JS 26

INTELLIGENT TOOLS AND METHODS FOR RADIO SCIENCE

Temporal Prediction of RF-EMF Exposure Using Surrogate Modeling with Environment-based Clustering

Paul Lagouanelle[1], *Shanshan Wang*[1], *Yukun Liu*[1], *Yarui Zhang* [2], *Joe Wiart*[1]

[1] *Chaire C2M, LTCI, Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France*
[2] *SATIE, ENS Paris-Saclay, CNRS, Université Paris-Saclay, Gif-sur-Yvette, France*

Keywords: Surrogate modeling, Data clustering, Exposure, EMF, In situ measurements

Abstract

This paper addresses the radio-frequency electromagnetic fields exposure modeling using data-driven surrogate models as an alternative to costly real-world measurements. It relies on data from large-scale EMF sensor networks deployed in urban environments. While neural-network-based models have shown promising results, they are insufficient to capture complex spatio-temporal exposure. Therefore, alternative approaches such as kriging and polynomial chaos are explored, but requiring prior data clustering. Clustering is based on environmental complexity extracted from OpenStreetMap, using variables such as building density and urban occupation rate. Although current models achieve acceptable safety-level accuracy, further improvements are needed for high-precision exposure mapping and complete time-series prediction.

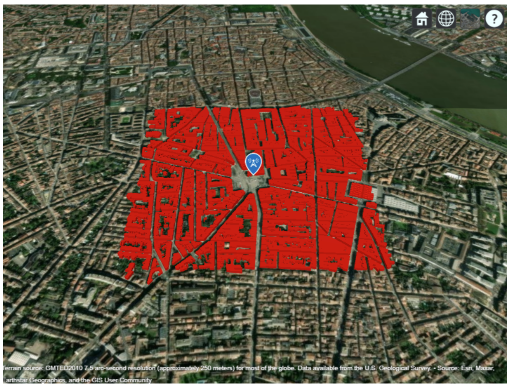
1 Introduction

Radio-frequency electromagnetic fields (RF-EMF) from wireless networks has become an integral part of everyday life, with mobile connections now far exceeding the global population and smartphone usage reaching very high levels. Although international exposure guidelines [1] and standardized compliance testing methods exist, public concern about potential health risks from RF-EMF remains strong, particularly with the recent deployment of 5G networks. This has fueled societal debate and motivated research efforts aimed at better understanding mobile network characteristics, including signal propagation, coverage, and performance across different environments.

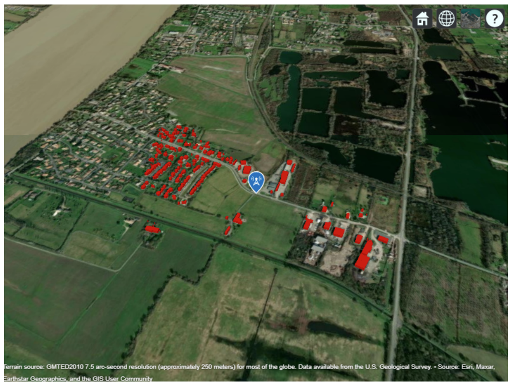
This paper aims at exploiting data from a network of EMF sensors deployed across French cities [2]. These sensors designed by EXEM measure the spatial components of the electric field in the range (250 kHz - 6 GHz). Every 2 hours, the sensor is activated to conduct one 6 min measurement. Over recent decades, significant work has been devoted to RF-EMF exposure assessment and monitoring through these fixed probes and spot measurements using Convolutional Neural Networks (CNNs)-based models. [3, 4]. Although particularly suited to treat these huge number of data, CNNs have been efficient at estimating the spatial distribution and magnitude variations of electric field levels across multiple frequency bands, but are still unable to predict time series data. The work presented here aims at expanding these recent developments to another surrogate model for the prediction of time series data: the Polynomial-Chaos based Kriging metamodel [5]. Unlike CNNs, the PCK surrogate model used here needs smaller training datasets obtained through prior data clustering for enhanced prediction performance.

2 Environment-based clustering

After studying several pure statistical approaches using the signal distribution to cluster the time series from the 108 sensors, the resulting metamodels were unable to properly approximate the temporal variation of the electric field on most sensors. As the temporal variation of the EMF is highly dependent on the environment surrounding the sensor, the most conclusive approach aims to directly exploit the environmental complexity in order to group data into disjoint training sets. A sensor's environment can be easily extracted using the collaborative tool OpenStreetMap, which allows data to be collected within a $400m$ square around each sensor, including the number of buildings and for each building: its average height, its surface area, its function (residential,



(a) Bordeaux_01 sensor
$n_{buildings} = 2186$ $OccupationRate = 0.2163$



(b) Saint-Louis-de-Montferrand_01 sensor
$n_{buildings} = 116$ $OccupationRate = 0.0377$

Figure 1: Example of two different environments (buildings in red) around 2 sensors in the Bordeaux region

commercial, religious, agricultural...). After an in-depth analysis of this environmental data, two environmental variables were chosen to partition the data: the number of buildings $n_{buildings}$ and the occupation of the urban space in volume $OccupationRate = \frac{\sum BuildingsVolume}{400m^2 \times MaxBuildingHeight}$. An example of two different environments within the Bordeaux region with the associated environmental variables is shown on Figure 1: the highly urbanized center of the city of Bordeaux and a rural village.

Based on these two environmental variables, each sensor can be associated with one of the five clusters defined by a classical k-means method [6] (see figure 2). The two previous environments are clearly reflected in two very different clusters (1 and 4). However, this graph clearly shows that the clusters seem complex to define because many data pairs are close to each other. One could imagine to mix distribution-based and environmental-based clustering by adding extra layers for computing clusters based on the mean or the variance. But due to the high influence of the environment on the moments of the distribution, this does not provide any improvement to the existing clusters of figure 2. Therefore, one possible improvement is to add a third environmental variable or split the $OccupationRate$ in two variables which would help in defining smaller clusters and thus, in improving both computation time of the surrogate models and their prediction performance.

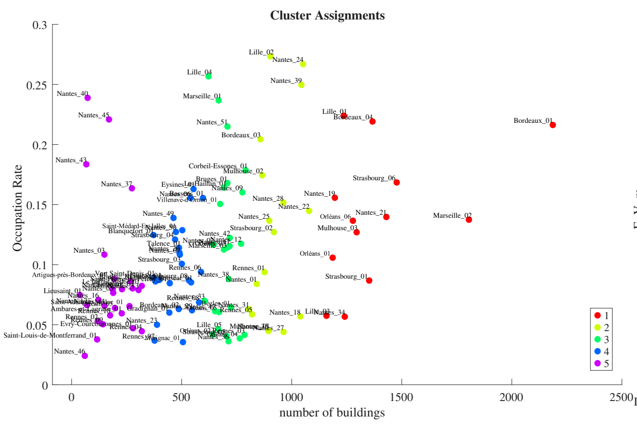


Figure 2: Partitioning of the 108 sensors

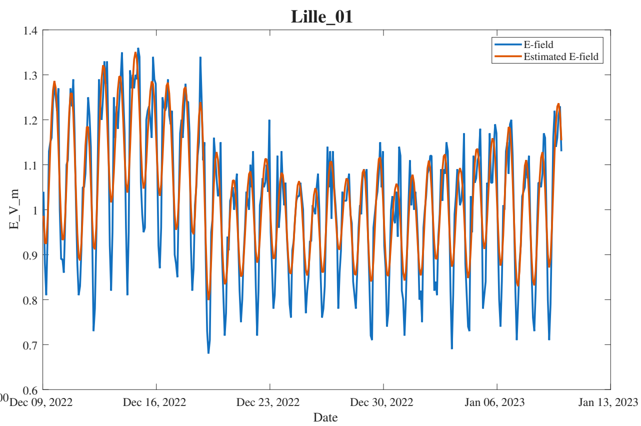


Figure 3: Estimation (blue) and true value (red) of the EMF variation for the Lille_01 sensor

3 Surrogate modeling

Based on the clusters defined on figure 2, sensors from each cluster are used separately to train 5 different surrogate models on a short time period: 09 Dec 2022 - 09 Jan 2023. For each sensor the training dataset consists of:

$$\left[n_{\text{buildings}},\ \text{OccupationRate},\ t_{\text{date}} = \begin{pmatrix} \text{09 Dec 2022 00:01} \\ \vdots \\ \text{09 Jan 2022 23:59} \end{pmatrix} \right] \Longrightarrow E_{V_m} = \begin{pmatrix} E_m(\text{09 Dec 2022 00:01}) \\ \vdots \\ E_m(\text{09 Jan 2022 23:59}) \end{pmatrix}$$

Although, each sensor provides a time series made of 384 samples for the training, the amount of datapoints is insufficient for the use of classical neural network surrogate models. Hence, in order to build an exact interpolator, the chosen surrogate model combines the advantages of the well-known Krigeage and Chaos Polynomial surrogate models: the Chaos Polynomial Kriging (PCK) metamodel. The Kriging part is used to interpolate local variations in the model output, while the PCE approximates the overall behavior of the system. A PCK metamodel is defined by [5]: $\mathcal{M}(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}} y_{\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) + \sigma^2 Z(\boldsymbol{x},\omega)$ where $\sum_{\boldsymbol{\alpha}\in\mathcal{A}} y_{\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$ is a weighted sum of orthonormal polynomials (*trend* of the PCK) and σ^2 is the variance of the stationary Gaussian process $Z(\boldsymbol{x},\omega)$. In order to assess the PCK metamodel performance, 11 cross-validation (CV) surrogate models have been built for each sensor using the Root Mean Squared Error (RMSE) and the Mean Absolute Percentage Error (MAPE) as evaluation metrics. These results are presented on figure 4, which shows the median (black line), the interquartile range (box: 25th–75th percentiles), and whiskers extending to 1.5×IQR with outliers beyond this range shown as red dots. Table 1 displays the MAPE and RMSE for every CV surrogate model.

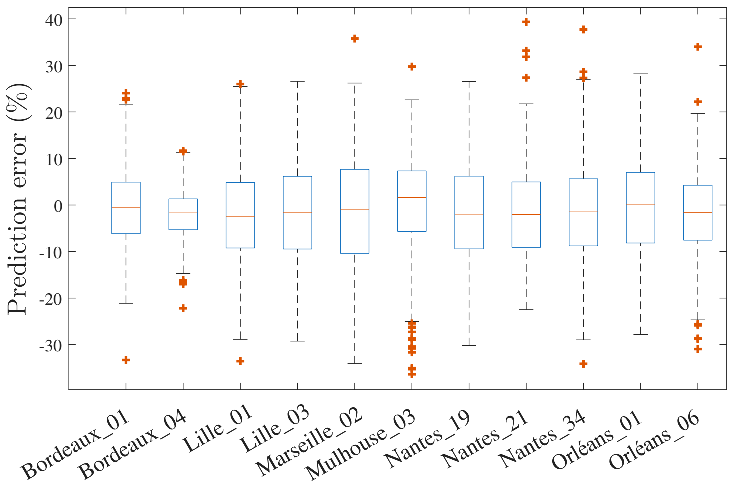


Figure 4: Box plots of prediction rates of CV surrogate models in the first cluster

Table 1: RMSE and MAPE for the CV surrogate models among the first cluster

Sensor	RMSE	MAPE
Bordeaux_01	0.184	5.69%
Bordeaux_04	0.0385	7.49%
Lille_01	0.0716	6.00%
Lille_03	0.0912	8.48%
Marseille_02	0.0943	8.47%
Mulhouse_03	0.229	6.89%
Nantes_19	0.0952	10.1%
Nantes_21	0.0935	7.67%
Nantes_34	0.100	14.5%
Orléans_01	0.226	7.11%
Orléans_06	0.0693	7.78%

When looking at the validation metrics, their values are sufficient to define the safety of the environment for human exposure for every sensor (an example is displayed on figure 3 for estimating E-field values for the Lille_01 sensor). Thus, the global metamodel built on the full cluster with 11 sensors is able to predict accurately the temporal variation of any environment within the cluster defined by $n_{buildings}$ and the $OccupationRate$. This clearly shows that data partitioning is of great interest for our substitution models for predicting time series. Although some sensors have a few outliers out of the predictive range, this work on our first environmental-based cluster is a success and will be expanded on the other clusters and tested with sensor data outside the 108 sensors available.

Acknowledgment

This work was supported in part by European project SEAWAVE adn GOLIAT through the Horizon Europe Research and Innovation Programme under grants 101057622 and 101057262.

References

[1] International Commission on Non-Ionizing Radiation Protection. "Guidelines for limiting exposure to electromagnetic fields (100 kHz to 300 GHz)". In: *Health Physics* 118.5 (2020), pp. 483–524.

[2] Ourouk Jawad et al. "Monitoring of the exposure to electromagnetic fields with autonomous probes installed outdoors in France". In: *Comptes Rendus. Physique* 25.S1 (2024), pp. 41–61.

[3] Shanshan Wang, Taghrid Mazloum, and Joe Wiart. "Prediction of RF-EMF exposure by outdoor drive test measurements". In: *Telecom*. Vol. 3. 3. MDPI. 2022, pp. 396–406.

[4] Yarui Zhang, Shanshan Wang, and Joe Wiart. "ExposNet: A deep learning framework for EMF exposure prediction in complex urban environments". In: *arXiv preprint arXiv:2503.02966* (2025).

[5] Roland Schobi, Bruno Sudret, and Joe Wiart. "Polynomial-chaos-based Kriging". In: *International Journal for Uncertainty Quantification* 5.2 (2015).

[6] Aristidis Likas, Nikos Vlassis, and Jakob J Verbeek. "The global k-means clustering algorithm". In: *Pattern recognition* 36.2 (2003), pp. 451–461.