Cognitive radar applied to forest and kriging-based metamodel

Panagiotis Piteros, Laetitia Thirion-Lefevre, Régis Guinvarc'H, Marc Lambert

COGNITIVE RADAR APPLIED TO FOREST AND KRIGING-BASED METAMODEL

P. Piteros, L. Thirion-Lefevre, R. Guinvarc'h

Supélec/ SONDRA
3 rue Joliot Curie, 91190 Gif-sur-Yvette, France

M. Lambert

L2S (CNRS-Supélec-Univ Paris Sud)/ DRE
3 rue Joliot Curie, 91190 Gif-sur-Yvette, France

ABSTRACT

The idea of applying the cognitive radar principle to the forest characterization is presented. It implies an adaptive design of experiments (DOE) that will allow us to construct approximate models. These models reduce the computational cost, which is mandatory when running many numerical simulations in various configurations.

Index Terms— Cognitive Radar, Forest, Metamodel Adaptive DOE, Kriging, Prediction Uncertainty.

1. INTRODUCTION

1.1. Context

Radar waveforms are usually designed without taking into account the effect of the environment. Cognitive radar is a new framework of radar systems introduced by S. Haykin [1]. His idea is that the waveform emitted by the radar at time $t+1$ will be dependent on the radar return at time t and of the analysis that has been done of this return. A further analysis shows that the emitted waveform $s_e(t)$ and the waveform illuminating the target $s_t(t)$, can actually be quite different for complex environments. As illustrated in Fig. 1, the environment, characterized by its transfer function H, can filter the waveform $s_t(t) = s_e(t) \cdot H(t)$. The signal received by the radar will be altered a second time, in a typical monostatic configuration. Thus, to evaluate these alterations, the transfer function has to be estimated based on our knowledge of the environment. Two applications are considered: target detection and forest parameters retrieval. In the first case, the goal is to compensate for the effect of the forest in order to maximize the intensity backscattered by the target and received by the sensor. For the second application, the objective is to use the capability to modify the radar waveform in order to probe a part of or the whole forest.

In this paper, we first recall the work we performed on a forest of trunks with a kriging metamodel for forest descriptive parameters retrieval. Then we turn to the new technique we propose to implement this work to the case of a forest of branches and trunks and also to support new functions. One of our objectives is to estimate the impact of uncertainties on forest descriptive parameters on the transfer function.

Fig. 1: A radar emits the signal $s_e(t)$ which the forest modifies to $s_t(t)$ before it reaches the target.

1.2. Previous results and limitations

We presented in [2] a new method for forest characteristics inversion based on a functional surrogate model, used instead of COSMO[3] a scattering model dedicated to forest, but too complex to be used for inversion process. This surrogate model was developed to simulate the polarimetric backscattering coefficients as COSMO would do, but faster. However, it can only

handle a few input parameters, because of its structure, leading to a strong simplification of the forest under study. In this particular case, only the trunks were considered for the simulation and their dimensions were derived from the age of the trees using allometric equations [4]. The ground moisture was also taken into account and with the age of the trees, they represented the two input parameters we needed. Considering a set of polarimetric backscattering coefficients obtained for a given radar configuration and derived from the backscattered fields, we addressed the inverse problem of determining the forest age and the ground moisture by using the surrogate model. We produced error maps as in Fig. 2, that illustrate the uncertainty we obtained on the retrieved values of the age and moisture. The combination of the different polarimetric channels allowed us to obtained a very good estimate of these two parameters, despite the low values of the cross-polarization due to the forest description.

Our objective here is first to extend this method to a complete forest of trunks and branches of different sizes and then to take into account additional parameters as the vegetation moisture, that were assumed to be constant in the previous example. Nevertheless, it is not straightforward to achieve this goal, as the number of input parameters is actually strongly limited by the two-level kriging technique we previously implemented. So to overcome this difficulty, we decided to rather apply a scalar technique. This will allow us to consider a more detailed description of the forest, and also new functionalities as the sensitivity analysis and the propagation of uncertainties. This technique is detailed in the next section.

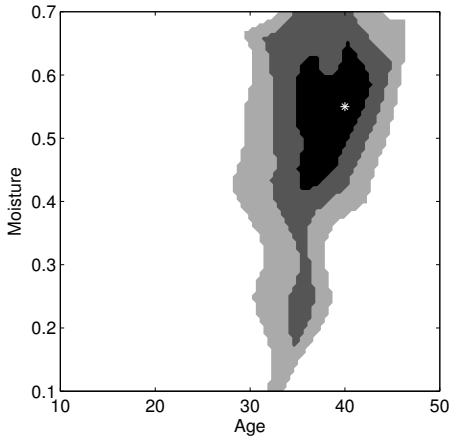

Fig. 2: Example of an error map: the point to retrieve is indicated by a white star at $(30 \text{ years}, 0.55)$. The black area refers to all the retrieved values leading to an error less than 1dB on the backscattering coefficients. In this example, all the polarization channels have been considered. The retrieved age is estimated between 35 and 42 years, and the retrieved volumetric moisture ranges between 0.42 and 0.65.

2. SCALAR KRIGING-BASED METAMODEL

Kriging is a probabilistic method originated from Geostatistics designed for interpolating multidimensional functions that shows certain spatial correlation. Its metamodels are fitted to data that are obtained for large experimental areas, so they can be regarded as global rather than local. They provide a cheap and accurate approximation of the objective function together with an estimation of the potential error. Kriging's main idea is to model the I/O function by a Gaussian random Process (GP) in order to provide a probabilistic framework[1] to account for the uncertainty stemming from the lack of information on the system [5]. The GP is fully characterized by its mean, which is assumed to be constant, and its covariance, with the latter reflecting the correlation among the observation points.

The kriging predictor is the so-called Best Linear Unbiased Predictor (BLUP) of the forward function in the vector space generated by the observations. Its performance depends on both the choice of the input values, where the observation are to be done, and the choice of the covariance. Taking into account the fact that the spatial correlation of the observation points and hence the covariance is unknown, in order to improve the predictor's performance an adaptive sampling procedure will be performed. Also a flexible covariance function will be chosen, so that we can "tune" it to best fit our data. In Fig. 3 we can see a general description of the adaptive sampling procedure that we are going to perform.

[1] J.P.C. Kleijnen's interpretation that the deviations of the simulation output f from its mean μ form a random process can explain why we use a random metamodel for a deterministic simulation model. f is assumed to be a sample path of a second-order Gaussian random process F.

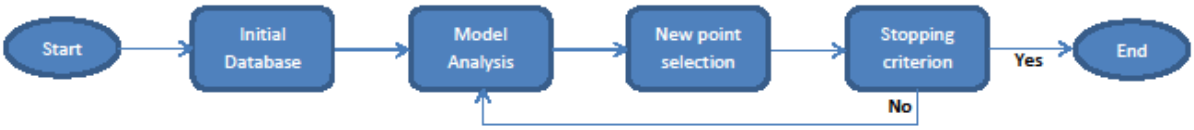

Fig. 3: The general plan of our adaptive sampling procedure.

2.1. Initial Prediction and Estimation of Uncertainty

Kriging is an exact interpolation method, so it needs the results of several simulation runs of the model in order to predict the function under study. The adaptive procedure requires as a first step an initial prediction based on the results of a sample adequately spread in the input space. In order to do that, we will generate a Latin Hypercube (LH) maximin design. The resulting set contains n vectors $\mathbb{X}_n = \{\mathbf{x}_i\}_{i=1}^n$ which is mapped to the vector of the n observations $\mathbf{f}_{\mathbb{X}_n} = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n))^T$.

The next step is to define an appropriate covariance function, which is assumed stationary for computational reasons, that will reflect the spatial correlation between the input samples. The Matérn covariance function is chosen, because it's flexible and it offers the possibility to adjust its regularity and the prediction's smoothness with just a single parameter [5]. This stationary covariance has the form:

$$k(\mathbf{x}_i, \mathbf{x}_j) \equiv k(h) = \frac{\sigma_0^2}{\Gamma(\nu) \cdot 2^{\nu-1}} \cdot \left(2\sqrt{\nu}h\right)^{\nu} \mathcal{K}_{\nu}\left(2\sqrt{\nu}h\right) \tag{1}$$

where h is the inhomogeneous distance between the observation points $\mathbf{x}_i, \mathbf{x}_j$, $\mathcal{K}_{\nu}$ is the modified Bessel function of the second kind, ν is the parameter that controls the regularity of the covariance function and $\sigma_0^2 = k(0)$. Because of the fact that the covariance is unknown, we will estimate its parameters by Restricted Maximum Likelihood Estimation (RMLE) so that they will better fit our model i.e. our observations.

To proceed to the prediction of the I/O function at a point $\mathbf{x} \notin \mathbb{X}_n$ we need to compute the $n \times n$ covariance matrix $\underline{\underline{\mathbf{K}}}$ of the sample space and also the prediction point's covariance vector $\mathbf{k}(\mathbf{x})$:

$$\underline{\underline{\mathbf{K}}} = [k(\mathbf{x}_i, \mathbf{x}_j)], \; i, j = 1, \ldots, n, \qquad \mathbf{k}(\mathbf{x}) = (k(\mathbf{x}, \mathbf{x}_1), \ldots, k(\mathbf{x}, \mathbf{x}_n))^T$$

After these we are ready to compute the so-called kriging coefficients (or weights) $\lambda_i(\mathbf{x})$ and the Lagrange multiplier $\mu(\mathbf{x})$ for the prediction point $\mathbf{x}$ by solving the following linear system (in Lagrangian form)

$$\begin{pmatrix} \underline{\underline{\mathbf{K}}} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{\lambda}(\mathbf{x}) \\ \mu(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \mathbf{k}(\mathbf{x}) \\ 1 \end{pmatrix}$$

The initial prediction for the I/O function is then given by (see [5]):

$$\hat{f}(\mathbf{x}) = \boldsymbol{\lambda}(\mathbf{x})^T \mathbf{f}_{\mathbb{X}_n} \tag{2}$$

The kriging property of exact interpolation results that $\forall \mathbf{x}_i \in \mathbb{X}_n$, it is $\hat{f}(\mathbf{x}_i) = f(\mathbf{x}_i)$ while for all the other $\mathbf{x} \in \mathbb{X}$ it is $\hat{f}(\mathbf{x}) \simeq f(\mathbf{x})$.

A major advantange of the kriging metamodeling procedure is that it provides the prediction together with an estimation of the its uncertainty. This prediction unertainty is given by the variance of the prediction error[2] (assuming a constant and unknown mean) which is: $\hat{\sigma}^2(\mathbf{x}) := \mathbb{E}\left[\left(\hat{f}(\mathbf{x}) - f(\mathbf{x})\right)^2\right] = \mathbb{E}\left[\left(\boldsymbol{\lambda}(\mathbf{x})^T \mathbf{f}_{\mathbb{X}_n} - f(\mathbf{x})\right)^2\right] = k(0) - \lambda(\mathbf{x})^T \mathbf{k}(\mathbf{x}) - \mu(\mathbf{x})$. The disadvantage of that measure is that there is no direct relation between the variance and the input samples, since it's only the covariance function that links them indirectly via the hyperparameters. So, it underestimates the prediction error and it slowly improves our prediction during the adaptive sampling. As an alternative, we can use the jackknife estimation $\hat{\sigma}^2_{\text{jack}}$ of $\hat{\sigma}^2$ to locate the new observation points that have to be added to our database. This estimation method, similar to the Leave One Out Cross-Validation (LOOCV), is based on the evaluation of the *reduced predictions* for every prediction point, by leaving out one observation each time [6]. It is more expensive computationally than the kriging variance, but also more effecient in improving the initial prediction.

[2]The prediction error cannot be evaluated without running simulations at the prediction points and hence its variance is used instead.

2.2. Adaptive Sampling and Final Prediction

When constructing a metamodel, the goal is to generate an input database which will be a kind of a discrete represantation of the I/O function. This can significantly contribute in a better prediction accuracy, but it cannot take place with the traditional fixed sampling designs. The best way to do so, is to adaptively implement the initial sample space. The basic idea behind the adaptive sampling is to locate where should the evaluation of the model be carried out optimally to improve our knowledge on the forward function, based on the previous observations and predictions [3]. This will result in a stepwise uncertainty reduction. The main tool for the uncertainty reduction is the estimation of the variance, whether it is the kriging one or the jackknife. We will choose as a new sample point, the point where the prediction is poor i.e. the estimated variance is high. In order to avoid new points that are very close to the already existing ones[4], we use the Euclidean distance in the input space and we select each new point as follows[5]:

$$\mathbf{x}_{n+1} = \arg\max_{\mathbf{x}\in\mathbb{X}} \left[\left(\min_{i=1,\ldots,n} \|\mathbf{x} - \mathbf{x}_i\| \right) \cdot \hat{\sigma}^2(\mathbf{x}) \right]$$

We continue selecting new points and reducing the prediction uncertainty by iterating the procedure described above, until a terminal criterion is met. The terminal criterion will be the total number of simulations to run at a first stage and then as an extension, a desired level of accuracy. For each new point we have to re-evaluate both the kriging prediction by (2) and its variance. As soon as the optimal input space $\mathbb{X}$ is complete the final prediction will be:

$$\hat{f}(\mathbf{x}) = \boldsymbol{\lambda}(\mathbf{x})^T \mathbf{f}_{\mathbb{X}} \tag{3}$$

3. CONLUSION AND PERSPECTIVES

We proposed to apply a scalar kriging technique to build a new surrogate model of COSMO in order to fast but accurately simulate the backscattering by whole forests. This work relies on a previous study we briefly recalled that shows the interest of developing such a method, in particular in the frame of forest parameters retrieval. However, for this first trial, the description of the forest was too simple to be of great interest for real applications. We were actually restricted by the heavy structure of the functional two-level kriging technique we applied. Here the objective is to fast take into account the effect of the forest when designing a radar waveform for FoPEN for instance. This new surrogate model will have more input parameters and more functionalities. In the frame of cognitive radar, it is necessary to easily and fast propagate the uncertainties on the input parameters. This new technique we are currently implementing has been developed to reach these goals and the first results on a complete forest will be further presented.

4. REFERENCES

[1] Simon Haykin, "Cognitive radar: a way of the future," IEEE Signal Processing Magazine, vol. 23, no. 1, pp. 30–40, 2006.

[2] Andras Vasko, Laetitia Thirion-Lefevre, Sandór Bilicz, Isabelle Champion, Marc Lambert, and Szabolcs Gyimóthy, "Metamodel-based adaptive use of a coherent polarimetric backscattering simulator for the characterization of forested areas at low frequencies," in Progress In Electromagnetics Research Symposium (PIERS 2011), Suzhou, China, Sept. 2011, pp. 818–821, Contribution invitée.

[3] L. Thirion, E. Colin, and C. Dahon, "Capabilities of a forest coherent scattering model applied to radiometry, interferometry, and polarimetry at P- and L-band," IEEE T. Geosci. Remote., vol. 44, pp. 849 – 862, 2006.

[4] K. Saleh, A. Porte, D. Guyon, P. Ferrazzoli, and J.-P. Wigneron, "A forest geometric description of a maritime pine forest suitable for discrete microwave models," IEEE T. Geosci. Remote., vol. 43, pp. 2024 – 2035, 2005.

[5] Julien Villemonteix, Emmanuel Vazquez, and Eric Walter, "An informational approach to the global optimization of expensive-to-evaluate functions," J. of Global Optimization, vol. 44, no. 4, pp. 509–534, Aug. 2009.

[6] Jack P. C. Kleijnen, Design and Analysis of Simulation Experiments, Springer Publishing Company, Incorporated, 1st edition, 2007.

[3] This reflects the main idea of the cognitive radar principle described in [1].

[4] The jackknife variance, because of the leave-one-point-out reduced predictions, can be unusually high close to some of the observation points.

[5] In order to ensure both a high variance and a high minimum distance from the already observed points