Uncertainty quantification in 2D morphodynamic model: application to the Gironde estuary.

Leroux Romain[2], Goeury Cédric[1], El Kadi Abderrezzak Kamal[1,2], and Tassi Pablo[1,2]

[1] EDF R&D National Laboratory for Hydraulics and Environment, Chatou, France
`romain-r-externe.leroux@edf.fr`
[2] Saint-Venant Laboratory for Hydraulics, Chatou, France.
`cedric.goeury@edf.fr`
`pablo.tassi@edf.fr`
`kamal.el-kadi-abderrezzak@edf.fr`

Abstract

The purpose of uncertainty propagation is the quantification of input data uncertainties on the output results. This involves understanding **(i)** how uncertainty is represented in the model structure and the input data ? **(ii)** how are uncertainties propagated in the model ? **(iii)** Which uncertainties affect mostly the model outputs ? The propagation analysis begins with the identification and characterization of the uncertainties of the input data. The aim of this work is to estimate the uncertainties pertaining the parameters of a 2D morphodynamic model so as to characterize the probability distribution $\mathbb{P}\big[h(x, y, t) \leq h_{\text{critical}}\big]$ of the water depth $h(x, y, t)$ over the Gironde Estuary, where h_{critical} is a critical threshold of the water depth $h(x, y, t)$ that allows navigation. To handle this purpose, we propose an original approach that includes sediment parameters and bathymetry data, through the use of probabilistic methods, imprecise probability and non-linear regression. The proposed strategy offers flexibility to handle the variability of these data are also suitable for data-driven applications since the uncertainty quantification can also be conducted from a small set of parameters of the 2D morphodynamic model.

1 Introduction

This present work is part of a mulltidisciplinary project aiming at improving the capacity of the navigation channel on the Gironde estuary. The Port of Bordeaux receive ships and has to cope with the natural estuarian constraints, such as sedimention (clay and sand). In order to satisfy the demand of the market for increasing ships size, while ensuring navigation safety, the water-depth evolution $h(x, y, t)$ in the estuary needs to be predicted with a certain accuracy.

In environmental sciences, experts are regularly confronted with inaccurate or incomplete knowledge of the data needed to model a physical phenomenon. Numerical predictions are usually tainted with uncertainties. The propagation of uncertainties on the input data through simulations might affect strongly the simulation results, and thus the estimation of $h(x, y, t)$. It is necessary to quantify the contributions of input uncertainties to the morphodynamic model

results to appraise the sensitivity of the model to these parameters. Uncertainty analysis can help finding the input parameters that cause the largest output uncertainty, and can identify the most uncertain locations and the most uncertain time periods in hydro-mophodynamic model predictions.

The origin of uncertainties is diverse and may be related **(i)** to the measurement errors carried out to calibrate the numerical model **(ii)** limited available measurements, which make it difficult, for example, to estimate spatial heterogeneity in sediment properties **(iii)** to governing equation's assumption. Thus, it may be that the only information on the value of a parameter θ involved in a model is an interval between the maximum and minimum values $[\min, \max]$ of the θ. The value of this parameter is generally imprecise. Such information does not express random variability but inaccuracy. The parameter can be a random variable described by a single probability distribution (Gaussian, Log-Gaussian, etc) but generally the available measurements do not allow defining this distribution. The quality and quantity of available information determine the type of uncertainty representation.

In the probabilistic framework, the parameters of a model can be represented in a coherent way by a single probability distribution. For example, the sediment grain size d follows generally a log-normal distribution and the uncertainty is propagated using the Monte Carlo method. Nevertheless, the assumption about the nature of the probability distribution can lead to errors if the available information is incomplete. The selection of the true probability distribution is thus limited by the incomplete or inaccurate nature of the information. For an experimental context, sufficient data can not be available for a single model. In this case, an approach is used for a given variable where the mean and standard deviation are assumed. To represent inaccurate and incomplete information, formal frameworks for the representation of information have been developed: **(i)** the p-boxes: pair of high and low cumulative probabilities [2] **(ii)** the theory of belief functions of Shafer [5] **(iii)** the theory of possibilities [1].

Probability and p-box theory are used in this work to represent the data uncertainty. These theories are flexible in relation to the quantity and quality of available information. This report is organized as follow. The most important parameters of the 2D morphodynamic model under consideration and the uncertainties associated are described in Section 2. The characterization of the probability density distributions of the sediment size d_{50} and d_{90} are presented in Section 3. Section 4 is dedicated to the characterization of the probability density distributions of the settling velocity W_s for cohesive sediment. Section 5 presents the method used to handle the nugget effect due to measurement errors in the bathymetry to generate multiple random sets of the bathymetry.

2   Model inputs

The uncertainty propagation analysis begins with the identification and characterization of the uncertainties of the input data. The most important parameters to consider in a 2D morphodynamic model are: **(i)** the grain size diameters d_{50}, d_{90}, and **(ii)** parameters associated to the sediment transport, such as the settling velocity W_s, dimensionless critical shear stress for particle motion Θ_c, critical shear velocity for clay deposition $u^*_{c,clay}$, critical shear stress for erosion τ_{ce} and Krone-Partheniades erosion law constant M.

3   Grain size distribution analysis

Herein we present the methodology for estimating:

    - the evolution of the percentage of cohesive and non-cohesive sediments for each area of the Gironde estuary from years 2000 to 2016. Cohesive properties appear for fine particle (silts and clay), with a diameter less than $63\mu m$, depending on the physico-chemical properties of the fluid and salinity [7]. From the mean diameter and grain size distribution, two types of bed material are considered: cohesive, for which $d_{50} < 63\mu m$ and non-cohesive, for which $63\mu m < d_{50} < 2mm$.

    - the probability density distribution of d_{50} and d_{90} for each area of the Gironde estuary.

Our approach combines probability bound analysis to represent the uncertainty of the data and an estimation by maximum likelihood of its pdf. This methodology was applied for the four parts describing the Gironde estuary. The results obtained for the grain size analysis are represented with **(i) pie charts** for the evolution of the percentage of cohesive and non-cohesive sediments and **(ii) graphs of the pdfs** for the median size d_{50} and the d_{90}. We present here the results obtained for the fluvial part, as shown in Figures 1(a), 1(b) and 4. In order to analyze the temporal variability of the proportions of cohesive and non-cohesive sediments through the years, the data d of all sites for one area are gathered together for each year. Figure 1(a) shows 17 CDFs min and max of d for all the years 2000 to 2016 for the fluvial part of the estuary. Since the sediment size $\Phi = \log_{10}(d)$ becomes imprecise, its distribution becomes an imprecise CDF. In each area of the estuary, for every year, the data d are represented with a p-box that are included themselves between a minimal and maximal CDF. The distance between these two CDFs represents the uncertainty due to the imprecise variable Φ. For the fluvial part, one can see that $\Phi = [-3, -0.6]$ with few outside values. For ϕ_{90}, this distance becomes larger with values comprised between 3 and 0. Therefore there is a large imprecision on the data d. Under the assumption that Φ follows a Gaussian distribution $\Phi \sim \mathcal{N}\left(\mu_\Phi, \sigma_\Phi^2\right)$, one can write

$$p(\Phi) = \frac{1}{\sqrt{2\pi\sigma_\Phi^2}} \exp\left[-\frac{1}{2}\left(\frac{\Phi - \mu_\Phi}{\sigma_\Phi}\right)^2\right] \tag{1}$$

where the mean μ_Φ and the variance σ_Φ^2 are determined from all measurements used to defined the distribution. μ_Φ and σ_Φ^2 are given as:

$$\mu_\Phi = \sum_i p_i\left(\Phi_i + \Phi_{i+1}\right)/2 \qquad \sigma_\Phi = \sqrt{\sum_i \left(\left(\Phi_i + \Phi_{i+1}\right)/2 - \mu_\Phi\right)^2} \tag{2}$$

where p_i is the weight percentage in each Φ interval and $\left(\Phi_i + \Phi_{i+1}\right)/2$ is the mid point of each Φ interval. Since for one area and one year Φ is represented by a p-box, the distribution $p(\Phi)$ is imprecise, and becomes a p-box:

$$\Phi \sim \mathcal{N}\left(\left[\underline{\mu_\Phi}, \overline{\mu_\Phi}\right], \left[\underline{\sigma_\Phi^2}, \overline{\sigma_\Phi^2}\right]\right) \tag{3}$$

where the support of the p-box is given by $\Phi \in \left[\underline{\Phi}, \overline{\Phi}\right]$ since the range of the measured d is $[0, 2]\ mm$. For ine area, and seventeen years, seventeen p-boxs Φ are available. Since one p-box Φ is a parametric p-box, the next step find the one PDF in $\Phi \in \left[\underline{\Phi}, \overline{\Phi}\right]$, among all the probability densities that the p-box prescribes, that is most likely to have produced the data. The maximum likelihood method is used to find μ_Φ and σ_Φ characterizing the probability distribution that bettter describes the p-box (3). Natural sediments in estuaries and coastal areas are usually characterised by a mixture of sand, mud and organic matters. This heterogeneity can be

modeled by a mixing of cohesive and non-cohesive sediments. A sand-mud mixture can be therefore represented by two classes of bed material: the mud fraction and the sand fraction.

The non-cohesive sediment is represented by the median diameter d_{50}. The settling velocity W_s is assumed a function of the relative sediment density ($s = 1.65$) and grain diameter d. For cohesive sediment (with grain diameter $d \leq 63\mu m$), the settling velocity W_s is a function of various parameters and needs to be specified through measurements. The percentage of cohesive and non cohesive sediment in the area of the estuary are established by the quantiles $\mathbb{P}[d \geq 63\mu m]$ and $\mathbb{P}[d \leq 63\mu m]$ of this probability distribution (Figure 4). Based on these results, one can see that the 2D morphodynamic model used for the uncertainties propagation must handled mixed sediment transport. Nevertheless, more insight should be conducted on these results, especially the abrupt changes between the percentages of cohesive and non cohesive sediments trough the years. For example, the fluvial part in 2011, the sediment is mainly cohesive whereas it remains mostly non-cohesive in 2013. At the end of this procedure, a vector of Φ_{50} quantile is estimated, representing d_{50} for all years for one area. Since the limited available information does not allow estimating robustly too many parameters, the Bayesian Information Criterion (BIC) is used to find the probability distribution of the vector that contained d_{50} for all years, as shown in Figure 1(b) for the fluvial part. The d_{50} and the d_{90} distributions exhibit quite different shapes and supports. Contrary to the d_{50} pdf, the d_{90} pdf tends to be more uniform. The pdf mode of the d_{50} is located on lower values of the sediment size.

4   Sediment settling velocity analysis

Hereafter, we present the methodology estimating the regression model $Y = f(X)$ when X and Y are given and thus characterize the pdfs of X and Y. The purpose is to handle the variability of the data through the sampling of pdfs of the coefficients of the model f. The proposed strategy is based on two sets of experimental values of two sediment-related parameters X and Y. A nonlinear regression model is used to relate the set of response variables $Y = \{y_i\}_{i=1}^n$ to the set of explanatory variables $X = \{x_i\}_{i=1}^n$. The nonlinear regression model g that map variables Y to variables X is assumed to take the form $y_i = g(x_i, \beta) + \sigma \varepsilon_i$. The vector $Y = (y_1, y_2, \cdots, y_n)$ is a vector of iid (*independent identically distributed*) Gaussian random variables and indexed by the m-dimensional parameter $\theta = (\beta, \sigma^2)$ with $\theta \in \Theta$. Consequently, an estimate of the mean response at the point of interest x_0 is $\hat{y} = g(x_0, \hat{\beta})$ and the $\gamma = 95\%$ Wald confidence interval on the mean response at the point x_0 is $\hat{y} \pm 1.96 \cdot \text{se}(\hat{y})$. Figure 2 shows the predictions of the mean settling velocity for cohesive sediment with confidence intervals at 95% for the nonlinear model described hereafter. A transition occurs concentrations larger than $3g \cdot L^{-1}$. This regime corresponds to *the hindered regime* for which we use a new piecewise continuous model:

$$W_s = \left\{ \begin{array}{l} a_1 C^{b_1} \text{ for } C < 3g \cdot L^{-1} \\ a_2 C^{b_2} \text{ for } C \geq 3g \cdot L^{-1} \end{array} \right.$$

The distributions of the parameters a_i, b_i are given by $a_i \sim \mathcal{N}(\hat{a}_i, \sigma_{\hat{a}_i}^2)$ et $b_i \sim \mathcal{N}(\hat{b}_i, \sigma_{\hat{b}_i}^2)$, for $i = 1, 2$.
This methdology was also applied to the critical Shield number $\Theta_c = f(D_*)$ of non-cohesive sediment, where D_* is the adimensional particle diameter given by $D_* = (d_{50}/0.9)((s-1)g)\nu^{-2/3}$, the critical shear stress for erosion $\tau_{c,e} = f(C)$ of cohesive sediment and the Krone-Partheniades erosion law constant $M = f(C)$, for cohesive sediment.
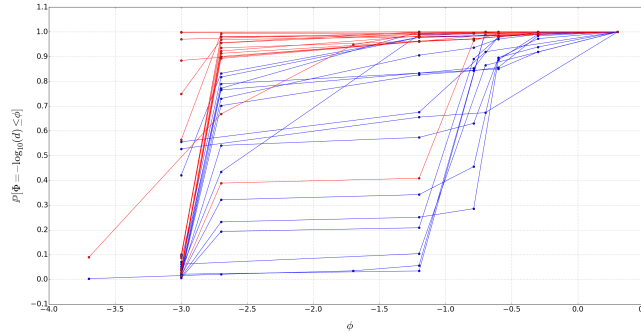
(a)
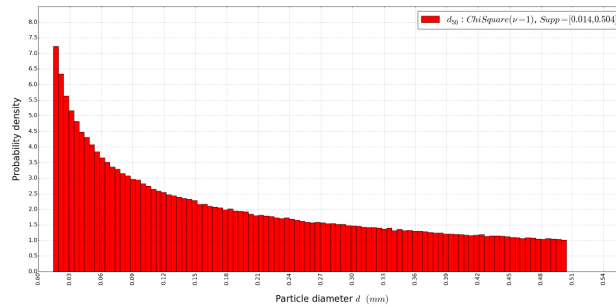
(b)

Figure 1: 1(a) CDFs of the granulometry data d for the fluvial part for all years 2000 to 2016. Red : upper CDFs, Blue : lower CDFs 1(b) Probability densities for the d_{50} and d_{90} for the fluvial part of the Gironde estuary.

5   Bathymetry analysis

In spatial interpolation, the nugget effect can be attributed to measurement errors and/or spatial sources of variation at distances smaller than the shortest sampling interval. The bathymetry is then modelled applying Universal Kriging [3]. Kriging was originally introduced as a spatial interpolation method in the earth sciences, and is also known as Gaussian Process regression in machine learning. At every measurement points $\mathbf{x}$, the Kriging simultaneously provides an approximation of a partially observed function z, the kriging mean predictor $\mu(\mathbf{x})$, and a measure of prediction uncertainty, the kriging variance $\sigma^2(\mathbf{x})$. The principle of Universal Kriging is to see $z(x,y)$ as one realizations of a random process $Z(\mathbf{x})$, and to make optimal linear predictions of $Z(\mathbf{x})$ given the observations values (noisy or not) at the measurement points $\mathbf{x}$. In the Universal Kriging framework, Z is generally assumed to be of the form:

$$Z(\mathbf{x}) = \mu(\mathbf{x}) + G(\mathbf{x}) = \sum_{i=1}^{n}\beta_i f_i(\mathbf{x}) + G(\mathbf{x}) \tag{4}$$

where f_i are known basis functions and G is a centered Gaussian process. In order take account of the nugget effect of measurement error in z, the bathymetry is first perturbed by Gaussian noise, in each node of the mesh used for the simulation of the 2D hydrosedimentary model. The kriging variance is used to construct confidence interval around the estimates. Under the
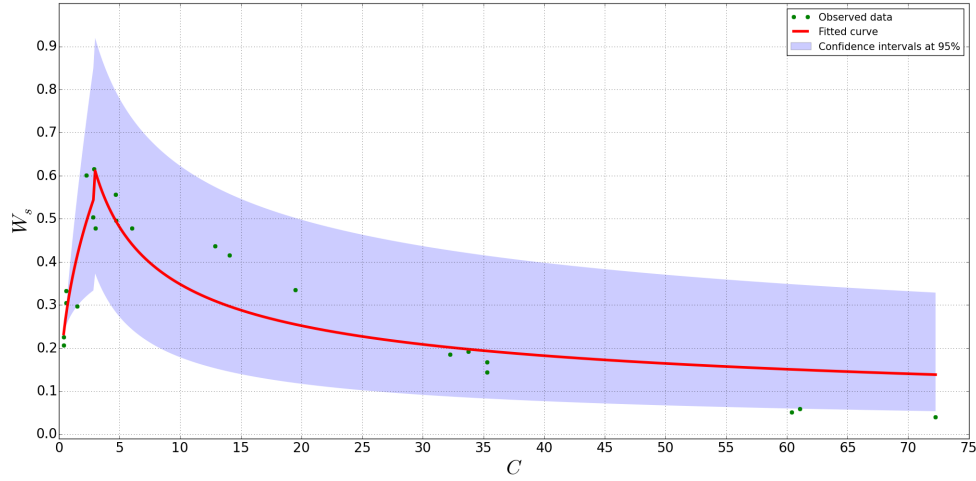
Figure 2: Predictions of the mean settling velocity W_s with confidence intervals at 95% (cohesive case), as a function of the sediment concentration C.

assumption that the error distribution at location $\mathbf{x}$ is Gaussian with zero mean and variance σ^2, the 95% confidence interval for $Z(\mathbf{x})$ with kriging estimate $z^*(\mathbf{x})$ is calculated as:

$$\mathbb{P}[Z(\mathbf{x}) \in z^*(\mathbf{x}) - 1.96\sigma(\mathbf{x}), z^*(\mathbf{x}) + 1.96\sigma(\mathbf{x})] = 95\% \tag{5}$$

which gives upper and lower bounds for a 95% confidence interval to construct confidence envelope maps. They represent the kriging model estimation of the 95% range of potential values for any location. A set of n_{map} maps of the bathymetry spatial distribution is generated by sampling a random field $Z^*(\mathbf{x})$ with a truncated Gaussian distribution of support $[z^*(\mathbf{x}) - 1.96\sigma(\mathbf{x}), z^*(\mathbf{x}) + 1.96\sigma(\mathbf{x})]$:

$$Z^*(\mathbf{x}) \sim \text{TruncatedGaussian}\big(z^*, \sigma(\mathbf{x}), [z^*(\mathbf{x}) - 1.96\sigma(\mathbf{x}), z^*(\mathbf{x}) + 1.96\sigma(\mathbf{x})]\big) \tag{6}$$

This overall procedure is repeated with n different levels of noise to generate bathymetry $n \times n_{map}$ maps of the bathymetry with the Universal Kriging. Figure 3(a) shows the original bathymetry of the Gironde estuary and Figure 3(b) shows a random field of the bathymetry generated from our method. These maps will jointly used as inputs with the other uncertain parameters in the Monte-Carlo simulations of the 2D hydrosedimentary model to estimate the uncertainties that pertained the water depth $h(x, y, t)$.

6   Conclusion

A new methodologie processing sediment data, hydrosedimentary parameters and bathymetry data, through the use of probabilistic method, imprecise probability and non-linear regression, was presented. The main findings of our work are: **(i)** the highlight of the main parameters of the 2D morphodynamic models : mainly granulometry parameters (such as the sediment size d_{50}, d_{90} and parameters related to the cohesive properties of the sediment), **(ii)** the evolution of the percentage of cohesive and non-cohesive sediments for each area of the Gironde estuary through the years 2000 to 2016, **(iii)** characterization of the probability density distribution of
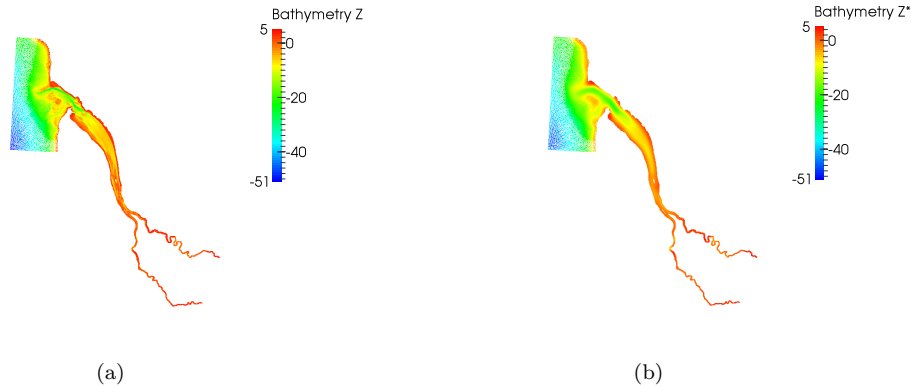
Figure 3: 3(a) Original bathymetry of the Gironde estuary. 3(b) Random field $Z^*(\mathbf{x})$ of the bathymetry generated from our method.

d_{50} and d_{90} for different areas of the Gironde estuary, **(iv)** the 2D hydrosedimentary model used for uncertainties propagation must handled mixed sediment transport **(v)** the characterization of the probability density distributions of the critical Shields number Θ_c, the settling velocity W_s for cohesive and non-cohesive sediments, the critical shear stress for erosion τ_{ce} and the Krone-Partheniades erosion law constant M for cohesive sediment, **(vi)** a new model for the settling W_s for cohesive sediment, **(vii)** the generation of multiple sets of the bathymetry z of the estuary that handle the nugget effect due to measurements errors, **(viii)** numerical codes using OpenTurns uncertainty library [4] to process the data in Sections **2,3,4,5**. Theses strategies offer flexibility to handle the variability of these experimental data are also suitable for data-driven applications since the uncertainty quantification can also be conducted from a small set of the most important parameters of the 2D morphodynamic model under consideration. They also allow to generate random sample of inputs, using their estimated probability distributions. In order to quantify the impact of the inputs uncertainties on the estimation of the water-depth $h(x, y, t)$, the Monte Carlo method will be used to propagate these samples in a 2D morphodynamic model [6].

References

[1]   Dubois, D. and Prade, H., 1988 Possibility theory. *New York Plenum Press.*

[2]   Ferson, S. and Ginzburg, L., 1996 Different methods are needed to propagate ignorance and variability. *Reliability Engineering and System Safety.* **54**, 133–144.

[3]   Montero, José and Fernández-Avilés, Gema and Mateu, Jorge, 2015 Spatial and Spatio-Temporal Geostatistical Modeling and Kriging. *Princeton University Press.* **08**

[4]   Airbus-EDF-IMACS-Phimeca., 2016 Reference Guide - OpenTurns 1.7

[5]   Shafer, G., 1976 A mathematical theory of evidence. *Princeton University Press.*

[6]   Hervouet JM; Ata R., 2017 User manual of opensource software TELEMAC-2D *EDF-R&D* V7P2

[7]   Thorn, M.F.C., 1981 Physical processes of siltation in tidal inlets. *Proceedings of Hydraulic Modelling applied to Maritime Engineering Problems, ICE, London.* 47–55.
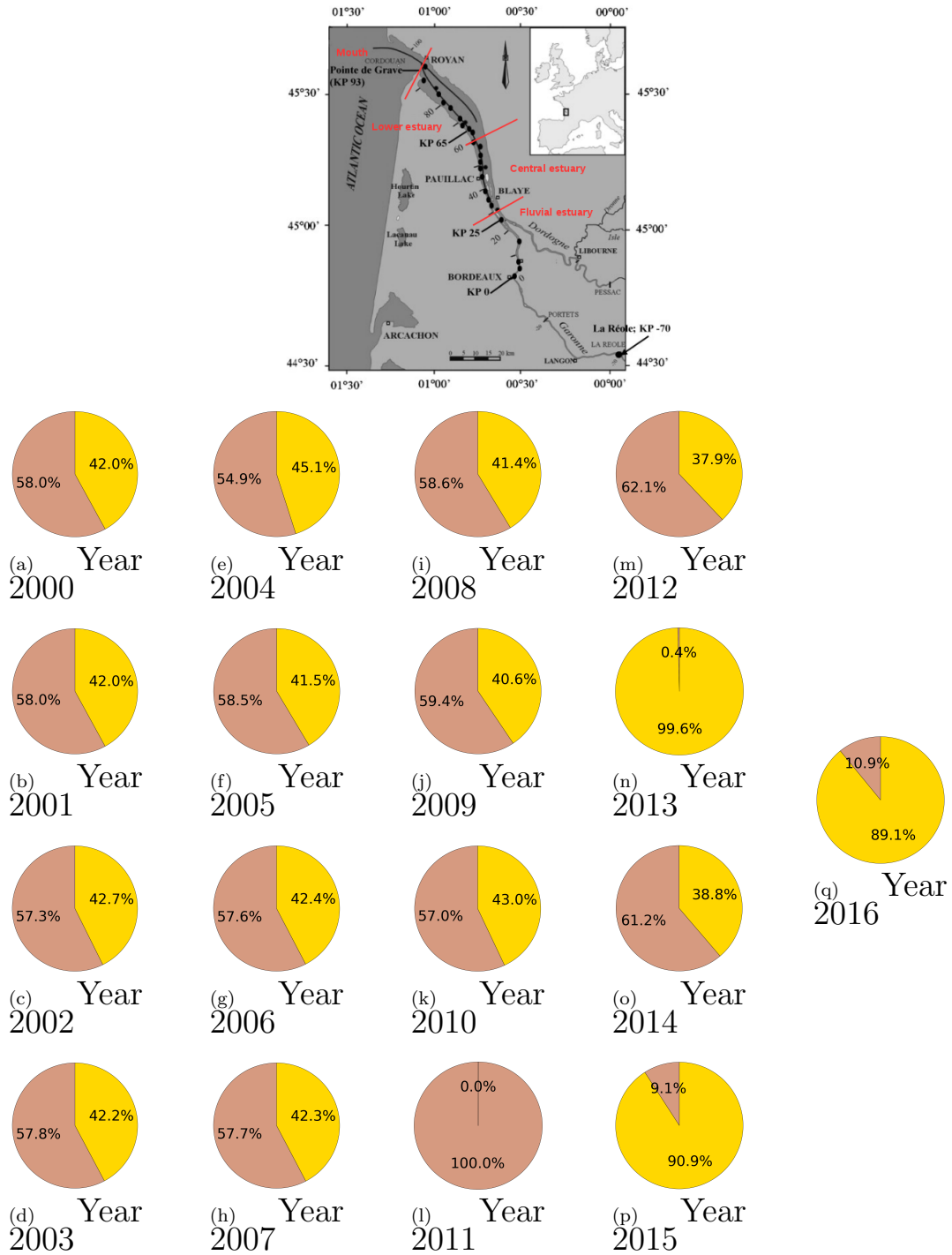
(a) Year 2000

(b) Year 2001

(c) Year 2002

(d) Year 2003

(e) Year 2004

(f) Year 2005

(g) Year 2006

(h) Year 2007

(i) Year 2008

(j) Year 2009

(k) Year 2010

(l) Year 2011

(m) Year 2012

(n) Year 2013

(o) Year 2014

(p) Year 2015

(q) Year 2016

Figure 4: Percentage of cohesive ■ and non-cohesive sediment ■ in the fluvial part of the Gironde estuary for the years 2000 to 2016.

1154