The Role of Geographic Information Systems for Analyzing Infestations and Spread Of Invasive Termites (Isoptera: Rhinotermitidae And Termitidae) in Urban South Florida

Authors: Hartwig H. Hochmair, Francesco Tonini, and Rudolf H. Scheffrahn

Source: Florida Entomologist, 96(3) : 746-755
Published By: Florida Entomological Society
URL: https://doi.org/10.1653/024.096.0307
THE ROLE OF GEOGRAPHIC INFORMATION SYSTEMS FOR ANALYZING INFESTATIONS AND SPREAD OF INVASIVE TERMITES (ISOPTERA: RHINOTERMITIDAE AND TERMITIDAE) IN URBAN SOUTH FLORIDA

HARTWIG H. HOCHMAIR, FRANCESCO TONINI AND RUDOLF H. SCHEFFRAHN
University of Florida, Fort Lauderdale Research and Education Center, 3205 College Avenue, Davie, Florida, U.S.A. 33314

*Corresponding author; E-mail: hhhochmair@ufl.edu

†Summarized from a presentation and discussions at the “Native or Invasive - Florida Harbors Everyone” Symposium at the Annual Meeting of the Florida Entomological Society, 24 July 2012, Jupiter, Florida.

ABSTRACT
The ability to manage geospatial data has made Geographic Information Systems (GIS) an important tool for a wide range of applications over the past decades, including management of natural resources, analysis of wildlife movement, ecological niche modeling, or land records management. This paper illustrates, using invasive termite species as examples, how GIS can assist in identifying their potential sources of infestations and model their spread in urban South Florida. The first case study shows that the Formosan subterranean termite, Coptotermes formosanus Shiraki, and the Asian subterranean termite, Coptotermes gestroi (Wasmann) (Isoptera: Rhinotermitidae), were introduced into and dispersed across South Florida by sailboats and yachts. The second case study shows an agent-based model to simulate the natural spread of Nasutitermes corniger (Motschulsky) (Isoptera: Termitidae) in Dania Beach, Florida. This paper provides an overview of basic functionalities in GIS and demonstrates how they can be customized for advanced modeling and simulation.

Key Words: Spatial analysis, GIS functionality, modeling, South Florida, exotic termites

RESUMEN
La capacidad de manejar datos geoespaciales ha hecho de los Sistemas de Información Geográfica (SIG) una herramienta importante para una amplia gama de aplicaciones en las últimas décadas, incluyendo el manejo de los recursos naturales, análisis de movimiento de la fauna, modelos de nichos ecológicos, o manejo de los registros de terrenos. Se ilustra, utilizando especies invasoras de termitas como ejemplos, cómo el SIG puede ayudar a identificar las posibles fuentes de infestación y modelar su diseminación en las zonas urbanas del sur de la Florida. El primer caso de estudio muestra que la termita subterránea de Formosa, Coptotermes formosanus Shiraki y la termita subterránea asiática, Coptotermes gestroi (Wasmann) (Isoptera: Rhinotermitidae), fueron introducidas y dispersadas en todo el sur de la Florida por medio de los barcos veleros y yates. El segundo caso de estudio muestra un modelo basado en agentes para simular la diseminación natural de Nasutitermes corniger (Motschulsky) (Isoptera: Termitidae) en Dania Beach, Florida. Este documento provee una visión global de las funciones básicas de SIG y demuestra la forma como puede ser adaptado para producir modelos avanzados y simulaciones.

Palabras Clave: análisis espacial, funcionalidad GIS, modelos, sur de la Florida, termitas exóticas

Termites are destructive insect pests which cause billions of US dollars in structural damage and control within the United States alone (Su & Scheffrahn 1998). Worldwide, more than a dozen exotic termite species have become established (Evans 2011), of which 6 can be found in Florida (Scheffrahn et al. 2002). Typical research questions related to invasive species are: (Q1) How did non-native species invade and establish non-endemic populations? (Q2) How fast do invasive species disperse without human assistance within a new environment? In the case of termites, maritime transport has long been suspected to be responsible for the transport and nonendemic establishment of numerous termite species across ocean barriers (Gay 1967; Scheffrahn et al. 2009). Two invasive termites, the Formosan subterranean termite (FST), Coptotermes formosanus Shi-
raki (Isoptera: Rhinotermitidae), and the Asian subterranean termite (AST), *Coptotermes gestroi* (Wasmann), are now well-established pests in urban South Florida. The paper by Hochmair & Scheffrahn (2010) illustrates how a Geographic Information System (GIS) was utilized to show with spatial analysis that these two invasive species were introduced through boat traffic. In a second example herein, we demonstrate how GIS functions together with an agent-based model simulate the natural dispersal of an arboreal invasive termite, *Nasutitermes corniger* (Isoptera: Termitidae), in Dania Beach, Florida. The focus of this paper is on describing how a GIS can be used to run spatial analyses in order to answer the aforementioned research questions.

Geographic Information Systems

A GIS comprises an integrated suite of software components containing (1) a data management system, (2) a mapping system for display and interaction with maps, and (3) a spatial analysis and modeling system (Longley et al. 2011). It uses georeferenced data, that is, having unique location information, such as postal addresses, or point coordinates. Geovisualization is used to explore, analyze, and present spatial data, however, a GIS mapping system also supports on-screen digitizing of spatial features on top of background maps. Point, line, and polygon object features are often displayed over a base map, e.g. a satellite image, which can be provided through a Web mapping service within the GIS. Geovisualization often involves a map projection, which is a process that transforms the spherical Earth’s surface into a plane (Slocum et al. 2005). The resultant map allows use of a Euclidean coordinate system with Eastings and Northings instead of geographic coordinates with longitude and latitude. This simplifies the computation of distance, direction, and area compared to spherical geometry.

A raster dataset represents geographic features by dividing the Earth into discrete square or rectangular cells laid out in a grid. Cells, also called pixels, are arranged in rows and columns, and each cell has a value that is used to represent some characteristic of that location. Resampling of both background maps and raster datasets in general is necessary whenever raster data must be transformed to another coordinate grid system or the cell size between input and output raster changes, like in the case of registering remotely sensed data to a ground coordinate system. During this process cell values in the new grid are filled with cell values derived from the original grid using a resampling technique. Resampling is also necessary in the context of map projections. Let us assume that a background image (Fig. 1) is provided in Albers Equal Area Conic projection for North America, as shown to the left. The image has a pixel resolution of approximately one meter, and Eastings and Northings are given in meters. Let us further assume that the GIS uses the Universal Transverse Mercator (UTM) projection (Zone 17) to display the image together with other data layers. For this task, the background image needs to be re-projected to UTM. The conversion between these 2 projected grids is done through different sets of equations, where projected coordinates are converted to geographical coordinates. For the output image in UTM projection, pixel brightness values need to be determined for each pixel from the input image (Albers projection). Since there is no direct one-to-one relationship between pixels of the input and output image, the output image often requires a value from a location of the input pixel grid that does not fall neatly on a cell center. This is illustrated in a zoomed portion of the images, shown with a highlighted grid cell as an example (lower right portion of Fig. 1), where the brightness value of a pixel in the pool area is sought (yellow dot). Using mapping equations, the point coordinate of that cell center in the output image can be converted back to the point coordinates in the coordinate system of the input image (dashed arrow), giving the position indicated by the brown dot. Since the brown dot in the input image is not on a cell center, a mechanism for determining the brightness value from neighboring cells is used. This mechanism is called intensity interpolation and is the core of resampling techniques (solid arrow). Widely used interpolation methods include the nearest neighbor interpolation and the cubic convolution (Jensen 2005).

GIS Analysis

Two conceptual schemas are used to represent the Earth, which are (i) discrete object view and (ii) continuous-field view. Both schemas have implications on the GIS data models used for GIS analysis. In the discrete object view the world is empty except where occupied by stationary or moving objects with well-defined boundaries, including lakes, roads, buildings, or animals. While this schema works for many everyday applications, it becomes difficult to provide definitions for all kinds of objects to be mapped, e.g. to distinguish between a hill and a mountain. In the discrete object view geographic objects are defined by their dimensionality and represented in the GIS as a vector data model. The vector data model uses points and their x-, y- coordinates to construct polygons (for area-like objects such as counties), lines (for linear features such as roads), and zero-dimensional points (such as termite-infested boats). In the continuous-field view, the world is described by a number of variables where each variable can be measured at every position on Earth. This conceptual scheme is realized as
raster data in a GIS, where a surface is overlaid with a raster grid that has attributes assigned to its cells, such as elevation or land cover class. The large number of spatial analysis functions in a GIS can be divided into (1) analyses based on a single location, and (2) analyses based on distance between separate places (Longley et al. 2011). The first group compares different properties of the same place and calculates relationships and correlations between them. For example, this...
includes the analysis of attribute tables, spatial joins, point-in-polygon operations, polygon overlays (such as Union and Erase), and raster analysis. The Union operation creates a new layer of vector polygons (called coverage) by overlaying two input polygon coverages. The resultant coverage contains the combined polygons and attributes of both input coverages. The Erase operation removes the part inside of the first coverage which is covered by the outline of the overlaid coverage. The second group of analysis functions includes the measurement of distances between points, buffering, cluster detection, computation of autocorrelation, density estimation, and spatial interpolation. All these methods can be combined into more complex tasks both in the raster and vector data model. One example is surface analysis, which includes the computation of both slope and aspect of a surface grid cell, finding the least cost path between two grid cells, delineation of watersheds, or determining viewsheds (a viewed area is an area of land or water that is visible to the human eye from a fixed vintage point) based on topography. Another example is network analysis which includes routing and logistics problems in transportation networks, such as optimizing the routing of delivery trucks. Observation of social insect behavior has allowed computer scientists and engineers to improve network based optimization algorithms, such as the Travelling Salesman Problem which consists of finding the shortest tour between a given set of cities visiting each city once only and ending at the starting point. This problem has been tackled by the use of artificial pheromones (with a decay of the pheromone concentration over time) that artificial ants use to mark travelled paths along completed routes. This approach has been adopted from ant colony behavior, where ants use pheromones during foraging for food to collectively discover the shortest path between nest and food source (Bonabeau et al. 2000). A GIS provides sampling tools for spatial statistical testing, e.g., generating random points within a pre-defined area (de Smith et al. 2010).

GIS Modeling

A GIS can also be used to create and visualize dynamic simulation models. A simulation shows the evolution of the phenomenon of interest through time and may involve multiple sub processes. Dynamic modeling allows scientists to experiment with policy options and what-if scenarios. It also allows them to implement ideas about the behavior to the world (Longley et al. 2011). Typical case study applications include: Planning for emergency evacuations, e.g. in the case of hurricanes or wildfires; urban growth scenarios and its impact on food resources and environment; assessment of the effect of planning policies on deforestation area; modeling competition for canopy space in forest ecosystems for better informed silvicultural prescriptions.

Widely used model types include: (a) analytical models, e.g. diffusion-type processes, which use differential equations (Holmes et al. 1994); (b) agent-based models (ABM) (Judson 1994), a.k.a. individual based models (IBM), which study the fate and movement of single individuals using both physiological and behavioral rules; (c) landscape models (Mladenoff 2004), which consider each cell as group of individuals; and (d) cellular automata models (Ermentrout & Edelstein-Keshet 1993), which represent the surface of the earth as a raster where each cell has a fixed number of states that change through transition rules based on each cell’s neighborhood.

Since these models are usually more complex than what is provided through standard GIS functionality, they need to be implemented from scratch or through customization of existing functions. Customization is the process of modifying GIS software through adding new functionality, embedding GIS functions to other applications, or creating specific-purpose applications (Longley et al. 2011). Numerous programming languages, such as C, C++, C#, Java, or Python are available for customizing both desktop GIS software and Web GIS applications (Zaragozá et al. 2012; Zandbergen 2013). Integrated development environments (IDEs) combine various software development tools, including a visual programming language, an editor, and a debugger. To support customization, a vendor must expose details on the software’s functionality to the developers. A key feature of such software components is that they have well-defined programming interfaces that allow the functionality to be called by programming tools in an IDE. One example is ESRI’s ArcObjects model and all its functionality which can be accessed through any programming language that supports the Microsoft Component Object Model (COM), such as VB.NET, C#, C++, or Java. The R programming language for statistical computing (R Development Core Team 2012) has recently improved its spatial functionalities and added a package called RPYGeo providing access to most of ESRI’s ArcGIS geoprocessing tools from within R. Within the ArcGIS software suite, customization can also be done with the Model-Builder. This allows the user to build a customized workflow of geoprocessing operations from existing tools using a graphical interface. The same can also be accomplished by using Python scripting. Finally, geospatial analyses and a number of image processing tasks can also be carried out using either open source software, e.g. R, GRASS GIS, SAGA GIS, or GeoDa, or proprietary software, e.g. Matlab, SAS, ENVI, or ERDAS. These software platforms typically exchange data through common GIS formats, such as shapefiles or geotiffs.
MATERIALS AND METHODS

For enhanced clarity, the figures in this report are also displayed online in color in supplementary material for this article in Florida Entomologist 96(3) (2013) at http://purl.fcla.edu/fcla/entomologist/browse.

Termite Samples

All termite samples used for analysis in the two case studies are preserved in 85% ethanol and are deposited in the University of Florida termite collection at the Fort Lauderdale Research and Education Center in Davie, Florida. Samples were collected by operatives in the pest control industry and property owners and submitted to R.H.S. The collection database records contain among other data the genus and species, geographic latitude and longitude of the termite infestation, collection date, and collecting conditions. For the analysis of the first case study, samples were used with collection dates ranging from Apr 1996 to Mar 2009 for AST, and from May 1990 to Jul 2008 for FST. For that study, termite samples were externally submitted for identification and location georeferencing, but no additional field surveys were conducted.

For the second case study, collection dates of used samples for N. corniger ranged between Jan 2003 and Feb 2012. From early 2003 until early 2011, a previously delineated area was targeted for a deliberate eradication campaign of this invasive pest. Sample locations were recorded in the field using a GPS device and later imported into a database. The surveyed area covered around 200 acres (81 ha). In 2012, new infestations were found in areas that were not surveyed since 2005 due to cancellation of project funding, providing some insight into the dynamics of natural dispersal.

Case Study 1: Infestation Source Analysis

The first GIS example comes from a study that compares the distances between 190 terrestrial point records for FST (first infestation reported 1980 in Broward County), 177 records for AST (first discovered 1996 in Miami), and random points locations in the surrounding urban areas to the nearest marine dockage (Hochmair & Scheffrahn 2010). The hypothesis is that both species are significantly closer to potential infested boat locations, i.e., marine docks, than random points in these urban areas. It is further hypothesized that a larger median distance between FST infestations and proximal dockage can be observed than for AST.

Using the ArcGIS software suite and as depicted in Fig. 2, the two point sets of termite sightings were first projected from geographic to UTM coordinates. Next, the study region was tessellated (i.e., divided into non-overlapping squares), where squares containing a dockage location, assessed through a background image on the screen, were marked as having a dock. Then a random point pattern in urban areas was generated. Urban areas as defined by the U.S. Census Bureau were utilized. Urban areas can be split into two categories, which are Urbanized Area (UA) and Urban Cluster (UC) (U.S. Census Bureau 2002). Only those UA and UC polygons for which at least one termite collection was recorded, serve as the reference area for the generation of a spatially random point pattern. Finally, for each point in both termite point sets and the random point set the closest dock location (i.e., the nearest center of a square dock polygon) was identified and the straight line distance determined. These last 2 steps were accomplished through the Spatial Join function in the GIS.

Case Study 2: Spread Model

The second case study combines GIS functionality with a computer simulation (Fig. 3) that uses an individual based model to predict the dispersal of N. corniger. A sample of 189 termite sightings between Jan and Apr 2003 in Dania Beach, FL (Tonini 2013) was used as the starting point for the simulation which was run for 10 yr between 2003 and 2012. The simulation algorithm is realized through a set of R functions that implement an individual based model (IBM) for natural termite dispersal. The model considers a variety of biological parameters, such as overall survival rates of alates, mean dispersal flight distance, age of colony maturity, and maximum density of colonies per hectare. Before the simulation, ArcGIS was used to determine areas suitable for the establishment of a new colony. A Union overlay operation (Fig. 3A) was applied to polygon coverages showing hydrographic features, buffered roads, airport runways, and agricultural fields, which gave an unsuitable area coverage. By means of the Erase operation (Fig. 3B), this coverage was converted to a suitable area coverage, which was utilized in the simulation (Fig. 3C). While overlay functions in ArcGIS provided the suitable area polygon layer, a set of R packages (libraries), including ‘sp’, ‘spatstat’, ‘rgdal’, and ‘raster’, was used in order to carry out further basic geoprocessing operations in preparation for the simulation. These included point-in-polygon overlays (to make sure that termite colonies do not fall within unsuitable habitat), the creation of a reference simulation grid over the study area (to control the maximum density of termite colonies), point-to-raster conversions (to represent the approximate area covered by one or more colonies based on the aforementioned simulation grid), and raster overlays (to count the number of times a given pixel was occupied after 100 Monte Carlo simulations).

To assess the sensitivity of the model with respect to model parameters, the simulation was run with one or two alternative values for each parameter, giving a total of 12 alternative model
Fig. 2. GIS operations involved in distance analysis of termite collection sites and random points to nearest dockage: A) Point coordinate conversion from geographic to UTM coordinates; B) Digitize dock locations and generate random point pattern in urban areas; C) Determine distance to nearest dockage through Spatial Join function, as illustrated for selected termite and random points.

Projection, mapping

Digitizing, random points

Spatial join
realizations in addition to the baseline simulation. The uncertainty associated with the outcome of a stochastic simulation was estimated through the Monte Carlo technique and 100 simulation runs. The GIS mapped the simulation result of the baseline model for 2012 on a background im-

Fig. 3. Spatial analysis functions supporting termite spread model: A) Union overlay to determine unsuitable area; B) Erase operation to identify suitable area; C) Suitable area coverage together with 2003 termite collection sites used as input for spread model in R, with mapped model result to the right.
age provided through a Web service (Fig. 3C), illustrating the spread of termite colonies from original points of infestation. Yellow, orange, and red cells indicate the >0%, >=50%, and 100% occupancy envelopes, respectively. The 100% occupancy envelope groups areas that are predicted as infested by all runs.

Statistical Analysis

Case study 1 resulted in three sets of distances which are Euclidean (straight line) distances to the nearest dockage for FST, AST, and the random points, respectively. A one-sample Kolmogorov Smirnov test indicated that distances in each set were not normally distributed. Therefore the non-parametric Mann-Whitney U test was used to compare median distances between the three distance sets.

Data analysis in case study 2 had two goals. The first goal was to validate the model fitness by comparing the predicted infested area for 2012 with actually observed infestation data occurring up to 2012. The predicted area was visualized through visualization of >0%, >=50%, and 100% occupancy envelopes resulting from a Monte Carlo simulation. This was visually compared to 74 new infestation sightings in 2012, all of which were found in previously Insecticide-untreated areas. The second goal was to conduct a sensitivity analysis for 12 alternative model realizations in addition to the baseline simulation to assess the effect of model parameters on predicted infested area. The sensitivity analysis comprises relative and absolute growth rates in infested area compared to the base line model for a given year between 2004 and 2012, and also a comparison of total predicted infested area for the different years.

RESULTS

Case Study 1: Infestation Analysis

Table 1 provides the descriptive statistics for distances associated with the 3 point patterns. Mann-Whitney U tests showed that the median distances to nearest docks associated with AST and FST, respectively, were significantly smaller than for the random point set ($p < 0.0001$, 2-tailed) for AST and FST), indicating that AST and FST are significantly closer to potential infested boat locations, i.e., marine docks, than random points in considered urban areas. A Mann-Whitney U test further showed that observed median distances to nearest docks were significantly smaller for AST than for FST ($p < 0.0001$, 2-tailed). Since FST was discovered in Florida about two decades before AST, larger median distances for FST than for AST can be expected if assuming that these two invasive termite species were first establishing near boat dockage, and with later generations, colonizing areas further away. In summary, statistical comparison of median distances provides strong evidence that the two exotic termite species were introduced and spread by boat in South Florida. A more detailed discussion of analysis results can be found in (Hochmair & Scheffrahn 2010).

Case Study 2: Spread Model

To validate the simulation model, the locations of newly infested sites from 2012 were compared to the occupancy envelopes which are based on 2003 sample sites as initial infestation points. Fig. 4A shows the infested areas predicted by the baseline simulation model with in all three occupancy envelopes (Tonini 2013). The 100% occupancy envelope overlaps well with the 2012 sighted collection points, while the > 0% and >= 50% envelopes overestimate termite spread.

Sensitivity analysis (Fig. 4B) found that the prediction model was most sensitive to the following parameters: Number of alates generated by a colony over the colony’s lifetime, survival rate of alates, maximum mating pheromone attraction distance, and mean dispersal flight distance. Only small effects were observed for the following parameters: prevalence of male alates in the colony, age of first production of alates, density of colonies/ha. While the solid line in Fig. 4B indicates the predicted infested area when using a parameter baseline value of 200 m for mean flight distance of alates, the dashed lines show the effect on the predicted infested area when changing this parameter to 100 m and 300 m (Tonini 2013). The results of this stochastic individual-based simulation model provide a means for regulatory agencies to anticipate possible areas of infestation. It must be noted that

<table>
<thead>
<tr>
<th>Table 1. Distance of Land-based Infestations of Coptotermes gestroi and Coptotermes formosanus to Nearest Marine Docks (in Meters).</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Standard deviation</td>
</tr>
<tr>
<td>Median</td>
</tr>
</tbody>
</table>
it is crucial to calibrate model parameters if the spread of other species is modeled, and that the current model does not take into account anthropogenic dispersal.

**SUMMARY**

A Geographic Information System (GIS) is a useful tool in various stages of analyzing the infestation and spread of invasive termites. The 2 examples herein illustrate how mapping and spatial analysis operations in the GIS, including map projection, on-screen digitizing, spatial join, and overlay functions, facilitate analysis in concert with external statistical software packages. Due to continued advances in spatial analysis, improved customization capabilities of GIS functions, and steadily increasing computer processing power, the role of GIS for spatial analysis of invasive pests is likely to increase further in the near future.

**REFERENCES CITED**


