

# Application of satellite image to the implementation of two stochastic models for modeling the transport of chlorophyll-a on Lake Valencia (Venezuela) \*

**Titre:** Application des images satellitaires à la mise en œuvre de deux modèles stochastiques pour modéliser le transport de chlorophylle-a sur le Lac de Valence du Venezuela

Maira Valera-López<sup>1</sup>, Angie Pineda<sup>1</sup> and José R. León<sup>1</sup>

**Abstract:** In this article, two statistical methods of diffusion of particles are proposed for modeling the transport of pollutants in Lake Valencia (Venezuela). Both methods use as velocity field on the lake surface solutions of the Saint-Venant equations. The two procedures are compared through Monte Carlo simulation. Furthermore, as a novelty an algorithm to randomly generate the initial positions of the particles using information obtained from a satellite image is designed.

**Résumé :** Dans cet article deux méthodes de diffusion de particules sont proposés pour modéliser le transport de polluants à la surface du Lac de Valencia (Venezuela). Les deux méthodes utilisent comme champ de vitesses sur la surface du lac les solutions de l'équation de Saint-Venant. Ces deux procédures sont comparées à l'aide d'une simulation de Monte Carlo. De plus, et comme une nouveauté, un algorithme qui utilise l'information obtenue d'une image satellite est construit pour générer les positions initiales des particules.

**Keywords:** diffusion of particles, random flight, contamination of water surface, Saint-Venant equations, satellite image

**Mots-clés :** diffusion de particules, vol aléatoire, pollution de surface d'eau, équation de Saint-Venant, image de satellite

**AMS 2000 subject classifications:** 60H10, 60H35, 60J60, 62G07, 62H35

## 1. Introduction

The increase in environmental awareness and the need to predict and improve water quality in lakes, estuaries and coastal seas, has led to significant advances in the modeling of contaminant transport. Hydrodynamic and pollutants transport models in water bodies, require a detailed understanding of transport processes that exist within it. Essential elements of life and productivity such as oxygen, heat and nutrients are transported and dispersed through these processes.

Some geometries of enclosed water bodies such as lakes, reveal that scale of the horizontal length is much greater than the vertical length. This fact justifies a two-dimensional simulation of

\* The work of José R. León was partially supported by the Inria International Chairs program.

<sup>1</sup> Escuela de Matemáticas. Facultad de Ciencias. Universidad Central de Venezuela.

E-mail: [maira.valera@ciens.ucv.ve](mailto:maira.valera@ciens.ucv.ve)

E-mail: [angie.pineda@ciens.ucv.ve](mailto:angie.pineda@ciens.ucv.ve)

E-mail: [jose.leon@ciens.ucv.ve](mailto:jose.leon@ciens.ucv.ve)

the current flow in these waters bodies using shallow water model by Saint-Venant equations. The current flows can be induced by hydrodynamic circulation (wind-drive circulation), and, in this case, the vertical current distribution is almost uniform depth with strong gradient existing only in the bottom.

This current flow calculated, can be implemented by the advection-diffusion equation, random walk and random flight, which are classical models applied for describing pollutants dispersion in shallow water. The first has a deterministic implementation that combines the Fick's law, the law of conservation of mass and the advection equation to determine the concentration of pollution, and the last two are stochastic models that determine the trajectories at the fluid of each particle of pollutant.

On the other hand, random walk model and the scheme by advection-diffusion equation accurately describe particles dispersion if these have been in the flow over a Lagrangian time scale and have spread to cover a distance larger than the largest scale of a turbulent fluid flow (Fisher et al., 1979). The random flight model corrects this defect by modifying the diffusion terms present in random walk model. This modification allows a more precise description of the particles deployment and can be shown that, after long periods of simulation, this model is consistent with the advection-diffusion equation (Heemink, 1990).

The particle tracking method has been studied and applied in many other areas such as kinetic theory, physical of plasma, oilfields patterns, etc. These models are also popular in some field of flow dynamics for velocity field visualization and gaining an insight into various phenomena of flow. In 1985, Van Dop et al. (1985), developed a one-dimensional random flight model to describe unstable turbulent flows. An application of the model to predict the dispersion in convective boundaries layers is discussed by De Baas et al. (1980).

This research was inspired by the ideas established by Heemink in Heemink (1990), but in this time, we applied the numerical scheme described before (random walk and flight models), on Lake Valencia, Venezuela ( $10^{\circ}11'00''\text{N}$ ,  $167^{\circ}44'00''\text{W}$ ), which has an area of  $350\text{ km}^2$  and mean depth of 18 m, then being the largest fresh water body of Venezuela with no outlet to the sea. Specifically, we apply those two stochastic models mentioned, to describe the dispersion of organic wastes in this lake because its high pollution caused by wasting water from the surrounding urban areas, makes it an appropriate place to do this study. The two algorithms are constructed to solve each stochastic differential equation (SDE), which governs the pollutant particles dispersion. The solution of the first stochastic differential equation is an ordinary diffusion and the second one is a hypoelliptic diffusion. Another important achievement of our work is the implementation of reflection algorithm at the border, developed by Bossy et al. (2004). We use this algorithm for reflecting toward the lake's interior the pollutant particles arriving at its frontier. Given the complicate geometry of the Lake Valencia boundary, we also built a slight modification of this algorithm.

An important and significant novelty of this research is that we build an algorithm that randomly generates the position of the particles of pollutant using satellite image information. This simulates the actual contaminant dispersion into the lake, specifically chlorophyll-a, on the time when the image were captured by the satellite. The importance of this random generation method is that the initial position of the particles are modeled by random pairs of a probability density of organic waste into the lake. To carry out this, we construct a density from the vegetation index corresponding to a satellite image of Lake Valencia. With this density we define and execute the

Metropolis-adjusted version of the Langevin algorithm developed in Roberts and Tweedie (1996), from which are obtained the pairs of random numbers (weakly dependent pairs of numbers), mentioned above that model the chlorophyll-a distribution observed in the satellite image. It is important to point out that the present work offers a comprehensive and pedagogical development about how making predictions associated with the dispersion of pollutants. It should be noted that although the goodness of the estimation is untested at this time, this should be the task of another research.

In order to summarize the structure of this research, then we present a sketch thereof. We first describe theoretically the random walk and flight models. Then, we present a short review about the hydrodynamic model with which we obtained the Lake Valencia's advective flow rates, generated in Valera-López et al. (2014, 2016). Additionally, to explain the process involving a digital processing of the satellite image to obtain the vegetation index, we expose the satellite image section. Thereupon, we describe the applied Monte Carlo method. Then, the results section is developed. In such a section we discuss the implementation of these stochastic methods with two types of initial distributions: initial distribution concentrated at a fixed point and initial distribution that simulate the proportions of chlorophyll-a on Lake Valencia. Finally, conclusion section is presented. Moreover, the Langevin equation is described in an appendix. This appendix is dedicated to establishing elements to simulate efficiently probability densities on relatively compact subsets of  $\mathbf{R}^d$ , for any positive integer  $d$ . This justifies the generation of pairs of random numbers, using the Langevin equation constrained by boundary conditions in a domain that is determined by the presence of chlorophyll-a on Lake Valencia.

## 2. Stochastic model: random walk

To model the dispersion of pollutants in shallow water media, we develop a random walk model consistent with the advection-diffusion equation with integrated depth. In the model that will be described below, the pollutant is conceived as a particles collection on the surface of a medium shallow water with depth  $H(x,y)$  and mean flow velocities  $u(x,y)$  and  $v(x,y)$  in  $x$  and  $y$  directions respectively; accordingly, the path of the contaminant will be molded by the corresponding particles path on the water surface.

The speed  $u$  and  $v$  are obtained from a hydrodynamic model to be developed in section 4.

As we know, the advection-diffusion equation with integrated depth is,

$$\frac{\partial(HC)}{\partial t} + \frac{\partial(CuH)}{\partial x} + \frac{\partial(CvH)}{\partial y} = \frac{\partial}{\partial x} \left( HD \frac{\partial C}{\partial x} \right) + \frac{\partial}{\partial y} \left( HD \frac{\partial C}{\partial y} \right), \quad (1)$$

where  $t$  is the time variable,  $D(x,y)$  denotes the dispersion coefficient and  $C(x,y,t)$  the particle concentration; it is a deterministic model that describes with precision particle dispersion because it responds to fundamental laws of natural science. The  $C$  concentration satisfying this differential equation is associated with the density  $p(x,y,t)$  corresponding to the distribution of particles in water by the identity

$$C(x,y,t) = \frac{p(x,y,t)}{H(x,y)}. \quad (2)$$

Substituting (2) in the advection-diffusion equation (1), the Fokker-Planck equation is obtained,

$$\begin{aligned} \frac{\partial p}{\partial t} = & -\frac{\partial}{\partial x} \left( \left[ u + \left( \frac{\partial H}{\partial x} D \right) \frac{1}{H} + \frac{\partial D}{\partial x} \right] p \right) - \frac{\partial}{\partial y} \left( \left[ v + \left( \frac{\partial H}{\partial y} D \right) \frac{1}{H} + \frac{\partial D}{\partial y} \right] p \right) \\ & + \frac{1}{2} \frac{\partial^2}{\partial x^2} (2Dp) + \frac{1}{2} \frac{\partial^2}{\partial y^2} (2Dp). \end{aligned} \quad (3)$$

The differential equation (3) with initial condition  $p(x, y, t_0) = \delta(x - x_0)\delta(y - y_0)$  coincides with the forward Kolmogorov equation. The solution of this Kolmogorov equation is the transition probability density of the Markov process  $\{(X_t, Y_t)\}_{t \geq t_0}$  which represents particles position on the surface water body under study and satisfies the following stochastic differential equations (SDEs):

$$dX_t = \left[ u + \left( D \frac{\partial H}{\partial x} \right) \frac{1}{H} + \frac{\partial D}{\partial x} \right] dt + \sqrt{2D} dW_t^1, \quad (4)$$

$$dY_t = \left[ v + \left( D \frac{\partial H}{\partial y} \right) \frac{1}{H} + \frac{\partial D}{\partial y} \right] dt + \sqrt{2D} dW_t^2, \quad (5)$$

where  $t \geq t_0$ ,  $X_{t_0} = x_0$ ,  $Y_{t_0} = y_0$ ,  $(x_0, y_0) \in \mathbf{R}^2$  and  $\mathbf{W}_t = (W_t^1, W_t^2)^T$  is a two-dimensional Brownian motion which verifies,

$$\mathbf{E}(d\mathbf{W}_t, d\mathbf{W}_t^T) = I_2 dt, \quad (6)$$

where  $I_d$  is the  $d \times d$  identity matrix. These results and others similar can be consulted in [Karatzas and Shreve \(1991\)](#). The consequence of the description above is that the distribution of  $(X_t, Y_t)$  coincides with the particles distribution whose concentration  $C(x, y, t)$  satisfies the advection-diffusion equation; this proves the consistency between the two models and justifies the use of the paths of the diffusion  $\{(X_t, Y_t)\}_{t \geq t_0}$  to simulate the trajectories of each particle of the pollutant.

Formally, random walk model associates to each particle of pollutant injected at time  $t = t_0$ , at the point  $(x_0, y_0)$  of the surface water study, a path of the diffusion  $\{(X_t, Y_t)\}_{t \geq t_0}$ , with initial condition  $(X_{t_0}, Y_{t_0}) = (x_0, y_0)$  to describe their movement.

When the water body surface considered, denoted by  $\mathcal{D}$ , has a closed border  $\partial\mathcal{D}$ , the Neumann boundary condition of the advection-diffusion equation  $\nabla C \cdot \mathbf{n} = \partial C / \partial \mathbf{n} = 0$ , is used to simulate the fact that there is no particles transfer across the boundary, i.e. any solid particle crosses the border to be reflected back into the flow domain thus keeping the conservation of mass.

For the random walk model, the corresponding Neumann condition is incorporated to its SDEs (4)-(5) in the form of reflection term as follows:

$$dX_t = \left[ u + \left( D \frac{\partial H}{\partial x} \right) \frac{1}{H} + \frac{\partial D}{\partial x} \right] dt + \sqrt{2D} dW_t^1 + \gamma_1 dk_t, \quad (7)$$

$$dY_t = \left[ v + \left( D \frac{\partial H}{\partial y} \right) \frac{1}{H} + \frac{\partial D}{\partial y} \right] dt + \sqrt{2D} dW_t^2 + \gamma_2 dk_t, \quad (8)$$

where  $t \geq t_0$ ,  $X_{t_0} = x_0$ ,  $Y_{t_0} = y_0$  and  $\gamma = (\gamma_1, \gamma_2)$  is a vector field with norm one, that defines the direction of reflection and  $\{k_t\}_t$  is a process increasing only when  $(X_t, Y_t)$  is on  $\partial\mathcal{D}$  such that  $k_t = \int_0^t \mathbf{1}_{\{(X_s, Y_s) \in \partial\mathcal{D}\}} dk_s$ . The process  $\{k_t\}_t$  determines the size of the reflection (see [Bossy et al., 2004](#)).

### 3. Stochastic model: random flight

According to Fisher et al. (1979), in reality, for short periods of time after particles injection in a turbulent water medium, their speeds are very close to a constant, and for long periods of time after, their movements become less and least correlated at longer and longer times. The random walk model (and consequently advection-diffusion equation), models this behavior only for long periods of simulation because, in this model, random particles displacement are simulated by increments of a Brownian motion which are uncorrelated due to its independence. The random flight model corrects this defect by replacing the diffusion terms of the SDEs in (4) - (5) by an Ornstein-Uhlenbeck processes, which accurately describe the actual behavior of the particle velocities at times  $t$  less than  $T_L$ , where  $T_L$  denotes a Lagrangian time scale. This scale, which models the time from which the particles movements are increasingly uncorrelated, it is interpreted as a measure of how long the particle takes to lose memory of its initial turbulent velocity (Griffa, 1996; Heemink, 1990). This last article is our main reference for the equation and their parameters. The equations that describe the random flight model are as follow:

$$dX_t = \left[ u + U_t + \left( D \frac{\partial H}{\partial x} \right) \frac{1}{H} + \frac{\partial D}{\partial x} \right] dt, \quad t \geq t_0, \quad (9)$$

$$dY_t = \left[ v + V_t + \left( D \frac{\partial H}{\partial y} \right) \frac{1}{H} + \frac{\partial D}{\partial y} \right] dt, \quad t \geq t_0, \quad (10)$$

$$dU_t = - \left( \frac{1}{T_L} \right) U_t dt + (\hat{K})^{1/2} dW_t^1, \quad t \geq t_0, \quad (11)$$

$$dV_t = - \left( \frac{1}{T_L} \right) V_t dt + (\hat{K})^{1/2} dW_t^2, \quad t \geq t_0, \quad (12)$$

with  $\hat{K} = 2D/T_L^2$  where  $D$  is again the dispersion coefficient,  $X_{t_0} = x_0$ ,  $Y_{t_0} = y_0$  and,  $U_{t_0}$  and  $V_{t_0}$  are independent random variables both with Gaussian distribution of zero mean and variance  $\frac{\hat{K}T_L}{2}$ . In the above system  $(X_t, Y_t)$  represents the particle position on the surface of the water body under study, and  $(U_t, V_t)$  represents the stochastic particle velocities induced by the turbulent fluid flow and the two process are independent.

Diffusion processes  $(U_t)_{t \geq t_0}$  and  $(V_t)_{t \geq t_0}$ , are Gaussian processes centered and stationary with autocovariance function given by:

$$\mathbf{E}(U_{t+\tau}U_t) = \mathbf{E}(V_{t+\tau}V_t) = \frac{\hat{K}T_L}{2} e^{-|\tau|/T_L}. \quad (13)$$

The identity in (13) shows that the autocorrelations of  $U_\tau$  and  $V_\tau$  are close to one for  $0 < \tau \ll T_L$ , for instance if for some  $\delta > 0$  small enough, we have  $\frac{\tau}{T_L} < \delta$  then  $\text{Corr}(U_\tau, U_0) \geq e^{-\delta}$ . However, if  $\tau \gg T_L$  the autocorrelations tend to zero. This simulates very well the behavior of the random velocities of the particles in the flow.

For times  $t$  such that  $t \gg T_L$ , this model is consistent with the advection-diffusion equation (1), since for  $(t - t_0)/T_L \rightarrow \infty$  or  $T_L \rightarrow 0$ , equations (9) - (12) reduce to stochastic differential equations:

$$dX_t = \left[ u + \left( D \frac{\partial H}{\partial x} \right) \frac{1}{H} + \frac{1}{2} \frac{\partial D}{\partial x} \right] dt + \sqrt{2D} dW_t^1 \quad (14)$$

$$dY_t = \left[ v + \left( D \frac{\partial H}{\partial y} \right) \frac{1}{H} + \frac{1}{2} \frac{\partial D}{\partial y} \right] dt + \sqrt{2D} dW_t^2, \quad (15)$$

where  $t \geq t_0$ ,  $X_{t_0} = x_0$  and  $Y_{t_0} = y_0$ .

We can sketch the procedure for obtaining the above limit. Let us consider for simplicity that  $D$  is a constant and we will work with only one coordinate. Hence the random fly model writes

$$dX_t = \left[ u + U_t + \frac{D}{H(X_t)} \frac{\partial H(X_t)}{\partial x} \right] dt \quad (16)$$

$$dU_t = -\frac{1}{T_L} U_t dt + \frac{\sqrt{2D}}{T_L} dW_t.$$

In this form we have

$$\int_s^t U_u du = T_L U_s - T_L U_t + \sqrt{2D}(W_t - W_s).$$

Yielding in (16)

$$X_t^{T_L} - X_s^{T_L} = u(t-s) + T_L U_s - T_L U_t + \sqrt{2D}(W_t - W_s) + D \int_s^t \frac{1}{H(X_u^{T_L})} \frac{\partial H(X_u^{T_L})}{\partial x} du.$$

Let assume that function  $\frac{1}{H} \frac{\partial H}{\partial x}$  is Lipschitz. Then making the difference between the above process and the following diffusion

$$Y_t - Y_s = u(t-s) + \sqrt{2D}(W_t - W_s) + D \int_s^t \frac{1}{H(Y_u)} \frac{\partial H(Y_u)}{\partial x} du,$$

it can be shown that for all compact  $A \subset \mathbb{R}$  it holds that  $\sup_{t \in A} |X_t^{T_L} - Y_t| \rightarrow 0$  when  $T_L \rightarrow 0$ , see [Nelson \(1967\)](#) pages 71 – 76 for a similar result.

The equations (14) - (15) have to be interpreted, by definition, in the Stratonovich's sense ([Arnold, 1974](#)), but this fact occurs only when  $D$  depends on  $(x, y)$ . The Itô equations corresponding to the above equations are (4) - (5), which define the random walk model. Consequently, the asymptotic model (14) - (15) is consistent with advection-diffusion equation (1).

The diffusion process  $\{(X_t, Y_t, U_t, V_t)\}_{t \geq t_0}$  is a Markov process whose transition probability density satisfies the hypoelliptic Fokker-Planck equation:

$$\begin{aligned} \frac{\partial p}{\partial t} = & -\frac{\partial}{\partial x} \left( \left[ u + u_1 + \left( \frac{\partial H}{\partial x} D \right) \frac{1}{H} + \frac{\partial D}{\partial x} \right] p \right) \\ & -\frac{\partial}{\partial y} \left( \left[ v + v_1 + \left( \frac{\partial H}{\partial y} D \right) \frac{1}{H} + \frac{\partial D}{\partial y} \right] p \right) \\ & + \frac{1}{2} \frac{\partial^2}{\partial u_1^2} (\hat{K} p) + \frac{1}{2} \frac{\partial^2}{\partial v_1^2} (\hat{K} p) + \frac{\partial}{\partial u_1} \left( \frac{p u_1}{T_L} \right) + \frac{\partial}{\partial v_1} \left( \frac{p v_1}{T_L} \right), \end{aligned} \quad (17)$$

with initial condition

$$p(x, y, u_1, v_1, t_0) = \delta(x - x_0) \delta(y - y_0) \frac{1}{\pi \hat{K} T_L} e^{-\frac{u_1^2 + v_1^2}{\hat{K} T_L}}. \quad (18)$$

The reader is referred to the book of Soize (1994) and the references therein for more about Fokker-Planck equations.

Models valid for more general assumptions are available: a generalization of random flight model to inhomogeneous and non stationary turbulence can be reviewed in subsection 4.2 of Griffa (1996) and the autocovariance functions of the speed processes in (13) can be modeled more accurately by entering equations for acceleration and higher order derivatives of the position (see Griffa, 1996).

Analogous to the case of random walk model, the Neumann boundary condition for the random flight model is incorporated as reflection terms of the particle position in its displacement equations (9) and (10).

#### 4. Hydrodynamic Model

Hydrodynamic modeling in closed water bodies requires a detailed knowledge of the transport process within the body.

The Saint-Venant equation, simulates the propagation of disturbances in water and other incompressible flows. The underlying assumption is that the depth of the medium is small compared to the wavelength of the disturbance. Lake Valencia retains this condition, so that the use of the Saint-Venant equation provides a reasonable model for this implementation.

We developed a two dimensional hydrodynamic model that solves the Saint-Venant equations:

$$\begin{aligned} \frac{\partial}{\partial t} \begin{pmatrix} H \\ uH \\ vH \end{pmatrix} + \frac{\partial}{\partial x} \begin{pmatrix} uH \\ u^2H + gH^2/2 \\ uvH \end{pmatrix} + \frac{\partial}{\partial y} \begin{pmatrix} vH \\ uvH \\ v^2H + gH^2/2 \end{pmatrix} \\ = \begin{pmatrix} 0 \\ -gH \frac{\partial b}{\partial x} + \varepsilon \Delta(Hu) + \rho^{-1} \tau_x \\ -gH \frac{\partial b}{\partial y} + \varepsilon \Delta(Hv) + \rho^{-1} \tau_y \end{pmatrix}, \end{aligned} \quad (19)$$

where  $H = H(x, y, t)$  is the depth and  $b = b(x, y)$  is the bathymetry that is positive measured downward from the geoid. Usually, depth  $H$  is define as  $H = b + \eta$ , where  $\eta = \eta(x, y, t)$  represent the free surface elevation relative to the geoid, but in this research we consider  $\eta = 0$  because it is observed that waves on the lake have very small height, thus  $H$  does not depend on  $t$ , i.e.,  $H = H(x, y)$ .

The source term contains the following expressions:  $g$  is gravity,  $\tau_x = \tau_{w_x} - \tau_{s_x}$  and  $\tau_y = \tau_{w_y} - \tau_{s_y}$  determine shear, with  $\tau_{w_x}$ ,  $\tau_{w_y}$  shear stream at the bottom of the lake and  $\tau_{s_x}$ ,  $\tau_{s_y}$  wind shear over

the lake surface in  $x$  and  $y$  directions, respectively,  $\rho$  is the density,  $\varepsilon$  is viscosity parameter or regularization and  $\Delta$  is the Laplacian (García and Kahawita, 1986; Tsanis and Saied, 2007).

These differential equations must be solved with an appropriate set of initial and boundary conditions. Initial conditions on the velocities field  $(u, v)$  in  $(t = 0)$  are introduced:  $u(x, y, 0) = 0$  and  $v(x, y, 0) = 0$ . The boundary conditions are closed and defined the calculation field to where the water surface is in contact with the ground. This type of boundary condition is sometimes called waterproof condition, and is interpreted as a null normal speed in respect to the normal edge defining the integration region. However, in this study the viscosity parameter is not null, which makes this formulation of the Saint-Venant equations (19), slightly parabolic. Consequently, it becomes necessary to change or supplement the impermeability boundary condition with the zero speed, sometimes referred to as non-slip condition (García and Kahawita, 1986).

The explicit finite-difference method used in this work is based on the MacCormack time-splitting scheme, see MacCormack (1971). The MacCormack scheme is a fractional-step method where a complicated finite difference operator is "split" into a sequence of simpler ones. The splitting process reduces the number of calculations during each time's step and achieves second-order accuracy in space and time when a symmetric sequence of operators is used.

In order to incorporate numerically the Saint-Venant equations on the physical continuous region are overlapped with a computational grid where all dependent variables are defined at the cell centers (fully dense grid). These values are taken to represent average cell properties. The use of fully dense grids are conceptually more consistent than staggered grids which usually generate excessive numerical diffusion, owing to the need to take averages across cell faces.

This adaptation of MacCormack numerical scheme (predictor / corrector) is developed in Valera-López et al. (2014, 2016).

Scheme stability is normally determined by the Courant-Friederichs-Lewy (Courant et al., 1967) condition. However, the particular character of the discretized equation results in a somewhat relaxed criterion, i.e.:

$$\Delta t \leq \min \left[ \left( \frac{1}{\Delta x^2} + \frac{1}{\Delta y^2} \right)^{-1/2} (gH)^{-1/2} \right], \quad (20)$$

where  $\Delta x$ ,  $\Delta y$  and  $\Delta t$  are the dimensions of the blocks of the mesh and the simulation time's step, respectively.

The results obtained after implementation on Lake Valencia, which are used for the development of this work can be seen in the Figures 1(a) and 1(b).

## 5. Satellite image

Since the organic wastes in water are mainly housed in elements containing chlorophyll (such as algae), our main interest is to model the transport of organic wastes on the lake, determined by the presence of chlorophyll therein, which is obtained from satellite image. This image is obtained from a satellite that shows the geography of a specific territory, in our case Lake Valencia. The photograph taken by a satellite is supplied along with the images of each spectral band. The image of each band is represented by a matrix whose inputs are associated with values on a discretization of the captured surface. An image can be analyzed and interpreted by comparing images from



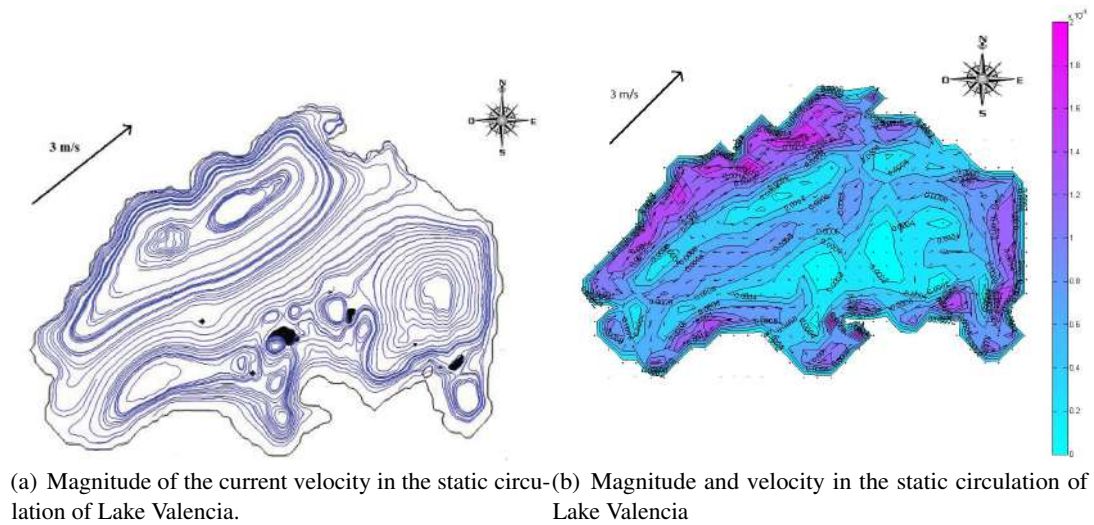


FIGURE 1. Graphical representation of the result obtained by Hydrodynamic model

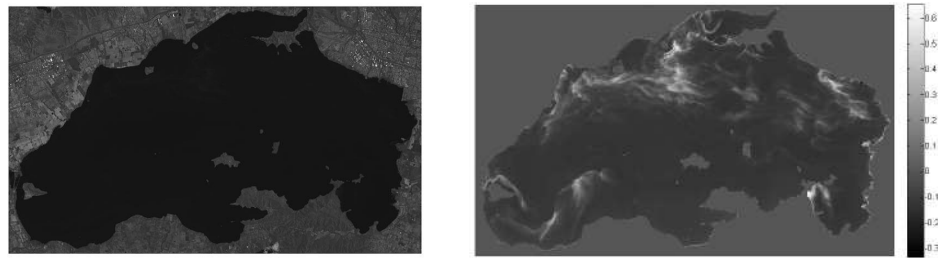
different spectral combinations. Satellites have different sensors to capture images which bands are located in a visible region (blue, green or red) and an infrared region (near, medium or far).

To carefully measure the wavelength and intensity of visible and near-infrared light that reflects the Earth's surface into space, a "vegetation index" is used to quantify the concentrations of leafy green vegetation on the planet. To determine the density of green in a region, you must observe the different colors (wavelengths), of visible and near-infrared light reflected by plants. The pigment of the leaves of plants, chlorophyll, strongly absorbs visible light for use in photosynthesis. We use the presence of chlorophyll in Lake Valencia to model the dispersion of organic waste in the lake because these contain nutrients that are housed in some bioindicators such as algae, see [Arcos et al. \(2005\)](#); [Lanza et al. \(2000\)](#); [Fan et al. \(2014\)](#); [Fei et al. \(2012\)](#); [Li et al. \(2010\)](#); [Ostapczuk et al. \(1997\)](#); [Peña et al. \(2005\)](#) and [Zhuowei et al. \(2010\)](#). On the other hand, the cellular structure of the leaves reflects a light near enough infrared light. If there is much more reflected radiation in the near infrared than in the visible wavelengths, it is likely that pixel vegetation is dense. If there is little difference in intensity of the reflected wavelengths of visible and near-infrared light, the vegetation is probably low. When sunlight collides with the object, certain wavelengths of this spectrum are absorbed and others are reflected. All the algae contain chlorophyll-a, therefore, the detection of these bioindicators over the lake is done by considering the absorption of visual radiation of this type of chlorophyll-a through a satellite image ([Awad, 2014](#)).

Almost all vegetation indices from satellite employ this difference formula to quantify the density of plant growth on Earth. The result of this formula is called normalized difference vegetation index (*NDVI*), and is the result of the following operation between red and infrared bands,  $B3$  and  $B4$ , respectively:

$$NDVI = \frac{B4 - B3}{B4 + B3}. \quad (21)$$

The array *NDVI* of the processed image, measures the vegetation index in a rectangular

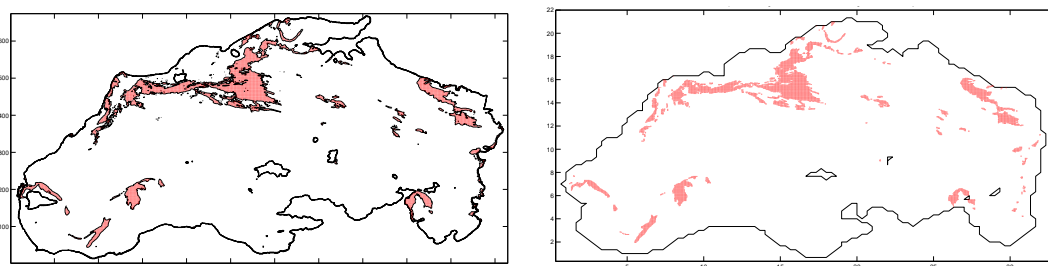


(a) Band 6 image of the satellite images down-loaded. (b) Processed image where the  $NDVI$  is evidenced.

FIGURE 2. Data obtained from the satellite images of the Lake Valencia captures by the satellite Landsat 8 at 17 : 15 : 59, April 07, 2014.

discretization of the region  $\mathcal{R}$ , captured by the satellite images, that contains the lake surface  $\mathcal{D}$ , (see Figure 2(a)). This matrix determines a coordinate system over  $\mathcal{R}$  considering to each pairs of indices  $(i, j)$  of  $NDVI$  as the coordinates of the point  $(x, y) \in \mathcal{R}$  where  $NDVI(i, j)$  was computed.

Values of  $NDVI$  in the interval  $(0, 1)$  represent the presence of chlorophyll in the lake (see Figure 3(b)), and determine a domain  $\mathcal{C} \subset \mathcal{D}$  formed by a collection of relatively compact subsets (see Figure 3(a)). This domain contains all the possible initial positions for the chlorophyll particles whose trajectories can be simulated to describe the dispersion of this substance from the time in that the images were captured by the satellite. To approximate these trajectories, their initial positions and the mean flow velocity of the lake would be defined on the same cartesian coordinate system. For this reason, the coordinates of the initial positions described above, are transformed to their corresponding coordinates on a coordinate system set to the approximate dimensions of the lake. This coordinate system corresponds to the mesh where the static circulation on the lake is calculated (see Figure 3(a) and 3(b)).



(a) All possible initial positions for the particles of chlorophyll.

(b) Cartesian coordinate system adjusted.

FIGURE 3. Data obtained from the satellite images downloaded.

Note that for best results, chlorophyll particles were discarded on the border. Thus, particles that are modeled strictly falls within the domain of the lake (see Figure 3(a)).

## 6. Monte Carlo method

To describe the pollutants transport over Lake Valencia, consider  $N$  particles to which are associated  $N$  paths of the diffusion processes that satisfy the stochastic differential equations (4) - (5) and (9) - (12). In general, these stochastic differential equations have no analytical solution since the existence thereof, depends on its coefficients, therefore, the numerical approximation of the solution is used. Hereinafter, the numerical approximation scheme employed for each stochastic model and the contaminant distribution on the lake surface are described for each time instant  $t$ .

### 6.1. Numerical approximation of particle transport

To perform the numerical approximation of pollutant transport over Lake Valencia, two types of initial particles distribution are considered: an initial distribution concentrated at a fixed point and an initial distribution generated from a satellite image.

#### 6.1.1. Initial distribution concentrated at a fixed point

In this part, we consider an initial distribution with density  $p(x, y) = \delta(x - x_0)\delta(y - y_0)$ . The implementation consists of: fix a point  $(x_0, y_0)$  on the surface of the lake were  $N$  particles are injected, whose displacements from this point will be described by Euler schemes corresponding to the random walk and flight models.

The Euler schemes corresponding to the equations (4) - (5) are:

$$\hat{X}_{t_{k+1}} = \hat{X}_{t_k} + \left[ u + \frac{\partial H}{\partial x} \frac{D}{H} + \frac{\partial D}{\partial x} \right] (\hat{X}_{t_k}, \hat{Y}_{t_k}) \Delta t + \sqrt{2D(\hat{X}_{t_k}, \hat{Y}_{t_k})} \varepsilon_{t_{k+1}}^1, \quad (22)$$

$$\hat{Y}_{t_{k+1}} = \hat{Y}_{t_k} + \left[ v + \frac{\partial H}{\partial y} \frac{D}{H} + \frac{\partial D}{\partial y} \right] (\hat{X}_{t_k}, \hat{Y}_{t_k}) \Delta t + \sqrt{2D(\hat{X}_{t_k}, \hat{Y}_{t_k})} \varepsilon_{t_{k+1}}^2, \quad (23)$$

In both equalities,  $\hat{X}_{t_k}$  and  $\hat{Y}_{t_k}$  denote the numerical approximation of  $X_t$  and  $Y_t$ , respectively, where  $\hat{X}_{t_0} = x_0$ ,  $\hat{Y}_{t_0} = y_0$  and  $t_k = k\Delta t$  for integers  $k \geq 0$  and  $\Delta t$  representing the time step. In addition,  $u$  and  $v$  are the approximations of the mean flow velocities of the lake in  $x$  and  $y$  directions, respectively,  $H$  and  $D$  are the approximation of its depth and dispersion coefficient, respectively. The above approximations are available in a rectangular grid adjusted to the exact dimensions of the lake surface  $\mathcal{D}$ .

The increments of Brownian motion  $\varepsilon_{t_{k+1}}^i = W_{t_{k+1}}^i - W_{t_k}^i$ ,  $i = 1, 2$ , necessary for the construction of the numerical scheme, are constructed by an array of  $N$  vectors (one per particle), whose entries are independent random numbers with the normal distribution of parameters  $(0, \Delta t)$ .

In the developed numerical scheme, it is always necessary to determine the relative position of the particle in the  $x$  and  $y$  direction respect to the cell that is located this particle. This calculation is necessary to perform two-dimensional linear interpolation of  $u$ ,  $v$  and  $H$ , in the point from which will determine the following coordinates of the discretization of the particle trajectory. With this information the code determines the new position of the particle by the approaches defined in equations (22) and (23).

However, it is important to check if the position obtained is outside or within the domain where is going to calculate the trajectory because if the particle falls outside the domain it should be immediately reflected back into the flow domain. To do this, a subroutine that executes a border control was built based in the reflection procedure corresponding to the symmetrized Euler scheme developed in [Bossy et al. \(2004\)](#), but adjusted to the characteristics of the boundary of the lake surface  $\partial\mathcal{D}$ .

Specifically, we start with  $(\hat{X}_{t_0}, \hat{Y}_{t_0}) = (x_0, y_0)$  and assume that we have obtained  $(\hat{X}_{t_k}, \hat{Y}_{t_k}) \in \mathcal{D}$ ; afterward, we compute the next possible position  $(Xp, Yp)$  of the particle, executing the above Euler scheme for  $t_{k+1}$  as follows

$$Xp = \hat{X}_{t_k} + \left[ u + \frac{\partial H D}{\partial x H} + \frac{\partial D}{\partial x} \right] (\hat{X}_{t_k}, \hat{Y}_{t_k}) \Delta t + \sqrt{2D(\hat{X}_{t_k}, \hat{Y}_{t_k})} \varepsilon_{t_{k+1}}^1, \quad (24)$$

$$Yp = \hat{Y}_{t_k} + \left[ v + \frac{\partial H D}{\partial y H} + \frac{\partial D}{\partial y} \right] (\hat{X}_{t_k}, \hat{Y}_{t_k}) \Delta t + \sqrt{2D(\hat{X}_{t_k}, \hat{Y}_{t_k})} \varepsilon_{t_{k+1}}^2. \quad (25)$$

Then, if  $(Xp, Yp) \in \mathcal{D}$  we define  $(\hat{X}_{t_{k+1}}, \hat{Y}_{t_{k+1}}) = (Xp, Yp)$  and follow the calculation of the next point of the discretization of the particle trajectory, but;

- If  $(Xp, Yp) \in \mathcal{D}^c \cap V_{\partial\mathcal{D}}(r)$  with  $V_{\partial\mathcal{D}}(r) = \{z \in \mathbf{R}^2 : d(z, \partial\mathcal{D}) \leq r\}$  for any  $r > 0$ , defined from the characteristics of  $\partial\mathcal{D}$  and the size of the grid with respect to which the discretization of the particle trajectory is being computed, then

$$\hat{X}_{t_{k+1}} = Xp - 2F^\gamma(Xp, Yp) \gamma_1(Xp, Yp), \quad (26)$$

$$\hat{Y}_{t_{k+1}} = Yp - 2F^\gamma(Xp, Yp) \gamma_2(Xp, Yp), \quad (27)$$

where  $\Gamma = (\gamma_1(z), \gamma_2(z))$  denotes the unit inward normal vector to  $\partial\mathcal{D}$  corresponding to the normal projection of  $z$  onto  $\partial\mathcal{D}$  for  $z \in \mathcal{D}^c \cap V_{\partial\mathcal{D}}(r)$ , and  $F^\gamma(z)$  denotes the algebraic distance of  $z$  to  $\partial\mathcal{D}$  parallel to  $\gamma$ .

- If  $(Xp, Yp) \in \mathcal{D}^c \cap V_{\partial\mathcal{D}}^c(r)$  the simulation of  $(Xp, Yp)$  is restarted.

In accord with [Bossy et al. \(2004\)](#), for domains with boundary bounded and class  $\mathbf{C}^5$  there exists  $r > 0$  such that, for any point outside the domain the normal projection from the point at the border is unique. The boundary of Lake Valencia is bounded but is not class  $\mathbf{C}^5$ , for this reason, there exist points  $z \in V_{\partial\mathcal{D}}(r)$  to which the normal projection from the point to  $\partial\mathcal{D}$  is not unique. For  $z \in V_{\partial\mathcal{D}}(r)$  with this condition, we select the normal reflection corresponding to the smaller of the distances of  $z$  to  $\partial\mathcal{D}$ .

The procedure is similar, in terms of numerical approximation of the diffusion corresponding to stochastic differential equations that define the random flight pattern. To estimate  $(X_t, Y_t)$ , the Euler scheme is implemented as follows:

$$\hat{X}_{t_{k+1}} = \hat{X}_{t_k} + \left[ u + \frac{\partial H D}{\partial x H} + \frac{\partial D}{\partial x} \right] (\hat{X}_{t_k}, \hat{Y}_{t_k}) \Delta t + \hat{U}_{t_k} \Delta t, \quad (28)$$

$$\hat{Y}_{t_{k+1}} = \hat{Y}_{t_k} + \left[ v + \frac{\partial H D}{\partial y H} + \frac{\partial D}{\partial y} \right] (\hat{X}_{t_k}, \hat{Y}_{t_k}) \Delta t + \hat{V}_{t_k} \Delta t, \quad (29)$$

where  $u$ ,  $v$ ,  $H$  and  $D$  are as before and, in view that (11) and (12) are analytically solved for  $(U_t)_{t \geq t_0}$  and  $(V_t)_{t \geq 0}$ , defined by:

$$U_t = U_0 e^{-\frac{1}{T_L} t} + \hat{K}^{1/2} \int_0^t e^{-\frac{1}{T_L}(t-s)} dW_s^1, \quad (30)$$

$$V_t = V_0 e^{-\frac{1}{T_L} t} + \hat{K}^{1/2} \int_0^t e^{-\frac{1}{T_L}(t-s)} dW_s^2, \quad (31)$$

we obtain the estimations  $\hat{U}_{t_k}$  and  $\hat{V}_{t_k}$  of  $U_{t_k}$  and  $V_{t_k}$ , respectively, using the Monte Carlo integration method.

The initial random velocities  $U_{t_0}$  and  $V_{t_0}$  are random numbers with normal distribution with parameters  $(0, \frac{\hat{K} T_L}{2})$ .

For random flight scheme, the procedure is analogous to the one described above for random walk model.

### 6.1.2. Initial distribution generated from a satellite image

As was explained before, we are interested in simulating the chlorophyll dispersion in Lake Valencia during a specific time interval  $[t_0, t_k]$ , using stochastic models. To do this, we need pairs of random numbers from a probability distribution that models the chlorophyll distribution in the lake at  $t_0$ , to use them as initial condition in the Euler schemes implementation of the stochastic models. These schemes are defined by the equations (22)-(23) and (28)-(31), respectively. The result of this implementation after  $k$  time steps, will be the estimations of the chlorophyll dispersion in the lake at  $t_k = k\Delta t$ .

To construct the above random sample, we consider three steps.

First, we download satellite images of Lake Valencia surface  $\mathcal{D}$  and use the values between 0 and 1 of their vegetation index ( $NDVI$ ), to determine the subset  $\mathcal{C}$  of  $\mathcal{D} \setminus \partial \mathcal{D}$  characterized by the presence of chlorophyll because these  $NDVI$  values indicate the presence of chlorophyll in the positions where they were computed. Actually, the information of the  $NDVI$  array only allows us to identify the nodes  $(i, j)$  of the mesh corresponding to the satellite images such that  $NDVI(i, j) \in (0, 1)$ , then it assumes that  $\mathcal{C}$  is formed by cells whose nodes that identifies them, are located in the lower left corner of each grid square where  $NDVI$  is between 0 and 1. Based in the last assumptions,  $\mathcal{C}$  is a collection of connected subsets of  $\mathcal{D}$  (see Figure 3(a)).

Second, we define the function  $\pi : \mathcal{C} \rightarrow \mathbf{R}$  by

$$\pi(i, j) = \frac{NDVI(i, j)}{\sum_A NDVI(i, j)}, \quad (32)$$

with  $A = \{(i, j) : NDVI(i, j) \in (0, 1)\}$ , for the nodes  $(i, j)$  of the mesh corresponding to the satellite images that are contained in  $\mathcal{C}$ , and by bilinear interpolation for the rest of the points in  $\mathcal{C}$ . We interpret  $\pi$  as the density of the probability distribution that measures the proportions of chlorophyll in the lake, for the instant that the satellite images were captured.

Third, we solve the two-dimensional Langevin stochastic differential equation (SDE) numerically:

$$d\mathbf{L}_t = d\mathbf{W}_t + \frac{1}{2} \nabla \log(\pi(\mathbf{L}_t)) dt, \quad t \geq t_0, \quad (33)$$

where  $\mathbf{W}_t$  is a standard two-dimensional Brownian motion, and select, of this discretized trajectory, the pairs of random numbers with density  $\pi$  that will be assigned to  $(x_0, y_0)$  in the implementation of the stochastic models. We approximate this SDE using the Metropolis- adjusted version of its Euler scheme, with symmetric and normal reflection on the border  $\partial\mathcal{C}$  (see [Bossy et al., 2004](#) and [Roberts and Tweedie, 1996](#)).

Specifically, for all connected subset of  $\mathcal{C}$ , is chosen a point from which the approximation of a path of the diffusion process  $\mathbf{L}_t$  that solves the equation in (33) is generated; then a collection of trajectory points are selected, spaced by a fixed number of time steps to ensure their approximate independence. All the points so built, form the random sample of points in  $\mathbf{R}^2$  with density  $\pi$  that we need to use as initial condition in the implementation of the stochastic models. The points of the random sample constructed have density  $\pi$  because this density is invariant for the diffusion  $\mathbf{L}_t$  (see Appendix for the theoretical justification).

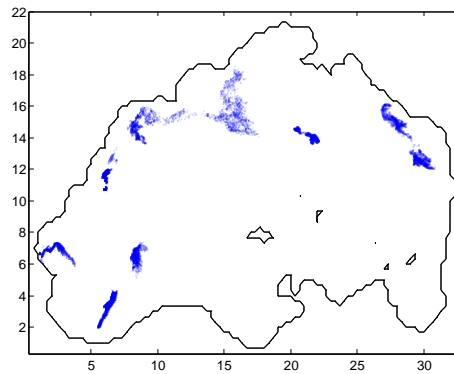


FIGURE 4. *Random pairs of the initial distribution generated from satellite image.*

Figure 4 shows a random sample over Lake Valencia surface constructed by the procedure described before, from images of Lake Valencia captures by the satellite *Landsat 8* particularly at 17 : 15 : 59, April 07, 2014.

## 6.2. Numerical approximation of the particles distribution in an instant $t$

In accordance with the previously developed, the position of  $N$  particles injected into the flow at time  $t = j\Delta$  where  $j$  is non-negative integer, is approximated by the random sample  $(X_j^1, Y_j^1), \dots, (X_j^N, Y_j^N)$ , obtained from the implementation of Euler scheme (22) - (23) in the random walk model, or Euler scheme (28) - (29) corresponding to the random flight model.

From this sample, the probability density  $p(x, y, t)$ , corresponding to the particles dispersion at time  $t$  is estimated using a Gaussian kernel estimator defined as follows:

$$\hat{p}_N(x, y, j\Delta) = \frac{1}{h^2 N} \sum_{i=1}^N K\left(\frac{x - X_{j\Delta}^i}{h}, \frac{y - Y_{j\Delta}^i}{h}\right), \quad (34)$$

where  $K$  is a two-dimensional Gaussian kernel and  $h = h(N) \rightarrow 0$  when  $N \rightarrow \infty$ . We use the optimal window that is  $h = N^{-1/3}$  in dimension two, one can also consider the use of the automatically chosen window, but this may slow the running of the algorithms. This estimator (34) is asymptotically unbiased and for sufficiently large samples is distributed as a normal distribution. The reader can consult the monograph (Silverman, 1986) for more about Kernel density estimation.

The estimation of this density allows us to determine an approximation of the distribution of particles on the fluid at any time  $t$ .

## 7. Results

In this section, we show the empirical behavior of the stochastic models that we have introduced. In order to present the results of the implementation, this section is divided in three parts. First, an analysis of both models, random and flight, are presented. Second, a presentation by the data used for implementation is exposed. Third and final, a comparison between these numerical schemes is done through Lake Valencia implementation with two types of initial distributions: initial distribution concentrated at a fixed point and initial distribution generated from a satellite image. These distributions are described in section 6.

### 7.1. Analysis of the models

As it was presented in section 2 and 3, random walk model, simulates the behavior of the particles in a domain considered only for long periods of simulation. This is because the random particles displacements are simulated by Brownian motions that have independent increments. However, the random flight model corrects this shortcoming through replacement of the dispersion terms in (4) and (5) by Ornstein-Uhlenbeck processes that describe the behavior of the particles velocities at time  $t$  less than  $T_L$ .

Is important to recall that  $T_L$  can be interpreted as a measure of how long the particle takes to lose memory of its initial turbulent velocity. Also, it is relevant to analyze the simulation behavior respect different values of  $T_L$ . The random flight model is consistent with the advection-diffusion equations (1) for times  $t$  such that  $t \gg T_L$ , since for  $\frac{t-t_0}{T_L} \rightarrow \infty$  or  $T_L \rightarrow 0$ , the stochastic differential equations that define the random flight model, (9)-(12), are reduced to the Itô equations established for the random walk model. As discussed in sections 2 and 3 both models are consistent with advection-diffusion equation (1).

### 7.2. Data used in the implementation

In testing is considered static circulation obtained with wind speed and direction is 3 m/s from south-south-west to the north-north-east. The current velocity over the lake used for the calculation corresponds to that obtained by the hydrodynamic model presented in section 4 (Valera-López et al., 2014, 2016).

To make the simulations we consider that the dispersion coefficient is a constant and then we use the following value for this coefficient in Lake Valencia:  $D = 5 \times 10^{-5} \text{ km}^2/\text{h}$ . This value has the same order of magnitude of one experimental value found for the dispersion coefficient in

Lake Huron that is  $4 \times 10^{-2} m^2/s$  (Fisher et al., 1979, chapter 6). In effect  $D = 5 \times 10^{-5} km^2/h = 1.39 \times 10^{-2} m^2/s$ . This data is used because there is no experimental data of this kind over Lake Valencia and was not found on similar lakes. The references consulted only contain experimental values about ocean, rivers, large lakes (such as the case of Lake Huron in USA), and reservoirs. Several tests before to choose the value of  $D$  were performed. The value  $D = 5 \times 10^{-5} km^2/h$  gave a better adjustment regarding the expected diffusive behavior of the particles. All that in function of the characteristics of the mean flow velocity and the vortices in Lake Valencia that are observed in Figure 1.

Bathymetry (function  $H$ ) was obtained from existing geographical information in a report by the Ministry of Environment of Venezuela (Ministerio del ambiente, 1995) and it was then digitized using an automatic software.

Along the simulations we use the time's step  $\Delta t = 0.3573h$ .

To implement the numerical schemes of the stochastic models using an initial distribution generated from satellite images, we download these images from satellite *Landsat 8* for free through the website of the US Geological Survey - EarthExplorer. This image has 12-bit radiometric resolution (240 m spatial resolution), eight bands of 30 meters and the panchromatic band of 15 meters. The downloaded images were captured by the sensor Enhanced Thematic Mapper (ETM), which has six bands in the electromagnetic spectrum from the visible to middle infrared.

### 7.3. Models implementation

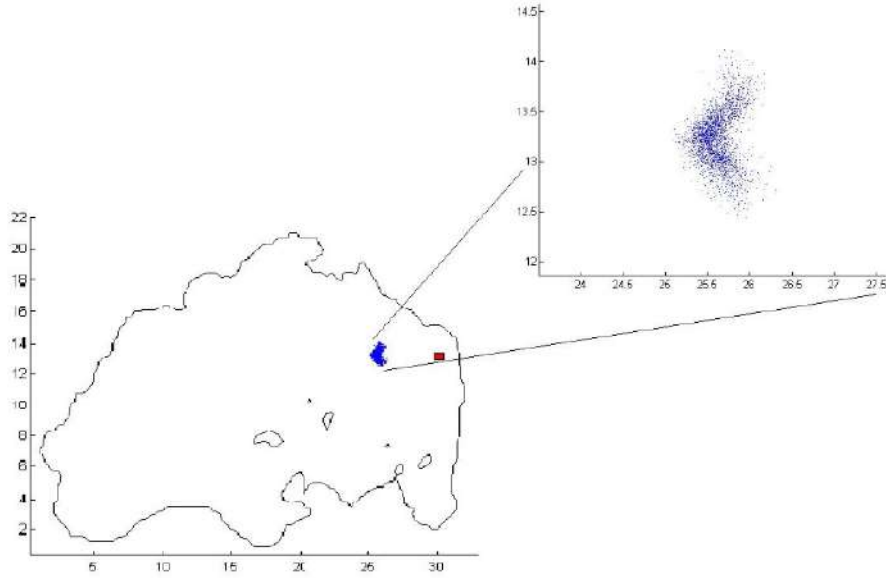
Hereinafter, the empirical behavior of both stochastic models through implementation on Lake Valencia are introduced. Here, we present two examples. Each of them corresponds to a type of an initial condition: one concentrated at a fixed point and other generated from a satellite image.

#### 7.3.1. Example 1: Initial distribution concentrated at a fixed point

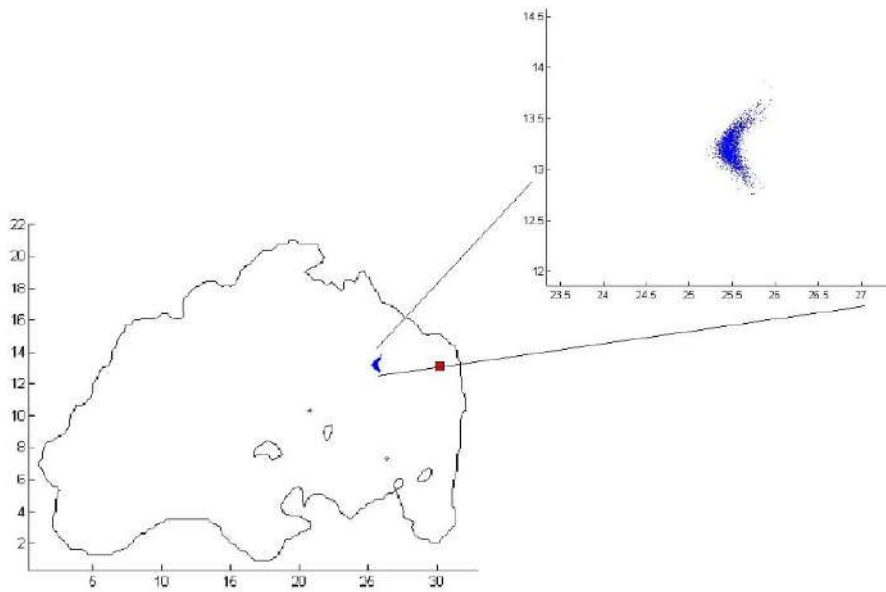
In what follows we present the obtained results when the particles depart at a fixed point. Thus, some graphical representations of the particles behavior before and after of certain Lagrangian time  $T_L$  are shown. The dispersion coefficient is as we have said  $D = 5 \times 10^{-5} km^2/h$ .

Figures 5 and 6 show the results obtained by injecting 2500 particles at point (30.2, 13.1). For this purpose, the position of each particle after 20000 time's steps corresponding to their trajectories is determined. To implement the random flight model is considered  $T_L = 500h$ . The value chosen for  $T_L$  is approximately  $\frac{T_f}{14}$ , where  $T_f = 7146h$  represents the total propagation time of the particles on the lake. Figures 5(a) and 5(b) show the position of the particles at time  $t = 357.3h$  lower than  $T_L = 500h$ . In both figures, a less dispersive behavior of random flight model versus random walk model is noticeable. However, in Figures 6(a) and 6(b), we see that in  $t = 2858.4h$ , which is much greater than  $T_L = 500h$ , the behavior of both models is similar, showing the trend towards behavior described theoretically. Under conditions as above, but with  $T_L = 20h$ , Figures 8(a) and 8(b) show that with the same value of  $t$ , the similarity between models becomes much more evident. This is because the autocorrelations of the particle stochastic speed, at the random flight model equation (13) tend faster to zero when  $T_L = 20h$  compared to  $T_L = 500h$ , and this fact ensures the fastest convergence of random flight to random walk, (see section 3).





(a) Random walk.



(b) Random flight with  $T_L = 500h$ .

FIGURE 5. Particles position at  $t = 357.3h$  generated from the implementation with  $D = 5 \times 10^{-5} km^2/h$ , after being injected into the coordinate  $(30.2, 13.1)$  represented in the figure by a square.

Figures 7 and 9 show the kernel density corresponding to time  $t$  defined above for  $T_L = 500h$  and  $T_L = 20h$  of random walk and flight models, respectively. These figures also show that random flight model approximates better random walk model, when the first is executed with  $T_L = 20h$ .

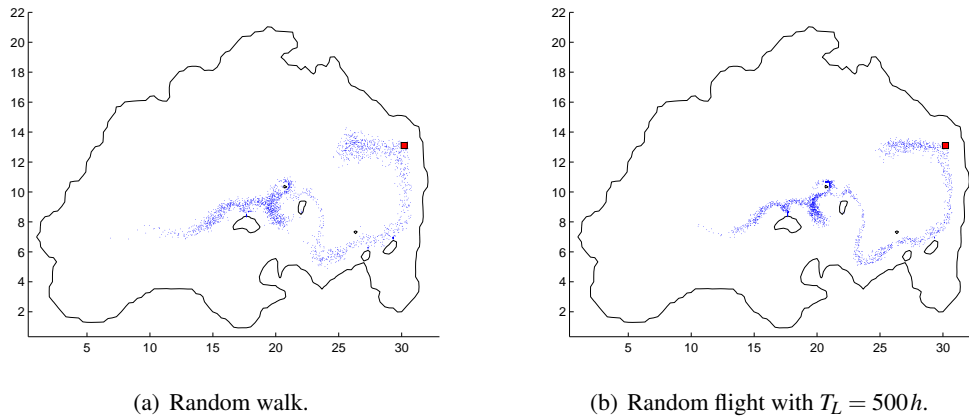


FIGURE 6. Particles position at  $t = 2858.4h$  generated from the implementation with  $D = 5 \times 10^{-5} km^2/h$ , after being injected into the coordinate  $(30.2, 13.1)$  represented in the figure by a square.

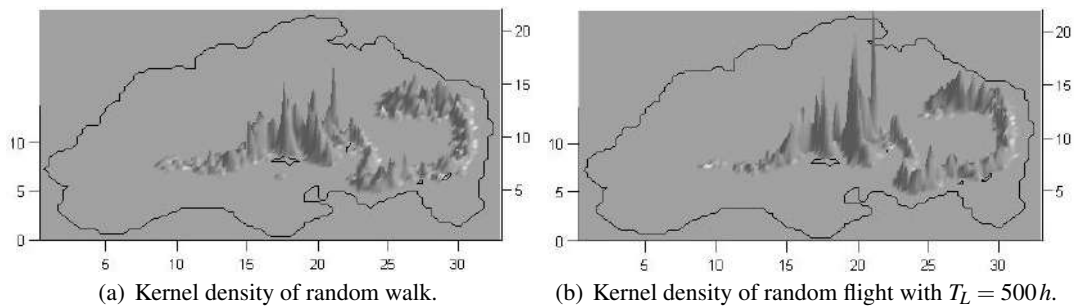


FIGURE 7. Kernel density of particles distribution at  $t = 2858.4h$  generated from the implementation with  $D = 5 \times 10^{-5} km^2/h$ , after being injected into the coordinate  $(30.2, 13.1)$ .

### 7.3.2. Example 2: Initial distribution generated from a satellite image

This example is devoted to describe the estimations given by the stochastic models of the real chlorophyll dispersion in Lake Valencia during  $1200.2h$ , this is approximately fifty days, from  $t_0 = 17 : 15 : 59$ , April 7, 2014.

To make these estimations, we follow the general procedure described in section 6.1.2.

As a first step, we download images of the lake surface captured by the satellite *Landsat 8* at time  $t_0$  and, using the information of the bands 3 and 4 of these images, we calculate the vegetation index *NDVI*, of the lake.

Second, with this *NDVI* we identify the areas with presence of chlorophyll in Lake Valencia (see Figure 3(b)), and define upon them, a probability density  $\pi$  that approximates the chlorophyll

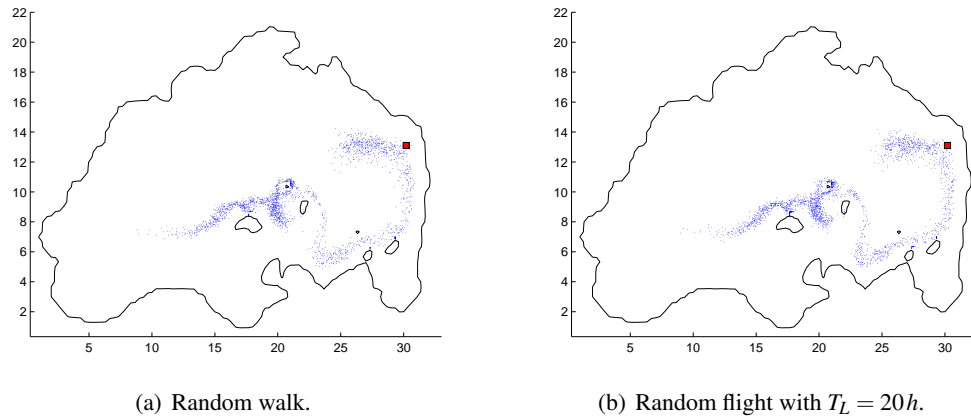


FIGURE 8. Particles position at  $t = 2858.4h$  generated from the implementation with  $D = 5 \times 10^{-5} km^2/h$ , after being injected into the coordinate  $(30.2, 13.1)$  represented in the figure by a square.

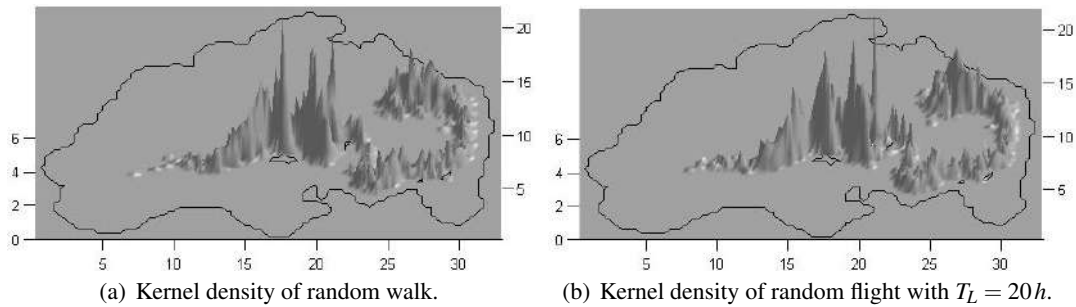


FIGURE 9. Kernel density of particles distribution at  $t = 2858.4h$  generated from the implementation with  $D = 5 \times 10^{-5} km^2/h$ , after being injected into the coordinate  $(30.2, 13.1)$ .

density and, as consequence, defines a probability distribution that simulates the chlorophyll distribution at  $t_0$ .

Third, for all connected spot of chlorophyll shown in Figure 3(a), we select a point from which the approximation of a path of the Langevin process  $L_t$ , defined by the equation (33) from the density  $\pi$ , is generated; then a collection of 15000 trajectory points are selected, spaced by 100 time steps to ensure their approximate independence. All the points so built, form the random sample of points in  $\mathbf{R}^2$  with density  $\pi$  that we need to implement the stochastic models.

Finally, we transform the coordinates of each element of the random sample constructed before to their corresponding coordinates in the mesh where the static circulation on the lake was approximated, and use, this new coordinates, as initial condition of the Euler schemes corresponding to the random walk and flight models. After 3359 time steps for numerical schemes, the simulation of the chlorophyll dispersion is obtained for 1200.2h.

Figure 4 shows a random sample constructed by the procedure described before, from images of Lake Valencia captured on the date and time stated above. Figures 10(a) and 10(b) show the positions of the particles 1200.2h after, determined by the random walk and flight models,

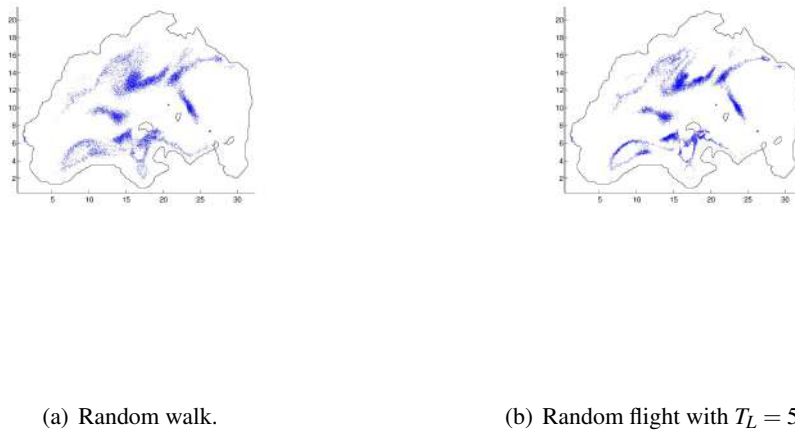


FIGURE 10. *Particles position at  $t = 1200.2h$  generated from the implementation with  $D = 5 \times 10^{-5} \text{km}^2/h$ , with initial distribution generated from satellite image.*

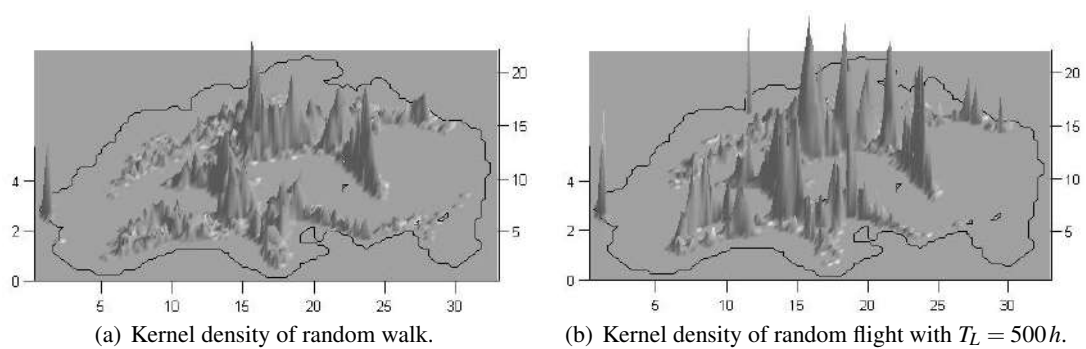


FIGURE 11. *Kernel density of particles distribution at  $t = 1200.2h$  generated from the implementation with  $D = 5 \times 10^{-5} \text{km}^2/h$ , with initial distribution generated from satellite image.*

respectively. The implementation of random flight model was made by using  $T_L = 500h$ . In the last two figures cited, we observe that, as in the case of the initial distribution concentrated at a fixed point, when  $D = 5 \times 10^{-5} km^2/h$  the approximations given by the models are similar, despite the fact that, the random flight model is less dispersive than random walk model. Figures 11(a) and 11(b) show that the kernel densities corresponding to the stochastic models also exhibit the behavior described above.

## 8. Conclusions

In this work is carried out for the first time, a study on the transport of pollutants in Lake Valencia (Venezuela) through the implementation of two stochastic models: random walk and flight.

This study was started by us with the determination of circulation flow on the surface of the lake (Valera-López et al., 2014, 2016), and continued with the model that describes the movement of pollutants, the main product of this research.

To achieve the goal of describing the distribution patterns of sediment in Lake Valencia by these two models, the velocities of static circulation were obtained from a hydrodynamic model, which is simulated by using Saint-Venant equations. Interpreting the advection-diffusion equation as a Fokker-Planck equation, it is possible to describe the particle trajectory of contaminant with both stochastic models. The random flight model is introduced to improve the behavior of random walk model, shortly after deployment of the particles. In the result obtained by the random flight model can be observed least dispersion of particles in comparison with the result of random walk before the Lagrangian time, demonstrating consistency with the behavior described in sections 2 and 3. In addition, after a long time to run and considering a smaller Lagrangian time, one observes greater similarity between the two models. Ensuring, faster convergence of the random flight model towards the random walk.

An important novelty of this research is that we build an algorithm that randomly generates the position of the particles from the contaminant using satellite image information. This simulates contaminant dispersion from a true initial distribution into the lake since the time when the images were captured by the satellite. The importance of this method of random generation is that the initial positions of the particles are modeled by random pairs of a probability density that accurately approximates the actual density of organic waste into the lake. In addition, this method ensures that the simulation of contaminant transport is not only determined by the initial position of the particles but also by the density of pollution in the area determined by these positions. The initial condition constructed by the random sample, allows models to better approximate the real pollutants transport on the lake. In implementing this initial condition in the stochastic schemes, we observed that the behavior of the results obtained when considering the punctual concentration is maintained, i.e. random flight is still less dispersive than random walk. Although using the shallow water model (from the Saint-Venant equation), the random walk model and random flight model are standard in the study of the circulation of lakes and the distribution of pollutants by sediments, its application is new to Lake Valencia, as well as the use of stochastic and statistical methods using satellite images to model the contaminant distribution (actually the chlorophyll-a dispersion).

We think that this study may lead to an automatic procedure for the determination of areas with the highest pollution and therefore ensure monitoring of areas closed or coastal water. This can be

done considering, on these models, dynamic speeds (which vary according to changes in wind speed or direction) or obtaining satellite images capture time nearest each other.

Even though our application has been made in the Lake of Valencia, developed method can be applied to any surface water closed.

## Appendix

### *Simulation of a density defined on a set $\Omega \subset \mathbf{R}^d$ with boundary in $C^2$*

The text that follows is an adaptation of the methods presented in [Fattler and Grothaus \(2007\)](#), to the conditions that satisfies our problem. We include it for sake of completeness and for giving to the non specialist reader some elementary insight over a so abstract matter. Thus, let us consider a relatively compact connected set  $\Omega \subset \mathbf{R}^d$ . Let us also assume that its boundary  $\partial\Omega$  is a rectifiable set and that its parametrization has two continuous derivatives and also the set has Lebesgue measure zero. Let  $\varphi^2$  be a density on  $\bar{\Omega}$  that is differentiable satisfying  $\varepsilon \leq \varphi^2(x) \leq \|\varphi^2\|_\infty$  in  $\bar{\Omega}$  (it can be assumed that this function only vanishes in the boundary but, in such a case, our diffusion can not start in a point of this set). It can be defined the symmetric Dirichlet form (for an introduction to such a notion the reader can consult the reference [Fukushima et al., 2010](#)), that is the closure in  $\mathbf{L}^2(\bar{\Omega}, \mu)$ , where  $d\mu(x) = \varphi^2(x)dx$ , of the following symmetric bilinear form

$$\mathcal{E}(f, g) = \frac{1}{2} \int_{\Omega} \langle \nabla f(x), \nabla g(x) \rangle \varphi^2(x) dx, \quad (35)$$

that is defined for  $f, g \in \mathcal{W}$  with  $\mathcal{W} = \{f \in C(\bar{\Omega}) \mid f \in W_{loc}^{1,1}(\Omega), \mathcal{E}(f, g) < \infty\}$ , where  $W_{loc}^{1,1}(\Omega)$  denotes the Sobolev space of functions having a locally integrable weakly derivative on  $\Omega$ . Given that the function  $\varphi^2$  does not vanish in no point of  $\bar{\Omega}$  then the bilinear form (35) with domain  $\mathcal{D}$  is closable in  $\mathbf{L}^2(\bar{\Omega}, \mu)$  and its closure  $(\mathcal{E}^{\varphi^2}, D(\mathcal{E}^{\varphi^2}))$  is a Dirichlet form conservative, regular and symmetric. Moreover there exists a self adjoint operator  $\mathcal{L}^{\varphi^2}$  with dense domain  $D(\mathcal{L}^{\varphi^2}) \subset \mathbf{L}^2(\bar{\Omega}, \mu)$  that generates the Dirichlet form  $(\mathcal{E}^{\varphi^2}, D(\mathcal{E}^{\varphi^2}))$  acting in the following form:

$$\mathcal{L}^{\varphi^2} f = \frac{1}{2} \Delta f + \langle \nabla(\log \varphi), \nabla f \rangle, \quad (36)$$

for functions belonging to  $\{f \in W^{2,\infty}(\Omega) \mid \partial_\nu f(x) = 0 \text{ for all } x \in \partial\Omega\} \subset D(\mathcal{E}^{\varphi^2})$ , where  $\partial_\nu$  denotes the derivative with respect to the outward normal of  $\partial\Omega$ . Also,  $W^{2,\infty}(\Omega)$  denotes the Sobolev space of functions two times weakly differentiable and essentially bounded in  $\Omega$ .

In fact the Dirichlet forms defined in (35) have different closures, depending on the core of the initial forms. Two of such closures are important, the minimal associated to Dirichlet boundary condition and the maximal linked to the Neumann boundary condition. We only refer here to the second one.

Now it can be defined the isometry  $\mathbf{L}^2(\bar{\Omega}, \mu) \rightarrow \mathbf{L}^2(\bar{\Omega}, dx)$  by using the properties of  $\varphi$ , such that for  $f \in \mathbf{L}^2(\bar{\Omega}, \mu) \rightarrow g = \varphi f \in \mathbf{L}^2(\bar{\Omega}, dx)$ . In this manner, denoting  $\mathcal{T}_t^{\varphi^2}$  the semigroup of self adjoint and bounded operators generated by  $-\mathcal{L}^{\varphi^2}$ , one can show, using the isometry, that the transformed semigroup denoted by  $T_t$  acting on  $\mathbf{L}^2(\bar{\Omega}, dx)$  satisfies  $T_t g = \varphi \mathcal{T}_t^{\varphi^2}(\varphi^{-1}g)$ , for all

$g \in \mathbf{L}^2(\overline{\Omega}, dx)$ . Furthermore, if  $\mathbf{1}$  denotes the constant function equal to 1, it holds that  $\mathcal{T}_t^{\varphi^2} \mathbf{1} = \mathbf{1}$ . By using the relationship between the two semigroups, it has  $T_t \varphi = \varphi$ . This implies that function  $\varphi$  is an eigenfunction for the infinitesimal generator  $A$  of  $T_t$  associated to the 0 eigenvalue. Besides, it can be proved that

$$Ag = \mathcal{L}^{\varphi^2}(\varphi^{-1}g) = \frac{1}{2}\Delta(\varphi^{-1}g) + \langle \nabla \log \varphi, \nabla(\varphi^{-1}g) \rangle .$$

The domain of  $A$  results the set  $D(A) = \varphi D(\mathcal{L}^{\varphi^2})$ .

Then for all continuous and bounded  $h$  defined on  $\overline{\Omega}$  and defining  $d\mu(x) = \varphi^2(x)dx$  it holds

$$\langle \mathcal{T}_t^{\varphi^2}(\mu), h \rangle = \langle \mu, \mathcal{T}_t^{\varphi^2}(h) \rangle = \int_{\Omega} \mathcal{T}_t^{\varphi^2}(h)(x)\varphi^2(x)dx.$$

Moreover, one needs to consider the boundary conditions. Hence, by using the Green formula one has

$$\begin{aligned} \int_{\Omega} (g\mathcal{L}^{\varphi^2}f - f\mathcal{L}^{\varphi^2}g)\varphi^2 dx &= - \int_{\partial\Omega} (\varphi^2(x)g\partial_\nu f - f\partial_\nu(\varphi^2(x)g) + \partial_\nu(\log \varphi^2)\varphi^2 fg)dx \\ &= - \int_{\partial\Omega} (g\partial_\nu f - f\partial_\nu g)\varphi^2 dx = 0. \end{aligned}$$

The last equality is a consequence of the boundary condition. This property is transferred onto the semigroup yielding

$$\begin{aligned} \langle \mathcal{T}_t^{\varphi^2}(\mu), h \rangle &= \int_{\Omega} \varphi(x)\mathcal{T}_t^{\varphi^2}(\varphi^{-1}g)(x)\varphi(x)dx \\ &= \int_{\Omega} T_t(g)\varphi(x)dx = \int_{\Omega} g(x)T_t(\varphi)(x)dx \\ &= \int_{\Omega} g(x)\varphi(x)dx = \int_{\Omega} h(x)\varphi^2(x)dx = \langle \mu, h \rangle . \end{aligned}$$

Getting  $\mathcal{T}_t^{\varphi^2}(\mu) = \mu$ , assuring that  $\mu$  is the invariant measure. Further the operator  $\mathcal{L}^{\varphi^2}$  is self-adjoint and, under the hypothesis of our problem, it has a discrete spectra.

Let  $\alpha > 0$  and let define the resolvent

$$(\alpha - \mathcal{L}^{\varphi^2})^{-1}f(x) = \int_0^\infty e^{-\alpha t} \mathcal{T}_t^{\varphi^2}f(x)dt.$$

Let us quote the following theorem shown in [Reed and Simon \(1978\)](#) Theorem XIII. 64. Only the assertions that are interesting for this work will be written.

**Theorem.** Let  $A$  be a self-adjoint operator that is bounded from below. The following are equivalent:

- i)  $(A - \alpha)^{-1}$  is compact for at least a

$$\alpha \in \rho(A) = \{ \alpha \in \mathbf{C} : (A - \alpha)^{-1} \text{ is bounded and injective} \}.$$

ii) There exists an complete orthonormal system  $\{\varphi_i\}_{i=0}^\infty$  in  $D(A)$  such that  $A\varphi_n = \lambda_n\varphi_n$  with  $\lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_n \dots$  and  $\lambda_n \rightarrow \infty$ .

To apply this result to our problem we only need to prove that  $(\alpha - \mathcal{L}^{\varphi^2})^{-1}$  is a compact operator defined  $\mathbf{L}^2(\overline{\Omega}, \mu) \rightarrow \mathbf{L}^2(\overline{\Omega}, \mu)$ . But we know by the general theory of semigroups that  $(\alpha - \mathcal{L}^{\varphi^2})^{-1}(\mathbf{L}^2(\overline{\Omega}, \mu)) = D(\mathcal{L}^{\varphi^2})$ . Let  $u_n$  be a sequence of bounded vectors in  $\mathbf{L}^2(\overline{\Omega}, \mu)$ . Let define  $g_n = (\alpha - \mathcal{L}^{\varphi^2})^{-1}u_n$ , then denoting as  $\|\cdot\|_{op}$  the operator norm it holds

$$\|g_n\| \leq \|(\alpha - \mathcal{L}^{\varphi^2})^{-1}\|_{op}\|u_n\| < K,$$

and moreover,  $\mathcal{L}^{\varphi^2}g_n = \alpha g_n - u_n$ , then

$$\|\mathcal{L}^{\varphi^2}g_n\| \leq \alpha\|g_n\| + \|u_n\| \leq (\|(\alpha - \mathcal{L}^{\varphi^2})^{-1}\|_{op} + 1)\|u_n\| < 2K.$$

Let us introduce also the norm generated by the operator  $\mathcal{L}^{\varphi^2}$  as

$$\|f\|_{\mathcal{L}^{\varphi^2}} = \|f\|_{\mathbf{L}^2(\overline{\Omega}, \mu)} + \|\mathcal{L}^{\varphi^2}f\|_{\mathbf{L}^2(\overline{\Omega}, \mu)} \quad \forall f \in D(\mathcal{L}^{\varphi^2}).$$

Given that the function  $\varphi^2$  is bounded by below and above, this norm is equivalent to the norm of the Sobolev space  $W^{1,2}(\overline{\Omega})$  with respect to the Lebesgue measure:

$$\|f\|_{W^{1,2}} = \|f\|_{\mathbf{L}^2(\overline{\Omega}, dx)} + \left( \int_{\overline{\Omega}} \|\nabla f\|^2 dx \right)^{1/2}.$$

This entails that  $\|g_n\|_{W^{1,2}} \leq C$  for some constant  $C$ . The compact embedding  $W^{1,2} \rightarrow \mathbf{L}^2(\overline{\Omega}, dx)$  implies that there exists a subsequence  $g_{n_k}$  that converges in  $\mathbf{L}^2(\overline{\Omega}, dx)$  and also in  $\mathbf{L}^2(\overline{\Omega}, \mu)$ , so the compactness of the resolvent hold.

Let us consider  $\{\lambda_i\}_{i=0}^\infty$  the eigenvalues for  $\mathcal{L}^{\varphi^2}$ , with  $\lambda_0 = 0$  and eigenfunctions  $\{\varphi_i\}_{i=0}^\infty$ . In this manner all function  $f \in \mathbf{L}^2(\overline{\Omega}, \mu)$  can written as

$$f = \sum_{i=0}^{\infty} \langle f, \varphi_i \rangle_{\mu} \varphi_i.$$

Recall that  $\varphi_0 = \mathbf{1}$ , in this form  $c_0 := \langle f, \mathbf{1} \rangle_{\mu} = \int_{\Omega} f(x)\varphi^2(x)dx$ . One has by using the spectral theorem,

$$\mathcal{T}_t^{\varphi^2}(f)(x) = \sum_{i=0}^{\infty} e^{-\lambda_i t} \langle f, \varphi_i \rangle_{\mu} \varphi_i.$$

This equality implies the following inequality

$$\left\| \mathcal{T}_t^{\varphi^2}(f) - c_0 \right\|_{\mathbf{L}^2(\mu)} \leq e^{-2\lambda_1 t} \|f\|_{\mathbf{L}^2(\mu)}. \quad (37)$$

For justifying this inequality, first we know that zero is an eigenvalue and has multiplicity one because the invariant measure is unique. Moreover, this implies that  $\lambda_1 > 0$ .

Also, the theory of Dirichlet forms allows obtaining a Markov process  $\{X_t\}_{t \geq 0}$  taking values on  $\overline{\Omega}$  such that it satisfies the SDE:



$$d\bar{X}(t) = d\bar{B}(t) + 2\nabla(\log \varphi)(\bar{X}(t))dt. \tag{38}$$

with reflecting boundary condition and  $X_0 = x \in \bar{\Omega}$ .

The expression “with reflecting condition in the boundary” must be understood according Lions and Sznitman (1984) as the SDE whose solution is the process  $\bar{X}_t$  and has the form

$$d\bar{X}(t) = d\bar{B}(t) + 2\nabla(\log \varphi)(\bar{X}(t))dt + \nu(\bar{X}(t))d\bar{k}(t), \tag{39}$$

where  $\bar{k}(t)$  is the local time that increasing only whenever  $\bar{X}_t$  is in  $\partial\Omega$ . In fact, the definition of this last process is

$$\bar{k}(t) = \int_0^t 1_{\partial\Omega}(\bar{X}(s))d|\bar{k}|(s).$$

The transition density of such a process starting at  $y \in \bar{\Omega}$  satisfies the following Fokker-Planck equation

$$\begin{aligned} \partial_t p(t, x) &= \mathcal{L}^{\varphi^2} p(t, x), \\ \partial_\nu p(t, x) &= 0, \end{aligned}$$

for all  $x \in \partial\Omega$  and  $p(0, x) = \delta_y(x)$  for  $y \in \bar{\Omega}$ . The result obtained in (37) implies that the process  $\bar{X}$  is ergodic and  $\beta$ -mixing (see Doukhan, 1994 for the definition and properties) with exponential rate and, furthermore,

$$\frac{1}{t} \int_0^t f(\bar{X}(u))du \xrightarrow{c.s.} \mathbf{E}[f(\bar{X}_0)] = \int_{\Omega} f(x)\varphi^2(x)dx.$$

### Simulation Algorithm

We can try now to implement a simulation process that permits the generation of a sample  $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_N$  with density  $\varphi^2$ . We consider first the case where  $\varphi^2$  is a density in  $\mathbf{R}^d$ . In Roberts and Tweedie (1996) it was established that the diffusion solution of the following SDE

$$d\bar{X}(t) = d\bar{B}(t) + \nabla \log \varphi(\bar{X}(t))dt \tag{40}$$

has as invariant measure  $d\mu(x) = \varphi^2(x)dx$ . In this form to generate an approximated sample it is enough to approximate the solution of the SDE with some numerical scheme. If one uses the Euler’s scheme the equation (40) can be approximated by

$$\bar{X}_{(k+1)h} = \bar{X}_{kh} + h(\nabla(\log(\varphi(\bar{X}_{kh})))) + \sqrt{h}\varepsilon_k.$$

The constant  $h$  is the discretization step and the  $\varepsilon_k$  are a sequence of independent standard Gaussians r.v. If one can ensure that for a sufficiently small  $h$  the Markov chain  $\bar{X}_{(k+1)h}$  has an invariant measure  $\mu_h$  and moreover  $\mu_h$  and  $\mu$  are near in a certain metrics, then we only need to run the scheme for generating the sample. However, Roberts and Tweedie in Roberts and Tweedie (1996) pointed out that such a procedure does not give always good results, proposing then a

modification that we sketch in that follows. In the Euler's scheme given the value  $\bar{U}_k = \bar{X}_{kh}$ , we must generate a Gaussian random variable  $\bar{U}_{k+1} \stackrel{d}{=} N(\bar{U}_k + h\nabla \log \varphi(\bar{U}_k), hI_d)$ . This fact implies that the transition density of the chain is

$$q(u_1, u_0) = \frac{1}{(2\pi h)^{\frac{d}{2}}} e^{-\frac{1}{2h} \|u_1 - (u_0 + h\nabla \log \varphi(u_0))\|^2}.$$

As is remarked in the article, citing Julien Besag, this chain could be non ergodic and also the convergence towards the invariant density could be pretty slow. As a consequence of this last situation the algorithm must be modified by using a procedure of acceptance or reject the same type of the one of the Hasting-Metropolis algorithm. This procedure is used for building a chain  $\bar{M}_k$ . In first place given the value of  $\bar{M}_k$ , a variable  $\bar{U}_{k+1}$  is generated with Gaussian distribution  $N(\bar{M}_k + h\nabla \log \varphi(\bar{M}_k), hI_d)$ . The value  $\bar{U}_{k+1}$  is accepted with probability equal to

$$\alpha(\bar{M}_k, \bar{U}_{k+1}) = 1 \wedge \frac{\varphi^2(\bar{U}_{k+1})q(\bar{U}_{k+1}, \bar{M}_k)}{\varphi^2(\bar{M}_k)q(\bar{M}_k, \bar{U}_{k+1})},$$

where  $q$  is the transition density of the discrete chain. If  $\bar{U}_{k+1}$  is accepted we set  $\bar{M}_{k+1} = \bar{U}_{k+1}$ , otherwise we set  $\bar{M}_{k+1} = \bar{M}_k$ . The properties of the Hasting-Metropolis algorithm imply that this algorithm converges towards the invariant measure.

Turning to the problem of simulate a density defined on  $\Omega$ . We approximate the solution of the SDE (39) by using the Euler scheme of Bossy et al. (2004). However, one must admit that this discrete solution approaches the invariant measure  $\varphi^2(x)dx$ , when the step of discretization tends towards zero. Therefore, two questions remain open and require further study.

- For simulation of a density defined on  $\Omega$  the Euler scheme of Bossy et al. (2004), can be modified to approach the invariant measure?
- Can we work with a simulation scheme similar to the one of Roberts and Tweedie (1996) to have a convergent algorithm?

## References

- Arcos, M. P., Ávila, S. L., Estipinan, S. M., and Gómez, A. (2005). Indicadores microbiológicos de contaminación de las fuentes de agua. *Nova-Publicación Científica*, 3:69–79.
- Arnold, L. (1974). *Stochastic differential equations: Theory and applications*. Wiley, 1 edition.
- Awad, M. (2014). Sea water chlorophyll-a estimation using hyperspectral images and supervised artificial neural network. *Elsevier Ecol. Inf.*, 24:60–68.
- Bossy, M., Gobet, E., and Talay, D. (2004). A symmetrized Euler scheme for an efficient approximation of reflected diffusions. *J. Appl. Prob.*, 41:877–889.
- Courant, R., Friedrichs, K., and Lewy, H. (1967). On the partial difference equations of mathematical physics. *IBM journal of Research and Development*, 11(2):215–234.
- De Baas, A. F., Van Dop, H., and Nieuwstadt, F. T. M. (1980). An application of the Langevin equation for inhomogeneous conditions to dispersion in a convective boundary layer. *Quart Journ. Roy. Met. Soc.*, 12:165–180.
- Doukhan, P. (1994). Mixing: properties and examples. *Lecture Notes in Statistics* 85.
- Fan, D., Huang, Y., Song, L., Liu, D., Zhang, G., and Zhang, B. (2014). Prediction of chlorophyll-a concentration using HJ-1 satellite imagery for Xiangxi Bay in Three Gorges Reservoir. *Water Science and Engineering*, 7(1):70–80.
- Fattler, T. and Grothaus, M. (2007). Strong feller properties for distorted Brownian motion with reflecting boundary condition an application to continuous  $n$ -particle systems with singular interactions. *Journal of Functional Analysis*, 246:217–241.

- Fei, W., Xuan, W., Ying, Z., and Zhifeng, Y. (2012). Long-term changes of water level associated with chlorophyll-a concentration in Lake Baiyangdian, North China. *Procedia Environmental Sciences*, 13:1227–1237.
- Fisher, H. B., List, E. J., Koh, R. C. Y., Imberger, J., and Brooks, N. H. (1979). *Mixing in inland and coastal waters*. Academic Press.
- Fukushima, M., Oshima, Y., and Takeda, M. (2010). Dirichlet forms and symmetric Markov processes. *De Gruyter studies in Mathematics*, 19.
- García, R. and Kahawita, R. A. (1986). Numerical solution of the St. Venant equations with the MacCormack finite difference scheme. *International Journal for Numerical Methods in Fluids*, 6:259–274.
- Griffa, A. (1996). Applications of stochastic particle models. In *Stochastic Modelling in Physical Oceanography*, pages 113–140.
- Heemink, A. W. (1990). Stochastic modelling of dispersion in shallow water. *Stochastic Hydrology and Hydraulics*, 4:161–174.
- Karatzas, I. and Shreve, S. E. (1991). *Brownian motion and stochastic calculus*. Springer GTM, 113.
- Lanza, G. D. L., Hernández, S., and Carbajal, J. L. (2000). Organismos indicadores de la calidad de agua y de la contaminación (bioindicadores). *Universidad Nacional Autónoma de México*.
- Li, L., Binghu, Z., and Lusan, L. (2010). Biomonitoring and bioindicators used for river ecosystems: definitions, approaches and trends. *Procedia Environmental Sciences*, 2:1510–1524.
- Lions, P. and Sznitman, A. (1984). Stochastic differential equations with reflecting boundary conditions. *Comm. Pure and Applied Math.*, XXXVII:511–537.
- MacCormack, R. W. (1971). Numerical solution of the interaction of a shock wave with a laminar boundary layer. *Lectures Notes in Physics, Springer-Verlag*, 8:151–163.
- Ministerio del ambiente (1995). Estudio del lago de Valencia : Parámetros fisicoquímicos y biológicos (1971-1995). *Caracas, Venezuela : Informe del 2do Seminario Técnico del Programa de Saneamiento Ambiental Integral de la Cuenca del Lago de Valencia, Caracas*.
- Nelson, E. (1967). *Dynamical theories of Brownian motion*, volume 3. Princeton university press Princeton.
- Ostapczuk, P., Burow, M., May, K., Mohl, C., Froning, M., Sübembach, B., Widmann, E., and Emons, H. (1997). Mussels and algae as bioindicators for long-term tendencies of element pollution in marine ecosystems. *Chemosphere*, 37:2049–2058.
- Peña, E., Palacio, M., and Ospina-Alvarez, N. (2005). Algas como indicadores de contaminación. *Universidad de Valle-Programa Editorial*.
- Reed, M. and Simon, B. (1978). *Analysis of operators*. Academic Press, IV.
- Roberts, G. O. and Tweedie, R. (1996). Exponential convergence of Langevin distribution and their discrete approximations. *Bernoulli*, 2(4):341–363.
- Silverman, B. (1986). *Density estimation for statistics and data analysis*. Chapman & Hall.
- Soize, C. (1994). *The Fokker-Planck equation for stochastic dynamical systems and its explicit steady state solutions*, volume 17. World Scientific.
- Tsanis, I. K. and Saied, U. (2007). A wind-driven hydrodynamic and pollutant transport model. *Global NEST Journal*, 9(2):117–131.
- Valera-López, M., J. Guevara-Jordan, R. G., Saavedra, I., and León, J. R. (2014). Understanding circulation in Lake Valencia, Venezuela by a shallow-water model. In *Ingeniería y Ciencias Aplicadas: Modelos Matemáticos y Computacionales*, pages MM37–MM42.
- Valera-López, M., J. Guevara-Jordan, R. G., Saavedra, I., and León, J. R. (2016). Modeling wind-driven circulation and pollution in lake Valencia. *Journal of Applied Fluid Mechanics*, 9(6).
- Van Dop, H., Nieuwstadt, F. T. M., and Hunt, J. C. R. (1985). Random walk models for particle displacements in inhomogeneous unsteady turbulent flows. *Physics of Fluids*, 28:1639–1653.
- Zhuowei, H., Hongqi, L., and Z. Lin, L. F. (2010). Quantitative inversion model of water chlorophyll-a based on spectral analysis. *Procedia Environmental Sciences*, 10:523–528.