Low Complexity Spatial Interpolation For Cellular Coverage Analysis

Hajer Braham, Sana Ben Jemaa and Berna Sayrac
Orange Labs
38, rue du Général Leclerc
92130 Issy les moulineaux, France
email: {hajer.braham,sana.benjemaa,
berna.sayrac}@orange.com

Gersende Fort and Eric Moulines
LTCI, Télécom ParisTech & CNRS
46, rue Barrault
75634 Paris cedex 13, France
email:{gersende.fort,eric.moulines}
@telecom-paristech.fr

***Abstract*—During the last decade a lot of effort has been spent on cellular network optimization to improve network capacity and end-user Quality of Service (QoS). Coverage analysis remains as one of the essential topics on which mobile operators still need innovation in terms of performance and cost. Manual coverage analysis is an inefficient and costly task. Radio Environment Maps (REMs) is an efficient coverage analysis solution for present-day cellular networks. REM concept consists of spatially interpolating geo-located measurements to build the whole coverage map using a spatial interpolation technique originating from geo-statistics. Kriging is such a powerful technique which results in high performance in terms of prediction quality. However, this method is costly in terms of computational complexity especially for large datasets: computational complexity of Kriging is $O(n^3)$ where n is the number of measurements. This paper proposes the application of a variant of Kriging, Fixed Rank Kriging (FRK), to coverage analysis in order to reduce the computational complexity of the spatial interpolation while keeping an acceptable prediction error.**

***Keywords*—Radio Environment Maps (REMs), Coverage analysis, Fixed Rank Kriging (FRK), Kriging, Cellular network.**

I. Introduction

Network optimization has always been a major operation for a cellular operator in order to improve the capacity of its networks and the Quality-of-Service (QoS) offered to the end users. Among the network optimization tasks, coverage optimization is the most crucial and fundamental one since it has a determining impact on the perceived QoS. The first step towards an efficient coverage optimization is an accurate coverage analysis. For this reason, planning tools use sophisticated propagation models that take as input, *a priori* knowledge on the terrain profile and on the buildings. These propagation models are then calibrated with field measurements which are obtained through drive tests. Both the acquisition of the *a priori* knowledge on the terrain (which is not even always available) and the drive tests for the collection of field measurements are costly. Minimization of Drive Tests (MDT), a feature introduced by the 3rd Generation Partnership Project (3GPP), in Release 9 [1] consists of collecting geo-located measurements from User Equipments (UEs) and reporting them to the operator (stored in the *MDT server* at the management plane). With the MDT feature, the operator can request geo-located coverage measurements from UEs in a certain geographical area where a coverage analysis is needed.

Radio Environment Map (REM) approach proposed in this paper is intended to be an efficient tool for coverage analysis. The concept of REM was first introduced in [2] as an integrated database for Cognitive Radio systems and one application was on the TV white spaces [3]. The REM of this paper, however, is different from the integrated database of [2] and [3], it consists of building a *coverage map* based on the MDT measurements for an automated coverage analysis [4]. More precisely, the REM applies powerful spatial interpolation techniques (coming from geo-statistics) to the reported geo-located measurements in order to predict the measured quantity metric on the locations where there are no available measured data. The main idea behind this way of predicting unavailable measurements is to benefit from the spatial correlation that exists in the measurement data to build a complete map over a geographical area and *with a given prediction quality*.

In this paper, we focus on the *construction* of the REM based on the Kriging interpolation technique which is known to provide accurate predictions on spatially correlated data [5]. Existing work on REM construction uses Bayesian Kriging applied to coverage hole detection in cellular networks which yields promising results [6]–[9]. It is obvious that the prediction quality and precision of Kriging increase with increasing the number of measurement samples. However, it is also known that the computational complexity of Kriging increases geometrically with number of measurement samples. In particular the computational complexity of Kriging is $O(N^3)$ where N is the number of measurements. This is a considerable disadvantage for Kriging, especially for large datasets. Therefore, in this paper, we propose to benefit from the excellent prediction performance of Kriging without paying the cost of its computational complexity by using a recently developed version of Kriging, called as Fixed Rank Kriging (FRK) [10], [11]. We show that this method enables coverage prediction from massive datasets (consisting of millions of measurement samples) within reasonable computational times. The main contribution of this paper can be summarized as follows: (1) the application of FRK to cellular coverage measurement data and the evaluation the prediction performance and (2) the modification of the original FRK algorithm proposed in [10] to adapt it to our problem, in particular by using the Expectation Maximization (EM) algorithm [12] for fitting the Kriging

978-3-901882-63-0/2014 - Copyright is with IFIP

model parameters to the measurement data. The EM algorithm is a well-known practical procedure in statistical theory for its excellent fitting performance.

The paper is structured as follows: section II introduces the statistical model of the cellular coverage measurement data. Then section III presents the key idea of FRK and details the prediction process together with the estimation of the model parameters. Finally, interpolation results on realistic measurements data are presented in section IV and conclusions are given in section V.

II. Model description

We consider the DownLink (DL) transmission of a cellular radio access network with a given Base Station (BS) transmitter equipped with an omni-directional antenna. Let $Y(s_i)$ denotes the DL received power (in dBm) at location $s_i \in D$ and D is the set of the spatial locations in the considered geographic area. Assuming that the fast fading effects are averaged out by the receivers, $Y(s_i)$ can be expressed as

$$Y(s_i) = \underbrace{p_0 - 10\kappa \log_{10} d_i + \nu(s_i)}_{Z(s_i)} + \varepsilon(s_i), \tag{1}$$

where p_0 denotes the path-loss coefficient expressed in dBm, d_i is the distance between the transmitter antenna and the receiver location $s_i \in D$, $\nu(s_i)$ is the shadowing term (in dB), and $\varepsilon(s_i)$ is the zero-mean additive error term which incorporates the uncertainties of the measurement process and all other random effects due to the propagation environment. $Z(s_i)$ denotes the received power at the location s_i without the measurements error term.

We assume that the power measurements are carried out by a set of N receiving terminals, located at the set of locations $s_1, \ldots, s_N$. Arranging these measurements in a $N \times 1$ column vector $\boldsymbol{Y}$, we obtain the vector-matrix relation

$$\boldsymbol{Y} = \boldsymbol{Z} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\boldsymbol{\varepsilon} = [\varepsilon(s_1) \ldots \varepsilon(s_N)]^T$ and $\boldsymbol{Z} = \boldsymbol{T}\boldsymbol{\alpha} + \boldsymbol{\nu}$ and the terms $\boldsymbol{T}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\nu}$ are given by

$$\boldsymbol{T} = \begin{bmatrix} 1 & -10\log_{10}(d_1) \\ \vdots & \vdots \\ 1 & -10\log_{10}(d_N) \end{bmatrix}, \ \boldsymbol{\alpha} = \begin{bmatrix} p_0 \\ \kappa \end{bmatrix}, \text{ and } \boldsymbol{\nu} = \begin{bmatrix} \nu(s_1) \\ \vdots \\ \nu(s_N) \end{bmatrix}. \tag{3}$$

Here $\boldsymbol{T}$ is a $N \times 2$ deterministic matrix of known functions of the measurement locations, $\boldsymbol{\alpha}$ is a 2×1 parameter vector. It is assumed that the shadowing process $(\nu(s), s \in D)$ is a centred Gaussian process with covariance function $\boldsymbol{C}$ and independent of the noise measurements $(\varepsilon(s), s \in D)$. Note that the noise process is assumed to have a normal distribution $\mathcal{N} = (0, \sigma^2 \boldsymbol{\Xi}(s))$. Therefore, $\boldsymbol{Y}$ is $N \times 1$ multivariate Gaussian vector whose means $T\boldsymbol{\alpha}$ and covariance matrix $\boldsymbol{\Sigma}$ is given by

$$\boldsymbol{\Sigma}_{ij} = \begin{cases} \boldsymbol{C}(s_i, s_j), & \text{if } i \neq j \\ \boldsymbol{C}(s_i, s_i) + \sigma^2 \boldsymbol{\Xi}(s_i), & \text{if } i = j \end{cases} \tag{4}$$

From equation (2), we can model the wireless channel in a statistical manner, where $\boldsymbol{Y}$ can be considered as the sum of a deterministic linear path-loss term and two stochastic terms: shadowing and error. Such a model can be viewed as a Spatial Mixed Effects (SME) model, it was suggested in [6] and was used to construct the REM for cellular coverage purposes.

III. Coverage prediction with Fixed Rank Kriging

A. Prediction

In this section we want to develop the Kriging prediction in terms of the covariance function $\boldsymbol{\Sigma}$. Let s_0 denote the location where we want to predict the coverage metric Z. Kriging prediction aims at minimizing the Mean Squared Error (MSE) between the real and the predicted coverage metrics at s_0. Since $[Z(s_0), \boldsymbol{Y}]^T$ is a Gaussian vector, the minimum mean-square error prediction of $Z(s_0)$ is the conditional expectation of $Z(s_0)$ given $\boldsymbol{Y}$, denoted as $\mathbb{E}[Z(s_0)|\boldsymbol{Y}]$. Thus we have,

$$\mathbb{E}[Z(s_0)|\boldsymbol{Y}] = \underset{Z^*(s_0)}{\operatorname{argmin}} \mathbb{E}\{(Z(s_0) - Z^*(s_0))^2\}, \tag{5}$$

where $Z^*(s_0)$ is the set of possible prediction of Z from $\boldsymbol{Y}$ at the location s_0. The prediction of $Z(s_0)$, denoted as $\hat{Z}(s_0)$, is obtained by minimizing the mean-square error (MSE) of equation (5), yielding [5]:

$$\hat{Z}(s_0) = t^T(s_0)\boldsymbol{\alpha} + \boldsymbol{C}^T(s_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{\alpha}), \tag{6}$$

where $\boldsymbol{t}^T(s_0) = [1 \ ; -10\log_{10}(d_{s_0})]$ and $\boldsymbol{C}(s_0) = [\boldsymbol{C}(s_0, s_1) \ldots \boldsymbol{C}(s_0, s_N)]^T$.

We notice that the use of (6) will necessitate the computation of the $N \times N$ matrix $\boldsymbol{\Sigma}^{-1}$. When N is very large, the inversion of $\boldsymbol{\Sigma}$ is not possible, because it involves a huge computational cost. In order to reduce the computational complexity, we propose to use the Fixed Rank Kriging (FRK) [10]. This method proposes a decomposition of the matrix $\boldsymbol{\Sigma}$ which reduces the dimension of the matrix inversion to $r \times r$, where r is fixed by the user. The basic idea behind the FRK is to capture the scales of spatial dependence through an appropriate set of r basis functions chosen to be located at points $s'_1, \ldots, s'_r$ (the choice of the basis functions is detailed in IV-A). These basis functions are denoted as,

$$\boldsymbol{S}(s) \equiv [S_1(s) \ldots S_r(s)]^T, \qquad s \in D \tag{7}$$

where r is fixed and in practice $r << N$. More precisely, the spatially correlated random (shadowing) process ν is projected onto the r basis functions such that,

$$\nu(s) = \boldsymbol{S}(s)^T \boldsymbol{\eta}, \tag{8}$$

where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_r)^T$ is a random projection coefficient vector modelled as a zero mean process with a covariance matrix denoted as $\boldsymbol{K}$. Thus $\operatorname{cov}\{\nu(s_1), \nu(s_2)\}$ is written as

$$\boldsymbol{C}(s_1, s_2) = \boldsymbol{S}^T(s_1)\boldsymbol{K}\boldsymbol{S}(s_2), \qquad s_1, s_2 \in D. \tag{9}$$

Now, the modified model of $Y(s_i)$ is given by,

$$Y(s_i) = \boldsymbol{t}(s_i)^T\boldsymbol{\alpha} + \boldsymbol{S}(s_i)^T\eta + \varepsilon(s_i), \qquad s_i \in D \tag{10}$$

assuming $\eta(.)$ and $\varepsilon(.)$ are two independent processes. Using (9), we get

$$\begin{aligned} \boldsymbol{\Sigma} &= \boldsymbol{S}\boldsymbol{K}\boldsymbol{S}^T + \sigma^2\boldsymbol{\Xi} \\ &= \sigma\boldsymbol{\Xi}^{1/2}\left(I + \sigma^{-1}\boldsymbol{\Xi}^{-1/2}\boldsymbol{S}\boldsymbol{K}\boldsymbol{S}^T\sigma^{-1}\boldsymbol{\Xi}^{-1/2}\right)\sigma\boldsymbol{\Xi}^{1/2}, \end{aligned} \tag{11}$$

where $\boldsymbol{S}$ is the $N \times r$ matrix whose $(i,l)^{\text{th}}$ element is $S_l(s_i)$. This implies

$$\boldsymbol{\Sigma}^{-1} = \sigma^{-1}\boldsymbol{\Xi}^{-1/2}\{I + (\sigma^{-1}\boldsymbol{\Xi}^{-1/2})\boldsymbol{S}\boldsymbol{K}\boldsymbol{S}^T(\sigma^{-1}\boldsymbol{\Xi}^{-1/2})\Big\}^{-1}\sigma^{-1}\boldsymbol{\Xi}^{-1/2}. \quad (12)$$

Employing the following standard matrix result [13], we get

$$\left(\boldsymbol{I} + \boldsymbol{P}\boldsymbol{K}\boldsymbol{P}^T\right)^{-1} = \boldsymbol{I} - \boldsymbol{P}\left(\boldsymbol{K}^{-1} + \boldsymbol{P}^T\boldsymbol{P}\right)^{-1}\boldsymbol{P}^T, \quad (13)$$

where P and K are respectively $N \times r$ and $r \times r$ matrices such that K and $K^{-1} + P^TP$ are invertibles. Using equation (13) in (12) gives

$$\boldsymbol{\Sigma}^{-1} = \left(\sigma^2\boldsymbol{\Xi}\right)^{-1} - \left(\sigma^2\boldsymbol{\Xi}\right)^{-1}\boldsymbol{S}\left\{\boldsymbol{K}^{-1} + \boldsymbol{S}^T\left(\sigma^2\boldsymbol{\Xi}\right)^{-1}\boldsymbol{S}\right\}^{-1}\boldsymbol{S}^T\left(\sigma^2\boldsymbol{\Xi}\right)^{-1}. \quad (14)$$

In addition, using equation (9) we have $\boldsymbol{C}(s_0)^T = \boldsymbol{S}(s_0)^T\boldsymbol{K}\boldsymbol{S}^T$. Combining this with equation (6) we get

$$\hat{Z}(s_0) = \boldsymbol{t}^T(s_0)\boldsymbol{\alpha} + \boldsymbol{S}(s_0)^T\boldsymbol{K}\boldsymbol{S}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{\alpha}), \quad (15)$$

where $\boldsymbol{\Sigma}^{-1}$ is computed according to equation (14).

As can be noticed by comparing equation (6) with equations (14) and (15), FRK involves the inversion of the matrix $\boldsymbol{K}$ which is $r \times r$ instead of $\boldsymbol{\Sigma}$ which is $N \times N$ where $r << N$. Noting that $\boldsymbol{\Xi}$ is diagonal so that $\boldsymbol{\Xi}^{-1/2}$ is easily computable.

The computational complexity reduction is achieved by projecting the shadowing part of the measurements on a given number r of basis functions. We assume that the resulting projection coefficients, ν, have the same spatial characteristics as the shadowing terms. Since the shadowing can be modelled as a zero-mean Gaussian random variable that is spatially correlated according to an exponential correlation model [14], the projection coefficients are similarly modelled. Thus, the matrix $\boldsymbol{K}$ is an exponential covariance matrix, defined as,

$$K_{i,j} = \frac{1}{\beta}\exp\left(-\frac{\left\|s_i' - s_j'\right\|}{\exp(\phi)}\right), \quad (16)$$

where $\left\|s_i' - s_j'\right\|$ is the Euclidean distance between two locations s_i' and s_j' (more details on the basis functions will be given in the following section). $\frac{1}{\beta}$ and ϕ are the parameters of $\boldsymbol{K}$ which are respectively analogous to the shadowing variance and correlation distance of the exponential shadowing model. Note that we are using $\exp(\phi)$, to be sure that the term in the denominator is always positive. Notice that the FRK (equation (15)) assumes that the model parameters $(\boldsymbol{\alpha}, \sigma^2, \boldsymbol{K})$ are known. Yet, in order to perform the prediction, these parameters have to be estimated from the observation $\boldsymbol{Y}$.

B. Model parameter estimation

A first option to estimate the model parameters is by using method of moments approach, suggested in [10]. When applied to our cellular data, this method has turned out to be inconvenient: the estimation of $\boldsymbol{K}$ does not respect the correlation matrix constraint ($\boldsymbol{K}$ must be positive definite). The solution proposed to this problem in [11] is to 'lift'[1] the eigenvalues of $\boldsymbol{K}$ after estimation, but this approach seems to be inconvenient for our cellular coverage problem. Thus, we have refrained from using the eigenvalue lifting and we propose to use the Expectation Maximization (EM) algorithm [12]. It is a well-known and efficient iterative method which performs Maximum Likelihood Estimation (MLE) of model parameters whose details will be given in the rest of this subsection.

Let's denote $\boldsymbol{\theta} = [\boldsymbol{\alpha}, \sigma^2, \beta, \phi]^T$, as the vector of unknown parameters. Knowing that the log-likelihood expression of the problem with the existing data set (which is called as the *incomplete data*) is computationally non-tractable, the basic idea of the EM algorithm is to associate to the given incomplete data problem, an augmented data set (which is called the *complete data*) with which ML (Maximum log-likelihood) parameter estimation is computationally more tractable. For our case, the log-likelihood with the incomplete data, $L_{\boldsymbol{Y}}(\boldsymbol{\theta}) = \log\Pr(\boldsymbol{Y}|\boldsymbol{\theta})$ has a closed-form expression. But the maximization of this expression is not straightforward and cannot be computed analytically, because we will need to inverse and perform multiplication by the covariance matrix $\boldsymbol{\Sigma}$, which is an $N \times N$ matrix and N is considerably high. Therefore, instead of maximizing directly the incomplete data log-likelihood, we maximize a mean complete data log-likelihood $L_{\boldsymbol{Y},\boldsymbol{\eta}}(\boldsymbol{\theta}) = \log\Pr(\boldsymbol{Y}, \boldsymbol{\eta}|\boldsymbol{\theta})$, which is computationally tractable when the mean over the missing data β is with respect to some adequate expectation (see Eq. (17)).

The EM algorithm proceeds iteratively, where at each iteration, there are two steps called as the Expectation step (E-step) and the Maximization step (M-step). Let $\boldsymbol{\theta}_{(0)}$ be some initial value of $\boldsymbol{\theta}$. Then in the first iteration, the E-step requires the calculation of the expected value of the complete-data log-likelihood with respect to the unknown data $\boldsymbol{\eta}$, given the observed data $\boldsymbol{Y}$ and the current parameter estimates $\boldsymbol{\theta}_{(0)}$, written as:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{(0)}) = \mathbb{E}_{\boldsymbol{Y},\boldsymbol{\theta}_{(0)}}\left[\log\Pr(\boldsymbol{Y}, \eta|\boldsymbol{\theta})\right]. \quad (17)$$

The M-step requires choosing a new $\boldsymbol{\theta}_{(1)}$ that maximizes the quantity $Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{(0)})$ with respect to $\boldsymbol{\theta}$. It is equivalent to saying that $\boldsymbol{\theta}_{(1)}$ satisfies the following property,

$$Q(\boldsymbol{\theta}_{(1)}, \boldsymbol{\theta}_{(0)}) \geqslant Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{(0)}). \quad \forall\boldsymbol{\theta}. \quad (18)$$

The two steps are then carried out again (for the next iteration), replacing $\boldsymbol{\theta}_{(0)}$ with $\boldsymbol{\theta}_{(1)}$. Thus the E- and M-steps are alternated repeatedly until the difference between $\boldsymbol{\theta}_{(l)}$ and $\boldsymbol{\theta}_{(l+1)}$ changes by an arbitrarily small amount which is determined by the user. We propose to perform the M-step in four steps, which consist in updating each component of $\boldsymbol{\theta}$ in turn. This yields the following algorithm where the details can be found in appendices A and B.

E-step. Given the current estimate of $\boldsymbol{\theta}_{(l)}$, calculate $Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{(l)})$

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{(l)}) = \mathbb{E}_{\boldsymbol{Y},\boldsymbol{\theta}_{(l)}}\left[\log\Pr(\boldsymbol{Y}, \eta|\boldsymbol{\theta})\right]. \quad \forall\boldsymbol{\theta} \quad (19)$$

[1] lift means shifting the eigenvalues to be all positive

M-step 1. Calculate $\boldsymbol{\alpha}_{(l+1)}$ by maximizing $Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{(l)})$ with respect to $\boldsymbol{\alpha}$ keeping σ^2, ϕ and β fixed respectively at $\sigma^2_{(l)}$, $\phi_{(l)}$ and $\beta_{(l)}$. This maximization with respect to $\boldsymbol{\alpha}$ can be done analytically and yields the following expression:

$$\boldsymbol{\alpha}_{(l+1)} = \left(\boldsymbol{T}^T\boldsymbol{T}\right)^{-1}\boldsymbol{T}^T\left(\boldsymbol{Y} - \boldsymbol{S}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}[\boldsymbol{\eta}]\right). \quad (20)$$

M-step 2. Calculate $\sigma^2_{(l+1)}$ by maximizing $Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{(l)})$ with respect to σ^2, keeping ϕ, β fixed respectively at $\phi_{(l)}$ and $\beta_{(l)}$ and $\boldsymbol{\alpha}$ fixed at $\boldsymbol{\alpha}_{(l+1)}$. Again, the maximization can be done analytically and yields the following expression:

$$\sigma^2_{(l+1)} = \frac{1}{n}\left\|\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{\alpha}_{(l+1)}\right\|^2 + \frac{1}{n}\mathrm{Tr}(\boldsymbol{S}^T\boldsymbol{S}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}[\boldsymbol{\eta}\boldsymbol{\eta}^T]) - \frac{2}{n}(\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{\alpha}_{(l+1)})^T\boldsymbol{S}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}[\boldsymbol{\eta}]. \quad (21)$$

M-step 3. Calculate $\beta_{(l+1)}$ by maximizing $Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{(l)})$ with respect to β, keeping ϕ fixed at $\phi_{(l)}$ and $\boldsymbol{\alpha}$ and σ^2 fixed respectively at $\boldsymbol{\alpha}_{(l+1)}$ and $\sigma^2_{(l+1)}$. Once more, this maximization can be done analytically and yields the following expression:

$$\beta_{(l+1)} = \frac{r}{\mathrm{Tr}\left(\tilde{\boldsymbol{K}}(\phi)^{-1}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}[\boldsymbol{\eta}\boldsymbol{\eta}^T]\right)}, \quad (22)$$

where $\tilde{\boldsymbol{K}}(\phi)$ is defined in appendix B.

M-step 4. Finally for the parameter ϕ, an analytical solution for maximizing $Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{(l)})$ with respect to ϕ is not possible. Therefore, we use a one-iteration Newton-Raphson method [12]. Details of this method are given in appendix B, where we have considered that $\boldsymbol{\alpha}$, σ^2 and β are fixed respectively at $\boldsymbol{\alpha}_{(l+1)}$, $\sigma^2_{(l+1)}$ and $\beta_{(l+1)}$ for the computation of $\phi_{(l+1)}$.
Note that the quantities $\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}[\boldsymbol{\eta}]$ and $\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}[\boldsymbol{\eta}\boldsymbol{\eta}^T]$ have explicit expressions (see appendix C).

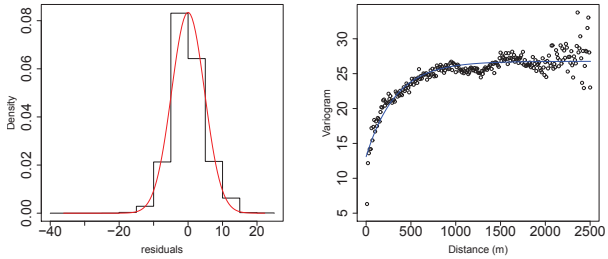
IV. Evaluation results of the coverage prediction algorithm

A. Data set used for evaluation

The geo-located measurements used in this paper are obtained with a very accurate planning tool [15], which uses a sophisticated ray-tracing propagation model developed and used for operational network planning. This tool uses specific propagation model related to supplied environment information (like antenna properties, terrain profile etc.) and it is calibrated through repeated drive tests. Therefore the data produced from the tool is considered as realistic radio measurements reflecting the ground-truth on the coverage situation over the area of interest and can be used in our work.

We consider an urban area located in the southwest of Paris where we construct an LTE coverage map. This map is composed of received pilot powers which are computed at a known location with a 5 m resolution on a surface of 1000 m×1000m (Figure 3a). The environment is covered by a macrocell with an omnidirectional antenna. We have the complete coverage map constructed by the simulator, in a total 40401 measurements are available in a regular square grid. We proceed by obtaining an observed vector which is extracted uniformly over the entire data (considering a given density p of the observed data). Then we will have two datasets, one set chosen randomly to perform model fitting, denoted as the learning set and the rest of the data considered as a second set called test set. We predict the measurements over the locations of the test set and then we compare the obtained prediction to the real value of the measurements that we have already in the test set. The difference between the two values represent the error over it we will build our analysis in the rest of this paper.

To justify our model choice and assumptions done in section II, we compute the histogram of a shadowing part of our measurement data shown in Figure 1a [11]. Fitting a Gaussian curve to this histogram, we obtain a goodness-of-fit metric of $R^2 = 0.9848$. Based on this result, we conclude that our model assumption on the shadowing part is valid. In order to verify the model assumption on the shadowing spatial correlation, we have computed the empirical variogram of the shadowing component, shown in Figure 1b. As we mentioned before, the random process can be modelled as an exponential process. In Figure 1b, the blue line presents an exponential fit to the empirical variogram, with a goodness-of-fit measure of $R^2 = 0.9157$, confirming our model assumption on the shadowing spatial correlation.



(a) Histogram of measurements with a fitted Gaussian distribution (in red).

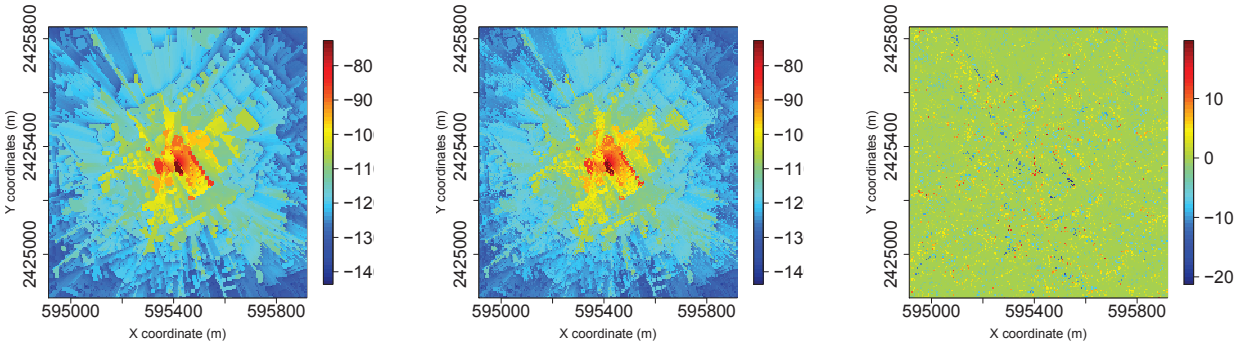
(b) Variogram of measurements and a fitted exponential model (in blue).

Fig. 1. A histogram (left) and a variogram(right) for the detrended data.

Finally, we need to make a choice on the basis function $\boldsymbol{S}$. Since no orthogonality assumptions are made on $\boldsymbol{S}$ and considering our spatial covariance model, we can choose simple bi-square functions also considered in [10],

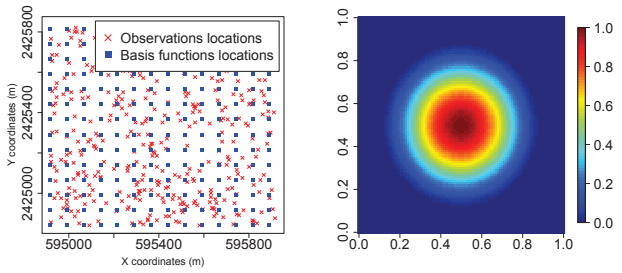
$$S_j(s_i) = \begin{cases} \left[1 - \left(\left\|s_i - s'_j\right\|/r_l\right)^2\right]^2, & \left\|s_i - s'_j\right\| \leqslant r_l \\ 0, & \text{otherwise} \end{cases}$$

where the parameter r_l represents the function expansion and s'_j is the center of the j^{th} basis function. Before computing $\boldsymbol{S}$, we need to fix a grid of center points of the basis functions. It yields to a new discrete grid map depicted in Figure 2a. Figure 2b depicts a two-dimensional view of the used bi-square function. In Figures 3a and 3b, the realistic coverage map, provided by the planing tool, and the REM after FRK prediction are respectively plotted. Figure 3c shows the empirical error map obtained by taking the difference between the real and the predicted measurements.



(a) Real coverage map. (b) Predicted map. (c) Error.

Fig. 3. Real coverage, interpolated and the error maps for 5×5 resolution grid of size $1000\text{m} \times 1000\text{m}$.



(a) Basis functions center points. (b) 2-dimensional bi-square function.

Fig. 2. Basis functions locations on $1000\text{m} \times 1000\text{m}$ map grid (left) and bi-square function in 2D view (right).

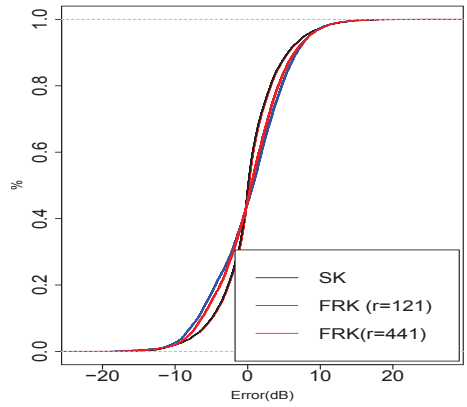


Fig. 4. CDF of the error in dB for a map size $500\text{m} \times 500\text{m}$.

B. Computational complexity evaluation

In this section, we present the computational gain provided by FRK. FRK is a low-complexity variant of the simple Kriging, where we reduce the rank of the covariance matrix to simplify its inversion. In fact, if we use simple Kriging, we need to evaluate equation (6), which requires the inversion of the $N \times N$ covariance matrix $\boldsymbol{\Sigma}$, an operation with a computational complexity of $O(N^3)$. As our coverage measurement samples are typically 'massive' (in the order of thousands to millions), this inversion operation rapidly becomes intractable. FRK, on the other hand, relies on equations (15) and (14) where the computational complexity is $O(nr^2)$. Thus, FRK reduces the computational complexity from $O(N^3)$ to $O(Nr^2)$ with $r << N$. Nevertheless, this technique induces a performance degradation as shown in Figure 4. In this figure, we consider a reduced map size ($500\text{m} \times 500\text{m}$), where we compare the cdf of the prediction error using simple Kriging and FRK, with $N = 2000$, $r = 121$ and $r = 441$. One can see that the use of the FRK interpolation decreases slightly the prediction quality.

C. Prediction evaluation

In this section we evaluate the FRK prediction quality. We start by presenting the assumptions made on the EM algorithm to ensure its convergence. We choose the OLS initialisation for $\boldsymbol{\alpha}$:

$$\boldsymbol{\alpha}_0 = \left(\boldsymbol{T}^T\boldsymbol{T}\right)^{-1}\boldsymbol{T}^T\boldsymbol{Y}, \tag{23}$$

The initial values of σ^2, β and ϕ are chosen randomly with the constraint that $\beta, \sigma^2 > 0$. We consider the following convergence condition: $\left\|\boldsymbol{\theta}_{(l)} - \boldsymbol{\theta}_{(l-1)}\right\| < \zeta$, where $\boldsymbol{\theta}_{(l)} = [\boldsymbol{\alpha}_{(l)}, \sigma^2_{(l)}, \phi_{(l)}, \beta_{(l)}]^T$ is the parameter vector at the l^{th} iteration and $\zeta = 10^{-5}$. It is also important to fix the parameters of the chosen basis functions. For the bi-square functions detailed in section IV-A, we fix the expansion parameter r_l to be to be equal to the distance separating two neighbouring basis functions.

In Figure 5, we show the box plot of the prediction error, for several choices of r. It can be noticed that the prediction quality increases with increasing the number of basis functions. But based on section IV-B, we know that the computational complexity is $O(Nr^2)$ and $r << N$. As a consequence the choice of the parameter r defines the trade-off between computational complexity and prediction quality. Figure 6 presents the prediction error for different numbers of available measurements, considering the same rank $r = 1200$. One can see that when the number of available measurements is much

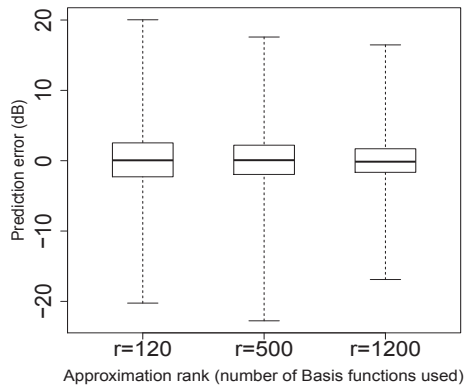


Fig. 5. Influence of the number of basis functions on the prediction error with a fixed number of observed measurements 20000 ($p = 50\%$).

higher than r ($>= 6000$), we do not obtain any additional gain on the prediction quality by increasing the number of measurements. This is due to the fact that fixing the rank by using basis functions with truncated shapes results in decreasing the spatial correlation impact. Notice from Figures 5 and

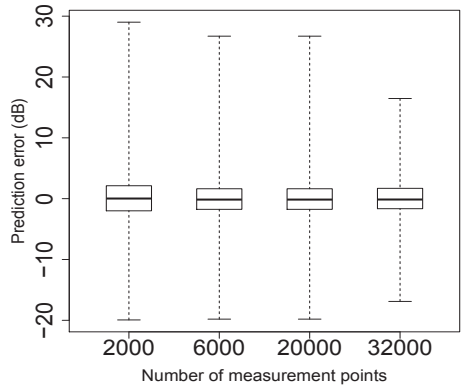


Fig. 6. Influence of the number observed of measurements on the prediction error with a fixed number basis functions ($r = 1200$).

6, that the prediction errors involved in FRK are considerably low (variance in the order of 1-3 dB) when compared to, for example, errors in typical propagation models (RMSE between 10-50 dB) with the exception of ray-tracing models that have substantially high computational complexities [16].

V. Conclusion

In this paper we have studied the performance of the Fixed Rank Kriging (FRK) algorithm applied to coverage analysis in cellular networks. This method is attractive when performing prediction using massive data sets (order of thousands and higher) as it offers a good trade-off between prediction quality and computational complexity compared to classical Kriging techniques. We have also adapted the FRK algorithm to the considered radio measurements by introducing the EM algorithm to estimate the model parameters. To evaluate the performance of our prediction and estimation method, we have analysed the prediction error for different numbers of available measurements and for different values of r, r being the "rank" corresponding to the number of basis functions used in FRK. We have shown that the choice of r impacts the prediction quality, and that we obtain *reasonable* prediction errors with a relatively low computational complexity (compared to typical errors obtained by existing propagation models, except ray-tracing models which have very high computational complexities). We have also shown that once r is fixed, limited gain is obtained by this technique when we increase the number of measurements. This study has been performed using field-like measurements obtained from an accurate and realistic planning tool which uses a ray-tracing propagation model. The next step of this work consists of applying our algorithm to real field measurements.

References

[1] 3GPP TR 36.805 v1.3.0 1, "Study on minimization of drive-tests in next generation networks; (release 9)," tech. rep., 3rd Generation Partnership Project, 2009.

[2] B. A. Fette, *Cognitive radio technology (Communications Engineering)*. Newnes, 2006.

[3] Y. Zhao, B. Le, and J. H. Reed, "Network support-the radio environment map," *Cognitive radio technology*, pp. 325–366, 2006.

[4] "Flexible and spectrum aware radio access through measurements and modelling in cognitive radio systems," Tech. Rep. 73, FARAMIR Deliverable D4.1, 2011.

[5] N. A. Cressie, "Statistics for spatial data," 1993.

[6] B. Sayrac, A. Galindo-Serrano, S. B. Jemaa, J. Riihijärvi, and P. Mähönen, "Bayesian spatial interpolation as an emerging cognitive radio application for coverage analysis in cellular networks," *Trans on Emerging Telecommunications Technologies*, 2013.

[7] S. Grimoud, B. Sayrac, S. Ben Jemaa, and E. Moulines, "Best sensor selection for an iterative rem construction," in *VTC Fall*, pp. 1–5, IEEE, 2011.

[8] B. Sayrac, J. Riihijärvi, P. Mähönen, S. Ben Jemaa, E. Moulines, and S. Grimoud, "Improving coverage estimation for cellular networks with spatial bayesian prediction based on measurements," in *SIGCOMM*, pp. 43–48, 2012.

[9] S. Grimoud, S. Ben Jemaa, B. Sayrac, and E. Moulines, "A rem enabled soft frequency reuse scheme," in *GLOBECOM Workshops*, pp. 819–823, 2010.

[10] N. Cressie and G. Johannesson, "Fixed rank kriging for very large spatial data sets," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 70, no. 1, pp. 209–226, 2008.

[11] M. Katzfuss and N. Cressie, "Tutorial on fixed rank kriging (frk) of co2 data," 2011.

[12] J. M. Geoffrey and K. Thriyambakam, *The EM algorithm and extensions*, vol. 382. Wiley-Interscience, 2007.

[13] H. V. Henderson and S. R. Searle, "On deriving the inverse of a sum of matrices," *Siam Review*, vol. 23, no. 1, pp. 53–60, 1981.

[14] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electronics letters*, vol. 27, no. 23, pp. 2145–2146, 1991.

[15] "Asset tool." http://www.aircominternational.com/Products/Planning/asset.aspx.

[16] C. Phillips, D. Sicker, and D. Grunwald, "Bounding the error of path loss models," in *DySPAN*, pp. 71–82, 2011.

[17] P. Billingsley, *Probability and measure*, vol. 939. Wiley, 2012.

[18] G. A. Seber and A. J. Lee, *Linear regression analysis*, vol. 936. John Wiley & Sons, 2012.

[19] A. Galindo-Serrano, B. Sayrac, S. Ben Jemaa, J. Riihijarvi, and P. Mahonen, "Automated coverage hole detection for cellular networks using radio environment maps," in *WiOpt*, pp. 35–40, 2013.

[20] T. Cai, J. van de Beek, B. Sayrac, S. Grimoud, J. Nasreddine, J. Riihijarvi, and P. Mahonen, "Design of layered radio environment maps for ran optimization in heterogeneous lte systems," in *PIMRC*, pp. 172–176, 2011.

[21] K. B. Petersen, M. S. Pedersen, J. Larsen, K. Strimmer, L. Christiansen, K. Hansen, L. He, L. Thibaut, M. Baro, S. Hattinger, V. Sima, and W. The, "The matrix cookbook," tech. rep., 2006.

APPENDIX A
E STEP

Our model is

$$\boldsymbol{Y} = \boldsymbol{T}\boldsymbol{\alpha} + \boldsymbol{S}\boldsymbol{\eta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$ play the role of random-effect process. Since $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$ are two independent process, we have

$$\boldsymbol{Y}|\boldsymbol{\eta},\boldsymbol{\theta} \sim \mathrm{N}\left(\boldsymbol{T}\boldsymbol{\alpha} + \boldsymbol{S}\boldsymbol{\eta}, \sigma^2\boldsymbol{\Xi}\right). \tag{24}$$

In addition,

$$\boldsymbol{\eta}|\boldsymbol{\theta} \sim \mathrm{N}\left(0, \boldsymbol{K}\right), \tag{25}$$

where $\boldsymbol{\theta} = [\boldsymbol{\alpha}, \sigma^2, \phi, \beta]^T$ and the matrix $\boldsymbol{K}$ is defined in equation (16). Then, the EM Q-function is written as

$$\begin{aligned} Q(\boldsymbol{\theta},\boldsymbol{\theta}_{(l)}) &= \mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\ln \Pr(\boldsymbol{Y},\boldsymbol{\eta}|\boldsymbol{\theta})\right], \\ &= \mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\ln(\Pr(\boldsymbol{Y}|\boldsymbol{\eta},\boldsymbol{\theta})\Pr(\boldsymbol{\eta}|\boldsymbol{\theta}))\right], \end{aligned} \tag{26}$$

where under the expectation $\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}$, the distribution of η is a Gaussian distribution with mean and covariance matrix given in appendix C. Using (24) and (25), we have

$$\begin{aligned} &\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\ln \Pr(\boldsymbol{Y}|\boldsymbol{\eta},\boldsymbol{\theta})\right] \\ &= -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\left\|\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{\alpha} - \boldsymbol{S}\boldsymbol{\eta}\right\|^2\right] \end{aligned} \tag{27}$$

and

$$\begin{aligned} &\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\Pr(\boldsymbol{\eta}|\boldsymbol{\theta})\right] \\ &= -\frac{r}{2}\ln(2\pi) - \frac{1}{2}\ln(\det(\boldsymbol{K})) - \frac{1}{2}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\boldsymbol{\eta}^T\boldsymbol{K}^{-1}\boldsymbol{\eta}\right] \end{aligned} \tag{28}$$

Combining (27) and (28) in expression (26) gives

$$\begin{aligned} Q(\boldsymbol{\theta},\boldsymbol{\theta}_{(l)}) = &-\frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\left\|\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{\alpha} - \boldsymbol{S}\boldsymbol{\eta}\right\|^2\right] \\ &- \frac{1}{2}\ln(\det(\boldsymbol{K})) - \frac{1}{2}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\boldsymbol{\eta}^T\boldsymbol{K}^{-1}\boldsymbol{\eta}\right] + c, \end{aligned} \tag{29}$$

where c is a constant, independent of $\boldsymbol{\theta}$. Developing the square term in expression (29), we get:

$$\begin{aligned} Q(\boldsymbol{\theta},\boldsymbol{\theta}_{(l)}) = &-\frac{n}{2}\ln(\sigma^2) - \frac{1}{2}\ln(\det(\boldsymbol{K})) - \frac{1}{2\sigma^2}\left\|\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{\alpha}\right\|^2 \\ &+ \frac{1}{\sigma^2}(\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{\alpha})^T\boldsymbol{S}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\boldsymbol{\eta}\right] \\ &- \frac{1}{2}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\boldsymbol{\eta}^T\left(\frac{\boldsymbol{S}^T\boldsymbol{S}}{\sigma^2} + \boldsymbol{K}^{-1}\right)\boldsymbol{\eta}\right] + c. \end{aligned}$$

We introduce one of the matrix expectation properties (See [18]),

$$\begin{aligned} \mathbb{E}\left[\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}\right] = &\,\mathrm{Tr}(\boldsymbol{A}\mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^T\right]) \\ &+ \mathbb{E}[\boldsymbol{X}]^T\boldsymbol{A}\mathbb{E}[\boldsymbol{X}], \end{aligned} \tag{30}$$

assuming that $\boldsymbol{A}$ is a symmetric matrix and $\boldsymbol{X}$ is a vector of random variables. Applying this property, equation (26) takes the form:

$$\begin{aligned} Q(\boldsymbol{\theta},\boldsymbol{\theta}_{(l)}) = &-\frac{n}{2}\ln(\sigma^2) - \frac{1}{2}\ln(\det(\boldsymbol{K})) - \frac{1}{2\sigma^2}\left\|\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{\alpha}\right\|^2 \\ &- \frac{1}{2}\mathrm{Tr}\left(\left(\frac{\boldsymbol{S}^T\boldsymbol{S}}{\sigma^2} + \boldsymbol{K}^{-1}\right)\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\boldsymbol{\eta}\boldsymbol{\eta}^T\right]\right) \\ &+ \frac{1}{\sigma^2}(\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{\alpha})^T\boldsymbol{S}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\boldsymbol{\eta}\right]. \end{aligned} \tag{31}$$

APPENDIX B
M-STEP

In the M step, we need to compute the update $\boldsymbol{\theta}_{(l+1)}$ which is under the following constraint,

$$Q(\boldsymbol{\theta}_{(l+1)},\boldsymbol{\theta}_{(l)}) \geqslant Q(\boldsymbol{\theta}_{(l)},\boldsymbol{\theta}_{(l)}). \tag{32}$$

Then $\boldsymbol{\theta}_{(l+1)}$ can be any value that increases $Q(\boldsymbol{\theta},\boldsymbol{\theta}_{(l)})$. If the maximum of $Q(\boldsymbol{\theta},\boldsymbol{\theta}_{(l)})$ has a close form the update of $\boldsymbol{\theta}_{(l+1)}$ can be given by

$$\boldsymbol{\theta}_{(l+1)} = \underset{\boldsymbol{\theta}}{\mathrm{argmax}}\, Q(\boldsymbol{\theta},\boldsymbol{\theta}_{(l)}). \tag{33}$$

To maximize $\boldsymbol{\theta} \mapsto Q(\boldsymbol{\theta},\boldsymbol{\theta}_{(l)})$, we can maximize terms containing each parameter ($\boldsymbol{\alpha}$, σ^2,$\boldsymbol{K}$). In fact, we start by computing first derivative with respect to each parameter independently and then set the obtained derivation equation to zero. The solution of this equation can be a minimum, a maximum, or an inflection point. Therefore, we proceed by computing the second order derivative with respect to the given parameter evaluated for the obtained solution and check that gives a negative number.

a) Update $\boldsymbol{\alpha}$: Deriving equation (31) with respect to $\boldsymbol{\alpha}$ gives

$$\partial_{\boldsymbol{\alpha}}Q(\boldsymbol{\theta},\boldsymbol{\theta}_{(l)}) = \frac{1}{\sigma^2}\boldsymbol{T}^T(\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{\alpha}) - \frac{1}{\sigma^2}\boldsymbol{T}^T\boldsymbol{S}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\boldsymbol{\eta}\right]. \tag{34}$$

We can easily compute $\boldsymbol{\alpha}_{new}$ which is the solution the equation (34) setted equal to zero. It is given by,

$$\boldsymbol{\alpha}_{new} = \left(\boldsymbol{T}^T\boldsymbol{T}\right)^{-1}\boldsymbol{T}^T\left(\boldsymbol{Y} - \boldsymbol{S}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\boldsymbol{\eta}\right]\right). \tag{35}$$

To ensure that the solution of $\partial_{\boldsymbol{\alpha}}Q(\boldsymbol{\theta},\boldsymbol{\theta}_{(l)}) = 0$ is a maximum, we compute the second order derivative of $Q(\boldsymbol{\theta},\boldsymbol{\theta}_{(l)})$ with respect to $\boldsymbol{\alpha}$ and check it is negative. This gives

$$\partial^2_{\boldsymbol{\alpha}}Q(\boldsymbol{\theta},\boldsymbol{\theta}_{(l)})\big|_{\boldsymbol{\theta}=(\boldsymbol{\alpha}_{new},\sigma^2,\boldsymbol{K})} = -\frac{1}{\sigma^2}\boldsymbol{T}^T\boldsymbol{T} < 0, \qquad \forall\sigma^2, \forall\boldsymbol{K}$$

As consequence, the computed update $\boldsymbol{\alpha}_{new}$ satisfies the M-step condition, written as

$$Q(\boldsymbol{\alpha}_{new},\sigma^2,\boldsymbol{K};\boldsymbol{\theta}_{(l)}) \geqslant Q(\boldsymbol{\theta};\boldsymbol{\theta}_{(l)}). \quad \forall\boldsymbol{\theta} \tag{36}$$

b) update σ^2: To find the update of σ^2, we derive the equation (31) with respect to σ^2, we get:

$$\begin{aligned} &\partial_{\sigma^2}Q(\boldsymbol{\theta},\boldsymbol{\theta}_{(l)}) \\ &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\mathrm{Tr}\left(\boldsymbol{S}^T\boldsymbol{S}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\boldsymbol{\eta}\boldsymbol{\eta}^T\right]\right) \\ &+ \frac{1}{2(\sigma^2)^2}\left\|\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{\alpha}\right\|^2 - \frac{1}{(\sigma^2)^2}(\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{\alpha})^T\boldsymbol{S}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\boldsymbol{\eta}\right]. \end{aligned} \tag{37}$$

Setting it equal to zero, we obtain

$$\sigma^2_{new} = \frac{1}{n}\|\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{\alpha}_{new}\|^2 + \frac{1}{n}\mathrm{Tr}(\boldsymbol{S}^T\boldsymbol{S}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\boldsymbol{\eta}\boldsymbol{\eta}^T\right]) - \frac{2}{n}(\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{\alpha}_{new})^T\boldsymbol{S}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\boldsymbol{\eta}\right]. \quad (38)$$

Since σ^2 is a positive number, we need to check that $\sigma^2_{new} > 0$. Therefore, we re-arrange the terms in equation (38) based on the property introduced in (30), we get

$$\sigma^2_{new} = \frac{1}{n}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\|\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{\alpha}_{new} - \boldsymbol{S}\boldsymbol{\eta}\|^2\right] \geqslant 0$$

and

$$\partial^2_{\sigma^2}Q(\boldsymbol{\theta},\boldsymbol{\theta}_{(l)})\big|_{\boldsymbol{\theta}=(\boldsymbol{\alpha}_{new},\sigma^2_{new},\boldsymbol{K})} = -\frac{n}{2(\sigma^2_{new})^2} < 0. \quad (39)$$

Then σ^2_{new} satisfies the M-step assumption, which is, $\forall\sigma^2, \boldsymbol{K}$

$$Q(\boldsymbol{\alpha}_{new},\sigma^2_{new},\boldsymbol{K};\boldsymbol{\theta}_{(l)}) \geqslant Q(\boldsymbol{\alpha}_{new},\sigma^2,\boldsymbol{K};\boldsymbol{\theta}_{(l)}). \quad (40)$$

c) Update β: Based on the definition of the matrix $\boldsymbol{K}$ introduced in equation (16), we denote $\boldsymbol{K} = \frac{1}{\beta}\tilde{\boldsymbol{K}}(\phi)$. We derive (31) with respect to β, it gives

$$\partial_\beta Q(\boldsymbol{\theta},\boldsymbol{\theta}_{(l)}) = \frac{1}{2}\frac{r}{\beta} - \frac{1}{2}\mathrm{Tr}\left(\tilde{\boldsymbol{K}}(\phi)^{-1}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\boldsymbol{\eta}\boldsymbol{\eta}^T\right]\right) \quad (41)$$

Setting this derivative equal to zero, we compute the update of β. It is given by,

$$\beta_{new} = \frac{r}{\mathrm{Tr}\left(\tilde{\boldsymbol{K}}(\phi)^{-1}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\boldsymbol{\eta}\boldsymbol{\eta}^T\right]\right)}, = \frac{r}{\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\boldsymbol{\eta}^T\tilde{\boldsymbol{K}}(\phi)^{-1}\boldsymbol{\eta}\right]} > 0 \quad (42)$$

where,

$$\partial^2_\beta Q(\boldsymbol{\theta},\boldsymbol{\theta}_{(l)})\big|_{\boldsymbol{\theta}=(\boldsymbol{\alpha},\sigma^2,\beta_{new},\phi)} = -\frac{r}{\beta^2_{new}} < 0 \quad (43)$$

Considering $\boldsymbol{\alpha}_{new}$ and σ^2_{new} computed before, we obtain: $\forall\phi,\beta$

$$Q(\boldsymbol{\alpha}_{new},\sigma^2_{new},\beta_{new},\phi;\boldsymbol{\theta}_{(l)}) \geqslant Q(\boldsymbol{\alpha}_{new},\sigma^2_{new},\beta,\phi;\boldsymbol{\theta}_{(l)}). \quad (44)$$

d) Update ϕ: We notice that when deriving equation (31) with respect to ϕ, we get a complicated expression with no explicit solution when setted equal to zero. In such a case when no closed form exists it may be feasible to attempt to find the value ϕ that globally maximizes the function $Q(\boldsymbol{\theta},\boldsymbol{\theta}_{(l)})$. This case is defined as the generalized EM algorithm (GEM), for which the M-step requires only that $\phi_{(l+1)}$ satisfies

$$Q(\boldsymbol{\alpha},\sigma^2,\beta,\phi_{(l+1)};\boldsymbol{\theta}_{(l)}) \geqslant Q(\boldsymbol{\alpha},\sigma^2,\beta,\phi_{(l)},\boldsymbol{\theta}_{(l)}), \forall\boldsymbol{\alpha}\forall\beta,\forall\sigma^2. \quad (45)$$

In this situation, where we don't have a closed form for updating parameter ϕ, we can use one step of the Newton-Raphson (NR) method [12], as the M-step. This gives

$$\phi_{(l+1)} = \phi_{(l)} - a_{(l)}\frac{\partial_\phi Q(\boldsymbol{\theta},\boldsymbol{\theta}_{(l)})\big|_{\boldsymbol{\theta}=\left(\boldsymbol{\alpha}_{(l+1)},\sigma^2_{(l+1)},\beta_{(l+1)},\phi_{(l)}\right)}}{\partial^2_\phi Q(\boldsymbol{\theta},\boldsymbol{\theta}_{(l)})\big|_{\boldsymbol{\theta}=\left(\boldsymbol{\alpha}_{(l+1)},\sigma^2_{(l+1)},\beta_{(l+1)},\phi_{(l)}\right)}}, \quad (46)$$

where $0 < a_{(l)} \leqslant 1$. This parameter controls the convergence rate, we can choose $a_{(l)} = 1$. We compute derivatives involved in equation (46) based on some results from matrix derivation theory detailed in [21] and we get

$$\partial_\phi Q(\boldsymbol{\theta},\boldsymbol{\theta}_{(l)}) = -\frac{1}{2}\mathrm{Tr}\left(\boldsymbol{K}^{-1}\frac{\partial\boldsymbol{K}}{\partial\phi}\right) + \frac{1}{2}\mathrm{Tr}\left(\boldsymbol{K}^{-1}\frac{\partial\boldsymbol{K}}{\partial\phi}\boldsymbol{K}^{-1}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\boldsymbol{\eta}\boldsymbol{\eta}^T\right]\right), \quad (47)$$

$$\partial^2_\phi Q(\boldsymbol{\theta},\boldsymbol{\theta}_{(l)}) = \frac{1}{2}\mathrm{Tr}\left(\boldsymbol{K}^{-1}\frac{\partial^2\boldsymbol{K}}{\partial^2\phi}(\boldsymbol{K}^{-1}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\boldsymbol{\eta}\boldsymbol{\eta}^T\right] - Id_r)\right) + \frac{1}{2}\mathrm{Tr}\left(\boldsymbol{K}^{-1}\frac{\partial\boldsymbol{K}}{\partial\phi}\boldsymbol{K}^{-1}\frac{\partial\boldsymbol{K}}{\partial\phi}(Id_r - 2\boldsymbol{K}^{-1}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}_{(l)}}\left[\boldsymbol{\eta}\boldsymbol{\eta}^T\right])\right).$$

Finally the derivatives of $\boldsymbol{K}$ with respect to ϕ are defines as,

$$\left[\frac{\partial\boldsymbol{K}}{\partial\phi}\right]_{i,j} = \left[\frac{\partial\boldsymbol{K}(\beta,\phi)}{\partial\phi}\right]_{i,j} = \frac{\|s_i - s_j\|^2}{\exp(\phi)}\boldsymbol{K}_{i,j}$$

and

$$\left[\frac{\partial^2\boldsymbol{K}}{\partial^2\phi}\right]_{i,j} = -\left[\frac{\partial\boldsymbol{K}(\beta,\phi)}{\partial\phi}\right]_{i,j} + \frac{\|s_i - s_j\|^2}{\exp(\phi)}\left[\frac{\partial\boldsymbol{K}(\beta,\phi)}{\partial\phi}\right]_{i,j}.$$

Appendix C
Characterization of η

We want to identify the conditional distribution of η given $\boldsymbol{Y}$ when the value of the parameter is $\boldsymbol{\theta}$, it is denoted by $\Pr(\eta|\boldsymbol{Y},\boldsymbol{\theta})$. We have,

$$\Pr(\eta|\boldsymbol{Y},\boldsymbol{\theta}) = \frac{\Pr(\boldsymbol{Y}|\eta,\boldsymbol{\theta})\Pr(\eta|\boldsymbol{\theta})}{\Pr(\boldsymbol{Y}|\boldsymbol{\theta})}.$$

Using expression (24) and (25) and up to a multivariate constant independent of η, we can write:

$$\Pr(\eta|\boldsymbol{Y},\boldsymbol{\theta}) \propto \exp(-\frac{1}{2\sigma^2}\|\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{\alpha} - \boldsymbol{S}\eta\|^2) \times \exp(-\frac{1}{2}\eta^T\boldsymbol{K}^{-1}\eta) \quad (48)$$

$$\propto \exp\left(\frac{1}{\sigma^2}(\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{\alpha})^T\boldsymbol{S}\eta - \frac{1}{2}\eta^T(\frac{\boldsymbol{S}\boldsymbol{S}^t}{\sigma^2} + \boldsymbol{K}^{-1})\eta\right), \quad (49)$$

We observe that $\eta \mapsto \Pr(\eta|\boldsymbol{Y},\boldsymbol{\theta})$ follows a Gaussian distribution with mean μ and covariance matrix $\boldsymbol{C}$. We have

$$\Pr(\eta|\boldsymbol{Y},\boldsymbol{\theta}) \propto \exp\left(-\frac{1}{2}(\eta-\mu)^T\boldsymbol{C}^{-1}(\eta-\mu)\right), \propto \exp\left(-\mu^T\boldsymbol{C}^{-1}\eta - \frac{1}{2}\eta^T\boldsymbol{C}^{-1}\eta\right). \quad (50)$$

Using (48) and (50), we can identify the mean and the covariance of $\eta|\boldsymbol{Y},\boldsymbol{\theta}$:

$$\begin{cases} \mu = \mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}}[\boldsymbol{\eta}] = (\boldsymbol{S}^T\boldsymbol{S} + \sigma^2\boldsymbol{K}^{-1})^{-1}\boldsymbol{S}^T(\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{\alpha}), \\ \boldsymbol{C} = \mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}}\left[(\boldsymbol{\eta}-\mu)(\boldsymbol{\eta}-\mu)^T\right] = (\frac{\boldsymbol{S}^T\boldsymbol{S}}{\sigma^2} + \boldsymbol{K}^{-1})^{-1}, \end{cases}$$

which yields,

$$\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}}\left[\boldsymbol{\eta}\boldsymbol{\eta}^T\right] = (\frac{\boldsymbol{S}^T\boldsymbol{S}}{\sigma^2} + \boldsymbol{K}^{-1})^{-1} + \mu\mu^T. \quad (51)$$