarters $s \in [1, 4]$ or for daily $s \in [1, 365]$. After this, we regroup the uncertainty series seasonally $H_{jts}(x_{jts})$, being $j = 1, 2, \dots, J$ spatial location, $t = 1, 2, \dots, T$ the temporal extension of data and $s = 1, 2, \dots, S$ the temporary subdivision t . The centroid value of the j seasonal time series entropy is the seasonal reference series, $H_{rts}(x_{rts})$ and the rest of the entropy series to classify would be $H_{jts}(x_{jts})$.

Part 2: Distance Measures. Calculating the distances with $H_{rts}(x_{rts})$ it is one of the pillars of any clustering algorithm. To classify and avoid measures of outliers or extreme values, we will use the concept of Median Seasonal Clustering Entropy ($MdSCEN_{rjts}$), which measures the Euclidean distance among the seasonal entropy series j and the seasonal reference series. Mathematically, the measure of seasonal difference according to the j series is defined as follows:

$$MdSCEN_{rjts} = Median\left(\sqrt{(H_{rts}(x_{rts}) - H_{jts}(x_{jts}))^2}\right) \quad (2)$$

Part 3: Automatic Seasonal Clustering Entropy. *AutoCEs* is an automatic classification rule that finds the optimal number of clusters for a series of seasonal data. A sample $N = \sum_{i=0}^{AutoSCEs} \binom{AutoSCEs - 1}{i} = (1 + 1)^{AutoSCEs - 1}$ contains the binomial coefficient for *AutoCEs* for N observations (numbers of entropy series once RDTs is done) for locations j and seasons s , where the cluster number i includes the number of samples. Finally, applying logarithms shows that the optimal number of clusters is expressed by $AutoSCEs = 1 + \log_2(N)$.

Part 4: Cluster evaluation. The clustering process is commonly considered unsupervised for this reason; the internal evaluation is usually entirely subjective; some validation mechanism is required to evaluate the quality of the obtained cluster [19]. The research areas largely determine the Cluster Validity Indices (CVIs) [21, 22]. The CVIs we propose are a relative internal validation in which we seek to verify the similarity of the cluster members. In particular, we work with the Neighbourhood's Internal Verification Coefficient— $NIVC_{sq}$ [4]. Finally, this task is subdivided into finding the width of the interval and then deriving $NIVC_{sq}$:

- The maximum distance intracluster is defined as the breadth of a cluster i $BC_i \rightarrow 0$ if $N \rightarrow \infty$
- The following expression will determine the Neighbourhood's Internal Verification Coefficient:

$$NIVC_{sq} = \frac{\max[d(MdSCEN_{srjt}^u, MdSCEN_{srjt}^l)]}{BC_i} \quad (3)$$

The numerator expression $\max[d(MdSCEN_{srjt}^u, MdSCEN_{srjt}^l)]$ includes the elements of the cluster found and may coincide with the maximum value of the breadth of a group $\left(\max[d(MdSCEN_{srjt}^u, MdSCEN_{srjt}^l)] \leq BC_i\right)$. Verifying

$0 \leq NIVC_{sq} \leq 1$, the most remarkable intra-group similarity will occur when $NIVC_{sq} = 0$; otherwise, we will find greater diversity with $NIVC_{sq} = 1$. Under the internal verification criterion, there may be limited values that meet the conditions. However, we consider this study a valid tool to establish clusters and eliminate acyclicity to make objective decisions.

This clustering $NIVC_{sq}$ allows the series to be grouped objectively based on seasonal entropy. Likewise, it is possible to identify the outliers of interest in different research fields. In our case, we will use a definition that is easy to interpret, and it is the use of divided intervals in neighbourhoods for non-grouped data. There are optimisation arguments and more in-depth readings in the literature, but it is not within the scope of this work [21–23].

Final aggregation: decision-making. The interpretation of clustering technique consists of interpreting the patterns identified in the algorithm. The identified space–time entropy trajectories allow us to know how the information systems transmit information to the cluster. Designs allow identifying laws and rules about trends and prediction, characterising individual and group behaviour [24]. Finally, the neighbourhood-based clusters are obtained: $cluster_q = cluster_1, cluster_2, \dots, cluster_Q$.

The criterion based on medians allows overcoming statistical problems based on outliers that could distort the cluster’s seasonal interpretation. This methodology proves its value in understanding the spatial uncertainty among two time series for different series j . In the next section, we will study the actual use of this analysis method in uncertainty models applied to the tourism sector.

4 Empirical Results

In this fourth empirical section of this work, we connect entropy and information with the automatic clustering algorithm developed. The entropy clustering of a macrostate (high dimension) will allow us to reduce the data to a microstate (dimensional reduction) for the understanding of the decision-making of economic agents based on behaviour patterns [25]. The empirical results show a work scheme described in Fig. 1. Next, we will detail the data pre-processing, uncertainty modelling and cluster processing for the sample period analysed. The empirical results obtained from the methodological application exposed above come from the Spanish hotel occupancy survey; in particular, it is data on overnight stays of foreigners in Spain from January 2001 to June 2022 collected by official Spanish statistics.

Data pre-processing. In this paper, the temporal database used comes from the Spanish National Institute of Statistics (INE)—it contains multichoice data from the spatial extension $l = (1 = \text{overnight stay in hotels}, 2 = \text{campsites}, 3 = \text{tourist apartments and } 4 = \text{rural apartments})$ according to Spanish Autonomous Communities from January 2001 to June 2022 ($t = 1, 2, \dots, T$). The data with 20,640 observations represent the spatial location j of the overnight stays by foreign tourists visiting Spain for $j = (1 = \text{Andalusia}, 2 = \text{Aragon}, 3 = \text{Asturias}, 4 = \text{Balearic Islands (Balears)})$,

5 = Canary Islands, 6 = Cantabria, 7 = Castile Leon (C. Leon), 8 = Castilla la Mancha (C. Mancha), 9 = Catalonia, 10 = Valencia, 11 = Extremadura, 12 = Galicia, 13 = Madrid, 14 = Murcia, 15 = Navarra, 16 = Basque Country, 17 = La Rioja) and the added variable called “Total” is the reference series r , variable “Total” collects all the seasonal and cyclical behaviours of all the regions analysed. INE data reveal seasonal behaviour and tourist flows with similar patterns by geographic area. The distances of each nationality j concerning the total number of overnight stays in Spain will be the reference measure r (denoted as *Total* in the rest of the document). We indicate that we group the seasonal classification values s = January, February, March, ..., December. In the next section on entropy modelling, we will represent the entropy time series for all the analysed time series.

Uncertainty modelling. Figure 2 represents the spatio-temporal uncertainty modelling for Autonomous Communities; this is the dimensional reduction of the original 20,640 observations collected in the sampling period. The multi-optional entropy function results show a cyclical behaviour with a periodicity of less than one year (monthly seasonality), compatible with the seasonality of the time series for the nationalities of origin described above. Based on our study, we have selected the reference series r as the one called Total. This series shows the global behaviour of all series, and the measurement/comparison of this series with the rest will allow us to cluster the information before decision-making.

Automatic Clustering Processing. The automated clustering processing results are presented in Table 1 analyses in detail grouped for January; the results of the automatic procedure have been a maximum of seven clusters in which finally, in the last group, zero elements have been included. The constant width of the intervals is 0.1176, starting with the lowest value of 0.0493 and ending with 0.8724; the number of members (for the 17 regions or Autonomous Communities) that make up the automatic seasonal grouping are: cluster 1 are two elements (*Andalucia*,

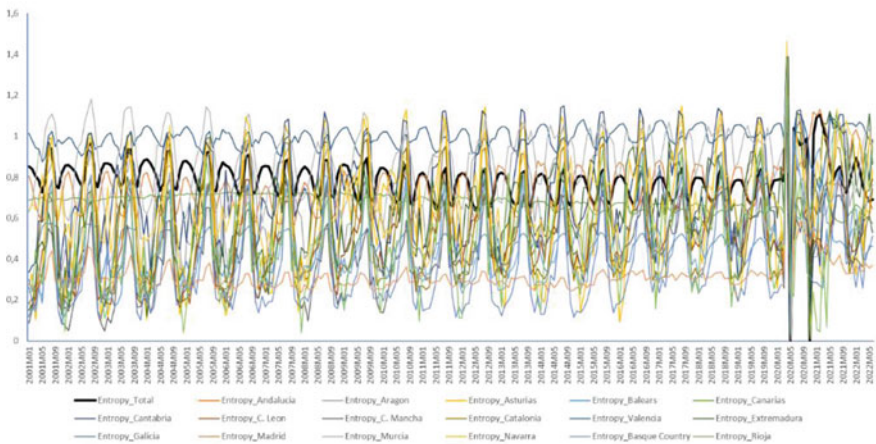


Fig. 2 Time series of spatial location entropy j from January 2001 to June 2022. Own elaboration

Canary Islands); cluster 2 are two elements (*Murcia, Valencia*); cluster 3 are four elements (*Navarra, Cantabria, Aragon, Extremadura*); cluster 4 are three elements (*Balears, C. Leon, Catalonia*); cluster 5 are five elements (*C. Mancha, Basque Country, Madrid, Asturias, Rioja*); cluster 6 is an element (*Galicia*); cluster 7 being an empty set proposed by the automatic system.

Regarding the similarity of seasonal behaviour of the series, we can say that the verifier $NIVC_q$ in all groups is less than one ($NIVC_q \leq 1$); in this way, we quantify dissimilarities among elements of the same cluster. Note that the groups with the most remarkable disparity accepted by the automatic methodology among features system are $NIVC_1 = 0.5090$, $NIVC_3 = 0.5416$, and $NIVC_5 = 0.56$. Cluster 2 presents a value of $NIVC_2 = 0.1368$, and cluster 4 has a dissimilarity of $NIVC_4 = 0.3756$; finally, group 6 does not show significant differences ($NIVC_7 = 0$) because only one member belongs. The internal verification of the regions allows an understanding of the cyclical fluctuations for January; this way, the stakeholders could act as a consequence of the international tourism demand. This last analysis of similarity verifies what was studied in part 4 of the methodology of the previous sections; it provides the theoretical-practical connection of the study presented in this work.

The following Table 2 collects the results of the automatic grouping of peaks of high tourist demand in Spain. This table is intended not to analyse all the months of the automated algorithm; with the interpretation of the previous table and Table 2, the reader could elaborate and interpret their data for the rest of database. Table 2 jointly represents the grouping months for July and August; it is presented simultaneously because they show similar automatic grouping behaviour. For both months, cluster 1 (for July: *Catalonia, Extremadura, Navarra, C. Leon, Basque Country, Aragon, Rioja, Andalucia, Valencia, Galicia*; for August: *Catalonia, Basque Country, Galicia*,

Table 1 $AutoCEs$ and $NIVC_{1q}$ for the clusters found in the treatment process according to Autonomous Communities visited by foreigners ($AutoCEs$ for January). Own elaboration

Automatic Cluster January					
	Limits		Elements	NIVC	Regions
	Lower	Upper			
1	0.0493	0.1669	2	0.5090	<i>Andalucia, Canarias</i>
2	0.1669	0.2845	2	0.1368	<i>Murcia, Valencia</i>
3	0.2845	0.4020	4	0.5416	<i>Navarra, Cantabria, Aragon, Extremadura</i>
4	0.4020	0.5196	3	0.3756	<i>Balears, C. Leon, Catalonia</i>
5	0.5196	0.6372	5	0.5600	<i>C. Mancha, Basque Country, Madrid, Asturias, Rioja</i>
6	0.6372	0.7548	1	0.0000	<i>Galicia</i>
7	0.7548	0.8724	0	0.0000	–

Navarra, Aragon, C. Leon, Rioja, Extremadura, Andalucia, Valencia) presents 10 regions with seasonal similar behaviour to the reference series; the month of July presents 4 elements in cluster 2 (*Canarias, Murcia, Asturias, Cantabria*) and August 5 elements (*Asturias, Murcia, Canarias, C. Mancha, Cantabria*); Cluster 3 (*C. Mancha*), 4 (*Balears*) and 6 (*Madrid*) only contains one region in every group for July; Cluster 3 (*Balears*) and 6 (*Madrid*) only include one region for August. Finally, we can say that the number of groups in July is 5 and in August it is 4, this means that in the peaks of these months, the global behaviour is highly concentrated. The internal verification values are similar and, in all cases, less than 0.4083. For July, the internal verification presents the most significant disparity in the first cluster, and this is due to a large number of elements within the group. The second group shows an internal difference of $NIVC_2 = 0.3476$; in the rest, there is no disparity because there are no elements or only one region. For the case of August, the values are similar to July with $NIVC_1 = 0.3657$ and $NIVC_2 = 0.1566$; we can confirm that there is less dissimilarity among members.

To close this empirical analysis section and avoid repeating similar comments for the reader, we encourage the reader to download the data referenced in this work and create their groups. The interpretation of the results is similar to that carried out previously. This way, researchers and practitioners could analyse an actual data set, this work is methodological, and we intend that users replicate the theoretical content. The following discussion and conclusion section with theoretical and practical implications will expose this tool's usability in making data-driven decisions in a digital environment. It will always be done under the central hypothesis of groupings: intra-group similarity and dissimilarity among groups. The robust criteria illustrate measurable behaviours that create value for public or private institutions.

5 Discussions and Conclusions

This study addresses unsupervised automatic clustering for seasonal time-series (*AutoCEs*) data in contexts of uncertainty and information gaps. The method assists practitioners and researchers with final analysis tasks in finance, economics, pattern recognition and information theory, among other fields of the final application. Automatic clustering adjusts its decision based on the number of spatial locations for a seasonal extent of global space t . The spatio-temporal information is verified with the so-called $NIVC_{sq}$ final generated clusters. The dimensional reduction of the proposed technique helps the objective interpretation of the data, avoids transferring cognitive biases of the investigation and improves understanding of the data-generating processes [26].

The results for a time-space problem of multi-locational tourist accommodation decisions show that the numbers of clusters are dynamic, adjusting to the characteristics of the entropy series and verifying intra-group similarity. Representing this knowledge in groups suggests policymakers' applications and decisions based on similarities of spatio-temporal behaviour based on entropy. However, the grouping is

Table 2 *AutoCEs* and $NIVC_{7q}$ for the clusters found in the treatment process according to Autonomous Communities visited by foreigners (*AutoCEs* for July and August). Own elaboration

<i>Automatic Cluster July</i>					
	Limits		Elements	NIVC	Regions
	Lower	Upper			
1	0.0915	0.1717	10	0.4083	<i>Catalonia, Extremadura, Navarra, C. Leon, Basque Country, Aragon, Rioja, Andalucia, Valencia, Galicia</i>
2	0.1717	0.2519	4	0.3476	<i>Canarias, Murcia, Asturias, Cantabria</i>
3	0.2519	0.3320	1	0.0000	<i>C. Mancha</i>
4	0.3320	0.4122	1	0.0000	<i>Balears</i>
5	0.4122	0.4924	0	-	-
6	0.4924	0.5726	1	0.0000	<i>Madrid</i>
7	0.5726	0.6528	0	-	-
<i>Automatic Cluster August</i>					
1	0.0854	0.1694	10	0.3657	<i>Catalonia, Basque Country, Galicia, Navarra, Aragon, C. Leon, Rioja, Extremadura, Andalucia, Valencia</i>
2	0.1694	0.2535	5	0.1566	<i>Asturias, Murcia, Canarias, C. Mancha, Cantabria</i>
3	0.2535	0.3375	1	0.0000	<i>Balears</i>
4	0.3375	0.4216	0	-	-
5	0.4216	0.5056	0	-	-
6	0.5056	0.5897	1	0.0000	<i>Madrid</i>
7	0.5897	0.6738	0	-	-

informative without a previous step to be applied by the researcher or policymaker. The empirical section results respond to the non-dynamic limitations of dynamic clustering of the literature in the tourism sector using 20,640 observations from INE [27]. We can state that different groupings are created by focusing on a reference series; it is noteworthy that March (usually Easter Week) and the months of July or August show a substantial similarity with the reference series, being 9, 10 and 10 intracluster elements, respectively. The latter shows temporal fluctuations similar to the reference series called Total. From an administrative perspective, this partition allows us to classify Autonomous Communities according to their decision-making ability from an objective perspective for analysis.

In addition, using external metrics or a combination of these can help balance the methods to generate profits in the possible field of application. Not all application areas are the same; each will present its empirical application and its characteristics. However, this method contributes to conglomerate techniques and the dimensional reduction in contours of global uncertainty before the analysis. The application in digital markets and a globally connected economy represent future investigations of this type of unsupervised cluster. The empirical results of this study verify the theoretical aspects raised in the methodological section, verifying an automatic system for grouping time series and dynamic data depending on the seasonal patterns studied.

In its current state, the approach of this unsupervised machine learning of clusters presented in this study offers several limitations and restrictions. They are not the only ones, but perhaps the most significant ones could be as follows: (1) The processing method requires time to analyse data arrays that generally must have the same length. Although this is not a limiting assumption since seasonal groupings are finally carried out. (2) The number of clusters is generated automatically; in this work, we have worked with spatial location information to divide up to a maximum of seven clusters. More clusters can be obtained with a more significant number of data, and intra-cluster similarity would be checked automatically. The introduction of a visual analysis could help to understand the data better. (3) Introducing specialised and sophisticated graphs could help/complement the previous understanding of dimensional reduction. However, this work is in the context of initial and exploratory uncertainty for researchers. (4) This work is framed in an unsupervised learning context, so it might be interesting to compare it with other clustering methods. Finally, the data in periods of confinement due to the coronavirus in Spain indicate increased entropy and disorder values. This argument can be the basis for studying structural changes in the data sets. Perhaps, the periods elapsed until the elaboration of this work would not resolve conclusive results. In the future, it will be an option for researchers.

Some of these issues may need in-depth study. In unsupervised learning and dimensional reduction of time series, they can vary according to the results of the mixed metrics used. The plausibility of specific methods will likely depend on the data set to be analysed; the truth is that the machine learning method is robust and presents its validation criteria. Finally, numerous real cases in which seasonal and spatio-temporal clustering algorithms are applied—that is, any analysis with time series data volumes—are of interest to researchers and practitioners for transport, electricity markets, operation research or finance, among others.

References

1. Kisilevich, S., Mansmann, F., Nanni, M., Rinzivillo, S.: Spatio-temporal clustering, *Data Mining and Knowledge Discovery Handbook*, pp. 855–874 (2009). https://doi.org/10.1007/978-0-387-09823-4_44.
2. Ruiz Reina, M.Á.: Multichoice Entropy Clustering for Time Series and Seasonality. In: *International Conference on Time Series and Forecasting* (2022)
3. Shannon, C.E.: A Mathematical Theory of Communication. *Bell System Technical J.* pp. 379–423, (1948). <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.
4. Ruiz-Reina, M.Á.: Spatio-temporal clustering: Neighbourhoods based on median seasonal entropy. *Spat Stat* **45**, 100535 (2021). <https://doi.org/10.1016/J.SPASTA.2021.100535>
5. UNWTO, “UNWTO World Tourism Barometer and Statistical Annex, March 2021,” UNWTO World Tourism Barometer, **19**(2), 1–32 (2021). <https://doi.org/10.18111/WTOBAROMETER.ENG.2021.19.1.2>.
6. INE, INEbase / Servicios /Hostelería y turismo /Cuenta satélite del turismo de España / Últimos datos (2022). https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=estadistica_C&cid=1254736169169&menu=ultiDatos&idp=1254735576863. Accessed Aug. 31, 2022
7. Qin, C., Song, S., Huang, G., Zhu, L.: Unsupervised neighborhood component analysis for clustering. *Neurocomputing* **168**, 609–617 (2015). <https://doi.org/10.1016/J.NEUCOM.2015.05.064>
8. Chen, B., Yin, H.: Learning category distance metric for data clustering. *Neurocomputing* **306**, 160–170 (2018). <https://doi.org/10.1016/J.NEUCOM.2018.03.048>
9. Cox, M., Ellsworth, D.: Managing big data for scientific visualisation. *ACM Siggraph* **97** (1997)
10. Aghabozorgi, S., Seyed Shirkhorshidi, A., Ying Wah, T.: Time-series clustering – A decade review. *Inf Syst*, vol. 53, pp. 16–38 (2015). <https://doi.org/10.1016/J.IS.2015.04.007>.
11. Scotto, M.G., Alonso, A.M., Barbosa, S.M.: Clustering time series of sea levels: Extreme value approach. *J Waterw Port Coast Ocean Eng* **136**, 215–225 (2010). [https://doi.org/10.1061/\(ASCE\)WW.1943-5460.0000045](https://doi.org/10.1061/(ASCE)WW.1943-5460.0000045)
12. Maharaj, E.A., Alonso, A.M., D’Urso, P.: Clustering seasonal time series using extreme value analysis: An application to Spanish temperature time series. *Commun Stat Case Stud Data Anal Appl* **1**, 175–191 (2015). <https://doi.org/10.1080/23737484.2016.1179140>
13. Alonso, A.M., Berrendero, J.R., Hernández, A., Justel, A.: Time series clustering based on forecast densities. *Comput Stat Data Anal* **51**, 762–766 (2006). <https://doi.org/10.1016/j.csda.2006.04.035>
14. Stuetzle, W.: Estimating the cluster tree of a density by analysing the minimal spanning tree of a sample. *J. Classif.* **20**(1), 25–47 (2003). <https://doi.org/10.1007/s00357-003-0004-6>
15. Leibovici, D.G., et al.: Spatio-temporal variations and uncertainty in land surface modelling for high latitudes: univariate response analysis. *Biogeosciences* **17**(7), 1821–1844 (2020). <https://doi.org/10.5194/BG-17-1821-2020>
16. Leibovici, D.G., Claramunt, C.: On Integrating Size and Shape Distributions into a Spatio-Temporal Information Entropy Framework *Entropy* Vol. 21, Page 1112, vol. 21, no. 11, p. 1112 (2019). <https://doi.org/10.3390/E21111112>
17. Peters, G.W., Nevat, I., Nagarajan, S.G., Matsui, T.: Spatial Warped Gaussian Processes: Estimation and Efficient Field Reconstruction. *Entropy* 2021, Vol. 23, Page 1323, vol. 23, no. 10, p. 1323 (2021). <https://doi.org/10.3390/E23101323>
18. Zhou, S., et al.: Travel Characteristics Analysis and Traffic Prediction Modeling Based on Online Car-Hailing Operational Data Sets. *Entropy* 2021, Vol. 23, Page 1305, vol. 23, no. 10, p. 1305 (2021). <https://doi.org/10.3390/E23101305>
19. Ansari, M.Y., Ahmad, A., Khan, S.S., Bhushan, G., Mainuddin, F.: Spatiotemporal clustering: a review. *Artif Intell Rev*, vol. 53, no. 4, pp. 2381–2423 (2020). <https://doi.org/10.1007/S10462-019-09736-1>
20. Vázquez, I., Villar, J.R., Sedano, J., Simić, S., de la Cal, E.: An ensemble solution for multivariate time series clustering. *Neurocomputing* **457**, 182–192 (2021). <https://doi.org/10.1016/J.NEUCOM.2020.09.093>

21. Wang, W., Zhang, Y.: On fuzzy cluster validity indices. *Fuzzy Sets Syst* **158**(19), 2095–2117 (2007). <https://doi.org/10.1016/J.FSS.2007.03.004>
22. Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J.M., Perona, I.: An extensive comparative study of cluster validity indices. *Pattern Recognit* **46**(1), 243–256 (2013). <https://doi.org/10.1016/J.PATCOG.2012.07.021>
23. Honarkhah, M., Caers, J.: Stochastic simulation of patterns using distance-based pattern modeling. *Math Geosci* **42**, 487–517 (2010). <https://doi.org/10.1007/s11004-010-9276-7>
24. Manco, G., Baglioni, M., Giannotti, F., Kuijpers, B., Raffaetà, A., Renso, C.: Querying and Reasoning for Spatiotemporal Data Mining. *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery*, pp. 335–374 (2008). https://doi.org/10.1007/978-3-540-75177-9_13.
25. Jakimowicz, A.: The Role of Entropy in the Development of Economics. *Entropy* **22**(4), 452, (2020). <https://doi.org/10.3390/E22040452>
26. Kliegr, T., Bahník, Š., Fürnkranz, J.: A review of possible effects of cognitive biases on interpretation of rule-based machine learning models. *Artif Intell* **295**, 103458 (2021). <https://doi.org/10.1016/J.ARTINT.2021.103458>
27. Yong-Jin, A.L.J., Jang, S., Jinwon, K.: Impacts of Peer-to-Peer Accommodation Use on Travel Patterns. *Annals of Tourism Research* **83**, 102960 (2020). <https://doi.org/10.1016/j.annals.2020.102960>

Modelling and Predicting the Dynamics of Confirmed COVID-19 Cases Based on Climate Data



Yuzhi Cai, Fangzhou Huang, and Jiao Song

Abstract We examine the impact of climate variables on the transmissibility of coronavirus (COVID-19) and predict the dynamics of confirmed COVID-19 cases using time series models and climate data. Although our method can be used in other parts of the world, our research focusses on Wales, UK. We find that temperature changes have a significant negative impact on the transmissibility, implying that as winter approaches, the spread of COVID-19 is likely to be aggravated. However, our findings also show that, due to the particular climate condition in Wales with considerably less volatile seasonal temperature change, the economic impact of the climate variables on suppressing the spread of COVID-19 in Wales is limited, suggesting that effective measures and potential intervention are crucial, particularly, during another wave. We also find that the accuracy of our forecasting results is high, indicating that our models are robust for forecasting and the climate variables are significant predictors of the COVID-19 daily cases in Wales. Our findings provide numerical evidence and are of importance when informing evidence-based policy-making on public health in Wales.

Keywords Temperature changes · COVID-19 · Forecasting · Climate variables · Public health

1 Introduction

Two and a half years have passed since the outbreak of the coronavirus (COVID-19). The spread of COVID-19 has triggered a global public health crisis. Therefore, there is still an urgent need to reduce the transmissibility of the pandemic. Predicting

Y. Cai (✉) · F. Huang
School of Management, Swansea University, Swansea, UK
e-mail: Y.Cai@swansea.ac.uk

F. Huang
e-mail: F.Huang@Swansea.ac.uk

J. Song
Public Health Wales, Cardiff, UK
e-mail: Jiao.Song@wales.nhs.uk

the spread of COVID-19 and determining the factors that impact the spread of the disease have played an important role in the decision-making process of public health officials around the world.

Different methods have been applied to obtain more accurate forecasts. For example, Tomar and Gupta [16] used neural network method to predict the number of COVID-19 cases in India. Kufel et al. [8] used autoregressive integrated moving average models to predict the COVID-19 cases in several European countries, while Barría-Sandoval et al. [2] also used autoregressive integrated moving average models to predict the COVID-19 cases in Chile. The official data modelling method was used by Li et al. [9] to study the transmission process and prediction of the COVID-19.

On the other hand, some researchers studied various factors that may affect the spread of the disease. For example, Kuchler et al. [7] used the structure of social networks as measured by Facebook to study the geographic spread of COVID-19. Sajadi et al. [14] studied the impact of climate variables on the spread of COVID-19 and found that several cities with substantial community spread had consistently similar weather patterns. Mollalo et al. [10] used several socio-economic factors to explain the variation of the spread of COVID-19 in the US. Zhao et al. [17] studied the impact of the policies that local authorities adopted on the spread of COVID-19. Shanmugam et al. [13] used a binomial-based approach to predict COVID-19 cases after a family gathering. Chen et al. [3] found that a country, which was located 1000 km closer to the equator, could expect 33% fewer cases per million inhabitants. A review on the recent research related to COVID-19 can be found in Shah et al. [15]. There are a number of more recent studies specifically focussed on the respiratory infections and climate factors in England and Wales. Nichols et al. [11] investigated the seasonality of coronaviruses infection, and suggested that infection was more likely to happen when temperature was low and humidity was high. Nottmeyer and Sera [12] also concluded that meteorological factors, such as temperature and humidity, were associated with COVID-19 case development. Lacobucci (2021) pointed out the “double risk” of outbreak of COVID-19 and flu in winter.

Due to considerably mild weather conditions in Wales, it is not clear how climate variables affect the spread of COVID-19 in Wales. Understanding the effect of the climate variables in relation to the spread of COVID-19 can provide solid evidence to policy-makers in Wales when new policies are to be informed. Although a number of models exist in the literature, very few specifically focussing on predicting the spread of COVID-19 in Wales. Therefore, it is of importance to develop regional-specific models for prediction. In this paper, we examine the impact of climate variables on the spread of COVID-19 in Wales, and predict the dynamics of confirmed COVID-19 cases based on time series models and climate data.

We find that temperature changes have a significant negative impact on the spread of coronavirus in Wales. Therefore, lower temperatures will accelerate the spread. We also find that the economic impact of the climate variables on the spread of COVID-19 in Wales is limited. Therefore, once the number of daily cases of COVID-19 is high, relying on temperature change to suppress the wave would be of little effectiveness, and if other measures and interventions are not taken, the daily cases would remain at that level for a long period of time. In addition, our forecasting

results for the dynamics of COVID-19 cases in Wales between 1 July 2021 and 31 August 2021 show high accuracy, indicating that our models are robust for short-term forecasting and the climate variables are good predictors of COVID-19 cases in Wales. Our findings also contribute to policy-making in Wales to control the spread of COVID-19.

In the next section, we discuss the data that we have collected, and in Sect. 3 we discuss our models. Our findings are discussed in Sect. 4, and the final section is the conclusion.

2 The Data

In this study, the climate data were collected from the Centre for Environmental Data Analysis (CEDA). Four climate variables are used in this paper, namely, daily precipitation, daily sunshine hours, daily maximum surface temperature and daily minimum surface temperature, where the daily precipitation measures the liquid that falls to the ground or condenses on the ground, the daily sunshine hours measures the amount of sunshine per day and the minimum and maximum temperatures measure the lowest and highest temperatures near the ground, respectively.

The daily number of new cases of COVID-19 is reported by the Wales Government and can be downloaded from the government website (see COVID-19 data, 2021). Due to the availability of the data, both climate data and COVID-19 data cover the time period from 31 January 2020 to 31 August 2021. We show the time series plots of the data in Fig. 1. It can be seen that the second wave of COVID-19 started around October 2020, and the third wave started near the start of June 2021 when lockdown was fully eased.

The summary statistics of the data are given in Table 1. It can be seen that during this period, the daily minimum number of COVID-19 cases was 0 and the maximum was 3502. The daily average number of COVID-19 cases is 448 and the standard deviation is 651.9, indicating that the number has changed significantly. During this period, 25% of the days had fewer than 49 daily cases, and 25% of the days had more than 697 daily cases.

For the climate conditions during this period, the average daily sunshine hours was 4.8 h and the standard deviation was 3.1 h. The average daily precipitation was 3.9 mm and the standard deviation was 5.6 mm. The average daily minimum and maximum temperatures were 6.2 °C and 13.9 °C, respectively, 25% of the daily minimum temperature was lower than 2.5 °C, 25% of the daily maximum temperature was higher than 18.6 °C. Therefore, the climate condition in Wales was relatively stable with little sign of extreme. In the next section, we will examine how this climate condition affects the spread of COVID-19 in Wales.

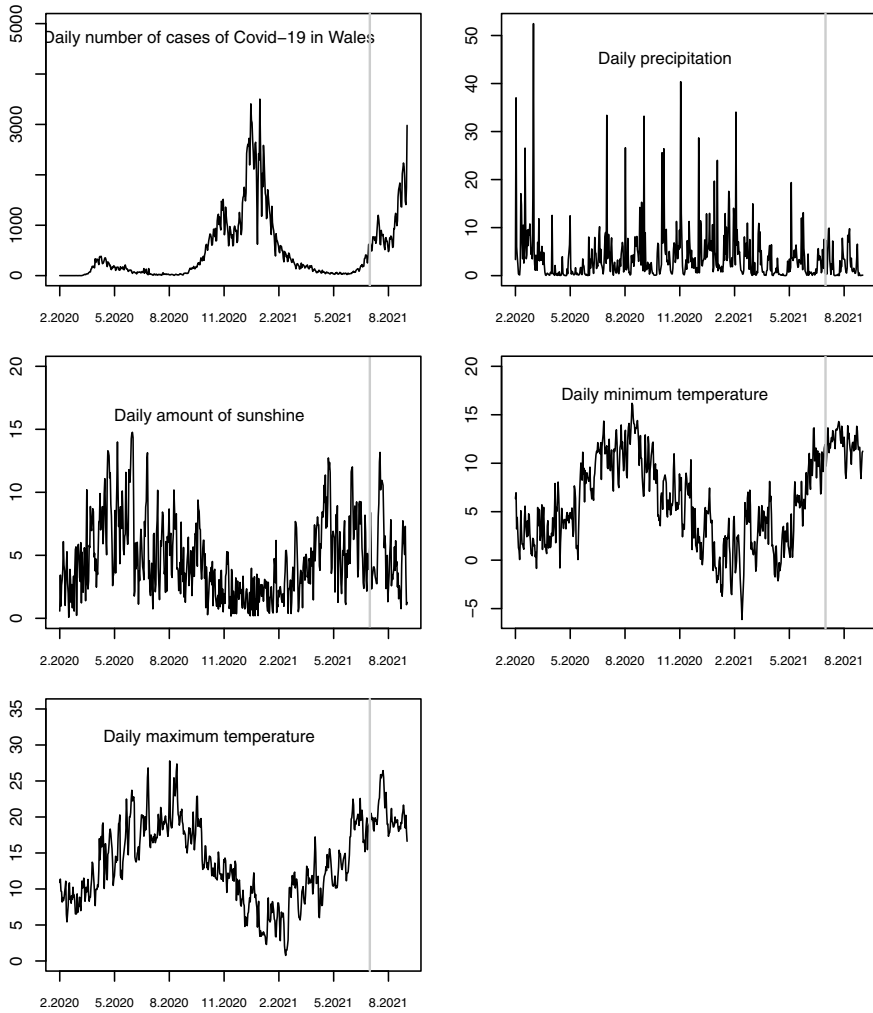


Fig. 1 Plots of the observed data, where the vertical line indicates the date of 1 July 2021

Table 1 Summary statistics of the data

	Mean	SD	Min	Q1	Median	Q3	Max
COVID cases	481.1	651.9	0.0	49.0	177.0	697.5	3502.0
Precipitation	3.9	5.6	0.0	0.6	2.2	4.9	52.5
Sunshine	4.8	3.1	0.1	2.5	4.0	6.7	14.8
Min-temp.	6.2	4.6	-6.1	2.5	5.8	10.0	16.2
Max-temp.	13.9	5.5	0.8	10.0	13.6	18.6	27.8

3 The Model

Within the time series modelling framework, a number of issues need to be considered when developing a model that will allow us to examine how the climate variables affect the spread of COVID-19 in Wales and how to obtain a short-term forecast of the future spread of COVID-19 in Wales.

The first issue is that all time series should be stationary. Figure 1 clearly shows that these time series are not stationary, which has also been confirmed by the Augmented Dickey–Fuller test (see, Cheung and Lai [4]). Therefore, directly examining how climate variables affect the spread of COVID-19 in Wales cannot be applied, but examining how short-term climate changes affect the spread of COVID-19 in Wales can be applied instead. This would require the change in each variable to be defined.

Let u_t be the number of new COVID-19 cases on day t , and w_{jt} be the climate variable, where $j = 1, 2, 3$ and 4 correspond to daily precipitation, daily sunshine hours and daily minimum temperature and daily maximum temperature, respectively. Then, we define the change in the number of new COVID-19 cases during the time period $(t - h_1, t)$ as the difference quotient

$$y_t = \frac{u_t - u_{t-h_1}}{h_1}, \tag{1}$$

where h_1 is a positive integer. So, y_t is the slope of the line segment joining the points $(t - h_1, u_{t-h_1})$ and (t, u_t) . For example, if $h_1 = 7$ and $y_t = 10$, then the slope of the line segment on $(t - 7, t)$ is 10. This means that the number of COVID-19 cases on day t will be 70 more than the number of COVID-19 cases a week ago. Similarly, we define the change of climate variables in the time period $(t - h_2, t)$ as

$$x_{jt} = \frac{w_{jt} - w_{jt-h_2}}{h_2}, \quad j = 1, \dots, 4, \tag{2}$$

where h_2 is a positive integer.

Since it is crucial to examine the impact of past climate change on the current spread of COVID-19, we let $h_1 \geq h_2$. On the other hand, according to the latest scientific and public health advice (<https://gov.wales/self-isolation>), “everyone must continue to self-isolate for 10 days if they test positive for Covid-19”, we let $1 \leq h_2 \leq h_1 \leq 10$. This defines that the minimum time period for changes in u_t and w_{jt} is $(t - 1, t)$, and the longest time period is $(t - 10, t)$.

After defining y_t and x_{jt} , ADF test is employed to check their stationarity. The results show that for all possible values of h_1 and h_2 , they are stationary. Therefore, y_t and x_{jt} can be used to examine the impact of climate variables on the spread of COVID-19 in Wales.

The second issue is that past values of y_t , for example, y_{t-1}, \dots, y_{t-p} , can also affect y_t , where p is a positive integer. Therefore, for similar reasons for h_1 and h_2 , in this study, we define the maximum value of p equal 10.

Table 2 Correlation coefficient between climate variables

	Precipitation	Sunshine	Min-temp.	Max-temp.
Precipitation	1.000	-0.314	0.024	-0.145
Sunshine	-0.314	1.000	0.135	0.467
Min-temp.	0.024	0.135	1.000	0.891
Max-temp.	-0.145	0.467	0.891	1.000

The final issue is that correlation among the four climate variables needs to be examined. The results are given in Table 2. It is seen that the correlation between the daily minimum and maximum temperatures is considerably high (> 0.89), which suggests that they should not be employed in the same model to avoid potential multicollinearity. Therefore, we consider the following two models:

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p} + \alpha_0 + \alpha_1 x_{1t-d} + \alpha_2 x_{2t-d} + \alpha_3 x_{3t-d} + e_t, \quad (3)$$

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p} + \alpha_0 + \alpha_1 x_{1t-d} + \alpha_2 x_{2t-d} + \alpha_4 x_{4t-d} + e_t, \quad (4)$$

where $1 \leq h_1, h_2, p \leq 10, h_2 \leq h_1$ and $d = h_1 - h_2 + 1$. It is seen that d also takes values from 1 to 10. For ease of discussion of our results, we refer to model (3) and model (4) as Model A and Model B, respectively. Therefore, Model A does not include the change in daily maximum temperature and Model B does not include the change in daily minimum temperature. Each model defines a set of similar type of models, these models have the same climate variables, but h_1, h_2 and p take different values.

4 Findings and Discussions

We divide the data into two parts: the first part contains data between 31 January 2020 and 30 June 2021 (inclusive), and the second part contains the data between 1 July 2021 and 31 August 2021 (inclusive). The first part of the data is used to estimate models and examine the impact of climate variables on the spread of COVID-19 in Wales, and the second part of data is used to check the accuracy of the forecasting results.

4.1 Estimated Models and Impact of Climate Variables

Since different values of h_1, h_2 and p correspond to different models, we use the Bayesian Information Criterion (BIC) to determine the best model and use the best model to examine the impact of short-term climate changes on the spread of COVID-

19 in Wales. BIC is one of the commonly used criteria for model selection. According to BIC, the best model corresponds to the model with the smallest BIC value. We present the estimated models in Table 3.

It is interesting to see that the best Model A and the best Model B have the same values of h_1, h_2, p , and hence d . Both models show that the effect of the changes in daily precipitation and sunshine hours is not significant, but the effect of the changes in minimum and maximum temperatures is statistically significant. The negative coefficients of the two temperature variables indicate that an increase in daily temperature changes will result in a decrease in the spread of COVID-19. However, due to the less volatile weather conditions in Wales, the daily temperature changes are rather modest. The statistical inference of estimated models is further scrutinized.

Consider Model A, where $h_1 = 7, h_2 = 2$ and $d = h_1 - h_2 + 1 = 6$. When change in daily minimum temperature $x_{3t-d} = x_{3t-6}$ increases by 1 unit, the slope of the line segment between $(t - d - h_2, w_{3t-d-h_2}) = (t - 8, w_{3t-8})$ and $(t - d, w_{3t-d}) = (t - 6, w_{3t-6})$ will increase by 1 unit. Therefore, the minimum temperature will increase by 2°C at time $t - 6$ from the minimum temperature at time $t - 8$. On the other hand, when x_{3t-6} increases by 1 unit, y_t will decrease 1.889 units. This means that the slope of the line segment between $(t - h_1, u_{t-h_1}) = (t - 7, u_{t-7})$ and (t, u_t) will decrease 1.889 units. Therefore, the number of COVID-19 cases at time $t - j$ will be $1.889(h_1 - j) = 1.889(7 - j)$ less than the number of COVID-19 cases at time $t - 7$, where $j = 6, 5, \dots, 0$. Table 4 shows the number of cases that will be decreased in the next 7 days if the temperature increases by 2°C at time $t - 6$. The

Table 3 Parameters of the estimated models

	Model A		Model B	
	$h_1 = 7, h_2 = 2$ $p = 7, d = 6$		$h_1 = 7, h_2 = 2$ $p = 7, d = 6$	
Variables	Estimate	P-value	Estimate	P-value
y_{t-1}	0.546	0.000***	0.552	0.000***
y_{t-2}	0.213	0.000***	0.207	0.000***
y_{t-3}	0.151	0.002***	0.147	0.003***
y_{t-4}	0.054	0.274	0.058	0.234
y_{t-5}	-0.215	0.000***	-0.218	0.000***
y_{t-6}	0.235	0.000***	0.236	0.000***
y_{t-7}	-0.207	0.000***	-0.205	0.000***
Intercept	1.143	0.792	1.140	0.793
Precipitation	-0.200	0.366	-0.287	0.186
Sunshine	-0.611	0.344	-0.135	0.824
Min-temp.	-1.889	0.016**		
Max-temp.			-1.371	0.061*

Note ***, **, and * Indicate significance at the 1%, 5%, and 10% levels, respectively

Table 4 COVID-19 cases reduced daily for the next 7 days when temperature increases by 2°C

	$t - 6$	$t - 5$	$t - 4$	$t - 3$	$t - 2$	$t - 1$	t
Model A	2	4	6	8	9	11	13
Model B	1	3	4	5	7	8	10

results of Model A show that the total number of COVID-19 cases reduced within 7 days is limited. Similar results for Model B are also given in Table 4.

As the weather conditions in Wales are fairly stable, we usually expect that the temperature will not change significantly within a short period of time (e.g. 2 days), and therefore the number of daily cases will change very little accordingly. This indicates that once the number of COVID-19 cases reaches a high level, intervention needs to be adopted to suppress the spread of the disease and climate variables, especially changes in temperature would have little impact on confining the transmissibility of COVID-19.

4.2 Forecasting Results

We now discuss how to use the models in Table 3 to predict future number of COVID-19 cases from 1 July 2021 to 31 August 2021, based on the information we have up to 30 June 2021. We let T represent the date of 30 June 2021. The vertical line in Fig. 1 indicates the position of T .

Let \hat{y}_{T+m} be the predicted value of y_{T+m} and let \hat{u}_{T+m} be the predicted value of u_{T+m} , where $m = 1, \dots, 62$, i.e. 62 days after 30 June 2021. Then, it follows from (1) that

$$\hat{u}_{T+m} = \hat{y}_{T+m}h_1 + \hat{u}_{T+m-h_1}, \tag{5}$$

where $m = 1, 2, \dots, 62$ and $\hat{u}_{T+m-h_1} = u_{T+m-h_1}$ if $T + m - h_1 \leq T$, because in this case we have the observed data. It is seen that once \hat{y}_{T+m} can be obtained from the estimated models, we can use the formula (5) to calculate \hat{u}_{T+m} , which is the forecast of the number of COVID-19 cases at time $T + m$. Figure 2 shows the predicted and the observed number of COVID-19 cases from 1 July 2021 to 31 August 2021. We can see that the number of COVID-19 cases observed and predicted during this period is very close.

To quantify the accuracy of the forecasts, as widely accepted measures, Mean Absolute Deviation (MAD), Mean Absolute Percent Error (MAPE) and Root Mean Square Error (RMSE) are employed to measure the accuracy of our prediction results, where the MAD is defined by the average of $|\hat{u}_{T+m} - u_{T+m}|$, the MAPE of the prediction results is defined by the average of $|\hat{u}_{T+m} - u_{T+m}|/u_{T+m} * 100$ and the RMSE is defined by the square root of the average of $(\hat{u}_{T+m} - u_{T+m})^2$. Values of MAPE, MAD and RMSE of the forecasting results are shown in Table 5. Table 5

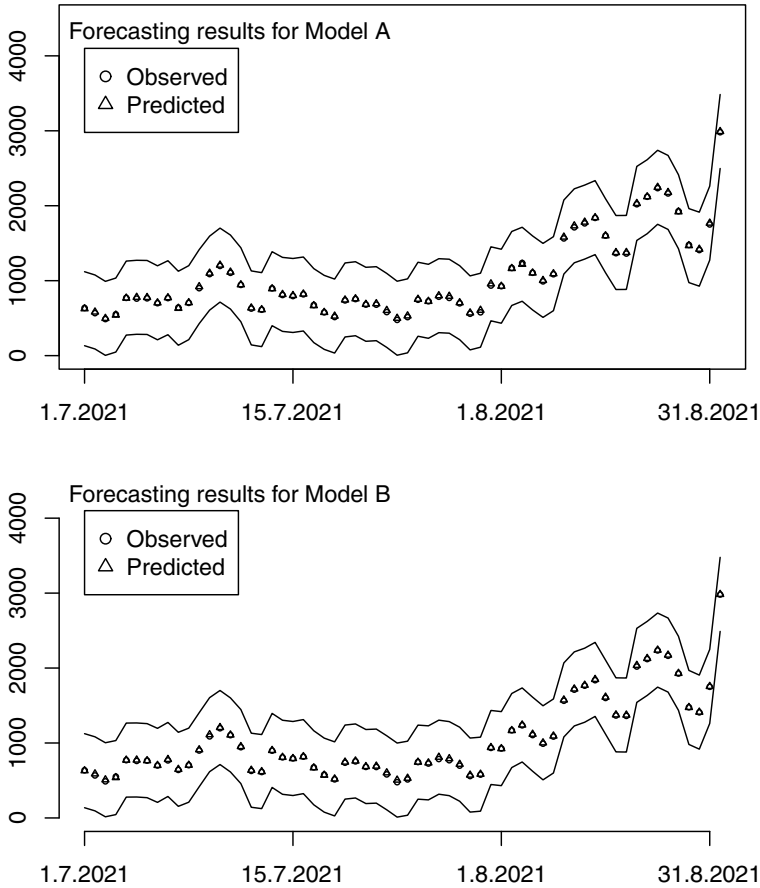


Fig. 2 The prediction results of the daily number of COVID-19 cases in Wales from 1 July 2021 to 31 August 2021

shows that Model B is slightly better than Model A, which shows that when predicting the dynamics of COVID-19 cases, the daily maximum temperature change is better than the daily minimum temperature change. According to MAPE, on average, the deviation of the prediction from the actual observed number of cases does not exceed 1.099%. Based on MAD, the average absolute difference between the observed and predicted number of cases is less than 12 cases. Finally, according to RMSE, the square root of the average squared difference between the observed and predicted number of cases is less than 14 cases. All these results show that the accuracy of our prediction results is high, which indicates that our models are robust for short-term prediction of the dynamics of COVID-19 cases in Wales, and the climate variables show good predictive power in forecasting.

Table 5 Results of forecasting accuracy

	MAPE	RMSE	MAD
Model A	1.099	13.419	11.464
Model B	0.906	10.714	9.026

5 Conclusion

We used actual observable data from 31 January 2020 to 30 June 2021 to study the impact of climate variables on the spread of COVID-19 in Wales. We found that temperature changes had a statistically significant negative impact on the spread of COVID-19 in Wales. However, since the weather conditions in Wales are less volatile, exceptional large temperature changes within a short period of time is not commonly expected. Therefore, the economic impact of the climate variables on the spread of COVID-19 in Wales is limited, which shows that once the number of daily cases of COVID-19 reaches a high level, effective measures are crucial to confine the spread of COVID-19. We also found that the accuracy of the forecasting results obtained from our models was high, which implies that our models are robust for short-term forecasting and the climate variables show good predictive power in predicting the dynamics of confirmed COVID-19 cases in Wales.

The implications of the findings are: high temperature has impact on preventing the spread of COVID-19, while the impact is marginal. Therefore, from public health and socio-economic perspective, taking effective measures and intervention to suppress the daily cases is the key to a re-opened economy. Nonetheless, climate variables, especially temperature changes, should be included as risk factors to the future public health surveillance framework. Our results are of value to policy-makers in Wales to inform evidence-based policy-making, specifically to the recovery of regional economy and sustaining healthcare system.

References

1. Armstrong, J.S.: Principles of Forecasting: a Handbook for Researchers and Practitioners. Kluwer Academic Publishers (2001)
2. Barría-Sandoval, C., Ferreira, G., Benz-Parra, K., López-Flores, P.: Prediction of confirmed cases of and deaths caused by Covid-19 in Chile through time series techniques: a comparative study. Plos One (2021). <https://doi.org/10.1371/journal.pone.0245414>
3. Chen, S., Prettnner, K., Kuhn, M., Geldsetzer, P., Wang, C., Bärnighausen, T., Bloom, D.E.: Climate and the spread of Covid-19. Sci. Rep. **11**, 9042 (2021). <https://doi.org/10.1038/s41598-021-87692-z>
4. Cheung, Y.W., Lai, K.S.: Lag order and critical values of the augmented Dickey-Fuller test. J. Bus. Econ. Stat. **13**, 277–280 (1995)
5. Covid data for the United Kingdom. <https://public.tableau.com/app/profile/public.health.wales.health.protection/viz/RapidCOVID-19virology-Public/Headlinesummary>

6. Iacobucci, G.: Covid and flu: what do the numbers tell us about morbidity and deaths? (2021). <https://www.bmj.com/content/375/bmj.n2514>
7. Kuchler, T., Russel, D., Stroebel, J.: The geographic spread of Covid-19 correlates with structure of social networks as measured by Facebook. Technical report, National Bureau of Economic Research (2020)
8. Kufel, T.: ARIMA-based forecasting of the dynamics of confirmed Covid-19 cases for selected European countries. *Equilib. Q. J. Econ. Econ. Policy* **15**, 181–204 (2020)
9. Li, L., Yang, Z., Dang, Z., Meng, C., Huang, J., Meng, H., Wang, D., Chen, G., Zhang, J., Peng, H., Shao, Y.: Propagation analysis and prediction of the Covid-19. *Infect. Dis. Modell.* **5**, 282–293 (2020)
10. Mollalo, A., Vahedi, B., Rivera, K.M.: GIS-based spatial modeling of Covid-19 incidence rate in the continental United States. *Sci. Total Environ* **728**, 138884 (2020). <https://doi.org/10.1016/j.scitotenv.2020.138884>. Epub 2020 Apr 22. PMID: 32335404; PMCID: PMC7175907
11. Nichols, G.L., Gillingham, E. L., Macintyre, H. L., Vardoulakis, S., Hajat, S., Sarran, C. E., et al.: Coronavirus seasonality, respiratory infections and weather. *BMC Infect. Dis.* **21**(1), 1–15 (2021). <https://doi.org/10.1186/s12879-021-06785-2>
12. Nottmeyer, L.N., Sera, F.: Influence of temperature, and of relative and absolute humidity on COVID-19 incidence in England-A multi-city time-series study. *Environ. Res.* **196**, 110977. <https://doi.org/10.1016/j.envres.2021.110977>
13. Shanmugam, R., Ledlow, G., Singh, K.P.: Predicting Covid-19 cases with unknown homogeneous or heterogeneous resistance to infectivity. *Plos One* (2021). <https://doi.org/10.1371/journal.pone.0254313>
14. Sajadi, M.M., Habibzadeh, P., Vintzileos, A., Shokouhi, S., Miralles-Wilhelm, F., Amoroso, A.: Temperature, humidity, and latitude analysis to estimate potential spread and seasonality of coronavirus disease 2019 (Covid-19). *JAMA Netw. Open* **3**(6), e2011834 (2020). <https://doi.org/10.1001/jamanetworkopen.2020.11834>
15. Shah, S., Mulahuwaish, A., Ghafoor, K.Z., Maghdid, H.S.: Prediction of global spread of Covid-19 pandemic: a review and research challenges. *Artif. Intell. Rev.* (2021). <https://doi.org/10.1007/s10462-021-09988-w>
16. Tomar, A., Gupt, N.: Prediction for the spread of Covid-19 in India and effectiveness of preventive measures. *Sci. Total Environ.* **728**, 138762 (2020). <https://www.sciencedirect.com/science/article/pii/S0048969720322798>
17. Zhao, Z., Li, X., Liu, F., Zhu, G., Ma, C., Wang, L.: Prediction of the Covid-19 spread in African countries and implications for prevention and controls: a case study in South Africa, Egypt, Algeria, Nigeria, Senegal and Kenya. *Sci. Total Environ.* **729**, 138959 (2020). <https://www.sciencedirect.com/science/article/pii/S0048969720324761>
18. Author, F.: Article title. *Journal* **2**(5), 99–110 (2016)
19. Author, F., Author, S.: Title of a proceedings paper. In: Editor, F., Editor, S. (eds.) CONFERENCE 2016, LNCS, vol. 9999, pp. 1–13. Springer, Heidelberg (2016). 10.10007/1234567890
20. Author, F., Author, S., Author, T.: Book title, 2nd edn. Publisher, Location (1999)
21. Author, A.-B.: Contribution title. In: 9th International Proceedings on Proceedings, pp. 1–2. Publisher, Location (2010)
22. LNCS Homepage. <http://www.springer.com/lncs>. Accessed 4 Oct. 2017

Least-Squares Wavelet Analysis of Rainfalls and Landslide Displacement Time Series Derived by PS-InSAR



Ebrahim Ghaderpour, Claudia Masciulli, Marta Zocchi, Roberta Marini, Giandomenico Mastrantoni, Francesca Reame, Gianmarco Pantozzi, Nicol o Belcecchi, Gabriele Scarascia Mugnozza, and Paolo Mazzanti

Abstract Time series analysis of Interferometric Synthetic Aperture Radar (InSAR) data is a crucial step for monitoring the displacement of the Earth's surface. The Persistent Scatterer InSAR (PS-InSAR) is a multi-temporal InSAR method that provides the displacement time series that can be used for studying ground deformation. From a hazard assessment perspective, the rapid detection of deformation patterns is crucial for identifying the areas that will be affected by damage due to landslides. Understanding the relationship between triggering factors, such as rainfall and the occurrence of mass movements from the interpretation of SAR time series is still a major challenge. Herein, we first review some of the traditional methods, such as Pearson correlation analysis for investigating whether there is any possible linear dependency between rainfall and ground deformation measurements. Then, we describe the time series analysis tools in the least-squares wavelet software that can be used for processing non-stationary time series which may not be evenly sampled. We demonstrate how these tools can be utilized to understand more about the relationships between displacement and rainfall time series which have different sampling rates without any need for filtering and/or aggregation.

Keywords Coherency analysis · Displacement · Interferometry · Landslide · Persistent scatterer (PS) · Rainfall · Remote sensing · Spectral analysis · Synthetic aperture radar (SAR) · Trend analysis

E. Ghaderpour (✉) · C. Masciulli · M. Zocchi · G. Mastrantoni · F. Reame · G. Scarascia Mugnozza · P. Mazzanti

Department of Earth Sciences, Sapienza University of Rome, P.le Aldo Moro, 5, 00185 Rome, Italy
e-mail: ebrahim.ghaderpour@uniroma1.it

G. Pantozzi · N. Belcecchi · G. Scarascia Mugnozza · P. Mazzanti
NHAZCA s.r.l., Spin-off Sapienza University of Rome, Via Vittorio Bachelet, 12, 00185 Rome, Italy

R. Marini · G. Scarascia Mugnozza · P. Mazzanti
CERI Research Centre, Sapienza University of Rome, P.le Aldo Moro, 5, 00185 Rome, Italy

1 Introduction

Ground deformation can be studied by satellite Interferometric Synthetic Aperture Radar (InSAR) [1]. In the Differential InSAR (DInSAR) techniques, two radar images capturing the same region but taken at different times are compared to produce a map (i.e., interferogram). The map shows ground displacement along the sensor's line of sight between two acquisition times [2]. However, its main limitation is related to the accuracy caused by spatial and temporal decorrelation (atmospheric delay, topographic errors, etc.) between the two signals. This can be overcome (at least for some particular points, called Persistent Scatterers), with the Advanced-DInSAR (A-DInSAR) technique, which consists of combining a large number of SAR images (multi-image technique). The Persistent Scatterer InSAR (PS-InSAR) technique is an Advanced Differential Synthetic Aperture Radar Interferometry (A-DInSAR) method to obtain the temporal evolution of displacements for stable radar reflectors, in terms of temporal coherence and amplitude stability, in the investigation area: the so-called Persistent Scatterers PSs [3]. PS-InSAR has been widely applied for different hazard investigations, from local-scale deformation processes (e.g., those affecting a single building or structure) to large-scale deformation processes (e.g., landslides, subsidence, tectonic deformations, etc.). The potential of A-DInSAR measurements for landslides detection, mapping, and characterization has been widely investigated to reconstruct the history of deformations through PS time series analysis [4–7].

A-DInSAR technique allows the study of trend changes in time series, reflecting the deformations observed in differential interferograms. The time series of deformation processes are non-stationary, i.e., exhibit trends, jumps, and wavelike components that change frequency and amplitude over time. Several methods were examined to analyze and evaluate the relations between landslides and triggering phenomena, in particular, heavy rainfalls [8, 9]. Heavy rainfall can determine an increase in velocity trends of time series, providing interesting information about the processes that governed the slope behavior [10]. A possible correlation between the rain and the activation of a deformation phenomenon has been highlighted in [11] through the analysis of Sentinel-1 interferograms and rainfall data. The deformation phenomena detected during the interferometric analysis were mostly concentrated during high rainfall periods, showing that rain could be considered the main triggering/accelerating factor of the studied landslides.

To investigate the possible influence of one phenomenon on another, one may process their time series by several statistical methods. For example, the Pearson correlation coefficient is a common statistical metric that shows the strength and direction of a linear relationship between two variables, denoted by r [12]. However, this metric only shows the possible linear dependency between two phenomena, e.g., deformation processes and rainfall. The deformation processes such as landslides, discovered by the PS time series, are often sampled unevenly because of many reasons, such as sensor defect, storage, and economy. Therefore, methods that can rigorously consider irregularities in sampling and values of such time series are

extremely demanding. To analyze the correlation of PS and rainfall in non-stationary time series, we have to take into account that they may also have different sampling rates with missing values. In addition, there might exist some time delay between the wavelike components of the time series, e.g., a time delay between the annual cycles of vegetation and climate or between the annual cycles of interferometric baseline length and temperature [13, 14].

The traditional wavelet coherence and cross-wavelet transform have been used in many applications to study possible relationships between two phenomena in the time-frequency domain [15, 16]. However, these techniques have some limitations. For example, at higher frequencies or small wavelet scales, peaks in the spectrograms/scalograms and cross-spectrograms/scalograms get smoothed out, resulting in a reduction of power that can be misleading [17]. Another limitation of these methods is that the time series should be evenly spaced with no missing values. The Least-Squares Wavelet (LSWAVE) software [18, 19] is designed to process any type of time series regardless of how they are sampled. This software contains several tools each designed for a particular purpose described in more detail in Sects. 2.4, 2.5, 2.6, and 2.7.

The main goal of this study is to highlight the potential of the LSWAVE software for analyzing and investigating the possible relationships between PS-InSAR and precipitation time series. A rainfall time series, obtained from the nearest weather station to an area in the Municipality of Borghi, Italy, affected by landslides, is selected. Then its possible impact on the ground deformation is investigated by performing correlation and coherency analyses with the ascending- and descending-orbital geometries of the PS-InSAR time series. The results of these analyses are demonstrated in Sect. 3. Finally, the discussion and conclusions are provided in Sects. 4 and 5, respectively.

2 Materials and Methods

2.1 Study Region

The study area, which comprises the municipality of Borghi, is located in the typical hilly landscape of the eastern part of the Forlì-Cesena province (Emilia Romagna Region, Northern Italy, [21]) (Fig. 1). The local geology is represented by a turbiditic sequence (flysch) composed of marly and pelitic rocks in alternation with fractured sandstone layers [22, 23]. The consequence of such a heterogeneous alternation of hard and soft rock layers is an intense slope instability that affects the study area, expressed in relatively small and shallow slides, earth flows, and complex landslides. According to the national-scale historical landslides archive IFFI (Inventario Dei Fenomeni Franosi in Italia [24]), the majority of these landslides are classified as active or dormant. The main triggering factor for the activation or reactivation of

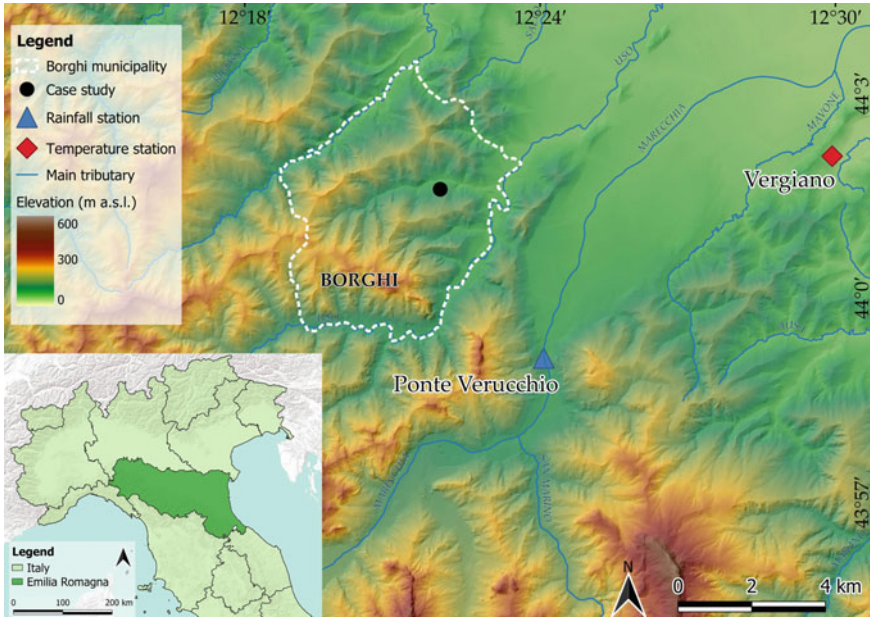


Fig. 1 Geographical location of the case study within the municipality of Borghi on 10 m resolution digital elevation model [20]. The closest rainfall and temperature stations to the studied area are represented by a blue triangle and a red diamond respectively (arpae database <https://www.arpae.it/>). This map was generated with QGIS which is a free and open-source geographic information system (GIS).

landslides is related to long and intense rainfall during the whole year, or rapid snowmelt during the springtime (March and April [25–27]).

2.2 Datasets

The meteorological data (rainfall and temperature) used herein are provided by the Arpae agency and freely downloaded from <https://www.arpae.it/>. The A-DInSAR dataset used in this study consists of 265 SAR images from the archives of the Italian Space Agency (ASI) acquired from 2010 to 2019 in both orbital geometry:

- Ascending orbital geometry: 119 images in Single Look Complex (SLC) format acquired by the COSMO-SkyMed satellites from 18 February 2011 to 03 August 2019;
- Descending orbital geometry: 146 images in Single Look Complex (SLC) format acquired by the COSMO-SkyMed satellites from 23 August 2010 to 01 September 2019.

The COSMO-SkyMed constellation consists of four satellites, launched from 2007 to 2010, at an orbital height of 619.6 km and an orbital inclination of 97.86° . All satellites are equipped with high-resolution X-band radar (3.1 cm wavelength), capable of observing through cloud cover and during the night. The repeat cycle is 16 days, and the revisit capability is variable because every single satellite will have a near revisit time of 6 days. The policy on data acquisition of satellite missions, e.g., background missions (continuous acquisition for the same areas) versus on-demand acquisitions, affects the availability of data. Since the COSMO-SkyMed satellite mission has both military and civil purposes, the available archived images are not exactly one every 6 days. Therefore, the dataset is not continuous over time. The PS-InSAR time series values often contain outliers because of the technical issues of the PSI method during the stack gathering from the SAR data, the atmospheric phase estimation, and the deformation model used [28].

The selected PS-InSAR time series exhibit a mean cumulative displacement (expressed in mm) of -91.3 mm (DESC) and -61.7 mm (ASC) in the time interval of data acquisition (2011–2019 for the ascending geometry and 2010–2019 for the descending one). The area of interest was carefully chosen within a quiescent landslide reported in the IFFI database (IFFI-0400794100 [24]). Furthermore, the possible correlation with intense rain events was emphasized by the presence of strong accelerations within a small time window in the time series.

2.3 Pearson Correlation Method

The correlation between pairs of series can be measured to determine how much two time series vary together. The most common quantitative measure of correlation is the Pearson correlation coefficient, denoted by r , which can be computed to determine the strength and direction of the relationship between two variables [29]. The Pearson correlation coefficient is essentially a normalized measurement of the covariance that indicates how far away the data points are from the best fitting line, defined by Eq. (1):

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (1)$$

where \bar{x} and \bar{y} are the means of variables x and y , respectively. The Pearson correlation coefficient has a value between -1 and $+1$. This metric reflects only the linear association between two sets of data to find how well they are related, ignoring other types of relationships or correlations. Furthermore, x_i and y_i should ideally be measurements corresponding to time t_i when applying Eq. (1) to find the linear dependency between two time series. This means that if the two time series have different sampling rates, one should first resample the time series to match their corresponding times before Pearson correlation analysis. Thus, the correlation result will also depend on the resampling method.

2.4 Least-Squares Spectral Analysis (LSSA)

The LSSA is a spectral analysis method for processing unevenly sampled time series that may have trends or jumps ([30]). A least-squares spectrum (LSS) can be estimated by choosing a frequency set and trend constituents, such as linear, quadratic, or cubic. The LSS can be plotted as frequency vs. amplitude or frequency vs. percentage variance. To obtain a LSS, trend and sinusoidal functions at each given frequency are fitted to the time series via the least-squares optimization. Mathematically, let

$$f = \begin{bmatrix} f(t_1) \\ f(t_2) \\ \vdots \\ f(t_n) \end{bmatrix}, \quad c_k = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_q \\ c_{k,1} \\ c_{k,2} \end{bmatrix},$$

$$\Phi_k = \begin{bmatrix} \phi_1(t_1) & \phi_2(t_1) & \cdots & \phi_q(t_1) & \cos(2\pi\omega_k t_1) & \sin(2\pi\omega_k t_1) \\ \phi_1(t_2) & \phi_2(t_2) & \cdots & \phi_q(t_2) & \cos(2\pi\omega_k t_2) & \sin(2\pi\omega_k t_2) \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ \phi_1(t_n) & \phi_2(t_n) & \cdots & \phi_q(t_n) & \cos(2\pi\omega_k t_n) & \sin(2\pi\omega_k t_n) \end{bmatrix}, \quad (2)$$

where $f(t_j)$ is the measurement estimated at time t_j , ω_k is a cyclic frequency, ϕ_ℓ is a constituent of known form, and c_k is the coefficient vector being estimated by the least-squares method. For example, if $\phi_1(t_j) = 1$ and $\phi_2(t_j) = t_j$ for $1 \leq j \leq n$, then c_1 and c_2 will be the intercept and slope of a linear trend to be estimated. The following cost function is minimized in LSSA after an optimization process.

$$\Psi(c_k) = (f - \Phi_k c_k)^T (f - \Phi_k c_k), \quad (3)$$

where T is transpose. Thus, the estimated coefficient vector will be

$$\hat{c}_k = (\Phi_k^T \Phi_k)^{-1} \Phi_k^T f. \quad (4)$$

The amplitude of the sinusoid at ω_k is the square root of the sum of squares of the last two elements of \hat{c}_k . Since the first q columns in Φ_k are fixed, the amplitude estimation using Eq. (4) is not computationally efficient, especially when estimating LSS for a large set of frequencies. More details for computational optimization of Eq. (4) can be found in [17, Supplementary Materials].

To obtain the percentage variance LSS, the time series may first be de-trended. Then the LSS may be estimated for the residual time series, a process that takes into account the correlation between the removed trend and the sinusoids at each frequency. More precisely, let $r(t_j)$ be the estimated residual at time t_j , i.e., the original measurement minus the fitted trend value at t_j . The normalized or percentage variance LSS (after multiplying by 100) of the residual series is calculated by the following formula:

$$s_k = \sum_{j=1}^n r(t_j) \left(\hat{c}_{k,1} \cos(2\pi\omega_k t_j) + \hat{c}_{k,2} \sin(2\pi\omega_k t_j) \right) / \sum_{j=1}^n r(t_j)^2, \quad (5)$$

where $\hat{c}_{k,1}$ and $\hat{c}_{k,2}$ are the last two elements of \hat{c}_k in Eq. (4).

The graph of estimated amplitudes of the fitted sinusoids versus their frequencies shows the amplitude LSS while the percentage variance LSS shows the contribution amount of the estimated sinusoids to the residual time series versus the frequencies. Note that when the trend is also fitted, both amplitude and percentage variance LSS are obtained for the residual series while the effect of trend removal is considered in the LSS estimation.

Assuming the normality of time series values, LSS in Eq. (5) will follow the beta distribution [31]. From this statement, if a spectral peak value is larger than the critical value at a certain significance level (e.g., 0.01), then the peak is statistically significant at 99% confidence level.

2.5 *Least-Squares Cross-Spectral Analysis (LSCSA)*

The LSCSA is a time series decomposition method that simultaneously processes two time series together for calculating coherency and phase differences between the harmonic components of the time series [32]. To obtain the least-squares cross-spectrum (LSCS) for two time series, first, the LSS of each time series is obtained, and then the LSSs are multiplied by each other. The stochastic significance of a peak in LSCS is also based on the normality assumption of the two time series whose values are also statistically independent. The discrepancy between the estimated phases can determine the time delay between the harmonics. For example, the phase difference of 60° at frequency 2 cycles/year indicates that the harmonic in the first time series leads/lags by about 30 days from the harmonic in the other time series. Note that the season-trend fit is applied to the entire time series in both LSSA and LSCSA. Therefore, the estimation of components whose frequencies and/or amplitudes change over time is an overall average and so not accurate locally [32, 33].

2.6 *Least-Squares Wavelet Analysis (LSWA)*

The LSWA, an extension of LSSA, can process non-stationary time series that may not be evenly sampled [32, 33]. In LSWA, a least-squares wavelet spectrogram (LSWS) is computed by sliding a window over time whose size is inversely proportional to the frequency, i.e., as the frequency increases the window size decreases, allowing a more accurate estimation of short-duration waves. The number of observations or measurements within a window is called the window size or segment size.

Spectrograms are usually displayed as time versus frequency versus amplitude or time versus frequency versus percentage variance. The LSWA also estimates the phases of sinusoids. To obtain LSWS, a windowing technique is implemented. Within each sliding window, LSSA is applied to estimate a spectral peak corresponding to that window using Eqs. (4) or (5). In other words, for each time and each frequency, the amplitude or percentage variance is estimated by simultaneously fitting the trend and sinusoids at the given frequency to the segment within the sliding window located at the given time. The percentage variance shows how much the residual segment contains the sinusoids at a given frequency.

A Gaussian function may be used to define weights for time series values within each window. Therefore, the values toward the window center get higher weights (more important) than the values toward the margins of the window (less important) during the season-trend estimation. This results in a smooth spectrogram, i.e., an optimal time-frequency resolution. Furthermore, the weights are useful when there exist missing values in non-stationary time series, so the values further away from the window center receive relatively lower weights. In fact, the Gaussian weights and harmonics in the LSWA model are like the Morlet wavelet in the least-squares sense [17, 32]. Note that the window location and window center are the same when the time series is evenly sampled or an equally spaced set of times is selected for estimating the spectrogram.

Like LSSA, the spectrogram peaks can be statistically assessed with the normality assumption. Note that in some applications, this assumption may not be valid but has no effect on the estimation of a spectrogram. In other words, regardless of whether the time series values are normally distributed or not, a spectrogram can still be estimated. A stochastic confidence level surface, shown herein as a gray surface, can show which spectrogram peaks are significant stochastically [32]. In other words, if a peak stands above the surface, then it is statistically significant.

2.7 *Least-Squares Cross-Wavelet Analysis (LSCWA)*

The least-squares cross-wavelet analysis (LSCWA) is a time-frequency decomposition technique proposed for coherency analysis and estimating phase differences between the harmonics of two time series [32, 33]. LSCWA can be directly applied to time series that are sampled at different time intervals, and it can account for the measurement errors. Moreover, the cross-spectrograms in LSCWA have higher time-frequency resolution compared to the ones in XWT [17].

The Least-squares cross-wavelet spectrogram (LSCWS) is obtained from the multiplication of the spectrograms of the two time series [32]. Since the time series may have different sampling rates, a common time vector is selected first that can be either the union of the time vectors in both time series or any equally spaced time vector whose values are within the common time interval of the two time series. The cross-spectrograms are plotted as time versus frequency versus percentage variance (coherency). The percentage variance shows the coherency amount between the har-

monics of two time series segments—the higher the percentage variance, the higher the coherency and vice versa [17, Supplementary Materials].

The LSCWA also estimates the phase differences between the harmonics fitted to time series segments. The phase difference is a number between -180 and 180° , and it is usually plotted by an arrow on the 2D view of the cross-spectrogram. For example, arrows pointing to the positive and negative directions of the time axis mean that the harmonics are in-phase and out-of-phase, respectively. Arrows pointing to the positive and negative direction of the frequency axis mean that the harmonic in the second time series lags and leads the one in the first time series by 90° , respectively [14]. Given the frequency, the estimated angle can be converted to time.

When the time series values are statistically independent and normally distributed, a confidence level surface can identify whether an estimated peak in the cross-spectrogram is statistically significant [32]. The MATLAB and Python software packages developed for LSSA, LSCSA, LSWA, and LSCWA are comprehensively described in [18, 19].

3 Results

3.1 Results of Traditional Methods

The relationship between temperature and rainfall variability was investigated by using monthly averages of temperature and cumulative rainfall (Fig. 2). These data were calculated from the daily cumulative temperature and rainfall for the 2011–2019 time period. Borghi municipality climate is classified as warm and temperate (Cfa, the acronym for humid subtropical climate, according to the Köppen-Geiger climate classification [34]). Temperatures are at their highest from June to August (summer) while they are at their lowest from December to February (winter) (Fig. 2a). The average temperature is 11.9°C . Rainfall amount is significant throughout the year, with an average annual rainfall of 74.3 mm. Even the driest month (August) has an average rainfall of 41.8 mm.

There is no statistically significant linear correlation between mean temperature and cumulative rainfall, as highlighted by Fig. 2b. This lack of correlation can be due to the different trends of temperature and rainfall. As underlaid by the annual mean temperature and cumulative rainfall trends, the rainfall seems to show no seasonal tendency (Fig. 2c), while the temperature is characterized by an annual periodicity (Fig. 2d). Both rainfall and temperature exhibit significant monthly variability.

The analysis of the correlation is not straightforward for unevenly spaced time series, as PS-InSAR. To face this problem, the original unevenly spaced time series were butter-filtered and resampled daily in order to remove outliers and obtain equidistant points of measure (Top panels of Fig. 3). The correlations between PS (ascending and descending) and rainfall time series are underlined in the bottom panels of Fig. 3. On one hand, the Pearson correlation coefficient shows that

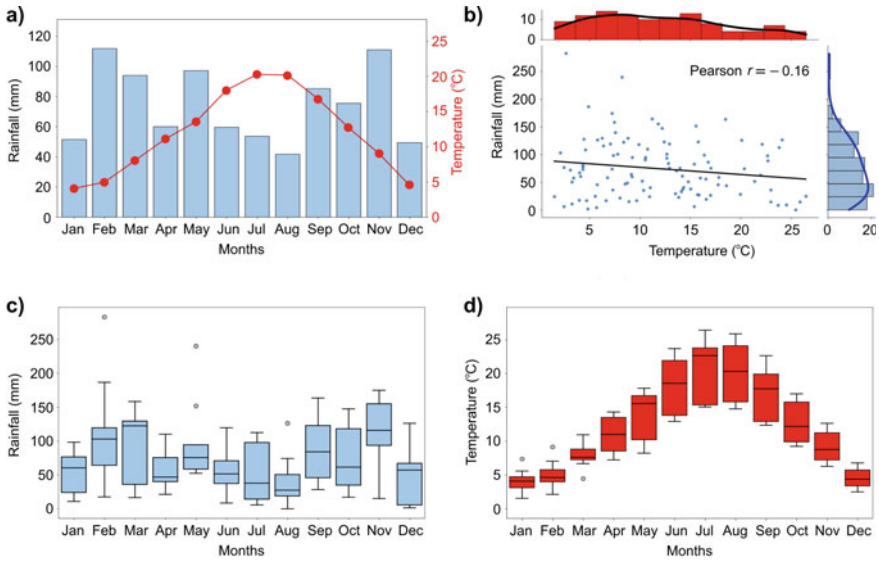


Fig. 2 The climate charts showing the temperature and rainfall of Borghi municipality during the 2011–2019 time period. The monthly mean temperature and cumulative rainfall, **a** show a decrease in the rain during the summer season (from June to August) when the mean temperature increases. The Pearson coefficient, **b** indicates a poor negative correlation between the two variables. The box plots of rainfall **c** and temperature, **d** reveal a great variability during months. The outliers are shown by gray hexagons, and the horizontal lines inside the boxes show the median values

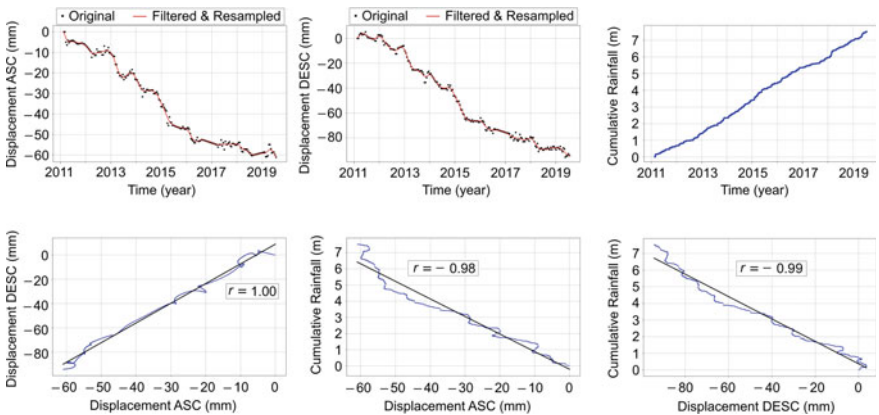


Fig. 3 Top three panels: PS-InSAR time series for both ascending (ASC) and descending (DESC) satellite orbits and cumulative rainfall time series. The red curves show the filtered and daily resampled time series. Bottom three panels: Pearson correlation results between the selected daily time series

the ascending and descending PS time series are positively correlated with $r \simeq 1$ (p -value < 0.01). On the other hand, a strong negative relationship between PS and rainfall time series is highlighted by the r -value close to -1 (p -value < 0.01). Since the occurrence of landslides can be directly caused by intensive rainfall, identifying an increasing change in rainfall could determine the identification of a change in the occurrence of landslides. Overall, traditional methods of correlation do not allow the analysis of the quantitative relationship between the volume of rainfall and landslide occurrence for the identification of triggers.

3.2 The LSWAVE Results

The percentage variance and amplitude spectra and spectrograms for rainfall and temperature time series are illustrated in Fig. 4. The linear trend is used for both LSSA and LSWA. Interestingly, there is no statistically significant annual component at 99% confidence level in the rainfall time series (Fig. 4a). Furthermore, from the LSS, there is no statistically significant seasonal component in the rainfall time series while the LSWs shows short-duration seasonal components in the years 2012, 2014, 2015, 2016, and 2018 (shown in reddish). The amplitude spectrogram in Fig. 4b also shows relatively higher estimated amplitudes for the seasonal components at 4–5 cycles/year (period of 2–3 months) in 2015 and 2018.

The temperature time series shows a dominant annual component that is statistically significant in both LSS and LSWs (cf. Fig. 4c). On the other hand, the amplitude spectrogram illustrated in Fig. 4d clearly shows the amplitude of the temperature has decreased since 2011 while this cannot be observed from LSS amplitude because LSS only shows frequency versus amplitude not time–frequency versus amplitude.

Figure 5 shows the LSCSs and LSCWSs of the climate and PS-InSAR displacement time series. A linear trend was estimated and removed from each segment when estimating LSCWSs in panels (a)–(d) and from each time series when estimating LSCSs in panels (a) and (b) while a cubic trend was fitted and removed from the displacement time series when estimating LSCSs in panels (c) and (d).

One can see from Fig. 5a that the annual cycles of temperature and rainfall time series are coherent at 99% confidence level though the percentage variance is very low (about 2%). Arrows displayed on the significant annual peaks are pointing to the left, meaning that the annual cycles of temperature and rainfall are almost out-of-phase. The estimated phase difference using LSCSA is approximately -170° . This means that when the annual cycle of the temperature reaches its maximum value, the annual cycle of the rainfall reaches its minimum value and vice versa. Note that Fig. 3b only showed a weak negative correlation ($r \simeq -0.16$) but neither showed the seasonality nor temporal change.

From Fig. 5b, a statistically significant coherency can be observed for the annual components, where the annual cycle of time series for the ascending geometry leads the one for the descending geometry by about one to two months over time. The

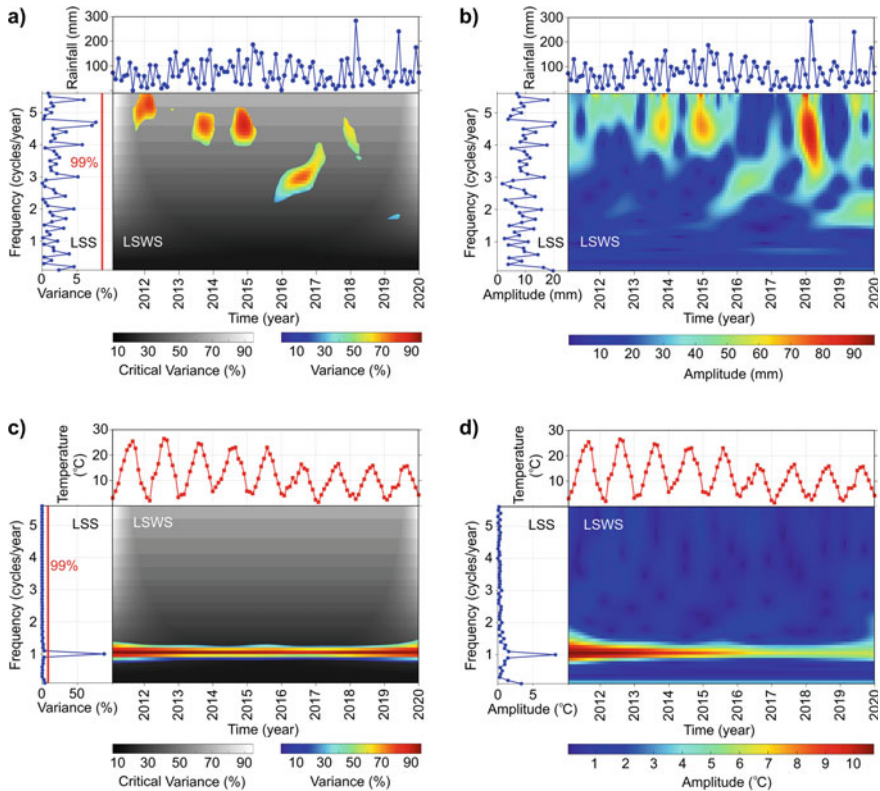


Fig. 4 The percentage variance and amplitude spectra and spectrograms for **a, b** rainfall and **c, d** temperature time series. The red lines in LSSs displayed in panels, **a** and **c** show the critical values at 99% confidence level. Also, the gray color map is the stochastic surface at 99% confidence level

cross-spectral peaks are not estimated for the locations with data gaps (e.g., between the end of 2016 and early 2017).

Interestingly, the short-duration seasonal components of displacement (ascending geometry) and rainfall at 4–5 cycles/year (about 2–3 months period) toward the end of 2014 are coherent at 99% confidence level (cf., Fig. 5c), with approximately 120° or one month phase difference. In other words, as the rainfall value increases, the displacement value decreases. It is also known that landslides occurred toward the end of 2014.

The annual cycle in the years 2013 and 2018 for rainfall lags the one for the displacement (descending geometry) by about one month, see Fig. 5d. This is in agreement with the fact that there were landslides in the years 2013 and 2018. Furthermore, in the year 2016 both LSCWSs, illustrated in Fig. 5c, d, show statistically significant coherence at 3 cycles/year with about 90° phase difference. This indicates that the four-month cycle of the displacement time series leads the one in the rainfall

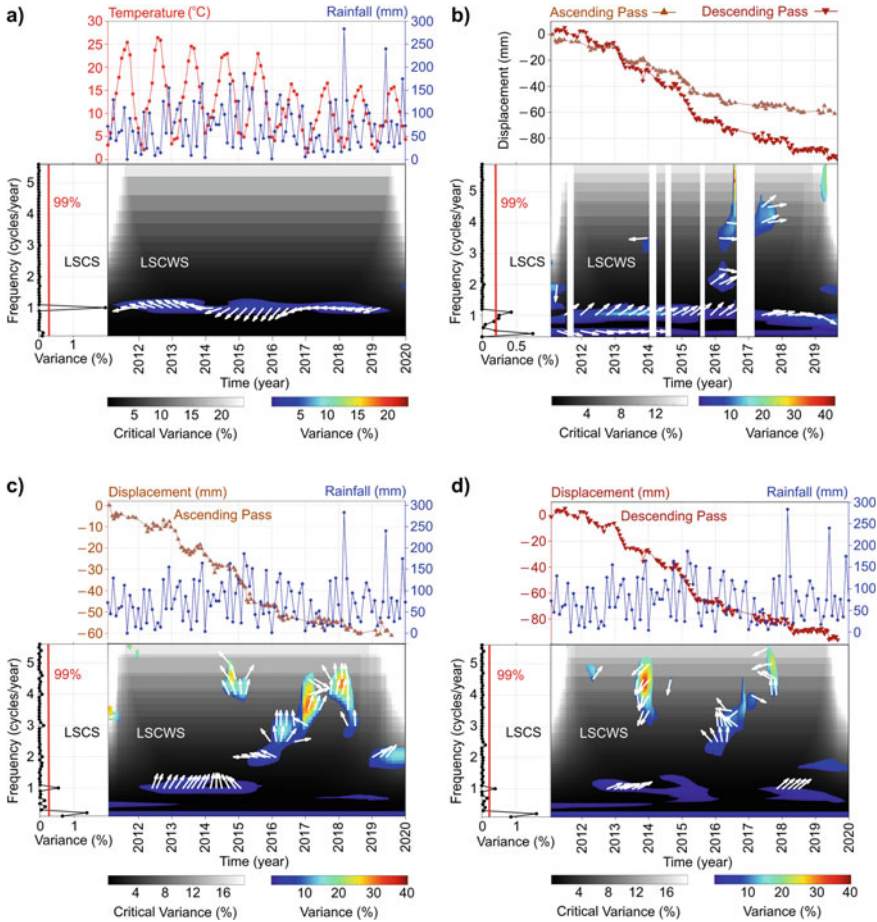


Fig. 5 The cross-spectra and cross-spectrograms of climate and PS-InSAR time series. The red lines in LSCSs show the critical values at 99% confidence level. Also, the gray color map represents the stochastic surface at 99% confidence level. The white arrows on LSCWSs show the local phase differences. Arrows pointing to the right, left, top, and bottom indicate that the seasonal cycles of the segments in the first time series are in-phase, out-of-phase, leads, and lags with/from the ones in the second time series, respectively. To avoid displaying too many arrows on LSCWSs, arrows are displayed only for some of the most significant peaks

by approximately one month which could mean that the rainfall might have played a significant role in the ground deformation. Note that landslides also occurred during 2016.

4 Discussion

As observed in Fig. 4a, there was no statistically significant annual component at 99% confidence level in the rainfall time series as opposed to a significant annual component that usually exists in rainfall time series in other regions (e.g., [13, 17]). The annual amplitude attenuation over time in the temperature time series was also an interesting observation. From the negative Pearson correlation ($r \simeq -0.16$) and the direction of the arrows displayed on the annual peaks in Fig. 5a, it can be deduced that the warm summers tended to be drier and cool winters tended to be wetter in this study region—like the results in some other regions (e.g., [35, 36]).

As demonstrated in Fig. 5, the seasonal cycles of the displacement time series generally led the ones in the rainfall time series by nearly one month during major landslides. The least-squares spectral and wavelet analyses did not show any statistically significant annual component in the rainfall time series while the annual component of the rainfall was weakly coherent with the one in the displacement time series which could have also triggered the landslides. Before applying the Pearson correlation analysis, the time series had to be aggregated, i.e., the displacement time series had to be regularized and resampled so that both displacement and rainfall data align in time. However, the LSWAVE tools did not require any such pre-processing and were directly applied to process the climate and displacement time series with different sampling rates and gaps.

The impact of rainfall on ground deformation requires further investigation. For example, the wavelike components of the displacement time series could have been created by several factors, such as possible biases created during the InSAR data pre-processing steps, and land cover and climate change. Applying the methods mentioned herein to process displacement and rainfall time series for other regions may help in a better understanding of their relationships which is subject to future work.

5 Conclusions

We briefly reviewed the tools in the LSWAVE software and showed how they may be utilized to investigate possible relationships between rainfalls and displacement time series derived by PS-InSAR (both ascending and descending geometry) in the municipality of Borghi, Italy. We also investigated the relationship between rainfalls and temperature in the study region. We highlighted what additional information one can obtain when using these tools as compared to traditional ones like Pearson correlation analysis. The tools in the LSWAVE software were directly applied to unevenly sampled time series having different sampling rates without any filtering and/or aggregation. As for future work, we shall generate geospatial maps using these tools for investigating possible spatiotemporal relationships between rainfalls and displacement time series. We hope that such analyses help geologists to better understand the pattern of landslides and their triggering factors.

References

1. Hanssen, R.F.: Radar Interferometry: data Interpretation and Error Analysis, vol. 2. Springer Science & Business Media (2001)
2. Kampes, B.M.: Radar Interferometry—Persistent Scatterer Technique, vol. 12. Springer Dordrecht (2006). <https://doi.org/10.1007/978-1-4020-4723-7>
3. Ferretti, A., Prati, C., Rocca, F.: Permanent scatterers in SAR interferometry. *IEEE Trans. Geosci. Remote Sens.* **39**, 8–20 (2001)
4. Bozzano, F., Mazzanti, P., Perissin, D., Rocca, A., De Pari, P., Discenza, M.E.: Basin scale assessment of landslides geomorphological setting by advanced InSAR analysis. *Remote Sens.* **9**, 267 (2017)
5. Moretto, S., Bozzano, F., Esposito, C., Mazzanti, P., Rocca, A.: Assessment of landslide pre-failure monitoring and forecasting using satellite SAR interferometry. *Geosciences* **7**, 36 (2017)
6. Antonielli, B., Mazzanti, P., Rocca, A., Bozzano, F., Dei Cas, L.: A-DInSAR performance for updating landslide inventory in mountain areas: an example from Lombardy region (Italy). *Geosciences* **9**, 364 (2019)
7. Antonielli, B., Monserrat, O., Bonini, M., Cenni, N., Devanthery, N., Righini, G., Sani, F.: Persistent Scatterer interferometry analysis of ground deformation in the Po Plain (Piacenza-Reggio Emilia sector, Northern Italy): seismo-tectonic implications. *Geophys. J. Int.* **206**, 1440–1455 (2016)
8. Tanyaş, H., Kirschbaum, D., Lombardo, L.: Capturing the footprints of ground motion in the spatial distribution of rainfall-induced landslides. *Bull. Eng. Geol. Environ.* **80**, 4323–4345 (2021)
9. Martino, S., Antonielli, B., Bozzano, F., Caprari, P., Discenza, M.E., Esposito, C., Fiorucci, M., Iannucci, R., Marmoni, G.M., Schilirò, L.: Landslides triggered after the 16 August 2018 Mw 5.1 Molise earthquake (Italy) by a combination of intense rainfalls and seismic shaking. *Landslides* **17**, 1177–1190 (2020)
10. Moretto, S., Bozzano, F., Mazzanti, P.: The role of satellite InSAR for landslide forecasting: limitations and openings. *Remote Sens.* **13**, 3735 (2021)
11. Barra, A., Monserrat, O., Mazzanti, P., Esposito, C., Crossetto, M., Mugnozza, G.S.: First insights on the potential of Sentinel-1 for landslides detection. *Geomat. Nat. Hazards Risk* **7**(6), 1874–1883 (2016)
12. Rodgers, J.L., Nicewander, W.A.: Thirteen ways to look at the correlation coefficient. *Am. Stat.* **42**, 59–66 (1988)
13. Ghaderpour, E., Vujadinovic, T.: The potential of the least-squares spectral and cross-wavelet analyses for near-real-time disturbance detection within unequally spaced satellite image time series. *Remote Sens.* **12**, 2446 (2020)
14. Ghaderpour, E.: Least-squares wavelet and cross-wavelet analyses of VLBI baseline length and temperature time series: Fortaleza-Hartebeesthoek-Westford-Wetzell. *Publ. Astron. Soc. Pac.* **133**(014502), 10 (2021). <https://doi.org/10.1088/1538-3873/abcc4e>
15. Torrence, C., Compo, G.P.: A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* **79**, 61–78 (1998)
16. Grinsted, A., Moore, J.C., Jevrejeva, S.: Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Process Geophys.* **11**, 561–566 (2004)
17. Ghaderpour, E., Vujadinovic, T., Hassan, Q.K.: Application of the least-squares wavelet software in hydrology: Athabasca River Basin. *J. Hydrol. Reg. Stud.* **36C**, 100847 (2021). <https://doi.org/10.1016/j.ejrh.2021.100847>
18. Ghaderpour, E., Pagiatakis, S.D.: LSWAVE: a MATLAB software for the least-squares wavelet and cross-wavelet analyses. *GPS Solut.* **23**, 50 (2019)
19. Ghaderpour, E.: JUST: MATLAB and python software for change detection and time series analysis. *GPS Solut.* **25**, 85 (2021)
20. Tarquini, S., Isola, I., Favalli, M., Battistini, A.: TINITALY, a digital elevation model of Italy with a 10 m-cell size (Version 1.0) [Data set]. Istituto Nazionale di Geofisica e Vulcanologia (INGV) (2007). <https://doi.org/10.13127/TINITALY/1.0>

21. MultiMedia LLC, Servizio Geologico, Sismico e dei Suoli, Geological and Soil Maps of the Emilia-Romagna Region. Historical Landslide Databasem (2006). <https://ambiente.regione.emilia-romagna.it/en/geologia/cartography/webgis/soil-cartography>. Accessed 5 July 2022
22. Farabegoli, E., Benini, A., Martelli, L., Onorevoli, G., Severi, P.: Geologia dell' Appennino Romagnolo da Campigna a Cesenatico. Mem. Descr. Carta Geol. It **56**, 165–184 (1992)
23. Benini, A., Cremonini, G., Martelli, L., Cibir, U., Severi, P., Bassetti, M.A., Ghiselli, F., Vaiani, S.C.: Carta geologica d'Italia. Scala. 1:50000, Foglio. n. 255 Cesena, Servizio Geologico d'Italia. Selca, Firenze, pp. 40 (2009). <http://hdl.handle.net/11585/98630>
24. ISPRA - Regione Emilia Romagna, IFFI - Inventario dei Fenomeni Franosi in Italia (2005). <https://www.progettoiffi.isprambiente.it/cartografia-on-line/>. Accessed 5 July 2022
25. Bertolini, G., Guida, M., Pizziolo, M.: Landslides in Emilia-Romagna region (Italy): strategies for hazard assessment and risk management. *Landslides* **2**, 302–312 (2005)
26. Rossi, M., Witt, A., Guzzetti, F., Malamud, B.D., Peruccacci, S.: Analysis of historical landslide time series in the Emilia-Romagna region, northern Italy. *Earth Surf. Process. Landforms* **35**, 1123–1137 (2010). <https://doi.org/10.1002/esp.1858>
27. Piacentini, D., Troiani, F., Daniele, G., Pizziolo, M.: Historical geospatial database for landslide analysis: the Catalogue of Landslide Occurrences in the Emilia-Romagna Region (CLOCKER). *Landslides* **15**, 811–822 (2018)
28. Crosetto, M., Monserrat, O., Cuevas-González, M., Devanthéry, N., Crippa, B.: Persistent scatterer interferometry: a review. *ISPRS J. Photogramm. Remote Sens.* **115**, 78–89 (2016)
29. Boslaugh, S., Watters, P.A.: Statistics in a Nutshell—A Desktop Quick Reference. O'Reilly Media, p. 480 (2008)
30. Vaníček, P.: Further development and properties of the spectral analysis by least-squares. *Astrophys. Space Sci.* **12**, 10–33 (1971)
31. Pagiatakis, S.: Stochastic significance of peaks in the least-squares spectrum. *J. Geod.* **73**, 67–78 (1999)
32. Ghaderpour, E., Ince, E.S., Pagiatakis, S.D.: Least-squares cross-wavelet analysis and its applications in geophysical time series. *J. Geod.* **92**, 1223–1236 (2018)
33. Ghaderpour, E.: Least-squares wavelet analysis and its applications in geodesy and geophysics. Ph.D. thesis, York University, Toronto, ON, Canada (2018)
34. Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F.: Present and future Köppen-Geiger climate classification maps at 1 km resolution. *Sci. Data* **5**, 180214 (2018)
35. Zhao, W., Khalil, M.A.K.: The relationship between precipitation and temperature over the contiguous United States. *J. Clim.* **6**, 1232–1236 (1993)
36. Nkuna, T.R., Odiyo, J.O.: The relationship between temperature and rainfall variability in the Levubu sub-catchment. *South Africa Int. J. Environ. Sci.* **1**, 65–75 (2016)

Time Series Forecasting

Macroeconomic Forecasting Evaluation of MIDAS Models



Nicolas Bonino-Gayoso  and Alfredo Garcia-Hiernaux 

Abstract We compare the nowcasting and forecasting performance of different variants of MIDAS models (ADL-MIDAS, TF-MIDAS and U-MIDAS) when predicting the GDP growth of the four largest Euro Area economies between 2011Q4 and 2020Q3. We consider various high-frequency indicators, horizons and sub-periods, each of the latter with a distinct level of uncertainty. A meta-regression, with an average error metric as exogenous variable, is estimated to account for potential differences in performance by country, indicator, sample period or method. The results obtained with the whole sample do not reveal any difference in the predictive accuracy of the models under comparison. The findings are robust to the forecasting error metric used, RMSFE or MAFE.

Keywords Nowcasting · Forecasting · Mixed-frequency models · MIDAS · U-MIDAS · TF-MIDAS · RMSFE · MAFE

1 Introduction

Access to real-time assessments of the state of the economy as well as to forecasts of its expected evolution is essential in the decision-making process, either for policy-makers or businesspeople. Up-to-date macroeconomic projections are critical inputs for designing and adjusting economic policy. This becomes even more important in non-stable and challenging economic environments, such as those faced with COVID-19 pandemic and the Russo-Ukrainian War.

However, data offered by the System of National Accounts is delivered with considerable delay. In the case of European countries, Eurostat provides the value of EU and Euro Area GDP 70 days after the end of each quarter, preceded by a

N. Bonino-Gayoso (✉)
Banco de España, Madrid, Spain
e-mail: alvaro.bonino@bde.es

A. Garcia-Hiernaux
DANAE and ICAE, Universidad Complutense de Madrid, Madrid, Spain
e-mail: agarciah@ucm.es

first preliminary estimate and a second estimate 30 and 45 days after the end of the quarter, respectively. This delay, as pointed out by [8], is the consequence of the difficulties in producing timely and accurate low-frequency aggregates: (i) not all disaggregated data are available when needed to compute a relevant aggregate; (ii) many disaggregated time series are only preliminary estimates, subject to substantial revisions, so they are not accurate descriptions of the current conditions.

Nevertheless, a considerable number of short-term economic indicators available at a much earlier stage could be used to capture information about the state of the economy. For instance, we have access to monthly data from consumer surveys or the industrial production index, daily data from financial markets and even more frequently observed variables, such as Google or Twitter trends and, sometimes, phone mobility data. Although likely not as complete as data from the System of National Accounts, these high-frequency (HF) indicators can improve the prediction of relevant economic aggregates, such as GDP growth, inflation or unemployment (see, e.g., [33, 35]).

Classical models usually require using data observed at the same frequency, representing a setback when working with a more complete dataset integrated by variables observed at different frequencies. The way to extract information from the available high-frequency indicators is not always a simple task; there are several methodologies, with different levels of complexity, to do it. Various classes of models have been proposed to work explicitly with mixed-frequency datasets, some of the most widely used being the so-called MIDAS (MIXed DATA Sampling) models. These models have attracted considerable attention recently, even being adopted by many official institutions.

Original MIDAS models [22, 23, 25] were defined in terms of a Distributed Lag (DL) polynomial, explicitly modeling the relationship between variables observed at different frequencies. In order to keep parsimony, standard MIDAS models are built in terms of a few parameters. In this chapter, to differentiate them from other classes of MIDAS models, we will name this standard MIDAS model as ADL-MIDAS (Autoregressive Distributed Lag-MIXed DATA Sampling). This model has been used to nowcast GDP, private consumption and corporate bond spreads, among other variables (see [10, 11, 14, 26, 36]).

A specific variation of the previous standard model, known as Unrestricted MIDAS (U-MIDAS), was first introduced by [30] and later deeply analyzed by [17]. Based on a set of simulation exercises, these authors state that U-MIDAS nowcasting precision outperforms the standard MIDAS when the difference in sampling frequencies is not large, specially for monthly to quarterly frequencies, as it is usually the case of macroeconomic nowcasting.

A more recent variation of the original MIDAS model called TF-MIDAS, standing for Transfer Function MIDAS, is introduced by [4]. The authors demonstrate that TF-MIDAS is a general version of U-MIDAS and show that this model beats the latter in terms of out-of-sample nowcasting performance for several HF data generating processes in a set of simulation exercises.

A considerable number of studies compare the different variants of MIDAS models in terms of their nowcasting and forecasting accuracy when projecting macroeconomic variables.

A first group of papers focuses on the nowcasting and predictive performance of the original ADL-MIDAS model in comparison with a set of alternative non-MIDAS models. Specifically, [37] compares ADL-MIDAS model to bridge equations and an AR model through an empirical exercise, in which Euro Area GDP growth is nowcast for the period 2010Q1–2014Q4. The author finds that MIDAS models tend to outperform bridge equations for most predictors, but only for a few indicators do these models beat a simple AR model.

Jansen et al. [28] consider Euro Area and its five largest economies (Germany, France, Italy, Spain and the Netherlands) to test the predictive capacity of a wide range of models, including random walk, AR model, Bayesian VAR model (BVAR), bridge equation, dynamic factor model, MF-VAR and ADL-MIDAS. They analyze the predictive accuracy of these models in two disjoint evaluation samples: 1996Q1–2007Q4 and 2008Q1–2011Q3, which allows them to consider a stable and a volatile period. They conclude that MF-VAR and MIDAS models yield better predictions after the financial crisis, but this does not occur in stable times.

More recently [6] employ data from six developed countries (US, UK, Japan, France, Germany, Italy) and the Euro Area to obtain empirical evidence on the predictive performance of five classes of models (AR, Factor Augmented DL, MIDAS, BVAR, and DSGE model). They consider a general evaluation sample that goes from 1993Q1 up to 2011Q3 and also split that sample into 5-year windows for most of the considered countries. The conclusions are that MIDAS models work better at 1-period ahead horizons. Nevertheless, they also show *t*-statistics with large spreads, meaning that they work well for the median country but poorly for some individual countries. For 4-period ahead forecasts, BVAR clearly performs better.

Other papers comparing ADL-MIDAS with other forecasting models are [32], for the Euro Area with an evaluation sample 1999Q2–2008Q1, [15], again for the Euro Area with evaluation sample 2003Q1–2009Q1, [10], for US on 1985Q2–2005Q1, [9], for Canada on 2002Q1–2016Q2, [38], for Singapore on 2001Q1–2010Q4, [18], for Switzerland on 2005Q1–2015Q2, or [12], for Turkey on 2010Q2–2015Q1.

A second and less numerous group of papers compares the nowcasting and forecasting performance of U-MIDAS model with alternative non-MIDAS models. For example [2], consider U-MIDAS model versus the classical bridge equation model to build a daily indicator of growth for the Euro Area. The results show that forecasts obtained from U-MIDAS considering different indicators present a higher forecasting accuracy when they are combined with inverse Mean Square Error (MSE) weights.

In this sense, [31] compare the predictive performance of U-MIDAS versus Dynamic mixed-frequency Factor Model (DFM) for Baden Württemberg's regional GDP growth. The evaluation sample, in this case, spans from 2012Q1 to 2019Q3. The paper finds MIDAS-based predictions to be more robust and to outperform DFM slightly.

Last, a third set of papers is formed by those simultaneously comparing ADL-MIDAS and U-MIDAS predictive precision against other non-MIDAS models. For

instance, [13] analyses the nowcasting and short-term forecasting power of U-MIDAS and ADL-MIDAS against an AR and bridge equations benchmark models for the Euro Area in the evaluation sample 2007Q1–2012Q4. The results show that MIDAS models contribute to increasing predictive capacity. Additionally, differences in forecasting precision between ADL-MIDAS and U-MIDAS tend to vanish as the forecasting horizon increases.

Similarly, [17] compare U-MIDAS with the traditional ADL-MIDAS and an AR benchmark model in terms of their power to nowcast and short-term forecast US and Euro Area GDP growth. They use two evaluation samples for US (1985Q1–2006Q4 and 1985Q1–2011Q1) and one for the Euro Area (2003Q1–2010Q4). Similar results are observed for the two regions. Neither U-MIDAS nor ADL-MIDAS have a significantly superior performance during the more stable pre-crisis period, even failing against the AR benchmark. However, ADL-MIDAS and U-MIDAS significantly outperform the AR model for the crisis sample.

Additional papers simultaneously comparing ADL-MIDAS and U-MIDAS with other forecasting models are [29], employing Korean data with evaluation data 2000Q1–2013Q4, and [34], for the Philippines on 1999Q1–2019Q4.

In brief, researchers have yet to reach a consensus on which model, if any, presents the best performance at predicting macroeconomic variables. As this literature review suggests, it seems that MIDAS models (either, ADL- or U- MIDAS) have better results at nowcasting and short-term forecasting than most of the alternatives, although results may depend on the sample, country and HF indicator applied. Additionally, differences between MIDAS models' accuracy still need to be clarified. For this reason, we run a comparative exercise of forecasting performance of MIDAS models in which a new MIDAS-class model, not yet considered by the literature, is added.

Therefore, this chapter assesses the empirical nowcasting and forecasting performance of the three variants of MIDAS models: ADL-MIDAS, TF-MIDAS and U-MIDAS. With this aim, we use data from the four major Euro Area economies, France, Germany, Italy and Spain, covering the period 1995Q1–2020Q3, analyze the out-of-sample forecast for the evaluation sample 2011Q4–2020Q3, and consider several sub-periods with different levels of uncertainty. When the forecasting errors obtained in the whole sample are observed, we find a slightly higher accuracy, ranged between 2.4 and 4.1% in terms of the root mean squared forecast error (RMSFE), of TF-MIDAS models. However, when these errors are analyzed in a meta-regression, where we include model, country, indicator, horizon and sample dummies, we do not find a statistically significant difference in the predictive performance of the models under comparison, nor in terms of RMSFE or the mean absolute forecast error (MAFE). Part of the content presented in this paper and further conclusions drawn from the country, indicator, horizon and sample effects (not included here due to the lack of space) can be found in the unpublished Chap. 4 of the Ph.D. Thesis [3].

The chapter is organized in five sections, including the present introduction. In Sect. 2, compared MIDAS models are briefly reviewed, describing the main differences among the three variants considered. Section 3 details the empirical forecasting performance evaluation exercise and presents the first results of the relative out-of-

sample nowcasting performance. In Sect. 4, we estimate a meta-regression to obtain the individual effects of several variables on two accuracy forecasting measures. Finally, Sect. 5 summarizes the main results and concludes.

2 MIDAS Models Under Comparison

This section reviews the main theoretical features of the mixed-frequency models compared in the chapter: ADL-MIDAS, U-MIDAS and TF-MIDAS. Additionally, we briefly discuss the identification, estimation and how the nowcasts and forecasts have been computed.

Throughout the chapter we follow the notation used in [4], which is summarized in the following lines. A high-frequency (HF) indicator is denoted by letter x . Let t be the time index for this variable x , $t = 1, \dots, T$ (i.e., in this chapter, months), being T the last period for which data of variable x is available. L denotes the lag operator for this HF indicator. If x_t is the monthly industrial production index, then Lx_t will be the previous month's value of the index.

Similarly, let y be the low-frequency (LF) variable that is aimed to be nowcast, sampled at periods denoted by time index $t_q = 1, \dots, T_q$ (i.e., in this chapter, quarters), being T_q the last period for which data of variable y is available. Usually, $T \geq kT_q$, as observations of HF indicators are available earlier than LF ones. Past realizations of the LF variable will be denoted by the lag operator Z , where $Z \equiv L^k$. So, if y_{t_q} is quarterly GDP, then Zy_{t_q} will be the previous quarter's GDP value.

The HF indicator x is sampled k times between samples of y . For example, for quarterly GDP and monthly indicator, $k = 3$.

Finally, both the target variable y and the indicator x are assumed to be stationary, so these variables often correspond to a (log) differenced version of some raw series.

2.1 ADL-MIDAS Model

The original MIDAS (MIXed DATA Sampling) model was introduced by [22, 23, 25]. From now on, we will refer to it as ADL-MIDAS (or simply ADL-M) to emphasize the differences with other variants. In the ADL-M model, the response of the LF variable to an HF explicative variable is modeled through a Distributed Lag polynomial, but particular attention is paid to parsimony. To avoid the so-called "parameter proliferation" problem, lag coefficients are not free, but are defined as a function of a vector of a few parameters, θ , known as hyperparameters.

Andreou et al. [1] extend the DL specification of MIDAS model introducing an autoregressive term. In the case of only one HF indicator and only one autoregressive LF term, which is the most widely used form in practice, ADL-M equation for h -steps ahead nowcast can be written as

$$y_t = \beta_0 + \lambda y_{t-p_y, k} + \beta \sum_{j=1}^{K_{\max}} b(j; \boldsymbol{\theta}) x_{t-j-h+1} + \epsilon_t \quad t = k, 2k, \dots, Tk, \quad (1)$$

where K_{\max} is the maximum number of lags of the HF variable included in the model and p_y is the lag of the autoregressive term, which is a function of the nowcasting/forecasting horizon (i.e., $p_y = s \in \mathbb{N}$, such that h satisfies the condition $(s-1)k \leq h \leq sk$). The process ϵ_t is assumed to be a white noise in weak sense.

Function $b(j, \boldsymbol{\theta})$, a component of the lag polynomial, is used to model the weights assigned to each lag of the HF indicator. This function depends on the indicator's period, j , and the vector of hyperparameters $\boldsymbol{\theta}$. An overview of different weighting functions proposed in the literature is provided by [20], the most popular being Exponential Almon and Beta functions.¹ Several variations have been built upon the basic MIDAS model. A detailed summary of their main features can be found in [21].

Once ADL-M model is estimated by Non-Linear Least Squares (NLS), nowcasts and forecasts for y_{Tk+k} conditional on information available at period $Tk+k-h$ (i.e., h -steps ahead nowcasts and forecasts) are calculated as

$$\begin{aligned} \hat{y}_{Tk+k | Tk+k-h} &= \hat{\beta}_0 + \hat{\lambda} y_{Tk-(p_y-1)k} \\ &+ \hat{\beta} \sum_{j=1}^{K_{\max}} b(j; \hat{\boldsymbol{\theta}}) x_{Tk+k-j-h+1} \end{aligned} \quad (4)$$

where h is the forecasting horizon (notice that Eq. (4) generates nowcasting if $0 \leq h < k$ and forecasting if $h \geq k$) and p_y is the lag of the autoregressive term. In the forecasting performance exercise, the ADL-M predictions will be built from Eq. (4). We will consider several values for K_{\max} , ranging from 1 to 24, in order to account for ADL-M models with different levels of parsimony.

¹ The Exponential Almon weighting function was proposed in [24], and it has the following expression, with Q shape parameters:

$$b(j; \boldsymbol{\theta}) = \frac{\exp(\theta_1 j + \dots + \theta_Q j^Q)}{\sum_{j=0}^{K_{\max}} \exp(\theta_1 j + \dots + \theta_Q j^Q)}, \quad \text{where } \boldsymbol{\theta} = \{\theta_1, \theta_2, \dots, \theta_Q\}. \quad (2)$$

Beta weighting function, proposed for the first time in [23], includes only two shape parameters:

$$b(j; \boldsymbol{\theta}) = \frac{\beta\left(\frac{j}{K_{\max}}; \theta_1, \theta_2\right)}{\sum_{j=0}^{K_{\max}} \beta\left(\frac{j}{K_{\max}}; \theta_1, \theta_2\right)}, \quad \text{where } \boldsymbol{\theta} = \{\theta_1, \theta_2\}, \quad (3)$$

and $\beta(\cdot)$ is the Beta probability density function.

2.2 U-MIDAS Model

Koenig et al. [30] introduced a variant of MIDAS model known as U-MIDAS (hereinafter U-M), which was later thoroughly studied by [16, 17]. U-M does not employ functional distributed lag polynomials to model the relationship between x and y , but a linear lag polynomial.

The idea behind this variant is that when the difference in sampling frequencies is not large the risk of falling into the curse of dimensionality becomes less relevant, and so it does the need to resort to functional DL polynomials.

Similarly to the previous model, U-M model with one HF indicator and one autoregressive term is defined by:

$$y_t = \beta_0 + \lambda y_{t-p_y,k} + \sum_{j=1}^{K_{\max}} \beta_j x_{t-j-h+1} + \epsilon_t \quad t = k, 2k, \dots, Tk, \quad (5)$$

where, again, h is the forecasting horizon, K_{\max} is the maximum number of lags of the HF variable, and p_y is the lag of the autoregressive term.

Foroni et al. [17] state that basic ADL-M model can be considered nested in U-M specification because it is the result of imposing a particular dynamic pattern on it. An important computational advantage of U-M model over the basic MIDAS models is that it can be estimated by OLS, as long as lag orders p_y and K_{\max} are long enough to make the error term, ϵ_t , uncorrelated and so the weak white noise assumption can be validated.

Once U-M model is fitted by OLS, nowcasts and forecasts for y_{Tk+k} conditional on information available at period $Tk + k - h$ (i.e., h -steps ahead nowcasts and forecasts), when considering only one autoregressive term, are computed as

$$\hat{y}_{Tk+k | Tk+k-h} = \hat{\beta}_0 + \hat{\lambda} y_{TK-(p_y-1)k} + \sum_{j=1}^{K_{\max}} \hat{\beta}_j x_{t-j-h+1} \quad (6)$$

Similarly to ADL-M, the U-M predictions will be built from Eq. (6), and values ranged from 1 to 24 will be considered for K_{\max} to apply models with different levels of parsimony.

2.3 TF-MIDAS Model

Recently, [4] have proposed another variant of the basic MIDAS model, called TF-MIDAS model (or simply TF-M), in which instead of a DL polynomial expression, an alternative representation based on a transfer function is applied.

General TF-M model with one HF indicator is easier defined using two equations:

(1) the equation that models the relation between y and x :

$$y_t = \beta_0 + \sum_{j=1}^k \frac{a_j(Z)}{b_j(Z)} x_{t-j} + \eta_t, \text{ and} \quad (7a)$$

(2) the equation that models the noise:

$$\phi(Z)\eta_t = \theta(Z)\epsilon_t, \quad t = k, 2k, \dots, Tk. \quad (7b)$$

In Eq. (7a, 7b), $a_j(Z)$ and $b_j(Z)$ are finite lag polynomials and $Z \equiv L^k$, while $\phi(Z)$ and $\theta(Z)$ are, respectively, autoregressive and moving average polynomials of order p and q . Additionally, let $\phi(Z)$ and $\theta(Z)$ have all their zeros lying outside the unit circle and do not have common factors [see, e.g., 5], and ϵ_t be a weak white noise process.

For monthly-quarterly data (i.e., $k = 3$), one HF indicator and 1-step ahead nowcast, the previous equations become

$$y_t = \beta_0 + \frac{a_1(Z)}{b_1(Z)} x_{t-1} + \frac{a_2(Z)}{b_2(Z)} x_{t-2} + \frac{a_3(Z)}{b_3(Z)} x_{t-3} + \eta_t \quad (8a)$$

$$\phi(Z)\eta_t = \theta(Z)\epsilon_t, \quad t = k, 2k, \dots, Tk, \quad (8b)$$

where x_{t-1} is the second monthly observation of the current quarter, x_{t-2} is the first monthly observation of the current quarter, and x_{t-3} is the third monthly observation of the previous quarter.

TF-M model is estimated by exact ML. To do so, it is transformed into its equivalent state space formulation.² As ML convergence sometimes depends on the initial values of the parameters and TF-MIDAS usually has a considerable number of them, we suggest applying a procedure to get consistent estimates for those values prior to the ML estimation. Here we use the procedure by [19]. Then, the exact ML is computed using the standard Kalman filter equations for a state space model with stochastic inputs [see, 7] by iterating on the set of parameters. Obviously, the estimation through iterative methods may entail some drawbacks with respect to LS techniques, as computational cost and stability issues.

Once the TF-M model is estimated, nowcasts and forecasts for y_{Tk+k} conditional on information available at period $Tk + k - h$ (i.e., h -steps ahead nowcasts and forecasts) are calculated as

² In order to keep focused on the models' performance evaluation and comparison, we do not present in this chapter the ML function and its corresponding Kalman filter equations, as these are standard in the state space models literature. However, for readers unfamiliar with this type of formulations, all the equations needed to compute the ML can be found in [7], Sect. 5.3.2, where expression (5.50) specifically shows the log-likelihood function used.

$$\begin{aligned}
\hat{y}_{T_{k+k} | T_{k+k-h}} &= \hat{\beta}_0 + \frac{\hat{a}_1(Z)}{\hat{b}_1(Z)} x_{T_{k+k-h}} \\
&+ \frac{\hat{a}_2(Z)}{\hat{b}_2(Z)} x_{T_{k+k-h-1}} + \frac{\hat{a}_3(Z)}{\hat{b}_3(Z)} x_{T_{k+k-h-2}} \\
&+ \frac{\hat{\theta}(Z)}{\hat{\phi}(Z)} \hat{\epsilon}_{T_{k+k}}
\end{aligned} \tag{9}$$

where $\hat{\phi}(Z) = 1 + \hat{\phi}_1 Z + \hat{\phi}_2 Z^2 + \dots + \hat{\phi}_p Z^p$, $\hat{\theta}(Z) = \hat{\theta}_1 Z + \hat{\theta}_2 Z^2 + \dots + \hat{\theta}_q Z^q$, and $\hat{a}_j(Z)$ and $\hat{b}_j(Z)$, with $j = 1, 2, 3$, are finite lag polynomials, whose order will be specified by means of information criteria (see Sect. 3.2 and Table 3 for more detail).³

3 Forecasting Performance Evaluation

This section first describes the data used in the forecasting evaluation exercise. Later, it details how the performance evaluation has been designed. Finally, a discussion of the unconditional distribution of the forecasting errors is also presented.

3.1 Data Description

We employ data from the four major economies of the Euro Area (France, Germany, Italy and Spain) in the period 1995Q1–2020Q3. In all cases, we have transformed GDP data to make it stationary, so our target variable is the quarterly change in seasonally adjusted log real GDP. The source of GDP data is Eurostat.⁴

We consider a set of fifteen monthly-observed economic indicators for each GDP, whose description is reported in Table 1. Each indicator series is seasonally adjusted and, as GDP, transformed to induce stationarity. These data were also obtained from Eurostat.⁵

³ Notice that the polynomial $\hat{\theta}(Z)$ does not include the unit term as $\hat{\epsilon}_{kT_q+es}$ is not known at period $T_q k + es$.

⁴ GDP data were downloaded from the webpage: <https://ec.europa.eu/eurostat/web/national-accounts/data/database>.

⁵ Volume Index of Industrial Production indicators were downloaded from the webpage: https://ec.europa.eu/eurostat/databrowser/view/sts_inpr_m/default/table?lang=en
Consumer Confidence Indicators were downloaded from the webpage: https://ec.europa.eu/info/business-economy-euro/indicators-statistics/economic-databases/business-and-consumer-surveys/download-business-and-consumer-survey-data/time-series_en.

Table 1 Description of the indicators

	Name	Description
1	IPI	Volume index of industrial production (Mining and quarrying, manufacturing, electricity, gas, etc.)
2	IPI2	Volume index of industrial production (IPI; construction)
3	COF	Consumer confidence indicator: total value
4	COF1	Consumer confidence indicator: financial situation over last 12 months
5	COF2	Consumer confidence indicator: financial situation over next 12 months
6	COF3	Consumer confidence indicator: general economic situation over last 12 months
7	COF4	Consumer confidence indicator: general economic situation over next 12 months
8	COF5	Consumer confidence indicator: price trends over last 12 months
9	COF6	Consumer confidence indicator: price trends over next 12 months
10	COF7	Consumer confidence indicator: unemployment expectations over next 12 months
11	COF8	Consumer confidence indicator: major purchases at present
12	COF9	Consumer confidence indicator: major purchases over next 12 months
13	COF10	Consumer confidence indicator: savings at present
14	COF11	Consumer confidence indicator: savings over next 12 months
15	COF12	Consumer confidence indicator: statement on financial situation of household

Source European Commission

3.2 Evaluation Design

As [27] show evidence of changing predictive capacity over time, we decide to separate our out-of-sample GDPs forecasts in three disjoint periods of three years (12 quarterly forecasts) each. The dates are 2011Q4–2014Q3, 2014Q4–2017Q3 and 2017Q4–2020Q3. We choose these periods deliberately to analyze the behavior of the methods in three substantially different economic contexts: the European sovereign debt crisis in the first period, a recovery and more stable phase during the second period, and a third convulsive period struck by the COVID-19 pandemic. We will check if changes in the underlying structure of the economies and the exogenous shocks affect the methods' relative forecasting performances.

In addition to the different periods, each prediction is calculated for four countries (Germany, France, Italy and Spain), seven forecast horizons, fifteen indicators and nine methods. Horizons have been chosen to investigate if nowcasting and forecast-

Table 2 Model acronyms

	Name	Description
1	ADL-M ₃	ADL-MIDAS with up to 3 lags of HF variable
2	ADL-M ₆	ADL-MIDAS with up to 6 lags of HF variable
3	ADL-M ₁₂	ADL-MIDAS with up to 12 lags of HF variable
4	ADL-M ₂₄	ADL-MIDAS with up to 24 lags of HF variable
5	U-M ₃	U-MIDAS with up to 3 lags of HF variable
6	U-M ₆	U-MIDAS with up to 6 lags of HF variable
7	U-M ₁₂	U-MIDAS with up to 12 lags of HF variable
8	U-M ₂₄	U-MIDAS with up to 24 lags of HF variable
9	TF-M	TF-MIDAS (see Table 3 for models structure)

ing affect the methods' performance differently. We set the forecasting horizons to 0, 1, 2, 3, 6, 9 and 12.

On the other hand, Table 2 presents the acronyms and a short description of the models employed. The methods are divided into four ADL-MIDAS models, ADL-M₃, ADL-M₆, ADL-M₁₂ and ADL-M₂₄, four U-MIDAS models, U-M₃, U-M₆, U-M₁₂ and U-M₂₄ and the TF-M model.

Each ADL-M _{K_{\max}} and U-M _{K_{\max}} method considers a set of models that range from 1 to K_{\max} lags of the HF variable and one autoregressive term with lag p_y . In the case of TF-M, a set of sixteen models are considered, see Table 3, covering different orders of lag polynomials $a_j(Z)$, $b_j(Z)$, $\theta(Z)$ and $\phi(Z)$. Every specification for ADL-M _{K_{\max}} , U-M _{K_{\max}} and TF-M is then fitted. We choose one specification for each method with the in-sample lowest BIC information criterion for each new observation. These chosen models are then used to compute the corresponding predictions.⁶

Finally, each sub-period analyzed is made up of twelve out-of-sample forecasts. Every prediction is computed using a recursive (expanding) forecasting scheme, i.e., we use all observations available from the beginning of the sample up to the forecasting origin in both the identification and estimation process.

In order to check the robustness of the results obtained, two measures of point forecasting performance are used: (1) the root mean squared forecast error (RMSFE), and (2) the mean absolute forecast error (MAFE). Each of these measures is calculated with the previous twelve observations.

⁶ In some specific cases, probably due to the presence of outliers, data was adjusted in order to not alter subsequent results and conclusions. In practice, the detection and treatment of these extreme nowcasts/forecasts would be easily addressed by an analyst. In summary, less than 0.9% of predictions were adjusted, most of them corresponding to Italy and Spain. Regarding the methods applied, the adjustments distribute uniformly, except for U-M₃ and ADL-M₃ that account for half of the adjusted values corresponding to each one of the rest of the methods. The exact same estimations have been calculated without these corrections and conclusions do not vary significantly.

Table 3 Polynomials structure of the TF-M model (8a–8b)

Model structure	
1	$a_j(Z) = a_{j,0}; b_j(Z) = 1, j = 1, 2, 3; \phi(Z) = 1 + \phi Z; \theta(Z) = 1 + \theta Z$
2	$a_j(Z) = a_{j,0}; b_j(Z) = 1, j = 1, 2, 3; \phi(Z) = 1 + \phi Z; \theta(Z) = 1$
3	$a_j(Z) = a_{j,0}; b_j(Z) = 1, j = 1, 2, 3; \phi(Z) = 1; \theta(Z) = 1 + \theta Z$
4	$a_j(Z) = a_{j,0}; b_j(Z) = 1, j = 1, 2, 3; \phi(Z) = \theta(Z) = 1$
5	$a_j(Z) = a_{j,0}, j = 1, 2, 3; b_j(Z) = 1 + b_{j,1}Z, j = 1, 2; b_3(Z) = 1; \phi(Z) = 1 + \phi Z; \theta(Z) = 1 + \theta Z$
6	$a_j(Z) = a_{j,0}, j = 1, 2, 3; b_j(Z) = 1 + b_{j,1}Z, j = 1, 2; b_3(Z) = 1; \phi(Z) = 1 + \phi Z; \theta(Z) = 1$
7	$a_j(Z) = a_{j,0}, j = 1, 2, 3; b_j(Z) = 1 + b_{j,1}Z, j = 1, 2; b_3(Z) = 1; \phi(Z) = 1; \theta(Z) = 1 + \theta Z$
8	$a_j(Z) = a_{j,0}, j = 1, 2, 3; b_j(Z) = 1 + b_{j,1}Z, j = 1, 2; b_3(Z) = 1; \phi(Z) = \theta(Z) = 1$
9	$a_j(Z) = a_{j,0} + a_{j,1}Z; b_j(Z) = 1, j = 1, 2, 3; \phi(Z) = 1 + \phi Z; \theta(Z) = 1 + \theta Z$
10	$a_j(Z) = a_{j,0} + a_{j,1}Z; b_j(Z) = 1, j = 1, 2, 3; \phi(Z) = 1 + \phi Z; \theta(Z) = 1$
11	$a_j(Z) = a_{j,0} + a_{j,1}Z; b_j(Z) = 1, j = 1, 2, 3; \phi(Z) = 1; \theta(Z) = 1 + \theta Z$
12	$a_j(Z) = a_{j,0} + a_{j,1}Z; b_j(Z) = 1, j = 1, 2, 3; \phi(Z) = \theta(Z) = 1$
13	$a_j(Z) = a_{j,0} + a_{j,1}Z, j = 1, 2, 3; b_j(Z) = 1 + b_j Z, j = 1, 2; b_3(Z) = 1; \phi(Z) = 1 + \phi Z; \theta(Z) = 1 + \theta Z$
14	$a_j(Z) = a_{j,0} + a_{j,1}Z, j = 1, 2, 3; b_j(Z) = 1 + b_j Z, j = 1, 2; b_3(Z) = 1; \phi(Z) = 1 + \phi Z; \theta(Z) = 1$
15	$a_j(Z) = a_{j,0} + a_{j,1}Z, j = 1, 2, 3; b_j(Z) = 1 + b_j Z, j = 1, 2; b_3(Z) = 1; \phi(Z) = 1; \theta(Z) = 1 + \theta Z$
16	$a_j(Z) = a_{j,0} + a_{j,1}Z, j = 1, 2, 3; b_j(Z) = 1 + b_j Z, j = 1, 2; b_3(Z) = 1; \phi(Z) = \theta(Z) = 1$

We use E4 [7] and Midas [20] MatLab Toolboxes to perform the estimation and prediction processes.⁷

3.3 Unconditional Distribution of the Forecasting Errors

This section investigates the relative performance of the considered MIDAS methods in terms of their nowcasting and forecasting accuracy according to the two measures, RMSFE and MAFE. The results obtained for RMSFE are summarized in Fig. 1.⁸

According to this figure, TF-M presents the lowest average RMSFE (in white in Fig. 1), although the discrepancy with respect to the other methods is not very large. The percentage difference in terms of average RMSFE between the best performing

⁷ Matlab code to estimate TF-MIDAS model is available from the authors.

⁸ The analogous figure for MAFE can be found in [3]. The main conclusions do not differ significantly from those for RMSFE.

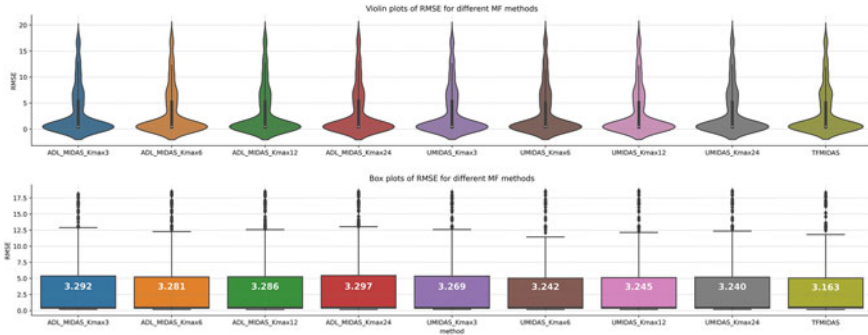


Fig. 1 Violin and box plots for RMSFE by nowcasting/forecasting method. Values displayed in white inside the boxes are the average RMSFE

method (TF-M) and the second best (U-M₂₄) is 2.4%, and with the worse one (ADL-M₂₄) is 4.1%. Moreover, the RMSFE distributions displayed by the violin and box plots look really similar compared to the differences observed by countries, indicators and horizons.⁹ Similar results are obtained for MAFE measure.

4 Forecasting Performance: A Meta-Regression Analysis

As mentioned, when looking at the mean of the forecasting performance measures by methods in the previous section, we do not account for the potential effect of the rest of the variables. In this section, we address this matter through a meta-regression.

4.1 Description of the Meta-Regression Analysis

We aim to study how the forecasting performance measure (RMSFE or MAFE) varies with the method applied, the source country, the HF indicator, the horizon and the specific sample considered. For that, we will run a meta-regression with the whole sample, containing 11,340 observations.

The main feature we consider is the model applied. To analyze its influence on the forecasting performance metric, we use eight dummy variables, each corresponding to a specific ADL-M and U-M model. The benchmark model is thus TF-M.

The impact of the country origin of the dataset is evaluated by including three dummy variables, choosing Germany as the baseline country.

⁹ Analogous figures for RMSFE by countries, indicators and horizons can be found in [3].

The third feature whose impact on the forecasting performance is assessed is the HF indicator used, so we create fourteen dummy variables, one for each indicator, being IPI the baseline indicator.

We also include six dummy variables corresponding to horizons 1, 2, 3, 6, 9 and 12. The effect of horizon 0 is captured by the constant.

Finally, we also want to study the effect of the specific sample chosen on the forecasting performance measure. This way, we evaluate how stable or uncertain periods affect the nowcasting or forecasting performance metrics. So, we create two dummy variables: sample1 for the period 2011Q4–2014Q3 and sample2 for 2014Q4–2017Q3, being 2017Q4–2020Q3 the benchmark period.

From these definitions the meta-regression equation is

$$\begin{aligned}
 AM_i = & \beta_0 + \sum_{j=1}^8 \beta_j^M D_{ij}^M + \sum_{j=1}^3 \beta_j^C D_{ij}^C \\
 & + \sum_{j=1}^{14} \beta_j^I D_{ij}^I + \sum_{j=1}^6 \beta_j^H D_{ij}^H \\
 & + \sum_{j=1}^2 \beta_j^S D_{ij}^S + \epsilon_i,
 \end{aligned} \tag{10}$$

where D_i^M , D_i^C , D_i^I , D_i^H and D_i^S are, respectively, Model, Country, Indicator, Horizon and Sample dummy variables for each AM_i (Accuracy Measure) observation obtained. AM_i is either the RMSFE or MAFE discrepancy quantity, computed with 12 observations characterized by the variables in Eq. (10) for the observation i . Obviously, dummy variables do not include TF-M, Germany, IPI, horizon0 and sample3, as these effects are captured by β_0 . The value of the rest of coefficients (all β s different from β_0) are interpreted as gains/losses relative to the benchmark model.

4.2 Meta-Regression Main Results

Table 4 presents estimates of the regression in Eq. (10), for RMSFE and MAFE accuracy measures. Results for both metrics are very similar, except for the expected different scale. As the stars denoting the statistical significance show, regressors' significance do not depend on the metric considered.

First, as the unconditional analysis showed, the β s corresponding to the estimation methods are all greater than zero, suggesting a better performance of the TF-M model. In fact, for RMSFE, the loss of precision when not using TF-M ranges from 2.4% to 4.1% (U-M₂₄ and ADL-M₂₄ coefficients, respectively, in terms of the mean dependent variable). The percentages for MAFE, although lower, show the same picture. However, when looking at the statistical significance, none of the methods

Table 4 Main meta-regression

Variable	RMSFE	MAFE	Variable (cont.)	RMSFE	MAFE
constant	6.8072***	3.0862***	IPI2	0.2153**	0.0875**
ADL-M ₃	0.1291	0.0230	COF	1.1614***	0.4269***
ADL-M ₆	0.1180	0.0236	COF1	1.1972***	0.4493***
ADL-M ₁₂	0.1226	0.0267	COF2	1.1451***	0.4112***
ADL-M ₂₄	0.1336	0.0332	COF3	1.0932***	0.3939***
U-M ₃	0.1056	0.0150	COF4	0.6538***	0.2418***
U-M ₆	0.0785	0.0057	COF5	0.9766***	0.3555***
U-M ₁₂	0.0812	0.0085	COF6	0.9856***	0.3440***
U-M ₂₄	0.0765	0.0072	COF7	0.9574***	0.3454***
France	1.8110***	0.6613***	COF8	0.7321***	0.2704***
Italy	1.4176***	0.5161***	COF9	1.1802***	0.4218***
Spain	2.7054***	0.9578***	COF10	1.0755***	0.4012***
Horizon1	-0.0485	-0.0243	COF11	1.0528***	0.3748***
Horizon2	-0.1978**	-0.0765***	COF12	1.0638***	0.3928***
Horizon3	0.1251	0.0385	Sample1	-7.9969***	-3.3184***
Horizon6	-1.2200***	-0.3369***	Sample2	-8.3018***	-3.4809***
Horizon9	-1.4023***	-0.3963***			
Horizon12	-1.4122***	-0.3926***			
N	11340	11340			
Mean dep. variable	3.2534	1.5275			
R ²	0.7899	0.8434			

Notes **, *** and **** mean rejection of the Null of the corresponding coefficient equal to zero at 10%, 5% and 1% level, respectively

presents a different nowcasting/forecasting accuracy, at least at a 10% significance level; see Table 4.

Regarding country dummies, all of them have a statistically significant and positive effect on the error quantity, either RMSFE or MAFE, meaning that predictions for France, Italy and particularly Spain are less accurate than those computed for Germany, apparently no matter the sample, method, horizon or indicator. This result would be originated in a more stable and thus predictable economic environment in Germany compared to the other countries in the sample.

All indicator dummies present a statistically significant increase in RMSFE and MAFE with respect to the use of IPI as HF indicator. First, this reveals that the inclusion of construction's production in IPI2 does not contribute to improving the accuracy of GDP's predictions. Second, it also shows that COF-related indicators do not provide more valuable information than IPI to reduce prediction errors.

Concerning the effect of the horizon on the accuracy measure, only horizon1 and horizon3 dummies have no statistically significant effect with respect to horizon0.

The rest of horizon's dummies present a negative relevant impact on RMSFE and MAFE compared to horizon0. Contrary to the literature, and leaving aside horizon2, whose effect deviates from the rest of shorter horizons, these results seem to point out that MIDAS models have a better performance at longer horizons, i.e., at forecasting much more than at nowcasting. This somehow baffling result might be induced by the interaction of these dummies with others in the meta-regression. Therefore, a more profound analysis in this direction is needed here to understand the roots of this finding.

Finally, β s associated to both period dummies reveal a strong negative effect on RMSFE and MAFE compared to the benchmark sample 3. This result is easily understandable, as sample 3 corresponds to the period including COVID-19 pandemic and the last observations in the sample are much more unpredictable than any in the other two sample periods.

In summary, the main findings obtained from this meta-regression analysis are: (i) the TF-M lowest error quantities are not large enough to show statistically significant coefficients of the rest of the model dummies, implying that either ADL-M or U-M models report a similar forecasting performance relative to TF-M, once we control for the other factors; (ii) all country dummies are relevant variables, indicating that nowcasts and forecasts for Italy, France and Spain are less accurate than the ones for Germany; (iii) using other indicator than IPI results in a statistically significant increase in RMSFE and MAFE; (iv) horizon dummies result to be most relevant, although with differences: while for shorter horizons (i.e., nowcasts), only horizon 2 presents a clear gain in accuracy in comparison to horizon 0, longer horizons (i.e., forecasts) do show in all cases a statistically significant reduction in RMSFE and MAFE relative to horizon 0; (v) predictions for the first period and second period (i.e., 2011Q4–2014Q3 and 2014Q4–2017Q3) are much more accurate than the ones for the third period (i.e., 2017Q4–2020Q3), which is consistent with the fact that this period involves a huge uncertainty associated with COVID-19 pandemic; and, finally, (vi) conclusions do not depend on the specific accuracy measure applied, as RMSFE and MAFE yield very similar results.

5 Conclusions

This chapter attempts to shed some light on the use of the different variants of MIDAS models in forecasting. To do so, we assess the empirical nowcasting and forecasting performance of the ADL-, U- and TF- MIDAS family models. We use data from the four main Euro Area economies (France, Germany, Italy and Spain) for the period 1995Q1–2020Q3, accounting for different HF indicators and horizons. We report the results of the out-of-sample forecasting analysis for the sample 2011Q4–2020Q3 and three disjoint sub-periods characterized by different levels of volatility and uncertainty: 2011Q4–2014Q3 (European sovereign debt crisis), 2014Q4–2017Q3 (recovery and stable phase) and 2017Q4–2020Q3 (including COVID-19 pandemic shock).

The predictive accuracy of the distinct MIDAS models is compared according to two accuracy measures: RMSFE and MAFE.

The results of an unconditional analysis reveal a better performance of TF-MIDAS in terms of lowest RMSFE and MAFE. However, a meta-regression with the whole sample does not show a statistically significant difference in the predictive accuracy of the compared methods at standard significance levels. Some other interesting features were found instead: (i) German GDP seems to be more predictable than Italy, France and, specially, Spain's; (ii) IPIs turn out to be the best HF indicators; (iii) contrary to the literature, MIDAS models seem to perform better at forecasting (longer horizons) than nowcasting; and, (v) as expected, the studied sub-periods can be decreasingly sorted according to predictability as 2014Q4–2017Q3, 2011Q4–2014Q3 and 2017Q4–2020Q3. All these findings are robust to the error metric used, either RMSFE or MAFE.

Finally, these results were obtained without including interaction terms in the meta-regression, which could cast some light on the conclusions drawn. This will be the object of future research.

References

1. Andreou, E., Ghysels, E., Kourtellis, A.: Should macroeconomic forecasters use daily financial data and how? *J. Bus. Econ. Stat.* **31**(2), 240–251 (2013). <https://doi.org/10.1080/07350015.2013.767199>
2. Aprigliano, V., Foroni, C., Marcellino, M., Mazzi, G., Venditti, F.: A daily indicator of economic growth for the Euro Area. *Int. J. Comput. Econ. Econom.* **7**(1/2), 43–63 (2017). <https://ideas.repec.org/a/ids/ijcome/v7y2017i1-2p43-63.html>
3. Bonino-Gayoso, N.: Mixed-frequency models. Alternative comparison and nowcasting of macroeconomic variables. Ph.D. thesis. Universidad Complutense de Madrid (2022). <https://docta.ucm.es/entities/publication/4379324f-9ddb-4578-99b3-a971b8fd47ee>
4. Bonino-Gayoso, N., Garcia-Hiernaux, A.: TF-MIDAS: a transfer function based mixed-frequency model. *J. Stat. Comput. Simul.* **91**(10), 1980–2017 (2021). <https://doi.org/10.1080/00949655.2021.1879082>
5. Box, G.E.P., Jenkins, G.M.: *Time Series Analysis: forecasting and Control*. Holden-Day (1976)
6. Carriero, A., Galvão, A.B., Kapetanios, G.: A comprehensive evaluation of macroeconomic forecasting methods. *Int. J. Forecast.* **35**(4), 1226–1239 (2019). <https://doi.org/10.1016/j.ijforecast.2019.02.007>
7. Casals, J., Garcia-Hiernaux, A., Jerez, M., Sotoca, S.: *State-Space Methods for Time Series Analysis: theory, Applications and Software*. Chapman & Hall (2016)
8. Castle, J., Hendry, D.: Forecasting and nowcasting macroeconomic variables: a methodological overview. Economics Series Working Paper 674. University of Oxford, Department of Economics (2013). <https://ideas.repec.org/p/oxf/wpaper/674.html>
9. Chernis, T., Sekkel, R.: A dynamic factor model for nowcasting Canadian GDP growth. *Empir. Econ.* **53**(1), 217–234 (2017). <https://doi.org/10.1007/s00181-017-1254-1>
10. Clements, M.P., Galvão, A.B.: Macroeconomic forecasting with mixed-frequency data: forecasting output growth in the United States. *J. Bus. Econ. Stat.* **26**(4), 546–554 (2008). <https://doi.org/10.1198/073500108000000015>
11. Clements, M.P., Galvão, A.B.: Forecasting US output growth using leading indicators: an appraisal using MIDAS models. *J. Appl. Econom.* **24**(7), 1187–1206 (2009). <https://doi.org/10.1002/jae.1075>

12. Dogan, B.S., Midiliç, M.: Forecasting Turkish real GDP growth in a data-rich environment. *Empir. Econ.* **56**, 367–395 (2019). <https://doi.org/10.1007/s00181-017-1357-8>
13. Duarte, C.: Autoregressive augmentation of MIDAS regressions. Tech. Rep. w201401, Banco de Portugal, Economics and Research Department (2014). <https://ideas.repec.org/p/ptu/wpaper/w201401.html>
14. Duarte, C., Rodrigues, P.M.M., Rua, A.: A mixed frequency approach to the forecasting of private consumption with ATM/POS data. *Int. J. Forecast.* **33**(1), 61–75 (2017). <https://doi.org/10.1016/j.ijforecast.2016.08.003>
15. Foroni, C., Marcellino, M.: A comparison of mixed frequency approaches for nowcasting Euro Area macroeconomic aggregates. *Int. J. Forecast.* **30**(3), 554–568 (2014). <https://doi.org/10.1016/j.ijforecast.2013.01.010>
16. Foroni, C., Marcellino, M., Schumacher, C.: U-MIDAS: MIDAS regressions with unrestricted lag polynomials. CEPR Discussion Paper 8828 (2012). <http://econpapers.repec.org/paper/cprceprdp/8828.htm>
17. Foroni, C., Marcellino, M., Schumacher, C.: Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *J. R. Stat. Soc. Ser. A* **178**(1), 57–82 (2015). <https://doi.org/10.1111/rssa.12043>
18. Galli, A., Hepenstrick, C., Scheufele, R.: Mixed-frequency models for tracking short-term economic developments in Switzerland. *Int. J. Central Bank.* **15**(2), 151–178 (2019). <https://ideas.repec.org/a/ijc/ijcjou/y2019q2a5.html>
19. Garcia-Hiernaux, A., Casals, J., Jerez, M.: Fast estimation methods for time-series models in state-space form. *J. Stat. Comput. Simul.* **79**(2), 121–134 (2009). <https://doi.org/10.1080/00949650701617249>
20. Ghysels, E.: Matlab toolbox for mixed sampling frequency data analysis using MIDAS regression models. Tech. rep. (2014)
21. Ghysels, E., Kvedaras, V., Zemlys-Balevičius, V.: Mixed data sampling (MIDAS) regression models. In: *Handbook of Statistics*, vol. 42, chap. 4, pp. 117–153. Elsevier (2020). <https://doi.org/10.1016/bs.host.2019.01.005>
22. Ghysels, E., Santa-Clara, P., Valkanov, R.: The MIDAS touch: mixed data sampling regression models. Working paper, UNC and UCLA (2002). <http://econpapers.repec.org/paper/circirwor/2004s-20.htm>
23. Ghysels, E., Santa-Clara, P., Valkanov, R.: Predicting volatility: getting the most out of return data sampled at different frequencies. Anderson School of Management working paper and UNC Department of Economics working paper (2003). <https://papers.ssm.com/abstract=440941>
24. Ghysels, E., Santa-Clara, P., Valkanov, R.: There is a risk-return trade-off after all. *J. Financ. Econ.* **76**(3), 509–548 (2005). <https://doi.org/10.1016/j.jfineco.2004.03.008>
25. Ghysels, E., Santa-Clara, P., Valkanov, R.: Predicting volatility: getting the most out of return data sampled at different frequencies. *J. Econom.* **131**(1), 59–95 (2006). <https://doi.org/10.1016/j.jeconom.2005.01.004>
26. Ghysels, E., Sinko, A., Valkanov, R.: MIDAS regressions: further results and new directions. *Econom. Rev.* **26**(1), 53–90 (2007). <https://doi.org/10.1080/07474930600972467>
27. Giacomini, R., Rossi, B.: Forecast comparison in unstable environments. *J. Appl. Econom.* **25**, 595–620 (2010). <https://doi.org/10.1002/jae.1177>
28. Jansen, W.J., Jin, X., de Winter, J.M.: Forecasting and nowcasting real GDP: comparing statistical models and subjective forecasts. *Int. J. Forecast.* **32**(2), 411–436 (2016). <https://doi.org/10.1016/j.ijforecast.2015.05.008>
29. Kim, H.H., Swanson, N.R.: Methods for backcasting, nowcasting and forecasting using factor-MIDAS: with an application to Korean GDP. *J. Forecast.* **37**(3), 281–302 (2017). <https://doi.org/10.1002/for.2499>
30. Koenig, E.F., Dolmas, S., Piger, J.: The use and abuse of real-time data in economic forecasting. *Rev. Econ. Stat.* **85**(3), 618–628 (2003). <https://doi.org/10.1162/003465303322369768>
31. Kuck, K., Schweikert, K.: Forecasting Baden-Württemberg’s GDP growth: MIDAS regressions versus dynamic mixed-frequency factor models. *J. Forecast.* **40**(5), 861–882 (2020). <https://doi.org/10.1002/for.2743>

32. Kuzin, V., Marcellino, M., Schumacher, C.: MIDAS versus mixed-frequency VAR: nowcasting GDP in the Euro area. *Int. J. Forecast.* **27**(2), 529–542 (2011). <https://doi.org/10.1016/j.ijforecast.2010.02.006>
33. Lehrer, S., Xie, T., Zhang, X.: Social media sentiment, model uncertainty, and volatility forecasting. *Econ. Modell.* **102**(105556) (2021). <https://doi.org/10.1016/j.econmod.2021.105556>
34. Mariano, R.S., Ozmucur, S.: Predictive Performance of Mixed-Frequency Nowcasting and Forecasting Models (with Application to Philippine Inflation and GDP Growth). SSRN Pap. (2020). <https://doi.org/10.2139/ssrn.3666196>
35. Mulero, R., Garcia-Hiermaux, A.: Forecasting Spanish unemployment with Google Trends and dimension reduction techniques. *SERIEs* **12**(3), 329–349 (2021). <https://doi.org/10.1007/s13209-021-00231-x>
36. Schumacher, C.: MIDAS regressions with time-varying parameters: an application to corporate bond spreads and GDP in the euro area. In: Annual Conference 2014 (Hamburg): evidence-Based Economic Policy, Verein für Socialpolitik/German Economic Association (2014). <http://econpapers.repec.org/paper/zbwvfsc14/100289.htm>
37. Schumacher, C.: A comparison of MIDAS and bridge equations. *Int. J. Forecast.* **32**(2), 257–270 (2016). <https://doi.org/10.1016/j.ijforecast.2015.07.004>
38. Tsui, A.K., Yang, Xu., C., Zhaoyong, Z.: Macroeconomic forecasting with mixed data sampling frequencies: evidence from a small open economy. *J. Forecast.* **37**(6), 666–675 (2018). <https://doi.org/10.1002/for.2528>

Relative Measures of Forecasting: Lambda-Family-Measures



Miguel Ángel Ruiz Reina 

Abstract We develop a robust theoretical and practical scheme for automatic forecasting model selection using R programming. We present Lambda-Family-Measures for forecasting by point in time series to solve the problem of lack of definition. Through the work scheme, we can objectively quantify which model has the “best” forecasting capacity with dimensionless measures. The empirical results show no significant differences among predictive models for 53 series analysed from the tourism sector in Spain. Using data from the Spanish National Institute of Statistics, we compare forecasting capacity data from January 1999 to December 2019. Our results make it possible to empirically measure and quantify the accuracy of the models through multiple series and forecasting models. This proposed methodology supposes a homogenisation for the forecasting model comparisons. In particular, we use measures based on minimising the prediction error, the measure of the Root Mean Square Error criterion for this work, not the only option. The proposed solution is a contribution to statistics and time-series analysis.

Keywords Forecast accuracy · Time series · Lambda theorem of accuracy

1 Introduction

Forecasting modelling and evaluation have always been at the core of scientific research; traditionally, forecasting has been a field of statisticians and econometricians. Nowadays, companies are hiring data scientists with computational forecasting tasks [1], and the application techniques have evolved in the last 40 years depending on the researchers’ skills [2]. Theoretical and practical growth since the beginning of the twenty-first century [3] has allowed significant advances in applying computational techniques in analytics and data science [4]. Traditional methods of forecasting, despite being considered older methods, many of these

M. Á. R. Reina (✉)

Department of Applied Economics (Statistics and Econometrics), University of Malaga, Malaga, Spain

e-mail: ruizreina@uma.es

approaches are robust and not prone to overfitting more modern techniques. These models lay the foundation for any comparison in the field of forecasting based on uncorrelated forecast errors. Machine learning techniques have been a competitor against traditional models in the literature [5, 6]. The combination and assembly of the techniques formally contribute to the improvements in computer science [4]. In practice, to propose accuracy measures and compare predictive models, but generally, they are not applicable from one definition, undefined results, and can cause problems of misleading results [7]. On other occasions, ruptures occur, caused by parameter estimation errors and forecast modelling [8].

This work aims to provide a model selection strategy, so-called *Lambda-Family-Measures*, for forecasting by point in time series. The benefits are immediate and can be summarised as follows: first, establish a work scheme referenced to forecasting horizon windows; in this way, it is possible to see the behaviour of the models in the different scenarios h . Second, it introduces a robust contrast for differences in variances using the median as a reference based on non-Gaussian normality [9]. Third, contribute to the solution when the predictive models contain zeros or exact predictions not covered by literature [7], overcoming limitations of interpretation among relative measures or even problems of lack of definition. This solution is the *Lambda-Relative-Measures* that supposes a contribution to one of the main limitations of the relative accuracy measures. In our theoretical development, we propose using the Root of Mean Square Error - RMSE for predictions in time series windows horizons h . This fact will allow the development of the *Lambda-Relative-Measures Theorem*.

To demonstrate *Lambda-Relative-Measures'* accuracy, we proved a series of actual data on hotel overnight stays in Spain and 52 provinces from January 1999 to December 2019. We are interested in comparing the accuracy of six methods with the R package automatically: (1) SARIMA [10]; (2) Singular Spectrum Analysis—SSA [11]; (3) ARAR algorithm—ARAR [12, 13]; (4) Hierarchical Neural Networks—HNN [14]; (5) Holt—Winter Methods, Holt—Winter Additive (HWA) and Holt—Winter Multiplicative (HWM) [15].

This work presents a tool to select the “best” point forecast depending on the error or accuracy [16], and the objective is to establish a criterion for the selection under the criteria of reliability, robustness and interpretability, avoiding undefined results. To test model selection criteria and overcome the limitations above, we propose estimates in the sample period from January 1999 to December 2018 to subsequently evaluate the ex-post predictive capacity from January 2019 to December 2019. The models are tested for monthly windows horizons $h = 3, 6, 12$.

The development of this document consists of four more sections added to this first introductory section. The second section describes the background where basic concepts of accuracy and the most relevant are discussed. The third section develops the methodological aspect of choosing the “best” model under the three principles of reliability, robustness and interpretability. The fourth section forms the empirical section with the previous modelling for the aggregate series. The fifth section is dedicated to the conclusions of this study and the final lines on future research. Finally, we include an appendix with empirical results after the bibliographical references.

2 Forecast Errors: A Short Literature Review

The forecast errors are generated by the mismatch of the modelling estimate with its Data Generating Process (DGP). These errors are explained by contamination from outliers and multiple breaks at unknown times. Every analysis process presents the costs and benefits of selection for forecasting. We could deduce that the costs of model selection are less than the benefit of finding systematic flaws in the forecast [4]. From a scientific point of view, we can consider that the first formal predictions occurred in the early 1700s, giving rise to a vast field of knowledge [17]. Proof of this, there have been forecasting competitions in the last 40 years, allowing the development of new techniques for accuracy [18].

Currently, forecasting techniques are a vast number of options, making it impossible to gather all of them in a single work of state-of-the-art theoretical and practical forecasting. The forecasting theory in time series is based on theoretical approaches to future values by identifying past patterns [19]. It is finding two main types of predictions, point estimation (future point values) and interval estimation (based on probability windows). On large scales, we distinguish theoretical models: statistical and econometric; Bayesian forecasting; variable and model selection; combining forecasts; data-driven methods; methods for intermittent demand; or reasoning and mining. Forecasting techniques improve decision-making in practice by reducing the uncertainty of the exogenous factors that determine the value to be forecast. The empirical applications are varied: operations and supply chain management, economics and finance; energy markets and consumption; environmental applications; social good and demographic forecasting; systems and humans; tourism demand forecasting; forecasting for aviation; traffic flow forecasting; car arrival forecasting; election forecasting; sports forecasting; agricultural forecasting, among other fields [4].

In analysing time-series data, comparing models to evaluate the predictive capacity is expected. When a new forecasting technique is proposed, it is common for its performance to be benchmarked according to some measure of forecast accuracy against other methods using a sub-sample of some particular time series [4]. The scientific literature has extensively discussed forecast and accuracy measures. The authors have focused their accuracy analysis on errors [20]. Other authors have focused their interest on analysing the direction of structural changes, concluding that the cost can be significantly ignored [21]. The debate has taken place in the literature with discussions among error measures and forecast direction analysis in decision-making, concluding the usefulness of the study of direction changes as a complement to the traditional criteria [22].

2.1 *Most Common Error Measures and Others in Forecasting*

Hyndman and Koehler [7] studied three examples of series within-sample and out-sample (ex-post) predictions, finding limitations of undefined or infinite solutions. They compared estimation methods for (a) series 472 from M-3 competition; (b) 10 years of monthly log stock returns for the Walt Disney Corporation, 1990–1999; (c) 3 years of monthly sales of a lubricant product sold in large containers. In particular, Mean Absolute Percentage Error (MAPE) forecasting accuracy measures were compared [23]; Geometrical Mean Relative Absolute Error (GMRAE), symmetric MAPE (sMAPE), Median Relative Absolute Error (MdRAE), Median Absolute Percentage Error (MdAPE) and symmetric Median Absolute Percentage Error (sMdAPE) [24, 25]; Mean Absolute Scaled Error (MASE) [7].

A considerable forecasting measure has been used in the applied modelling: MAPE, MdAPE, Root Mean Square Percentage Error (RMSPE), Root Mean Square Percentage Error (RMSPE); measures based on relative errors: Mean Relative Absolute Error (MRAE), Median Relative Absolute Error (MdRAE), GMRAE; relatives measures or scaled errors. Even more recent articles with new measures such as Unscaled Mean Bounded Relative Absolute Error (UMBRAE) [26] or trigonometry measures of forecast called the Mean Arcotangent Absolute Percentage Error (MAAPE) [27]. Researchers seem to prefer unit-free measures to compare forecasting methods [24]. Root Mean Squared Error (RMSE) was the most used criterion in M-competition [23]. However, the researchers have preferred additional measures such as the MAPE, which is relevant only by ratio-scaled data and penalised by extreme values in the predictions [24]. Using the different forecast accuracy measures in the modelling is difficult; the authors reveal difficulties of consistent homogenisation. After extensive reading and investigation, we can conclude that there is no agreement on the best forecast measure [17].

In the following sections of this paper, we will propose a solution to the relative measures among models. This work aims to provide the researchers with a decision tool to improve their analysis in point estimation [28]. The proposal is the development of a working scheme to compare models that meet three principles (reliability, robustness and interpretability) that allow the identification of the “best” model on a set of predictive models. We understand that in practice, forecasting scientific articles try to provide a better model than techniques used in the literature. Our contribution is for new researchers to have a dimensionless measure of comparison of predictive models, and the robustness of this measure will be given by the *Lambda-Relative-Measures Theorem* [29]. We consider our proposal applicable for new estimation methods compared to existing ones and is usable on different time granularities [30].

3 Theoretical Framework

Time series present pitfalls in evaluating forecasts associated with the characteristics of the DGP to be analysed [1]. We will differentiate the training period $t = 1, 2, \dots, T$ and the out-sampling training period (ex-post) $t + h$ with the window of time horizon h . Starting from the definition of the ex-post forecast, we will denote an actual value y_{t+h} on the horizon $t + h$, where $t \in T$ is the last value of the training period of the models that we use to predict on a horizon $h \in H$. Similarly, we can denote the predictive value as \hat{y}_{t+h} this prediction can be made by any existing model in the next univariate forecasting techniques section. Prediction errors are defined by $\varepsilon_{t+h} = \hat{y}_{t+h} - y_{t+h}$; giving the prediction results; the goal is to minimise the expectation of the error $\arg \min E(\varepsilon_{t+h})$. The limitation for the relative accuracy measures appears when $\hat{y}_{t+h} = y_{t+h}$ being obtained $\varepsilon_{t+h} = 0$ for the numerator or denominator for forecasting relative accuracy measures [7].

These models will provide support for the development of the empirical section. A second subsection will be the contribution of this paper with the theoretical development of *Lambda-Relative-Measures*, giving rise to the *Lambda-Relative-Measures Theorem* of accuracy. These theoretical aspects will be verified in the empirical section; in particular, we will work with data applied to the Spanish tourism industry, in which the importance of anticipating future demand is “mandatory” and a significant area of research [4, 31, 32].

Figure 1 represents the steps for choosing the “best” model under reliability, robustness, and interpretability characteristics. The initial data must be used to estimate forecasting and control tasks by the researcher—the models to be used by researchers and the desired window time horizon h selection. The empirical time horizon will be determined by the science in which the methodology is applied. Testing the robustness of the projections will involve knowing whether the models are significantly different from the current predicted values (ex-post). From this step, we can find the problem and redirect the modelling. Finally, we will propose the “best” model with the Lambda decision matrix and the theoretical support of the *Lambda-Relative-Measures Theorem*.

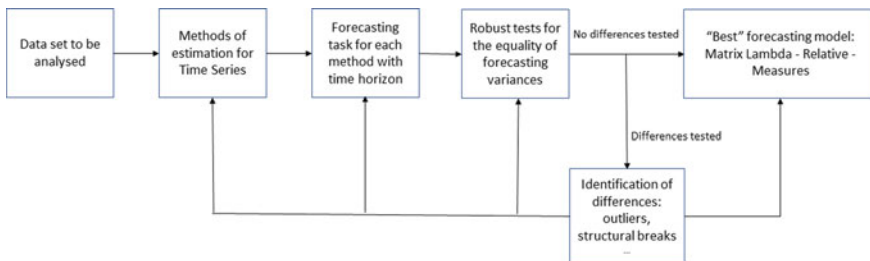


Fig. 1 A scheme to find the time-series model with the least forecast error. Own Elaboration

3.1 Automatic Forecasting and Testing: Univariate Forecasting Techniques

For the estimation of models, we use the R package, and the techniques are as follows SARIMA models [10], HWA and HWM [15], ARAR model [12], SSA [11] and HNN [14]. One of the aspects to be dealt with in evaluating predictions is the “forecast distribution” [17]. Our work scheme (Fig. 1.) tests the equality of variances to find representative differences among forecasting models based on medians. The primary motivation is that predictions can range from short to long terms [33].

Once the prediction models have been proposed for the analysis, it is convenient to compare the variances for the evaluation period horizon h . We apply the Brown-Forsythe test’s robust homogeneity of variance [9].

Suppose we have l_v prediction methods $\forall v = 1, 2, 3 \dots L$ and we evaluate the same time horizon $h (\forall h = 1, 2, 3 \dots H)$. Once the prediction tasks with the l_v techniques, we know the standard deviations of all models $\sigma_1, \sigma_2, \sigma_3, \dots, \sigma_L$. The Brown-Forsythe test calculation formula for the predictions would be defined as follows:

$$W = \frac{(Hl - l) \sum_{k=1}^l h_k (Z_k - \bar{Z})^2}{(l - 1) \left\{ \sum_{k=1}^l \sum_{i=1}^{h_k} (Z_{ki} - \bar{Z}_k)^2 \right\}} \sim F_{(l-1)(Hl-l)} \quad (1)$$

The W^1 statistic is distributed as approximately F-Fisher-Snedecor under the null hypothesis of equality of variances among time series compared [34]. In the case of rejecting the null hypothesis, we must consider whether the differences with the present value are significant or not. Otherwise, the option of not rejecting the null hypothesis should be measured. With the equality of statistical variances certified, it is convenient to quantify the improvement using the different models. The final decision criterion is developed in the following subsection: “Family forecasting Evaluation: Lambda-Relative-Measures”.

Family forecasting Evaluation: Lambda-Relative-Measures. In this methodological section, we will develop the *Lambda-Relative-Measures* of accuracy. We would not encounter problems with exact predictions in measures based on the errors for an individual model $\hat{y}_{t+h} = y_{t+h} \rightarrow \varepsilon_{t+h} = 0$. The problem occurs when our decision on the choice of the model is based on the dimensionless relative ratio among different models for the same forecasting capacity [19]. An example of this could be the Relative Ratio of RMSE (RRMSE), which can be expressed as follows:

¹ $Z_{ki} = |y_{kl_v,t+h} - \text{median}(y_{kl_v,t+h})|$; $\bar{Z} = (1/Hl) \sum_{k=1}^l \sum_{i=1}^{h_k} (Z_{ki})$; $\bar{Z}_k = (1/h_k) \sum_{i=1}^{h_k} (Z_{ki})$.

$$RRMSE_{i,j} = \frac{\sqrt{\sum_{h=1}^H (\hat{y}_{t+h}^i - y_{t+h})^2}}{\sqrt{\sum_{h=1}^H (\hat{y}_{t+h}^j - y_{t+h})^2}} \tag{2}$$

The advantage of using relative methods lies in easy interpretability; i.e., in ex-post forecasting, we will say that model i is better than the model j when $RRMSE_{i,j} < 1$, otherwise, we will say that j is better than the model i . In this case of $RRMSE_{i,j} = 1$, both have the same predictive capacity.

Its relative expression has been proposed in the literature, meaning a change of scale in the measurement [7]. We suggest *Lambda-Relative-Measures* and the decision matrix. The ratio is easy to interpret, based on results around 1. If $\hat{y}_{t+h}^j - y_{t+h} = 0$, we find a problem of mathematical undefinition that invalidates the selection process of predictive models. The relative measure ratio cited is a solution to the undefined problem using RMSE. For our development, we will assume that \mathbb{R}^t the variable is ordered $t = 1, 2, \dots, T$. Thus, the errors ε_{t+h} are distributed with $E[\varepsilon_{t+h}] = 0$ and $\text{var}[\varepsilon_{t+h}] = \sigma^2$. From which it can be deduced that $RRMSE_{i,j} = \sigma_i^2 / \sigma_j^2$ with $\sigma_i^2, \sigma_j^2 \geq 0$. Undefined problems occur when $\sigma_j^2 = 0$, being a plausible and not improbable result.

The proposed solution is the *Lambda-Relative-Measures*, which expression for RMSE is as follows:

$$\lambda - RRMSE_{i,j} = \frac{\lambda + \sqrt{\frac{\sum_{h=1}^H (\hat{y}_{t+h}^i - y_{t+h})^2}{h}}}{\lambda + \sqrt{\frac{\sum_{h=1}^H (\hat{y}_{t+h}^j - y_{t+h})^2}{h}}} \tag{3}$$

Being λ a constant value that we introduce ad-hoc, and that can take any value $\lambda \neq 0$. To demonstrate the validity of this criterion, we must show the limits of the expression:

$$\lim_{RRMSE_j \rightarrow 0} \lambda - RRMSE_{i,j} = \frac{\lambda + \sigma_i^2}{\lambda} \tag{4}$$

Given any value of $\sigma_i^2 \geq 0$ and $\sigma_j^2 = 0$ with $\lambda \neq 0$, we can indicate that we have found a solution to $RMSE_j \rightarrow 0$. Analogously, we can find an expression in the limit of $RMSE_i \rightarrow 0$ being $\lim_{RRMSE_i \rightarrow 0} \lambda - RRMSE_{i,j} = \lambda / (\lambda + \sigma_j^2)$. The λ values will depend on the dimensions of the variables to be analysed. In our empirical work, we use a $\lambda = 1$, and the main idea is not to distort the interpretability of the ratio, this criterion being the basis for solving the problem of mathematical indefiniteness.

A decision is based on *Matrix* $\lambda - RRMSE_{i,j}$ establishing a better model with a more significant predictive capacity for the DGP. This matrix is developed with robust laws avoiding undefined results. The idea of replicating different automatic models to the same data framework and obtaining a prediction result for each window time horizon h [35]. For this, we define the decision tool called *Matrix* $\lambda - RRMSE_{i,j}$:

$$Matrix \lambda - RRMSE_{i,j} = \begin{pmatrix} \lambda - RRMSE_{1,1} & \lambda - RRMSE_{1,2} & \cdots & \lambda - RRMSE_{1,j} \\ \lambda - RRMSE_{2,1} & \lambda - RRMSE_{2,2} & \cdots & \lambda - RRMSE_{2,j} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda - RRMSE_{i,1} & \lambda - RRMSE_{i,2} & \cdots & \lambda - RRMSE_{i,j} \end{pmatrix} \quad (5)$$

The robustness of this matrix is determined by the fulfilment of three axioms: completeness, transitivity and rationality. Completeness implies that a model can be rated on a scale of good, bad or the same as that model itself. Transitivity allows us to organise the models according to the criterion of least $\lambda - RRMSE_{i,j}$. Based on the above, rationality is fulfilled when the best ex-post predictive capacity model presents the lowest $\lambda - RRMSE_{i,j}$. The matrix results are easy to interpret; we will say that model i is better than the model j when $\lambda - RRMSE_{i,j} < 1$, otherwise, we will say that j is better than the model i . In this case of $\lambda - RRMSE_{i,j} = 1$, both have the same predictive capacity.

The theoretical results could be summarised in the following *Lambda Theorem of Accuracy*:

For any pair of time series predicted using forecast models ($\hat{y}_{t+h}^i, \hat{y}_{t+h}^j$) on a variable under study out-sample of the training period y_{t+h} . If $\exists \text{ Lambda} - \text{Relative} - \text{Measures} \neq 0$. Then, an ordering allows quantifying the closeness or distance to the actual value y_{t+h} .

In the following empirical section, we will develop the conceptual scheme of Fig. 1. For a data set, we will make the predictions with six models, and finally, we will be able to quantify which model is better in terms of accuracy avoiding inconsistency decisions.

4 An Empirical Application to Evaluate Automatic Forecasting Modelling of Tourism Data

The empirical application of the methodological framework is carried out with official data from the Spanish National Institute of Statistics (INE in Spanish). The modelling period runs from January 1999 to December 2019. The variable to study is the number of foreign tourists' hotel overnight stays in Spain. The results have been tested in 52 Spanish provinces, but we will show the aggregate "Spain" time series. We will take January 1999 to December 2018 as the training period of the models with an ex-post forecasting training period for the year 2019. For reasons of avoiding repetitive

calculations and interpretations, in this work, we will present the results of the Spain time series for the horizons $h = 3, 6, 12$. For the rest of the 52 Spanish provinces, we will present the decision matrix for $h = 3$ (see appendix); it is a similar application and interpretation for $h = 6, 12$. In appendix 3, we offer the comparative decision matrix with the SARIMA model; however, any researcher could select the desired comparative model.

Figure 2 shows the application of ex-post estimation and forecasting methods for the data mentioned (Spain) in the previous section. The results observed in the graph make it impossible to decide at a glance which model presents the minimum forecasting error with a seasonal component observed throughout the data training period. This fact makes it sense to continue with the proposed work scheme to contrast significant differences in variances and subsequently classify the models using the *Matrix* $\lambda - RRMSE_{i,j}$. This way, we improve decision-making based on objective data and robust statistical processes.

In Table 1, the robust variances test based on the forecasting methods' medians indicates no significant differences among the processes. The p-values in Table 1 verify the non-existence of differences between prediction methods with windows $h = 3, 6, 12$ for the Spain time series.

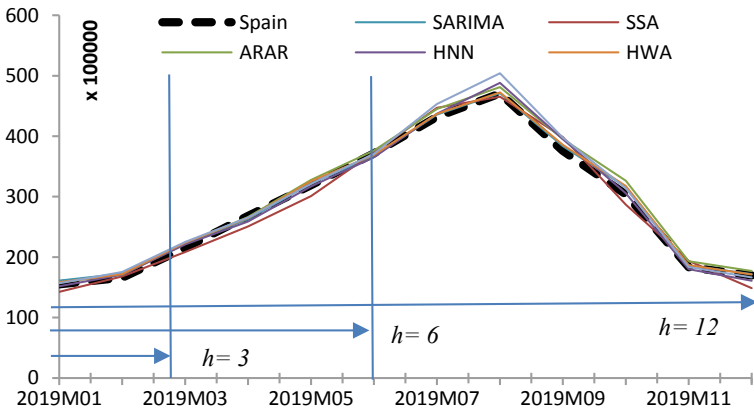


Fig. 2 Out-sample forecast for the aggregate series for foreign tourists visiting Spain $h = 3, 6, 12$ (Jan. 2019 to Dec. 2019). Own Elaboration

Table 1 Brown-Forsythe test results. Horizons (January 2019 to December 2019). Null hypothesis: the variances of the forecasting methods are the same for the time horizons h —own elaboration

Brown-Forsythe test			
Horizon	df	Value	p-value
$h = 3$	(6, 14)	0.0057	1.00
$h = 6$	(6, 35)	0.0108	1.00
$h = 12$	(6, 77)	0.0221	1.00

The non-existence of variance differences has been empirically demonstrated in the Spanish case. The application of differences of variances Brown-Forsythe test for the 52 provinces (outside of this work) shows no significant differences with a high confidence level (p -values = 1). We are interested in mentioning other time horizons for the rest of the 52 provinces ($h = 6, 12$); this test shows only statistically significant differences among prediction methods for the cases of the Balearic Islands ($h = 12$; HWM), Palencia ($h = 6, 12$; ARAR) and Melilla ($h = 12$; ARAR) provinces. According to these results, we can deduce there are no differences among the variances of the estimation methods for a total of 159 predictions with the horizon windows $h = 3, 6, 12$.

The last step is the calculation of the “best” model. The results of the models used in the literature are very limited to the data set and the homogeneity [17]. The results of table 2 verify the validity of each model for the time horizon for the time series analysed for the Spanish case according to the methodological section. For $h = 3$, the ranking is (1) ARAR, (2) HWA, (3) HNN, (4) SARIMA, (5) HWM and (6) SSA; for $h = 6$, the order is (1) ARAR, (2) SARIMA, (3) HWM, (4) HNN, (5) HWA and (6) SSA; finally the best models for a time horizon of 12 months are (1) SARIMA, (2) HWA, (3) HNN, (4) ARAR, (5) HWM and (6) SSA (Table 2).

Once the entire evaluation process of forecasting models with tourist accommodation data for Spain has been empirically described, we proceed to the conclusions, discussions on the procedure, future research lines of investigation and recommendations.

5 Conclusions and Discussions

This work aims to offer a forecast model selection strategy, aware of the limitations and uncertainty. The most significant advantage is identifying mathematical patterns and avoiding excess optimism or pessimism with critical approaches based on scientific literature. Not forgetting that the forecasts’ accuracy and the uncertainties’ correction will depend on the patterns in the DGP period analysed [4]. Using benchmarks to evaluate models meaningfully quantifies the model’s advantages compared to similar tools.

We have proposed a granularity working scheme for selecting models in time horizon windows ($h = 3, 6, 12$). Our optimal model will minimise the ratio $\lambda - RRMSE_{i,j}$, previous the analysis described under the principles of reliability, robustness and interpretability. These three principles represent a theoretical and empirical contribution of precision comparison among methods or hypothesis validation. The criteria developed in this work contribute to the traditional problem of selecting the “best” model [36], and its granularity allows quantifying the comparative analysis of predictive models [30]. In this work with *Lambda-Relative-Measures*, we can measure the “profits” and the costs of using different models [17].

Table 2 Matrix $\lambda - RRMSE_{i,j}$ for Spanish forecasting methods. January 2019 to December 2019). Rankings in brackets. Own elaboration

Horizon	For. methods	Forecasting methods					
		SARIMA (4)	SSA (6)	ARAR (1)	HNN (3)	HWA (2)	HWM (5)
h = 3	SARIMA (4)	1.00	0.57	1.39	1.03	1.12	0.92
	SSA (6)	1.75	1.00	2.42	1.81	1.95	1.61
	ARAR (1)	0.72	0.41	1.00	0.75	0.80	0.67
	HNN (3)	0.97	0.55	1.34	1.00	1.08	0.89
	HWA (2)	0.90	0.51	1.24	0.93	1.00	0.89
	HWM (5)	1.08	0.62	1.50	1.12	1.21	1.00
		SARIMA (2)	SSA (6)	ARAR (1)	HNN (4)	HWA (5)	HWM (3)
h = 6	SARIMA (2)	1.00	0.49	1.05	0.99	0.74	1.00
	SSA (6)	2.05	1.00	2.15	2.04	1.52	2.05
	ARAR (1)	0.95	0.46	1.00	0.95	0.70	0.95
	HNN (4)	1.01	0.49	1.06	1.00	0.74	1.01
	HWA (5)	1.35	0.66	1.42	1.35	1.00	1.35
	HWM (3)	1.00	0.49	1.05	0.99	0.74	1.00
		SARIMA (1)	SSA (6)	ARAR (4)	HNN (3)	HWA (2)	HWM (5)
h = 12	SARIMA (1)	1.00	0.42	0.54	0.66	0.93	0.43
	SSA (6)	2.36	1.00	1.27	1.56	2.18	1.02
	ARAR (4)	1.85	0.79	1.00	1.23	1.71	0.80
	HNN (3)	1.51	0.64	0.81	1.00	1.40	0.65
	HWA (2)	1.08	0.46	0.58	0.72	1.00	0.47
	HWM (5)	2.31	0.98	1.25	1.53	2.14	1.00

From a theoretical point of view, studying differences in medians for windows temporal horizons implies knowing the significant differences among predictive models. The analysis of the difference of medians overcomes the theoretical limitations on the distribution of the predictions since each predicted observation is unique and unrepeatable [16]. In cases where there are no significant differences, the *Lambda-Relative-Measures* represent an efficient decision-making tool and a statistical measure with the optimal accuracy option [36].

The debate is quite broad in the scientific literature, and we currently consider that there is no standard criterion concerning the “best” predictive model. However, our contribution to a robust model selection criterion must be evaluated on several practical problems: (a) The prediction of seasonal time series is questioned for several reasons. There are almost as many model applications as there are series to study in practice. Seasonality depends on the temporal frequency and the repetitive cycles of each series. So, it is complicated to determine previously “the best” predictive model;

(b) according to De Gooijer and Hyndman in 2006, forecasting analysis for time series with seasonal patterns presents robustness problems. The conclusions depend on the results carried out. Therefore, the researchers consider no consensus on which model is preferred; (c) ex-post forecasting allows quantifying the differences with actual values. Depending on the type of model, the weighting of the last observations. For this, model predictions are subject to possible structural changes not previously modelled [8, 37]. Examples of them can be economic crises in economics or health crises [38].

The robust criteria in decision-making overcome many limitations of the literature. This criterion opposes previous criticisms of using error measures depending on the situation since it reliably compares the relative differences with actual values [24]. In the empirical section, it has been observed that there is not always the best predictive model for the 53 series (Spain and 52 provinces) and different windows of time horizons. The *Lambda theorem of accuracy* allows us to quantify and decide the “best” model in the analysed window horizon. In this work, we have used some forecasting techniques, which are many more in the literature [20, 42]. Furthermore, in the empirical section, data have been used for tourism forecasting and are widely studied in the literature [40–44].

The implications of this study are varied, and the suggested future research lines: (1) We encourage demonstrating the *Lambda theorem of accuracy* with other measures of accuracy different from RMSE [7]; (2) We have found results for the tourism industry, and we suggest it replicate the methodology to other disciplines of science [4]; (3) New alternative robustness tests can be developed. We encourage practitioners and researchers to reflect deeply on the limitations of forecasting in order to improve theoretical and applied sciences and ultimately improve the knowledge society.

Data Availability Statement <https://www.ine.es/jaxiT3/Tabla.htm?t=2074>.

Appendix

See Table 3.

Table 3 $\lambda - RRMSE_{SARIMA,j}$ for all Spanish provinces, including the Spanish series ($h = 3$, January 2019 to March 2019). Own elaboration

$h = 3$	SARIMA versus SARIMA	SARIMA versus SSA	SARIMA versus ARAR	SARIMA versus HNN	SARIMA versus HWA	SARIMA versus HWM
Spain	1,0000	0,5716	1,3857	1,0349	1,1151	0,9232
Alava	1,0000	0,9994	0,9999	1,0599	1,0001	1,0000
Albacete	1,0000	0,9997	0,9999	1,0000	1,0050	1,0049
Alicante	1,0000	0,9981	0,9987	0,9986	0,9996	1,0009
Almería	1,0000	0,9968	1,0003	1,0001	1,0003	1,0003
Avila	1,0000	0,9996	0,9999	1,0000	1,0001	1,0001
Badajoz	1,0000	0,9996	0,9998	1,0000	1,0002	1,0003
Baleares	1,0000	0,9995	0,9977	0,9981	0,9987	0,9861
Barcelona	1,0000	1,0005	0,9998	0,9947	0,9982	0,9972
Burgos	1,0000	0,9995	1,0002	0,9999	1,0001	1,0001
Caceres	1,0000	0,9997	1,0001	1,0001	1,0001	1,0001
Cádiz	1,0000	0,9958	0,9999	0,9997	0,9997	0,9998
Castellon	1,0000	0,9985	1,0001	0,9997	0,9997	0,9991
Ciudad_Real	1,0000	1,0002	1,0000	1,0000	1,0001	1,0001
Cordoba	1,0000	0,9999	0,9999	1,0000	1,0001	0,9999
A_Coruña	1,0000	0,9985	1,0000	0,9995	0,9998	1,0001
Cuenca	1,0000	0,9999	0,9999	0,9999	1,0000	0,9999
Gerona	1,0000	0,9971	1,0000	1,0004	1,0008	1,0017
Granada	1,0000	0,9962	0,9997	1,0003	1,0007	1,0016
Guadalajara	1,0000	0,9999	1,0000	1,0000	1,0000	1,0000
Gipuzcoa	1,0000	0,9991	0,9998	0,9997	0,9998	0,9996
Huelva	1,0000	0,9975	0,9997	0,9999	1,0000	0,9970
Huesca	1,0000	0,9990	1,0005	0,9999	0,9999	0,9998
Jaen	1,0000	0,9995	1,0000	1,0001	1,0001	1,0002
Leon	1,0000	1,0003	1,0004	1,0003	1,0002	1,0002
Lleida	1,0000	0,9991	0,9999	1,0001	1,0000	1,0002
La_Rioja	1,0000	0,9998	1,0001	1,0000	1,0000	1,0001
Lugo	1,0000	0,9993	0,9997	0,9995	0,9996	0,9996
Madrid	1,0000	0,9938	0,9990	0,9979	0,9997	0,9987
Malaga	1,0000	0,9962	0,9996	0,9976	0,9985	0,9985
Murcia	1,0000	0,9973	1,0000	0,9998	1,0001	1,0000
Navarra	1,0000	0,9992	1,0000	0,9999	1,0000	0,9999
Ourense	1,0000	0,9999	1,0000	1,0000	1,0000	1,0000
Asturias	1,0000	0,9983	0,9997	0,9998	1,0002	1,0001
Palencia	1,0000	0,9999	1,0081	1,0000	1,0000	1,0000

(continued)

Table 3 (continued)

$h = 3$	SARIMA versus SARIMA	SARIMA versus SSA	SARIMA versus ARAR	SARIMA versus HNN	SARIMA versus HWA	SARIMA versus HWM
Las_Palmas	1,0000	1,0336	1,0033	1,0054	1,0044	1,0007
Pontevedra	1,0000	0,9983	1,0001	1,0000	1,0009	1,0012
Salamanca	1,0000	0,9993	0,9998	1,0000	1,0002	1,0001
S.C. Tenerife	1,0000	1,0053	1,0006	1,0009	0,9996	1,0031
Cantabria	1,0000	0,9981	0,9997	0,9998	1,0000	0,9999
Segovia	1,0000	0,9995	0,9999	1,0000	1,0001	0,9999
Sevilla	1,0000	0,9999	0,9994	0,9993	1,0000	0,9987
Soria	1,0000	0,9998	1,0000	1,0000	1,0000	1,0000
Tarragona	1,0000	0,9987	0,9996	0,9993	0,9983	0,9994
Teruel	1,0000	0,9995	0,9999	1,0000	1,0001	1,0001
Toledo	1,0000	0,9999	1,0001	1,0000	1,0001	1,0000
Valencia	1,0000	0,9948	1,0004	0,9995	0,9983	0,9997
Valladolid	1,0000	0,9998	0,9998	1,0000	1,0001	1,0001
Bizkaia	1,0000	0,9990	1,0001	0,9995	1,0001	0,9997
Zamora	1,0000	0,9996	0,9998	1,0000	1,0001	1,0000
Zaragoza	1,0000	0,9996	1,0008	1,0003	1,0007	1,0017
Ceuta	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Melilla	1,0000	1,0001	1,0739	2,0457	1,9483	2,0195

References

- Hewamalage, H., Ackermann, K., Bergmeir, C.: Forecast Evaluation for Data Scientists: Common Pitfalls and Best Practices (2022). <https://doi.org/10.48550/10.1101/2022.07.16.2203.10716>
- Ruiz-Reina, M.A.: An approximation to relative automatic time series forecasting measures. In: ITISE 2022 International Conference on Time Series and Forecasting, pp. 42–42 (2022)
- de Gooijer, J.G., Hyndman, R.J.: 25 Years of IIF Time Series Forecasting: A Selective Review. SSRN Electron. J. (2011). <https://doi.org/10.2139/ssrn.748904>
- Petropoulos, F., et al.: Forecasting: theory and practice. *Int. J. Forecast.* **38**(3), 705–871 (2022). <https://doi.org/10.1016/j.ijforecast.2021.11.001>
- Bontempi, G., ben Taieb, S., le Borgne, Y.E.: Machine learning strategies for time series forecasting (2013). https://doi.org/10.1007/978-3-642-36318-4_3
- Serengil, S.I., Ozpinar, A.: Workforce Optimization for Bank Operation Centers: A Machine Learning Approach. *International Journal of Interactive Multimedia and Artificial Intelligence* **4**(6), 81–87 (2017). <https://doi.org/10.9781/ijimai.2017.07.002>
- Hyndman, R.J., Koehler, A.B.: Another look at measures of forecast accuracy. *Int. J. Forecast.* **22**(4), 679–688 (2006). <https://doi.org/10.1016/j.ijforecast.2006.03.001>
- Inoue, A., Jin, L., Rossi, B.: Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics* **196**, 55–67 (2017). <https://doi.org/10.1016/j.jeconom.2016.03.006>
- Brown, M.B., Forsythe, A.B.: Robust tests for the equality of variances. *J Am Stat Assoc.* pp. 364–367 (1974). <https://doi.org/10.1080/01621459.1974.10482955>

10. Hyndman, R.J., Khandakar, Y.: Automatic time series forecasting: The forecast package for R. *J. Statistical Software* **27**(3) (2008). <https://doi.org/10.18637/jss.v027.i03>
11. Golyandina, N., Korobeynikov, A., Zhigljavsky, A.: *Singular Spectrum Analysis with R* (2018)
12. Brockwell, P.J., Davis, R.A.: *Introduction to Time Series and Forecasting - Second Edition* (2002)
13. Weigt, G.: *ITSM-R Reference Manual* (2018). <https://georgeweigt.github.io/itsmr-refman.pdf>
14. Athanasopoulos, G., Hyndman, R.J., Kourentzes, N., Petropoulos, F.: Forecasting with temporal hierarchies. *Eur. J. Oper. Res.* **262**(1), 60–74 (2017). <https://doi.org/10.1016/j.ejor.2017.02.046>
15. Hyndman, R.J., Athanasopoulos, G.: *Forecasting: Principles and Practice*. In: *Principles of Optimal Design* (2018)
16. Kolassa, S.: Why the ‘best’ point forecast depends on the error or accuracy measure. *Int. J. Forecast.* **36**, 208–211 (2020). <https://doi.org/10.1016/j.ijforecast.2019.02.017>
17. Makridakis, S., Hyndman, R.J., Petropoulos, F.: Forecasting in social settings: The state of the art. *Int. J. Forecast.* **36**(1), 15–28 (2020). <https://doi.org/10.1016/j.ijforecast.2019.05.011>
18. Hyndman, R.J.: A brief history of forecasting competitions. *Int. J. Forecast.* **36**, 7–14 (2020). <https://doi.org/10.1016/j.ijforecast.2019.03.015>
19. Reina, M.A.R.: Big Data: Forecasting and Control for Tourism Demand. In: *Theory and Applications of Time Series Analysis*. ITISE 2019, R. I. Valenzuela O., Rojas F., Herrera L.J., Pomares H., Ed. Springer, Cham, pp. 273–286 (2020). https://doi.org/10.1007/978-3-030-56219-9_18
20. Witt, C.A., Witt, S.F.: Measures of forecasting accuracy - turning point error v size of error. *Tour. Manage.* **10**(3), 255–260 (1989). [https://doi.org/10.1016/0261-5177\(89\)90087-3](https://doi.org/10.1016/0261-5177(89)90087-3)
21. Pesaran, M.H., Timmermann, A.: How costly is it to ignore breaks when forecasting the direction of a time series? *Int. J. Forecast.* **20**, 411–425 (2004). [https://doi.org/10.1016/S0169-2070\(03\)00068-2](https://doi.org/10.1016/S0169-2070(03)00068-2)
22. Blaskowitz, O., Herwartz, H.: On economic evaluation of directional forecasts. *Int. J. Forecast.* **27**, 1058–1065 (2011). <https://doi.org/10.1016/j.ijforecast.2010.07.002>
23. Makridakis, S., et al.: The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *J. Forecast.* **1**(2), 111–153 (1982). <https://doi.org/10.1002/for.3980010202>
24. Armstrong, J.S., Collopy, F.: Error measures for generalising about forecasting methods: Empirical comparisons. *Int. J. Forecast.* **08**, 69–80 (1992). [https://doi.org/10.1016/0169-2070\(92\)90008-W](https://doi.org/10.1016/0169-2070(92)90008-W)
25. Makridakis, S., Hibon, M.: The M3-competition: Results, conclusions and implications. *Int. J. Forecast.* **16**(4), 451–476 (2000). [https://doi.org/10.1016/S0169-2070\(00\)00057-1](https://doi.org/10.1016/S0169-2070(00)00057-1)
26. Chen, C., Twycross, J., Garibaldi, J.M.: A new accuracy measure based on bounded relative error for time series forecasting. *PLoS ONE* (2017). <https://doi.org/10.1371/journal.pone.0174202>
27. Kim, S., Kim, H.: A new metric of absolute percentage error for intermittent demand forecasts. *Int. J. Forecast.* **32**, 669–679 (2016). <https://doi.org/10.1016/j.ijforecast.2015.12.003>
28. Lehmann, E.L., Casella, G.: *Theory of Point Estimation*. Second Edition Springer Texts in Statistics. (1998). <https://doi.org/10.2307/1270597>
29. Chen, C.: Robustness properties of some forecasting methods for seasonal time series: A Monte Carlo study. *Int. J. Forecast.* **13**, 269–280 (1997). [https://doi.org/10.1016/S0169-2070\(97\)00014-9](https://doi.org/10.1016/S0169-2070(97)00014-9)
30. Dong, R., Pedrycz, W.: A granular time series approach to long-term forecasting and trend forecasting. *Physica A* **387**, 3253–3270 (2008). <https://doi.org/10.1016/j.physa.2008.01.095>
31. García-Ferrer, A., Queralt, R.A.: A note on forecasting international tourism demand in Spain. *Int. J. Forecast.* **13**(4), 539–549 (1997). [https://doi.org/10.1016/S0169-2070\(97\)00033-2](https://doi.org/10.1016/S0169-2070(97)00033-2)
32. Song, H., Hyndman, R.J.: Tourism forecasting: An introduction. *Int. J. Forecast.* **27**(3), 817–821 (2011). <https://doi.org/10.1016/j.ijforecast.2011.03.001>
33. Andrawis, R.R., Atiyya, A.F., El-Shishiny, H.: Combination of long term and short term forecasts, with application to tourism demand forecasting. *Int. J. Forecast.* **27**, 870–886 (2011). <https://doi.org/10.1016/j.ijforecast.2010.05.019>

34. Dag, O., Dolgun, A., Konar, N.M.: Onewaytests: An R package for one-way tests in independent groups designs. *R Journal* **10**(1), 175–199 (2018). <https://doi.org/10.32614/rj-2018-022>
35. Athanassopoulos, G., Hyndman, R.J., Song, H., Wu, D.C.: The tourism forecasting competition. *Int. J. Forecast.* **27**, 822–844 (2011). <https://doi.org/10.1016/j.ijforecast.2010.04.009>
36. Tofallis, C.: A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society* **66**(8), 1352–1362 (2015). <https://doi.org/10.1057/jors.2014.103>
37. Wunderlich, F., Memmert, D.: Are betting returns a useful measure of accuracy in (sports) forecasting? *Int. J. Forecast.* **36**(2), 713–722 (2020). <https://doi.org/10.1016/j.ijforecast.2019.08.009>
38. Rossi, B., Inoue, A.: Out-of-sample forecast tests robust to the choice of window size. *Journal of Business and Economic Statistics* **30**(3), 432–453 (2012). <https://doi.org/10.1080/07350015.2012.693850>
39. Hamilton, J.D.: What’s real about the business cycle? *Federal Reserve Bank of St. Louis Review*. pp. 435–452 (2005). <https://doi.org/10.20955/r.87.435-452>.
40. Makridakis, S., Spiliotis, E., Assimakopoulos, V.: The M4 Competition: 100,000 time series and 61 forecasting methods. *Int. J. Forecast.* **36**, 54–74 (2020). <https://doi.org/10.1016/j.ijforecast.2019.04.014>
41. Li, G., Song, H., Witt, S.F.: Recent developments in econometric modeling and forecasting. *J. Travel Res.* (2005). <https://doi.org/10.1177/0047287505276594>
42. Song, H., Li, G.: Tourism demand modelling and forecasting-A review of recent research. *Tour. Manage.* **29**(2), 203–220 (2008). <https://doi.org/10.1016/j.tourman.2007.07.016>
43. Peng, B., Song, H., Crouch, G.I.: A meta-analysis of international tourism demand forecasting and implications for practice. *Tour. Manage.* (2014). <https://doi.org/10.1016/j.tourman.2014.04.005>
44. Di Lascio, F.M.L., Durante, F., Pappadà, R.: Copula-based clustering methods (2017)
45. Jiao, E.X., Chen, J.L.: Tourism forecasting: A review of methodological developments over the last decade. *Tour. Econ.* (2019). <https://doi.org/10.1177/1354816618812588>

Recurrent Neural Networks for Forecasting Time Series with Multiple Seasonality: A Comparative Study



Grzegorz Dudek , Sławek Smył , and Paweł Pełka 

Abstract This paper compares recurrent neural networks (RNNs) with different types of gated cells for forecasting time series with multiple seasonality. The cells we compare include classical long short-term memory (LSTM), gated recurrent unit (GRU), modified LSTM with dilation, and two new cells we proposed recently, which are equipped with dilation and attention mechanisms. To model the temporal dependencies of different scales, our RNN architecture has multiple dilated recurrent layers stacked with hierarchical dilations. The proposed RNN produces both point forecasts and predictive intervals (PIs) for them. An empirical study concerning short-term electrical load forecasting for 35 European countries confirmed that the new gated cells with dilation and attention performed best.

Keywords LSTM · Multiple seasonality · RNN · Short-term load forecasting · Time series forecasting

1 Introduction

Forecasting time series (TS) with multiple seasonality is a challenging problem. To solve it, a forecasting model has to deal with short- and long-term dynamics as well as a trend and variable variance. Classical statistical methods such as autoregressive moving average (ARMA) and exponential smoothing methods can be extended to multiple seasonal cycles [1, 2] but they suffer from many drawbacks. The most

G. D and P. P. were partially supported by grant 020/RID/2018/19 from the Polish Minister of Science and Higher Education titled “Regional Initiative of Excellence”, 2019–22.

G. Dudek (✉) · P. Pełka

Electrical Engineering Faculty, Czestochowa University of Technology, Czestochowa, Poland

e-mail: grzegorz.dudek@pcz.pl

P. Pełka

e-mail: pawel.pelka@pcz.pl

S. Smył

Meta, 1 Hacker Way, Menlo Park, CA 94025, USA

important of these are: their linear nature, limited adaptability, limited ability to model complex seasonal patterns, problems with capturing long-term dependencies and problems with introducing exogenous variables.

To improve the ability of statistical models to capture multiple seasonality, various approaches have been applied, such as extending the model with Fourier terms [3, 4], TS decomposition [5] and local modeling [6, 7]. Machine learning (ML) gives additional opportunities to the models and makes them more flexible. The main idea behind ML is to learn from past observations any inherent structures, patterns or anomalies within the data, with the objective of generating future values for the series. The most popular ML models in the field of forecasting are neural networks (NNs) [8] as they can flexibly model complex nonlinear relationships with minimum a-priori assumptions and reflect process variability in uncertain dynamic environments. They offer learning of representation, cross-learning on massive datasets and modeling temporary relationships in sequential data. In particular, RNNs, which were designed for sequential data such as TS and text data, are extremely useful for forecasting. They form a directed graph along a temporal sequence which is able to exhibit temporal dynamic behavior using their internal state (memory) to process sequences of inputs.

Modern RNNs, such as LSTM and GRU, are capable of learning both short and long-term dependencies in TS [9]. They are equipped with recurrent cells that can maintain their states over time and, using nonlinear “regulators” called gates, can control the flow of information inside the cell. Recent works have reported that gated RNNs provide high accuracy in forecasting and outperform most of the statistical and ML methods, such as ARIMA (integrated ARMA), support vector machine, and shallow NNs [10]. A comparison of RNNs on multiple seasonality forecasting problems performed in Bianchi et al. [11] showed that LSTM, GRU and classical Elman RNN demonstrate comparable performance but are relatively slow in terms of training time due to the time-consuming backpropagation through the time procedure. To improve the learning capability and forecasting performance facing RNN, different mechanisms have been used such as residual connections [12] and dilated architecture [13], which solves the major challenges of RNN when learning on long sequences: i.e., complex dependencies, vanishing and exploding gradients, and efficient parallelization. Hybrid solutions have also been proposed combining RNN with TS decomposition [14] or other methods such as exponential smoothing. One such hybrid model won the reputed M4 forecasting competition in 2018, showing impressive performance [15].

Motivated by the superior performance of RNN in TS forecasting, in this study, we compare RNNs with different recurrent cells. We consider a problem of univariate forecasting TS with multiple seasonality on the example of short-term electrical load forecasting (STLF). We propose a stacked hierarchical RNN architecture trained globally across all series and equipped with recurrent cells of different types. We normalize TS input data and encode output data using coding variables determined from recent history. This is to better capture the current dynamics of the process. Such preprocessing has proven successful in other forecasting models for multiple seasonality, see our papers [6, 16, 17].

The contribution of this study is as follows:

1. We propose a new RNN architecture for forecasting TS with multiple seasonality. It is composed of three dilated recurrent layers stacked with hierarchical dilations to deal with multiple seasonality. It uses a combined asymmetrical loss function which enables the model to produce both point forecasts and PIs and also to reduce the forecast bias.
2. We compare five types of gated recurrent cells: classical LSTM and GRU, modified LSTM with dilation, and two new cells we proposed recently, which are equipped with dilation and attention mechanisms.
3. We empirically demonstrate on real data for the electricity demand for 35 European countries that our proposed model copes successfully with complex seasonality. The new attentive dilated recurrent cell significantly outperforms its competitors in terms of accuracy.

The remainder of the paper is organized as follows. Section 2 describes the forecasting problem and data representation. Section 3 presents the recurrent cells and Sect. 4 describes RNN architecture. Section 5 describes the results of experiments and discusses our findings. Finally, Sect. 6 concludes the paper.

2 Forecasting Problem and Data Representation

In this study, as an example of forecasting time series with multiple seasonality, we consider a problem of STLF. The hourly load time series, $\{z_\tau\}_{\tau=1}^M$, express triple seasonality: yearly, weekly and daily (see Smyl et al. [18] for details, where such time series are analyzed). Our goal is to forecast the daily profile (24 h) for the next day based on historical loads (univariate problem).

As input information, we introduce a weekly profile, which precedes the forecasted day. This profile is represented by the input pattern defined as follows:

$$\mathbf{x}_t = \frac{\mathbf{z}_t^w - \bar{z}_t^w}{\text{std}(z_t^w)} \quad (1)$$

where $\mathbf{x}_t \in \mathbb{R}^{168}$ is the t -th weekly pattern, $\mathbf{z}_t^w \in \mathbb{R}^{168}$ is the original sequence of the t -th week, and \bar{z}_t^w and $\text{std}(z_t^w)$ are its mean and standard deviation, respectively.

Note that Eq. (1) expresses standardization of the weekly sequence. Thus the weekly sequences for $t = 1, \dots, N$ are unified, i.e., they are centered around zero with a unit variance. This operation filters out the trend and yearly seasonality.

An output pattern represents a forecasted daily sequence as follows:

$$\mathbf{y}_t = \frac{\mathbf{z}_t^d - \bar{z}_t^d}{\text{std}(z_t^d)} \quad (2)$$

where $\mathbf{y}_t \in \mathbb{R}^{24}$ is the t -th daily pattern and $\mathbf{z}_t^d \in \mathbb{R}^{24}$ is the forecasted sequence (following directly weekly sequence \mathbf{z}_t^w which is encoded in \mathbf{x}_t).

Note that in Eq. (2), we encode the daily sequence using the mean and standard deviation of the preceding week. This enables us to decode the forecasted pattern, $\hat{\mathbf{y}}$, into the real sequence as follows:

$$\hat{\mathbf{z}}^d = \hat{\mathbf{y}}\text{std}(z^w) + \bar{z}^w \quad (3)$$

where \bar{z}^w and $\text{std}(z^w)$ are coding variables determined on the basis of the historical weekly sequence represented by query pattern \mathbf{x} .

Following Smyl et al. [18], to introduce more input information related to the forecasted sequence, we extend the input vector with the following components: $\log_{10}(\bar{z}_t^w)$, which informs about the level of the time series, $\mathbf{d}_t^w \in \{0, 1\}^7$, $\mathbf{d}_t^m \in \{0, 1\}^{31}$ and $\mathbf{d}_t^y \in \{0, 1\}^{52}$, which are binary one-hot vectors encoding day of the week, day of the month and week of the year for the forecasted day. The extended input pattern takes the form:

$$\mathbf{x}'_t = [\mathbf{x}_t, \log_{10}(\bar{z}_t), \mathbf{d}_t^w, \mathbf{d}_t^m, \mathbf{d}_t^y] \quad (4)$$

The paired input and output patterns constitute the training set, $\{(\mathbf{x}'_i, \mathbf{y}_i)\}_{i=1}^N$. The proposed model is trained in cross-learning mode, i.e., on many time series [15], which enables it to capture the shared features of the individual series and prevents over-fitting. The training sets for all L time series are combined: $\Phi = \Phi_1 \cup \dots \cup \Phi_L$.

3 Recurrent Cells

In our study, we explore RNNs with different gated recurrent cells. They include classical cells such as LSTM and GRU, modified LSTM, i.e., dilated LSTM (dLSTM), and two new solutions proposed recently, dRNNCell and adRNNCell.

3.1 LSTM

LSTM was proposed in Hochreiter and Schmidhuber [19] for learning problems related to sequential data. The main idea behind LSTM is a memory cell that carries relevant information throughout the processing of the sequence, and nonlinear gating units that regulate the information flow in the cell. Due to the memory, long-term temporal relationships can be captured and the effects of short-term memory can be reduced, i.e., even information from the earlier time steps can make its way to later time steps. Moreover, in LSTM, unlike in simple RNNs, the optimization problem with vanishing gradients was reduced, which improved learning capabilities.

Figure 1 shows a diagram of LSTM. LSTM uses two states: a cell state, \mathbf{c}_t , and a hidden state, \mathbf{h}_t . The states contain information learned from the previous time steps. At each time step t , information is added to or removed from the cell state. These

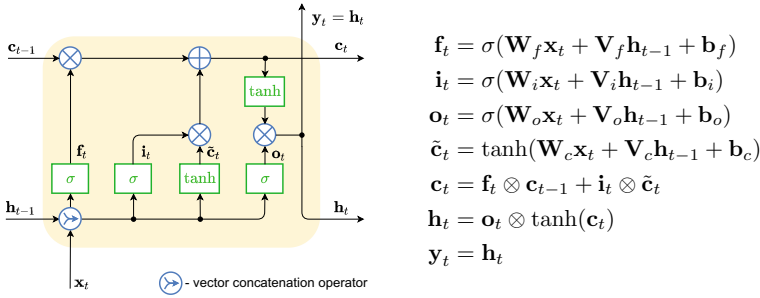


Fig. 1 LSTM

updates are controlled using three gates, which in fact are layers of learned nonlinear transformations. They comprise input gate (*i*), forget gate (*f*) and output gate (*o*). All of the gates receive the hidden state of the past cycle and the current time series sequence as inputs. They can learn what information is relevant to keep or forget during training. At time step *t*, the cell uses the recent states, \mathbf{c}_{t-1} and \mathbf{h}_{t-1} , and the input sequence, \mathbf{x}_t , to compute new updated states \mathbf{c}_t and \mathbf{h}_t . The hidden and cell states are recurrently connected back to the cell input. The new hidden state, \mathbf{h}_t , has two functions. It controls the gating mechanism in the next step and it is treated as the cell output, \mathbf{y}_t , which goes to the next layer.

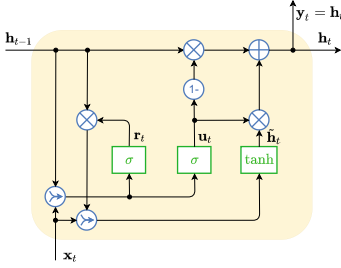
The compact form of the equations describing LSTM is shown in Figure 1, where: \mathbf{W} and \mathbf{V} are learned weight matrices, \mathbf{b} are learned bias vectors, \otimes denotes the Hadamard product and σ is a logistic sigmoid function.

3.2 GRU

In comparison to LSTM, in GRU the cell state was eliminated so the hidden state is used to both store information and control the gating mechanism [20]. GRU only has two gates, a reset gate (*r*) and an update gate (*u*). The update gate acts in a similar way as the forget and input gates in LSTM. It decides what information to remove and what new information to add. The reset gate decides how much past information to forget. The output gate was eliminated. The gating mechanism of GRU and the corresponding equations are shown in Fig. 2.

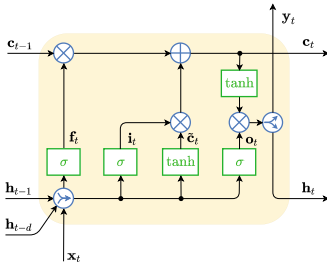
3.3 dLSTM

To improve the modeling of long-term dependencies in time series, we propose a dilated LSTM cell (Fig. 3). Our modification comes down to two elements. First, in addition to the hidden state \mathbf{h}_{t-1} , we introduce a delayed hidden state, \mathbf{h}_{t-d} , $d > 1$.



$$\begin{aligned} r_t &= \sigma(\mathbf{W}_r \mathbf{x}_t + \mathbf{V}_r \mathbf{h}_{t-1} + \mathbf{b}_r) \\ u_t &= \sigma(\mathbf{W}_u \mathbf{x}_t + \mathbf{V}_u \mathbf{h}_{t-1} + \mathbf{b}_u) \\ \tilde{\mathbf{h}}_t &= \tanh(\mathbf{W}_h \mathbf{x}_t + \mathbf{V}_h (r_t \otimes \mathbf{h}_{t-1}) + \mathbf{b}_h) \\ \mathbf{h}_t &= (1 - u_t) \otimes \mathbf{h}_{t-1} + u_t \otimes \tilde{\mathbf{h}}_t \\ \mathbf{y}_t &= \mathbf{h}_t \end{aligned}$$

Fig. 2 GRU



$$\begin{aligned} f_t &= \sigma(\mathbf{W}_f \mathbf{x}_t + \mathbf{V}_f \mathbf{h}_{t-1} + \mathbf{U}_f \mathbf{h}_{t-d} + \mathbf{b}_f) \\ i_t &= \sigma(\mathbf{W}_i \mathbf{x}_t + \mathbf{V}_i \mathbf{h}_{t-1} + \mathbf{U}_i \mathbf{h}_{t-d} + \mathbf{b}_i) \\ o_t &= \sigma(\mathbf{W}_o \mathbf{x}_t + \mathbf{V}_o \mathbf{h}_{t-1} + \mathbf{U}_o \mathbf{h}_{t-d} + \mathbf{b}_o) \\ \tilde{\mathbf{c}}_t &= \tanh(\mathbf{W}_c \mathbf{x}_t + \mathbf{V}_c \mathbf{h}_{t-1} + \mathbf{U}_c \mathbf{h}_{t-d} + \mathbf{b}_c) \\ \mathbf{c}_t &= \mathbf{f}_t \otimes \mathbf{c}_{t-1} + \mathbf{i}_t \otimes \tilde{\mathbf{c}}_t \\ \mathbf{h}'_t &= \mathbf{o}_t \otimes \tanh(\mathbf{c}_t) \\ \mathbf{h}_t &= [h'_{t,1}, \dots, h'_{t,s_h}] \\ \mathbf{y}_t &= [h'_{t,s_h+1}, \dots, h'_{t,s_h+s_y}] \end{aligned}$$

Fig. 3 dLSTM

This allows the data processing in time t to be controlled using not only information from the recent state but also using direct information from the delayed state. This can be useful for seasonal time series, in which case the dilation can correspond to the period of seasonal variations. Second, the output hidden state, \mathbf{h}'_t , is split into “real output” \mathbf{y}_t , which goes to the next layer, and a controlling output \mathbf{h}_t , which is an input to the gating mechanism in the following time steps. This solution was inspired by Ben-Ari and Shwartz-Ziv [21]. The size of the c -state is equal to the summed sizes of h -state and y -output, i.e., $s_c = s_h + s_y$.

The equations corresponding to dLSTM are shown in Fig. 3, where: \mathbf{W} , \mathbf{V} and \mathbf{U} are learned weight matrices, \mathbf{b} are learned bias vectors, and s_h , s_y are the lengths of hidden state and output vectors, respectively.

3.4 dRNNCell

A dilated recurrent NN cell, dRNNCell, was introduced in Smyl et al. [18] as a combination of GRU and LSTM cells, see Fig. 4. It was designed to operate as part of a multilayer dilated RNN [13]. Its output is split into \mathbf{y}_t and \mathbf{h}_t as in dLSTM.

As in LSTM, dRNNCell uses two states, i.e., c -state and h -state. But, unlike LSTM, dRNNCell is fed by both most recent states, \mathbf{c}_{t-1} and \mathbf{h}_{t-1} , and delayed states, \mathbf{c}_{t-d} and \mathbf{h}_{t-d} , $d > 1$. dRNNCell is equipped with three gates, which transform

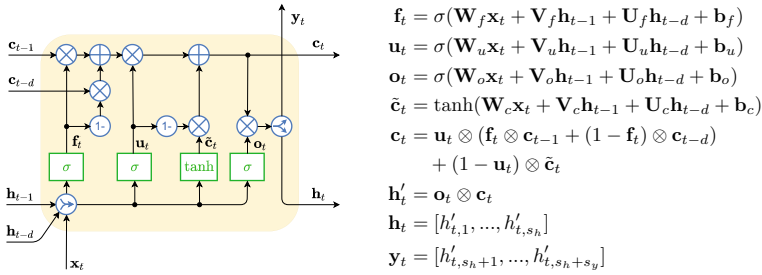


Fig. 4 dRNNCell

nonlinearly input vectors using logistic sigmoid function. They comprise fusion (f), update (u), and output (o) gates. A candidate c -state, $\tilde{\mathbf{c}}_t$, is produced by transforming input vectors using tanh nonlinearity. The operation of the cell is described by the equations shown in Fig. 4.

Note that the c -state is a weighted combination of past c -states and new candidate state $\tilde{\mathbf{c}}_t$ computed in the current step. Update vector, \mathbf{u}_t , decides in what proportion the old and new information are mixed in the c -state, while fusion vector \mathbf{f}_t decides about the contribution of recent and delayed c -states in the new state.

3.5 adRNNCell

An attentive dilated recurrent NN cell, adRNNCell, was proposed in Smyl et al. [22] as an extended version of dRNNCell. It combines two dRNNCells to obtain a more efficient cell, which is able to preprocess dynamically the sequence data. It is equipped with an attention mechanism for weighting the input information.

Figure 5 shows adRNNCell composed of lower and upper dRNNCells. The former produces attention vector \mathbf{m}_t of the same length as the input vector \mathbf{x}_t . The components of \mathbf{m}_t , after processing by exp function, are treated as weights for the inputs collected in \mathbf{x}_t . The weighted inputs, \mathbf{x}_t^2 , feed the upper cell. The goal of such an attention mechanism is to dynamically strengthen or weaken particular inputs depending on their relevance. Note that this process is dynamic, the weights are adjusted to the current inputs at time t . Both cells, lower and upper, learn simultaneously. Based on the weighted input vector, \mathbf{x}_t^2 , the upper cell predicts vector \mathbf{y}_t .

The mathematical model describing adRNNCell is shown in Fig. 5.

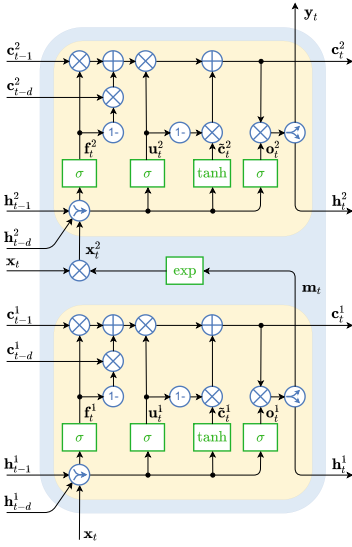


Fig. 5 adRNNCell

4 RNN Architecture

In this study, we adopt RNN architecture from Smyl et al. [22]. It is composed of three single-layer blocks, see Fig. 6. In each block, the cells are dilated differently, i.e., 2, 4 and 7, respectively. Delayed connections enable the direct input into the cell of information from a few time steps ago. This can be useful in modeling seasonal dependencies. To model the temporal dependencies of different scales, our architecture has multiple dilated recurrent layers stacked with hierarchical dilations. It also uses ResNet-style shortcuts between blocks to improve the learning process [12].

To reduce input dimensionality, the calendar variables, \mathbf{d}_t^w , \mathbf{d}_t^m and \mathbf{d}_t^y , are embedded using a linear layer into d -dimensional continuous vector \mathbf{d}_t . The second linear layer at the top of stacked recurrent layers, produces the point forecasts, $\hat{\mathbf{y}}_t$, and two vectors of quantiles, a lower one, $\hat{\mathbf{y}}_t \in \mathbb{R}^{24}$, and an upper one, $\hat{\mathbf{y}}_t \in \mathbb{R}^{24}$. These quantiles of assumed orders, q and \bar{q} , define the PI.

To enable our RNN to learn both point forecasts and PI quantiles, we employ the following loss function [18]:

$$L = \rho(y, \hat{y}_{q^*}) + \gamma(\rho(y, \hat{y}_q) + \rho(y, \hat{y}_{\bar{q}})) \quad (5)$$

where ρ is a pinball loss:

$$\rho(y, \hat{y}_q) = \begin{cases} (y - \hat{y}_q)q & \text{if } y \geq \hat{y}_q \\ (y - \hat{y}_q)(q - 1) & \text{if } y < \hat{y}_q \end{cases} \quad (6)$$

$$\begin{aligned} \mathbf{f}_t &= \sigma(\mathbf{W}_f \mathbf{x}_t + \mathbf{V}_f \mathbf{h}_{t-1} + \mathbf{U}_f \mathbf{h}_{t-d} + \mathbf{b}_f) \\ \mathbf{u}_t &= \sigma(\mathbf{W}_u \mathbf{x}_t + \mathbf{V}_u \mathbf{h}_{t-1} + \mathbf{U}_u \mathbf{h}_{t-d} + \mathbf{b}_u) \\ \mathbf{o}_t &= \sigma(\mathbf{W}_o \mathbf{x}_t + \mathbf{V}_o \mathbf{h}_{t-1} + \mathbf{U}_o \mathbf{h}_{t-d} + \mathbf{b}_o) \\ \tilde{\mathbf{c}}_t &= \tanh(\mathbf{W}_c \mathbf{x}_t + \mathbf{V}_c \mathbf{h}_{t-1} + \mathbf{U}_c \mathbf{h}_{t-d} + \mathbf{b}_c) \\ \mathbf{c}_t &= \mathbf{u}_t \otimes (\mathbf{f}_t \otimes \mathbf{c}_{t-1} + (1 - \mathbf{f}_t) \otimes \mathbf{c}_{t-d}) \\ &\quad + (1 - \mathbf{u}_t) \otimes \tilde{\mathbf{c}}_t \\ \mathbf{x}_t^2 &= \mathbf{x}_t \otimes \mathbf{m}_t \\ \mathbf{h}_t^{1'} &= \mathbf{o}_t^1 \otimes \mathbf{c}_t^1 \\ \mathbf{h}_t^1 &= [h_{t,1}^{1'}, \dots, h_{t,s_h}^{1'}] \\ \mathbf{m}_t &= [h_{t,s_h+1}^{1'}, \dots, h_{t,s_h+s_m}^{1'}] \\ \mathbf{h}_t^{2'} &= \mathbf{o}_t^2 \otimes \mathbf{c}_t^2 \\ \mathbf{h}_t^2 &= [h_{t,1}^{2'}, \dots, h_{t,s_q}^{2'}] \\ \mathbf{y}_t &= [h_{t,s_q+1}^{2'}, \dots, h_{t,s_q+s_y}^{2'}] \end{aligned}$$

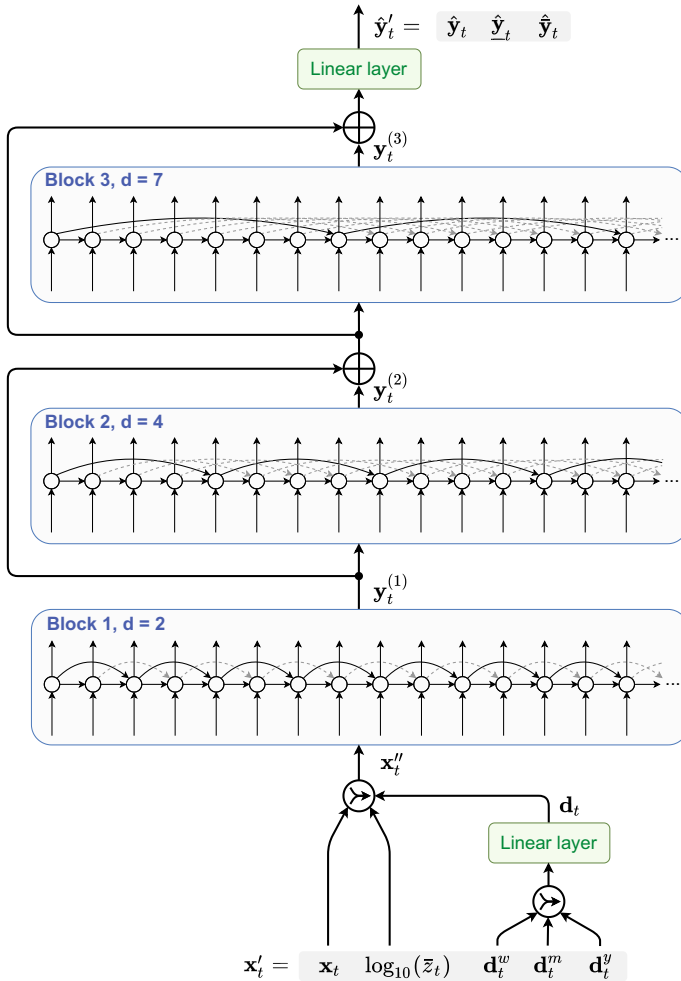


Fig. 6 RNN architecture

$q \in (0, 1)$ is a quantile order, y is an actual value (standardized), \hat{y}_q is a forecasted value of q -th quantile of y , $q^* = 0.5$ corresponds to the median, $\underline{q} \in (0, q^*)$ and $\bar{q} \in (q^*, 1)$ correspond to the lower and upper bound of PI, respectively, and $\gamma \geq 0$ is a parameter controlling the impact of the components related to PI on the loss function, typically between 0.1 and 0.5.

The first component in (5) is a symmetrical loss for the point forecast, while the second and third components are asymmetrical losses for the quantiles. The asymmetry level, which determines PI, results from the quantile orders. For example, we obtain a 90% symmetrical PI for $\underline{q} = 0.05$ and $\bar{q} = 0.95$.

Remarks:

1. In Sect. 5, we compare RNN with the different cell types which were described in Sect. 3. Note that RNN shown in Fig. 6 requires cells equipped with both recent and delayed connections. Classical LSTM and GRU are not equipped with delayed inputs. RNN with these cells are considered in two variants: (i) LSTM1, GRU1—the delayed connections are removed and cells are fed with only recent inputs $t - 1$, and (ii) LSTM2, GRU2—the recent connections are removed and cells are fed with only delayed inputs, $t - 2$, $t - 4$ or $t - 7$, depending in the layer.
2. The pinball loss gives the opportunity to reduce the forecast bias by penalizing positive and negative deviations differently. When the model tends to have a positive or negative bias, we can reduce the bias by introducing q^* smaller or larger than 0.5, respectively (see Smyl [15] and Dudek et al. [23]).

5 Experimental Study

We compare the performance of the proposed RNN with different recurrent cells on STLF problems for 35 European countries. The data, collected from ENTSO-E repository (www.entsoe.eu/data/power-stats), concerns real-world hourly electrical load time series. The data period is from 2006 to 2018 but a large amount of data is missing in this period (about 60% of the countries have complete data). For 35 countries, the data provides a variety of time series with triple seasonality expressing different properties such as different levels, trends, variance and daily shapes (see Section II in Smyl et al. [18] where these time series are analyzed). We treat data from 2018 as test data. We predict daily load profiles for each day of the test period and each country with the exception of three countries. For these three countries, due to missing data, the test periods were shorter, i.e., for Estonia and Italy (missing last month of data) and Latvia (missing last two months of data).

The RNNs were optimized on data from the period 2006–17. As performance metrics we use: mean absolute percentage error (MAPE), median of APE (MdAPE), interquartile range of APE (IqrAPE), root mean square error (RMSE), mean PE (MPE), and standard deviation of PE (StdPE). Below, we report results for an ensemble of five RNNs (average of five RNN runs). We use a similar training and optimization setup as in Smyl et al. [22]. The key hyperparameters were: $s_c = 250$, $s_h = s_q = s_y = 125$, $q^* = 0.5$, $\underline{q} = 0.05$, $\bar{q} = 0.95$, $\gamma = 0.3$, number of epochs: 10, learning rates: $3 \cdot 10^{-3}$ (epochs 1–5), 10^{-3} (epoch 6), $3 \cdot 10^{-4}$ (epoch 7), 10^{-4} (epochs 8–10), batch size: 2 (epochs 1–3), 5 (epochs 4–10).

Table 1 displays the forecasting quality metrics averaged over the 35 countries. The results indicate that, on average, adRNNCell is the best cell according to three accuracy measures, MAPE, MdAPE and RMSE. It also produces the least dispersed forecasts—see the lowest values of IqrAPE and StdPE. The second most accurate and precise cell is dRNNCell. The worst results are for GRU1.

Table 1 Forecasting quality metrics

Cell type	MAPE	MdAPE	IqrAPE	RMSE	MPE	StdPE
GRU1	2.31	2.10	2.23	318.69	-0.06	3.86
GRU2	2.26	2.04	2.19	308.92	-0.15	3.78
LSTM1	2.25	2.03	2.18	307.09	-0.19	3.78
LSTM2	2.16	1.94	2.10	293.00	-0.10	3.60
dLSTM	2.19	1.97	2.12	297.58	-0.19	3.66
dRNNCell	2.15	1.93	2.09	292.60	-0.15	3.57
adRNNCell	2.12	1.91	2.07	289.32	-0.14	3.52

Fig. 7 Results of the Giacomini-White test

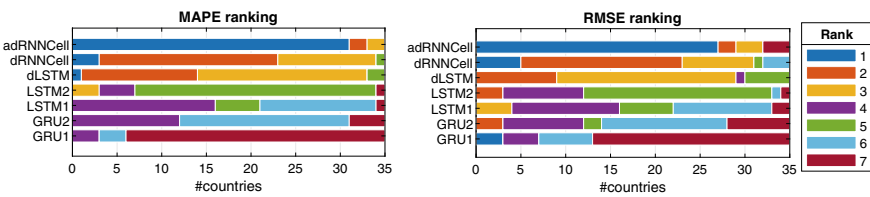
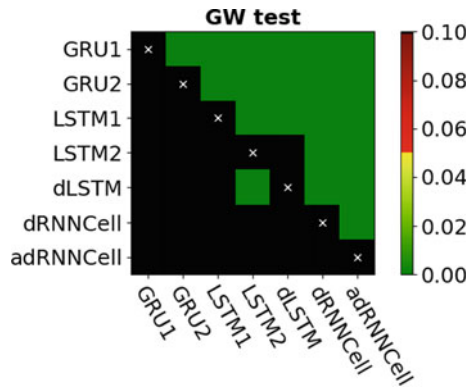


Fig. 8 Results of MAPE and RMSE rankings

To confirm the performance of adRNNCell, we perform a pairwise one-sided Giacomini-White test (GM test) for conditional predictive ability [24] (we used the multivariate variant of the GW test implemented in <https://github.com/jeslago/epftoolbox>; [25]). Figure 7 shows the obtained p -values of this test. The closer the p -values are to zero the significantly more accurate the forecasts produced by the model on the X -axis are than the forecasts produced by the model on the Y -axis. The black color is for p -values larger than 0.10, indicating rejection of the hypothesis that the model on the X -axis is more accurate than the model on the Y -axis. Figure 7 clearly shows that adRNNCell and dRNNCell performed best.

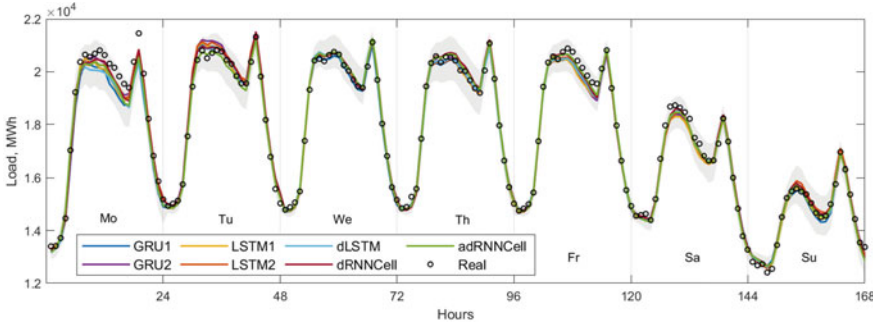


Fig. 9 Examples of the forecasts. 90% PIs for adRNNCell are shown as gray-shaded areas

Figure 8 shows rankings of the examined RNNs based on average errors for each country. Note the high position of adRNNCell. For 31 out of 35 countries this model gave the lowest MAPE, and for 27 countries it also gave the lowest RMSE. The second highest ranked model was dRNNCell.

In Table 1, we also show MPE, which is a measure of the forecast bias. Its negative values for all cases indicate over-prediction. The proposed model, thanks to the pinball-type loss function, can control the bias. For example, to reduce the bias for dRNNCell and adRNNCell we assumed $q^* = 0.485$. This resulted in a reduction of MPE to -0.04 without decreasing the forecast accuracy. So, it is possible to reduce the biases shown in Table 1 further, but we were trying to prevent over-tuning of the hyperparameters, so left them as they are reported.

Figure 9 shows example forecasts of daily profiles for different days of the week. Note that forecasts generated by RNN with different cells do not differ much from each other. Figure 9 also shows PIs for adRNNCell. To evaluate the accuracy of the PIs, we calculated the percentage of forecasts lying inside, above and below their PIs. The results are shown in Table 2. The predicted 90% PIs cover the forecasts most accurately in the case of GRU2. But note that our loss function (5) gives us the opportunity to tune further PIs. This can be performed by adjusting the quantiles determining the PI bounds, q and \bar{q} .

In Table 2, a Winkler score is also shown. For observations that fall within the PI, this score is simply the length of the PI, while for observations outside PI, the penalty applies, which is proportional to how far the observation is outside PI [26]. To bring the Winkler scores for different countries to a comparable level, we divide these scores by the mean loads of the corresponding countries in the test period. Such unified Winkler scores are shown in Table 2. Note that adRNNCell has the lowest Winkler score and dRNNCell the second lowest.

Our research shows that adRNNCell is the best gated cell for forecasting time series with multiple seasonality. In Smyl et al. [22] we compared RNN based on adRNNCells with a variety of forecasting models including statistical models (ARIMA, exponential smoothing, Prophet) and ML models (MLP, SVM, ANFIS,

Table 2 Evaluation of the PIs

Cell type	% in PI	% below PI	% above PI	Winkler score
GRU1	92.78±2.80	3.19±1.33	4.03±1.68	0.1524±0.2579
GRU2	90.40±3.40	4.79±1.74	4.81±1.80	0.1448±0.2710
LSTM1	89.16±4.07	5.30±2.05	5.53±2.21	0.1428±0.2760
LSTM2	88.52±3.88	5.53±1.96	5.96±2.04	0.1368±0.2741
dLSTM	88.84±3.74	5.59±1.96	5.57±1.94	0.1393±0.2737
dRNNCell	87.51±3.54	6.16±1.92	6.33±1.77	0.1363±0.2771
adRNNCell	88.41±3.14	5.68±1.67	5.91±1.62	0.1332±0.2683

LSTM, GRNN, nonparametric models). This comparison clearly showed that the adRNNCell-based approach outperforms all its competitors in terms of accuracy.

6 Conclusion

In this study, we explore the potential of RNNs with different cells for forecasting time series with multiple seasonality. The best RNN solutions use dRNNCells and adRNNCells, cells designed especially for such complex time series. They outperform classical GRU and LSTM cells as well as modified LSTM with dilation. adRNNCell, which is the most advanced cell with dilation and attention, combines two dRNNCells: one of which learns an attention vector while the other uses this vector to weight the inputs. The attention mechanism enables the cell to preprocess dynamically the sequence data while the delayed connections enable it to capture the long-term and seasonal dependencies in time series.

Apart from the dilation and attention mechanisms, the superior performance of the proposed RNN has its sources in the following mechanisms and procedures. First is the multilayer architecture, which is composed of several dilated recurrent layers stacked with hierarchical dilations to deal with multiple seasonality. Second is cross-learning on many time series, which enables RNN to capture the shared features of the individual series and helps to avoid over-fitting. Third is a time series representation using standardized weekly patterns as inputs and encoded daily patterns as outputs. The encoding variables are determined from the history, which enables decoding. Fourth is a composed asymmetrical loss function based on quantiles, which enables RNN to produce both point forecasts and PI and also to reduce the forecast bias.

In further research, we plan to enrich the input information with a learned context vector. This represents information extracted from other time series, which can help predict a given time series.

References

1. Box, G.E.P., Jenkins, G.M., Reinsel, G.C.: Time Series Analysis: forecasting and Control, 3rd edn. Prentice Hall, New Jersey (1994)
2. Taylor, J.W.: Triple seasonal methods for short-term load forecasting. *Euro. J. Oper. Res.* **204**, 139–152 (2010)
3. De Livera, A.M., Hyndman, R.J., Snyder, R.D.: Forecasting time series with complex seasonal patterns using exponential smoothing. *J. Am. Stat. Assoc.* **106**(496), 1513–1527 (2011)
4. Taylor, S.J., Letham, B.: Forecasting at scale. *Am. Stat.* **72**, 37–45 (2018)
5. Høverstad, B.A., Tidemann, A., Langseth, H., Öztürk, P.: Short-term load forecasting with seasonal decomposition using evolution for parameter tuning. *IEEE Trans. Smart Grid* **6**, 1904–1913 (2015)
6. Dudek, G.: Pattern-based local linear regression models for short-term load forecasting. *Electric Power Syst. Res.* **130**, 139–147 (2016)
7. Sharma, S., Majumdar, A., Elvira, V., Chouzenoux, E.: Blind Kalman filtering for short-term load forecasting. *IEEE Trans. Power Syst.* **35**, 4916–4919 (2020)
8. Benidis, K., Rangapuram, S.S., Flunkert, V., Wang, B., Maddix, D., Turkmen, C., Gasthaus, J., Bohlke-Schneider, M., Salinas, D., Stella, L., Callot, L., Januschowski, T.: Neural forecasting: introduction and literature overview (2020). [arXiv:2004.10240](https://arxiv.org/abs/2004.10240)
9. Hewamalage, H., Bergmeir, C., Bandara, K.: Recurrent neural networks for time series forecasting: current status and future directions. *Int. J. Forecast.* **37**, 388–427 (2021)
10. Yan, K., Wang, X., Du, Y., Jin, N., Huang, H., Zhou, H.: Multi-step short-term power consumption forecasting with a hybrid deep learning strategy. *Energies* **11**, 3089 (2018)
11. Bianchi, F.M., Maiorino, E., Kampffmeyer, M.C., Rizzi, A., Jenssen, R.: An overview and comparative analysis of recurrent neural networks for short term load forecasting (2017). [arXiv:1705.04378](https://arxiv.org/abs/1705.04378)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
13. Chang, S., Zhang, Y., Han, W., Yu, M., Guo, X., Tan, W., Cui, X., Witbrock, M., Hasegawa-Johnson, M., Huang, T.S.: Dilated recurrent neural networks. In: NIPS 2017 (2017)
14. Bandara, K., Bergmeir, C., Hewamalage, H.: LSTM-MSNet: leveraging forecasts on sets of related time series with multiple seasonal patterns. *IEEE Trans. Neural Netw. Learn. Syst.* **32**, 1586–1599 (2020)
15. Smyl, S.: A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *Int. J. Forecast.* **36**, 75–85 (2020)
16. Dudek, G.: Neural networks for pattern-based short-term load forecasting: a comparative study. *Neurocomputing* **205**, 64–74 (2016)
17. Dudek, G., Pelka, P., Smyl, S.: 3ETS+ RD-LSTM: a new hybrid model for electrical energy consumption forecasting. In: Proceedings of the International Conference on Neural Information Processing ICONIP, pp. 519–531. Springer (2020)
18. Smyl, S., Dudek, G., Pelka, P.: ES-dRNN: a hybrid exponential smoothing and dilated recurrent neural network model for short-term load forecasting. *IEEE Trans. Neural Netw. Learn. Syst.* (2023). <https://doi.org/10.1109/TNNLS.2023.3259149>
19. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997)
20. Cho, K., van Merriënboer, B., Gülçehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using RNN encoder-decoder for statistical machine translation (2014). [arXiv:1406.1078](https://arxiv.org/abs/1406.1078)
21. Ben-Ari, I., Shwartz-Ziv, R.: Sequence modeling using a memory controller extension for LSTM. In: NIPS 2017 Time Series Workshop (2017)
22. Smyl, S., Dudek, G., Pelka, P.: ES-dRNN with dynamic attention for short-term load forecasting. In: 2022 International Joint Conference on Neural Networks (IJCNN), pp. 1–8 (2022). <https://doi.org/10.1109/IJCNN55064.2022.9889791>

23. Dudek, G., Pełka, P., Smyl, S.: A hybrid residual dilated LSTM and exponential smoothing model for midterm electric load forecasting. *IEEE Trans. Neural Netw. Learn. Syst.* **33**, 2879–2891 (2022). <https://doi.org/10.1109/TNNLS.2020.3046629>
24. Giacomini, R., White, H.: Tests of conditional predictive ability. *Econometrica* **74**, 1545–1578 (2006)
25. Lago, J., Marcjasz, G., De Schutter, B., Weron, R.: Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Appl. Energy* **293**, 116983 (2021)
26. Hyndman, R.J., Athanasopoulos, G.: *Forecasting: principles and practice*. OTexts (2018)

Time Series Applications in Energy

Markov Processes for the Management of a Microgrid



Salvatore Vergine, César Álvarez-Arroyo, Guglielmo D'Amico,
Juan Manuel Escaño, and Lázaro Alvarado-Barrios

Abstract In this work, a stochastic modelization of wind and photovoltaic power productions is coupled with a two-level optimization in which the operating costs of a hybrid and isolated microgrid are minimized. First, the power demand is modeled and predicted using an autoregressive moving average model (ARMA), and the renewable productions are modeled using Markov reward processes. Then, the optimization problem is solved through a stochastic unit commitment and an economic dispatch. The results show that the stochastic models correctly capture the behavior of renewable sources in all system configurations proposed in the different scenarios. Furthermore, the different impacts caused by wind and photovoltaic sources and battery energy storage system on operating costs are also highlighted, which is more punctual for the first and more regular and smoother for the second.

Keywords Uncertainty · Economic dispatch · Stochastic unit commitment

1 Introduction

In recent years, the need to reduce carbon dioxide emissions and the use of fossil fuels has led to a transformation of centralized grids into decentralized networks, with the aim of increasing the penetration of renewable energy sources coupled with battery energy storage systems. In this context, microgrids have demonstrated their

S. Vergine (✉)

Department of Management, Marche Polytechnic University, 60121 Ancona, Italy
e-mail: s.vergine@univpm.it

C. Álvarez-Arroyo

Department of Electrical Engineering, Universidad de Sevilla, 41092, Sevilla, Spain

G. D'Amico

Department of Economics, University G. D'Annunzio, 65127 Pescara, Italy

J. Manuel Escaño

Department of Systems Engineering and Automatic Control, Universidad de Sevilla, 41092 Sevilla, Spain

L. Alvarado-Barrios

Department of Engineering, Universidad Loyola Andalucía, Avda. de las Universidades s/n, 41704 Dos Hermanas, Spain

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

189

O. Valenzuela et al. (eds.), *Theory and Applications of Time Series Analysis*,
Contributions to Statistics, https://doi.org/10.1007/978-3-031-40209-8_13

fundamental role in the management of photovoltaic and wind units [10]. Generally, the optimization problem applied to the microgrid is composed of the following two stages: the unit commitment to schedule the on/off state of the controllable power units and the battery behavior, and the economic dispatch to know how much power each available generating unit should supply to minimize the operating cost. This process is a mixed integer nonlinear optimization problem (MINLP) due to the binary variable representing the on/off decision and the nonlinearity of the operating cost. The intermittent nature of renewable energy sources and the resulting uncertainty make it essential to add a stochastic approach to this problem, with the Stochastic Unit Commitment [1, 13]. The deterministic formulation of this kind of problem considers only a predicted scenario and assumes that controllable power units have enough available capacity to supply the difference in power between real and forecast demand. Studies [8, 15] propose a deterministic unit commitment to solve optimization problems related to microgrids to minimize the objective function. This approach shows some limitations, such as not considering the stochastic nature of renewable energy sources or limiting the study only to some of them, such as power demand. In the study [17], the authors consider the uncertain nature of wind and photovoltaic power sources in a two-state optimization problem through Markov processes. These processes are one of the most important categories of stochastic models that have many applications in real-world problems, including wind and photovoltaic power predictions [6, 18]. They non-parametrically describe the sequencing of an event by assuming that what happens in the next state depends only on the current state [9]. The obtained results show that the stochastic models faithfully represent the generation of renewable energy, which has a significant impact on the microgrid's overall cost.

In this study, we apply the methodology proposed in [17], where, firstly, the uncertainties of wind and photovoltaic power productions are modeled according to general Markov processes, and, secondly, a microgrid optimization problem is solved. The latter considers a two-stage problem with the aim of minimizing the total operating costs under technical constraints related to the design of the microgrid. We extend this contribution by measuring the impact of each renewable unit and the battery energy storage system through the implementation of five different scenarios.

2 System Modeling

In this section, the microgrid structure is described and the stochastic modeling of power demand, wind, and photovoltaic power is presented and explained.

2.1 Microgrid Structure

The microgrid is composed of two controllable power units, a microturbine (MT) and a diesel engine (DE), two renewable sources, a wind turbine (WT) and a photovoltaic unit (PV), and a battery energy storage systems (BESS). This is a hybrid and isolated microgrid studied in an optimization problem to minimize its operating cost. As shown in Fig. 1, this study is divided into three stages: the first deals with the modeling of the two renewable sources and the demand for electricity, the second and third are part of the optimization problem and include a stochastic unit commitment and two economic dispatches.

2.2 ARMA Model for Demand Prediction

The data set to model and predict the power demand D_t is based on 2-year historical data from Sardinia (Italy). According to the recent literature [2, 17], we use an autoregressive moving average (ARMA) model. The ARMA model is composed of two polynomials: the first for the autoregressive part (AR), which regresses the variable on its past values, and the second for the moving average part (MA), which models the errors as a linear combination of their values from the present and the past; see, e.g. [4].

The ARMA model is defined as

$$D_t = \phi_1 D_{t-1} + \dots + \phi_p D_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}. \tag{1}$$

At this point, we introduce the back-shift operator B

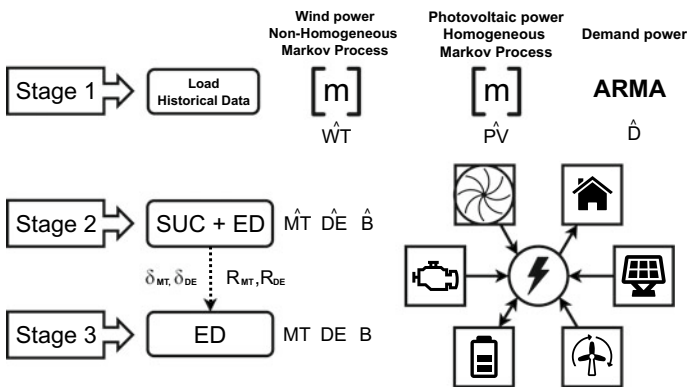


Fig. 1 Optimization problem structure

$$BD_t = D_{t-1}. \quad (2)$$

Applying B to D_{t-1} , we obtain $BD_{t-1} = D_{t-2}$ and if we replace it in Eq. (2), we obtain $B(BD_t) = B^2D_t = D_{t-2}$. In general, $B^jD_t = D_{t-j}$, with $j = 1, 2, \dots$

Using the operator B , we can write Eq. (1) as follows:

$$(1 - \phi_1B - \phi_2B^2 - \dots - \phi_pB^p)D_t = (1 - \theta_1B - \dots - \theta_qB^q)a_t, \quad (3)$$

or, equivalently, $\phi_p(B)D_t = \theta_q(B)a_t$, where ϕ_p and θ_q are polynomials of order p and q , respectively, in B . Therefore, we obtain the equation $D_t = \frac{\theta_q(B)}{\phi_p(B)}a_t$, where $\{a_t\} \sim WN(0, \sigma_a^2)$ is a white noise process.

D_t is the demand at time t that must be predicted, and ϕ_p and θ_q are the coefficients of the autoregressive model and the moving averages, respectively. a_t represents errors due to predictions. The value of the demand at any time t is a linear combination of its previous values p and the previous values q of errors. We use 2-year data to model power demand, and the Akaike Information Criterion (AIC) is applied to obtain the values of $p = 6$ and $q = 5$.

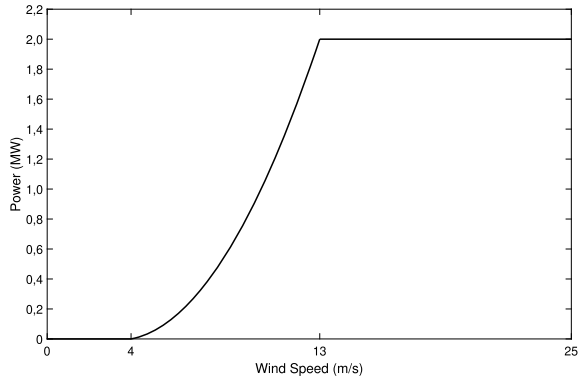
2.3 Non-homogeneous Markov Model for Wind Generation

We used a ten-year data set consisting of hourly wind speed from 01/08/2008 to 01/08/2018 provided by [5]. The considered location is in Sardinia (39.5N 8.75E). We use the following power curve to obtain wind power production for $\forall t \in \{1, 2, \dots, 24\}$ [12, 16, 17]:

$$WT(t) = \begin{cases} 0 & \text{for } v < v_{ci} \text{ and } v > v_{co} \\ P_r \frac{v^3 - v_{ci}^3}{v_r^3 - v_{ci}^3} & \text{for } v_{ci} < v < v_r, \\ P_r & \text{for } v_r < v < v_{co}, \end{cases} \quad (4)$$

where P_r is the rated power (in kW) and represents the maximum power that the WT can produce at the rated wind speed v_r (in m/s). v_{ci} is the wind speed cut-in and represents the speed at which the WT starts to produce power; vice versa, v_{co} is the wind speed cut-off, which is the point at which the WT must shut down due to the risk of damage, and v is the current wind speed. In this case, we use the power curve plotted in Fig. 2, where $P_r = 2$ MW, $13 \leq v_r < 25$ m/s, $v_{ci} = 4$ m/s and $v_{co} = 25$ m/s. We scaled the data obtained to have a maximum wind power production equal to 80 kW, which corresponds to about 30% of the maximum hourly power request. We model wind power production as done in [17], building a non-homogeneous Markov reward process with state space $E^W = \{1, 2, 3, 4, 5\}$ in which each state denotes a different level of wind power. In particular, a power production between 0 and 16 kW is identified with state 1, 16 and 32 kW which are the limits that identify state 2, and state 3 includes the power values between 32 and 48 kW, state 4 between 48 and 64 kW, and state 5 between 64 and 80 kW. We indicate the wind power at the

Fig. 2 Power curve of the considered WT



$n - th$ transition with J_n^W which occurs from one hour to the next and depends on the current hour. The relationship that governs the 24 transitions that occur in 24 h is the following:

$$P[J_{n+1}^W = j \mid J_n^W = i_n, J_{n-1}^W = i_{n-1}, \dots, J_1^W = i_1, J_0^W = i_0] = P[J_{n+1}^W = j \mid J_n^W = i_n] = p_{i_n, j}(n + 1). \tag{5}$$

Thus, the relation (5) affirms that the wind power at the next hour depends only on the wind power at the current hour and not on the hours before. However, the probabilities depend on time $n + 1$, and thus change according to the hour of the day.

In practice, the transition probabilities in formula (5) can be estimated with the maximum likelihood estimator as follows:

$$\hat{p}_{i_n, j}(n + 1) = \frac{N_{i_n, j}(n + 1)}{N_{i_n}(n)}, \tag{6}$$

where $N_{i_n, j}(n + 1)$ indicates the number of transitions from state i at hour n to state j at hour $n + 1$, and the denominator $N_{i_n}(n)$ represents the total number of visits to state i at hour n . In this way, we have a total of 24 probability transition matrices, one for each possible transition within a day, that is, $\mathbf{P}(k) = (p_{ij}(k))_{i, j \in E^W}$, $k = 1, 2, \dots, 24$. We define $\Phi(s, t) = (\phi_{ij}(s, t))_{i, j \in E^W}$, $s, t \in \mathbb{N}$ a matrix of functions with elements obtained by $\phi_{ij}(s, t) = \mathbb{P}[J_t^W = j \mid J_s^W = i]$ that denote the probability of having wind power j at time t , since wind power is i at time s . Transition probability functions are obtained according to the following equation:

$$\Phi(s, t) = \prod_{k=s+1}^t P(k), \quad s, t \in \mathbb{N}, \quad 0 \leq s \leq t. \tag{7}$$

Table 1 The Markov transition matrices refer to four different hours within a day

$ \begin{matrix} & 1 & 2 & 3 & 4 & 5 \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{pmatrix} 0.97 & 0.03 & 0 & 0 & 0 \\ 0.11 & 0.81 & 0.08 & 0 & 0 \\ 0 & 0.16 & 0.73 & 0.10 & 0.01 \\ 0 & 0 & 0.26 & 0.62 & 0.12 \\ 0 & 0 & 0.02 & 0.11 & 0.87 \end{pmatrix} \end{matrix} $	$ \begin{matrix} & 1 & 2 & 3 & 4 & 5 \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{pmatrix} 0.97 & 0.02 & 0.01 & 0 & 0 \\ 0.18 & 0.66 & 0.15 & 0.01 & 0 \\ 0 & 0.18 & 0.63 & 0.14 & 0.05 \\ 0 & 0.01 & 0.18 & 0.55 & 0.26 \\ 0 & 0 & 0.02 & 0.07 & 0.91 \end{pmatrix} \end{matrix} $
(a) 1:00 am - 2:00 am	(b) 5:00 am - 6:00 am
$ \begin{matrix} & 1 & 2 & 3 & 4 & 5 \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{pmatrix} 0.95 & 0.05 & 0 & 0 & 0 \\ 0.09 & 0.75 & 0.16 & 0 & 0 \\ 0 & 0.12 & 0.71 & 0.17 & 0 \\ 0 & 0 & 0.17 & 0.63 & 0.20 \\ 0 & 0 & 0 & 0.08 & 0.92 \end{pmatrix} \end{matrix} $	$ \begin{matrix} & 1 & 2 & 3 & 4 & 5 \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{pmatrix} 0.98 & 0.02 & 0 & 0 & 0 \\ 0.32 & 0.61 & 0.06 & 0.01 & 0 \\ 0 & 0.40 & 0.50 & 0.08 & 0.02 \\ 0 & 0.02 & 0.48 & 0.42 & 0.08 \\ 0 & 0 & 0.03 & 0.15 & 0.82 \end{pmatrix} \end{matrix} $
(c) 12:00 - 1:00 pm	(d) 6:00 pm 7:00 pm

In Table 1, we report four probability matrices referring to the transition from hour 1:00 am to hour 2:00 am (a), from 5:00 am to 6:00 am (b), from 12:00 to 1:00 pm (c) and from 6:00 pm to 7:00 pm (d). For example, we can notice that the system moves from state 4 to state 5 with very different probabilities (0.12; 0.26; 0.20; 0.08) depending on the time we consider (1:00 am; 5:00 am; 12:00 pm; 6:00 pm).

Whenever the system occupies a state $i \in E^W$ at any hour, the Markov chain gives us the range of possible power values. To establish the point value, we consider a r.v. R that describes the power conditionally in the occupied state i . In symbols,

$$F_i(x) = \mathbb{P}(R \leq x | J_n = i) \quad \forall n \in \{1, 2, \dots, 24\} \text{ and } i \in E^W. \tag{8}$$

The cumulative distribution function (cdf) $F_i(\cdot)$ can be empirically estimated:

$$\hat{F}_i(x) = \frac{\sum_{h=1}^{N_i} \mathbf{1}_{\{w_i \leq x\}}}{N_i}, \tag{9}$$

where $N_i = \sum_{k=1}^{24} N_i(n)$, and $\{w_i\}_{i=1}^{N_i}$ are the sample values of the powers belonging to the state range i . In Fig. 3 the empirical cdf of the wind powers is shown.

The curve related to state 1 has a probability mass equal to about 0.4 corresponding to zero production. This happens when the wind speed is below the cut-in speed. The curve referred to state 5 has a probability mass equal to about 0.7 corresponding to the case of maximum production when 13 m/s wind speed or greater occurs.

Fig. 3 Empirical cumulative distribution functions for each state of the Markov chain

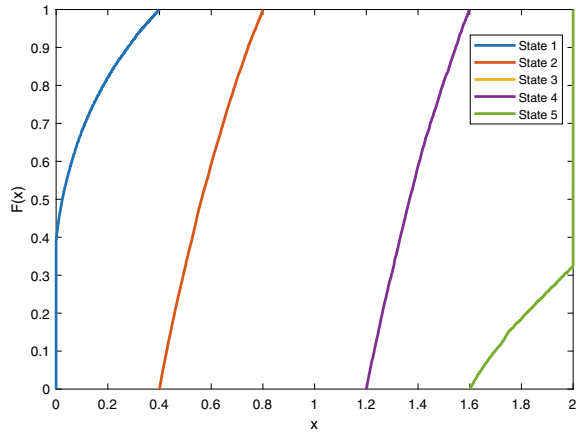
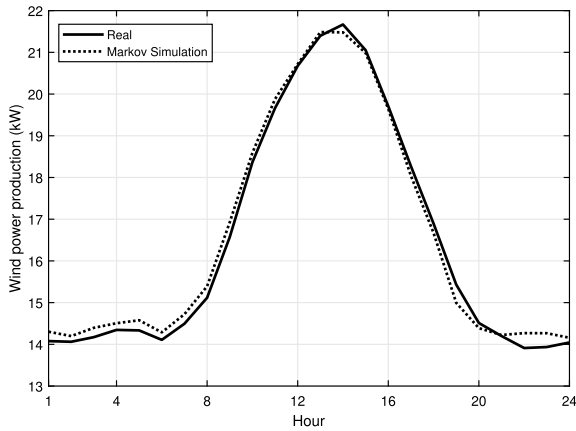


Fig. 4 Comparison between real and simulated hourly wind power averages



After obtaining the 24 transition probability matrices and the 5 cumulative distribution functions, we simulate power production in order to have synthetic data as follows:

1. Set $n = 0$ and $J_0 = i_0, R_0 = 0$.
2. Sample the r.v. $J \sim p_{J_n, \cdot}(n)$ and set $J_{n+1} = J(\omega)$.
3. Sample the r.v. $R \sim F_{J_{n+1}}(\cdot)$ and set $R_{n+1} = R(\omega)$.
4. Set $n = n + 1$ and continue to Step 2 until $n = 24$.

The application of this algorithm gives simulated wind power for one day. The algorithm can run for several days as requested in the application.

Figure 4 shows that the hourly averages of wind power for real and simulated data have equal behavior, confirming the accuracy of the used Markov reward process.

2.4 Homogeneous Markov Model for Photovoltaic Generation

We download ten-year hourly solar irradiation data set referring to the same period and geographical coordinates that characterize wind speed data [5]. We obtain the photovoltaic power $PV(t)$ using the following equation [11, 17]:

$$PV(t) = P_{STC} \frac{n \cdot E_{M,t}}{E_{STC}} [1 + h(T_{M,t} - T_{STC})], \quad \forall t \in \{1, 2, \dots, 24\}. \quad (10)$$

$PV(t)$ depends on solar irradiance $E_{M,t}$, on module temperature $T_{M,t}$, on maximum power P_{STC} , on irradiance E_{STC} and on temperature T_{STC} under Standard Test Conditions (STC). The coefficients n and h indicate the number of PV panels and the power temperature coefficient ($\%/^{\circ}\text{C}$), respectively. In this case, we scale the data to a maximum power equal to 40 kW, which corresponds to approximately 15% of the maximum hourly power request in the microgrid.

The photovoltaic production is modeled as done in [17]. First of all, it is detrended by calculating the difference between the average hourly power of the considered month and the power data. PV power presents a daily and a monthly periodicity. This operation allows us to obtain the residuals that we model using a homogeneous Markov reward process [7]. This is a particular case of the non-homogeneous process used to model the wind speed data, and it is characterized by only 1 probability transition matrix valid for each time n .

Let J_n^{PV} be the random values that represent the state of the system in the n -th transition and have values within the set $E^{PV} = \{1, 2, 3\}$. In this case, state 1 groups all residual values less than 0, state 2 for all residual values equal to 0, and state 3 for all residuals greater than 0. The transition probabilities satisfy the following relation:

$$P[J_{n+1}^{PV} = j \mid J_n^{PV} = i_n, J_{n-1}^{PV} = i_{n-1}, \dots, J_1^{PV} = i_1, J_0^{PV} = i_0] = P[J_{n+1}^{PV} = j \mid J_n^{PV} = i_n] = p_{ij}. \quad (11)$$

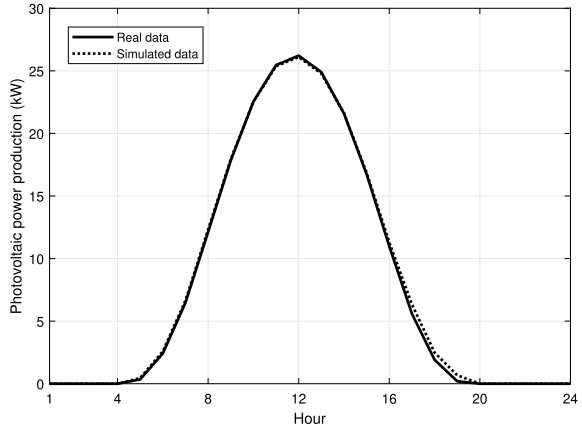
The transition probability matrix is as follows:

$$\begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} 0.91 & 0.04 & 0.05 \\ 0.08 & 0.82 & 0.10 \\ 0.08 & 0.10 & 0.82 \end{pmatrix} \end{matrix}. \quad (12)$$

Thus, if the system is in state 2, it will remain there with a probability of 82%.

After obtaining the simulated Markov chain, we estimate the cdf for states 1 and 3 with the empirical estimator (as done for wind power). This gives the punctual values of the residual. Finally, we sum the detrended power and the simulated residuals and obtain the simulated power over the day. In Fig. 5, the real and simulated photovoltaic hourly averages are compared over 24 h.

Fig. 5 Comparison between real and simulated hourly photovoltaic power averages



3 Optimization Strategy

In this section, the optimization strategy presented in [17] is described. The first subsection shows the system constraints, and the second subsection describes the two-step optimization process.

3.1 System Constraints

The total power generated by controllable units (MT and DE), renewable sources (WT and PV), and the power supplied or stored in the BESS must be equal to the demand. This constraint is valid for each time step t and is described by the following equation representing the power balance:

$$D(k) = PV(t) + WT(t) + BESS(t) + \delta_{MT}(t) \cdot MT(t) + \delta_{DE}(t) \cdot DE(t). \quad (13)$$

The power provided by MT and DE is bounded at any time, i.e.

$$DE_{\min} \leq DE(t) \leq DE_{\max} \quad (14)$$

$$MT_{\min} \leq MT(t) \leq MT_{\max} \quad (15)$$

where DE_{\max} and DE_{\min} , and MT_{\max} and MT_{\min} are the maximum and minimum powers of DE and MT, respectively. Regarding the battery energy storage system (BESS), it provides or stores energy depending on the demand power and its state of charge (SOC) is updated at each time t using the following equation $\forall t \in \{1, 2, \dots, 24\}$ [3]:

$$\text{SOC}(t) = \text{SOC}(t-1) - \begin{cases} \Delta t \cdot B(t) \cdot \eta_c & \text{for } \text{BESS}(t) < 0 \\ \frac{\Delta k \cdot B(t)}{\eta_d} & \text{for } \text{BESS}(t) > 0 \end{cases} \quad (16)$$

where Δt is the time between samples and η_c and η_d are the charging and discharging efficiency, respectively. The SOC limits must satisfy the inequality:

$$\text{SOC}_{\min} \leq \text{SOC}(t) \leq \text{SOC}_{\max} \quad (17)$$

for all $t \in \{1, \dots, 24\}$, where SOC_{\max} and SOC_{\min} are the maximum and minimum levels of SOC determined by the capacity of the BESS connected to the system.

3.2 Optimization Process

In this study, a two-step optimization process is applied to obtain the power delivered by each controllable power unit at each time step during the day. We indicate with $\hat{d}(t)$ the vector containing the forecast of demand $\hat{D}(t)$ and renewable sources $\hat{P}(t)$ and $\hat{W}(t)$, and $u(t)$ the vector associated with controllable sources BESS(t), MT(t) and DE(t), respectively. We denote them as follows:

$$\hat{d}(t) = [\hat{D}(t), \hat{P}(t), \hat{W}(t)], \quad (18)$$

$$u(t) = [\delta_{\text{MT}}(t), \delta_{\text{DE}}(t), \text{BESS}(t), \text{MT}(t), \text{DE}(t)], \quad (19)$$

where $\delta_{\text{MT}}(t)$ and δ_{DE} are binary variables that indicate whether the power unit is on or off. The first step of the optimization process (Stage 2) consists in solving the stochastic unit commitment in which a controllable power source is turned on or off at each time step t for a decision horizon of one day. Then, an economic dispatch is solved to know how much power each power unit must deliver to cover the predicted demand.

The second step of the optimization process (Stage 3) begins when the microgrid real-time operations are calculated and the difference between the demand prediction and the current value is known. The error is covered by using the spinning reserve delivered by the BESS and the two controllable units. It is assumed that these power units are always able to cover the error, which guarantees reliable microgrid operation. Then, in the second optimization problem, a second economic dispatch is solved, in which it is decided which power unit must cover the variation in demand to minimize the operating cost.

The spinning reserve represents an extra generation unit and is indicated as $R_{\text{MT}}^u(t)$ when provided by the DE and $R_{\text{MT}}^u(t)$ when provided by the MT. The equations that represent it analytically are the following $\forall t \in \{1, 2, \dots, 24\}$:

$$R_{DE}^u(t) = DE_{\max} - DE(t) \quad (20)$$

$$R_{MT}^u(t) = MT_{\max} - MT(t). \quad (21)$$

The prediction of demand $\hat{D}(t)$ is affected by an error with a normal distribution $\mathcal{N}(\mu_t, \sigma_t^2)$. Taking into account the ARMA model used to predict demand, in 99.73% of the cases, these errors are within the interval $[\mu_t - 3\sigma_t, \mu_t + 3\sigma_t]$, according to the criteria 3σ . For this reason, the spinning reserve must respect the following equation:

$$R_{DE}^u(t) + R_{MT}^u(t) \geq 3\sigma_t. \quad (22)$$

These constraints are reported considering the minimum power that can be generated:

$$R_{DE}^d(t) = DE(t) - DE_{\min} \quad (23)$$

$$R_{MT}^d(t) = MT(t) - MT_{\min} \quad (24)$$

$$R_{DE}^d(t) + R_{MT}^d(t) \geq 3\sigma_t. \quad (25)$$

MT and DE are the only power units with which a cost is associated. In particular, we indicate with C_{MT} the cost of the power produced by MT and with C_{DE} the cost of the power produced by DE for each time step and for $j = 0, \dots, N - 1$. Furthermore, we indicate with $C_{R_{MT}}$ the cost of the spinning reserve supplied by MT and with $C_{R_{DE}}$ the cost of the spinning reserve supplied by DE. The costs associated with controllable power units are the same as those considered in [2, 14, 17]. Therefore, we formulate the first optimization problem as follows:

$$\min_{\substack{\delta_{MT}, \delta_{DE}, \\ BESS, MT, DE}} \sum_{j=0}^{N-1} C_{MT}(\delta_{MT}(j), M(j)) + C_{MT}(\delta_{MT}(j), DE(j)) + C_{R_{MT}} R_{MT}^u(j) + C_{R_{MT}} R_{MT}^d(j) \quad (26a)$$

$$\text{s.t. } \hat{D}(t+j) = \hat{P}V(t+j) + \hat{W}T(t+j) + B(j) + MT(j) + DE(j) \quad (26b)$$

$$\hat{S}OC(j+1|t) = f(\hat{S}OC(j|k), BESS(j)) \quad (26c)$$

$$\hat{S}OC(0) = SOC(t) \quad (26d)$$

$$\hat{S}OC(j|t) \in \mathcal{S}, \quad (26e)$$

$$\delta_{MT}(j), \delta_{DE}(j) \in \{0, 1\} \quad (26f)$$

$$MT(j), DE(j), B(j) \in \mathcal{E} \quad (26g)$$

$$R^u(j) \geq 3\sigma_j \quad (26h)$$

$$R^d(j) \geq 3\sigma_j \quad (26i)$$

where $\hat{S}OC(j+1|t)$ is the SOC prediction for time $j+1$ at the current time t . Set \mathcal{S} indicates the BESS charge and set \mathcal{E} indicates the power limits.

In stage 3, the binary variables $\delta_{MT}(t)$, $\delta_{DE}(t)$ are constant and the demand is obtained by adding the error to the expected demand, as follows:

$$D(t) = \hat{D}(t) + \omega(t). \quad (27)$$

In this step, the power increments of each controllable power unit are decided to cover the current demand and minimize the operating cost:

$$\min_{\Delta MT, \Delta DE, \Delta BESS} \sum_{j=0}^{N-1} [a_{MT} \delta_{MT}(j) + b_{MT}(MT(j) + \Delta MT(j)) + c_{MT}(MT(j) + \Delta MT(j))^2 + a_{MT} \delta_{MT}(j) + b_{MT}(DE(j) + \Delta DE(j)) + c_{MT}(DE(j) + \Delta DE(j))^2] \quad (28a)$$

$$\text{s.t. } \omega(j) = \Delta MT(j) + \Delta DE(j) + \Delta BESS(j), \quad (28b)$$

$$DE_{\min} \delta_{MT}(j) \leq DE(j) + \Delta DE(j) \leq DE_{\max} \delta_{MT}(j), \quad (28c)$$

$$MT_{\min} \delta_{MT}(j) \leq MT(j) + \Delta MT(j) \leq MT_{\max} \delta_{MT}(j), \quad (28d)$$

$$\hat{SOC}(j+1|t) = f(\hat{SOC}(j|t), BESS(j) + \Delta BESS(j)), \quad (28e)$$

$$\hat{SOC}(0) = SOC(t), \quad (28f)$$

$$\hat{SOC}(j|t) \in \mathcal{S}, \quad (28g)$$

$$MT(j) + \Delta MT(j), DE(j) + \Delta DE(j), BESS(j) + \Delta BESS(j) \in \mathcal{E}. \quad (28h)$$

4 Application

In this section, six scenarios are presented and explained, characterized by different configurations of the system. Then, we present and comment on the results obtained.

4.1 Scenarios Studied

This methodology is applied to six different scenarios shown in Table 2. We consider a Base Scenario in which all the parameters of the different power units belonging to the microgrid are established. In Scenario 1 we assume that the contribution of renewable energies is absent. The meaning of this scenario is to have a visual measure of the maximum cost (or the maximum objective function) facing the microgrid in the worst case, which means using power supplied only from MT and DE. Scenario 2 presents where the maximum power of the WT is reduced to 40 kW and the maximum power of the PV is increased to 80 kW. The same quantities are set to the same value equal to 80 kW in Scenario 3. In Scenario 4, the capacity of the BESS is increased up to double compared to the Base Scenario, and in Scenario 5 both maximum charge and discharge power of the BESS are increased from 80 to 120 kW.

Table 2 Parameters of the base scenario and differences of all the other scenarios

Base scenario						
Source	P_{\max} (kW)	P_{\min} (kW)	P_{\max}^c (kW)	P_{\max}^d (kW)	SOC _{min} (kW)	SOC _{max} (kW)
MT	150	5	–	–	–	–
DE	150	15	–	–	–	–
WT	80	0	–	–	–	–
PV	40	0	–	–	–	–
BESS	–	–	80	80	70	280
Scenario 1	No use of renewable sources					
Scenario 2	Maximum power of WT unit equal to 40 kW and maximum power of PV unit equal to 80 kW					
Scenario 3	Maximum power of WT unit equal to 80 kW and maximum power of PV unit equal to 80 kW					
Scenario 4	Double capacity of the BESS with SOC _{min} = 70 kWh and SOC _{max} = 490 kWh					
Scenario 5	Maximum charge power of the BESS $P_{\max}^c = 120$ kW and maximum discharge power of the BESS $P_{\max}^d = 120$ kW					

4.2 Results

In this section, we present the result obtained by applying the two-level optimization process in each scenario studied for both real and simulated data and refer to 30-day sample considered (from 2008/08/01 to 2008/08/30). In Fig. 6, an example of the result of the economic dispatch for real (6a) and simulated (6b) data referred to August 23 is shown. This allows us to know exactly which power unit contributes to satisfy the power demand at each hour of the day, and it emphasizes the behavior of the system that charges the BESS when the power demand is low and discharges when it is high to decrease the operating cost.

Hereafter, all figures present the same setting and show the daily objective function and the total wind and photovoltaic powers for real and synthetic data. The results referred to the Base Scenario are shown in Fig. 7 where it is evident that the objective function is largely influenced by the amount of power supplied by the wind plant, while the photovoltaic power determines the general trend or slope acquired by the objective function. When the trend of photovoltaic power decreases with time, the objective function shows an increasing trend. Scenario 1 corresponds to the worst-case situation in which there is no contribution from renewable energies. In fact, the objective function develops in a range from 9940€ to 9915€, as shown in Fig. 8. Scenario 2 represents the situation in which the PV has a higher maximum power (equal to 80 kW), while the WT can deliver a maximum of 40 kW. The results are shown in Fig. 9.

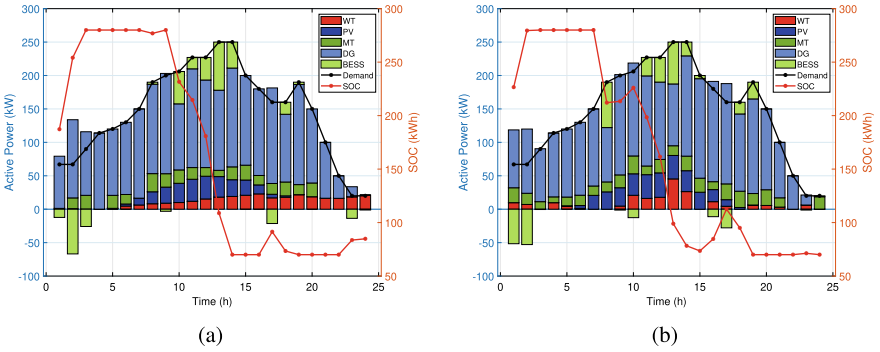


Fig. 6 Economic dispatch of the real (a) and (b) simulated data referred to August 23

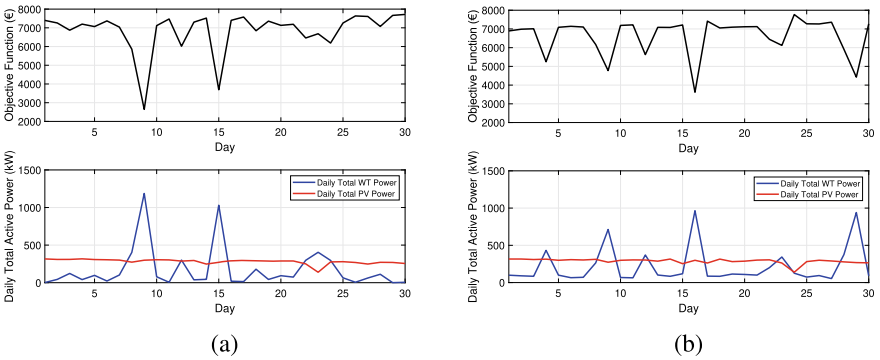
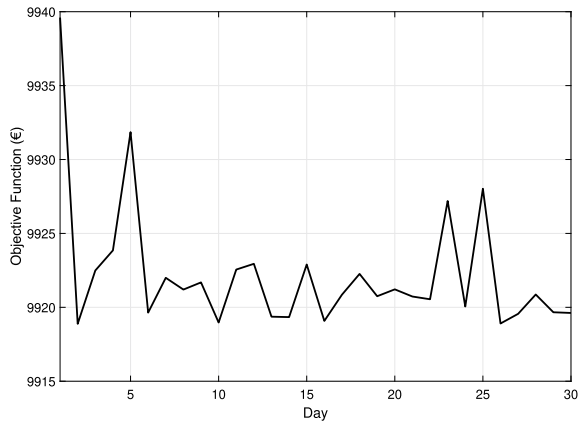


Fig. 7 Objective function and daily total wind and photovoltaic powers in the base scenario for real (a) and synthetic (b) data

Fig. 8 Objective function in scenario 1



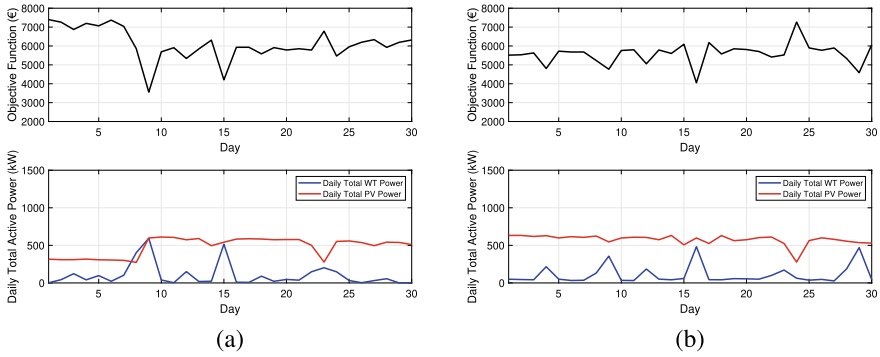


Fig. 9 Objective function and daily total wind and photovoltaic powers in Scenario 2 for real (a) and synthetic (b) data

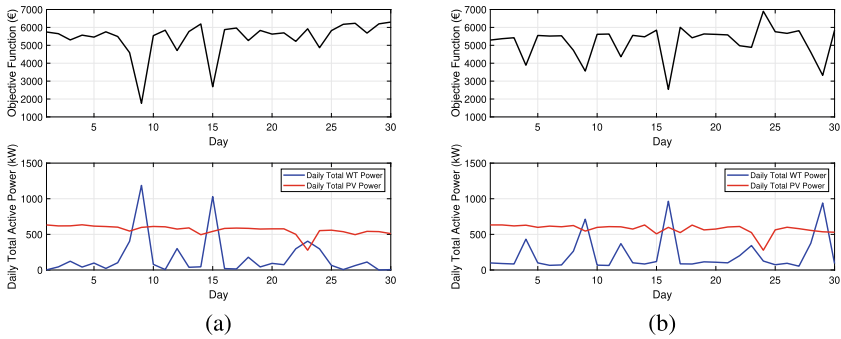


Fig. 10 Objective function and daily total wind and photovoltaic powers in Scenario 3 for real (a) and synthetic (b) data

In Scenario 3, both the wind and photovoltaic units are set with the same maximum power (80 kW) and in Fig. 10 it is evident that the objective function values decrease in the sample considered compared to the previous scenarios. In Scenario 4 a double BESS capacity is used, and in Fig. 11 the results are shown. It is noticeable that in Scenarios 3 and 4 a high availability of wind power corresponds to lower values of the objective function in comparison with the other scenario. This is because a greater amount of wind power can be used both for the larger BESS capacity and for the higher maximum power of WT and PV. In Scenario 5, the maximum charge and discharge powers of the BESS are increased to 120 kW, and the results are presented in Fig. 12. Table 3 shows the mean and standard deviation of the objective function referred to both real and synthetic data and for each scenario studied.

By comparing each scenario with the Base Scenario, in terms of mean, it is evident that the worst case occurs for Scenario 1 when there is no contribution from wind and solar sources. The situation improves in Scenario 2 when the highest maximum renewable power is assigned to the photovoltaic source, even if the total maximum

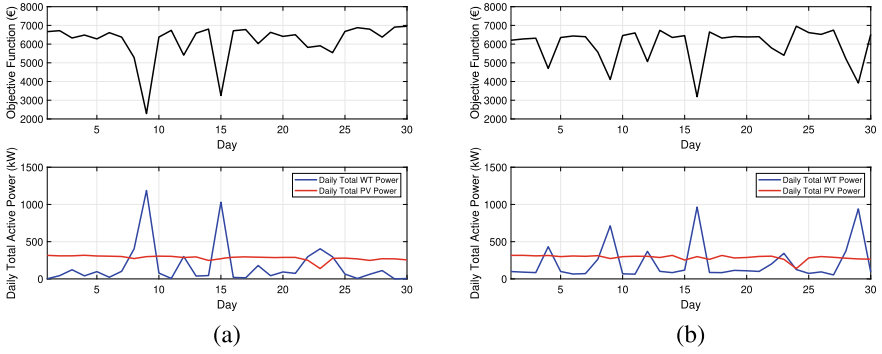


Fig. 11 Objective function and daily total wind and photovoltaic powers in Scenario 4 for real (a) and synthetic (b) data

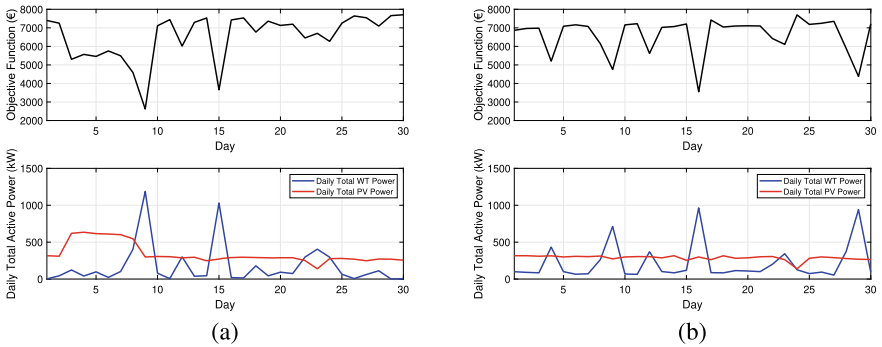


Fig. 12 Objective function and daily total wind and photovoltaic powers in Scenario 5 for real (a) and synthetic (b) data

Table 3 Objective function mean and standard deviation in € for real and simulated data and for each scenario studied

Scenario	Real		Simulation	
	Mean	Std.	Mean	Std.
Base	6733.85	1209.17	6394.39	1555.73
1	9922.21	4.4027	9926.66	14.09
2	6026.81	912.90	5508.90	653.48
3	5315.84	1051.08	5001.80	1296.02
4	6053.01	1133.97	5750.96	1425.78
5	6421.45	1314.70	6472.68	1182.20

installed renewable power is constant. This could be due to the fact that this power has a more constant feature during the sample period considered (which is August). Scenario 3 presents the lowest cost due to the highest power characterizing both renewable sources. Doubling the BESS capacity in Scenario 4 leads to a considerable benefit both for real and simulated data due to the most performing characteristic of the BESS, which allows better power management. Increasing the maximum charge and discharge powers in Scenario 5 leads to limited benefits compared to the base scenario. The results obtained from the simulated data are in line with those from the real data, and this indicates that the two Markov models correctly simulate the behavior of the wind and photovoltaic power sources. The best performance is obtained in Scenario 3 and this indicates that having a higher maximum power for both renewable sources is more important than having a double capacity of the BESS or increasing its maximum charge and discharge powers.

4.3 Conclusions

In this chapter, two stochastic modelizations referring to wind and photovoltaic powers are associated with the scheduling problem of an isolated hybrid microgrid. The problem consists in minimizing total operating costs while measuring how the uncertainty of renewable energy production influences the results. The ARMA model is used to forecast power demand, and wind and photovoltaic productions are modeled using Markov reward processes. The authors always juxtapose the real and simulated results to show the suitability of the models. Wind power strongly influences the objective function, and its low presence promptly increases the cost. Photovoltaic power has a more constant influence, determining the general trend of the objective function. The six scenarios studied provide a sensitivity analysis with respect to the parameters of the model and allow us to understand how they impact the results.

References

1. Alasali, F., Haben, S., Holderbaum, W.: Stochastic optimal energy management system for RTG cranes network using genetic algorithm and ensemble forecasts. *J. Energy Storage* **24**, 100759 (2019)
2. Alvarado-Barrios, L., del Nozal, A.R., Valerino, J.B., Vera, I.G., Martínez-Ramos, J.L.: Stochastic unit commitment in microgrids: influence of the load forecasting error and the availability of energy storage. *Renew. Energy* **146**, 2060–2069 (2020)
3. Chen, S.X., Gooi, H.B., Wang, M.: Sizing of energy storage for microgrids. *IEEE Trans. Smart Grid* **3**(1), 142–151 (2011)
4. Dagum, E.B.: *Analisi delle serie storiche: modellistica, previsione e scomposizione*. Springer Science & Business Media (2001)
5. Data, G.E.S., DISC, I.S.C.G.: Global modeling and assimilation office (GMAO), merra-2 (2015)

6. D'Amico, G., Petroni, F., Prattico, F.: Wind speed and energy forecasting at different time scales: a nonparametric approach. *Phys. A: Stat. Mech. Appl.* **406**, 59–66 (2014)
7. D'Amico, G., Petroni, F., Vergine, S.: An analysis of a storage system for a wind farm with ramp-rate limitation. *Energies* **14**(13), 4066 (2021)
8. Gabriel, L.G., Ruiz-Cruz, R., Zúñiga-Grajeda, V., Gurubel-Tun, K., Coronado-Mendoza, A., et al.: Optimizing the penetration of standalone microgrid, incorporating demand side management as a guiding principle. *Energy Rep.* **8**, 2712–2725 (2022)
9. Gagniuc, P.A.: *Markov Chains: from Theory to Implementation and Experimentation*. Wiley (2017)
10. Gerbaulet, C., von Hirschhausen, C., Kemfert, C., Lorenz, C., Oei, P.Y.: European electricity sector decarbonization under different levels of foresight. *Renew. Energy* **141**, 973–987 (2019)
11. Lasnier, F., Ang, T.: *Photovoltaic Engineering Handbook* (1990)
12. Lei, Y., Mullane, A., Lightbody, G., Yacamini, R.: Modeling of the wind turbine with a doubly fed induction generator for grid integration studies. *IEEE Trans. Energy Conv.* **21**(1), 257–264 (2006)
13. Marnieris, I.G., Biskas, P.N., Bakirtzis, A.G.: Stochastic and deterministic unit commitment considering uncertainty and variability reserves for high renewable integration. *Energies* **10**(1), 140 (2017)
14. Ramabhotla, S., Bayne, S., Giesselmann, M.: Operation and maintenance cost optimization in the grid connected mode of microgrid. In: 2016 IEEE Green Technologies Conference (GreenTech), pp. 56–61. IEEE (2016)
15. Rana, M.M., Romlie, M.F., Abdullah, M.F., Uddin, M., Sarkar, M.R.: A novel peak load shaving algorithm for isolated microgrid using hybrid PV-BESS system. *Energy* **234**, 121157 (2021)
16. Tapia, A., Tapia, G., Ostolaza, J.X., Saenz, J.R.: Modeling and control of a wind turbine driven doubly fed induction generator. *IEEE Trans. Energy Conv.* **18**(2), 194–204 (2003)
17. Vergine, S., Álvarez-Arroyo, C., D'Amico, G., Escaño, J.M., Alvarado-Barrios, L.: Optimal management of a hybrid and isolated microgrid in a random setting. *Energy Rep.* **8**, 9402–9419 (2022)
18. Yang, L., He, M., Zhang, J., Vittal, V.: Support-vector-machine-enhanced Markov model for short-term wind power forecast. *IEEE Trans. Sustain. Energy* **6**(3), 791–799 (2015)

Deep Learning Applied to Wind Power Forecasting: A Spatio-Temporal Approach



Rubén del Campo , Eloy Anguiano , Álvaro Romero ,
and José R. Dorronsoro 

Abstract The problem of wind energy production prediction has been one of the most prolific topics of study in the field of machine learning applied to the energy sector. Usually, these models receive data in tabular format. However, in this work we propose to solve the problem of predicting wind power like a spatio-temporal prediction problem as if it were an image or video analysis problem. On the one hand, energy production and the weather variables provided by Numerical Weather Prediction models (NWP) are time series, justifying the temporal treatment. On the other hand, NWP variables are provided in a regular grid format (in terms of latitude and longitude). Thus, the data are arranged as different meteorological variables in the shape of a grid, justifying the spatial treatment as if it were a low-resolution image, where the meteorological points are treated as pixels. For this reason, the goal of this article is to carry out an initial benchmark that compares the performance measured between different types of deep learning architectures that take advantage of these temporal and spatial features. The proposed architectures are CNN, LSTM, LSTM+CNN (Stacked), LSTM+CNN (Parallel), ConvLSTM and Vision Transformer.

Keywords Wind · Power · Forecasting · Time series · NWP · Deep learning · CNN · RNN · LSTM · ConvLSTM · ViT · Transformer

<https://www.iic.uam.es/>

R. del Campo (✉) · E. Anguiano · Á. Romero · J. R. Dorronsoro
Instituto de Ingeniería del Conocimiento, Francisco Tomás y Valiente, 11 Cantoblanco-UAM.
EPS Edificio B pt 5, 28049 Madrid, Spain
e-mail: ruben.delcampo@iic.uam.es

E. Anguiano
e-mail: eloy.anguiano@iic.uam.es

Á. Romero
e-mail: alvaro.romero@iic.uam.es

J. R. Dorronsoro
e-mail: jose.dorronsoro@iic.uam.es

1 Introduction

In recent years, the development of the electricity system toward a production system based on renewable energies has enabled society to move toward a more efficient and ecologically sustainable way of life. However, the generation of some of these energies cannot be controlled by humans, as their performance is inevitably linked to weather conditions. Both generating agents and the electricity system operators have always shown great interest in obtaining the best possible information on the production capacity that a station will be able to generate in consecutive hours or days. Being able to make this prediction reliably means great benefits, not only from an economic point of view, but also allows safeguarding the structural integrity of the entire transmission grid.

Given the great importance that this problem has gained in view of the large percentage that renewables already occupy in the energy mix, in recent decades a considerable amount of research has been carried out in the field of machine learning. However, classical machine learning techniques seem to have reached their maximum capacity for improvement with regard to the problem at hand. Thanks to the advances made both in the field of GPU computing, which make it possible to execute complex algorithms in reasonable times, and in the field of artificial intelligence and machine learning with the implementation of new, more efficient optimization algorithms, the state of the art has evolved into what is known as deep learning. However, all these techniques have in common that they use input data in a tabular form.

The objective of this work is to review several of the techniques that make up the state of the art of deep learning, applying them to the problem of predicting wind power production from a spatio-temporal point of view. Many of the implementations presented throughout this paper have in common the attempt to take advantage of the geospatial component of weather forecast data and also from the temporal information provided by both the production series and weather forecasts.

2 State of the Art

2.1 CNN

Inspired by human visual perception, CNNs [12] apply a series of filters where a kernel is slid through the bidimensional input image, thus extracting the most important features of the image. The fundamental operation here is the convolution which consists in applying to the input image a filter of smaller width and length and equal depth to that image, performing an element-by-element multiplication of each matrix and adding all the results so that we obtain a single value. In this way, it continues sliding in a sequential way going through all the input pixels several times, resulting in a new coded image with a new dimensionality.

Convolutional layers are often stacked multiple consecutive filters, so the dimensionality of the data increases dramatically. For this reason, it is very common to alternate convolutional layers with pooling layers. These layers are not trainable, but their only function is to compress the information always trying to preserve the most important features found by the filters of the convolutional layers.

2.2 RNN/LSTM

Deep neural networks have demonstrated a great ability to find complex relationships among variables. However, when we confront them with problems that have temporal information and future instants are related to past instants, we need a type of architecture that takes this feature into account. Recurrent networks (RNN) [13] were proposed with the aim of being able to encode this type of dependencies. In this type of network, each past input x_{t-1} is used in the next iteration when x_t arrives. The information propagated from previous states at time moment t is known as *hidden state*.

In [15], *bidirectional* recurrent networks were proposed. When they process time series data, the prediction is not made solely on the basis of the past instants following the time order, since in some problems the sequences are not always ordered in time. For example, in our particular case it is quite common for meteorology to have a certain time shift due to the complication of generating such forecasts. The units of this type of network are divided into two parts: one that processes the inputs in the same direction of time t and the other in the reverse direction.

Despite the great popularity of recurrent networks for solving time-dependent problems, they have proven to have major limitations when taking into account dependencies that extend widely in time. *Gradient vanishing* [8] occurs when taking the *backpropagation* so many steps back in time. For this reason, in [9] the LSTM architecture was proposed as a solution to this problem. The main idea behind this solution lies in the use of a memory that is propagated during different time instants [17]. The main parts of the LSTM are

- *Cell state* is responsible for propagating the information passed along the following time instants. We can understand that it gives the LSTM a kind of memory capacity. The information flow of the cell state is regulated by the different sigmoid gates that limit the information that passes through the different time instants.
- *Forget gate* has the function of establishing how much information from the input and the hidden state reaches the cell state.
- *Input gate* is in charge of deciding how much information from the cell output has to be propagated to the cell state, i.e., how much new information must be taken into account.
- *Output gate* is in charge of establishing the relevance of the new information calculated to be passed as hidden state to the next time instant.

2.3 Transformers

Transformers have become the state of the art of Natural Language Processing. This new architecture is used in this type of problems due to what is called self-attention [19]. This self-attention strategy is able to extract contextual information between the different points of a sequence of information, which fits very directly with text-related problems, since semantic and grammatical relations between words can be captured.

However, there are also adaptations of this architecture that have improved the state of the art in image classification problems, called *Vision Transformers* (ViT) [6]. In this approach, the image is divided into different groups of pixels as if it were a vector to be introduced into a *Transformer Encoder* with a fictitious parameterized vector added (usually known as CLS Token), so that the output of the encoding of that vector serves as input to the prediction of a series of fully connected layers. This trick of the CLS parameter is commonly used in document classification problems, since it allows us to summarize all the contextual information of the sequence (in this case the image) in a single vector that can be processed later.

Therefore, for our particular use case, we will use this approach as one of the models to consider for that spatial, but not temporal, approach to solving the problem in question.

2.4 Metrics

In this research, a wide variety of neural network architectures have been developed and therefore, each of them has an associated prediction error. This error must be quantifiable and ideally interpretable so that we can compare networks with each other and establish which one is better. Within the field of knowledge of machine learning, this problem falls into the category of *regression*, which is made up of those problems in which the objective is to achieve a single continuous numerical result. Classically, the metrics used in this type of problem are given in absolute terms. However, in the case of our problem, it is obvious that for a wind farm with a high maximum productive capacity, called *installed power*, these measurements will be much higher than for other farms with a lower installed power. It is for this reason that the previous measurements are always normalized by the installed power of the wind farm. This feature makes this type of metrics make the *target* insensitive to changes in installed power over time.

– Normalized Mean Absolute Error:

$$\text{NMAE}(Y, \hat{Y}) = \frac{1}{n} \sum \frac{|y - \hat{y}|}{\text{InstalledPower}} \quad (1)$$

– Normalized Mean Squared Error:

$$\text{NMSE}(Y, \hat{Y}) = \frac{1}{n} \sum \left(\frac{y - \hat{y}}{\text{InstalledPower}} \right)^2 \quad (2)$$

where $y \in Y$ are the observations given and $\hat{y} \in \hat{Y}$ are the predictions made by the model for which its performance is to be measured.

On the other hand, in the search for measures that are independent of the data domain, we will use measures such as

– Coefficient of determination (R^2 Score):

$$R^2(Y, \hat{Y}) = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2} \quad (3)$$

where \bar{y} is the mean of all observations.

– Explained Variance:

$$\text{Exp_Var}(Y, \hat{Y}) = 1 - \frac{\text{Var}[y - \hat{y}]}{\text{Var}[y]}. \quad (4)$$

2.5 Related Work

The problem of renewable production prediction has been widely studied in the field of machine learning. All kinds of papers have been published on the subject using a wide variety of techniques and algorithms.

In recent years, following the popularization of deep learning, we have started to see some research using this type of novel techniques. Focusing on publications that use data similar to those available for this work, we find some research such as [1], where they simply extend the idea of the multilayer perceptron by increasing the number of layers and neurons per layer. On the other hand, in [18] they continued with the idea by increasing the number of layers even more.

Architectures based on LSTM have also begun to be used in the renewable prediction problem, as they are particularly well suited to deal with time series. For example, in [7] they use an LSTM to predict the production of a photovoltaic farm using meteorological data.

One of the first publications that began to try to exploit the geospatial component of meteorological data was [5]. In it, a three-dimensional dataset is constructed where the first dimension is the meteorological variables, the second dimension is the latitude and the third is the longitude. It showed that this type of approach taking into account the location of each weather point brought new information to the model that the other traditional methods were unable to appreciate.

Extending this idea of encoding the input data by means of CNNs, some researches proposed that the output of the last convolutional, instead of attaching it to a fully connected layer, should be attached to an LSTM layer such as [3, 14]. We will call this type of architecture *CNN-LSTM Stacked*. Another approach of combining CNN with LSTM is found in [11], where instead of connecting the output of one network to the input of another, what they propose is that both process the same input in parallel, concatenating their output to pass it to a *fully connected* layer. We will call this type of architecture *CNN-LSTM Parallel*.

Following the approach of trying to encode both temporal and geospatial information of meteorological data, the idea of ConvLSTM [16] arises. In essence, what is proposed is an architecture similar to LSTMs except that the input data, instead of being one-dimensional data on which tensor multiplication operations are applied, are two-dimensional data on which the convolution operation is applied using convolutional layers. That is the reason why they have been used in video analysis problems.

3 Methodology

3.1 Data Sources

To solve the problem of renewable production forecasting, we need two sources of data: productions and weather forecasts.

- Production data: We have used the series of actual productions measured in a wind farm located in Spain for every hour of the years 2019, 2020 and 2021.
- Meteorological data: Meteorological forecasts made with Numerical Weather Prediction models (NWP) have been obtained for 24 horizons corresponding to the years 2019, 2020 and 2021. For all these time instants, meteorological variables have been obtained forming a 9×9 regular grid around the location of the park. We have used the following variables: u and v wind components at 10 and 100 m altitude with its module (V_{10} and V_{100}), the temperature at 10 m altitude (T) and the surface pressure (p). So we can understand this grid of points as if it were an “image” where each meteorological point would be a pixel and each variable would be a channel of it.

3.2 Exploratory Analysis

It has been previously emphasized throughout this work that one of the purposes of this work is to try to take advantage of the spatial characteristics of the weather forecast data. In order to justify this decision, an analysis has been carried out, which is shown in Fig. 1, where the Pearson correlation of each of the variables in the grid

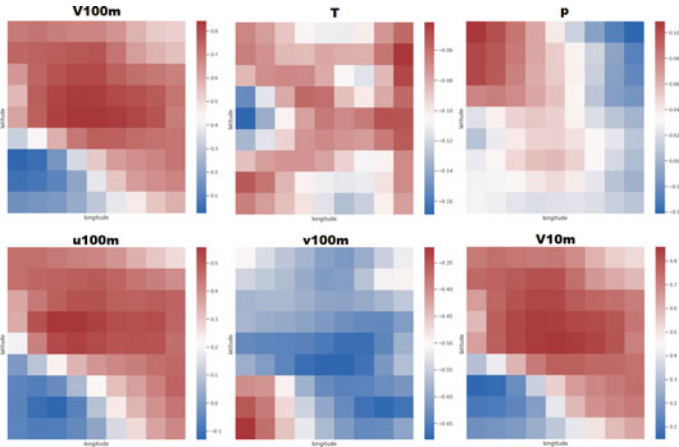


Fig. 1 Correlation of the meteorological variables with the wind farm production in each one of the coordinates

is measured with the wind production target. In this way, we can clearly see how, depending on the variable, there will be certain geographical areas that are much more relevant than others. The winds that have more impact on the target are those measured in points of the northwest-southeast diagonal. For this reason, neural network architectures that take into consideration this type of geospatial characteristics such as CNN or ViT seem more than adequate to solve this problem. In other variables such as pressure and temperature, although their correlation also occurs in specific areas, this value is very small. Nevertheless, it is convenient to keep in mind that the Pearson correlation only finds linear relationships, so we will not exclude these variables from the study.

3.3 Experimentation

In order to arrive at the neural network architectures that will be described later, we have carried out an exhaustive experimentation process. During this process, multiple variations on the networks have been tested, varying their structure (number of units, shape of the layers, lags passed, etc.), their hyperparameters (dropout, bidirectionality, hidden size, etc.) or even parameters affecting the training (learning rate, batch size, etc.).

4 Implemented Neural Architectures

4.1 LSTM

The first of the implemented architectures is the one based on LSTM networks, which, as previously explained, are especially appropriate for time series-related problems. We note that this network simply receives the data records in one-dimensional form by applying time delays of the received data. Furthermore, they have been used in their bidirectional mode which is capable of analyzing the data sequence in both temporal directions.

In this problem where the input data are meteorological predictions with their associated error, bidirectionality can be an improvement since it can happen that the predictions received are correct in the magnitude of the meteorological variables, but these are wrongly ordered in time. The bidirectional network, by evaluating the sequence in both directions, can mitigate this effect that could worsen the training capacity of our network.

4.2 CNN

It has been justified during the previous sections that considering the spatial component in meteorological data by treating our data in a two-dimensional form can help our models to find relationships that would be difficult to appreciate using a tabular format. For this reason, it is proposed to use a CNN as shown in the CNN parts of Fig. 2.

For this architecture, we start changing input data shape to a three-dimensional representation (D variables $\times L$ latitudes $\times M$ longitudes). In the first block, a convolutional layer is used applying *zero-padding* so that our network can take into consideration the information presented by the variables corresponding to the points located at the ends of the grid of points. Subsequently, an *average-pooling* layer is used to reduce the dimensionality of the input data. This is followed by a series of identical blocks consisting of two convolutional layers followed by batch normalization layers in charge of reducing the *internal covariate shift* and favoring a fast convergence during training [10]. Although the *batch normalization* layer may sometimes make it unnecessary to apply *dropout* layers [2], it has been seen that in this case they are convenient, since they help to reduce the *overfitting* that can be caused in a complex network such as the one proposed. Finally, before passing the data through *fully connected* layers, an *average-pooling* layer is applied, also with the aim of reducing the data dimension, which at that moment is very high after having applied a large number of convolutional filters. The activation function used is *ELU* [4] that has proved to be one of the most reliable activation functions in the state of the art.

4.3 Combinations

As discussed above, CNNs are particularly suitable for capturing the spatial information provided by the input data, while LSTMs are particularly suitable for capturing the temporal information of the input data. For this reason, it has been decided to apply combinations between the CNN and LSTM networks detailed in the previous subsections:

- CNN+LSTM Stacked: in this solution, the networks are connected in such a way that the output of the CNN is immediately directed to the input of the LSTM network, as can be seen in the first Fig. 2. It should be noted that the CNN output has to be flattened so that it can be treated by the LSTM input.
- CNN+LSTM Parallel: this solution is similar to the one previously explained except that, as shown in the second Fig. 2, in this case the input of the neural network is passed equally to the CNN and the LSTM, except that the former will receive the data in three-dimensional form and the latter will not. The output of both networks will be flattened and concatenated into a single one-dimensional vector.

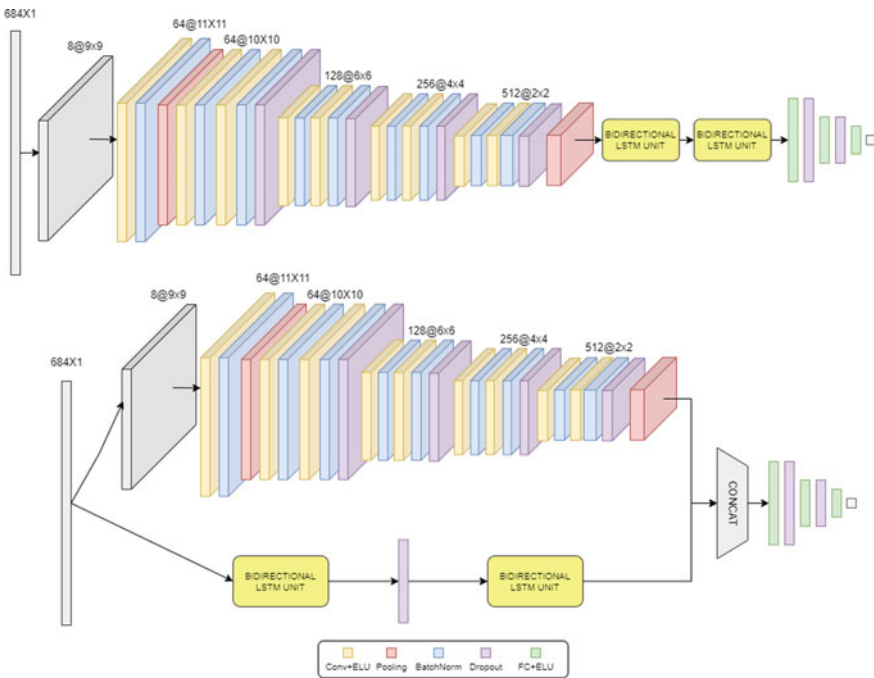


Fig. 2 Implemented CNN-LSTM parallel and stacked neural network architectures

4.4 ConvLSTM

The ConvLSTM presents an architecture that, instead of treating the CNN and the LSTM as independent networks, merges the underlying idea of both of them trying to achieve a conceptually more complete architecture that by itself is able to process both spatial and temporal information of meteorological data.

We have used two ConvLSTM units separated by dropout and *batch normalization* layers that help to maintain a stable training avoiding *overfitting*. In the first unit, the convolutional layers use a 5×5 filter, while in the second unit they use a 3×3 . In this way, the former can focus on more general features of the input data while the latter can focus on more specific ones.

4.5 ViT

As mentioned above, the ViT model architecture groups chunks of pixels disjointly. For the current implementation, we have grouped the pixels into three vectors to which a last parameterizable vector (called CLS) is added. To these four vectors, we add a vector representing their positional encoding to move the projection to the latent space of the next dense layer of the network. After encoding this information through a Transformer Encoder, the output corresponding to the position of the CLS vector is used for a fully connected network to give us the prediction value we are looking for (Fig. 3).

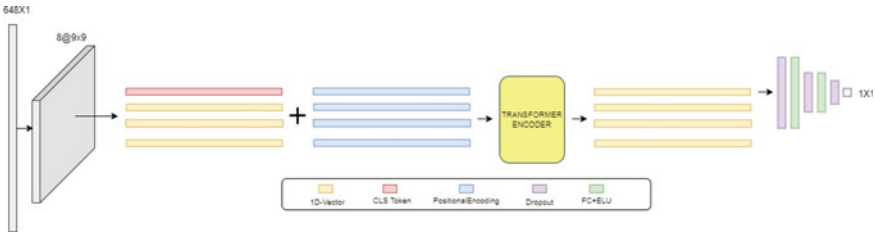


Fig. 3 Implemented ViT neural network architecture

Table 1 Test metrics of every model averaged over the first 24 prediction horizons

	NMAE	NMSE	R ²	VarScore
CNN	0.084	0.019	0.778	0.779
LSTM	0.080	0.017	0.802	0.801
CNN+LSTM Parallel	0.082	0.018	0.788	0.791
CNN+LSTM Stacked	0.082	0.018	0.795	0.796
ConvLSTM	0.081	0.018	0.788	0.789
ViT	0.080	0.017	0.799	0.802
Ridge	0.104	0.020	0.762	0.765
MLP	0.103	0.020	0.765	0.800
SVM	0.092	0.018	0.791	0.797

5 Results

5.1 Benchmark Results

During this research, we have carried out a multitude of executions with the aim of arriving at neural network architectures that offer the best performance in accordance with the metrics we have previously explained. All of them have been trained and tested under the same conditions: we have used the 2019 and 2020 data for training and validation and the 2021 data for testing.

In Table 1, we can see the comparison among all the implemented architectures. We can see that the results obtained in the comparison do not differ much between the different proposed solutions. In spite of this, the neural network architectures that have finally shown the best performance in all metrics are LSTM and ViT. LSTM, thanks to its bidirectional characteristic, is able to minimize the intrinsic temporal error in weather predictions. ViT also demonstrates excellent performance, which further validates the idea of using spatially-aware network architectures to address this problem. However, although this problem has both a spatial and a temporal component, the combinations between CNNs (spatial) and LSTMs (temporal) do not end up being the best performers. This leaves the door open for further investigation of new combination techniques.

Nevertheless, it has been proven that the deep learning methods proposed during this research are a considerable improvement over other classical machine learning methods applied to regression problems, such as Ridge, MLP and SVM.

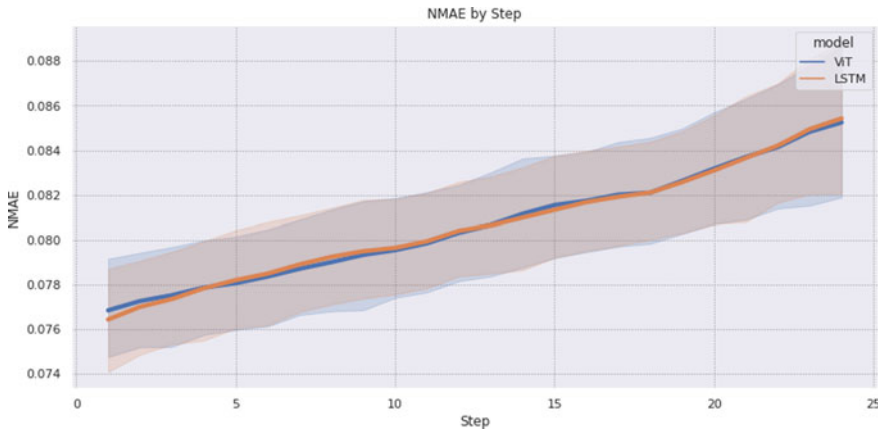


Fig. 4 Comparison of NMAE by prediction step between LSTM and ViT with its confidence interval

5.2 Degradation Study

Since the weather-based forecasting problem is based on source data that are also forecasts, models may have different performances at different forecast horizons. Therefore, comparing models solely by a metric averaged over such prediction horizons seems to us to lack all the analysis that a model's performance might receive, since one model might perform better for near-prediction horizons and another might perform better at far horizons (Fig. 4).

This is why this part of the study focuses on comparing the performance of the different models over the prediction horizons and the error distributions of the models.

6 Conclusion

The prediction of wind energy production is a complex task, as it is influenced by meteorological and geographical factors. For this reason, the aim of this paper is to provide a new perspective on the problem by applying the latest Deep Learning techniques that have been classically used in other disciplines such as NLP or Computer Vision.

After an extensive research on the state-of-the-art techniques in neural network architectures applied to time series and spatial data together with their combinations and variants, we have proposed many architectures to solve the problem. These approaches have been CNN, LSTM, LSTM+CNN (Stacked), LSTM+CNN (Parallel), ConvLSTM and Vision Transformer.

Despite the high computational cost of training this type of architectures based on deep neural networks, an exhaustive benchmark has been elaborated in order to compare, applying different metrics, the performance of the proposed architectures.

All the implementations have obtained results of similar quality, although the LSTM and ViT have stood out, obtaining an NMAE of 0.080. Thus, with the former, the temporal approach to the problem has been justified, while with the latter, the spatial approach has been justified. Moreover, all the proposed solutions have been shown to be a clear improvement over classical machine learning models using tabular data.

References

1. Badrinath Krishna, V., Wadman, W., Kim, Y.: Nowcasting: accurate and precise short-term wind power prediction using hyperlocal wind forecasts (2018)
2. Baldi, P., Sadowski, P.J.: Understanding dropout. In: Burges, C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K. (eds.) *Advances in Neural Information Processing Systems*, vol. 26. Curran Associates, Inc. (2013)
3. Chen, Y., Wang, Y., Dong, Z., Su, J., Han, Z., Zhou, D., Zhao, Y., Bao, Y.: 2-d regional short-term wind speed forecast based on CNN-LSTM deep learning model. *Energy Conv. Manag.* **244**, 114451 (2021)
4. Clevert, D.A., Unterthiner, T., Hochreiter, S.: Fast and accurate deep network learning by exponential linear units (elus) (2015)
5. Díaz-Vico, D., Torres-Barrán, A., Omari, A., Dorransoro, J.R.: Deep neural networks for wind and solar energy prediction. *Neural Process. Lett.* **46**(3), 829–844 (2017)
6. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16×16 words: transformers for image recognition at scale (2020)
7. Gao, M., Li, J., Hong, F., Long, D.: Short-term forecasting of power production in a large-scale photovoltaic plant based on LSTM. *Appl. Sci.* **9**, 3192 (2019)
8. Hochreiter, S.: The vanishing gradient problem during learning recurrent neural nets and problem solutions. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* **6**, 107–116 (1998)
9. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**, 1735–80 (1997)
10. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Bach, F., Blei, D. (eds.) *Proceedings of the 32nd International Conference on Machine Learning. Proceedings of Machine Learning Research*, vol. 37, pp. 448–456. PMLR, Lille, France (2015)
11. Karim, F., Majumdar, S., Darabi, H., Chen, S.: LSTM fully convolutional networks for time series classification. *IEEE Access* **6**, 1662–1669 (2018)
12. Lecun, Y., Haffner, P., Bengio, Y.: Object recognition with gradient-based learning (2000)
13. Lipton, Z.: A critical review of recurrent neural networks for sequence learning (2015)
14. Ren, J., Yu, Z., Gao, G., Yu, G., Yu, J.: A CNN-LSTM-lightgbm based short-term wind power prediction method based on attention mechanism. *Energy Rep. (iCPE 2021-The 2nd International Conference on Power Engineering)* **8**, 437–443 (2022)
15. Schuster, M., Paliwal, K.: Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **45**, 2673–2681 (1997)
16. Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.k., Woo, W.c.: Convolutional lstm network: a machine learning approach for precipitation nowcasting. In: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R. (eds.) *Advances in Neural Information Processing Systems*, vol. 28. Curran Associates, Inc. (2015)
17. Staudemeyer, R., Morris, E.: Understanding LSTM—A tutorial into long short-term memory recurrent neural networks (2019)
18. Torres, J., Aguilar, R., Zúñiga, K.: Deep learning to predict the generation of a wind farm. *J. Renew. Sustain. Energy* **10** (2018)
19. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need (2017)

Forecasting Natural Gas Prices with Spatio-Temporal Copula-Based Time Series Models



Sven Pappert and Antonia Arsova

Abstract In this work, we model and forecast commodity price time series using multivariate copula-based time series models. In particular, we consider the Gaussian copula, the t-copula and D-vine copulas to model both the contemporaneous and the temporal dependencies of the series. We focus on daily short-term gas, coal, oil and carbon emission futures in the period from March 2010 to February 2021. Further, we examine how the copula-based model gives rise to non-elliptical conditional probabilistic forecasts. During volatile periods, the conditional forecast density can become bimodal. This can be problematic when not only a distributional but also a mean forecast is desirable. Because for such probabilistic forecasts, it is not clear what represents a good mean forecast. A possible solution is to augment the forecasts by an artificial neural network that predicts the best quantile to use as mean forecast based on past values and past best quantiles. In a forecasting study, the predictive performances of the models are examined and compared with a benchmark. The distributional forecasts are examined by the CRPS. The mean forecasts are examined by the RMSE. The copula-based models are competitive with benchmark models.

Keywords Copula · Dependence modeling · Deep neural network · Energy forecasting · Probabilistic forecasting

1 Introduction

Modeling and forecasting commodity prices is important for trading and political decision-making. In this work, we focus on modeling and forecasting short-term natural gas future prices jointly with related commodity prices. We model the time

S. Pappert (✉)
Department of Statistics, TU Dortmund University, Vogelpothsweg 78, 44227 Dortmund,
Germany
e-mail: pappert@statistik.tu-dortmund.de

A. Arsova
Department of Statistics, TU Dortmund University, Vogelpothsweg 78, 44227 Dortmund,
Germany & RWI - Leibniz Institute for Economic Research, Hohenzollernstr. 1-3, 45128 Essen,
Germany
e-mail: arsova@statistik.tu-dortmund.de

series jointly to exploit the additional information carried by their mutual dependence. To this aim, we employ spatio-temporal copula-based time series models.

Copulas are popular choices to model the cross-sectional dependence in multivariate time series with the copula-GARCH approach in financial markets [16, 18], as well as in energy markets [2, 7]. In the copula-GARCH approach, the temporal dependence of each time series is modeled by univariate time series models, such as ARMA-GARCH. The cross-sectional dependence structure of the time series can be captured by the copula of the standardized residuals of the univariate time series model. However, such models are only able to allow for flexible dependence structures in the cross-sectional dimension. The mean process is modeled linearly.

On the other hand, temporal copula modeling (or ‘copula-based time series modeling’) as well as spatio-temporal copula time series modeling offers an alternative to classical linear time series approaches. The models are able to flexibly model contemporaneous dependencies across multiple time series, as well as temporal dependencies. There is emerging literature on the topic. Chen and Fan [9] investigate the estimation of univariate copula-based semiparametric time series models. The authors provide conditions for β -mixing and prove consistency as well as asymptotic normality of the model parameters. Beare [4] further investigates mixing conditions. Smith et al. [24] decompose the serial dependence of intraday electricity load using pair copula constructions. Simard and Rémillard [22] investigate the forecasting performance of the spatio-temporal t-copula, dependent on the strength and structure of the dependence as well as the marginal distributions. Loazia-Maya et al. [19] explore heteroskedasticity modeling using copulas. Beare and Seo [5] as well as Nagler et al. [20] examine spatio-temporal vine copula models.

Examples where flexible dependence modeling can be important are the following. The cross-sectional dependence between international stock markets can be asymmetric with dominant lower tail dependence, indicating the phenomenon of contagion in financial markets [16]. With regard to energy markets, it was found by Aloui et al. [2] that crude oil and gas markets comove rather during bullish periods, thereby also displaying an asymmetric cross-sectional dependence structure. A concise example for a possible emergence of non-linear dependence structures in the temporal domain and hence requiring sophisticated dependence modeling is given in the introduction of the work by Beare [4]: The continuous growth of financial time series contrasted with their sudden and quick decrease represents an asymmetric temporal relation. Thus, cross-sectional as well as temporal dependence modeling can be important in many fields.

In this paper, we explore the performance of spatio-temporal copula models for modeling energy commodity prices. The basic idea underlying spatio-temporal time series modeling with copulas is a decomposition of the joint distribution. Using Sklar’s theorem [23], the joint distribution of consecutive observations is decomposed into dependence and marginal structure, $F_{\mathbf{X}_t, \mathbf{X}_{t-1}}(\mathbf{a}, \mathbf{b}) = C[F_{\mathbf{X}_t}(\mathbf{a}), F_{\mathbf{X}_{t-1}}(\mathbf{b})]$. Various copula specifications can be employed. In this work, the t-copula (discussed in detail in [11]) and the Gaussian copula are considered. Vine copula models [1, 10] are also considered. Spatio-temporal forecasting with the t-copula was examined in

[22]. The spatio-temporal vine copula modeling of multivariate time series is, among others, explored in [5, 20, 24].

By employing conditional copulas, the models can be used for forecasting. The resulting probabilistic forecasts can be non-elliptical. It is not obvious what constitutes a sensible mean forecast in this case. The expectation value is a sensible mean forecast for elliptical or almost elliptical probabilistic forecasts. For non-elliptical forecasts, e.g. a bimodal probabilistic forecast, the expectation value predicts points that are unlikely. One possible solution to this problem would be to take the mode of the probabilistic forecast as the mean forecast. Another possibility is to augment the forecasting procedure by an artificial neural network (ANN).¹ The ANN predicts which quantile of the probabilistic forecast is optimal (in MSE sense) as mean forecast. The inputs of the ANN are past values of the times series and the last optimal quantiles. One advantage of the ANN-augmented forecast is that the ANN can be estimated and used for prediction completely independent from the probabilistic time series model used for forecasting. In this approach, the ANN mean forecasts are also equipped with an underlying probabilistic distribution, enabling the calculation of confidence intervals and other distributional properties. The models' performances are examined in a forecasting study. The ARMAX-AVTGARCH model, closely related to the model from [6], which was shown to outperform other popular models, is considered as a benchmark. We find that the spatio-temporal copula models with ANN-augmented mean forecasts are indeed competitive for natural gas and related commodity price forecasting.

The main contributions of this paper are two-fold. The first contribution is the application-oriented exploration of spatio-temporal copula-based time series models for predicting energy commodity prices. We evaluate the performance of the models in a forecasting study and find that they perform well. The second contribution is the methodological exploration of optimal mean forecasting from non-elliptical probabilistic forecasts by means of ANNs.

In the next Section, Sect. 2, the data used in this work is introduced briefly. Section 3 describes the statistical methods used in this work. The empirical results of the forecasting study are presented in Sect. 4, while Sect. 5 summarizes the results and outlines some avenues for future research.

2 Data Description

In this work, we analyze month-ahead natural gas futures (NGas) from the Netherlands (TTF Hub). The related commodities used for modeling are short-term carbon emission futures (CEF), short-term brent oil futures (oil) and short-term coal futures (coal). The analyzed time series are comprised of daily observations. The observation period spans from March 2010 to February 2021. In total, each time series comprises 2861 observations. Missing values, which occur especially during the holidays, are

¹ We refer to [14] for a concise introduction.

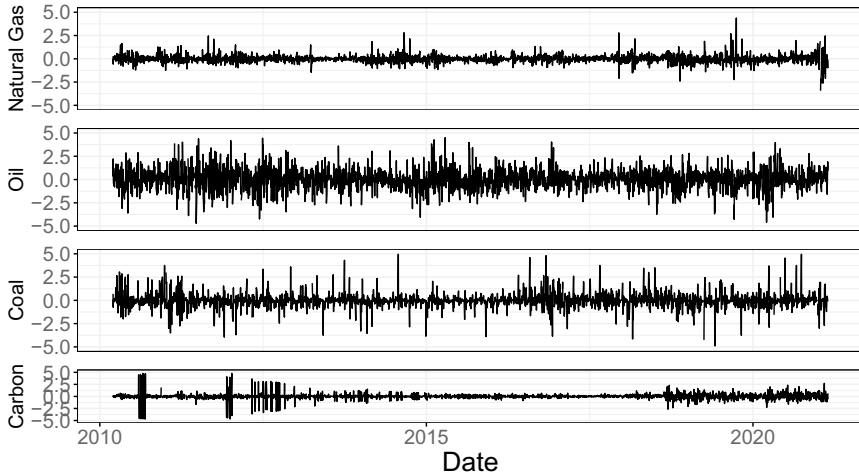


Fig. 1 First differences of the respective commodity price time series

trivially imputed as the last known value. The original time series are non-stationary. To obtain stationarity, which is necessary for the methods used in this report, the time series are differenced once. The differenced time series are displayed in Fig. 1. The hypothesis of non-stationarity in the first differences is rejected by the Dickey-Fuller test at the level $\alpha = 0.01$ for all time series.²

3 Statistical Methods

This section describes the methods used in this study. First copulas and related notions are introduced in Sect. 3.1. The copula specifications used in the analysis are presented. In particular vine copulas are presented in Sect. 3.2. The application of copulas to time series modeling is discussed in Sect. 3.3. The emergence of non-elliptical probabilistic forecasts is examined with regard to the t-copula in Sect. 3.4. It is shown how the t-copula may be used to model conditional heteroskedasticity. We note that Loazia-Maya et al. [19] also explored heteroskedasticity modeling based, among others, on the t-copula. The need for new mean forecast methods is presented.

3.1 Copulas

Copulas are distribution functions on the unit cube with uniform marginals:

² Results are omitted for brevity, but available upon request.

$$C : [0, 1]^d \rightarrow [0, 1]. \tag{1}$$

Copulas gain their relevance by Sklar’s theorem [23]. It states that every multivariate distribution can be decomposed into a copula and marginal distributions. Let X_1, \dots, X_d be real-valued random variables with joint distribution F_{X_1, \dots, X_d} and marginal distributions F_{X_1}, \dots, F_{X_d} . Then it holds that there exists a copula C such that

$$F_{X_1, \dots, X_d}(x_1, \dots, x_d) = C_{U_1, \dots, U_d} [F_{X_1}(x_1), \dots, F_{X_d}(x_d)], \tag{2}$$

where $(U_1, \dots, U_d) := (F_{X_1}(X_1), \dots, F_{X_d}(X_d))$. In the following, the indices of the copula will be dropped. If the random variables X_1, \dots, X_d are continuous, then the decomposition is unique [17]. The pseudo-observations $(u_1, \dots, u_d) := F_{X_1}(x_1), \dots, F_{X_d}(x_d)$ are, by virtue of the probability integral transformation, realizations from a uniform distribution, $U_i = F_{X_i}(X_i) \sim U[0, 1], i \in \{1, \dots, d\}$ [3]. This permits the copula to be interpreted as the dependence structure of the random variables X_1, \dots, X_d . The copula density c , which couples the joint density f_{X_1, \dots, X_d} and marginal densities f_{X_1}, \dots, f_{X_d} , can be derived directly from Eq. 2 by taking derivatives (if the copula is differentiable) :

$$f_{X_1, \dots, X_d}(x_1, \dots, x_d) = c [F_{X_1}(x_1), \dots, F_{X_d}(x_d)] f_{X_1}(x_1) \dots f_{X_d}(x_d), \tag{3}$$

$$c[u_1, \dots, u_d] = \frac{\partial^d C[u_1, \dots, u_d]}{\partial u_1 \dots \partial u_d}. \tag{4}$$

The copula density is important for estimation via maximum likelihood as well as for the visualization of dependence structures. In this paper, the copula density will also be used to introduce the notion of vine copula models. Another important notion for dependence modeling is the conditional copula of U_1, \dots, U_i given U_{i+1}, \dots, U_d , along with the respective conditional copula density. The conditional copula (density) can also be derived from Eq. 2. It is given as [22],

$$C[u_1, \dots, u_i | u_{i+1}, \dots, u_d] = \frac{\partial_{u_{i+1}} \dots \partial_{u_d} C[u_1, \dots, u_d]}{c[u_{i+1}, \dots, u_d]}, \tag{5}$$

$$c[u_1, \dots, u_i | u_{i+1}, \dots, u_d] = \frac{c[u_1, \dots, u_d]}{c[u_{i+1}, \dots, u_d]}. \tag{6}$$

Conditional copulas are especially relevant for conditional time series models as presented in this paper. The relation between the conditional density and the copula is as follows:

$$f_{X_1, \dots, X_i | X_{i+1}, \dots, X_d}(x_1, \dots, x_i | x_{i+1}, \dots, x_d) = \frac{c[F_{X_1}(x_1), \dots, F_{X_d}(x_d)]}{c[F_{X_{i+1}}(x_{i+1}), \dots, F_{X_d}(x_d)]} \times f_{X_1}(x_1) \dots f_{X_i}(x_i). \tag{7}$$

The copula approach to multivariate modeling allows for separate modeling of marginal properties and dependence structure. This feature renders the approach far more flexible than standard multivariate modeling. Joint distributions such as the multivariate normal or Students t-distribution restrict the choice of marginal distributions. In the copula approach, marginal distributions can be arbitrary. Also the dependence structure of random variables can exhibit various features that have to be accounted for by choosing an appropriate copula specification. In this work, the Gaussian, Clayton, Gumbel and t-copula are utilized. In the following, they will be introduced briefly as the joint distribution of random variables $U_i \sim U[0, 1], i \in \{1, \dots, d\}$. The Gaussian copula is a popular choice for modeling linear dependence structures. The Gaussian copula is constructed by extracting the dependence structure of the multivariate normal distribution and filtering the marginal influences [17],

$$C^{\text{Gaussian}}[u_1, \dots, u_d] = \Phi_{\Sigma}[\phi^{-1}(u_1), \dots, \phi^{-1}(u_d)], \tag{8}$$

where ϕ is the cumulative distribution function of the standard normal distribution and Φ_{Σ} is the d -variate cumulative distribution function of the normal distribution with correlation matrix Σ . The correlation matrix $\Sigma \in [-1, 1]^{d \times d}$ contains $\frac{d(d-1)}{2}$ dependence parameters, $\rho_1, \dots, \rho_{\frac{d(d-1)}{2}}$, governing the linear dependencies among the random variables U_1, \dots, U_d . The density of a bivariate Gaussian copula with dependence parameter $\rho = 0.4$ is displayed in the upper left panel of Fig. 2. The density only displays a linear relation between the variables. Similar to recovering linear dependence structures from the multivariate normal distribution, heavy-tailed dependence structures can be recovered from the multivariate t-distribution using the t-copula

$$C^t[u_1, \dots, u_d] = t_{\Sigma, \nu}[t_{\nu}^{-1}(u_1), \dots, t_{\nu}^{-1}(u_d)], \tag{9}$$

where t_{ν} is the cumulative distribution function of the t-distribution with degree of freedom ν and where $t_{\Sigma, \nu}$ is the cumulative distribution function of the multivariate t-distribution with correlation matrix Σ and degree of freedom ν . Incorporating the degree of freedom $\nu \in (2, \infty)$ permits heavy-tailed dependence structures. The heavy-tailedness can be interpreted as a higher probability of extreme events coinciding. A lower degree of freedom ν implies heavier tails. The density of a bivariate t-copula with dependence parameter $\rho = 0.4$ and degree of freedom $\nu = 4$ is displayed in the upper right panel of Fig. 2. The density displays the linear relation between the variables as well as the coincidence of extreme events. Another class of dependence structures can be described as asymmetric dependence structures. Two relevant copulas are the Gumbel and Clayton copula. The Gumbel copula exhibits dominant upper tail dependence while the Clayton copula exhibits dominant lower tail dependence. Their bivariate densities are displayed in the lower left, respectively

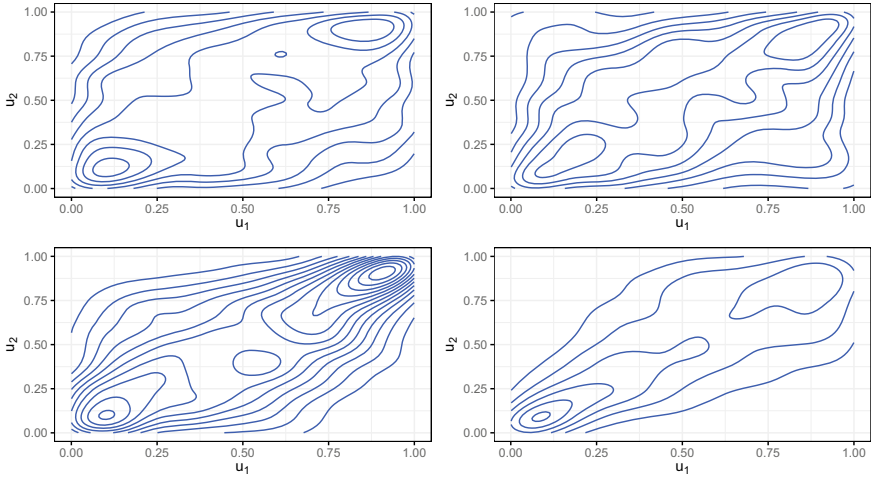


Fig. 2 Simulated density plots of four different two-dimensional copula specifications. The upper left plot shows the Gaussian copula density with dependence parameter set to $\rho = 0.4$. Upper right shows the t-copula density with dependence parameter $\rho = 0.4$ and degree of freedom $\nu = 4$. The lower left plot displays the Gumbel copula density with dependence parameter $\theta = 2$. The lower right plot displays the Clayton copula density with dependence parameter $\theta = 2$. All plots were created from simulations with 2000 samples

lower right panel of Fig. 2. Both copulas are part of the Archimedean copula family. They are constructed as $C[u_1, \dots, u_d] = \Psi^{-1}(\Psi(u_1) + \dots + \Psi(u_d))$ with a suitable generator function Ψ [12, 15]. The generators for the Gumbel, respectively Clayton copula, are given by

$$\Psi^{\text{Clayton}}(t) = (1 + t)^{-\frac{1}{\theta}}, \tag{10}$$

$$\Psi^{\text{Gumbel}}(t) = e^{-t^{\frac{1}{\theta}}}. \tag{11}$$

The dominant lower tail dependence of the Clayton copula can be interpreted as lower tail events coinciding more often than upper tail events and vice versa for the dominant upper tail dependence of the Gumbel copula. An even more flexible copula model is the vine copula model. Vine copula models will be explained next.

3.2 Vine Copulas

Vine copulas are special pair copula constructions. The idea of pair copula constructions is to decompose a d -variate dependence structure into a product of bivariate copulas. The joint density of d random variables can, by virtue of the law of total

probability, be decomposed into a product of conditional densities. Using the relation between conditional densities and copula densities (Eq. 7), one possible decomposition of a d -variate copula density can be derived as [1, 10]

$$c[u_1, \dots, u_d] = \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c[u_i, u_{i+j} | u_{i+1}, \dots, u_{j-1}], \quad (12)$$

where indices of the copula densities were dropped again. Furthermore, we neglect the influence of conditioning on the copula density. In the literature, this model then is called simplified vine [20]. The decomposition is not unique. The decomposition in Eq. 12 is called drawable vine (D-vine). The unconditional copulas in the product all capture the dependence structure of neighboring variables, e.g. $c[u_i, u_{i+1}]$, $c[u_{i+1}, u_{i+2}]$ and so forth. In a graphical representation, the connection between the variables resembles a straight line, hence the name D-vine.³ The vine copula approach allows for flexible dependence modeling. It is advantageous in contexts where the bivariate dependence structures between variables can take different forms, e.g. the dependence structure between variables U_1 and U_2 may be linear whereas the dependence structure between U_2 and U_3 may be heavy-tailed and so forth. Vine copula models can be estimated by maximum likelihood; we refer to [1] for details.

3.3 Modeling Time Series with Spatio-Temporal Copulas

So far, our discussion on copulas referred to modeling dependence in a general multivariate setting. In this subsection, the copula-based time series models are reviewed and summarized. Further, it is explained how these models can be used for forecasting. First, the temporal copula modeling of univariate time series (see, for example, [4, 9]) is introduced. Next, the copula-based modeling of multivariate time series, the spatio-temporal copula modeling, is discussed. The exposition on the spatio-temporal t-copula is based on [22], while for vine copula modeling we refer to [5, 8, 20, 24]. It will be examined how conditional temporal copula models offer an alternative approach to conditional heteroskedasticity modeling, which was also explored in [19]. The emergence of non-elliptical conditional probabilistic forecasts will be exemplified. The consequences for forecasting and the need for new mean forecasting methods will be discussed.

Let X_t be a univariate stationary Markov(1) time series. The temporal evolution of the time series is completely specified by the joint distribution of random variables from consecutive time points, i.e. $F_{X_t, X_{t-1}}$. Using Sklar's theorem (Eq. 2), the joint

³ Another special class of decompositions are canonical vines (C-vine). In this decomposition, the unconditional dependence structures are all centered around one variable, e.g. $c[u_i, u_{i+1}]$, $c[u_i, u_{i+2}]$, ... In a graphical representation, the unconditional connection between variables resembles a star. In this study only D-vines are used.

distribution can be decomposed into copula and marginal distributions as

$$F_{X_t, X_{t-1}}(a, b) = C [F_{X_t}(a), F_{X_{t-1}}(b)]. \tag{13}$$

By the stationarity of X_t , $F_{X_t} = F_{X_{t-1}} =: F_X$. Hence, the model can be determined by choosing an appropriate marginal distribution F_X and an appropriate copula specification. Note that the marginal distribution F_X is the unconditional distribution of X_t . Conditional properties of the time series are completely determined by the conditional copula. The conditional density of $X_t|X_{t-1}$ is given by

$$f_{X_t|X_{t-1}}(a|b) = c [F_X(a), F_X(b)] f_X(a). \tag{14}$$

Hence, for forecasting, the conditional density of $(X_t|X_{t-1} = x_{t-1})$ can be used as a probabilistic forecast. This model can be understood as a generalization of the AR(1) model [24].⁴ The Gaussian autoregressive model can be recovered by choosing $C = C^{\text{Gaussian}}$ and $F_X = \Phi$. When allowing other dependence structures, any temporal dependency representable by a copula can be reproduced. The concept of the copula-based time series models can be further illustrated by its conditional model equation:

$$X_t|(X_{t-1} = x_{t-1}) = F_X^{-1}(C^{-1} [w_t F_X(x_{t-1})]), \quad w_t \stackrel{iid}{\sim} U[0, 1]. \tag{15}$$

In this formulation, the non-linear connection between X_t and X_{t-1} becomes obvious. The generalization of the temporal copula time series model to d -variate time series, hence spatio-temporal time series models, is straightforward. Let $\mathbf{X}_t = (X_{1,t}, \dots, X_{d,t})$ be a stationary Markov(1) time series. The structure of the time series is completely captured by the joint distribution of \mathbf{X}_t and \mathbf{X}_{t-1} ,

$$F_{\mathbf{X}_t, \mathbf{X}_{t-1}}(\mathbf{a}, \mathbf{b}) = C [F_{X_1}(a_1), \dots, F_{X_d}(a_d), F_{X_1}(b_1), \dots, F_{X_d}(b_d)]. \tag{16}$$

The conditional density, given the observations from time point $(t - 1)$, is as follows:

$$f_{\mathbf{X}_t|\mathbf{X}_{t-1}}(\mathbf{a}|\mathbf{b}) = \frac{c [F_{X_1}(a_1), \dots, F_{X_d}(a_d), F_{X_1}(b_1), \dots, F_{X_d}(b_d)]}{c [F_{X_1}(b_1), \dots, F_{X_d}(b_d)]} \times f_{X_1}(a_1) \cdots f_{X_d}(a_d). \tag{17}$$

To sample from the conditional distribution, as necessary for Monte Carlo approximations of conditional forecasts, the following procedure as in [22] is employed. First transform the observations at time $t - 1$ to pseudo-observations. This is done by applying the probability integral transform to the observations, $(u_{t-1,1}, \dots, u_{t-1,d}) := (F_{X_1}(x_{t-1,1}), \dots, F_{X_d}(x_{t-1,d}))$. Then sample n d -dimensional realizations from

⁴ The generalization to $AR(p)$ models can be achieved by permitting the time series to be a Markov(p) process.

the conditional copula, Eq. 5 (Details on how to sample from the t-copula can be found in [22]. Details to sampling from vine copulas can be found in [1]). At last, the n realizations have to be quantile-transformed with their respective marginal distribution, yielding the n samples of the conditional distribution.

For our empirical study, namely forecasting energy commodity prices, we consider the following models: First, the spatio-temporal time series model where the copula is specified as the Gaussian copula (Eq. 8). The marginals are approximated non-parametrically by the empirical distribution:

$$F_{\mathbf{X}_t, \mathbf{X}_{t-1}}(\mathbf{a}, \mathbf{b}) = \Phi_{\Sigma}[\phi^{-1}(F_{X_1}^{\text{emp}}(a_1)), \dots, \phi^{-1}(F_{X_d}^{\text{emp}}(a_d)), \phi^{-1}(F_{X_1}^{\text{emp}}(b_1)), \dots, \phi^{-1}(F_{X_d}^{\text{emp}}(b_d))]. \quad (18)$$

This model is sensible to use when the dependence structure between the variables as well as the temporal dependence is linear. When the dependence structure exhibits heavy-tailedness, the spatio-temporal t-copula model with non-parametric marginals poses a viable alternative, which we consider as well,

$$F_{\mathbf{X}_t, \mathbf{X}_{t-1}}(\mathbf{a}, \mathbf{b}) = t_{\nu, \Sigma}[t_{\nu}^{-1}(F_{X_1}^{\text{emp}}(a_1)), \dots, t_{\nu}^{-1}(F_{X_d}^{\text{emp}}(a_d)), t_{\nu}^{-1}(F_{X_1}^{\text{emp}}(b_1)), \dots, t_{\nu}^{-1}(F_{X_d}^{\text{emp}}(b_d))]. \quad (19)$$

Third, to allow for more flexibility, we utilize the spatio-temporal D-vine copula with non-parametric marginals. For convenience, the model is presented in terms of its joint density and with variables $(\mathbf{a}, \mathbf{b}) =: \mathbf{p}$,

$$f_{\mathbf{X}_t, \mathbf{X}_{t-1}}(\mathbf{p}) = \prod_{j=1}^{2d-1} \prod_{i=1}^{2d-j} c[F_{X_i}(p_i), F_{X_{i+j}}(p_{i+j}) | F_{X_{i+1}}(p_{i+1}), \dots, F_{X_{j-1}}(p_{j-1})] \times f_{X_1}(p_1) \cdot \dots \cdot f_{X_d}(p_d) f_{X_1}(p_{d+1}) \dots f_{X_d}(p_{2d}). \quad (20)$$

This model is sensible to use when the dependence between variables differs in its structure or when the temporal dependence differs from the cross-sectional dependence.

Fourth, applicable solely for temporal modeling, the temporal t-copula is employed,

$$F_{X_t, X_{t-1}}(a, b) = t_{\nu, \Sigma}[t_{\nu}^{-1}(F_X^{\text{emp}}(a)), t_{\nu}^{-1}(F_X^{\text{emp}}(b))]. \quad (21)$$

The heavy-tailed temporal dependence that this model exhibits is suitable for conditional heteroskedasticity modeling as discussed in [19]. This, and the emergence of non-elliptical probabilistic forecasts, is what we focus on next.

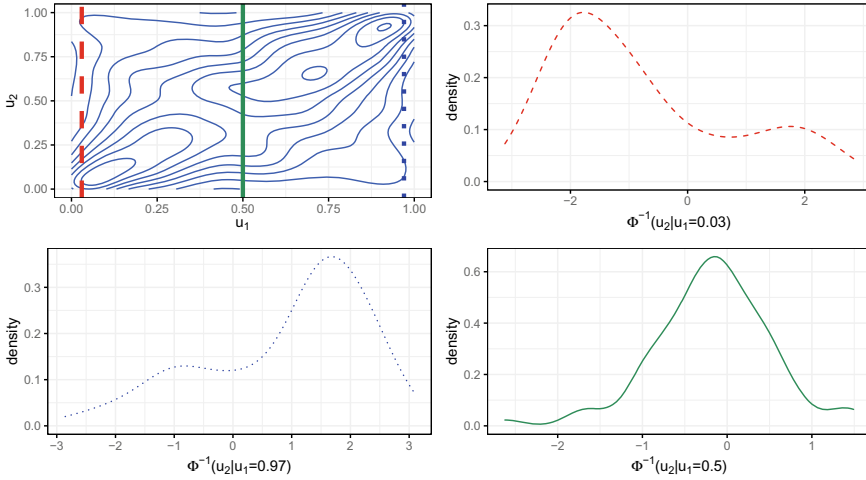


Fig. 3 Visualization of the conditional density structure depending on the value of the conditioning variable. The underlying model assumes the t-copula with dependence parameter $\rho = 0.4$ and degree of freedom $\nu = 2.1$. The marginal distribution is assumed as the standard normal distribution. The upper left panel shows the copula density of 2000 realizations of the before mentioned t-copula. The three lines indicate the three cases where the conditional density is examined. The conditional density is calculated by collecting all values in the neighborhood ($u_1 \pm 0.025$) of the conditioning variable and quantile-transforming them. The density in the upper right panel displays the conditional density given $u_1 = 0.03$. The lower panels display the conditional densities given $u_1 = 0.5$, respectively $u_1 = 0.97$

3.4 Emergence of Non-elliptical Probabilistic Forecasts

The conditional distributions, respectively probabilistic forecasts from (spatio)-temporal copula time series models, can be non-elliptical because of non-linear influences of the conditioning variable. In the following, the behavior of the conditional distributions will be examined with regard to the temporal t-copula with standard normal marginal distribution.⁵ The emergence of non-elliptical conditional densities from the heavy-tailed t-copula is visualized in Fig. 3. When the conditioning variable takes moderate values around $u_1 = 0.5$, the resulting conditional density is approximately elliptical. However, when the conditioning variable takes extreme values, e.g. $u_1 = 0.03$ and $u_1 = 0.97$, the conditional density becomes bimodal. Thus, depending on the value of the conditioning variable, the resulting conditional density can have fundamentally different structures. This behavior offers a new approach to conditional heteroskedasticity modeling. Instead of widening the conditional density as in GARCH models, the density gets bimodal. This can be viewed as a sensible approach to volatility because the extreme behavior in volatile phases is mirrored in this model: When the time series takes a very low value at time point ($t - 1$), it

⁵ The choice of the standard normal distribution is just for convenience. The example would still be valid with other marginal distributions, e.g. students t-distribution.

can be expected that the value at time point t will either be also very low or very high. The variance at time point t is increased nevertheless, but the mechanism for the increase in variance is a new one.

The temporal t-copula approach to conditional volatility, however, holds a problem. When the conditional density is non-elliptical, it is not clear what constitutes a sensible mean forecast. The expectation value may not be suitable in extreme cases where the conditional density is bimodal because the expectation value will take a value which is less probable than, e.g. the modes. Taking the mode as mean forecast could be a solution. Another possible solution to the problem of mean forecasting is to augment the forecast by an artificial neural network (ANN). The ANN predicts which quantile of the conditional distribution is best (in terms of MSE) to use as mean forecast. The inputs of the ANN are past values of the time series and the last optimal quantiles. The ANN architecture used in this work is the basic multi-layer perceptron (MLP) structure. We refer to [14] for an introduction to the topic.

4 Forecasting Study

This section comprises the results of an expanding window forecasting study, investigating the one-step-ahead forecasting performance of different models. The first 1000 observations are used as initial training data set. The following models are considered for evaluation.

- (1) The temporal t-copula model with non-parametric marginals, Eq. 21, henceforth denoted by Tem-t,
- (2) The spatio-temporal D-vine copula model, Eq. 20, henceforth denoted by S-Tem D-vine,
- (3) The spatio-temporal t-copula model, Eq. 19, henceforth denoted by S-Tem-t,
- (4) The spatio-temporal Gaussian copula model, Eq. 18, henceforth denoted by S-Tem-Gaussian,
- (5) The autoregressive moving-average model with external regressors and absolute value, threshold generalized autoregressive conditional heteroskedasticity model, henceforth denoted by ARMAX-AVTGARCH (closely related to the newly proposed model of [6]).

The models' distributional forecasting performance is examined by the continuous ranked probability score (CRPS) [13]. Further, the ANN-assisted mean forecasts of the S-Tem D-vine model and the Tem-t model are compared with the mean forecasts from the ARMAX-AVTGARCH model by RMSE. For each time series, the ARMAX-AVTGARCH model is fitted individually. All models are estimated via Maximum Likelihood. However, the marginals of the copula models are estimated non-parametrically to avoid transmitting estimation errors caused by two-step estimation [21]. The order of the variables in the S-Tem D-vine copula model is fixed as

Table 1 Aggregated CRPS values of the competing models for their one-day-ahead probabilistic forecast for the four commodities. The CRPS is evaluated for the period 12/19/2013 to 02/23/2021, comprising 1861 observations

Model/commodity	S-Tem D-Vine	ARMAX- AVTGARCH	Tem-t	S-Tem-t	S-Tem- Gaussian
Natural gas	0.236	0.227	0.230	0.234	0.234
Oil	0.564	0.548	0.551	0.559	0.558
Coal	0.400	0.389	0.392	0.398	0.397
CEF	0.236	0.222	0.229	0.234	0.234

$$\text{CEF} - \text{coal} - \text{oil} - \text{NGas} - \text{NGas lag} - \text{oil lag} - \text{coal lag} - \text{CEF lag}. \quad (22)$$

This order is chosen to enable the lagged natural gas price to directly interact with the non-lagged natural gas price. The Gaussian, Gumbel, Clayton and t-copula are allowed as bivariate copulas in the D-vine decomposition (Eq. 12). The probabilistic forecasts of all models are approximated by Monte Carlo simulations with 1000 samples for each forecast. Table 1 displays the models’ performances in terms of the CRPS. The ARMAX-AVTGARCH model performs best with regard to univariate distributional forecasting. However, the S-Tem D-vine model, the S-Tem-t and the Tem-t model are competitive. The performance of the copula models may be enhanced, when the marginal distributions are modeled parametrically. The empirical marginal distributions of the copula may not capture all marginal features of the time series. More versatile copula models could be used to enhance the forecast. The conditional dependence modeling may only be able to capture parts of the conditional effects.

The probabilistic forecasts from the Tem-t model during a volatile period are displayed in Fig. 4. During volatile times, the probabilistic forecasts are non-elliptical. During these times, the ANN-augmented mean forecasts can be valuable. The mean forecasting performance of the models can be found in Table 2. The evaluation starts at the 2001st observation, because the first 1000 probabilist forecasts are used to train the ANN. The hybrid, ANN-augmented S-Tem vine and the ANN-augmented Tem-t model generate the best mean forecasts. The mean forecasts of the ARMAX-AVTGARCH model are competitive though. Note that the ANN model used for forecasting is built according to the basic multi-layer perceptron architecture. It is not perfectly suitable for catching sequential patterns. Using recurrent neural networks, especially long short-term memory architectures could enhance the performance even more and could be subject to future research. Also incorporating a measure for the structure of the probabilistic forecast could enhance the performance. However, this would require more advanced architectures.

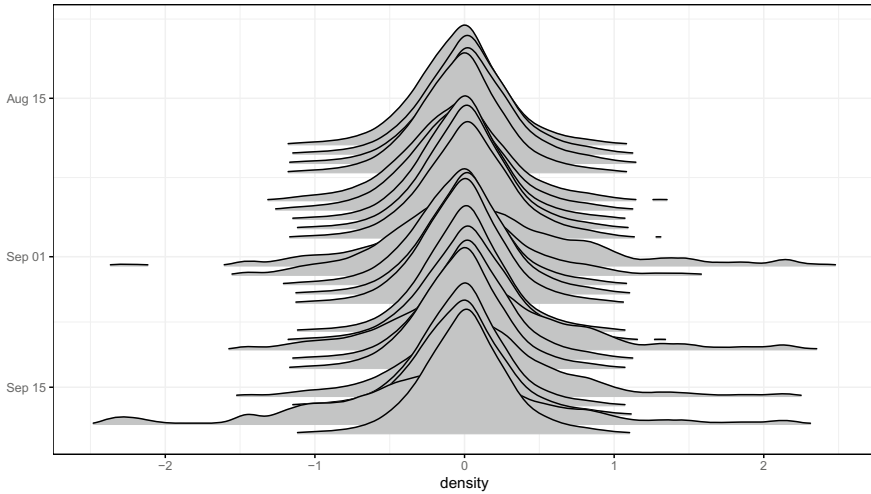


Fig. 4 Probabilistic forecasts for natural gas futures generated from the temporal t-copula model. The forecast densities can be non-elliptical during volatile times (August 2019–September 2019)

Table 2 Aggregated RMSE values of the competing mean forecasting procedures for the four commodities. The RMSE is evaluated for the period 10/23/2017 to 02/23/2021, comprising 861 observations

Model/commodity	S-Tem D-Vine ANN	ARMAX- AVTGARCH	S-Tem D-Vine mean	S-Tem D-Vine mode	Tem-t ANN
Gas	0.594	0.600	0.599	0.597	0.589
Oil	1.222	1.222	1.236	1.246	1.220
Coal	1.000	0.999	1.009	1.002	0.997
CEF	0.605	0.608	0.614	0.604	0.607

5 Conclusion

The application of copula-based time series models to natural gas and related commodity prices is explored in this work. An expanding window forecasting study is conducted. The time series comprises short-term future price series of natural gas, crude oil, coal and carbon emissions.

After introducing the basic notions of dependence modeling with copulas and the D-Vine copula, the copula-based time series models from the literature are reviewed. The emergence of non-elliptical probabilistic forecasts is exemplified using the temporal t-copula. It is visualized how the temporal t-copula offers an alternative approach to conditional heteroskedasticity modeling. It is not clear what constitutes a sensible mean forecast when the probabilistic forecast is non-elliptic. To this end, an artificial neural network is employed to predict what quantile of the probabilistic forecast is best to use as mean forecast. The inputs of the artificial neural network

are past values of the multivariate time series and the last best quantiles of the probabilistic forecast.

In a forecasting study, the predictive performance of the temporal t-copula, the spatio-temporal t-copula and the spatio-temporal D-Vine copula is examined. The marginal distributions are estimated by the respective empirical distribution. The performance is compared with the performance of an autoregressive moving-average model with external regressors and absolute value, threshold generalized autoregressive conditional heteroskedasticity model (ARMAX-AVTGARCH). A closely related model was recently shown to be the best model for natural gas forecasting. Hence, it is understood as a benchmark model. The distributional predictive performance is examined by the continuous ranked probability score (CRPS). We find that the copula-based time series models are competitive with the ARMAX-AVTGARCH model. The mean forecasts are evaluated by the root mean squared error (RMSE). The ANN-augmented mean forecasts perform best, although the forecasts from the ARMAX-AVTGARCH model are still competitive.

The performance of the copula-based time series models could be enhanced by modeling the marginal distributions parametrically. The non-parametric modeling may not catch all marginal features of the time series. However, this procedure requires the estimation to be conducted in one step to guarantee efficient estimation. Another possibility to enhance the performance is to consider more versatile copula models. The current modeling may not capture all conditional features of the time series. Another possibility, with regard to the vine copula model, is to consider other vine structures. In this work, the D-vine structure was imposed. Other structures may be able to capture the dependencies better. As for the mean forecasts, it was shown that the ANN-augmented forecasts perform well. Even though we choose to utilize the standard multi-layer perceptron architecture, which cannot model sequential information perfectly well, the precision was increased. Using more sophisticated architectures that are more suited to catch sequential information will be subject to future research. It would also be interesting to use other models to predict the best quantile for mean forecasting.

Acknowledgements The authors gratefully acknowledge the computing time provided on the Linux HPC cluster at Technical University Dortmund (LiDO3), partially funded in the course of the Large-Scale Equipment Initiative by the German Research Foundation (DFG) as project 271512359.

References

1. Aas, K., Czado, C., Frigessi, A., Bakken, H.: Pair-copula constructions of multiple dependence. *Insurance: Math. Econ.* **44**(2), 182–198 (2009)
2. Aloui, R., Aïssa, M.S.B., Hammoudeh, S., Nguyen, D.K.: Dependence and extreme dependence of crude oil and natural gas prices with applications to risk management. *Energy Econ.* **42**(332–243), 2014 (2014)
3. Angus, J.E.: The probability integral transform and related results. *SIAM Rev.* **36**(4), 652–654 (1994)
4. Beare, B.K.: Copulas and temporal dependence. *Econometrica* **78**(1), 395–410 (2010)

5. Beare, B.K., Seo, J.: Vine copula specifications for stationary multivariate Markov chains. *J. Time Ser. Anal.* **36**(2), 228–246 (2015)
6. Berrisch, J., Ziel, F.: Distributional modeling and forecasting of natural gas prices. *J. Forecast.* (2022)
7. Berrisch, J., Pappert, S., Ziel, F., Arsova, A.: Modeling volatility and dependence of European carbon and energy prices (2022). [arXiv:2208.14311](https://arxiv.org/abs/2208.14311)
8. Brechmann, E.C., Czado, C.: COPAR-multivariate time series modeling using the copula autoregressive model. *Appl. Stoch. Models Bus. Ind.* **31**(4), 495–514 (2015)
9. Chen, X., Fan, Y.: Estimation of copula-based semiparametric time series models. *J. Econom.* **130**(2), 307–335 (2006)
10. Czado, C.: Pair-copula constructions of multivariate copulas. In: *Copula Theory and its Applications*, pp. 93–109. Springer, Berlin, Heidelberg
11. Demarta, S., McNeil, A.J.: The t copula and related copulas. *Int. Stat. Rev.* **73**(1), 111–129 (2005)
12. Genest, C., Rivest, L.P.: Statistical inference procedures for bivariate Archimedean copulas. *J. Am. Stat. Assoc.* **88**(423), 1034–1043 (1993)
13. Gneiting, T., Raftery, A.E.: Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **102**(477), 359–378 (2007)
14. Higham, C.F., Higham, D.J.: Deep learning: an introduction for applied mathematicians. *Siam Rev.* **61**(4), 860–891 (2019)
15. Hofert, M.: Sampling archimedean copulas. *Comput. Stat. Data Anal.* **52**(12), 5163–5174 (2008)
16. Hu, L.: Dependence patterns across financial markets: a mixed copula approach. *Appl. Financ. Econ.* **16**(10), 717–729 (2006)
17. Joe, H.: *Dependence Modeling with Copulas*. CRC Press (2014)
18. Jondeau, E., Rockinger, M.: The copula-Garch model of conditional dependencies: an international stock market application. *J. Int. Money Finance* **25**(5), 827–853 (2006)
19. Loaiza-Maya, R., Smith, M.S., Maneesoonthorn, W.: Time series copulas for heteroskedastic data. *J. Appl. Econom.* **33**(3), 332–354 (2018)
20. Nagler, T., Krüger, D., Min, A.: Stationary vine copula models for multivariate time series. *J. Econom.* **227**(2), 305–324 (2022)
21. Patton, A.: Copula methods for forecasting multivariate time series. In: *Handbook of Economic Forecasting*, vol. 2, pp. 899–960 (2013)
22. Simard, C., & Rémillard, B.: Forecasting time series with multivariate copulas. *Depend. Model.* **3**(1) (2015)
23. Sklar, M.: Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris* **8**, 229–231 (1959)
24. Smith, M., Min, A., Almeida, C., Czado, C.: Modeling longitudinal data using a pair-copula decomposition of serial dependence. *J. Am. Stat. Assoc.* **105**(492), 1467–1479 (2010)