

Research Article

## Invasionsoft: A web-enabled tool for invasive species colonization predictions

Matthew W. Johnston\* and Samuel J. Purkis

National Coral Reef Institute, Nova Southeastern University Oceanographic Center, 8000 North Ocean Drive, Dania Beach, Florida 33004, USA

E-mail: [johnmatt@nova.edu](mailto:johnmatt@nova.edu) (MWJ), [purkis@nova.edu](mailto:purkis@nova.edu) (SJP)

\*Corresponding author

Received: 13 September 2011 / Accepted: 6 March 2012 / Published online: 22 March 2012

### Abstract

Predicting and averting the spread of invasive species is a core focus of resource managers in all ecosystems. Patterns of invasion are difficult to forecast, compounded by a lack of user-friendly species distribution model (SDM) tools to help managers focus control efforts. This paper presents a web-based cellular automata hybrid modeling tool developed to study the invasion pattern of lionfish (*Pterois volitans/miles*) in the western Atlantic and is a natural extension our previous lionfish study. Our goal is to make publically available this hybrid SDM tool and demonstrate both a test case (*P. volitans/miles*) and a use case (*Caulerpa taxifolia*). The software derived from the model, titled *Invasionsoft*, is unique in its ability to examine multiple default or user-defined parameters, their relation to invasion patterns, and is presented in a rich web browser-based GUI with integrated results viewer. The beta version is not species-specific and includes a default parameter set that is tailored to the marine habitat. *Invasionsoft* is provided as copyright protected freeware at <http://www.invasionsoft.com>.

**Key words:** invasive species; species distribution models; lionfish; cellular automata; *Caulerpa taxifolia*; web portal, software

### Introduction

Predicting colonization sequences of biological invaders has become a core focus of resource managers as invasions become more commonplace. The patterns of introduction and recruitment are hidden from casual view and often not well understood. However, the potential deleterious effects of non-native species in all habitats have been well documented (O'Neill 1997; Dittel and Epifanio 2009). Compounding this lack of understanding is a shortage of user-friendly tools for resource managers faced with newly introduced biological invaders.

This paper focuses on a new web-based simulation tool whose base algorithms were developed during a previous study (Johnston and Purkis 2011) to investigate the initial release and subsequent spread of *Pterois volitans* (Linnaeus, 1758) and *P. miles* (Bennett, 1828) (red lionfish and devil firefish, referred to collectively as lionfish), two invasive reef fish species which have become established in the eastern Atlantic and entire Caribbean probably through aquarium releases (Hare and Whitfield 2003). This paper is

a logical extension of that study. The lionfish itself is a voracious consumer of reef fish and crustaceans up to their own body size (Hare and Whitfield 2003; Morris and Atkins 2009). They have few predators in their introduced environment and are highly fecund, producing free-drifting larvae distributed on currents (Fishelson 1997; Freshwater et al. 2009). They are extremely successful invaders in the western Atlantic and Caribbean regions. We will also examine another successful invasive species, *Caulerpa taxifolia* ((M.Vahl), C.Agardh, 1817), which is a marine alga that is native to the tropical Atlantic, Indian, and Pacific oceans and was introduced to the Mediterranean Sea in the 1980s (Phillips and Price 2002). The introduced strain, linked to the aquarium industry, is highly robust and reproduces asexually via fragmentation (Boudouresque et al. 1995; Phillips and Price 2002). Our main objective with this paper is to provide an easy to use and accessible hybrid modeling tool for scientists or resource managers studying invasions that can be adapted to a specific study species. Herein we intend to demonstrate the use of *Invasionsoft* by providing two example modeling scenarios;

simulation creation based on the lionfish invasion test case, and an example use case with *Caluherpa taxifolia* as a sample species. We also present the methods used to verify model outputs and how an end user can utilize Invasionsoft to verify their own models.

### 1.1 Species Distribution Models (SDMs)

The use of species distribution models (SDMs) for invasive species biology is becoming more prevalent with the advent of such software packages as openModeller and Maxent (Phillips et al. 2004; Muñoz et al. 2009). These tools largely focus on the overall potential distribution of a species by algorithmic examination of environmental conditions to produce model types including the habitat suitability models (HSMs) used by Guisan and Thuiller (2005) and Franklin (2010). Recent aquatic modeling by Jacobs and MacIsaac (2009) have used a gravity model approach to predict the spread of *Cabomba caroliniana* (Green Cabomba) via active (human-mediated) and passive (advective flow) movement of propagules into relatively closed aquatic systems. The invasive vegetative weed *Myriophyllum spicatum* L. (Eurasian watermilfoil) has also been investigated using gravity models, which uncovered substantial deficiencies in this approach resulting in unreliable predictions (Rothlisberger and Lodge 2011). Gallien et al. in 2010 provides a thorough review of SDM methodologies including both mechanistic and phenomenological models as well as newly emerging hybrid models which combine the two.

A significant shortcoming in the overall potential distribution approach employed by many of these SDMs is the lack of representation of the dynamics and mechanism of spread. Mechanistic models, such as the Mechanistic Niche Model (MNM) utilized by Kearney et al. (2008) examine the progression of invasions, but often ignore environmental factors (Gallien et al. 2010). This has resulted in the movement toward development of hybrid models combining mechanistic and phenomenological approaches, the components of which are discussed in great detail by Gallien et al. (2010). Even with this movement, there remains a lack of readily accessible tools to easily create hybrid models. Invasionsoft engages such a spatial hybrid model, addressing both the dynamics of spread and overall spatial distribution through the use of a cellular automata (CA) algorithm-based

examination of environmental conditions. CA models have been previously employed in the study of invasive vegetation utilizing GIS (Cole and Albrecht 1999), and agricultural pests like *Ceratitidis capitata* (Wiedemann, 1824) (Mediterranean fruit fly) using AnyLogic™ (Parks et al. 2005). CA models have also been used to study the spread of a species and range expansion into adjacent habitats, often in the event of climate change events (Ostendorf et al. 2001; Engler and Guisan 2009; Wilson et al. 2009). The CA approach was chosen for our model because of its relative simplicity and proven success in a large variety of physical systems where examining simple parameters can often describe the behavior of a more complex system (Bolliger et al. 2003).

The Invasionsoft CA algorithm presented examines historical patterns of capture records combined with physical parameters present at the actual capture sites to produce a simulation replicating the spread of the study species. CA models in general consist of 4 elements; *conceptual cells*, *cell state*, *neighborhood* (the surrounding cells) and a *rule*. The study area is divided into a grid of equal sized cells (each a *conceptual cell*) which each contain discrete parameter values. One cell is initially marked colonized (the *cell state*) and subsequent cells in the *neighborhood* are marked colonized based on a predetermined set of conditions and stochastic random variable (the *rule*). For best results, our model relies on pattern fitting (producing a best-fit model – BFM) of a historical sighting or capture sequence to tune the inputs of the model. Models can also be produced without producing a BFM. When comparing the model output to a historical pattern, a resultant BFM can then be used to create simulations predicting future invasions in other locales. The model is simple but powerful and has been shown to be accurate in its predictive capabilities in the lionfish test study case (Johnston and Purkis 2011).

The initial version of the software contains a default parameter set, specifically tailored to the marine environment, and examines four prevalent physical parameters; salinity, temperature, depth, and ocean current in the western Atlantic, Caribbean, and Gulf of Mexico regions. Users can also upload custom parameter datasets, enabling modeling in other regions as well as terrestrial environments. Invasionsoft beta is unique in its simple web-based format, enabling the tool to be easily accessed with any modern web browser, requiring no software installation

and no cost. With Invasionsoft, we aim to bridge the gap between the need for more intuitive and simple hybrid modeling tools and more complex modeling tools like openModeller and Maxent.

## Methods

### 2.1 Technology

The Invasionsoft web portal was written primarily in ASP.Net technology, utilizing Visual Basic (VB.Net) as the main coding language of the presentation and data access layers. The web interface (Figure 1) uses JavaScript including AJAX technology for browser scripting requirements. Scaled Vector Graphics (SVG) are employed to display the results of the model in a custom viewer using Google Maps as a base layer to show spatial distribution of the resulting data points. Most algorithmic logic is performed via Structured Query Language (SQL) stored procedures and server-side VB.Net code which act upon a Microsoft SQL Server 2008 Express database. This database also houses all data and tables required for running the model. Some data processing was accomplished utilizing ArcMap 10 – especially in those cases which involved spatial referencing of parameter data points and lionfish record data points (ESRI 2011). The front end website is hosted on a Microsoft IIS server running the Windows XP operating system. To take full advantage of all of the technologies in the simulator, supported browsers for end users include Microsoft Internet Explorer version 8.0 or higher, Mozilla Firefox version 3.0 or higher, and all versions of Google Chrome. The Invasionsoft software can be accessed at the website <http://www.invasionsoft.com>.

### 2.2 High level model overview

When using the default parameter set, Invasionsoft examines 4 common oceanographic environmental characteristics; depth, temperature, salinity, and ocean current, to determine their effect on the distribution of an invasive marine species. These four default parameters are derived from the lionfish test case and were determined likely to be the most influential parameters on that species' invasion sequence (Johnston and Purkis 2011). Users can also opt to provide their own parameter datasets for evaluation by utilizing the custom dataset upload

feature. To run a simulation using the default parameter set, a user inputs values using the web-based interface (Figure 1) to define the acceptable range of values for temperature, salinity, and ocean depth, thus delineating a portion of the *CA rule*. An instructional tutorial is provided via an 'Instructions' link on the main page to guide the user in what input is necessary to run the model. Additionally, context-sensitive help is available for each input field by clicking on the "?" button next to the field. For the default parameters, the "?" button also presents valid value ranges. The parameter input values can be based on statistical analysis of actual historical records or other sources of data that define an invasive species' range tolerance to the parameters being examined. The user also assigns a weight factor (another portion of the *CA rule*) to each of these parameters. This weight factor is a proportional number to the other parameters' weights and is used to determine the influence that a parameter has on that cell (the *CA conceptual cell*, in the *CA neighborhood*) meeting the conditions for colonization (the *CA cell state*). A weight factor is also assigned to a 'null' cell in which none of the parameter conditions are satisfied and can be used to factor in a randomness value or account for influential parameters other than those currently being examined. The null cell weight value can also be set to zero to eliminate a null condition influence. Finally a 'required' value is assigned to the each parameter. If this value is checked, a cell is labeled as colonized (as opposed to a transport area) only if the checked parameter value falls within the acceptable range. In addition to the parameter values, the user is also required to enter a name for the simulation, the number of cycles to run the simulation (the model is iterative for this number of cycles, with the invasion algorithm run for each infected cell once per cycle), the coordinates of the initial introduction (vector) cell, and finally an email address to email the results which are contained in a .csv (comma separated values) file. Additionally, the output is displayed on a custom viewer in the web browser which includes the ability to animate the tracks from the initial introduction location.

To create simulations using a custom dataset, the user must compile their parameters into a .csv according to the format provided in the instruction file. Using the GUI the user will then choose the 'Use Custom Parameter Set' option and upload their pre-formatted .csv file.

**i**nvasionsoft Beta\*  
Simulations for Predicting Colonization of Invasive Species

[Click here for instructions/tutorial](#)

**Step 1** - Choose to use the default parameter dataset or define a custom parameter dataset. The default data set is a 1 x 1 degree dataset with monthly average variables for salinity and temperature, as well as depth and current values for a geographic area covering 4 to 45 N latitude and -100 to -50 W longitude.

Use Custom Parameter Set   Use Default Parameter Set   View Default Parameters   Reset Page

View Sample Simulation

**Step 2** - Provide all information required to run the simulation. When completed, click 'Validate Inputs'. If all data is complete and valid, the 'Run Simulation' button will become enabled. Click to run the simulation.

? Simulation Name:

? Cycles:  Simulation Quantity: 1

? Grid Size (degrees): 1   Set Default Values

? Vector Latitude:    Set Random Trajectory

? Vector Longitude:

? Current Weight:

? Null Weight:

Weight: ?   Lower Limit: ?   Upper Limit: ?   Required: ?

? Salinity (ppm):         

? Temperature (C):         

? Depth (m):         

Enter the email address where the results will be sent or your User ID:  ?

Message Center:

? Output File:    View   Validate Inputs   Run Simulation

Citation:  
Johnston, M.W., Purkis, S.J., 2011. Spatial analysis of the invasion of lionfish in the western Atlantic and Caribbean. *Marine Pollution Bulletin* 62 (6), 1218–1226. [Article](#)

\* Beta version. Please report bugs or issues.

[Contact](#)   [Copyright and License](#)

**Figure 1.** Invasionsoft Web Interface. User interface where parameter values and required inputs are entered and validated.

The software will auto-set the grid size being used (based on the values in the .csv) and provide text-box inputs for each parameter in the dataset (up to 10). The remaining steps to create a simulation are then identical to the simulation creation procedure using the default parameter set.

### 2.3 Parameter and record data compilation

The initial version of the Invasionsoft simulator (using the default parameter set) examines a

geographic area encompassing the western Atlantic Ocean, Caribbean Sea, and Gulf of Mexico from 45° – 5° N latitude and -100° to -50° W longitude, which corresponds to the approximate geographic extent of the current lionfish invasion. Parameter data were compiled on a 1° latitude × 1° longitude grid cell size (about 100 km × 100 km), with a total of 1,382 marine cells. Mean salinity and temperature data were obtained from the World Ocean Atlas 2005 (WOA05) database (Boyer et al. 2006). Values for water depth were obtained from the ETOPO1

1 Arc-Minute Global Relief Model which combines bathymetry and topography data based on underway hydrographic soundings and satellite altimetry estimates (Amante and Eakins 2009). Yearly average current data were obtained from the National Oceanic and Atmospheric Administration (NOAA) Ocean Surface Current Analysis – Real Time (OSCAR) database as well as from the NOAA Atlantic Oceanographic and Meteorological Laboratory (AOML) (Bonjean and Lagerloef 2002). Current velocity and current direction were calculated for each cell in the  $1^{\circ} \times 1^{\circ}$  grid based on the NOAA data. In those cases where NOAA current data were not available, direction and velocity were estimated from surrounding cells and prevailing currents. All default parameter data can be reviewed in a custom viewer from the main page via a button labeled ‘View Default Parameter Dataset’ (Figure 1).

For the lionfish test case, the United States Geological Survey – Non-indigenous Aquatic Species (USGS-NAS 2011) database was queried for historical lionfish capture records to use in development and verification of the model. The initial database contained 1,174 lionfish records at the time the test study began, with the first recorded observation on October 16th 1985 and the most recent record on January 2nd 2010. The records vary in their degree of accuracy and completeness, therefore only those records with complete geographic and date information were used for the test study, yielding a final dataset of 987 records. In order to establish acceptable value ranges to use in the model for salinity, ocean depth, and temperature, values for these parameters had to be calculated for each lionfish capture point in the USGS database. This was accomplished by using ArcMap to create a spatial join between the lionfish records and each parameters dataset. Once parameter values were obtained for each lionfish record, the mean parameter value was calculated for all records. For the lionfish test case, upper and lower ranges for temperature, salinity, and depth were then set to the mean value  $\pm 2$  standard deviations, sufficient to encompass 95% of the expected value range for lionfish.

A second use case was examined for an introduction of a highly invasive strain of the marine alga *Caulerpa taxifolia*, common in the Mediterranean (Boudouresque et al. 1995; Delgado et al. 1996). This strain of *C. taxifolia* is common in the aquarium industry and is thought more robust than the native strains that are

common to tropical seas worldwide, although this point has been somewhat contested (Phillips and Price 2002). The aquarium strain reproduces asexually from fragmentation and is responsible for large smothering-type outbreaks which have occurred in the western Mediterranean Sea. Small populations of the invasive weed have also been found in California where it has since been extirpated (CISR 2012). As an example of how the Invasionsoft model could be used without examining historical record sequences, a theoretical *C. caulerpa* invasion model was produced originating in the coastal waters of Louisiana, USA. Because historical records were not analyzed for this invasion, the model input relied on parameter ranges based on *C. caulerpa* tolerances from literature.

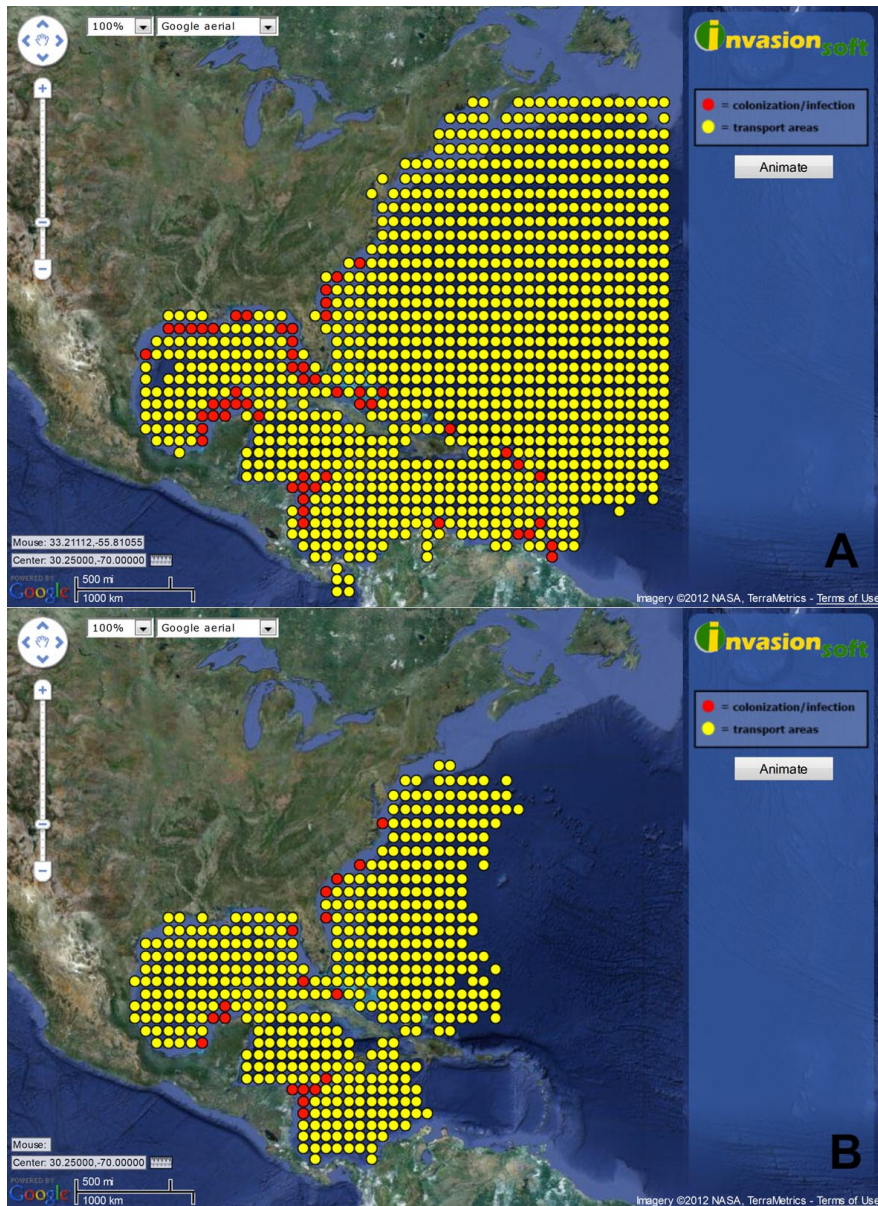
#### 2.4 Processing logic details

The process of creating a model begins with the user inputting all required data fields in the browser form as a series of steps. Instructions are provided in an instruction link at the top of the page (Figure 1) as well as context sensitive help via ‘?’ buttons located next to each input field via popups. Once the user has entered all data these must be validated in order to continue processing, which is accomplished by pressing the ‘validate fields’ button. All validation of input data occurs either in browser-based code (for required field checks) or calls to the SQL database via stored procedures to check for valid parameter data ranges and initial introduction points. Once all fields are validated, including a valid user ID or email address, a popup advises that processing can begin.

Validated input data is sent to a SQL stored procedure and VB.Net server-side code which initiates a set of algorithms performing the processing logic. The entire sequence logic is outlined in the following steps and in Figure 2 in Johnston and Purkis (2011). The algorithms are iterative and repeated for each cycle as defined in the web input.

1. A SQL stored procedure is called from the front-end code and the initial introduction (vector) cell and parameter weight/ranges are sent as parameters (defining the CA rule).

2. A SQL stored procedure cycles through all records in either the default or custom parameter dataset (PD) (the default dataset contains salinity, temperature, depth, and current values and coordinates) and calculates the score for each latitude/longitude in the PD based on the



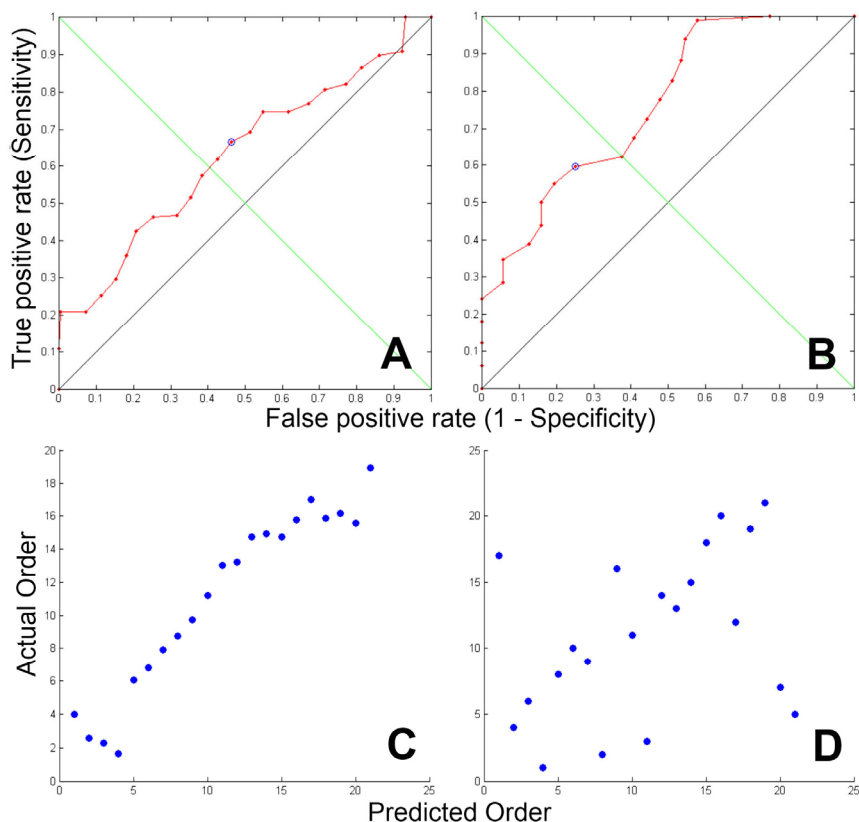
**Figure 2.** Model Simulation Maps. BFM simulation output for a lionfish invasion based on a south Florida introduction and a cycle time of 80 (A). Simulation output for a *Caulerpa taxifolia* invasion based on a Gulf of Mexico introduction with a 60-count cycle time (B).

parameter weights and ranges as defined in the user input. Example: Cell 2 (Figure 3 in Johnston and Purkis (2011); a CA conceptual cell) has a salinity and depth within range, so its score is the sum of the weight of salinity (.02) and depth (.02) for a total of .04. These values are stored in a temporary table (TT1) which will be used later in processing.

3. The initial cell denoted in the user input as the vector cell is marked as colonized (a positive CA cell state) in TT1.

4. The following process then repeats for each colonized cell in TT1 for the number of cycles defined in the user input:

a. If current is defined as a parameter (this is the normal behavior using the default parameter



**Figure 3.** ROC/SRCC Calculations. L-BFM simulation ROC calculation with an AUC value of 0.65 (A). L-BFM simulation ROC calculation with a one-step error factor and AUC value of 0.76 (B). L-BFM SPCC analysis with a correlation value of 0.97 (C). Random trajectory SPCC analysis with a correlation value of 0.37 (D).

set), a normalized velocity factor (NVF) is obtained by dividing 1 by the mean of all current velocity values in the PD. The NVF will be used later in processing when determining the influence of a particular cells' current velocity when combined with the current weight value. For example, a NVF of 1 means that cell has an average ocean current velocity and a NVF of 2.5 means the current is two and a half times as strong as the mean velocity for all parameter cells. Current direction is also obtained for the colonized cell from the PD.

**b.** Records for all 8 cells surrounding the colonized cell (the CA *neighborhood*) are selected and inserted into a temporary table (TT2) for further processing.

**c.** If current is being considered, a weighted current score factor (WCSF) is calculated by multiplying the velocity of the ocean current for

the colonized cell by the NVF. The WCSF is then multiplied by the current weight factor from the user input giving a standardized score for ocean current. If the cell contains a 'multiplier' factor (used to simulate a partial dispersal barrier in areas of narrow geographic spread but high current flow, such as the Florida straits between south Florida and the Bahamas), the standardized score is then multiplied by this score. The score of down-current cell (from the colonized cell) in TT1 is increased by the weighted current score.

**d.** The scores of the *neighborhood* cells (in TT2) are standardized to a value between 0 and 1 by dividing the score of each neighboring cell by the sum of all eight neighboring cell scores. Each cell is then assigned a number range from 0 to 1 based on the calculated number [Example: Cell 1 (Figure 4 in Johnston and Purkis (2011)) was given a value range of 0.0001 to 0.0148 based on

calculations in Table 1 in Johnston and Purkis (2011)]. This calculation ensures that the relative probability of a cell being chosen in the next step (4e) is based on the cumulative score of that cells parameter values.

e. A random number between 0 and 1 is generated (the stochastic variable). The corresponding cell in the CA *neighborhood* whose range contains that number is chosen as the infected cell (a positive CA *cell state*) in TT1. If the mean current velocity for the vector cell is more than 5 times that of the NVF, and current is being evaluated, then the cell is deemed a high dispersal cell. Accordingly, if the down-current cell from the vector was selected as an infected cell, then the neighboring down-current cell in the same direction as the vector current is also marked infected.

f. For the indicated cell(s), if one or more parameters are marked as a required field and the cells parameter values fall within range for the required fields, the cell is marked colonized as indicated with a red dot in the viewer. Infected cells whose required parameter values do not fall within the input range are marked transport cells and indicated with a yellow dot in the output and viewer. If no parameters are marked as required, all infected cells are marked as colonized.

5. Step 4 is repeated for the number of cycles as defined in the user input. Because the cycle is repeated for every infected cell, the trajectory of each cycle (and for each cell) is independent of the previous cycle. The trajectory therefore is entirely controlled by the weight factors and parameter values of the *neighborhood* cells.

6. The output of the model is a list of latitude/longitude data points and the cycle in which they were colonized. A .csv file is compiled by the software and a link emailed to the address as specified in the user input.

7. Using the generated .csv file, the custom data browser is launched which displays the relative location of each cell that has been colonized (Figure 2). This map utilizes a Google maps background to draw the points in client-side javascript code. Cells which are defined as colonized in the output are drawn with a red dot, the rest are drawn with a yellow dot. If no cells are marked as required in the input parameters, then all cells will be drawn with a red dot. The purpose of the browser viewer is to immediately display the result of the model.

## 2.5 Obtaining a Best Fit Model (BFM)

The ideal method when utilizing Invasionsoft is to pattern-match the results of the model output to actual historical records of an invasion to obtain a Best Fit Model (BFM). The input parameter values derived from this BFM can then be applied to a different geographical location to show the theoretical progression of an invasion starting at the selected introduction point. In invasions where no historical pattern is available to pattern-match, a BFM is not produced. A BFM is obtained by following these steps:

### 1 Pattern match to obtain the initial BFM:

The lionfish invasion pattern follows a distinct set of ‘keynote’ events which somewhat define the overall pattern of spread. These keynote events include: 1) The initial quick spread north from point of origin on the Gulf Stream, 2) slow spread gradually south in the Bahamas 3) introduction and spread in the mid-Caribbean to the Yucatan and south, and 4) eventual spread back to the Florida Keys. Through trial and error, or a predetermined test grid of parameter values, simulations should be created with the overall goal of obtaining a general visual pattern match based on a known historical invasion sequence

### 2 Verifying and tuning the BFM:

Once an initial pattern-matched BFM and resulting parameter input values have been identified, a series of simulations should be created using the same input values to further verify/tune the model (for our lionfish test case, we created 20 sample simulations). From these sample simulations, an aggregate Receiver Operating Characteristic (ROC) analysis should be performed and resulting Area Under the Curve (AUC) value calculated to account for false positive/false negative predicted sequences. The ROC analysis should be based on the actual order of progression of the historic invasion into predefined quadrants verses the predicted progression from the model. This process is described in more detail in section 2.6. Additionally, a Spearman’s Rank Correlation Coefficient (SRCC) should be calculated for the simulations to indicate relative fit value for the model versus the actual historic pattern. An AUC



value above 0.6 (where 0.5 is a random classifier) and high positive SRCC value ( $>0.5$ ) indicate a good fit (Rothlisberger and Lodge 2010; Tuite et al. 2011). Input parameter values should be further adjusted to tune the model and obtain the highest possible AUC and SRCC values.

Currently, Invasionsoft allows for multiple simulations (designated in the ‘Simulation Quantity’ input box) to be created using the same input parameters to facilitate the creation and validation of the BFM. Additionally, random trajectory simulations can be created for null model comparison (default parameters only). Future versions will automate the process to obtain the BFM based on the highest AUC and most significant SRCC values given an introduction point and parameter dataset with range values.

### *2.6 Lionfish – Best Fit Model (L-BFM) – model validation*

In order to validate the L-BFM, 20 identical simulations were created using the parameter input values obtained from our L-BFM. In order to evaluate the actual invasion sequence against the USGS records, the colonized area was then divided into a grid consisting of  $5^{\circ} \times 5^{\circ}$  (approximately  $500 \text{ km} \times 500 \text{ km}$ ) cells, based on the large geographic area of the invasion, lack of current support for variable rates of spread, and intent to depict general direction of spread and first occurrence sequences (a smaller grid pattern could be used if examining a smaller scale model). Each cell was assigned a column and row number and the sequence of colonization for the USGS database records was then recorded according to first occurrences in each of the 21 grid cells containing lionfish (3 cells were excluded as extreme outliers). The sequence of invasion was then recorded for each of the 20 sample simulations and the mean sequence value calculated for each step. Finally a ROC analysis was performed, including an AUC calculation. A SRCC value was also calculated comparing the actual invasion sequence to the mean predicted sequence.

### *2.7 Caulerpa taxifolia– model validation*

For the *C. taxifolia* use case example, we chose to demonstrate the use of the software without examining a historical pattern (ignoring the historical invasion sequence from the

Mediterranean) to show its predictive ability given a newly established invader. Because a BFM was not used in the *C. taxifolia* use case using a historical invasion sequence, a probability distribution of spread was produced by creating 20 simulations with the same input parameters. The sequence of spread for each simulation was then recorded using the  $5^{\circ} \times 5^{\circ}$  grid quadrants as defined when producing the lionfish BFM. To analyze the overall pattern of invasion, the quadrants were summed across all simulations and counted for the first 25 invasion steps. The quadrant with the highest count for each step was selected as the overall best fit sequence. Next each individual simulation was compared to the overall best fit sequence and scored based on fit for each step. The simulation with the highest score was then selected as the most representative best fit sequence. Finally, a null simulation model was produced using the same vector location and its sequence analyzed. A SRCC value was produced comparing the test case best fit sequence to the null simulation to evaluate any correlation.

## **Results**

### *3.1 L-BFM – lionfish test case*

For the lionfish test-case, the L-BFM was defined as the model whose output most closely matched the pattern and temporal sequence of captures recorded in the USGS database as well as the model whose validation resulted in an  $\text{AUC} > 0.60$  and  $\text{SRCC} > 0.90$ . In the resulting model, upper and lower ranges for temperature, salinity, and depth were set to the mean value  $\pm 2$  standard deviations, as determined by the statistical analysis of the parameter data and presented in Johnston and Purkis (2011) ( $26.5^{\circ}\text{C}$  for temperature, 36.11 psu for salinity, 35 m for depth). A current weighted value of 0.90 provided the best fit to the invasion pattern as described by the USGS database. Salinity, temperature, and depth were weighted at 0.02 based on the 95th percentile range for each parameter. A cell where salinity, temperature, and depth values were not in range (mean value  $\pm 2$  standard deviations) was weighted at 0.01 (sum of weight factors have no correlation to 1).

### *3.2 L-BFM model validation*

In order to validate the L-BFM, 20 identical simulations were created using the parameter

input values obtained from our L-BFM. The mean order of invasion for each step from these 20 simulations was compared to the actual order of invasion in each of the 21 quadrants by performing a ROC analysis and calculating a SPCC value. From the ROC analysis we obtained an AUC value of 0.65 (Figure 3A), and with a one-step error margin (the step before and after the current step were evaluated for a match), this value increased to 0.76 (Figure 3B). Our SPCC calculation resulted in a value of 0.97 (Figure 3C). We also calculated a SPCC value for a random trajectory from the same vector location and obtained a value of 0.37 (Figure 3D). An example simulation delivered by our L-BFM with a vector location of 25.5° latitude, -79.5° longitude, and with a cycle count of 80 predicts establishment along the east coast as far north as Cape Hatteras, North Carolina, Bermuda, throughout the Bahamas, the entire Caribbean as far south as northern South America, and the majority of the Gulf of Mexico (Figure 2A). From the results of our ROC and SPCC tests and overall pattern match, we conclude our model results to be significantly better than can be explained by random chance.

### 3.3 Lionfish test case predictions

Using the parameter values obtained from the L-BFM, sample simulations were created using two different origination points to demonstrate how the invasion would have progressed if lionfish had instead been introduced to Gulf of Mexico waters near Galveston, Texas, USA and waters off the north coast of Colombia (Figure 6D and 6F in Johnston and Purkis (2011)). Most simulations produced a similar end result after 70-90 iterative cycles with lionfish established throughout the Caribbean, limited in north latitudes by temperature and confined to depths of < 200 m. Most simulations, regardless of the point of initial release, also predict establishment throughout the Gulf of Mexico and are therefore in agreement with predictions made by Schofield (2010).

### 3.4 *Caulerpa taxifolia* use case example

For the *C. taxifolia* use case example, a model was produced without examining the historical configuration of invasion to show how Invasionsoft can produce a non-pattern matched model. Since the strain of *Caulerpa* being examined relies on fragmentation as a

reproduction method, instead of releasing free gametes into the water column, less emphasis (and thus a lower weight factor) was placed on current in the model. However, current still has an influence in transport of fragments so likewise the weight factor was set at a value of 0.50. The temperature range used was 10°C to 28.5°C (the maximum value in the default parameter set) based on approximate lethal temperature extremes in their introduced range (Komatsu et al. 1997). As temperature is one of the only limiting factors of this invader, the weight factor used in the model was set relatively high at 0.50. The depth range used in the model was a minimum of 2 m and a maximum of 30 m, based on the observation that new colonies typically occur in waters from 2–20 m and large colonies regularly occur to 30 m. (Meinesz et al. 1993; Boudouresque et al. 1995). Depth is perceived to be somewhat of a limiting factor as photosynthesis ceases to occur beyond the euphotic zone. As a result, the weight factor used for depth was also relatively high at .50. *Caulerpa taxifolia* is a marine alga and thus the salinity range used was 32–37 psu. This encompasses almost the entire range of default parameter values for salinity present in the study area (31.060–37.203 psu). *Caulerpa taxifolia* has been shown to be highly tolerant of low nutrient levels as well as adverse lighting conditions (Delgado et al. 1996; Komatsu et al. 1997). It is likely that salinity is not a limiting factor in the theorized introduced range; therefore, the weight factor used for salinity was low at 0.02. A cell where salinity, temperature, and depth values were not in range was weighted at 0.01.

The model colonization sequence produced by Invasionsoft indicates that an uncontrolled *C. taxifolia* invasion with a vector location of 29.5° latitude and -88.5° longitude (off the coast of New Orleans, Louisiana) would likely occur towards the east and south first, by fragments transported on the strong loop current, followed by a gradual spread west (Figure 2B). After 60 cycles, the simulation produced by Invasionsoft indicates spread into the entire Gulf of Mexico, around the tip of Florida and north to the coast of New York where overwintering temperatures would cause die-off. Many Caribbean nations were within the colonized area (as far east as the Dominican Republic) as well as the Yucatan Peninsula and the northwestern South American coast.

## Discussion

### 4.1 Lionfish test case

In our lionfish test-case scenario, neither depth nor temperature was shown as highly influential (low weight values of 0.02) in the spread of lionfish larvae into new areas within the L-BFM produced by the Invasionsoft software. We do acknowledge that temperature and depth are likely correlated; however any correlation is not considered to be detrimental to the study case as the influence of these parameters on the L-BFM is minor. Additionally, over the last 25 years, awareness of the lionfish invasion has increased dramatically and likely has resulted in increased sampling intensity in the affected areas, possibility introducing spatial bias to the USGS records invasion sequence upon which the lionfish test-case L-BFM is based. Other test cases where sampling was more tightly regulated likely would not have this spatial bias.

The resulting L-BFM output map shows lionfish colonization in the most of the Caribbean Sea, Gulf of Mexico, and the eastern shore of the United States north to Cape Hatteras, North Carolina where winter bottom temperatures are likely to cause die-off. This northern most limit corresponds to major benthic structure and habitat changes and was also found to be the likely northernmost limit for lionfish by Kimball et al. (2004). Lionfish are known to utilize a large variety of habitats including mangroves, hard bottom, artificial structures, deep reefs, and estuaries, all of which are encompassed in predicted the colonization area (Whitfield et al. 2002; Barbour et al. 2010). Additionally, although the algorithm currently does not contain a time component, each cycle represents approximately 2.5 months spread in the lionfish test case (208 months/80 cycles). On a large-scale, our results indicate current is likely a major determining factor in the distribution of lionfish larvae, however settling and survivorship of larvae is limited by temperature and depth requirements in the study region.

Of the 1,147 records examined in the test case, only 1 was indicated from the Gulf of Mexico – north of Tampa Bay in October of 2009. According to the L-BFM produced by Invasionsoft, up-stream populations which are established throughout the Caribbean will likely seed the Gulf-region based on ocean current patterns and the high influence currents have on the distribution of this species. Temperature,

salinity, and depths along coastal regions in the Gulf fall within lionfish tolerances as defined by the test case study and supported by literature (Kimball et al. 2004; Schofield 2010). At the time of writing, lionfish have been reported from the Florida panhandle, Alabama, Louisiana, and Texas offshore waters indicating the invasion has progressed to these areas, as predicted by the Invasionsoft model (Schofield 2010; USGS NAS database 2011).

### 4.2 *Caulerpa taxifolia* example use case

The model colonization sequence produced by Invasionsoft indicates a large area of colonization throughout the Gulf of Mexico, east coast of the United States north to New York, and many Caribbean nations. Of particular concern would be shallow shelf areas, such as those off the west coast of Florida and northeast coast of Florida as *C. taxifolia* has high affinity for shallow depths (< 50 m) (Meinesz et al. 1993). The results of the SRCC analysis between the best fit sequence and a random sequence resulted in a value of 0.85. This positive correlation was expected due to the sequential radial spread of the model (cells closer to the vector having a much higher chance of initial infect than further cells), lower current weight value used in the *C. taxifolia* case, and lack of historical invasion pattern match to refine the model outputs. We acknowledge that models produced without a historical matching component are likely less precise than those with, however these models can still provide useful insight to expected spread patterns.

### 4.3 Future directions and research

The Invasionsoft web portal is accessible to anyone with internet access making it widely available even in developing countries at no cost. Additionally, an instruction page, context sensitive help, and an intuitive interface negate the need for formal training. The Invasionsoft software would be a useful tool for resource managers to help predict patterns of colonization should a current invader (such as the lionfish) or a newly introduced invasive become established in a new ecosystem. For our lionfish test case, an example could be the Pacific coast of Panama (by passage through the Panama Canal) or the Mediterranean Sea. Early use would indicate areas where detection and prevention efforts should be focused. For our lionfish test case, the

western Gulf of Mexico is one such area as predicted by the software and supported by Schofield (2010). Additionally, the web-based portal enables users to access the software via any modern web-browser without a software download.

Additional features to future versions of Invasionsoft include further refinements to the prediction algorithms which would allow for time-sensitive cycles to gain a more accurate determination of the timeline of an invasion. Other enhancements will include enhanced capability to adjust for unusual physical conditions in custom datasets, as encountered and adjusted for during the lionfish test case and default dataset in the Florida straits (narrow geographic spread, fast current). Additionally, the ability to specify cells as dispersal barriers will be an important enhancement facilitating use in environments where these are more prevalent, as opposed to the marine environment where dispersal barriers are not as common.

The algorithm developed from our lionfish and *C. taxifolia* test cases has shown that a simple CA spatial model is capable of emulating the complex spread of a marine invasive species. A user-friendly web-based portal named Invasionsoft has been developed to allow to scientific community to tap into this resource (<http://www.invasionsoft.com>). We encourage its use and welcome suggestions for future enhancements to better hone its predictive capabilities.

## Acknowledgements

We would like to thank those who have provided their feedback during the initial testing of the Invasionsoft web-portal. We also thank the USGS for providing us the lionfish capture data on which the initial study and resulting software was based. We are also grateful the three anonymous reviewers who provided great insight into the refinement of this paper and our model. This is NCRI contribution #143.

## References

- Amante C, Eakins BW (2009) ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC 24: 19 pp, March 2009
- Barbour AB, Montgomery ML, Adamson AA, Díaz-Ferguson E, Silliman BR (2010) Mangrove use by the invasive lionfish *Pterois volitans*. *Marine Ecology Progress Series* 401: 291-294, <http://dx.doi.org/10.3354/meps08373>
- Bonjean F, Lagerloef GSE (2002) Diagnostic model and analysis of the surface currents in the tropical Pacific Ocean. *Journal of Physical Oceanography* 32(10): 2938-2954, [http://dx.doi.org/10.1175/1520-0485\(2002\)032<2938:DMAAOT>2.0.CO;2](http://dx.doi.org/10.1175/1520-0485(2002)032<2938:DMAAOT>2.0.CO;2)
- Bolliger J, Sprott JC, Mladenoff DJ (2003) Self-organization and complexity in historical landscape patterns. *Oikos* 100: 541-553, <http://dx.doi.org/10.1034/j.1600-0706.2003.12109.x>
- Boudouresque CF, Meinesz A, Ribera MA, Ballesteros E (1995) Spread of the green alga *Caulerpa taxifolia* (Caulerpales, Chlorophyta) in the Mediterranean: Possible consequences of a major ecological event. *Scientia Marina* 59 (Dec) Suppl. 1: 21-29
- Boyer TP, Antonov JI, Garcia HE, Johnson DR, Locarnini RA, Mishonov AV, Pitcher MT, Baranova OK, Smolyar IV (2006) World Ocean Database 2005. Levitus S (ed) NOAA Atlas NESDIS 60, U.S. Government Printing Office, Washington, D.C., USA, 190 pp, DVDs [http://www.nodc.noaa.gov/OC5/WOA05/pr\\_woa05.html](http://www.nodc.noaa.gov/OC5/WOA05/pr_woa05.html) (Accessed October, 2010)
- Center for Invasive Species Research (CISR) (2012) Killer Alga, *Caulerpa taxifolia* website. [http://cizr.ucr.edu/caulerpa\\_taxifolia.html](http://cizr.ucr.edu/caulerpa_taxifolia.html) (Accessed February, 2012)
- Cole V, Albrecht J (1999) Modeling the Spread of Invasive Species; Parameter Estimation Using Cellular Automata in GIS. Proceedings of the Second International Workshop on Dynamic and Multi-Dimensional GIS. Beijing, China, October 4-6, 1999, ISPRS, Vol. 32, part 4W12, pp 141-48
- Dittel AI, Epifanio CE (2009) Invasion biology of the Chinese mitten crab *Eriocheir sinensis*: A brief review. *Journal of Experimental Marine Biology and Ecology* 374: 79-92, <http://dx.doi.org/10.1016/j.jembe.2009.04.012>
- Delgado O, Rodriguez-Prieto C, Gacia E, Ballesteros E (1996) Lack of severe nutrient limitation in *Caulerpa taxifolia* (Vahl) C Agardh, an introduced seaweed spreading over the oligotrophic northwestern Mediterranean. *Botanica Marina* 39(1): 61-67, <http://dx.doi.org/10.1515/botm.1996.39.1-6.61>
- Engler R, Guisan A (2009) MigClim: Predicting plant distribution and dispersal in a changing climate. *Diversity and Distributions* 15: 590-601, <http://dx.doi.org/10.1111/j.1472-4642.2009.00566.x>
- ESRI (2011) ArcGIS Desktop: Release 10. Environmental Systems Research Institute, Redlands, CA, USA
- Fishelson L (1997) Experiments and observations on food consumption, growth and starvation in *Dendrochirus brachypterus* and *Pterois volitans* (Pteroinae, Scorpaenidae). *Environmental Biology of Fishes* 50 (4): 391-403, <http://dx.doi.org/10.1023/A:1007331304122>
- Franklin J (2010) Moving beyond static species distribution models in support of conservation biogeography. *Diversity and Distributions* 16: 321-330, <http://dx.doi.org/10.1111/j.1472-4642.2010.00641.x>
- Freshwater DW, Hines A, Parham S, Wilbur A, Sabaoun M, Woodhead J, Akins L, Purdy B, Whitfield PE, Paris CB (2009) Mitochondrial control region sequence analyses indicate dispersal from the US East Coast as the source of the invasive Indo-Pacific lionfish *Pterois volitans* in the Bahamas. *Marine Biology* 156: 1213-1221, <http://dx.doi.org/10.1007/s00227-009-1163-8>
- Gallien L, Münkemüller T, Albert CH, Boulangat I, Thuiller W (2010) Predicting potential distributions of invasive species: where to go from here? *Diversity and Distributions* 16: 331-342, <http://dx.doi.org/10.1111/j.1472-4642.2010.00652.x>
- Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8: 993-1009, <http://dx.doi.org/10.1111/j.1461-0248.2005.00792.x>
- Hare JA, Whitfield PE (2003) An integrated assessment of the introduction of lionfish (*Pterois volitans/miles* complex) to the western Atlantic Ocean. NOAA Technical Memorandum NOS NCCOS 2, 21 pp
- Jacobs MJ, MacIsaac HJ (2009) Modelling spread of the invasive macrophyte *Cabomba caroliniana*. *Freshwater Biology* 54: 296-305, <http://dx.doi.org/10.1111/j.1365-2427.2008.02108.x>

- Johnston MW, Purkis SJ (2011) Spatial analysis of the invasion of lionfish in the western Atlantic and Caribbean. *Marine Pollution Bulletin* 62(6): 1218-1226, <http://dx.doi.org/10.1016/j.marpolbul.2011.03.028>
- Kearney M, Phillips BL, Tracy CR, Christian KA, Betts G, Porter WP (2008) Modelling species distributions without using species distributions: the cane toad in Australia under current and future climates. *Ecography* 31: 423-434, <http://dx.doi.org/10.1111/j.0906-7590.2008.05457.x>
- Kimball ME, Miller JM, Whitfield PE, Hare JA (2004) Thermal tolerance and potential distribution of invasive lionfish (*Pterois volitans/miles* complex) on the east coast of the United States. *Marine Ecology Progress Series* 283: 269-278, <http://dx.doi.org/10.3354/meps283269>
- Komatsu T, Meinesz A, Buckles D (1997) Temperature and light responses of alga *Caulerpa taxifolia* introduced into the Mediterranean Sea. *Marine Ecology Progress Series* 146(1-3): 145-153, <http://dx.doi.org/10.3354/meps146145>
- Meinesz A, de Vaugelas J, Hesse B, Mari X (1993) Spread of the introduced tropical green alga *Caulerpa taxifolia* in northern Mediterranean waters. *Journal of Applied Phycology* 5: 141-147, <http://dx.doi.org/10.1007/BF00004009>
- Morris JA Jr, Akins JL (2009) Feeding ecology of the invasive lionfish in the Bahamian archipelago. *Environmental Biology of Fishes* 86: 389-398, <http://dx.doi.org/10.1007/s10641-009-9538-8>
- Muñoz MES, Giovanni R, Siqueira MF, Sutton T, Brewer P, Pereira RS, Canhos DAL, Canhos VP (2009) "openModeller: a generic approach to species' potential distribution modelling". *GeoInformatica* 15: 111-135, <http://dx.doi.org/10.1007/s10707-009-0090-7>
- O'Neill CR Jr (1997) Economic Impact of Zebra Mussels – Results of the 1995 National Zebra Mussel Information Clearinghouse Study. *Great Lakes Research Review* 3: 35-42
- Ostendorf B, Hilbert DW, Hopkins MS (2001) The effect of climate change on tropical rainforest vegetation patterns. *Ecological Modelling* 145: 211-224, [http://dx.doi.org/10.1016/S0304-3800\(01\)00392-1](http://dx.doi.org/10.1016/S0304-3800(01)00392-1)
- Parks SC, Garifullin M, Dronzek R (2005) Argus invasive species spread model constructed using agent-based modeling approach and cellular automata. In: Kuhl ME, Steiger NM, Armstrong FB, Joines JA (eds), Proceedings of the 2005 Winter Simulation Conference, 4-7 December 2005. IEEE, ACM Press, New York, USA, pp 1038-1042
- Phillips JA, Price IR (2002) How different is Mediterranean *Caulerpa taxifolia* (Caulerpales: Chlorophyta) to other populations of the species? *Marine Ecology Progress Series* 238: 61-71, <http://dx.doi.org/10.3354/meps238061>
- Phillips SJ, Dudík M, Schapire RE (2004) A maximum entropy approach to species distribution modeling. Proceedings of the Twenty-First International Conference on Machine Learning, 2004, Banff, Alberta, Canada, pp 655-662
- Rothlisberger JD, Lodge DM (2011) Limitations of gravity models in predicting the spread of Eurasian watermilfoil. *Conservation Biology* 25(1): 64-72, <http://dx.doi.org/10.1111/j.1523-1739.2010.01589.x>
- Schofield PJ (2010) Update on geographic spread of invasive lionfishes (*Pterois volitans* [Linnaeus 1758] and *P. miles* [Bennett 1828]) in the Western North Atlantic Ocean, Caribbean Sea and Gulf of Mexico. *Aquatic Invasions* 5 (Supplement 1): S117-S122, <http://dx.doi.org/10.3391/ai.2010.5.S1.024>
- Tuite AR, Tien J, Eisenberg M, Earn DJ, Ma J, Fisman DN (2011) Cholera epidemic in Haiti, 2010: using a transmission model to explain spatial spread of disease and identify optimal control interventions. *Annals of Internal Medicine* 154: 593-601
- USGS-NAS (2011) United States Geological Survey - Nonindigenous Aquatic Species database (USGS-NAS). <http://nas.er.usgs.gov> (Accessed 4 August 2011)
- Whitfield PE, Gardner T, Vives SP, Gilligan MR, Coutenay Jr WR, Ray GC, Hare JA (2002) Biological invasion of the Indo-Pacific lionfish *Pterois volitans* along the Atlantic coast of North America. *Marine Ecology Progress Series* 235: 289-297, <http://dx.doi.org/10.3354/meps235289>
- Wilson RJ, Davies ZG, Thomas CD (2009) Modelling the effect of habitat fragmentation on range expansion in a butterfly. *Proceedings of the Royal Society B-Biological Sciences* 276: 1421-1427, <http://dx.doi.org/10.1098/rspb.2008.0724>