# Chapter 1
# Introduction

**Abstract** This chapter provides an intuitive, easy to read explanation of what geospatial abduction is. It uses a set of examples to explain what geospatial abduction is, and how it can be used to solve real-world problems in many different domains. Our examples show how geospatial abduction can be used to (i) identify the locations of weapons caches supporting improvised explosive device attacks by terrorists and armed insurgents from information about the locations of the attacks, (ii) identify the possible locations of tigers from information about locations of their kills, (iii) identify habitats that support host animals that carry certain viruses from information about where diseases caused by those viruses occurred, and (iv) identify the location(s) of a burglar from information about where burglaries he carried out occurred. These four examples are used continuously throughout the book to illustrate the mathematical foundations and definitions that are presented in later chapters.

## 1.1 Motivation

There are numerous applications in the real world in which we observe that certain phenomena occur at various locations and where we wish to infer various "partner" locations that are somehow associated with those observations. Partner locations could be associated with entities that cause the phenomena we observe or facilitate the observations that we observe.

Informally speaking, a *geospatial abduction problem* (GAP) refers to the problem of finding partner locations that best explain a set of observations (at certain locations), in the context of some domain-specific information that tells us something about the relationship between the observations we make and the partner locations that cause, facilitate, support, or are somehow correlated with the observations. Geospatial abduction was first introduced by the authors in [4] and later studied by them in a series of papers [5, 6, 7, 8].

P. Shakarian and V.S. Subrahmanian, *Geospatial Abduction: Principles and Practice*,
DOI 10.1007/978-1-4614-1794-1_1, 

For instance, we have used geospatial abduction to find the locations of the weapons caches that allow insurgents and terrorists in Baghdad, Iraq to carry out improvised explosive device (IED) attacks both on Iraqi civilians, as well as on multinational troops situated there. In this application of the geospatial abduction technique, the observations correspond to the locations of the attacks and the partner locations we wish to find are the locations of the weapons caches that facilitate or support those attacks. Of course, to do so, we must take domain information into account. What kinds of places are suitable locations for weapons caches? Are there operational constraints on the insurgents that somehow constrain how far the weapons caches can be from the locations of the attacks? The answers to these questions constitute *application-specific* information related to the problem of detecting IED caches. A generic algorithm, computational engine, or software tool for geospatial abduction must support application development where the application (in this case, detection of IED weapons caches) developer can explicitly articulate such application-specific information to the GAP engine in a manner that is *uniform* and *application independent*.

As another application, consider the case of tiger conservation. The number of tigers in the world is dwindling rapidly and organizations such as the World Wildlife Fund (WWF) are making heroic efforts to save the tiger. Unfortunately, the tiger is not an easy animal to save. Unlike lions, they are solitary creatures that maintain a very stealthy existence. Their range can be over 100 square miles, often making it difficult to pinpoint exactly where they like to reside at a given time. We have been considering the prospect of identifying relatively small regions where tigers might like to reside based on observations (locations) of tiger kills. Fortunately, after eating its meal, the tiger does not drag away the carcass or skeleton that is left behind, providing researchers and conservationists valuable information on where the tiger has been. In this application as well, we need to take much domain specific information into account (*e.g.*, a wide open space is not a place where it is likely that a tiger will dwell, nor is a place where there is a paucity of prey [10]).

A third application we have worked on is an effort led by epidemiologists at UCLA that involves identifying the habitats of creatures that carry certain viruses. For instance, monkey pox [2] is a deadly disease that kills and/or irreparably damages many children—and even adults—in Africa. It is particularly widespread in the Democratic Republic of Congo. The disease is spread by host animals that are often eaten raw by a hungry, highly malnourished human population, who are desperate for food. Thus, a natural public health question arises. Can we somehow identify the habitats where the host animals live in large numbers so that appropriate public health measures (*e.g.*, extermination of the hosts or other environmentally appropriate actions) can be taken? As in the case of the tigers above, this requires application specific knowledge about the types of environments/habitats that the host animals prefer and/or flourish in.

A fourth application deals with crime. We are all painfully aware of the existence of burglars and home invasions. How can we identify the locations (home or office or even a significant other's house—as long as it is a place where the burglar spends a fair amount of time) of an unknown burglar or home invader by examining the locations where the burglaries or home invasions were committed? In this case, again, domain specific information can be taken into account. For instance, we know that burglaries are usually committed in neighborhoods that the burglar knows, but usually the burglar targets homes that are not too close to either his home or his office or places where he spends a lot of time and is known to others. How do we find the burglar's house or somehow narrow down the space of possible targets?

In the rest of this chapter, we explore these applications in further detail, clearly articulating the issues involved in further detail. In short, this chapter tries to explain what types of real-world problems geospatial abduction is supposed to solve, but not how. Following this chapter, most of the rest of this book will focus on the "how."

## 1.2 The IED Cache Detection Problem

Improvised Explosive Devices (IEDs) are crude bombs constructed by insurgents to attack an external force. The term IED was first introduced by the British Army in response to attacks by the Irish Republican Army (IRA) in the 1970s. Since then, it has been used by insurgent groups around the world to attack external forces.

Figure 1.1 shows a screenshot of real-world data gleaned from open sources about the locations of IED attacks in Baghdad during the February 2007–November 2008 time frame. The map was generated using the Spatio-Cultural Abductive Reasoning Engine (SCARE) system [4] which in turn used Google Maps to get geographic data. The red push pins show the locations of IED attacks during this time frame.

All the IED attacks shown in Figure 1.1 were believed to have been carried out by Shiite-militia supported by Iran. Experience has shown that these attacks were typically carried out by insurgents who placed their munitions in *weapons caches*. A weapons cache was then used to support one or more attacks.

Of course, the insurgents were not stupid and had no wish to get caught. Weapons caches were chosen carefully. In particular, it was clear that the insurgents could not locate weapons caches within US or international coalition bases. Likewise, they could not locate weapons caches within Sunni neighborhoods of Baghdad because of ongoing ethnic conflict between the Shiites and Sunnis. Last, but not least, we deemed that they could not place weapons caches on the Tigris river because of the probability of being spotted as well as the logistical difficulties involved in transporting munitions from a river to land. [9] contains further work on IED cache placement. The shaded regions in Figure 1.2 shows regions where the IED caches could not be located.

The job of a geospatial analyst is now clear. Is there a way to study the map of Figure 1.1 showing the locations of the IED attacks, together with the map overlays shown in Figure 1.2 showing where caches could not possibly occur, and infer the plausible locations of weapons caches used to support the IED attacks carried out by Shiite insurgents?

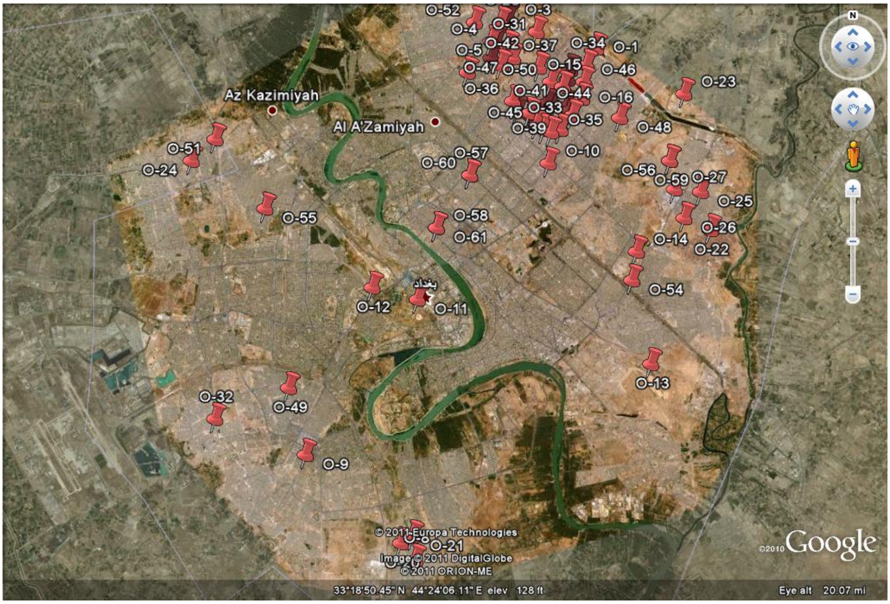


**Fig. 1.1** SCARE [4] screenshot showing locations of IED attacks in Baghdad during the Feb. 1, 2007 to Nov. 2008 time frame.

The problem is highly non-trivial to solve for several reasons. First, we do not know how many IED caches there are to find. Second, the zones where IED caches cannot be present (as shown in Figure 1.2) are highly irregular in shape—so simple geometric reasoning cannot be a solution. Third, the insurgents are constantly adapting their attack techniques to any counter-measures being taken to find and/or thwart them. Finally, as we shall show in Chapter 2, the problem of finding a set of such cache locations is NP-complete, making it intractable to compute in practice.

We have developed two systems called SCARE [4] and SCARE-S2 [7] that use geospatial abduction. SCARE used a version of geospatial abduction called *point-based geospatial abduction* (studied in Chapter 2) that was applied to the problem of finding IED weapons caches in Baghdad. Using 21 months of data (7 for training, 14 for evaluation), we were able to show that SCARE predicted cache locations that (on average) were within 0.45 miles of the actual locations of caches discovered in Baghdad by coalition forces.

SCARE-S2 was applied to the problem of discovering high value targets (or HVTs) in Helmand and Kandahar provinces of Afghanistan. HVTs were defined to be either depot-level weapons caches (as opposed to smaller caches designated for more immediate use) or insurgent commanders. SCARE-S2 used a different technique than SCARE called *region abduction*, described in detail in Chapter 3, to identify regions in these provinces that were highly likely to contain HVTs. Comparison with real-world data showed that the regions we discovered had a density of HVTs that was 35 times higher than the density of HVTs in the two provinces

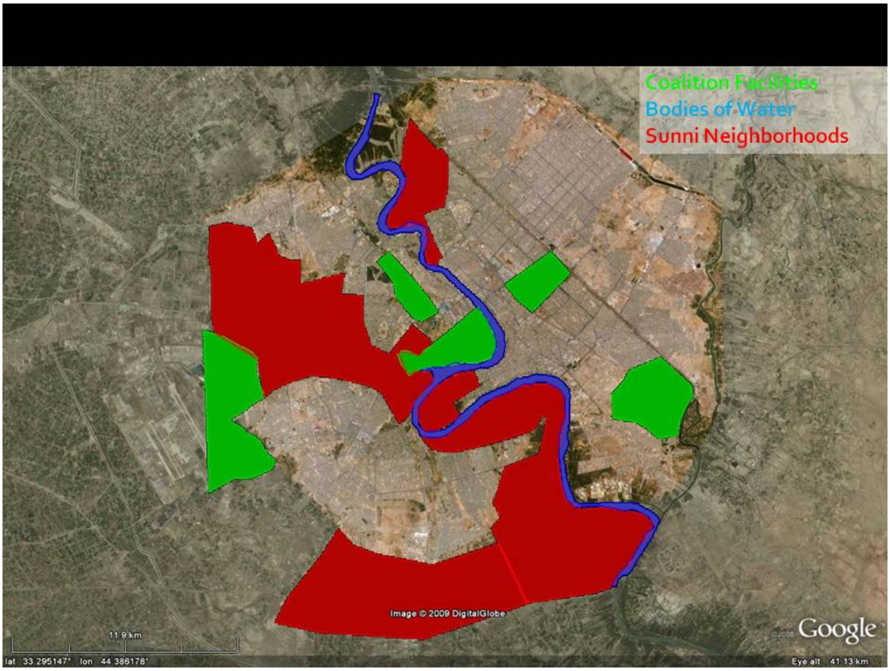


**Fig. 1.2** SCARE [4] screenshot showing coalition bases, Sunni neighborhoods, and the Tigris River.

considered as a whole. In addition, these regions contained on average 4.8 villages that needed to be searched by US and coalition forces.

## 1.3 The Tiger Detection Problem

As anybody who has ever gone to a tiger reserve knows, getting to the tiger reserve is easy, but spotting a tiger is hard. Tigers are hunters who live a largely solitary existence and depend on stealth attacks in order to capture prey. At the time this chapter was written, the World Wildlife Fund estimated that there are fewer than 3,200 tigers still living in the wild in the entire world.

Wildlife experts have considerable interest in identifying the precise region where the tigers are living so that appropriate conservation steps can be taken.[1] Consider the Achanakamar Wildlife Sanctuary (AMWLS) in the state of Chattisgarh, India. Tiger conservation experts would like to understand exactly where the tigers reside. In order to do so, the wildlife conservators looking at a map of AMWLS

[1] We thank Tom Snitch for suggesting we consider this problem using geospatial abduction techniques after a meeting he had with World Wildlife Fund officials who expressed concern about the need for better tracking of tigers.

need to identify locations in the sanctuary that are feasible for tigers to adopt as their range. This involves a number of issues. For instance:

- The places where the tigers live needs to have a high concentration of *prey* which in the case of this sanctuary includes chital, sambar, as well as wild boar.
- The placess where the tigers live need to have the right kind of vegetation, involving variables such as "canopy cover, canopy height, forest, shrub cover, shrub height" [10, page 563].
- The number of dung pellets found in a given region is also correlated with the suitability of a location for the tiger's habitat as this is closely correlated with the amount of prey in the area (more dung pellets implies more prey).

Thus, wildlife analysts may first plot a "habitat map" showing locations that are suitable for the tiger to live versus locations that are not suitable for the tiger to live, as shown in Figure 1.3 below.

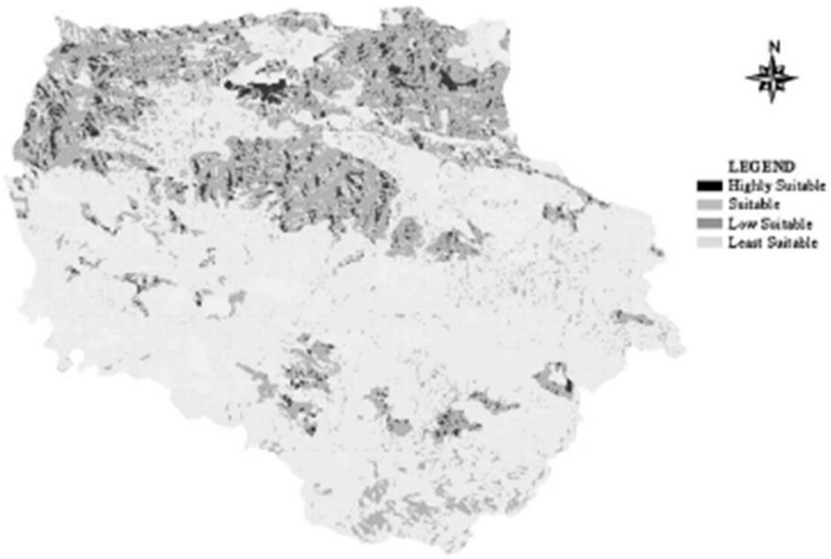


**Fig. 1.3** Tiger habitat suitability map for the Achanakamar Wildlife Sanctuary—figure taken from *M. Singh, P.K. Joshi, M. Kumar, P.P. Dash and B.D. Joshi. Development of tiger habitat suitability model using geospatial tools: a case study in Achankmar Wildlife Sanctuary (AMWLS), Chhattisgarh India, Env. Monitoring and Assessment journal, Vol. 155, pages 555-567, 2009.* and reprinted courtesy of Springer.

A wildlife analyst equipped with such a map (stored in the Keyhole Markup Language, or KML, format) can use geospatial abduction through SCARE or SCARE-S2 to upload an Excel file containing information on the location of various tiger kills. Figure 1.4 shows one such example (synthetic data) of locations of tiger kills in AMWLS.

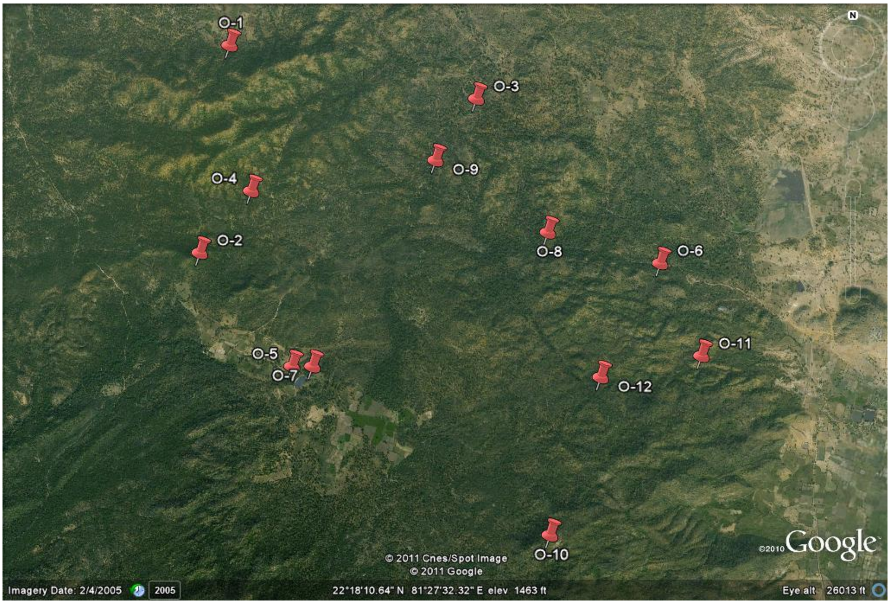


**Fig. 1.4** Tiger kill locations in AMWLS. Synthetic data used for example purposes only.

The goal is to now determine where the tiger responsible for the actual kills lives, given both the locations of its kills and the habitat suitability map. Region-based geospatial abduction studied in Chapter 3 provides a suite of techniques to address this problem. Figure 1.5 shows potential locations predicted by SCARE [4].

## 1.4 The Virus Host Habitat Identification Problem

A related potential application of geospatial abduction, similar to the tiger habitat problem, is that of identifying the habitats of animal hosts that carry certain viruses which cause diseases in human populations.[2] Many such diseases fall into the category of *vector-borne diseases* in which a host transmits a virus to humans, usually via a bite.

Realistic examples of such diseases include diseases spread through mosquito bites (*e.g.*, malaria, chikungunya fever, yellow fever, West Nile encephalitis and other types of encephalitis), diseases spread by rodents and rodent fleas (*e.g.*, plague, monkey pox), diseases caused by ticks and deer flies living on deer (*e.g.*, Lyme dis-

[2] We are grateful to Trevon Fuller for thinking of this application.

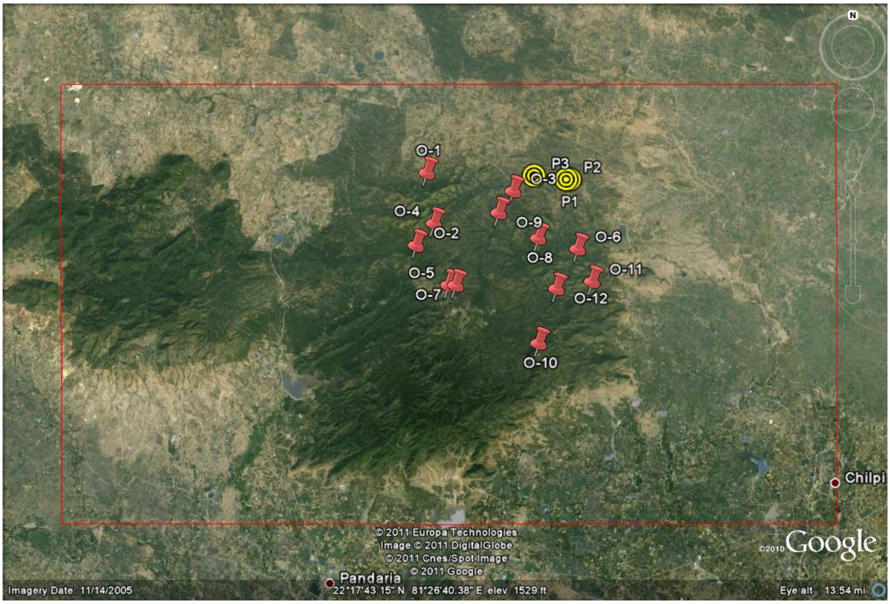


**Fig. 1.5** Tiger kill locations with predicted tiger locations in AMWLS. Synthetic data used for example purposes only.

ease, tularemia), diseases caused by various types of flies (*e.g.*, sleeping sickness), and many others.

In such cases, a public health expert might ask himself the question: How can I identify the locations of habitats of hosts (*e.g.*, deer, rodents) that support the organisms (*e.g.*, ticks) that spread these diseases? To do this, the public health expert can use geospatial abduction to carry out the following steps:

- Identify locations where the disease occurred or has been known to occur (perhaps at a certain level of occurrence or higher so that isolated cases do not skew the analysis).
- Identify the properties of habitats (*e.g.*, standing bodies of water in the case of mosquito-borne diseases or the existence of certain types of foliage in the case of deer) that support the host animals.

Based on these two analyses, the public health analyst can easily use a region-based geospatial abduction tool to identify regions which have a high probability of supporting the hosts that carry and spread the disease. Once these regions are identified, appropriate public health actions can be taken, possibly in conjunction with public health authorities.

## 1.5 The Burglar Detection Problem

Police all over the world are constantly confronted with burglaries. Using a number of forensic techniques, they can often identify which burglaries were committed by the same perpetrator(s). A natural question for criminologists and law enforcement agencies is to figure out how to find the places where the burglar lives or works.

It is well known in criminology [1, 3] that burglars, serial killers, and many other types of criminals often carry out their criminal activities in areas they know well. Typically, this condition of "knowing well" means that either the criminals live in the area where they carry out their crimes, or work there, or grew up there.

Figure 1.6 shows a map of St. Paul, Minnesota, with the locations of various church burglaries explicitly marked via red push pins. This data shows real church burglaries that occurred in 2008–2009, not synthetic information. Moreover, the police in St. Paul believed that these burglaries were all carried out by the same burglar.

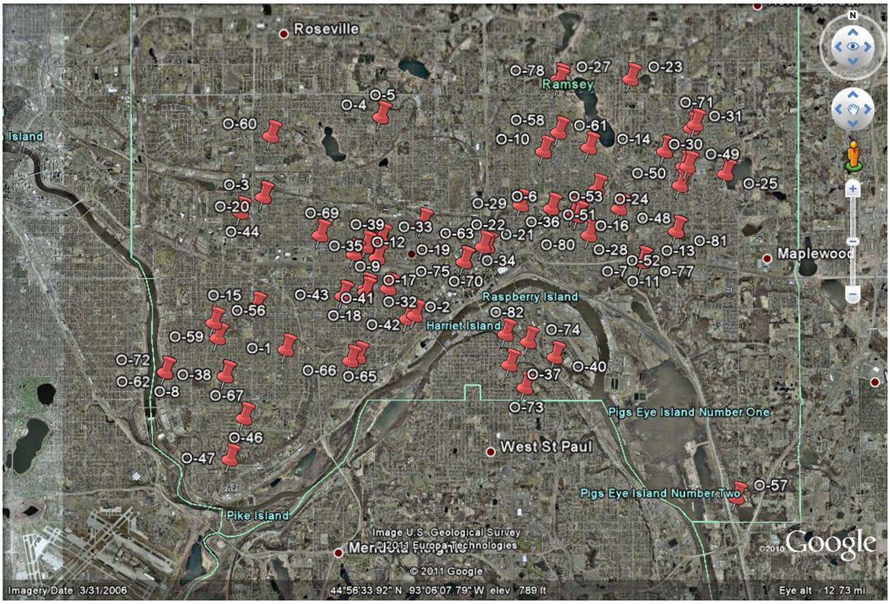


**Fig. 1.6** SCARE [4] screenshot showing locations of church burglaries in St. Paul, Minnesota in 2008–2009.

A criminologist or police officer investigating these burglaries might want to give a geospatial abduction system some information. For instance, he might say that he does not believe that a burglar would commit such crimes less than a kilometer from his house or more than seven kilometers from his house (these distances can also be automatically learned from historical data or explicitly provided by an expert). In addition, he might mark certain regions on the map as unlikely places for the burglar to have his home or office. Such *excluded* regions are shown in Figure 1.7. Note that in this example, these are only "notional" excluded regions and real excluded regions would need to be inserted by a domain expert (*e.g.*, a St. Paul, MN, police officer investigating the burglaries).

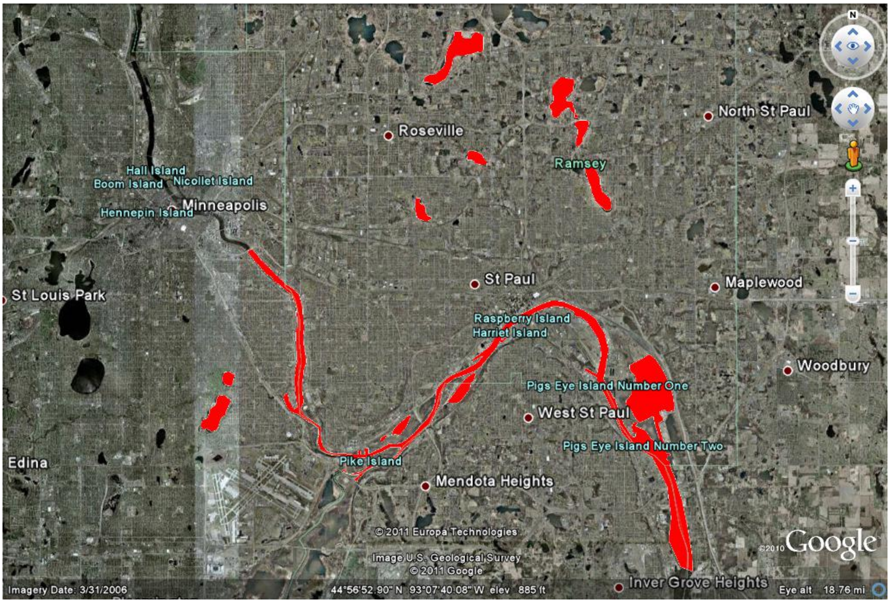


**Fig. 1.7** SCARE [4] screenshot showing regions in St. Paul, Minnesota, that were excluded as potential locations for the church burglar.

Last, but not least, we would like our geospatial abduction system to generate "predicted" locations for the church burglar. It is too hard to designate whether these predicted locations represent his home or his office—rather, they represent locations that are most likely to be locations where he has a significant presence. Figure 1.8 shows the St. Paul, Minnesota, map, together with yellow bull's-eyes reflecting predicted locations. Again, we emphasize that these are *notional* predicted locations; even though the church burglary data we use is real data, our exclusion zones shown in Figure 1.7 may not reflect police knowledge of the reality of crime in St. Paul, and hence, the results shown in Figure 1.8 may be incorrect. Our purpose in this example is to show how such a system should work.

## 1.6 Other Applications

The preceding sections highlight four real-world applications in which geospatial abduction is currently or could be employed. However, the space of possible ap-

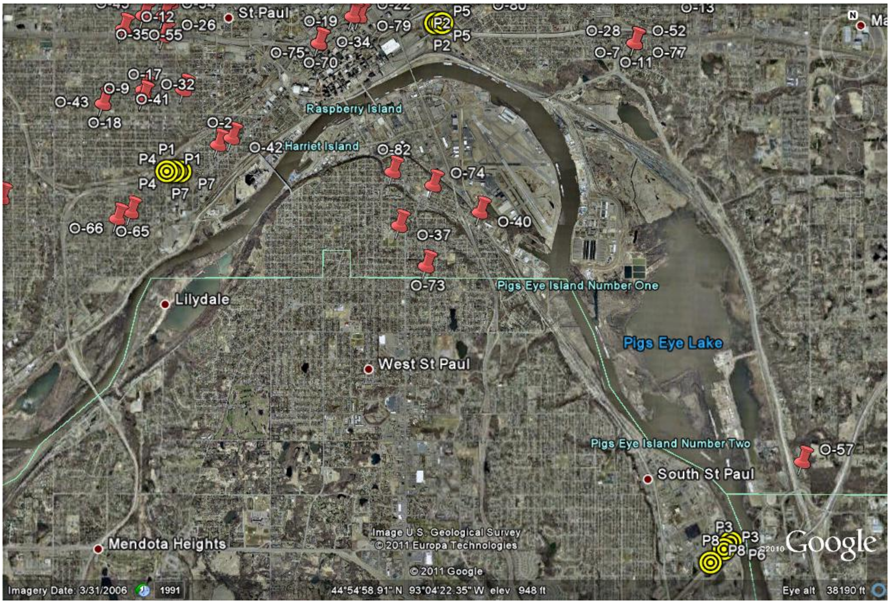


**Fig. 1.8** SCARE [4] screenshot showing locations of predicted locations of the burglar with respect to the church burglaries in St. Paul, Minnesota, in 2008–2009.

plications for geospatial abduction is really much larger; we highlight a few more examples here, though we will be unable to consider them in further detail in the rest of the book.

One important application area deals with environmental pollutants in a body of water. Often times, water is contaminated by unscrupulous organizations or companies that dump toxic waste into a body of water. We do not always know who the responsible party is, but identifying the location(s) where the dumping is likely to be occurring allows environmental authorities and police to target their surveillance efforts with a view to catching the culprits. In this case, the observations are the locations where the pollution was discovered (*e.g.*, contaminated water), and the partners we want to find are the locations where the polluting substances are introduced into the water or into the ground. Domain information specifies how the contamination spreads—either through the water or through the ground.

The same principle also applies to pollution in the ground: we see contaminants at various locations on the ground and we would like to infer the source of these contaminants. The source may be a leak in a network of pipes distributing the substance that is leaking, or an explicit attempt to dump pollutants, or simply an accident. Knowing the location from which the pollutant is coming can play a key role in helping solve the problem.

Another important application is identifying the location of illegal drug labs or distribution centers from information about the locations where various drug dealers were arrested. Alternatively, with aerial surveillance of the coca plant in countries like Colombia and Peru, we know the locations of the base crop that is converted into an illegal substance. Based on the locations of these fields, can we infer the locations of the labs that convert these crops into illegal drugs?

## 1.7 Conclusion

We see in this chapter that geospatial abduction in different forms can be used to help address a wide variety of problems that have a significant geospatial character. We have only described a small number of problems that geospatial abduction can help with. As the technique is studied more extensively, we believe there will be far more applications.

All of these geospatial abduction applications described have the following characteristics:

1. *Observations.* There is a set of observations that we start with. The set of observations could be the locations where IED attacks occurred, where disease outbreaks occurred, where tiger kills were observed, or where pollutants were spotted.
2. *Domain knowledge.* The domain knowledge involved in the class of examples we have discussed include two types of phenomena.
   a. Information about the distances between the locations or regions we are trying to find (*e.g.*, locations of IED weapons caches or regions where the tiger responsible for certain kills may be) and the observations that are causally linked to the observation; and
   b. Maps showing which locations or regions on the ground satisfy various "feasibility requirements" (*e.g.*, having the appropriate type and quantity of prey in the case of the tiger habitat identification problem, or having the right kinds of populations for insurgents to blend into after carrying out IED attacks).

These inputs can be specified in a variety of ways; however, in later chapters of this book, we will show that these inputs can often be specified in a highly (syntactically) restricted format that makes them easy to manipulate computationally.

Once these inputs are provided, we will show in the rest of this book, how we can find a set of places that best explains the observations in our application while being consistent with the provided domain knowledge.

## References

1. Brantingham, P., Brantingham, P. 2008. Crime Pattern Theory. In Enviromental Criminology and Crime Analysis, R. Wortley and L. Mazerolle, Eds., pp. 78–93.
2. Rimoin, A. *et al.* Endemic Human Monkeypox, Democratic Republic of Congo, 2001–2004, Emerging Infectious Diseases, 13, 6, pp. 934–937, 2007.
3. Rossmo, D. K., Rombouts, S. 2008. Geographic Profiling. In Enviromental Criminology and Crime Analysis, R. Wortley and L. Mazerolle, Eds. pages 136–149.

4. Shakarian, P., Subrahmanian, V.S., Sapino, M.L. SCARE: A Case Study with Baghdad, Proc. 2009 Intl. Conf. on Computational Cultural Dynamics (eds. D. Nau, A. Mannes), Dec. 2009, AAAI Press.
5. Shakarian, P., Subrahmanian, V.S., Sapino, M.L. 2012. GAPS: Geospatial Abduction Problems, ACM Transactions on Intelligent Systems and Technology (TIST), 3, 1, to appear.
6. Shakarian, P., Subrahmanian, V.S. Region-based Geospatial Abduction with Counter-IED Applications, accepted for publication in: Wiil, U.K. (ed.).Counterterrorism and Open Source Intelligence, Springer Verlag Lecture Notes on Social Networks, to appear, 2011.
7. Shakarian, P., Nagel, M., Schuetzle, B., Subrahmanian, V.S. 2011. Abductive Inference for Combat: Using SCARE-S2 to Find High-Value Targets in Afghanistan, in Proc. 2011 Intl. Conf. on Innovative Applications of Artificial Intelligence, Aug. 2011, AAAI Press.
8. Shakarian, P., Dickerson, J., Subrahmanian, V.S. 2012. Adversarial Geospatial Abduction Problems, ACM Transactions on Intelligent Systems and Technology (TIST), to appear.
9. Shakarian, P., Otstott, C. What is Old is New: Countering IEDs by Disrupting the Weapon Supply, Military Review, pp. 46–52, 2011.
10. Singh, M., Joshi, P.K., Kumar, M., Dash, P.P., Joshi, B.D. Development of tiger habitat suitability model using geospatial tools—a case study in Achankmar Wildlife Sanctuary (AMWLS), Chhattisgarh India, Env. Monitoring and Assessment journal, Vol. 155, pp. 555–567, 2009.