Cognitive kriging metamodels for forest characterization and target detection

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

Cognitive kriging metamodels for forest characterization and target detection

Panagiotis Piteros and Laetitia Thirion-Lefevre
and Régis Guinvarc'h
Supélec/ SONDRA
3 rue Joliot Curie, 91190 Gif-sur-Yvette, France
Email: panagiotis.piteros@supelec.fr
laetitia.thirion@supelec.fr
regis.guinvarch@supelec.fr

Marc Lambert
Département de Recherche en Électromagnétisme
Laboratoire des Signaux et Systèmes UMR8506
3 rue Joliot Curie, 91190 Gif-sur-Yvette, France
Email: marc.lambert@lss.supelec.fr

Abstract—The idea of applying the cognitive radar principle to the radar observations of forests is presented. It implies an adaptive design of experiments (DOE) that will allow us to construct probabilistic surrogate models with a known level of uncertainty. These models reduce significantly the computational cost, which is mandatory when running many numerical simulations in various configurations.

I. INTRODUCTION

A. 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. This principle can be useful for forest observations, where a further analysis shows that the emitted waveform $s_e(t)$ and the waveform $s_t(t)$ transmitted through the forest, can actually be quite different. As illustrated in Fig. 1, the environment, characterized by its transfer function H, can filter the waveform that finally reaches the target:

$$s_t(t) = s_e(t) \cdot H(t) \qquad (1)$$

The signal received later 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.

B. 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.
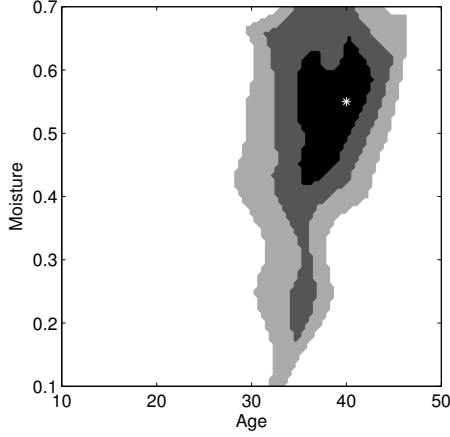

Fig. 2. Example of an error map: the point to retrieve is indicated by a white star at $(30 \text{ years}, 0.55)$. The black area refers to all the retrieved values leading to an error less than 1dB on the backscattering coefficients. In this example, all the polarization channels have been considered.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

II. SCALAR KRIGING-BASED METAMODEL

When dealing with expensive-to-evaluate functions, approximation methods based on probabilistic surrogate models have significant advantages. Mainly they require fewer function evaluations. The surrogate models, also called metamodels, can help us explain the undelying simulation model's behavior and predict the expected output $f(\mathbf{x})$ for combinations of input values $\mathbf{x}$ that have not yet been simulated. In order to make up for the lack of knowledge on the I/O function, the system is treated as a "black-box" [1] and our metamodels are based on Gaussian random Processes (GP). Apart from prediction, they can be used also for validation of the simulation model, Sensitivity Analysis (SA) and optimization [5]. In our work we are going to focus on the so-called *kriging* metamodels.

Kriging is a probabilistic method originated from Geostatistics designed for interpolating multidimensional functions that shows certain spatial correlation [6]. 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.

---

[1] In problems like these, no closed-form formulation or gradient information of the objective function are available. The only information that we have are the measurements of system performance, which can be updated as the metamodel can suggest the location of an additional point at which the objective function should be evaluated.

Kriging's main idea is to model the I/O function by a GP in order to provide a probabilistic framework [2] to account for the uncertainty stemming from the lack of information on the system [7]. 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) [8] 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.
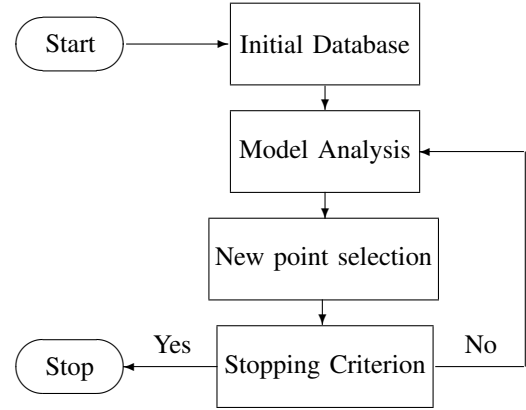

Fig. 3. The general plan of our adaptive sampling procedure.

A. Construction of the Metamodel and Initial Prediction

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. For the construction of this initial sample, a Latin Hypercube (LH) maximin design is chosen. The resulting set contains n vectors $\mathbb{X}_n = \{\mathbf{x}_i\}_{i=1}^n$ which are mapped to the vector of the n simulation results $\mathbf{f}_{\mathbb{X}_n} = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n))^T$. 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

---

[2] According to J.P.C. Kleijnen, the deviations of the simulation output f from its mean μ form a random process. Relying on this interpretation, a random metamodel can be used for a deterministic simulation model and f is assumed to be a sample path of a second-order GP.

smoothness with just a single parameter [7]. As the covariance is assumed stationary, a distance function that will reflect the spatial correlation has to be introduced. This distance has to be anisotropic since the model inputs are of different kind and their domains can vary a lot. We define it as:

$$h_{ij} \equiv h\left(\mathbf{x}_i, \mathbf{x}_j\right) = \sqrt{\sum_{k=1}^{d} \left(\frac{x_{i,k} - x_{j,k}}{\rho_k}\right)^2} \qquad (2)$$

where $x_{i,k}, x_{j,k}$ are the k-th components of the $\mathbf{x}_i$ and $\mathbf{x}_j$ vectors respectively and ρ_k the scaling factor for the k-th component of the input vectors that reflects the sensitivity of the model to that component.

The Matérn stationary covariance has now the form:

$$k\left(h_{ij}\right) = \frac{\sigma_0^2}{\Gamma\left(\nu\right) \cdot 2^{\nu-1}} \cdot \left(2\sqrt{\nu} h_{ij}\right)^{\nu} \mathcal{K}_\nu \left(2\sqrt{\nu} h_{ij}\right) \qquad (3)$$

where $\mathcal{K}_\nu$ is the modified Bessel function of the second kind, ν is the parameter that controls the regularity of the covariance function and $\sigma_0^2 = k\left(0\right)$. Since the covariance is unknown, we have to estimate its parameters ρ_k, σ_0^2 and ν so that they will better fit our model i.e. our observations. This is done by Restricted Maximum Likelihood Estimation (RMLE) [8].

To proceed to the initial 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 of the sample space:

$$\underline{\mathbf{K}} = \left[k\left(\mathbf{x}_i, \mathbf{x}_j\right)\right], \ i, j = 1, \ldots, n,$$

and also the prediction point's covariance vector:

$$\mathbf{k}\left(\mathbf{x}\right) = \left(k\left(\mathbf{x}, \mathbf{x}_1\right), \ldots, k\left(\mathbf{x}, \mathbf{x}_n\right)\right)^T$$

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

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

The prediction for the I/O function is then given by (see [7]):

$$\hat{f}\left(\mathbf{x}\right) = \boldsymbol{\lambda}\left(\mathbf{x}\right)^T \mathbf{f}_{\mathbb{X}_n} \qquad (4)$$

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

B. Estimation of the Uncertainty

A major advantage of the kriging metamodeling procedure is that it provides the prediction together with an estimation of its uncertainty. This prediction unertainty is given by the variance of the prediction error:

$$\varepsilon\left(\mathbf{x}\right) = \left\| \hat{f}\left(\mathbf{x}\right) - f\left(\mathbf{x}\right) \right\| \qquad (5)$$

since this error cannot be evaluated without running simulations at the prediction points. Thus, assuming it has a constant and unknown mean its variance is used instead. This variance

is actually the Mean Squared Error of the prediction given by the formula:

$$\begin{aligned} \mathrm{MSE} = \hat{\sigma}^2\left(\mathbf{x}\right) = & \\ = \mathbb{E}\left[\left(\hat{f}\left(\mathbf{x}\right) - f\left(\mathbf{x}\right)\right)^2\right] = & \\ = \mathbb{E}\left[\left(\boldsymbol{\lambda}\left(\mathbf{x}\right)^T \mathbf{f}_{\mathbb{X}_n} - f\left(\mathbf{x}\right)\right)^2\right] = & \\ = k\left(0\right) - \lambda\left(\mathbf{x}\right)^T \mathbf{k}\left(\mathbf{x}\right) - \mu\left(\mathbf{x}\right) \end{aligned} \qquad (6)$$

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. And since this uncertainty will be our tool in improving the initial prediction its disadvantage can cause an underestimation of the prediction error and a slow improvement for our prediction during the adaptive sampling. As an alternative, we can use the jackknife estimator $\hat{\sigma}_{\mathrm{Jack}}^2$ 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) [3], is based on the evaluation of the *reduced predictions* for every prediction point, by leaving out one observation each time [5]. It is more expensive computationally than the kriging variance, but also more efficient in improving the initial prediction.

Jackknife estimation consists of the following steps:

1) Evaluating the reduced predictions:

$$\hat{f}^{(-i)}\left(\mathbf{x}\right) = \boldsymbol{\lambda}^{(-i)}\left(\mathbf{x}\right)^T \mathbf{f}_{\mathbb{X}_n}^{(-i)}$$

2) Defining and calculating the *pseudovalues*:

$$\tilde{f}^{(-i)}\left(\mathbf{x}\right) = n\hat{f}\left(\mathbf{x}\right) - \left(n-1\right)\hat{f}^{(-i)}\left(\mathbf{x}\right), \quad i = 1, \ldots, n$$

3) Calculating the *modified prediction*:

$$\hat{f}_{\mathrm{Jack}}\left(\mathbf{x}\right) = \frac{1}{n}\sum_{i=1}^{n} \tilde{f}^{(-i)}\left(\mathbf{x}\right)$$

4) Estimating the jackknife variance:

$$\hat{\sigma}_{\mathrm{Jack}}^2\left(\mathbf{x}\right) = \frac{1}{n\left(n-1\right)}\sum_{i=1}^{n}\left(\tilde{f}^{(-i)}\left(\mathbf{x}\right) - \hat{f}_{\mathrm{Jack}}\left(\mathbf{x}\right)\right)^2 \quad (7)$$

where the index $\left(-i\right)$ implies that the i-th observation is being ignored during the calculations.

C. 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 representation 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

---

[3]Jackknife is actually an approximation of the bootstrapping estimation method, but it is preferred as it is easier to compute and its samples are more similar to the original data.

to improve our knowledge on the forward function, based on the previous observations and predictions[4]. 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 (6) or the jackknife (7). 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[5], we use the Euclidean distance in the input space and we select each new point as follows:

$$\mathbf{x}_{n+1} = \arg\max_{\mathbf{x}\in\mathbb{X}} \left[\left(\min_{i=1,\dots,n} \|\mathbf{x} - \mathbf{x}_i\| \right) \cdot \hat{\sigma}^2(\mathbf{x}) \right] \quad (8)$$

This way we ensure both a high variance and a high minimum distance from the already observed points. 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 (4) and its variance. As soon as the optimal input space $\mathbb{X}$ is complete the final prediction will be:

$$\hat{f}(\mathbf{x}) = \boldsymbol{\lambda}(\mathbf{x})^T \mathbf{f}_{\mathbb{X}} \quad (9)$$

D. Example

A simple example of the procedure described above is presented, in order to illustrate the capabilities and the results of the method and also to provide a visualisation of them. We generated a forest consisting only of trunks using COSMO and by setting constant all the forest and radar parameters apart from the frequency. We then simulated the polarimetric backscattering coefficient σ_{pq} of the forest, where $p, q \in \{V, H\}$ define the incident and scattered polarization respectively. We started with an input space consisting of 7 points, 5 chosen by the LH design and 2 on the border of the frequency band[6] and we adaptively added 10 extra points. The result of the initial prediction is presented in Fig. 4 (top), where we can see the forward function, the prediction and its confidence intervals. In order to improve this initial approximation, we added to the database one point each time using (8) exactly where the uncertainty shown in Fig. 4 (bottom) was high. The final prediction shown in Fig. 5 (top) is much closer to the forward function and the confidence intervals are very small. The final absolute error is much smaller than the initial one, reaching a maximum of about 0.25 dB as seen in Fig. 5 (bottom).

During this elementary example, the σ_{pq} coefficients under all polarizations were approximated. The electric field shows many oscillations to be adequately interpolated, while at the

same time all the backscattering coefficients are of a smoother nature. The results for all coefficients were similar and so we have just presented the ones corresponding to the HH polarization. The method can be easily extended to include more inputs for the metamodel. However as the inputs increase, the number of simulation points and the approximation error also increase as expected by the curse of dimensionality.

III. CONCLUSION 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 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.

REFERENCES

[1] S. Haykin, "Cognitive radar: a way of the future," IEEE Signal Processing Magazine, vol. 23, no. 1, pp. 30–40, 2006.
[2] A. Vasko et al., "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.
[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 et al., "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] J.P.C. Kleijnen, Design and Analysis of Simulation Experiments, Springer Publishing Company, Incorporated, 1st edition, 2007.
[6] J.P.C. Kleijnen, "Kriging metamodeling in simulation: A review," Discussion Paper 2007-13, Tilburg University, Center for Economic Research, 2007.
[7] J. Villemonteix, E. Vazquez, and E. 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.
[8] M.L. Stein, Interpolation of Spatial Data: Some Theory for Kriging, Springer Series in Statistics. Springer New York, 1999.

---

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

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

[6]Since kriging is an interpolation method it cannot extrapolate and we have to evaluate the I/O function on the border of the input space also.
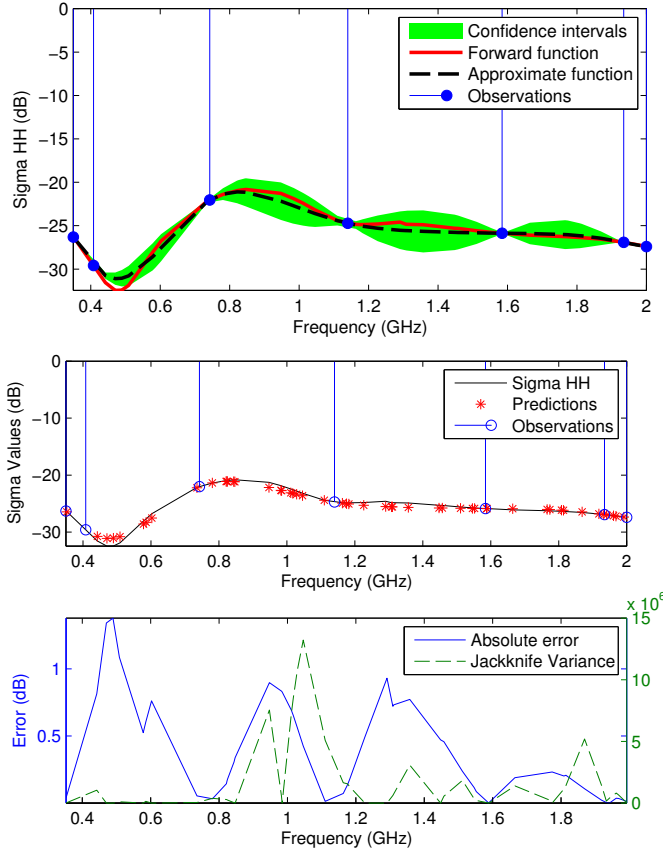
Fig. 4. The initial prediction: (top) with its confidence intervals, (middle) positions of observation points and prediction points (bottom) with the corresponding relative error and Jackknife variance multiplied by the minimum distance as seen in (8).
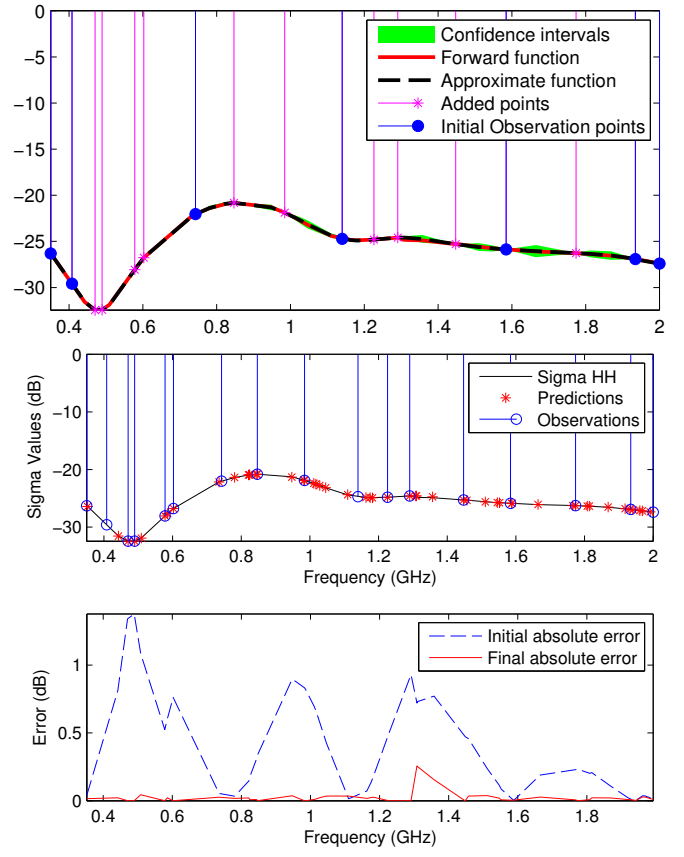
Fig. 5. The final prediction (to be compared with Fig. 4): (top) with the very limited confidence intervals, (middle) positions of observation points and prediction points (bottom) with its corresponding error evaluated for each prediction point before and after the uncertainty reduction.