Determination of the concentration Chlorophyll-a in the Valencia Lake (Venezuela)

José R. León R.
Universidad Central de Venezuela.

Introducción

The Valencia Lake is the second lake of Venezuela and for some time it has experienced a degradation of its waters. It is an inland lake with no outlet to the sea, having a surface of $3.140 \mathrm{km}^2$. It is localized between $67^o07'$ and $68^o12'$ west longitude and $09^o57'$ and $10^o26'$ north latitude. In the next figure a map of Venezuela is shown, signaling the placement of the Valencia lake.

Figure: Venezuela map with the lake

The Venezuela's environment ministry (Ministerio del Ambiente y de los Recursos Renovables) has been monitoring the quality of the lake's waters since 1997.

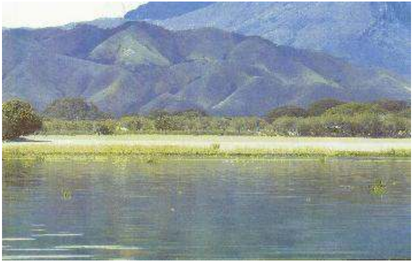


Figure: A shore of Lake

The concentration of several substances is studied in 13 stations. One of these substances is the concentration of Chlorophyll-a. This substance measures the quality of the inland waters. Chlorophyll-a is a green pigment found in plants. It absorbs sunlight and converts it to sugar during photosynthesis. Important concentrations of this substance is a sign of eutrophication. High levels often indicate poor water quality and low levels often suggest good conditions.

Below we present an image of the lake in Google map. In the south-west we can see a zone of green color where has been found, in certain times, important Chlorophyll-a concentration.

Figure: Valencia Lake in Google map

The pertinence of Chlorophyll-a as an indicator of the assessment of water quality is indicated below by means of a table, that classifies the waters by the levels of Chlorophyll-a.

Table 1. Eutrophication scale for chlorophyll *a* concentrations

Chloroplyll *a* (mg m^{-3})	Eutrophication Level
0.000 – 0.084	Oligotrophic
0.084 – 0.359	Lower-mesotrophic
0.359 – 0.793	Upper-mesotrophic
> 0.793	Eutrophic

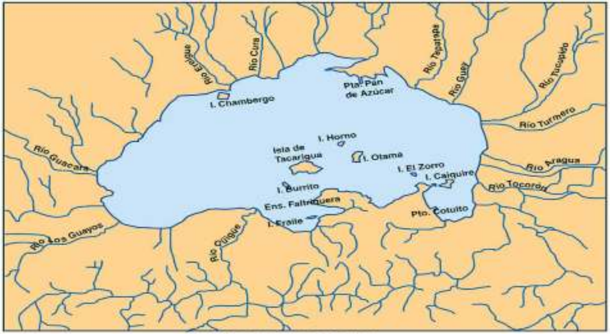
Table: Water's classification by the concentration of Chlorophyll-a

In the next table the mesures of Chlorophyll-a in the stations of the Valencia Lake are shown, these measurements were taken in 2004. Under these results the lake's waters, having a mean of 48.83, must be classified as upper-mesotrophic. Nevertheless, there are some stations: 0, 34, 16, whose levels indicating eutrophic waters.

station	Chlorofilla Mg/m^2
12A	20
0	71
39	43
40	50
3	32
5	41
6	17
24	39
33	67
34	76
16	93
17	37

Below we illustrate with an image the location of these stations.

We display in the next figure the basin of the lake and rivers that comprise it.



CUENCA DEL LAGO DE VALENCIA

There are two big cities nearby the lake. Maracay in the east side is a town with 800.000 habitants and in the other side we have Valencia with more than one million and half of habitants.
Recently there has been an interest in assessing the concentration of Chlorophyll-a via satellite images. This is done by establishing some regression models between the reflectance measured in the image and the Chlorophyll-a concentrations that have been taken in situ. In that follows we will establish a satellite monitoring system that uses these regression models and Kriging's method to extend the obtained concentration of Chlorophyll-a, for all the lake's surface.

Kriging and Chlorophyll-a concentrations

To extrapolate the values of a random quantity taken in some points of a region in $D \subset \mathbb{R}^2$, Kringing is a good method. Let $Z(x)$ be a random field showing spatial dependence. We suppose that the two first moments of Z exist. Let us define $\mu(x) = \mathbb{E}Z(x)$ the mean function and
$C(x,h) = \mathbb{E}\big((Z(x+h) - \mu(x+h))(Z(x) - \mu(x))\big)$ the covariance.
If these two quantities do not depend of the site x we say that the field is weakly stationary, and we assume that this is our case.

Thus we can extrapolate the values of several observations $\{Z(x_1), \ldots, Z(x_N)\}$ to another site $Z(x_0)$ as a linear combination of the known values i.e.

$$Z(x_0) = \sum_{i=1}^{N} \lambda_i Z(x_i) \text{ with } \sum_{i=1}^{N} \lambda_i = 1.$$

The weights λ_i are chosen such that they made minimum the mean quadratic error.
$MQE = \mathbb{E}(Z(x_0) - \sum_{i=1}^{N} \lambda_i Z(x_i))^2$.
By denoting $\overrightarrow{\lambda}$ and $\overrightarrow{w}$ the vectors whose coordinates are λ_i and $R(x_0 - x_i)$ respectively, and also the covariance matrix $C = (c_{ij}) = (R(x_i - x_j))$.

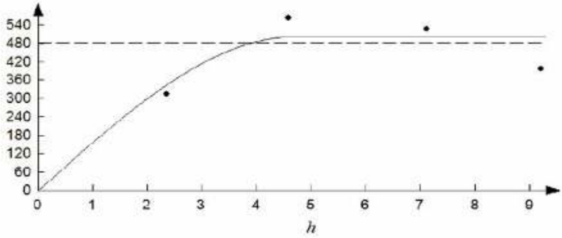
We get $\overrightarrow{\lambda} = C^{-1}\overrightarrow{w}$. The procedure consist in estimating the mean μ and the covariance matrix C. A procedure that works well consists in consider parametric models of the covariance, then the chosen model is that approximating best has to the estimated covariance.

In the case where the mean is not known one can consider the use of the Variogram. This function is defined as the increment's variance $\mathbb{E}(Z(x+h)-Z(x))^2$
and the semivariogram will be $\gamma(h) = \frac{\mathbb{E}(Z(x+h)-Z(x))^2}{2}$.

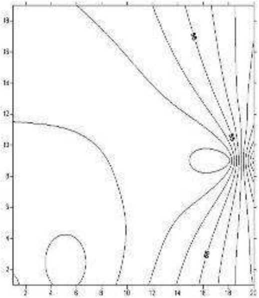
The estimation of this quantity does not require estimation of the mean. We shall estimate it by $2\hat{\gamma}(h) = \frac{1}{|N(h)|}\sum_{(i,j)\in N(h)}(Z(x_j+h) - Z(x_i+h))$ where $N(h)$ is the set of elements in the sample separated by h and $|N(h)|$ is its cardinal. We have applied this method to some Chlorophyll-a data taken in 2004. Among the parametric models the best fit was for the spherical model.

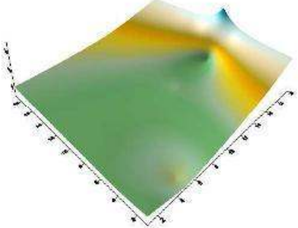
$$\gamma(h) = \begin{cases} C\left(\frac{3}{2}\frac{|h|}{a} - \frac{1}{2}\frac{|h|^3}{a^3}\right) & si \;\; |h| \le a \\ C & si \;\; |h| > a \end{cases}$$

Below we find a graph indicating the fit. It is important to point out that we do not have enough points in the sample. However, we have constructed the surface of the extrapolated values of $Z(x)$ for the entire lake surface.



In the figure below we show the result of the Kriging extrapolation. For the data of Chlorophyll-a in the table.





Here we have one extrapolation in a another season (the format is different sorry) red color implies hight concentration and blue very little.

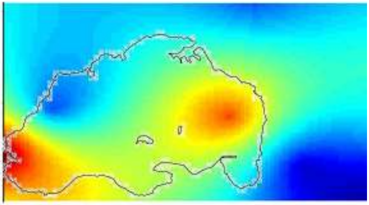


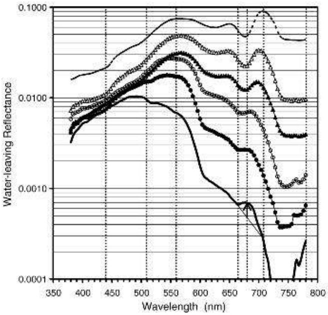
Figure: Surface of Chlorophyll-a obtained by Kriging year 2004

In the first of the last two images, we can see more Chlorophyll-a concentration in the northwest zone in the proximity of Maracay and their rivers, the pollution discharged for these rivers is huge. In the second one the concentration is more important nearby of Valencia. We can see this type of oscillation during all the year. With little information is dangerous to do statistics. That is why we decided to supplement the information with satellite imagery.

Reflectance vs Chlorophyll-a correlation

In my country Venezuela it is difficult to have enough experimental data in any area, in particular for environmental studies. This is why we look for another approach for assessment the water quality in the Valencia Lake. Our choice was optical remote sensing.
It is well known that Chlorophyll-a exhibits a unique spectral absorption signature with marked peaks in the blue and red wavebands. However, detect Chlorophyll in inland waters by means of reflectance in the red and blue wavebands is not successful this was well established in reference [2]. Some authors have proposed using the Medium Resolution Imaging Spectrometer (MERIS) in bands centred at 665, 681 and 709 nm.

The next figure taken of [2] indicates the spectra of water-leaving reflectance selected for incremental Chlorophyll-a concentrations (solid line without symbols: Chla=0.49 mg m^{-3}; solid circles: Chla=1.1 mg m^{-3}; open circles: Chla=4.2 mg m^{-3}; solid triangles: Chla=17 mg m^{-3}; open triangles: Chla=31 mg m^{-3}; broken line without symbols: Chla=131 mg m^{-3}). The vertical dotted lines indicate positions of MERIS wavebands 2, 4, 5, 7, 8, 9 and 12 centred at approximately 443, 510, 560, 665, 681, 709, and 779 nm, respectively. The arrow pointing at the spectrum for a concentration of 0.49 mg m^{-3}, indicates the fluorescence line height above the reflectance baseline drawn between the values at 665 nm and 709 nm.



The red-NIR (near infrared) reflectance peak has been shown to be a suitable indicator of Chlorophyll-a concentration when the analysis is restricted to eutrophic water bodies. The MERIS sensor features 3 bands in the region of the peak in the precedent figure and other bands in NIR that may also be used for Chlorophyll-a retrieval. In the literature it was adapted a semi-analytic regression algorithm to reflectance at the MERIS bands 7, 9 and 12:

$$[Chla] = \frac{\{R_M(0.70 + b_b) - 0.40 - b_b^{1.06}\}}{0.0016}, \tag{1}$$

where [Chla] = Chlorophyll-a concentration (mg m^{-3}),
$R_M = Rw(709)/Rw(665)$ MERIS band 9 to band 7 ratio of water-leaving reflectance,

and $b_b =$ is the apparent backscatter coefficient (m^{-1}) obtained from the MERIS band 12 (centred at 779 nm) according to:

$$b_b = \frac{1.61 R_w(779)}{0.082 - R_w(779)}.$$

Equation (1) adequately predicts [Chla] (residuals $<$ 35% of measured values) at all mesotrophic to highly eutrophic waters. Predicted values were inaccurate for [Chla]$<$ $5mg\ m^{-3}$, and even negative for the oligotrophic waters.
The foregoing discussion was intended to establish the close relationship between chlorophyll and reflectance. I think it has become clear.

Chlorophyll-a detection by using Modis-Terra satellite

We have decided to implement our study by using the Modis-Terra satellite, that has integrated the Modis (Moderate Resolution Imaging Spectroradiometer). We will use the reflectance values obtained in the web-site of the satellite.

The Terra satellite orbits the Earth passing over the poles in an orbit that allows the sensor to get daily images of most of the surface of our planet. The polar orbiting keeps Terra satellite constantly aligned with the sun, so every day the satellite passes over every place around the same local time, about 10:30 am.

When MODIS data are downloaded they are stored in a compressed format with extension HDF. For each HDF file the following data were taken.

- 500m 16 days red reflectance MODIS Band #1, 620-670nm
- 500m 16 days NIR reflectance MODIS Band #2, 841-876nm
- 500m 16 days blue reflectance MODIS Band #3, 459-479nm
- 4500m 16 days MIR reflectance MODIS Band #7, 2105-2155nm

These data allow us to built a table where the evolution of reflectance was stored for each year.

We have in the next image a time series showing this evolution.

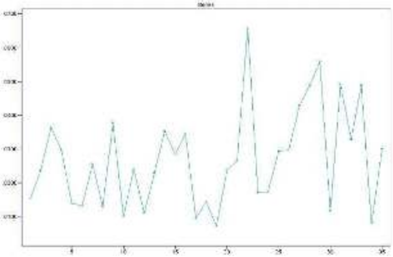


Figure: Reflectance in station 16 by month. Years 2001,2002,2003.

These time series were studied by using R time series software.

We display below an image containing the tendency of the above time series. The reflectance as function of the time grows. Accepting as true the regression model, we can conclude that in this individual place the Valencia Lake's waters are deteriorating their quality.

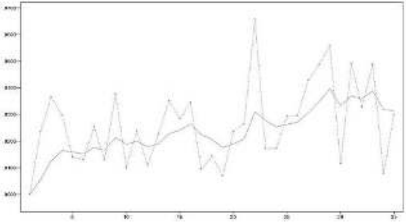


Figure: Reflectance and tendency in station 16 by month.

Another important feature is the existence of a remarkable seasonal component. During the dry season (November-May) the values of reflectance are bigger than during the rainy season. This behavior is maintained in all the years of the study.

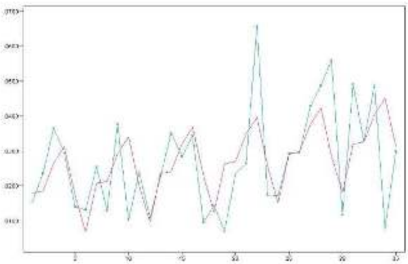


Figure: Reflectance and seasonal component for station 16 by month.

The study conducted in all the places where there are stations, gave different or sometimes contradictory results. So in this moment we are studying more satellite data combined with in situ data expecting to get more definitive conclusions.

Kriging and Reflectance

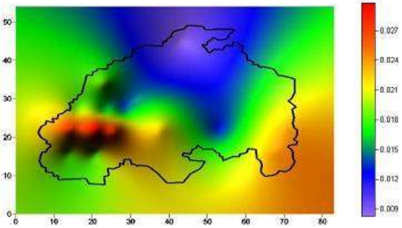
With the data of reflectance of satellite Meris, in the stations, and by using the regression model and Kriging we extrapolated, to the lake´s surface the predicted values of Chlorophyll-a. The procedure was

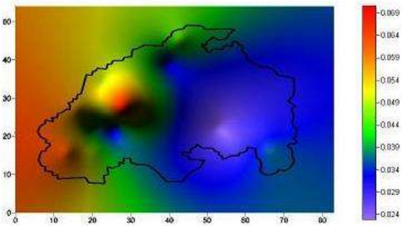
1. Compute for each month the actual values of the reflectance in the stations. Then compute the Chlorophyll-a concentration by using the regression model, built in each station.
2. Determine by using the empiric variogram the better fit model.
3. Extrapolate the values to the lake's surface.

Results

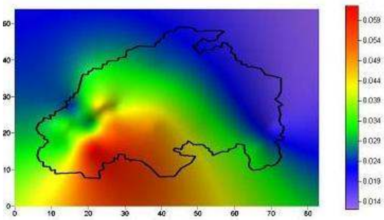
- The spherical variogam also gives the better fit here.
- We compute the Kringing extrapolation surface in both in the dry season as in the rainy season.
- We simulate the 2004 data comparing these results with the true values of reflectance.

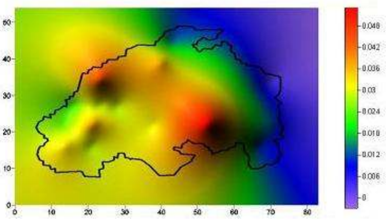
The next figure display two of the extrapolations in the dry season.





In the graph below we show the kriging result for two different months of the rainy season.





Models of Transport

In the previous sections we have studied some statistical tools to determinate the proportion of Chlorophyll-a in the Valencia Lake. In that follows we shall study the dynamic models of transport of Chlorophyll-a. Two approaches will be considered, the fist one will be the Diffusion model and the second one the Particles method. The goal is to combine the predicted concentrations obtained by using remote sensing with the solutions of the stochastic models representing the two methods of transport.

Diffusion model

This is the traditional method for computing the evolution of the particles in liquid advent by a velocity field in the surface. The deterministic equation of each particles is

$$\frac{dX(t)}{dt} = -\nabla V(t, X(t)) \text{ with } X(0) = X_0, \text{ and } X(t) \in \mathbb{R}^2, \nabla \cdot V = 0.$$

If we consider a diffusion effect we must introduce a standard Brownian motion $W = (W_1, W_2)$ in $\mathbb{R}^2$ and a diffusivity matrix

$$D(\mathbf{x}) = \begin{pmatrix} a_{11}(\mathbf{x}) & a_{22}(\mathbf{x}) \\ a_{12}(\mathbf{x}) & a_{22}(\mathbf{x}) \end{pmatrix} \text{ and } \Sigma^2 = DD^t(\mathbf{x}) = \begin{pmatrix} \sigma_{11}(\mathbf{x}) & \sigma_{22}(\mathbf{x}) \\ \sigma_{12}(\mathbf{x}) & \sigma_{22}(\mathbf{x}) \end{pmatrix}$$

The movement equation for the particle perturbed by the noise dW writes

$$dX_t = \Sigma(X_t)dW(t) - \nabla V(t, X(t))dt \quad X(0) = X_0.$$

If we consider $\int_B p(t, \mathbf{x})d\mathbf{x}$ the concentration of Chlorophyll-a in the region B of particles at time t, and $p(t, \cdot)$ is a density of probability. This last function satisfies the Fokker-Planck partial differential equation.

$$\partial_t p = \frac{1}{2}\partial_x(\sigma_{11}\partial_x p + \sigma_{12}\partial_y p) + \frac{1}{2}\partial_y(\sigma_{22}\partial_y p + \sigma_{12}\partial_x p) - V(\mathbf{x})\nabla p,$$

$$p(0, \cdot) = p_0(\cdot)$$

The existence and uniqueness of the solution of such an equation is guaranteed by standard hypothesis on Σ^2 and V.

Advection velocity

What is important to solve this equation is to know the last two functions.

We study first the advection velocity $\mathbf{u} = (u, v, w)$ in Fluid Mechanics notation (in fact we take $\nabla V = (u, v, 0)$).

The advection velocity $\mathbf{u}$ satisfies the following system of PDE:

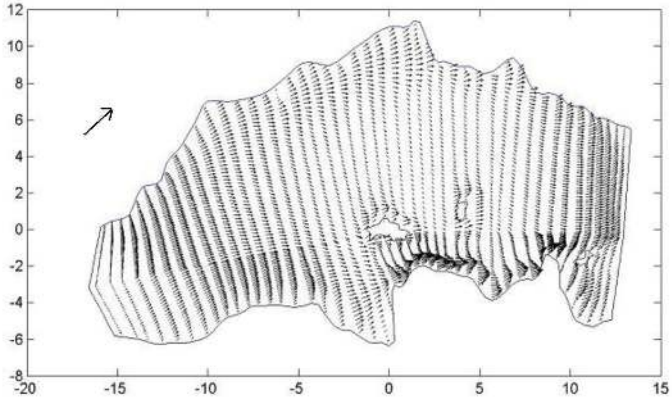
$$\partial_t \mathbf{u} + \mathbf{u}.\nabla \mathbf{u} = -\nabla P + \frac{(v\mathbf{e}_1 - u\mathbf{e}_2)}{R_o} - \frac{\rho \mathbf{e}_3}{F^2} + \frac{1}{R_e}\nabla^2 \mathbf{u}$$

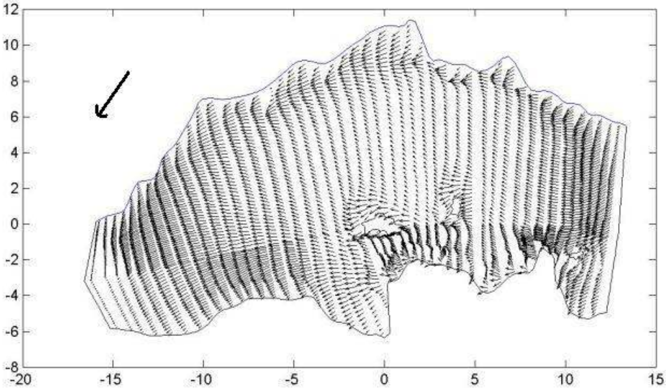
$$\partial_t \rho + \rho.\nabla \rho = 0$$

$$\nabla^2 P = -\frac{1}{F^2}\nabla.\rho\mathbf{e}_3 - \mathbf{u}.[(\mathbf{u}.\nabla)\mathbf{u}] + \frac{1}{R_e}\nabla^2 d - \partial_t d + \frac{1}{R_o}\nabla.(v\mathbf{e}_1 - u\mathbf{e}_2) \quad (*)$$

Where P is the pressure, ρ is the density, $d = \nabla.\mathbf{u}$ denotes the divergence of the velocity, $\mathbf{e}_i$ are the vectors of the canonical basis in $\mathbb{R}^2$. The dimensionless numbers are the Reynolds number, $R_e = UL/v$, the Froude number, $F = U/NL$ and the Rossby number $Ro = U/fL$. $N^2 = \frac{-g}{\rho_0}\partial_z\rho$ is the Brunt-Vaisala frequency. Equation (*) substitutes the incompressibility condition, $\nabla.\mathbf{u} = 0$.

Our group have worked jointly with a group of researchers from the Universidad of Carabobo (at Venezuela). This people have elaborated a code in Fortran to solve the precedent PDE system, in the real volume of the lake, they research has produced the article [1], where our description was taken. They used a system of curvilinear coordinates, meteorological conditions and a refined bathymetry. Their results are displayed in the two next figures. These images show the advection velocity in the lake's surface.


12
10
8
6
4
2
0
-2
-4
-6
-8
-20
-15
-10
-5
0
5
10
15


12
10
8
6
4
2
0
-2
-4
-6
-8
-20
-15
-10
-5
0
5
10
15

The first one with the wind directing towards the Northeast and the second one towards the Southwest. The violation of the condition $\nabla.\mathbf{u} = 0$ made more involve the model of SDE, but in sake of simplicity we do not consider here this additional difficulty.

Determination of Σ^2 matrix

The coefficient of diffusivity is a measure of the speed at which a chemical species spreads in a medium, this is expressed with a positive number, its unit of measure is $(length)^2/(time)^{-1}$, i.e., cm^2/sec^{-1} and can be classified into two types, molecular and turbulent. For the diffusion of Chlorophyll-a (or phytoplankton) in a lake only the turbulent diffusivity is important.

The Eddy diffusion coefficient, in the x direction (resp. in the y direction), E_x (resp. E_y) depends on the scale of the phenomenon, it can be defined as

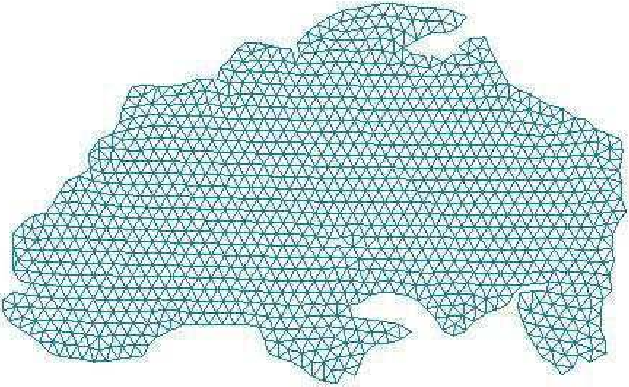
$$E_x := ku\frac{d\omega^2}{dx} = \sigma_{11} \text{ and } E_y := kv\frac{d\omega^2}{dy} = \sigma_{22},$$

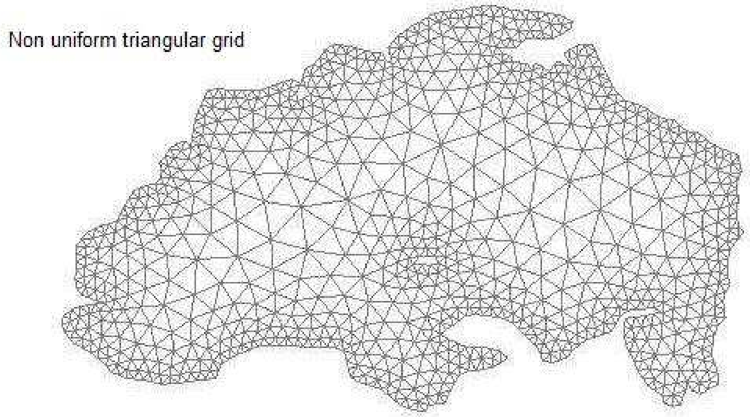
where (u, v) is the advection velocity vector, ω^2 is the radius of the stain of the pollutant substance and k is a constant some authors take the value $k = 0.03$. With this model and by using the observed Chlorophyll-a concentration we can obtain an approximate expression of these coefficients in the lake's surface.

Diffusion equation vs Lagrangian method

To obtain the concentration of Chlorophyll-a in Lake Valencia we must solve the diffusion equation. We take as initial condition $p_0(x)$ predicted by Kriging. We use a Crack-Nicholson for approximating the time derivatives and a second order differences finite method for the spatial derivatives.Two examples of the possible grids are displayed next.

Uniform triangular grid.




Non uniform triangular grid

We do not pursue that way because we are interested in answer another question: What is the kind of diffusions we deal? In that follows we shall compare two diffusion hypothesis: Brownian diffusion vs. fractionary pseudo-diffusion.
Let us to introduce the fractional Brownian motion (mBf) of parameter $0 < H < 1$ in notation B_H. This is a centered Gaussian process of stationary increments whose covariance is

$$\mathbb{E}B_H(t)B_H(s) = \frac{1}{2}[|t|^{2H} + |s|^{2H} - 2|t-s|^{2H}].$$

For $H = 1/2$ we have the Brownian motion. The fBm is self-similar and is continuous for all $0 < H < 1$. Whenever the parameter H is greater than $1/2$ a stochastic integral can be defined with respect this process, by means of Riemman sums. Moreover an Itô lemma also holds. Hence if $\mathbf{B}_H(t) = (B^1_H(t), B^2_H(t))$ is a bidimensional fBm. We can define the pseudo-fractional difusión as the solution of the following system of SDE.

$$dX(t) = \Sigma(X(t))d\mathbf{B}_H(t) - \nabla V(t, X(t))dt \quad X(0) = X_0.$$

The existence and uniqueness of such a system was studied by Nualart & Rascanu in [5]. The problem is now that the Focker-Planck equation for the density does not exist.

We will use, to compare both models, the Lagrangian method, which consists in to solve these systems of equations for an huge number of particles, by using some approximate method as: Euler's method or Milstein's method.

The actual density of the concentration of the substance will be the proportion of particles present in a surface region over the total number of particles.

For sake of brevity let us define as $M(t)$ one of the two process: W or $\mathbf{B}_H$. The Euler method consists to write for $k = 1, \ldots, [t/2^n]2^n$ the following recurrence

$$X(\frac{k}{2^n}) - X(\frac{k-1}{2^n}) = \Sigma(X(\frac{k-1}{2^n}))(M(\frac{k}{2^n}) - M(\frac{k-1}{2^n})) - \nabla V(X(\frac{k-1}{2^n}))\frac{1}{2^n}.$$

The increments of M, for the Brownian motion case, are iid Gaussian with variance $\frac{1}{2^n}$. However, in the fractional case we must simulate the increments of the process by using the Durbin-Levison algorithm.

We will show the results of this simulation but with a change of coordinates rendering the lake's surface in a square. First we interpolate by using a spline the two components of the velocity. We display in the next figure the u component.

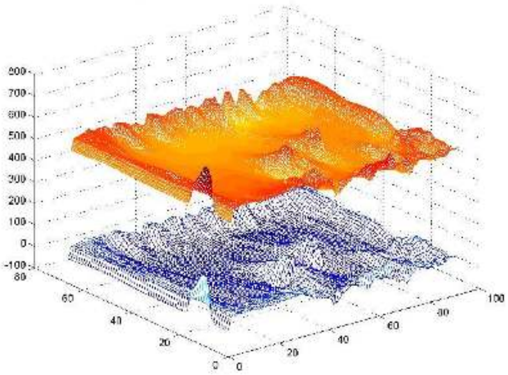


Figure: Coordinate u of the advection velocity.

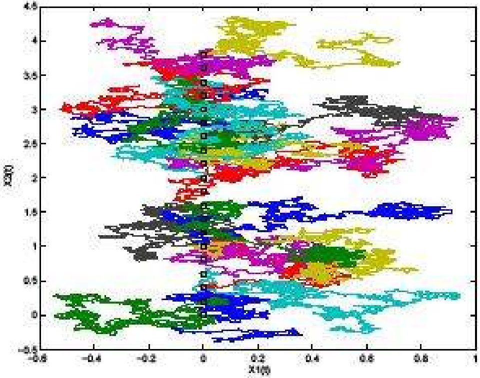


Figure: Euler method for Brownian diffusion.

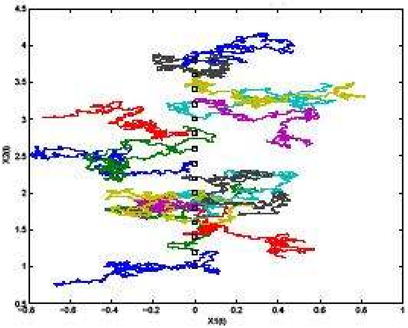


Figure: Euler method for fBm diffusion $H = 0.6$.

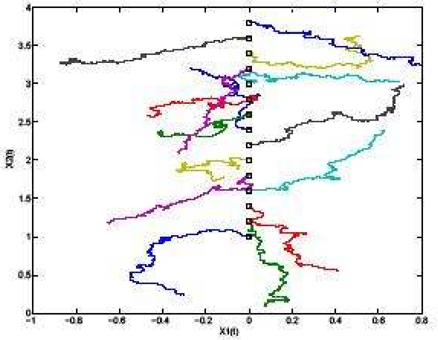


Figure: Euler method for fBm diffusion $H = 0.8$.

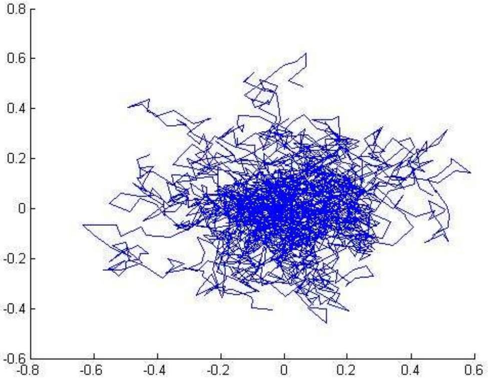


Figure: Milstein method for Brownian diffusion.

The following graph shows the result of a simulation on the surface of the lake. The advection and diffusivity coefficients are also simulated. These values are good approximations of actual values.

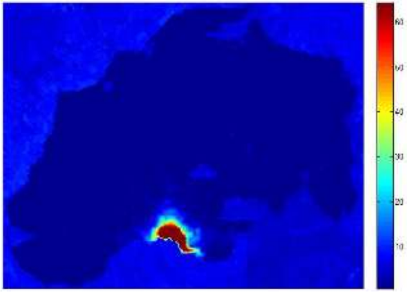


Figure: Simulation of a stain. Time 24 hours. 500 particles.

Conclusions and perspectives

- The Chlorophyll-a is a good indicator of water's quality.
- The concentration of Chlorophyll-a can be inferred by the index of reflectance in the satellite images.
- The Kringing method allows predicting the concentrations of Chlorophyll-a in all of the lake's surface.
- This prediction provides the initial conditions for algorithm of diffusion written in terms of SDE.

- It is important decide between Brownian diffusion and pseudo-fractional diffusion.

- For doing the above selection we need to estimate the Hurst parameter by a nonparametric method.

- To develop a system that meets our goals we need to transform our codes in a parallel computing framework.

- We want to end by offering an interactive system that can be used in different lakes and another pollutants substances.

Acknowledgments: I want to express my deepest gratitude to my master students: Claudia de la Hoz, Manuel Alejandro González, Alessandra Fariñas, Jocelyn León and María Fernández González, whose works have been the basis for this conference.

- C. Torres, G. Larrazabal, J. Castillo: Understanding three-dimesional circulation in Lake Valencia, Venezuela: A numerical apprach. Simulación y Modelado en Ingeniería y Ciencias, pp. MF49-MF56, SVMNI (2006).
- H. J. Gons, M. T. Auer, S. W. Effler: MERIS satellite chlorophyll mapping of oligotrophic and eutrophic waters in the Laurentian Great Lakes. Remote Sensing of Environment 112 (2008) 4098-4106.
- F. Gohin, S. Loyer, M. Lunven, C. Labry, J-M Froidefond, D. Delmas, M. Huret, A. Herbland: Satellite-derived parameters for biological modelling in coastal waters: Illustration over the eastern continental shelf of the Bay of Biscay. Remote Sensing of Environment 95 (2005) 29-46

- F. Gohin, B. Saulquin, H. Oger-Jeanneret, L. Lozac'h, L. Lampert, A. Lefebvre, P. Riou, F. Bruchon: Towards a better assessment of the ecological status of coastal waters using satellite-derived chlorophyll-a concentrations. Remote Sensing of Environment 112 (2008) 3329-3340.
- D. Nualart and A. Rascanu: Differential equations driven by fractional Brownian motion. Collectanea Mathematica 53, 55-81, (2001).
- C. Torres, G. Larrazabal, J. Castillo: Understanding three-dimesional circulation in Lake Valencia, Venezuela: A numerical apprach. Simulación y Modelado en Ingeniería y Ciencias, pp. MF49-MF56, SVMNI (2006).