GI_Forum 2016, Vol.2
Page: 36-45
Full Paper
Corresponding Author:
karin.wannemacher@fhwn.ac.at
DOI: 10.1553/giscience2016_02_s36

Using Geostatistical Methods to Automatically Verify Citizen Science Data on Alien Species

Karin Wannemacher[1] and Roland Grillmayer[2]

[1] University of Applied Sciences Wiener Neustadt, Austria
[2] Austrian Federal Environment Agency (Umweltbundesamt) Wien, Austria

Abstract

With the increase of travel and transportation of goods, the distribution and invasion of alien species have increased. While the majority of neobiota do not cause any problems, there are some that are problematic for nature conservation, have negative effects on the economy, or cause health problems.
In Regulation (EU) No. 1143/2014 of 22 October 2014, the European Parliament and the Council of the European Union published a set of rules to prevent, minimize and mitigate the adverse impacts caused by invasive alien species, and ordered Member States to implement a surveillance system of invasive alien species to prevent the spread of such species into or within the Union.
Citizen Science lends itself to the collation of high-quality information on a wide variety of species over a large scale and the long term. However, verification of the data collected is a key challenge for legal monitoring projects, especially regarding costs. To ensure that the data fits the quality standards while also minimizing the necessary budget for the project, an algorithm for the automated verification of observation data has been developed, based on geostatistical methods.

Keywords:

citizen science, alien species, geostatistics

1 Indroduction

In 1860 a fungal disease known as crayfish plague came to Europe from North America. It proved fatal for the noble crayfish (Astacus astacus) population, which went into a steep decline. In order to compensate such losses, the signal crayfish (Pacifastacus leniusculus) was introduced to Europe's rivers and crayfish farms in the mid-20th century. The species, which also comes from North America, is immune to the fungal disease. This move, however, turned out to be even more bad news for the indigenous species as the signal crayfish, while not affected by the disease, is a carrier of the lethal crayfish plague and also competes for the same habitats as Astacus astacus, where it often has the upper hand over the European species due to its high reproductive rates (Essl & Rabitsch, 2002).

Organisms which, intentionally or unintentionally, have been introduced into a specific area outside their natural range by humans since 1492 are generally known as alien species or neobiota. While the majority of neobiota do not cause any problems, there are some that are problematic for nature conservation, have negative effects on the economy, or cause health problems (Essl & Rabitsch, 2002).

In Regulation (EU) No. 1143/2014 of 22 October 2014 (Council of the European Union, 2014), the European Parliament and the Council of the European Union published a set of rules to prevent, minimize and mitigate the adverse impacts caused by invasive alien species. The regulation identifies three forms of intervention: prevention, early warning and rapid response, and management. EU member states are ordered to implement a surveillance system of invasive alien species, which collects and records data on the occurrence of such species, in order to prevent the spread of invasive alien species into and within the Union.

According to Article 14 (Surveillance system) of the regulation, (a) such a system must cover the whole territory, and determine the presence and distribution of new and already established invasive alien species of concern to the EU; (b) the process should be able to quickly detect any new invasive species; (c) it should comply with, but not duplicate, other regulations on species monitoring; (d) the system must, as far as possible, take relevant trans-boundary impacts into account. The surveillance system should also be used to confirm early detection of the introduction or presence of invasive alien species of concern to the EU. Any such early detection should be notified to the EU without delay (Article 16 - Early detection notifications).

2 Project Parameters

Species monitoring lends itself to citizen science, as economic and logistical factors prevent scientists from generating the volume of data they need for research on their own. One of the greatest concerns about citizen science is the quality of the data gathered. Yet studies have shown that, with proper instructions and statistical methods or expert verification in place, the data collected by citizen scientists can match the quality of data collected by experienced researchers (Jarvis et al., 2015).

Records collated by volunteers are often the only source for high-quality information on a wide variety of species over a large scale and the long term (Roy et al., 2012).

For this project the incoming data is verified by an algorithm, which derives a number of index values to decide whether the observation is credible. The experts on alien species who are involved in the project do not have to evaluate every observation but only those that fail to meet the criteria set by them as threshold parameters for the algorithm. This will reduce the overall costs and make it possible to design this as a long-term project.

While the overall aim is to provide a nationwide platform to observe alien species, it is also possible to connect local projects or projects with a limited list of taxa to the main kernel, so that all feed the same database. Moreover, the same framework can also be used to target white- or red-listed (endangered) species.

Monitoring potentially dangerous and invasive species will help us to establish an early warning system, which would make early responses and counter-measures to threats as effective as possible.

Project Species List

Many invasive species cause damage to managed and natural ecosystems or are responsible for the extinction of native species. The most problematic aliens are those that manage to fill a biological gap within an established ecosystem. Their habits, characteristics and lifespans can have an effect on the surrounding fauna and flora. Neobiota can severely affect nutrient cycles if they are competing for the same resources, or if they prey on native species. Another point of concern is the transmission of parasites and diseases (e.g., hepatitis E) to species that have not had a chance to develop resistances over many generations (Essl & Rabitsch, 2002).

As the list of invasive species mentioned in the EU Regulation 1143/2014 had not yet been developed at the time of our research project, we used a reduced list of species that had been agreed upon with specialists from the Austrian Federal Environment Agency. It included species that citizen scientists can easily identify and/or that are fairly common, such as Ailanthus altissima (tree of heaven), Ambrosia artemisiifolia (ragweed), Bunias orientalis (Turkish rocket), Potentilla indica (Indian mock strawberry) and Robinia pseudoacacia (false acacia). Ragweed, which also has a negative impact on human health (allergies), and false acacia are two of 17 alien plant species that are considered to pose a threat to biodiversity in Austria.

The project list also included Aethina tumida (small hive beetle), a species for which every occurrence in Austria ("Bienenseuchengesetz" [rules for the protection of bees against epizootic diseases], §3.1. and 2.) and the European Union is compulsorily notifiable.

Taxa are referenced internally via one of the three reference lists (EU-Nomen, EUNIS or Natura 2000) recommended in the INSPIRE scheme.

System Architecture

Users can submit their sightings by giving the location, species, date, and at least one picture of the specimen via a mobile-ready website. The uploading of a photo of the monitored taxon is mandatory so that the classification can, if necessary, be verified.

Once submitted, the data are processed by the pre-validation algorithm. The algorithm is designed to consider all the factors that an expert would inspect, apart from the observation picture, and return a value of certainty on whether the observation is credible or not (see section 3 and Figure 1). The combined value is compared with a species-specific acceptance threshold value. If the credibility index is below the threshold, the observation requires additional verification by an expert.

Additionally, other users or experts can confirm or oppose the categorization of observations. Once an expert confirms an observation, it is no longer open for user verification.

All observations are stored in a spatial database, and their derivatives (heat maps, gridded data and features, among others) are accessible as OGC Webservices via Geoserver.

The gridded observation data is modelled taking the concepts of the INSPIRE Data Specification on species distribution into account. In this context, species distribution is defined as a geographical distribution of the aggregated occurrences of animal and plant species by using grids or polygons (INSPIRE Thematic Working Group Species Distribution, 2013).

The base grid used in this project is one for floristic observation data provided by the Federal Environment Agency. If a quadrant contains observations verified by an expert, then the polygon is assigned the value 1. If a quadrant contains observations which have not been verified by an expert, then the polygon is assigned the value 2. Both 1 and 2 indicate the presence of a species.

In accordance with the INSPIRE Data Specification for species observation, a distinction has to be made between areas where a thorough search for a particular species yielded no results and areas that have not been searched at all. As this kind of citizen science project is not designed to confirm the true absence of alien species, quadrants that do not contain any observations are assigned the value 0, which means that the species was not searched for.

3 Pre-Validation Algorithm

Every observation submitted is sent through the pre-validation algorithm (Figure 1).

High-Alert Species

The first module in the algorithm, where every observation has to be forwarded to a professional, looks for species. All observations are flagged for expert validation and do not go through the rest of the components.

In their respective settings, some species can be marked as "high-alert". Any observation of such a species will be sent for expert validation (Figure 1 [A]). In addition, experts may receive notification via email as an immediate validation and response might be vital in such cases.

A species may be stamped as "high-alert" because:

- it is of particular interest to scientists
- it is dangerous to the environment
- it is highly invasive
- it has a high impact on human health or the economy.

Furthermore, this part of the process is carried out in accordance with Article 16 (Early detection notifications) of Regulation (EU) No. 1143/2014. The article states that member states have to notify the European Commission of any sighting of an alien species whose presence was not previously known in their territory (Council of the European Union, 2014).

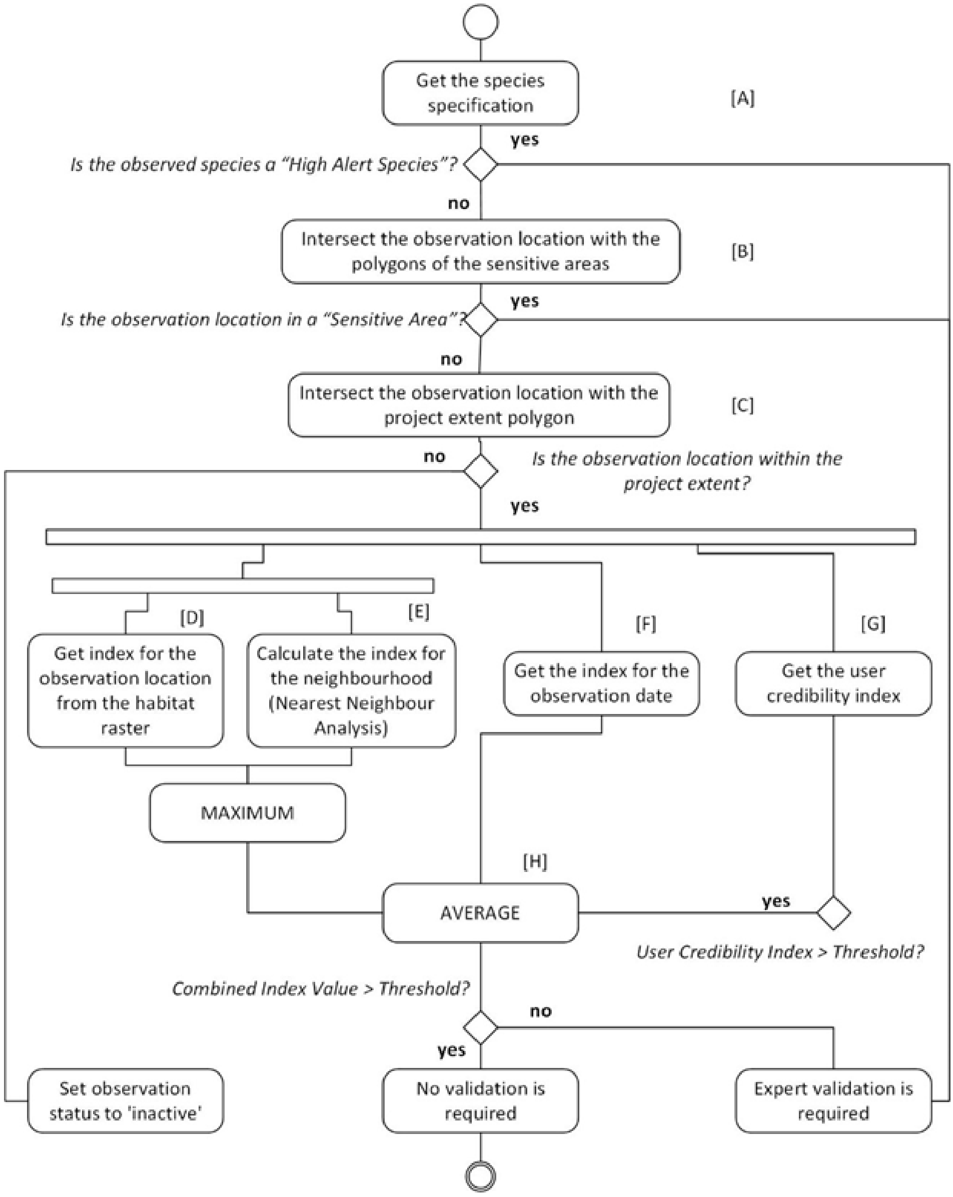
Get the species specification
[A]
Is the observed species a "High Alert Species"?
yes
no
Intersect the observation location with the polygons of the sensitive areas
[B]
Is the observation location in a "Sensitive Area"?
yes
no
Intersect the observation location with the project extent polygon
[C]
Is the observation location within the project extent?
no
yes
[D]
Get index for the observation location from the habitat raster
[E]
Calculate the index for the neighbourhood (Nearest Neighbour Analysis)
[F]
Get the index for the observation date
[G]
Get the user credibility index
MAXIMUM
[H]
AVERAGE
yes
User Credibility Index > Threshold?
Combined Index Value > Threshold?
no
yes
Set observation status to 'inactive'
No validation is required
Expert validation is required

Figure 1: Activity Diagram, Pre-Evaluation Algorithm

Within the pre-evaluation algorithm, the question regarding high-alert species deliberately occurs before the location of the observation is checked against the project's geographic extent (i.e. Austria). Invasive alien species are, after all, a cross-border challenge and affect not just a single country. Article 22 (Cooperation and coordination) of Regulation (EU) No. 1143/2014 instructs every member state to make every effort to ensure close coordination with other member states, especially if they share borders or belong to the same biogeographical region (Council of the European Union, 2014).

If an observation from outside the area covered by the project is sent via the website, an expert is able to evaluate the data, contact the user who sent it, and forward both the pictures and data to colleagues abroad and the authorities concerned.

Sensitive Areas

There may be areas where any observation of an alien species is of particular interest. Any sighting of a species in a sensitive area will immediately be flagged for expert validation (Figure 1 [B]). It is possible to define sensitive areas for every species and check the observation points against these areas.

Such areas could be:

- plantations, farms, etc.
- protected sites and sanctuaries
- habitats that house endangered or rare species
- highly contaminated areas
- sites with unique/specialized/rare ecosystems.

This section of the algorithm also occurs before the observation is checked against the extent of the project's geographical area. However, it will only have an impact if sensitive areas themselves are defined beyond the project scope — for example, transnational protected sites or national parks. Observations relating to such areas can be evaluated by a project expert and if necessary — and in compliance with the Article on "Cooperation and coordination" of the EU regulations regarding alien species (Council of the European Union, 2014) — forwarded to the authorities abroad.

Project Extent

The coordinates of the observation will be intersected with the polygon of the project extent (i.e. Austria) (Figure 1 [C]). If the point lies outside the project polygon, the observation will be flagged as inactive but still be kept in the database.

Observation Site

This section contains two blocks that are designed to produce a one-dimensional parameter for a spatial point by answering the following question: how likely is an observation in the area? The habitat module matches the observation coordinates against the raster of a habitat model. The neighbourhood module checks for similar sightings within a specified distance (which reflects, for example, the maximum home range of a species) of the observations. However, only the module that produces the higher index value will be used for the calculation of the overall value.

If, for a particular area, the habitat value is repeatedly smaller than the neighbourhood value, it might either indicate an insufficient habitat model or be the first evidence of a species extending its known habitat into a new territory. The habitat model itself acts as a safety net, especially during the early stages of the project when the density of observation points and known sightings is low.

Habitat Module

Habitat maps have to be modelled for each species (for this project, the habitats were modelled using geostatistical methods). The index values have to be scaled to decimal numbers ranging between 0.0 and 1.0. The numbers indicate how likely it is for a species to grow or live in a certain area. The value 0 means that the location is considered inadequate for the species, while an index of 1 indicates a very suitable location or habitat for the species (Figure 2). Those models have to be prepared by experts, taking known characteristics of a species into consideration.

The precision of the model should be proportional to the quality of the input data as well as the information about the existing potential habitats for the species. The data collected during the monitoring project can later help to enhance or review the existing models.

A raster of the model is transferred into a table in a spatial database. For each observation, the value for the given coordinates is extracted from the raster (Figure 1 [D]).

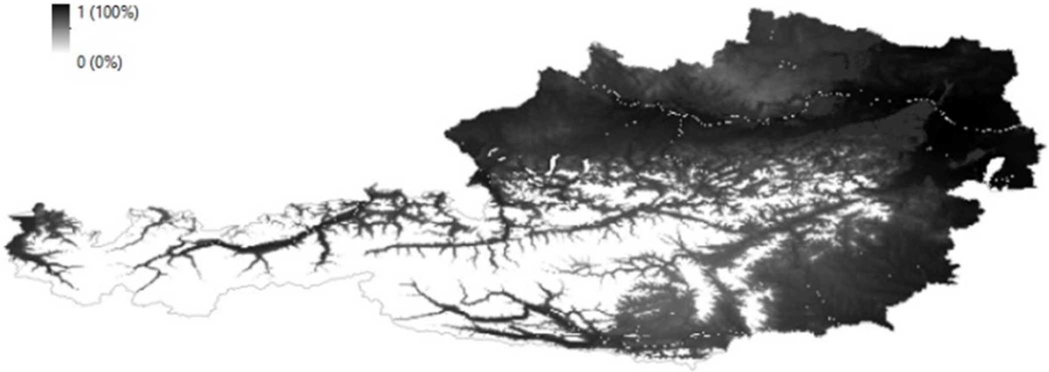

Figure 2: Habitat Model for Robinia pseudoacacia in Austria

Neighbourhood Module

Here the algorithm looks for sightings of the same species within a specified distance (the species action radius (r)) from the observation point (Figure 1 [E]). It checks for other observations as well as known occurrences that stem from other sources (e.g., open data tree maps). The sighting that is closest to the new observation (d_min) is used to calculate the probability value:

P(r,d_min)= (r-d_min)/r

The value derived indicates whether the observation may be connected to other occurrences of the same species in the vicinity (i.e. through seed dispersion or habitat size).

Observation Date

This module checks whether it is likely that the species was correctly identified on the sighting date (Figure 1 [F]).

This block matches the observation date against discrete probability values between 0 and 1 for each month, which indicate the likelihood of observing and correctly identifying the species at that time. While it may be very easy to correctly identify blooms during summer, it is unlikely that citizen scientists would be able to observe most flowers or hibernating animals during winter.

When using discrete values, there could be a considerable gap between the values of neighbouring months. However, even finer intervals cannot guarantee a better approximation of the actual number, as the current state of development of the plant or organism will depend heavily on local climate and location characteristics.

For larger surveillance areas, it may be necessary to generate different date-related indices for different sub-regions.

User Credibility

The last parameter reflects a user's credibility. A table in the database stores a user's observations by species, awarding 1 point for each categorization that has been confirmed by an expert. The maximum score is 100 credibility points per species. If a user's score is higher than the species-specific credibility threshold number, the score, divided by 100, will be taken into account for the combined index.

A total of less than the threshold number of accurate observations will not have any impact on the overall probability number, since many potentially knowledgeable users will only sporadically upload data.

Combined Index

The indices for the observation location, date and, if applicable, user credibility are combined as an arithmetic mean (Figure 1 [H]):

$$x_{combined} = \begin{cases} \frac{1}{3}(x_{location} + x_{date} + x_{user}) & x_{user} > x_{threshold}(Species) \\ \frac{1}{2}(x_{location} + x_{date}) & x_{user} \leq x_{threshold}(Species) \end{cases}$$

All the components are scaled to the same interval, where a value of 1 says that a species is very likely to be correctly identified at this time or by this user, and that it is very likely to grow (or live) at the location submitted. An index value of 0 indicates that it would have been impossible to observe a species, or that the location is not a suitable habitat. The threshold element marks for expert validation any observation with a combined index that falls below the species-specific threshold value.

Additionally, 1% of all observations will be randomly flagged for expert verification.

A feedback system was not included for the prototype model, but would be recommended so that users receive notification if their observation is confirmed by an expert, or an explanation if their observation has been reclassified or rejected.

4 Data evaluation

During a test period of 4 months (May – August 2015) for the project prototype, 488 observations were submitted to the project by 12 participants who had no advanced botanical or zoological training or experience in identifying neobiota. All observations were located in the easternmost part of Austria, where black locust and the tree of heaven in particular show very high habitat indices in their respective habitat models.

Four participants managed to achieve a user credibility index above the threshold level for at least one species. Their scores improved the combined pre-validation value for 226 further submissions for those species.

On 41 occasions, the neighbour index value was higher than the habitat value.

5 Conclusion and Outlook

The uncontrolled import of animals and plants over the centuries has shaped today's world, bringing with it access to a wide range of resources and food supplies. But we also have to deal with the impacts of new pests, diseases and invasive species that have come in its wake. Monitoring potentially dangerous and invasive species will help us to establish an early warning system, which could make early responses and counter-measures to threats as effective as possible.

The cornerstones of successful citizen science projects are the number of volunteers and the quality of the data that they submit. For this project, the latter is met by providing simple technology in combination with an algorithm that validates the incoming data through a number of derived index values. The experts on alien species who are involved in the project do not have to evaluate every observation, but only those that fail to meet the criteria set by them. This will reduce the overall costs and make it easier to design this as the long-term project which the EU regulation asks for.

References

Council of the European Union (2014). Regulation (EU) no. 1143/2014 of the European Parliament and of the Council of 22 October 2014 on the prevention and management of the introduction and spread of invasive alien species. Official Journal of the European Communities. Available online at eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2014:317:FULL&from=EN [accessed 2015–10–01]

Essl, F. & Rabitsch, W. (2002). Neobiota in Österreich. Vienna, Austria: Umweltbundesamt

INSPIRE Thematic Working Group Species Distribution (2013). Inspire data specification on species distribution – technical guidelines. Available online at http://inspire.ec.europa.eu/file/1526/download?token=agVvUHPi [accessed 2015–10–28]

Jarvis, R. M., Breen, B. B., Krägeloh C. U. & Billington, D. R. (2015). Citizen science and the power of public participation in marine spatial planning. Marine Policy (57), pp. 21–26

Roy, H., Pocock, M., Preston, C., Roy, D., Savage, J., Tweddle, J. & Robinson, L. (2012). Understanding Citizen Science Environmental Monitoring. Final report on behalf of UK-EOF. Available online at https://www.ceh.ac.uk/sites/default/files/citizensciencereview.pdf [accessed 2015–10–28]