RESEARCH ARTICLE

Location of sources in reaction-diffusion equations using support vector machines

Venecia Chávez-Medina[✉], José A. González^{✉*}, Francisco S. Guzmán[✉]

Laboratorio de Inteligencia Artificial y Supercómputo, Instituto de Física y Matemáticas, Universidad Michoacana de San Nicolás de Hidalgo. Edificio C-3, Cd. Universitaria, 58040 Morelia, Michoacán, México

[✉] These authors contributed equally to this work.

* gonzalez@ifm.umich.mx

OPEN ACCESS

Citation: Chávez-Medina V, González JA, Guzmán FS (2019) Location of sources in reaction-diffusion equations using support vector machines. PLoS ONE 14(12): e0225593. <https://doi.org/10.1371/journal.pone.0225593>

Editor: Enrique Hernandez-Lemus, Instituto Nacional de Medicina Genómica, MEXICO

Received: May 28, 2019

Accepted: November 7, 2019

Published: December 5, 2019

Copyright: © 2019 Chávez-Medina et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The numerical code use to solve the RD system is available at https://figshare.com/articles/Code_that_solves_the_RD_equation/10347836 (<https://doi.org/10.6084/m9.figshare.10347836.v1>).

Funding: This work was supported by: CIC-UMSNH-4.9, JAG: CIC-UMSNH-4.23, <https://www.cic.umich.mx> (FSG); and CONACyT Grant No.258726, <https://www.conacyt.gob.mx> (FSG). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

The reaction-diffusion equation serves to model systems in the diffusion regime with sources. Specific applications include diffusion processes in chemical reactions, as well as the propagation of species, diseases, and populations in general. In some of these applications the location of an outbreak, for instance, the source point of a disease or the nest of a vector spreading a virus is important. Also important are the environmental parameters of the domain where the process diffuses, namely the space-dependent diffusion coefficient and the proliferation parameter of the process. Determining both, the location of a source and the environmental parameters, define an inverse problem that in turn, involves a partial differential equation. In this paper we classify the values of these parameters using Support Vector Machines (SVM) trained with numerical solutions of the reaction-diffusion problem. Our set up has accuracy of classifying the outbreak location above 90% and 77% of classifying both, the location and the environmental parameters. The approach presented in our analysis can be directly implemented by measuring the population under study at specific locations in the spatial domain as function of time.

1 Introduction

The reaction-diffusion (RD) equation is applied to model various dynamical systems involving populations. For instance in mathematical biology, the RD equation is applied, in the frame of biological invasion, used to model the propagation of dengue via the *Aedes aegypti* vector at urban spatial scale [1]. More recently, a free boundary approach is used to model the behavior *Aedes aegypti* mosquito populations, based on reaction diffusion advection processes to determine the population spreading and vanishing regimes [2]. Also the propagation and proliferation of other types of viruses like the West Nile virus [3, 4] or the feline immunodeficiency virus [5], is studied using RD models. At metropolitan scale, the RD system is used to model the population dynamics of the tiger mosquito and has potential at the scale of a whole country [6]. Generalization to more than one population is constructed by coupling two or more RD equations, and is useful to explain the spread of seeds and animals. For instance, RD allows the description of cross-diffusion in multiple species systems with nonlocal interaction [7], and

Competing interests: The authors have declared that no competing interests exist.

modeling biological invasion [8, 9]. In fact, the particular case of cross-diffusion has a wide variety of applications in mathematical biology [10]. The RD is also being used to model the response of ecosystems to climate change, for instance through the rainfall decline [11].

With different degrees of complexity and additional conditions, the temporal and spatial evolution of a population u is assumed to be described by the RD equation

$$\frac{\partial u}{\partial t} = \nabla \cdot (D(\mathbf{x}) \nabla u) + s(u). \quad (1)$$

The evolution of the population u is described by the solution of an initial value problem for given initial conditions $u(\mathbf{x}, 0) = u_0(\mathbf{x})$ evolving in a domain $A \in \mathbb{R}^2$. The first term of the right hand side is the diffusion operator with a space-dependent diffusion coefficient $D(\mathbf{x})$, and $s(u)$ is a source term usually named proliferation term when u is interpreted as the density of a population. The space dependency of the diffusion D and s is useful in various scenarios. For example, the spread of a species of seeds or animals, depends on environmental variables that influence the diffusion and proliferation effects (see e.g. [12, 13]).

Of course, different initial conditions, as well as different diffusion and proliferation functions, give rise to different solutions $u(\mathbf{x}, t)$. The solution of the *direct problem* consists in the evolution of an initial profile $u_0(\mathbf{x})$, for instance a fluctuation representing the outbreak of a disease or a vector, with a specific diffusion coefficient function $D(\mathbf{x})$ representing the spatial conditions for the propagation of the population, and the proliferation function $s(u)$ associated to a population growth model. Solving the problem consists in calculating the value of u at a given set of particular points in the domain called detectors, and predict the population u in a spatial neighborhood of such detectors as function of time t . For the particular problem of the RD equation, this can be achieved by numerically solving the initial value problem (IVP) in (1).

The direct problem is of interest in itself when D and s are known for specific initial data. It happens though that locating the position of an outbreak's origin, the starting point of a plague, of an anomaly in seed spread based on measurements, are also important situations, with the additional difficulty of not knowing the initial conditions that provoked the observations and possibly including ignorance on the parameters of D and s of a specific disease and population. These scenarios correspond to *inverse problems* focused on the reconstruction of the initial conditions and environmental variables. Inverse problems related to determine parameters and initial conditions associated to partial differential equations using Artificial Neural Networks happen in different areas, for instance in fluid dynamics [14, 15]. The aim of this paper is to present a strategy to solve this inverse problem by combining the numerical solution of the IVP (1) and the use of SVMs to classify and bound the initial conditions and environmental variables.

The paper is organized as follows. In Section 2 we present the details for the solution of the initial value problem of the RD equation and the machine learning methods. In Section 3 we present the results for the location of an outbreak origin and the environmental parameters. Finally in Section 4 we draw some conclusions.

2 Direct and inverse problems

2.1 General description of the inverse and direct problems

We will study the solution of two inverse problems. The first one consists in the location of a fluctuation of the population u , that could represent the origin of an outbreak. One counts with measurements in time of u at various points of the domain $\mathbf{x}_{det} \in A$ called detectors and the knowledge that the process is governed by Eq (1) for given $D(\mathbf{x})$ and $s(u(\mathbf{x}))$. Should the

problem be linear, the location of the outbreak could make use of a triangulation, nevertheless the equation and the possible space dependency of the diffusion, complicate the implementation of a straightforward triangulation.

The second inverse problem we analyze consists in the use of the time series measured by the detectors, and the objective is to determine not only the location of the source, but also the parameters of $D(\mathbf{x})$ and $s(u(\mathbf{x}))$.

In this paper we use the SVM to solve these two problems using the same strategy. The approach to their solution consists in a classification scheme. This means that a training set is constructed with location and environmental parameters within specific ranges known as classes. The method used to solve these inverse problems is based on SVM method that requires training. Training is based on the solution of the direct problem in the following systematic way.

Training the SVM for the location of the outbreak. We solve the direct problem for a set of initial conditions with a fluctuation of the population u centered at known positions $\mathbf{x}_{src}^{ij}$ where i, j are tags for the location of the fluctuation on a two dimensional domain A . Then measure the value of the solution u at detector locations $\mathbf{x}_{det}^{k,l}$ as a function of time $u(\mathbf{x}_{det}^{k,l}, t)$, where k, l are now tags for the position of the detectors on A . The set of time-series measured by detectors $u(\mathbf{x}_{det}^{k,l}, t)$ corresponding to the different source locations $\mathbf{x}_{src}^{ij}$ are used to train the SVM and fix its internal parameters.

Training for the second inverse problem. For the additional determination of D and s we construct a training set of runs for different values of D and s that produce time series again measured at detector locations $\mathbf{x}_{det}^{k,l}$. These time-series are used to train the SVM as before. The difference is that in this case the locations of initial fluctuations $\mathbf{x}_{src}^{ij}$ are distributed in classes, and the values of D and s are distributed in classes as well.

After the training process, a set of measurements known as prediction set, constructed with location and environmental parameters different from those used to construct the training set, are presented to the SVM which will determine the class they should belong to, that is, a class associated with the position of the source, diffusion coefficient, and proliferation parameter. The classification can be correct or incorrect, and the amount of correct classifications determines the accuracy of the method for a prediction set.

2.2 The direct problem

The direct problem we solve is an IVP involving (1) on the spatial domain $B \subseteq \mathbb{R}^2$, for $t \in [0, t_{max}]$, provided initial conditions $u_0(\mathbf{x})$ and boundary conditions at ∂B . For our analysis we set the spatial domain to $B = \{(x, y) \in [-2, 2] \times [-2, 2]\}$ and impose the boundary conditions $\hat{n} \cdot \nabla u = 0$ at the boundary ∂B , that in theory allow the numerical boundaries to be transparent. Even if these boundary conditions are imposed to guarantee the uniqueness of the solution, we improve the transparency of the boundaries for our problem, by defining the subset $A = \{(x, y) \in [-1, 1] \times [-1, 1]\} \subset B$ as the domain of analysis. This means that we will use initial data consisting of localized fluctuations of $u_0(\mathbf{x})$ centered at positions $\mathbf{x}_{src}^{ij}$ and detectors $\mathbf{x}_{det}^{k,l}$ located also only within A .

The value of t_{max} is such that the diffusion of the initial fluctuation in the initial conditions $u_0(\mathbf{x})$, reaches the whole domain A , and at the same time the possible spurious noise reflected from ∂B does not reach the boundary ∂A . In this manner we prevent the values of u measured at the detectors from being influenced by boundary effects because ∂B is causally disconnected from the domain of interest A . For the estimates of time travel of the signals across the domain, we consider the velocity of the signal to be the Fischer-Kolmogorov velocity, suitable for the reaction diffusion equation [16, 17].

We then solve this IVP numerically using a finite differences approximation of (1), with second order accurate stencils, defined on a discrete domain version of B , with constant and uniform resolutions $\Delta x = \Delta y = 0.01$. The evolution in time is solved out using the method of lines with a third order accurate TVD Runge-Kutta integrator, and time resolution Δt [18]. In order to keep the stability of the evolution method we choose the time step such that $\Delta t \leq \min \left(\frac{\Delta x^2}{2D} \right)$.

Diffusion. In order to construct a more realistic scenario, with a region with a larger diffusion coefficient than in the rest of the domain, that could for instance represent a zone with favorable conditions for reproduction or diffusion, like humidity of a river that favors the diffusion of some bacteria or the proliferation of mosquitoes [19], also thought of as corridors [20], we define a model for the diffusion function to be

$$D(\mathbf{x}) = D_1 + D_2 e^{-\left(\frac{x^2 + y^2 - 2xy}{2\sigma^2} \right)}. \quad (2)$$

where $\mathbf{x} = (x, y)$. We set the width of the Gaussian corridor to $\sigma = 0.08$ and the reason for this value is that we consider the accuracy in the location of an outbreak is interesting within a scale size of the corridor. Now, since we will launch a set of simulations with initial fluctuations centered at positions $\mathbf{x}_{src}^{ij}$ separated by a distance of order ~ 0.2 , our value of σ is appropriate. In our analysis the value of σ is kept fixed in all of our runs, whereas there are two free parameters D_1 and D_2 that could depend on the environmental conditions or characteristics of the population.

Proliferation. For the proliferation we use the logistic population model $s(u) = \rho u(1 - u)$, where ρ is the normalized birth-death coefficient. We also want this coefficient to depend on the environmental conditions and use a function that enhances proliferation in the region with more diffusion, for which we use a similar Gaussian strip as before, across the numerical domain given by

$$\rho(\mathbf{x}) = 0.5 + \rho_0 e^{-\left(\frac{x^2 + y^2 - 2xy}{2\sigma^2} \right)}. \quad (3)$$

with the same value of σ as for the diffusion. In our analysis we assume σ is fixed, whereas ρ_0 is a parameter we vary in our analysis, that indicates how much the population u proliferates.

2.3 The training set

The training set is prepared with the solution of the direct problem for an educated set of values of the parameters D_1 or D_2 in the diffusion model (2), ρ_0 for the proliferation model (3) and the two coordinates of the center of the outbreak location $\mathbf{x}_{src}^{ij}$.

Initial conditions. Since we want to simulate an outbreak type of process we define initial conditions with a prominence at a given location $\mathbf{x}_{src}^{ij}$. For this we decided to use a Gaussian pulse

$$u_0(\mathbf{x}) = u_0(\mathbf{x}_{src}^{ij}) = \exp(-(\mathbf{x} - \mathbf{x}_{src}^{ij})^2 / \sigma_{src}^2) \quad (4)$$

where $\mathbf{x}_{src}^{ij}$ is the center of the outbreak location. Specifically, the initial conditions are characterized only by the location of the outbreak parametrized by the two coordinates of $\mathbf{x}_{src}^{ij}$. We define 9 values for each of the two coordinates, so that the center of the Gaussian is located at different points uniformly distributed across the domain A , so that for $i, j = 1, \dots, 9$ with

$$\mathbf{x}_{src}^{ij} = \left(-1 + \left(i + \frac{1}{2} \right), -1 + \left(j + \frac{1}{2} \right) \right) \Delta s \quad (5)$$

Fig 1. Numerical domain $A = [-1, 1] \times [-1, 1]$ and the position of the outbreak sources with coordinates $\mathbf{x}_{src}^{i,j}$ indicated with squares. For the sake of illustration we only label a few of them in the bottom-left corner. The shaded circle indicates the location of the outbreak location for the fluctuation case centered at $\mathbf{x}_{src}^{3,8}$.

<https://doi.org/10.1371/journal.pone.0225593.g001>

where $\Delta s = 2/10$ is the length of the domain along the x or y directions divided by 10. In order to avoid a considerable overlap in the initial fluctuations launched from neighboring outbreak center positions $\mathbf{x}_{src}^{i,j}$ we use $\sigma_{src} = 0.04$. In Fig 1 we present the domain A and the location of the 81 positions $\mathbf{x}_{src}^{i,j}$ where the initial fluctuation is centered. The gray circle illustrates the position of the initial Gaussian population profile for the particular case $\mathbf{x}_{src}^{3,8}$.

Classes. Nine squared equal size classes are defined for the location of the sources, that correspond to different non overlapping parcels of the domain that cover the whole area of A . These classes are delimited by the dashed lines in Fig 1. For example, sources at $\mathbf{x}_{src}^{1,3}$ and $\mathbf{x}_{src}^{3,1}$ belong to the bottom left class, whereas $\mathbf{x}_{src}^{4,4}$ belongs to the class in the center of the domain.

Environmental and population parameters. For the diffusion parameters, notice that in (2) D_1 is a background value whereas D_2 is the amplitude of the Gaussian prominence along a diagonal of the domain. We illustrate the distribution of the diffusion function in Fig 2.

Fig 2. Numerical domain $A = [-1, 1] \times [-1, 1]$ with a diagonal strip where the diffusion coefficient is bigger according to Eq (2).

<https://doi.org/10.1371/journal.pone.0225593.g002>

We fix the value of D_1 and set the value of D_2 free. For this we define the fixed value $D_1 = 0.1$ and let $D_2 \in [0, 1.9]$ so that the maximum ratio in different regions of the domain is 20, which is a wide range. Specifically we use the values $D_2^k = k\Delta D$ with $k = 0, \dots, 8$ and $\Delta D = 1.9/8$. The classes for the value of D_2 are defined as three equal sized segments of the domain $D_2 \in [0, 1.9]$.

We also define nine values for the proliferation parameter ρ_0 following the expression

$$\begin{aligned} \rho_0^l &= \rho_{min} + l\Delta\rho \quad l = 0, 1, \dots, 8 \\ \Delta\rho &= \frac{\rho_{max} - \rho_{min}}{8} \end{aligned} \quad (6)$$

where $\rho_{min} = -0.5$ and $\rho_{max} = 0.5$, which includes positive and negative population grow rate. Like in the previous cases we define three classes for the value of equal size segments of the domain $\rho_0 \in [-0.5, 0.5]$.

In total we have defined 81 classes for the location, diffusion and proliferation parameter space.

Example of a solution to the direct problem. In Fig 3 we show the iso-contours for the solution of the direct problem for the outbreak centered at position $\mathbf{x}_{src}^{1,5}$, with environmental

Fig 3. Snapshot for the isocontours of u at different times, for the case $D_2 = 0.95$, $\rho_0 = -0.25$. In this example the outbreak is centered at $\mathbf{x}_{src}^{0,1}$. The domain of the plots is A .

<https://doi.org/10.1371/journal.pone.0225593.g003>

parameters $D_2 = 0.95$ and $\rho_0 = -0.25$. At initial time u shows the profile in (4), which is illustrated with the first snapshot. Later on u propagates across the domain. The space-dependent diffusion, higher along the diagonal according to Eq (2) and Fig 2, allows a different velocity of propagation. Notice that the distortion of the lines is pronounced along the diagonal where the diffusion is higher from $t = 0.5$ and on. Notice that by the time of the last snapshot, the signals have reached all the detectors. This is the type of problem we solve for all the combinations of source locations, values of D_2 and ρ_0 described previously.

However the solution itself is not the information collected in measurements of a real case scenario. The information will consist in the value of u at certain given points of the domain as a function of time measured by detectors.

Detectors. We locate detectors that record the value of u at fixed specific locations. In an applied example, these could be the location of an infectious vector counter for instance. In our numerical example, we locate 16 of these detectors in the following positions $\mathbf{x}_{det}^{k,l} = (x_k, y_l)$ where $x_k, y_l = -0.6, -0.2, 0.2, 0.6$. In Fig 4 we show the location of the outbreak points and the location of the detectors $\mathbf{x}_{det}^{k,l}$ with filled circles. These detectors measure the function $u(\mathbf{x}_{det}^{k,l}, t)$ in the time domain. The specific data collected by these detectors is given within the time domain $t \in [0, 2.5]$ and registered every 0.01 units of time. In this way, each detector collects 250 values of u .

An example of the time series measured by the detectors is shown in Fig 5. The initial and environmental conditions are the same used to construct Fig 3. We indicate labels for the measurements corresponding to the detectors that are close to the outbreak location. It is helpful to compare the source center at $\mathbf{x}_{src}^{1,5}$ in Fig 1 and the location of the closest detectors to this

Fig 4. Location of the 81 outbreak positions indicated with small squares and the 16 detectors as filled circles. For illustration, only a few of the detector locations are labeled.

<https://doi.org/10.1371/journal.pone.0225593.g004>

point in Fig 4. The signal arrives later to the other detectors, with different time dependence and amplitudes.

Finally, with these specifications, the training set consists of the results generated by all the combinations of the parameters of position, diffusion and proliferation defined above. Each of them has 9 values, which means that we solve the direct problem 9^4 times, and produce the time series of 250 data values measured by each of the 16 detectors. These 16×9^4 time series constitute the set used to train the SVM in the most general case, for the second of the inverse problems. For the first inverse problem, the training set is constructed with the same settings as for the general case, except that D_7 and ρ_0 are set fixed.

2.4 The SVM classification method

Among the various machine learning methods, we account with implementations of artificial neural networks (ANNs) as regressors and classifiers, so as convolutional neural networks (CNNs). The use of ANNs has the advantage of easy implementation and high adaptability to non-linear problems. Nevertheless, the computational time required to train the network grows with the number of inputs, in our case, the data set in the 16 time series measured by the detectors; it also grows with the number of neurons in the hidden layers; finally it also

Fig 5. Time series of the variable u measured at the 16 detectors $u(x_{det}^{k,l}, t)$ measured by the 16 detectors for $D_2 = 0.95$, $\rho_0 = -0.25$, and outbreak centered at $x_{src}^{5,1}$. As described in the text, each detector captures 250 values of u in the time domain.

<https://doi.org/10.1371/journal.pone.0225593.g005>

increases with the number of classes. The number of inputs and classes used in our analysis is prohibitive for ANNs, both as classifiers or regressors. A second option are the CNNs, which are known for their efficiency at classifying images. In this case we deal with time-series instead of images, which is the reason we prefer to use Support Vector Machines.

The specific SVM method used in our analysis, consists in constructing the optimal decision function by obtaining the solution of an optimization problem subject to constraints with a radial basis function kernel with two free parameters. In order to select the best possible values for these parameters, we perform a parameter search known as *cross-validation*. Basically, the cross-validation separates the training set into a fixed number of subsets and, sequentially, each one of those subsets is used to test the accuracy of the decision function obtained using the rest of the subsets. For a detailed explanation of the methods the reader can consult [21, 22] and for a specific setup of our implementation in a similar analysis [23]. Instead of implementing our own version of the support vector machine, we use the library libSVM [22].

3 Results

We aim to classification of two inverse problems, the one focused on the location of the outbreak position only, and the second one intended for the classification of the four parameters.

The key performance measure used in our analysis is the percentage of correctly classified cases within the prediction set. The accuracy is the average of accuracy results obtained using various randomly generated prediction sets, with the aim of minimizing fluctuations.

3.1 Location of the source only

Training accuracy. The SVM was trained with all the time series registered by detectors considering the 81 source locations $x_{src}^{i,j}$ according to Eq (5). The accuracy in classification of

the SVM at training was above 95%. The SVM with this training is the one used to estimate the accuracy in prediction.

Prediction sets. In this case the prediction set consists of 225 simulations with randomly generated coordinate position of the source $\mathbf{x}_{src}^{pred}$ and fixed values of diffusion and proliferation parameters D_1, ρ_0 . We make sure the coordinates do not coincide with any of the positions $\mathbf{x}_{src}^{pred} \neq \mathbf{x}_{src}^{ij}$ used for training, which is a modelling assumption that avoids the trivial cases. When the SVM is used to classify the outbreak location only, the accuracy of the method at prediction is 90.0%.

3.2 Location of the source plus environmental parameters

Training accuracy. In this case the SVM was trained with the time series of the detectors considering the 81 source locations, the 9 values of D_2 and the 9 values of ρ_0 defined before. In total 81^2 simulations for 9^2 classes. The accuracy in classification at training was above 80%.

Prediction. In this case we prepared 5 different sets with 1000 runs each, using random values of location coordinates, D_2 and ρ_0 , within the domain values used so far, but unknown to the SVM. The average accuracy among the prediction sets in the four parameter case is of 77.72%.

4 Conclusions and final comments

We have presented the effectiveness of an SVM to classify the location of an outbreak, and parameters of diffusion and proliferation of a process ruled by the RD equation in a domain. These parameters are constructed out of measurements obtained at given detectors in the domain of interest.

When a set of parameters is correctly classified it means that such parameter values lie within the range associated with the class selected by the SVM. For example, the domain $A = [-1, 1] \times [-1, 1]$ is covered by nine classes corresponding to squares of sides $2/3$. When the SVM correctly classifies the location of the outbreak, it means that the possible location will be associated to the center of the class $\pm 1/3$ along each direction. The same argument is applied for the proliferation and diffusion parameters.

Improvement in accuracy needs classes of smaller sizes. For a fixed domain like A , it means the use of a bigger set of simulations for training, which is affordable if the problem consists only in the location of the outbreak. For example, it is easy to cover the domain with 6×6 classes instead of 3×3 to reduce the uncertainty from $\pm 1/3$ to $\pm 1/6$ in the location of the source. The price that has to be paid is to increase the number of elements in the training set by a factor of 2^2 . Nevertheless when the parameter space is four dimensional the number of elements in the training set has to be increased to 2^4 which is unpractical.

Instead, to decrease the uncertainty in the values of parameters, we can use a mesh refinement of the parameter space, which implies a reduction in size for a specific class, and consequently, the reduction of uncertainty. For example, in Fig 6 we illustrate with an example in two dimensions the case in which the SVM predicts that the source of an outbreak lies in the up-center class (see left-bottom square). A refinement of this domain is done and a new training within this subdomain determines that the source of the outbreaks will be located at the bottom-center class of that subdomain (see center-up square). In a further refinement, the SVM indicates that the source is located in the center-right class (see right-bottom square). With this process we have decreased the uncertainty in the prediction by a factor of $1/9$. This procedure has been successfully implemented in other classification problems such as [24, 25]. In fact, this is a strategy to tackle inverse problems associated with initial value problems ruled by partial differential equations with a potential variety of applications.

Fig 6. This figure illustrates how the method can be used on refined subdomains in order to reduce the uncertainty of the parameters in each class.

<https://doi.org/10.1371/journal.pone.0225593.g006>

Finally, considering a real scenario problem, like the spread of a vector that transmits a disease, the domain could be not as regular as the one used here, and the detectors may not be located in such a regular grid type of distribution. Nevertheless, the advantage of the method presented here is that it can be easily generalized to general shaped domains and irregular distributions of detectors. Two immediate applications of our method can be the propagation of mosquitoes at country and city scale including corridor components [6] and modeling the evolution of environmental parameters in ecological systems due to the impact of climate change [11].

With respect to the computing power required for this analysis, all the runs, those involving the numerical solution of the RD problem, so as the runs with SVM were carried out in regular computers equipped with 2.50GHz Xeon processors without any special requirements.

Acknowledgments

The simulations were carried out in the computer farm and the Big Mamma cluster at the IFM.

Author Contributions

Investigation: Venecia Chávez-Medina, José A. González, Francisco S. Guzmán.

Software: Venecia Chávez-Medina, José A. González, Francisco S. Guzmán.

Writing – original draft: Venecia Chávez-Medina, José A. González, Francisco S. Guzmán.

Writing – review & editing: Venecia Chávez-Medina, José A. González, Francisco S. Guzmán.

References

1. Takahashi LT, Maidana NA, Castro Ferreira W, Pulino P, Yang HM. Mathematical models for the *Aedes aegypti* dispersal dynamics: travelling waves by wing and wind Bull. Math. Biol. 2005; 67:509–528. <https://doi.org/10.1016/j.bulm.2004.08.005>
2. Tian C, Ruan S. A free boundary problem for *Aedes aegypti* mosquito invasion. Appl. Math. Modeling 2017; 46:203–217 <https://doi.org/10.1016/j.apm.2017.01.050>
3. Lewis MA, Renclawowicz J, van den Driessche P. Traveling waves and spread rates for a West Nile Virus model 2006; Bull. Math. Biol. 68:3–23. <https://doi.org/10.1007/s11538-005-9018-z> PMID: 16794919
4. Ou C, Wu J. Spatial spread of rabies revisited: influence of age-dependent diffusion on nonlinear dynamics. SIAM J. Appl. Math. 2006; 67:138–164. <https://doi.org/10.1137/060651318>
5. Hilker FM, Langlais M, Petrovskii SV, Malchow H. A diffusive SI model with Allee effect and application to FIV. Math. Biosci. 2006; 206:61–80. <https://doi.org/10.1016/j.mbs.2005.10.003> PMID: 16387332
6. Roques L, Bonnefon O. Modelling Population Dynamics in Realistic Landscapes with Linear Elements: A Mechanistic-Statistical Reaction-Diffusion Approach. PLoS ONE 2016; 11(3):e0151217. <https://doi.org/10.1371/journal.pone.0151217> PMID: 26986201
7. Berendsen J, Burger M, Pietschmann JF. On a cross-diffusion model for multiple species with nonlocal interaction and size exclusion. Nonlinear Analysis 2017; 159:10–39. <https://doi.org/10.1016/j.na.2017.03.010>
8. Lewis MA, Petrovskii S, Potts JR. The Mathematics Behind Biological Invasions, Springer 2016.
9. Soubeyrand S, Roques L. Parameter estimation for reaction-diffusion models of biological invasions. Pop. Ecology 2014; 56(2):427–434. <https://doi.org/10.1007/s10144-013-0415-0>
10. Madzvamuse A, Barreira R, Gerisch A. Cross-Diffusion in Reaction-Diffusion Models: Analysis, Numerics, and Applications. J. Math. Biol. 2015; 70(4):709–743.
11. Roitberg E, Shoshany M. Can spatial patterns along climatic gradients predict ecosystem responses to climate change? Experimenting with reaction-diffusion simulations. PLoS ONE 2017; 12(4): e0174942. <https://doi.org/10.1371/journal.pone.0174942> PMID: 28394914
12. Araujo MB, Guisan A. Five (or so) challenges for species distribution modeling. J. Biogeogr. 2006; 33:1677–1688. <https://doi.org/10.1111/j.1365-2699.2006.01584.x>
13. Austin MP, Van Niel KP. Improving species distribution models for climate change studies: variable selection and scale. J. Biogeogr. 38, 1–8 (2011) <https://doi.org/10.1111/j.1365-2699.2010.02416.x>
14. Carrillo M, Que U, González JA. Estimation of Reynolds number for flows around cylinders with lattice Boltzmann methods and artificial neural networks. Phys. Rev. E 2016; 94:063304. <https://doi.org/10.1103/PhysRevE.94.063304> PMID: 28085295
15. Carrillo M, Que U, González JA, López CE. Recognition of an obstacle in a flow using artificial neural networks. Phys. Rev. E 2017; 96:023306. <https://doi.org/10.1103/PhysRevE.96.023306> PMID: 28950464
16. Fisher RA. The wave of advance of advantageous genes. Annals of Eugenics 1937; 7(4):353–369. <https://doi.org/10.1111/j.1469-1809.1937.tb02153.x>
17. Kolmogorov A, Petrovsky I, Piskounov N. Travelling Wave Solution of the Fisher-Kolmogorov Equation with Non-Linear Diffusion. Moscow University Mathematics Bulletin 1989; p 105.
18. Shu CW, Osher S. Efficient implementation of essentially non-oscillatory shock-capturing schemes. J. Comput. Phys. 1989; 83:3278. [https://doi.org/10.1016/0021-9991\(89\)90222-2](https://doi.org/10.1016/0021-9991(89)90222-2)
19. Shigesada N, Kawasaki K. Biological Invasions: Theory and Practice. Oxford Series in Ecology and Evolution, Oxford: Oxford University Press; 1997.
20. Beier P, Noss RF. Do habitat corridors provide connectivity? Conserv Biol. 1998; 12(6):1241–1252. <https://doi.org/10.1111/j.1523-1739.1998.98036.x>
21. Abe S *Support Vector Machines for Pattern Classification*. Springer, 2005.
22. Chang CC, Lin CJ. LIBSVM: A library for support vector machines. ACM transactions on intelligent systems and technology 2017; 2(3):1–27. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
23. González JA, Guzmán FS. Classification of a black hole spin out of its shadow using support vector machines. Phys. Rev. D 2019; 99: 103002.
24. Rivera-Paleo FJ, López-Núñez CE, Guzmán FS, González JA. Classifying initial conditions of long GRBs modeled with relativistic radiation hydrodynamics. Phys. Rev. D 2017; 95:123005. <https://doi.org/10.1103/PhysRevD.95.123005>
25. González JA, Guzmán FS. Characterizing the velocity of a wandering black hole and properties of the surrounding medium using convolutional neural networks. Phys. Rev. D 2018; 97:063001. <https://doi.org/10.1103/PhysRevD.97.063001>