A Probabilistic Collocation Method for the Imperfection Propagation: Application to Land Cover Change Prediction

Wadii Boulila[1, 2], Amine Bouatay[1], Imed Riadh Farah[1, 2]
[1]Laboratoire RIADI, Ecole Nationale des Sciences de l'Informatique
Campus Universitaire de la Manouba
2010 La Manouba, Tunisia
[2]Laboratoire ITI, TELECOM-Bretagne
Technopole Brest Iroise CS83818
29238 Brest Cedex, France
Wadii.Boulila@riadi.rnu.tn

ABSTRACT: *The process of land cover change prediction is generally subjected to several types of imperfections which affect the reliability of decision about these changes. Several works in literature are carried out in an attempt to mitigate the issues caused by data related imperfections. Although new prediction models are created and improvement is done for the existing ones, both the imperfection related to the input of models and its propagation through models are disregarded. To bridge this research gap, we propose a methodology that propagates imperfection throughout a model of land cover change prediction. The proposed approach incorporates three steps: 1) computing membership functions for input variables of the model of land cover change prediction, 2) applying a sensitivity analysis technique to determine which input variables are the most influential in the overall imperfection model, and 3) propagating distributions of the most influential input variables throughout the model of land cover change prediction.*

Experiments are made on images representing the Saint-Denis region, capital of Reunion Island. Results show the effectiveness of the proposed methodology in improving both computation time and prediction of the land cover change.

Keywords: Imperfection Propagation, Sensitivity Analysis, Collocation Method, Membership Functions, Land Cover Change Prediction, Satellite Images

Received: 12 October 2013, Revised 25 November 2013, Accepted 30 November 2013

1. Introduction

Remote sensing data has made significant contributions to many environmental engineering and civil applications. It is being used increasingly in several fields to include land cover change prediction, land use monitoring and management, fire protection, deforestation, desertification and environmental systems.

Decisions based on remote sensing data in these fields strongly depend on the accuracy of these data. However, when remote sensing data is generated by using a measuring device and derived by various methods and models, it is generally marked by imperfection. Several types of imperfection can be listed such as uncertainty, imprecision, and ignorance. The imperfection can be related to data: remotely sensed errors, overlapping errors, position errors, scaling errors, measurement errors. It can be also related to processing methods or results.

In this paper, we are interested in predicting land cover changes. However, building reliable procedures to predict land cover changes requires taking into account imperfections related to the prediction process.

In literature, several studies have been devoted to predicting land cover changes while considering the imperfect aspect of data [8] [1]. The majority of these studies treat imperfection on three levels: the input data, the model and model outputs. However, the propagation of imperfections from one level to another is not well addressed.

The land cover change prediction is often marked by several types of imperfections that affect the accuracy of decisions. The source of this imperfection is not always due to an existing inconsistency for the data, but it may be due to the ambiguity and lack of precision. Epistemic imperfections are left for forthcoming works. We are interested in random imperfections (stochastic). That is, a probabilistic collocation method is used to propagate imperfection as it is a computationally efficient method for propagating imperfection on large complex models.

The objective of this paper is to present a methodology for imperfection propagation through a model for land cover change prediction. The prediction model is presented in previous works [3] [4] [5]. In these works, we presented a model to predict spatiotemporal changes in satellite image databases. The proposed model exploits data mining concepts to build predictions and decisions for several remote sensing fields. It takes into account imperfections related to the spatiotemporal mining process in order to provide more accurate and reliable information about land cover changes in satellite images. Nevertheless, the propagation of imperfection through this model is not considered. Decisions on land cover could change consequently.

Therefore, we propose an approach which intend to reduce imperfection related to the prediction process and therefore to improve decisions about land cover change. This approach is based on sensitivity analysis and imperfection propagation methods.

This paper is organized as follows: in Section II we discuss related works concerning propagation of imperfection and sensitivity analysis. Section III describes the proposed approach for propagating imperfection throughout the model of land cover change prediction. Finally, Section IV presents our experiments and results.

2. Related Works

Land cover change prediction contributes significantly to assess, manage and monitor different issues related to environmental changes. However, the prediction is generally marked by several types of imperfection. This imperfection can be propagated throughout the prediction process.

This section is divided into two parts. The first one describes concepts of propagation of imperfection. The second describes sensitivity analysis.

2.1 Propagation of Imperfection

The goal of imperfection propagation is to estimate the change's rate in the model output due to some input data or parameters change's model. In literature, most of studies are focusing on two families of uncertainty propagation: 1) probabilistic and 2) non-probabilistic methods [6]. Probabilistic methods have the advantage of being simple to represent. They have no restrictions on input attributes. The probability distribution functions defining these attributes are assumed to be known. In the current paper, we focus in studying probabilistic uncertainty propagation methods.

Among probabilistic methods, we list the Monte Carlo method (MC). It is a very common statistical and probabilistic method for propagating uncertainty [24] [25]. The probability density of the final result is obtained from compiling the results a large number of runs. However, MC has two problems:the computation time and complexity. Many methods based on polynomial chaos have been developed to reduce the computational effort.

Another probabilistic method is Galerkin Polynomial Chaos method (GPC) [6] [22]. It is an intrusive uncertainty propagation method. Intrusive means that the uncertainty propagation method requires modifying model in which uncertainty propagation method will be applied. It is based on the homogeneous chaos theory of Wiener [21], who constructed a chaos expansion using Hermite polynomials. The intrusive GPC approach requires the modification of the deterministic system that leads to a stochastic huge system.

Non-intrusive polynomial chaos methods (NIPCM) are proposed to overcome the model modificationproblem [23]. These methods use the model as a black-box. Both GPC and NIPCM methods use sampling to estimate coefficients of the polynomial chaos expansion. For the same accuracy, run's number is much less than Monte Carlo simulation.

2.2 Sensitivity Analysis

A sensitivity analysis can extend an uncertainty analysis by identifying which input parameters are important (due to their estimation uncertainty) in contributing to the prediction imprecision of the outcome variable. Therefore, a sensitivity analysis quantifies how changes in the values of the input parameters alter the value of the outcome variable [26]. This helps refine their modeling and therefore reducing the impact of the imperfection. The identification of the major sources of imperfection is an important practice in the assessment of model performance as it allows concentrating the available resources on the major sources of imperfection [15] [16]. Saltelli et al. in [15] lists three types of methods for sensitivity analysis. The first type of methods is local techniques such as Fourier amplitude sensitivity test (FAST). Local techniques allow estimating the response outputs to variation of individual inputs or parameters while fixing the other parameters at their initial values. The second type of sensitivity methods is qualitative technique denoted also as screening methods. Qualitative techniques are "*one-at-a-time*" experiments in which the impact of changing the values of each of the chosen factors is evaluated [17]. The last type of sensitivity analysis methods is a global technique denoted also as variance-based methods. Global techniques allow the variation of all inputs or parameters with the later implicitly accounting for parameter interactions [18].

2.3 Input Parameter Estimation

In order to take into account the propagation of imperfection through the prediction model, we need to estimate distribution of input data. In literature, there are three categories of estimation methods for probability densities [13]: 1) non-parametric approaches, 2) semi-parametric approaches and 3) parametric approaches.

Among methods for estimation of probability densities using the first family, we list histogram and kernel density estimator [14], and kernel differomorphism method [19]. Finite mixture models (FMM) belongs to the second family method which is the semi-parametric approaches [13].

For the third family, we list the maximum likelihood methods [27]. This method is simple and has a good convergence property but it is limited to the great complexity's problems. It allows giving an estimator of a parameter of an unknown probability law whose independent realizations are observed.

3. Proposed Approach

The proposed approach, as shown in Figure1, is based on two steps. The first one consists of a sensitivity analysis. The goal of this step is to identify among the input variables which are the variables that affect more the output of the model of land cover prediction. So refining the modeling of these variables helps reducing their impact in the overall model imperfection. In the second step, a probabilistic collocation method is used to propagate imperfection related to variables identified in the first step. The collocation method is a probabilistic polynomial chaos method [2] [7] [21]. It is a class of non-intrusive polynomial chaos methods, it uses the model as a black box [11] [20].

3.1 Review of the Module for Land Cover Change Prediction

In previous works [5] [3] [4], we presented an approach to predict spatiotemporal changes in satellite images. In order to better understand the process of land cover change prediction, let suppose that an object is extracted from a satellite image acquired at date *t* using previous work [5].This object can be a lake, vegetation zone, urban area, etc. Five features are considered for this object which are: the radiometry, geometry, texture, spatial relations, and acquisition context. Each feature is described through a set of attributes A_i $(1 \le i \le N)$. We note by a state the set of attribute values computed for the object at a given date. In [3], an object database is built by an offline process. This database is composed by a set of models. Each model is composed by a set of states. Each state represent the same object at a different date.

The prediction process is divided into three main steps. It starts by a similarity measurement step to find similar states (in the object database) to a query state (representing the query object at a given date). The second step is composed by three sub-steps: (1) finding the corresponding model for the state, (2) finding all forthcoming states in the model (states having dates superior to the date of the retrieved state), and (3) for each forthcoming date, build the spatiotemporal change tree for the retrieved state. The third step is to construct the spatiotemporal change for the query state. Interested readers can refer to [3] [4].

The module for the prediction of land cover changes allows taking into account imperfection related to the prediction process. However, the propagation of the input imperfection through this module is not considered.

Thus, we propose to develop a methodology for uncertainty propagation through the module of land cover change prediction. This methodology combines sensitivity and imperfection propagation methods to reduce imperfection related to the prediction process.

3.2 Imperfection Propagation Module

Let X be a random variable with a probability density $f(X, \theta)$ analytically known but one of the parameters θ is unknown (numerically). The theory of chaos polynomial proposes to characterize random variable X with chaos polynomial representation. The problem is to construct an analytical expression based on achievements of this variable in a sample size n, to find the most likely numerical value for the parameter θ [28].

If $\{x_1, \ldots, x_n\}$ are independent realizations of the random variable X. We can say that $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ is a random realization of vector $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ whose components X_i are independent pairs.

The a priori probability of occurrence of a sample can then be characterized by the product of the probabilities of occurrence of each of the outputs (since these are assumed to be independent pairs).

$$P(\vec{X} = \vec{x}) = \prod_{i=1}^{n} f(x_i, \theta) \tag{1}$$

The objective of the method of maximum likelihood is to find the value that leads to the maximal probability. The product of the values $f(x_i, \theta)$ is also noted L $\{x_1, \ldots, x_n, \theta\}$ and called likelihood function.The value of $\hat{\theta}$ that maximizes the likelihood function is the solution of the following equation:

$$\frac{\partial Ln\, L}{\partial \theta} = 0 \rightarrow \hat{\theta} : \frac{\partial^2 Ln\, L}{\partial \theta^2} < 0 \tag{2}$$

Let X be a normal distribution $N(\mu, \sigma)$ with σ known but μ unknown. The objective is to construct an estimator of the value μ, given a sample implementation $\vec{x} = (x_1, \ldots, x_n)$. To do this, we start from the likelihood function of this sample:

$$\begin{gathered} \vec{L}(x, \mu) = \prod_{i=1}^{n} f(x_i, \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2} \\ L(\vec{x}, \mu) = K \prod_{i=1}^{n} e^{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2} \\ Ln\, L(\vec{x}, \mu) = K' - \frac{1}{2} \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^2 \\ \frac{\partial Ln\, L}{\partial \mu} = 0 \rightarrow \hat{\mu} : \frac{1}{\sigma} \sum_{i=1}^{n} \left(\frac{x_i - \hat{\mu}}{\sigma}\right) = 0 \\ \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i \end{gathered} \tag{3}$$

The arithmetic mean is the most efficient estimator of the expected value in the case of the normal distribution.

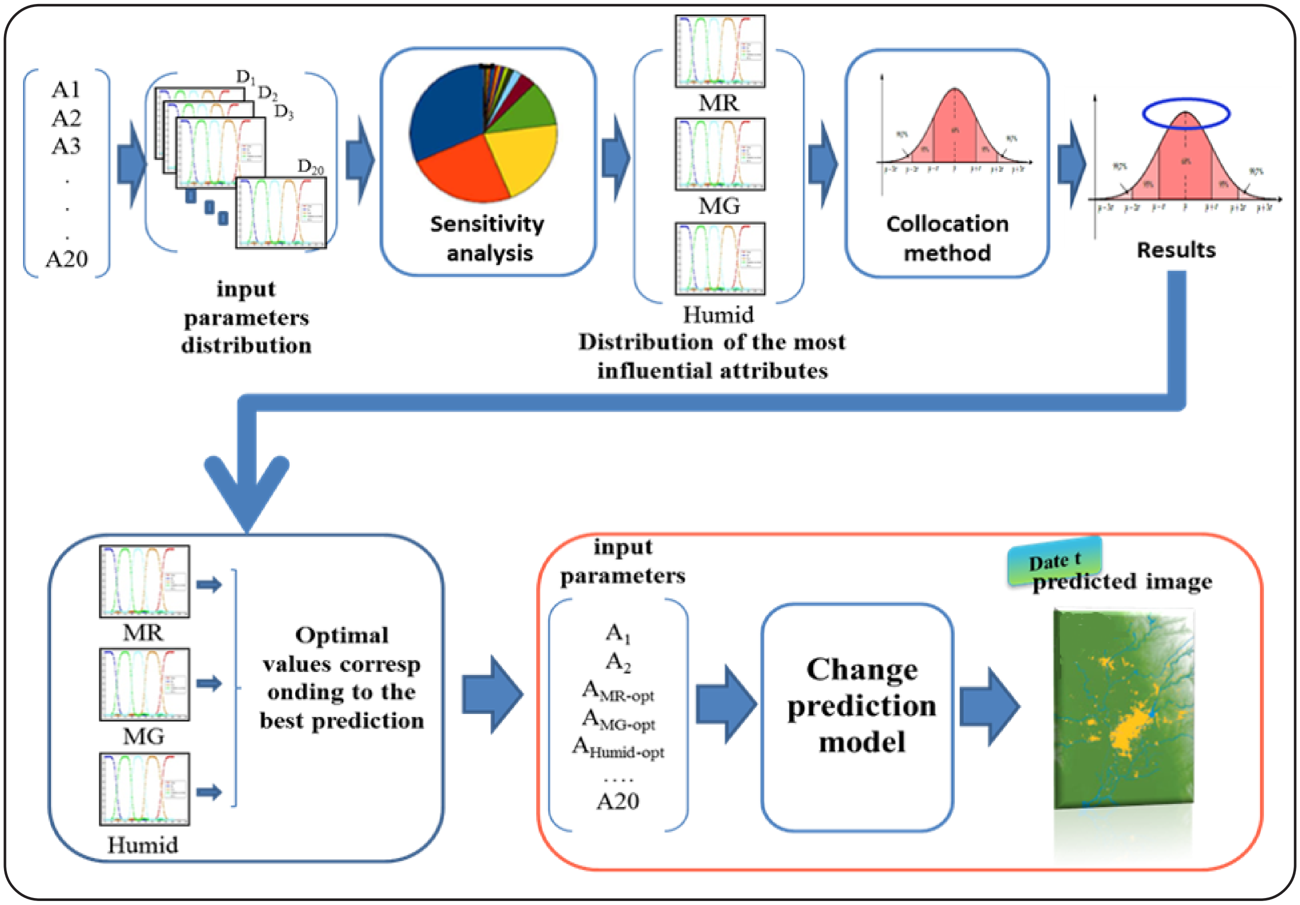

Figure 1. The Proposed Approach

-The mean estimator is the experimental average. It is given by the following equation:

$$\hat{\mu}_n(\vec{X}) = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (4)$$

- The estimator of the variance is the experimental variance. It is given by the following equation:

$$E(\hat{\mu}_n) = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \mu \qquad (5)$$

3.3 Sensitivity Analysis

The main objective of the analysis sensitivity is to determine what are the input attributes of describing an object that contribute most to the imperfection of the prediction model output. In the proposed approach, a global sensitivity analysis is used [12].

In this study, we focus more specifically on the methods based on the study of the variance of sensitivity indices for quantifying the influence of input attributes.

Let us denote by *M* the model for land cover change prediction, $S = (A_1, \ldots, A_p)$ p independent input variables with known distribution and *Y* the output of the prediction model. *Y* are decisions about land cover change.

Variance-based sensitivity indices, also called Sobol indices (S_J), are used in the proposed approach as an evaluation index for output variables. The index *SJ* for output variable *Y* according to input variables $A = (A_i)_{i=1,\ldots,p}$ is computed as follow:

$$S_J = \frac{V_J(Y_m)}{Var(Y)} \qquad (6)$$

Where $V_J\ (Y_m)$ is computed using equation (7) and *Var* (Y) is computed using equation (8).

$$\begin{aligned} V_J\,(Y_m) &= Var\,[\,\mathbb{E}\,(Y_m\,|\,A_i\,)] \\ &= Var\,\{\mathbb{E}\,[\mathbb{E}\,(Y\,|\,A\,)\,A_i\,]\} \\ &= Var\,[\mathbb{E}\,(Y\,|\,A_i\,)] \end{aligned} \tag{7}$$

The variance of Y, *Var* (Y), is the sum of contributions of all the input variables $A = (A_i)_{i=1,\dots,p}$ and A_ε. Here, A_ε denotes an additional input variable, independent of A, named seed variable. This variable is an uncontrollable variable used to make the prediction model stochastic.

$$Var\,(Y) = V_\varepsilon(Y) + \sum_{i=1}^{p} \sum_{|j|=i} [V_J\,(Y) + V_{J\varepsilon}(Y)\,] \tag{8}$$

$$V_\varepsilon\,(Y) = Var\,[\mathbb{E}\,(Y\,|A_\varepsilon)\,],\ = Var\,[\mathbb{E}\,(Y\,|A_i A_\varepsilon)\,] - V_i\,(Y) - V_\varepsilon\,(Y)\,,\dots. \tag{9}$$

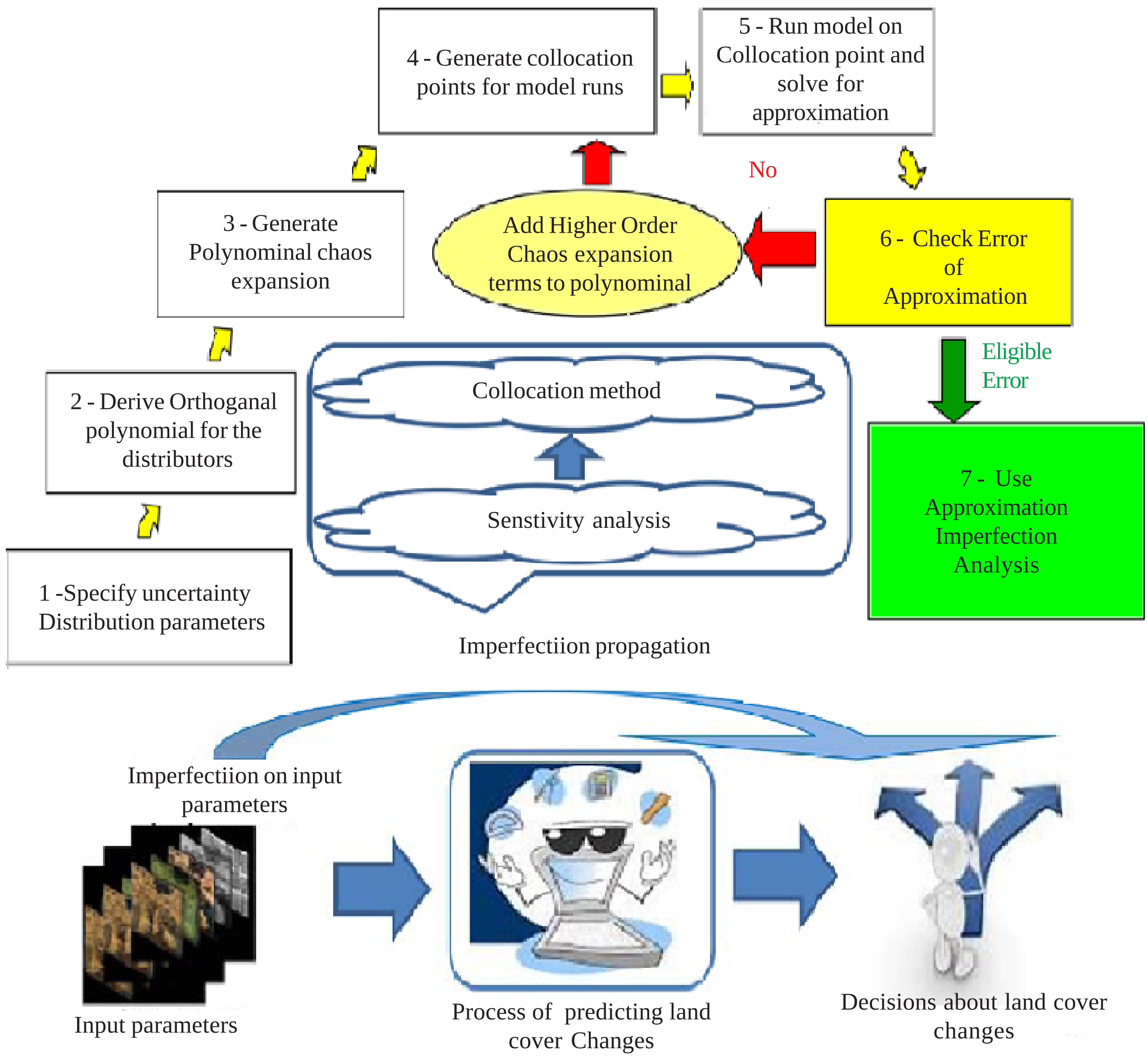

Figure 2.The Polynomial Collocation Method (PCM)

The interpretation of Sobol index is straightforward as their sum is equal to one. The larger an index value, the greater the importance of the variable or the group of variables linked to this index [12]. In the case of a model with p input variables, the number of Sobol indices is $2^p - 1$.

The basic concept of the Probabilistic Collocation Method (*PCM*) is to approximate the response of the prediction model to some polynomial function of uncertain parameters:

$$Y = f(I_1, \ldots, I_n)$$

The goal of the *PCM* is to approximate Y, by a simple polynomial function of I_k. Orthogonal polynomials are used by *PCM* approximation. They are also used to generate sets of parameter values, called collocation points [20].

Steps of the collocation approach are illustrated in Figure 2. First, input attributes are identified and their distributions of uncertainty are determined. The determination of the distribution may either be based on the expert knowledge, or be based on statistical data. Second, the orthogonal polynomial distributions are derived. If the approximation of the response of the prediction model has an order equal to p, orthogonal polynomials up to order p +1 are determined. Third, a polynomial expression is generated to represent the performance or output variable based on orthogonal polynomials of random variables $\{\xi_1, ..., \xi_n\}$. This is called the extension of polynomial chaos. Since the model is a black box, we can use a linear approximation in the first estimate.

$$y' = y_0 + \sum_{i1=1}^{n} y_1 \Gamma_1(\xi_{i1}) + \sum_{i1=1}^{n} \sum_{i2=1}^{i1} y_2 \Gamma_2(\xi_{i1}, \xi_{i2}) + \sum_{i1=1}^{n} \sum_{i2=1}^{i1} \sum_{i3=1}^{i2} + y_3 \Gamma_3(\xi_{i1}, \xi_{i2}, \xi_{i3}) + \sum_{i1=1}^{n} \sum_{i2=1}^{i1} \sum_{i3=1}^{i2} \sum_{i4=1}^{i3} y_4 \Gamma_4(\xi_{i1}, \xi_{i2}, \xi_{i3}, \xi_{i4}) + \ldots \quad (10)$$

Where

y_i are deterministic coefficients to be estimated.

$\{\xi_1\}_{i=1\ldots\infty}$ is the set of random variables associated with reduced centered Gaussian random variables A_i.

$\Gamma_p()$ denotes the multidimensional Hermite polynomials of degree p.

The random inputs and outputs are approximated by the PC expansions. These expansions contain unknown coefficients of the outputs. Calculating these coefficients is made by solving a linear system of equations that uses a selected number of collocations points. For a problem with n random variables, the total number of deterministic solutions required is given by equation (11).

$$T = \frac{(p+n)!}{p!\ n!} \quad (11)$$

Where p is the PC order.

For example: if n = 2 and p = 2 then T = 6, the output variable can be written as follow:

$$Y \cong Y' = y_0 + y_1 \Gamma_1(\xi_1) + y_2 \Gamma_1(\xi_2) + y_3 \Gamma_2(\xi_1, \xi_1) + y_4 \Gamma_2(\xi_1, \xi_2) + y_5 \Gamma_2(\xi_2, \xi_2) \quad (12)$$

In this study, collocation points are chosen as roots of the higher order polynomials. Specifically, the $n+1$ roots of the $(n+1)^{\text{th}}$ polynomial order corresponding to each parameter y_k are used to define collocation points. Thus, Y that is particularly good within the most probable range of values of input variables. Moreover, roots of the $(n+2)^{\text{th}}$ order polynomials are used to define another set of collocation points that can be used to estimate the error of the approximation.

After, we run the model for each of input sets, we get y_i as the corresponding result. Then, by replacing each ξ_I in the approximation of Y, we can solve the three simultaneous equations for the unknowns y_0, y_1, y_2, y_3, y_4 and y_5.

Before using the approximation of Y, the quality of the adjustment is computed. This is achieved by calculating the error between consecutive polynomial orders. In this work, we use the method proposed in [20] to determine collocation points. These points are obtained from the next orders of the orthogonal polynomial. If the error is greater than a given threshold, we need to pass to a higher order approximation and recalculate the error between consecutive approximations.

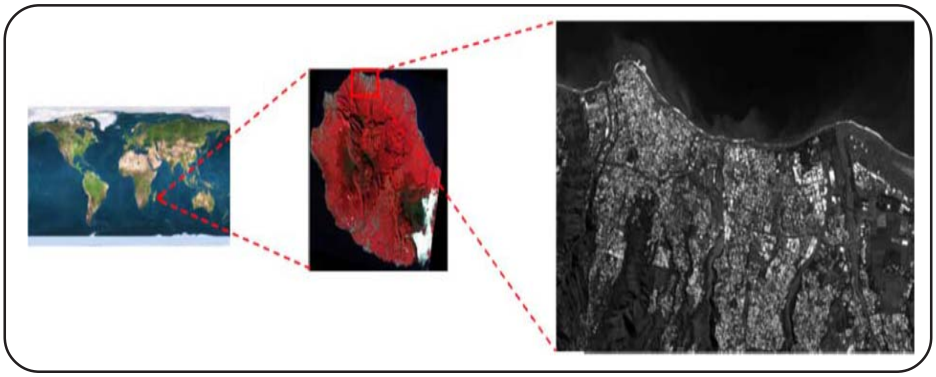

Figure 3. The studied area

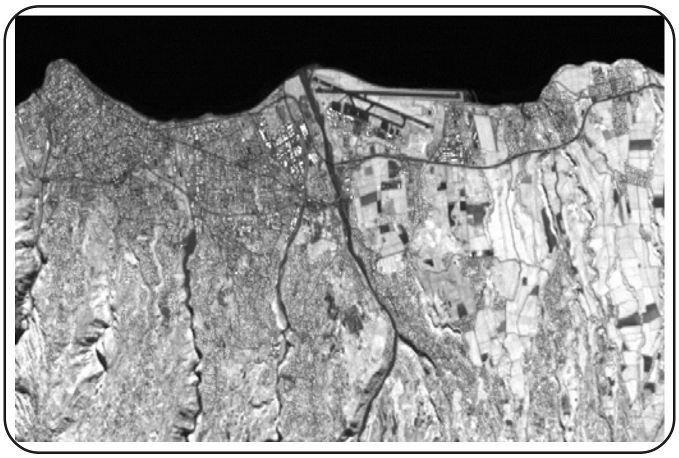

Figure 4. Satellite image acquired on June 12, 2007

The error term is denoted as e_{App} and calculated according to the equation (13):

$$e_{App} = \frac{\sqrt{\frac{1}{T} \left| \sum_{i=1}^{T} (y_i^{p+1} - y_i^{p}) \right|}}{\frac{1}{T} \sum_{i=1}^{T} y_i^{p}} \tag{13}$$

Where

T is the number of terms in the approximation y_i^p , equivalent to the number of collocation points.

y_i^p is the values of Y at the collocation points for a $(p+1)^{th}$ polynomial order approximation using a pth polynomial order approximation.

y_i^p is the values of Y at the collocation points for a $(p+1)^{th}$ polynomial order approximation using a $(p+1)^{th}$ polynomial order approximation. Steps 4, 5 and 6 in Figure 2 are repeated until obtaining an error lesser than a given threshold. Therefore, the obtained approximation can be used for the model of land cover change prediction.

Such a complex model can be reasonably approximated by a polynomial.

4. Experimental Results

The experimental result section is divided into two parts: validation of the proposed approach and evaluation of the proposed approach compared with existing ones.

The study area is located in the north-eastern Reunion Island in the Indian Ocean (55°13'1.07"E to 20°51'46.35"S), east of Madagascar (Figure 3).

Experiments were conducted on SPOT4 satellite images and belong to the Kalideos[1] database set up by the CNES[2]. Images dated June 12, 2007 and June 09, 2011 with a spatial resolution of 10 m, and a size of 1190 × 670 pixels were acquired (Figure 3 and Figure 4). Images were orthorectified and coregistered to the UTM coordinate system with a root mean square error of less than 0.5 pixel per image.

4.1 Validation of the proposed approach
The validation section is divided into three main steps:
1) Computation of Membership functions, 2) Sensitivity Analysis, 3) Imperfection Propagation.

4.1.1 Presentation of Uncertain Input Parameters
The aim of the proposed approach is to propagate imperfection through the model for land cover change prediction presented in [3].

In the current study, we are concerned by predicting urban changes of the study region between thetwo dates 2007 and 2011. To achieve this goal, we take as input to the prediction model the image acquired on June 12, 2007 (Figure. 4). However, the image acquired on June 09, 2011 (Figure 5) is only used for evaluation of results proposed by the proposed approach.

Let us consider that an urban object is extracted after a segmentation of the image in the Figure 4 using previous work [5]. The urban object is described by five features (radiometric, geometric, textural, spatial and acquisition context). Each feature is characterized by a set of attributes. 20 attributes are considered to describe the urban object work [5].

These attributes represent the input of the proposed approach and are:

• **Radiometry:** The mean radiometric, the standard deviation, the swekness, and the kurtosis.

• **Geometry:** We use length, width, perimeter, and area. These parameters are computed from theminimum bounding rectangle of the urban object.

• **Texture:** We use seven attributes from Gabor which are: energy, entropy, correlation, homogeneity, contrast, mean Gabor and variance Gabor.

• **Spatial localization:** Directional and metric relations are used to describe the spatial localization.

• **Acquisition context:** We use temperature, pressure, moisture.
Interested readers can refer to our previous work [5].

4.1.2 Computation of Membership Functions
The first step in the proposed methodology is to estimate the membership function for the urban attributes. These 20 attributes are random attributes (A_1,..., A_{20}) represented byprobability distributions (P_1, ..., P_{20}).

Figure 6 depicts results obtained after the application of the algorithm of the maximum likelihood method for the attribute mean radiometric.

4.1.3 Sensitivity Analysis
The second step in the proposed approach consists of the application of the sensitivity analysis to identify most influential input parameters. This helps determining input attributes that contribute most to the overall imperfection of prediction model. Figure 8 depicts that, the mean Gabor part of textural feature, the radiometric mean part of radiometric feature, and the moisture part of the acquisition context feature are the most involving and influential attributes to the overall imperfection of the prediction model.

After applying the sensitivity analysis process, we will consider only three input variables which are the mean Gabor, the mean

[1]http://kalideos.cnes.fr.

[2]Centre National d'Etudes Spatiales (French Space Agency).

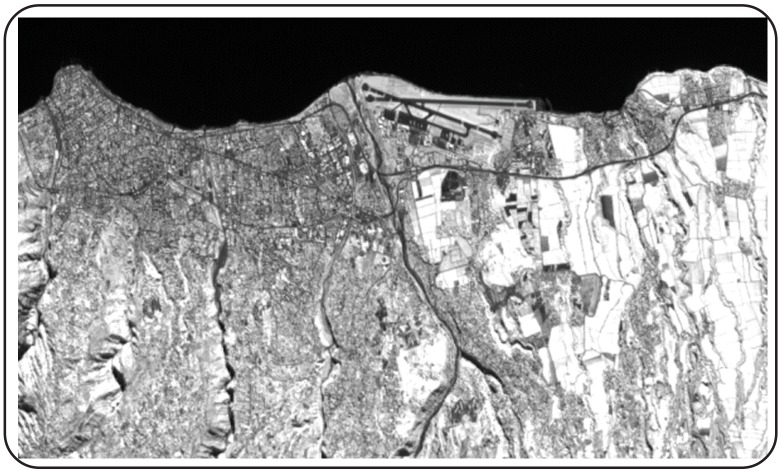

Figure 5. Satellite image acquired on June 09, 2011

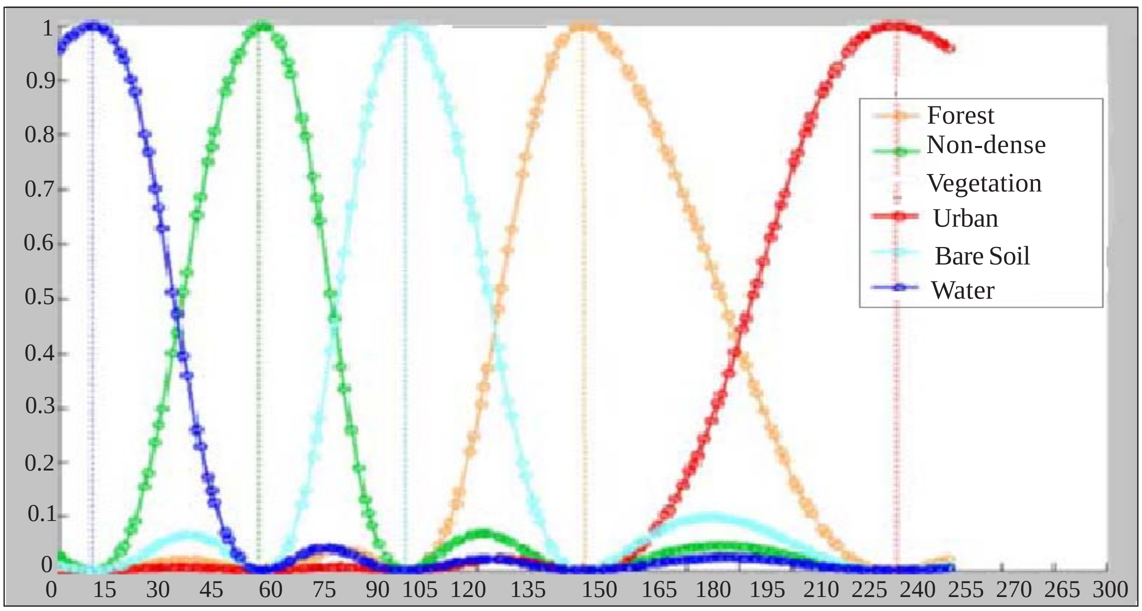

Figure 6. Membership degrees for the attribute mean radiometric

radiometric and the moisture. Then, the probabilistic collocation method can be applied by reducing the complexity of the model and therefore minimizing computational cost.

4.1.4 Probabilistic Collocation Method (PCM)

The PCM approach is composed by seven steps. In this paper, we will consider only the three input attributes identified by the sensitivity analysis process.

Figure 7 depicts results obtained after the application of the algorithm of the maximum likelihood method for the attribute mean texture.

Step 1: Specify Distribution of Uncertain Parameters

The first step is to determine the distribution of the three parameters: mean Gabor, radiometric mean and moisture. This is achieved by applying the algorithm of the maximum likelihood method for the three attributes.

First, we must specify the uncertainty in the parameters. The Gabor mean (I1) has a normal distribution with a mean of 175, and a standard deviation of 15.

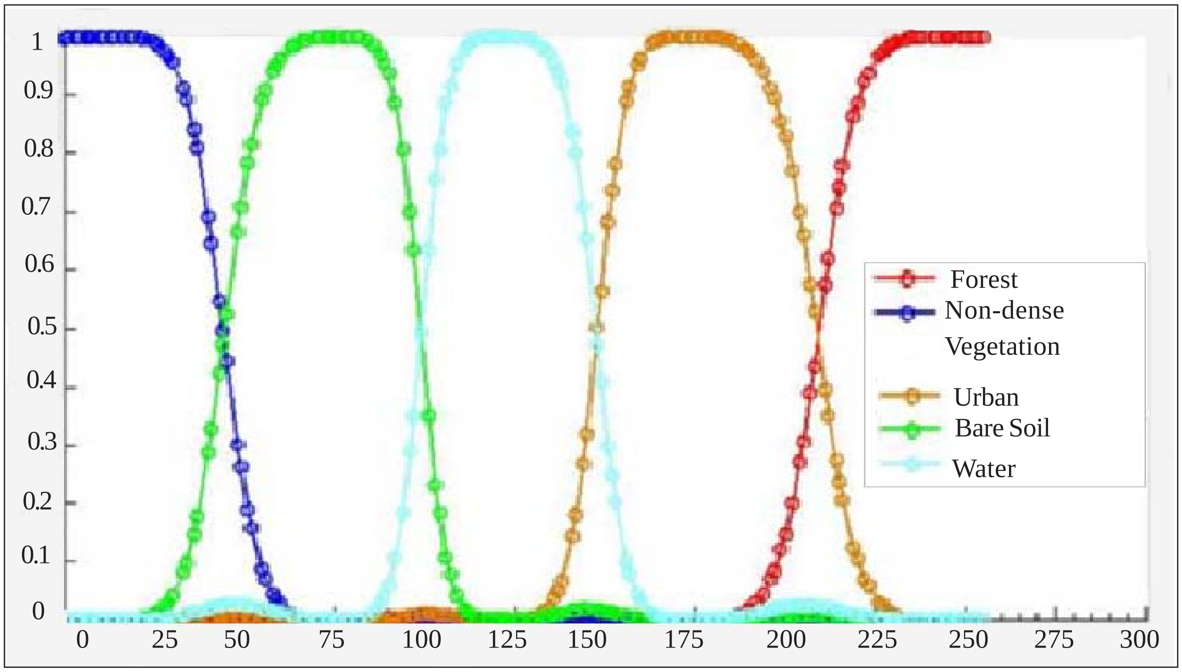

Figure 7. Membership degrees for the attribute mean texture

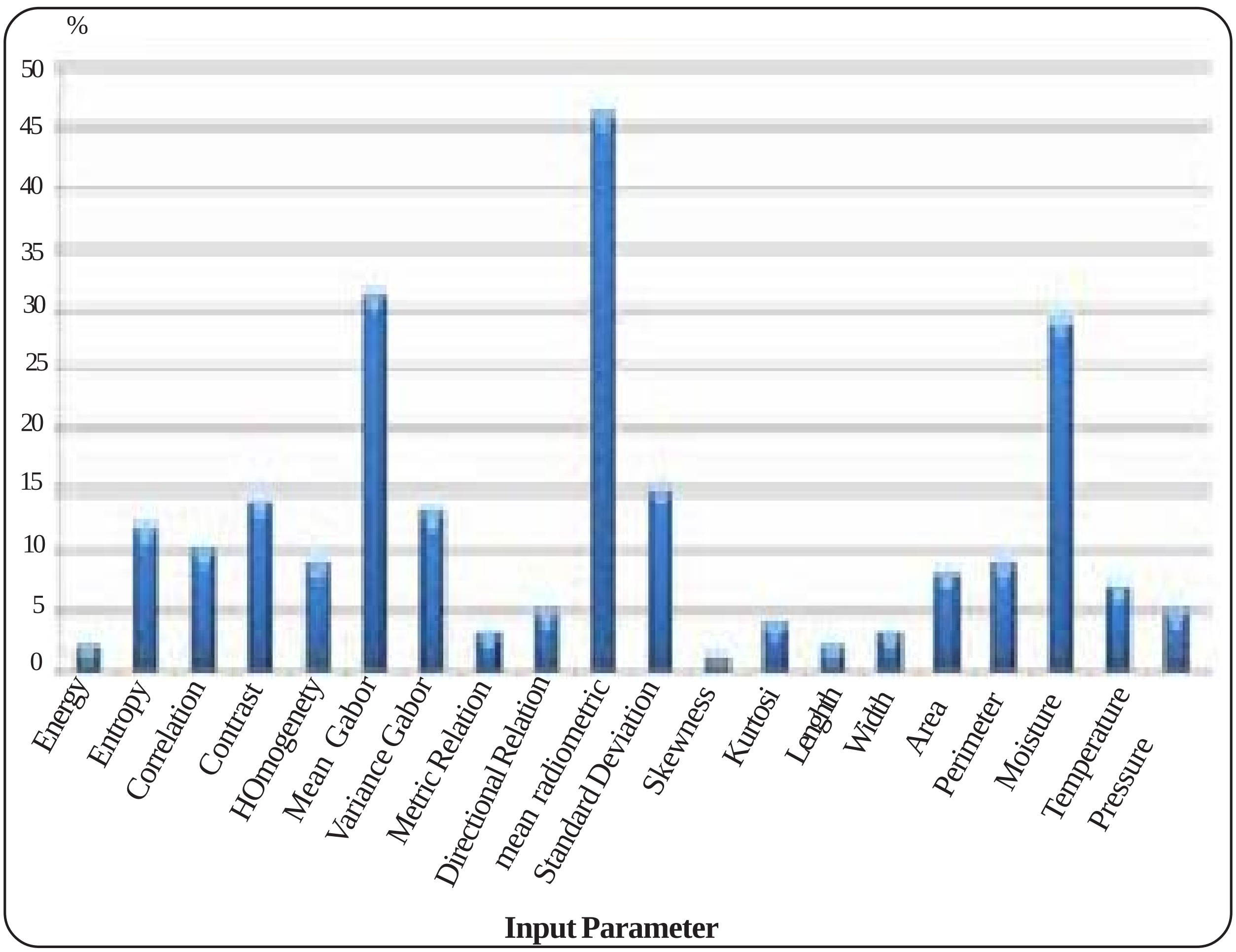

Figure 8. Sensitivity analysis of the input attributes to the model of land cover change prediction

$$N(\mu_1, \sigma_1) = N(175, 15) \tag{14}$$

The radiometric mean (I2) has a normal distribution with a mean of 170, and a standard deviation of 23.

$$N(\mu_2, \sigma_2) = N(175, 23) \tag{15}$$

The moisture (I3) has a normal distribution with a mean of 55, and a standard deviation of 8.

$$N(\mu_3, \sigma_3) = N(55, 8) \tag{16}$$

Step 2: Derive orthogonal polynomial for the distributions
In the second step, derive the set of orthogonal polynomials from distributions obtained in the step1. Since these distributions are Gaussian, the following simplification may be used, [6].

$$X = \mu + \sigma\,(H1(\xi)) \tag{17}$$

Where μ is the mean and σ is the standard deviation of X. $H1\ (\xi)$ is the first order Hermite polynomial. Hermite polynomials are a set of polynomials which are orthogonal to the standard normal distribution ξ, with an average of 0 and a variance of 1. This allows the use of the same set of orthogonal polynomials for all Gaussian distributions, instead of deriving orthogonal polynomials for each specific distribution. Since in our study, equation (18) to (22) describe the first five Hermite polynomials are:

$$H1\,(\xi) = \xi \tag{18}$$

$$H2\,(\xi) = \xi^2 - 1 \tag{19}$$

$$H3\,(\xi) = \xi^3 - 3\,\xi \tag{20}$$

$$H4\,(\xi) = \xi^4 - 6\,\xi^2 + 3 \tag{21}$$

$$H5\,(\xi) = \xi^5 - 10\,\xi^3 + 15\,\xi \tag{22}$$

Step 3: Generate polynomial chaos expansion
In this step, we generate a polynomial expression to represent the output parameter. The third-order approximation used in this case is given by the equation (23):

$$Y \approx Y' = y_0 + y_1\xi_1 + y_2\xi_2 + y_3\xi_3 + y_4(\xi_1^{\,2} - 1) + y_5(\xi_1\xi_2) + y_6\xi_1\xi_3 + y_7\ (\xi_2^{\,2} - 1) + y_8(\xi_2\xi_3) + \ldots. \tag{23}$$

This equation (23) is a development of the equation (10). y_i are the unknowns of the equation (23). To solve these variables, simulation points are needed. These points are called collocation points.

Table 1 describes the development of the equation (10) in the 20th element. This represents the third order polynomial chaos.

Step 4: Generate collocation points for model runs
The goal of applying the probabilistic collocation method is to find a good approximation with a reduced number of simulations. Collocation points are selected from the roots of orthogonal polynomials of next higher order (n + 1) for each uncertain parameter. At order p = 1, the equation (10) becomes:

$$Y \approx Y' = y_0 + y_1\xi_1 + y_2\xi_2 + y_3\xi_3 \ldots. \tag{24}$$

The unknowns in equation (24) are $y0$, $y1$, $y2$ and $y3$. To resolve this equation, we need four collocation points. These points are the roots of the Hermite polynomial of order p +1 = 2, as mentioned in the equation (25):

$$H_2(\xi_i) = 0 \rightarrow (\xi_i^2 - 1) = 0 \rightarrow \xi_i \in \{-1; 1\} \tag{25}$$

The number of available collocation points is always greater than the number of needed collocation points. A method for selecting collocation points is presented in [20]. In our work, we need four collocation points.

Step 5: Run the model to the collocation points

We run the model for each collocation point for the approximation of Y of order p. After saving the values we reruns for the approximation of order $p + 1$.

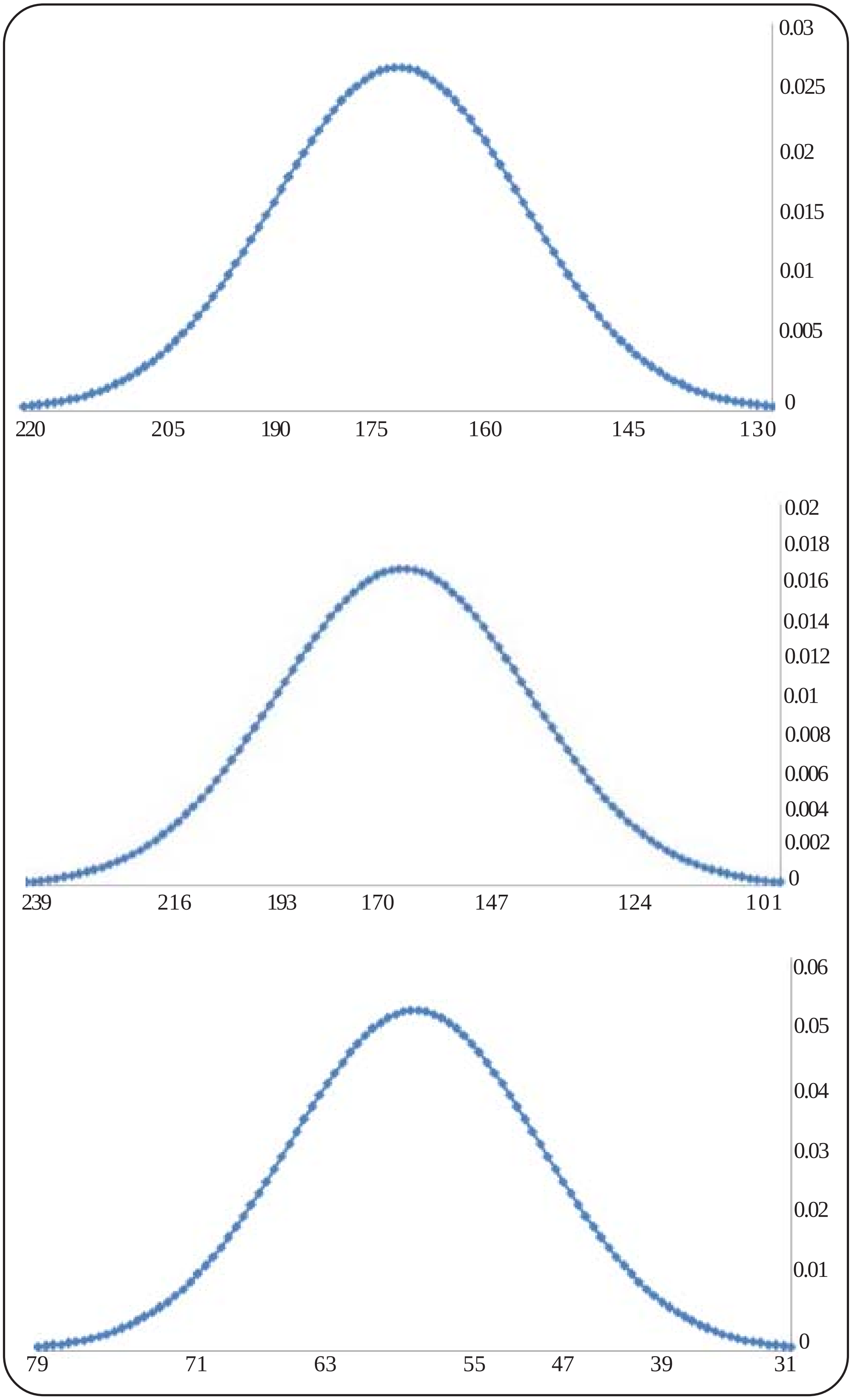

Figure 9.Probability density distribution of parameters Gabor mean (a), radiometric mean (b) and moisture (c) of the object forest

i^{th} **Polynomial Element**	Γp	**Polynomial Chaos Order**
0	1	0
1	ξ_1	1
2	ξ_2	1
3	ξ_3	1
4	$\xi_1^{\,2}-1$	2
5	$\xi_1\xi_2$	2
6	$\xi_1\xi_3$	2
7	$\xi_2^{\,2}-1$	2
8	$\xi_2\xi_3$	2
9	$\xi_3^{\,2}-1$	2
10	$\xi_1(\xi_1^{\,2}-3)$	3
11	$\xi_2(\xi_1^{\,2}-1)$	3
12	$\xi_3(\xi_1^{\,2}-1)$	3
13	$\xi_1(\xi_2^{\,2}-1)$	3
14	$\xi_1\xi_2\xi_3$	3
15	$\xi_1(\xi_3^{\,2}-1)$	3
16	$\xi_2(\xi_2^{\,2}-3)$	3
17	$\xi_3(\xi_2^{\,2}-1)$	3
18	$\xi_2(\xi_3^{\,2}-1)$	3
19	$\xi_3(\xi_3^{\,2}-1)$	3

Table 1. The First 20 Hermite Polynomials Approximation of Y

Step 6: Check the approximation error

Before using the approximation in uncertainty analysis, the quality of the retrieved approximation can be tested. We calculate the error term by the following equation:

$$e_{App} = \frac{\sqrt{\frac{1}{4}\left|\sum_{i=1}^{4}(y_i^2 - y_i^1)\right|}}{\frac{1}{4}\sum_{i=1}^{4} y_i^1} \tag{26}$$

Where Y_i^1 and Y_i^2 are the approximation of Y at the first and the second orders. We found an error of 57.4% for the first order. In order to reduce the error term, we consider a higher order for the approximation.

Step 7: Try a higher order approximation
Figure 10.a describes the convergence of error reduction by increasing the order of the polynomial chaos approximation. Figure 10.b shows the improvement of the reduction of the error rate for an i^{th} order approximation compared to previous approximation of the first order. We note that the important error rate reduction is for the third order. Then, we can consider this order as satisfying order for our PCM approximation.

The next step is to return to the probability distributions for the input attributes. Then, we determine the optimal values of these attributes corresponding to the order 3. These values are finally incorporated into the prediction model presented in [2].

For the urban area, the rate of change between 2007 and 2011 found by the model described in [2] is around 41.75 %. Real rate of change for the urban is about 42.89 %.

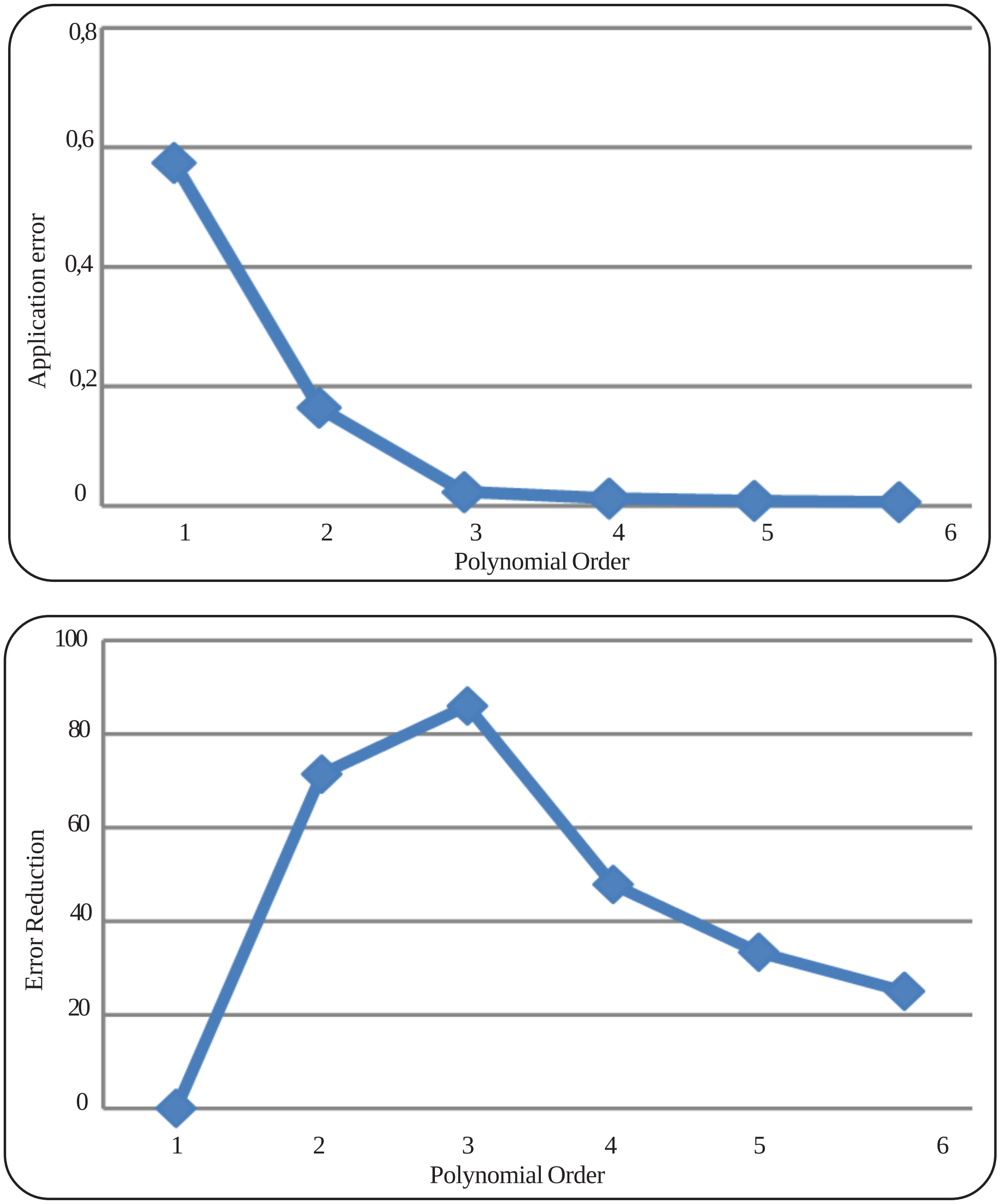

Figure 10.The approximation error variation (a) and the global error reduction (b) according to the order of the polynomial

For the first order, the average of predicting urban changes is 41.12%. Passing to the second order, the prediction rate of urban change is about 42.52%.

For the third order, the prediction rate of urban change is about 42.77 %. For the MC method, the prediction rate of urban change is about 42.68%. We can conclude that increasing the polynomial order allows to improve the prediction of urban changes.

4.2 Evaluation of the Proposed Approach

In order to evaluate the proposed approach in improving land cover change prediction, we apply the proposed propagation

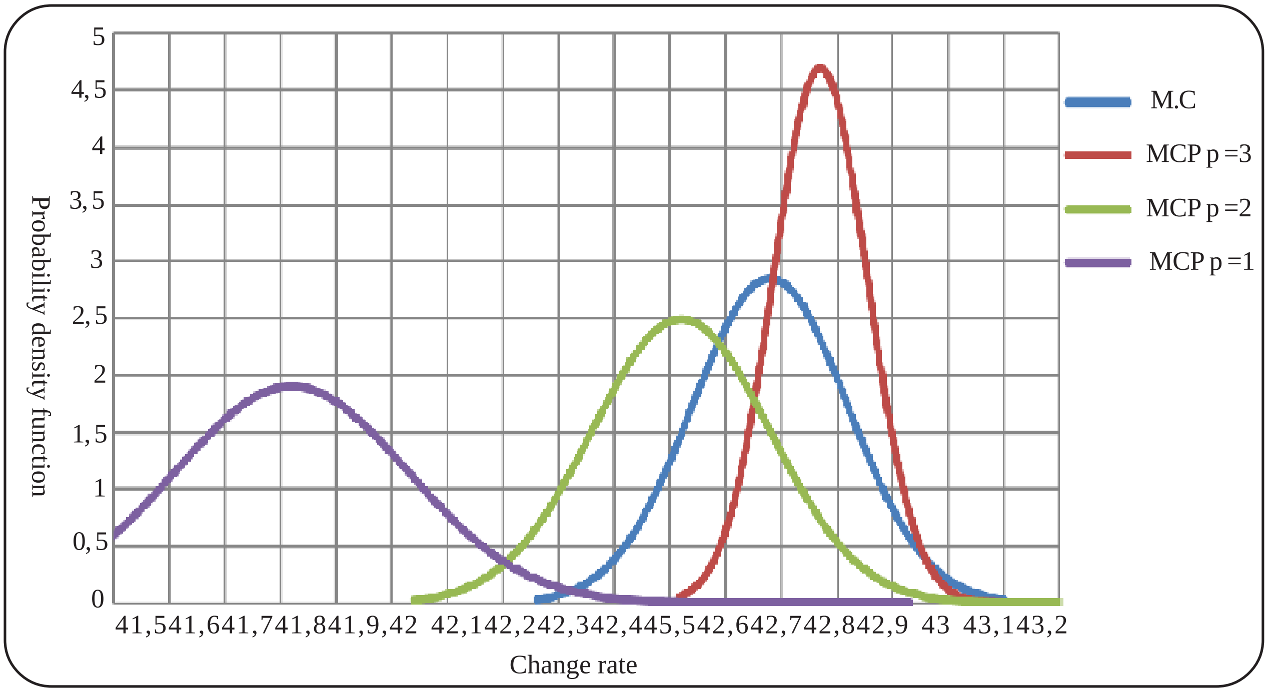

Figure 11. Comparison of prediction rate of urban change for MC and MCP with an order equal to 1, 2 and 3

method and the Monte Carlo method to the prediction model presented in [2]. Then, we compare the proposed prediction changes to the MC ones.

Monte Carlo method is applied in several domains. It is considered as one of the most used methods in literature for uncertainty propagation.

Figure 12 depicts the ground truth image at the date June 09, 2011. Information was extracted by experts over the studied are to construct the ground truth image. Polygons of the studied area of northwestern Reunion Island are digitized to derive the the matic information using a topographic map with the scale of 1/50,000.Topographic information is used to determine the matic classes in the studied area. Five the matic classes are identified which are the following: urban, water, forest, bare soil and non-dense vegetation areas. The ground truth image is used in this paper to compare land cover prediction obtained by the application of the proposed approach and MC method to the model presented in [3].

Approach	**Error for Predicting Urban Change (%)**
Monte Carlo Approach	0.35
Proposed Approach	0.22

Table 2. Error for the Prediction of Urban Changes between MCP With an Order Equal to 3 and Mc Between Dates 2007 and 2011

Table 2 depicts the error calculated between real urban changes, MCP with an order equal to 3 and MC between dates 2007 and 2011. As we note, the proposed approach provides a better results than the MC method in predicting urban changes. This shows the effectiveness of our approach in reducing imperfection related to the prediction process.

Table 3 illustrates percentages of change of the five land cover types (forest, water, bare soil, nondense vegetation and urban). Results show that the propagation of imperfection improves the land cover change prediction compared to the

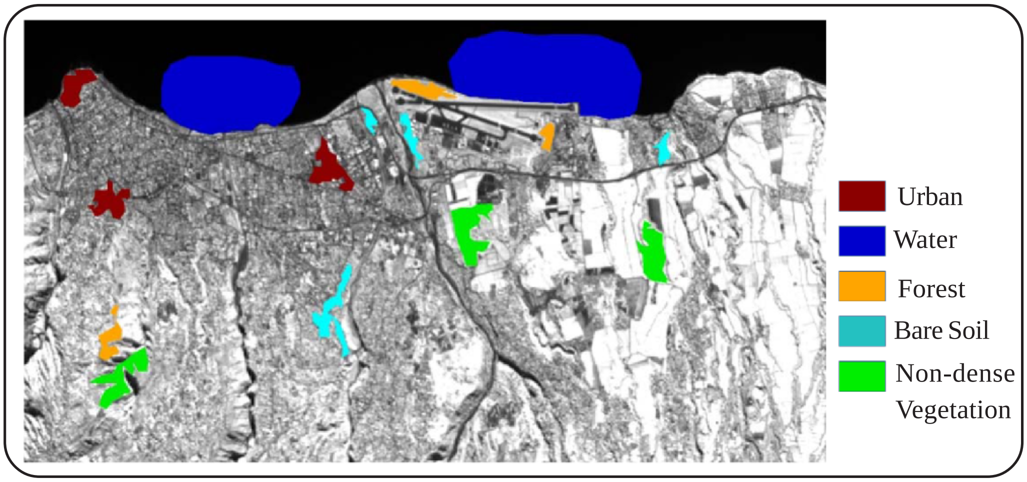

Figure 12. The Ground truth image at the date June 09, 2011

	Forest	Water	Bare Soil	Non Dense Vegetation	Urban
Method Proposed by [3]	70, 98%	2, 95%	66, 03%	32, 03%	41, 75%
Monte Carlo Approach	72, 14%	3, 38%	65, 23%	33, 28%	42, 68%
Proposed Approach	72, 19%	3, 49%	65, 17%	33, 35%	42, 77%
Real	72, 21%	3, 59%	65, 06%	33, 49%	42, 89%

Table 3. Comparison of Percentage of Change Prediction Between 2007and 2011 for the Prediction Model Proposed By [3], Monte Carlo Approach and Proposed Approach.

original model presented in [3]. In addition, we note that the proposed approach outperforms the MC method in predicting land cover changes.

Indeed, in 2011, the ground truth image shows a change of 33.49 % for non-dense vegetation, while the proposed approach predicts a change of 33.35 %. After the treatment of the propagation of uncertainty in the MC method, the prediction is about 33.28 %. This provides a difference between real changes and prediction of changes for the proposed approach in the order of 0.14 %. For the forest object, the ground truth image shows a change of 72.21 %, while the model proposed by [3] predicts a change of 70.98 %. After the treatment of the propagation of uncertainty in MC method, the prediction is about 72.14 %. After the treatment of the propagation of uncertainty in collocation approach, the prediction is about 72.19 %. This provides a difference between real changes and prediction of changes for the proposed approach in the order of 0.02 %. These results confirm the effectiveness of the proposed approach in improving land cover change prediction. This is made by reducing the effect of imperfection related to input variables and their propagation on the model of land cover change prediction.

In order to better evaluate performances of the proposed approach, 20 additional experiments are performed. 20 different periods are considered. Predicted land cover changes for these 20 periods are estimated through the proposed approach and MC method. Then, real urban changes are evaluated based on images representing the same dates in each period.

Table 6 depicts that, over the 20 areas studied, the proposed approach provides best results in 80% of cases compared to the MC method. In addition, the proposed approach provides an average rate of prediction equal to 0.344%. This average is greater than the average rate of prediction given by the MC method which is equal to 0.376%.

In addition to the improvement of the land cover changes prediction, we decide to evaluate the performance of the proposed approach in term of processing time.

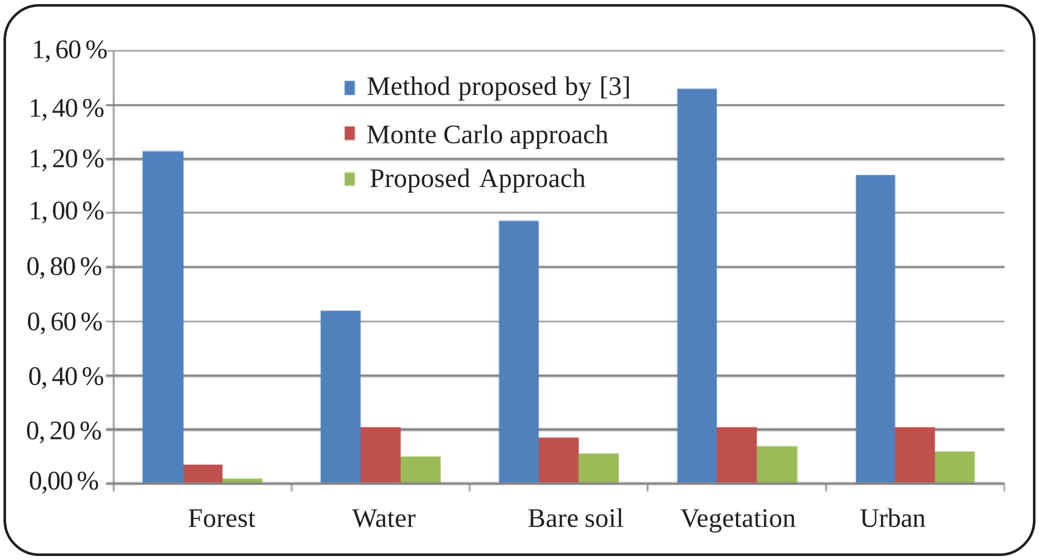

Figure 13.Comparing difference between real changes and prediction of changes for the three methods: method presented in [3], Monte Carlo and proposed approach

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20
M.C	0, 37	0, 37	0, 41	0, 26	0, 33	0, 4	0, 43	0, 39	0, 28	0, 37	0, 36	0, 41	0, 41	0, 43	0, 32	0, 39	0, 45	0, 37	0, 35	0, 42
P.C .M	0, 29	0, 31	0, 35	0, 28	0, 29	0, 36	0, 40	0, 36	0, 31	0, 33	0, 35	0, 37	0, 42	0, 40	0, 34	0, 32	0, 40	0, 33	0, 30	0, 38

Table 4. Comparison of Error Rate for Land Cover Change Prediction between the Proposed Approach and the Mc Method for 20 Period Tests

Table 5 provides a comparison of computational time for the three methods: MC method, approach of collocation without sensitivity analysis (while considering the 20 input attributes) and collocation approach with sensitivity analysis (the proposed approach). The calculations are performed on a Dell i7- 2670QM (2.2 GHz 6MB cache and 6GB of RAM).

Approach	**Polynomial Order**	**Number Total of Work**	**Total Time of Work**
Collocation method with sensitivity analysis N = 3	1	31	32min 8sec
	2	37	32min 20sec
	3	47	32min 40sec
	4	62	33min 10sec
	5	83	33min 52sec
Collocation method without sensitivity analysis N = 20	1	27	54sec
	2	378	756 = 12min 36sec
	3	3654	7308 = 2h 1min 48sec
	4	27405	15h 13min
MC method	5	169911	94h 23min
	-	10000	20000 = 6h 15min
	-	100000	200000 = 55h33min

Table 5. Comparison of Computational Time between Mc Method, Collocation with Sensitivity Analysis Collocation (Proposed Approach) and Collocation without Sensitivity Analysis

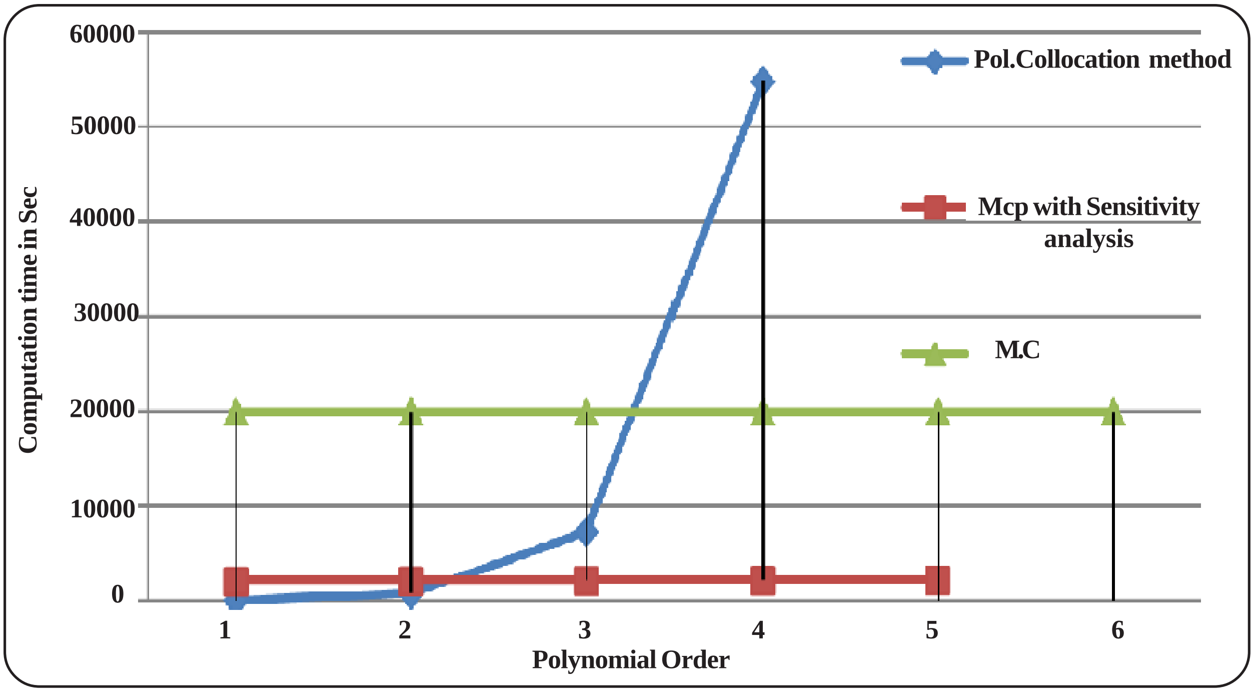

Figure 14. Convergence of computational time for three methods:
MC, Polynomial Collocation method and proposed approach

For an execution with 100000 simulations, 55 hours are needed to apply Monte Carlo method. For the classical collocation method (without applying the sensitivity analysis),we note that the computation time increase significantly from 2 hours to 15 hours to 94 hours when increasing the polynomial order from 3 to 4 to 5.

However, for the proposed approach (collocation method with sensitivity analysis), the computation time is about 32 minutes.

In Figure 14, we plotted the evolution of the computational time of the three methods tested according to the order of the polynomial chaos expansion. We note that the computational time increases dramatically with the order of polynomial chaos.

5. Conclusion

This paper presents a methodology for propagating imperfection throughout a model for land cover change prediction. The methodology is based on computing membership functions for input features for a given land cover type. Then, these membership functions are evaluated through a sensitivity analysis module to identify the most influential features in the overall imperfection of the prediction module. After that, influential features are propagated through the land cover prediction model. To achieve this, we use a probabilistic collocation method. This helps identifying the optimal values of those features that best reduce the overall imperfection. Finally, we take all attributes describing a particular object while considering the optimal values of influential attributes and we introduce them into the model of land cover change prediction. This helps obtaining more reliable decisions about land cover change.

The proposed approach was compared on error prediction and computing time to existing propagation methods. Results shows good performance of the proposed approach.

References

[1] Aksoy, S., Cinbis, R. G. (2010). Image mining using directional spatial constraints, *IEEE Geoscience and Remote Sensing Letters*, 7 (1) 33-37.

[2] Babuska, I., Nobile, F., Tempone, R. (2007). A stochastic collocation method for elliptic partial differential equations with random input data, *SIAM Journal on Numerical Analysis*, 123 (1) 1005-1034.

[3] Boulila, W., Farah, I. R., Ettabaa, K. S., Solaiman, B., Ben Ghézala, H. (2011). A data mining based approach to predict Spatio-temporal changes in satellite images, *International Journal of Applied Earth*, 13 (3) 386-395.

[4] Boulila, W., Farah, I. R., Ettabaa, K. S., Solaiman, B. (2011). Combining decision fusion and uncertainty propagation to improve land cover change prediction in satellite image databases, *Journal of Multimedia Processing and Technologies*, 2, (3) 127-139.

[5] Boulila, W., Farah, I. R., Ettabaa, K. S., Solaiman, B., Ben Ghézala, H. (2010). Spatiotemporal modeling for knowledge discovery in satellite image databases, *CORIA* 2010: *Conference en Recherche d'Information et Applications*, Sousse, p. 35–49.

[6] Darbra, R. M., Eljarrat, D., Barceló, D. (2008). How to measure uncertainties in envioronmental risk assessment, *Trends in Analytical Chemistry*, 27 (4) 377-385.

[7] Ghanem, R. G ., Spanos, P. D. (1991). Stochastic Finite Elements: a Spectral Approach, *Springer Verlag*.

[8] Hubert-Moy, L., Corgne, S., Mercier, G., Solaiman, B. (2002). Land use and land cover change prediction with the theory of evidence: a case study in an intensive agricultural region of France, *Fifth International Conference on Information Fusion*, p. 114-121.

[9] Cherry, M. R., Knopp, J. S., Blodgett, M. P. (2012). Probabilistic collocation method for NDE problems with uncertain parameters with arbitrary distributions, *Review of Progress in Quantitative Nondestructive Evaluation*, 1430, p. 1741-1748.

[10] Lesieutre, B. C., Hockenberry, J. R. (2004). Evaluation of uncertainty in dynamic simulations of power system models: the probabilistic collocation method, *IEEE Transactions on Power Systems*, 19 (3) 1483 - 1491.

[11] Loeven, G. J. A., Witteveen, J. A. S., Bijlz, H. (2007). Probabilistic Collocation: An Efficient Non- Intrusive Approach For Arbitrarily Distributed Parametric Uncertainties, 45th *AIAA Aerospace Sciences Meeting and Exhibit*, 8 - 11 January, Reno, Nevad.

[12] Marrel, A., Iooss, B., Veiga, S., Ribatet, M. (2012). Global sensitivity analysis of stochastic computermodels with joint metamodels, *Statistics and Computing*, 22 (3) 833-847.

[13] Matteoli, S., Veracini, T., Diani, M., Corsini, G .(2013). Models and Methods for Automated Background Density in Hyperspectral Anomaly Detection, *IEEE Transactions on Geosciences and Remote Sensing*, 51 (5) 2837-2852.

[14] Rousson, M., Cremers, D. (2005). Efficient Kernel Density Estimation of Shape and Intensity Priors for Level Set Segmentation, *Medical Image Computing and Computer Assisted Intervention* (MICCAI), Palm Springs, LNCS3750, p.757–764.

[15] Saltelli, A., Ratto, M., Tarantola, S., Campolongo, F. (2006). Sensitivity analysis practices: strategies for model-based inference, *Reliability Engineering & System Safety*, 91 (10-11) 1109- 1125.

[16] Saltelli, A., Ratto, M., Tarantola, S., Campolongo, F. (2000). Sensitivity analysis as an ingredient of modeling, *Statistical Science*, 15 (13) 377-395.

[17] Shen, Z. Y., Hong, Q., Yu, H., Niu, J. F. (2010). Parameter uncertainty analysis of non-point source pollution from 248 different land use types, *Science of the Total Environment*, 48 (8) 1971-1978.

[18] Song, X., Kong, F., Zhan, C., Han, J., Zhang, X. (2013). Parameter identification and global sensitivity analysis of Xin'anjiang model using meta-modeling approach, *Water Science and Engineering*.

[19] Troudi, M., Ghorbel, F. (2012). Optimisation of Smoothing Parameter of Diffeomorphism Kernel Estimate for Bounded Random Data, *SIGMAP*, p. 56-61.

[20] Webster, M., Tatang, M. A., McRae, G. J. (1996). Application of the Probabilistic Collocation Method for an Uncertainty Analysis of a simple Ocean Model, *MIT Joint Program on the Science and Policy of Global Change*, Cambridge.

[21] Wiener, N. (1938). The homogeneous chaos, *American Journal of Mathematics*, 60 (13) 897- 936.

[22] Xiu, D., Karniadakis, G. E. (2002). The Wiener-Askey polynomial chaos for stochastic differential equations, *SIAM J. Sci. Comput.*, 24 (2) 619-644 .

[23] Hosder, S., Walters, R.W., Balch, M. (2007). Efficient sampling for non-intrusive polynomial chaos applications with multiple uncertain input variables, 48th *AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, p. 2946–2961, Honolulu, Hawaii, USA, April.

[24] Eckhardt, R. (1987). Stan Ulam, John von Neumann, and the Monte Carlo method, *Los Alamos Science*, 15, p. 131-137.

[25] Beekhuizen, J., Heuvelink, G. B. M., Biesemans, J., Reusen, I. (2011). Effect of DEM Uncertainty on the Positional Accuracy of Airborne Imagery, *IEEE Transactions on Geoscience and Remote Sensing*, 49 (5) 1567-1577.

[26] Iman, K. L., Helton, J. C. (1988). An investigation of uncertainty and sensitivity analysis techniques for computer models, *Risk Analysis*, 8, p. 71-90.

[27] Malle, A. (1986). A maximum likelihood estimation method for random coefficient regression Models, *Biometrika*, 73, p. 645–656.

[28] Saha, S., Kay, S. M. (2002). Maximum Likelihood Parameter Estimation of Super imposed Chirps Using Monte Carlo Importance Sampling, *IEEE Transactions on Signal Processing*, 50 (2) 224-230.