Progress In Electromagnetics Research B, Vol. 87, 111–129, 2020

A Numerical Simulation System for Mobile Telephony Base Station EMF Exposure Using Smartphones as Probes and a Genetic Algorithm to Improve Accuracy

Pierre Combeau[1, *], Nicolas Noé[2], François Gaudaire[2], Steve J. Demeffo[1], and Jean-Benoit Dufour[3]

Abstract—With the increasing number of mobile phone users, new services, and mobile applications, the proliferation of radio antennas has raised concerns about human exposure to electromagnetic waves. This is now a challenging topic to many stakeholders such as local authorities, mobile phone operators, citizen, and consumer groups. Thus, the prediction of exposure map at urban scale is a very important requirement to find a relevant indicator of the real exposure. In this paper, we propose a monitoring solution for electromagnetic field (EMF) exposure based on a numerical modeling of the radio wave propagation radiated by mobile telephony base stations. The accuracy of this tool directly depends on the input data precision, such as location of base station antennas or their radiation pattern, which are often poorly known. These data are therefore refined by an optimization algorithm fed by a lot of information, such as the indication of the received signal strength (RSSI) measured directly from users' smartphones, which are used as probes. Results show that this method significantly improves the precision of unknown data concerning mobile base stations and the accuracy of exposure maps at urban scale.

1. INTRODUCTION

Since the emergence of the first cellular networks in the 1990s, the mobile communications market has grown steadily to reach today almost 5 billion users worldwide (see Figure 1).

To meet this demand, the infrastructures of mobile phone operators are constantly evolving. In particular, new transmitters are regularly installed, on the one hand to cope with increasing traffic, and on the other hand depending on the evolution of the technologies or frequency bands used. The exponential increase in radio frequency transmitters naturally causes concern among citizens about the potential harm of the generated waves on their health, although to date no studies have proven any health effects. Thus, monitoring of peoples exposure to electromagnetic waves from mobile telephone base stations is increasingly of interest to local communities, in order to communicate with residents or enforce local exposure regulations. This can be achieved from intense measurement campaigns [1–3], but the corresponding data quickly become outdated due to the constantly evolving network. Furthermore, this would be very expensive for a large number of measurements spots and for the numerous technologies and frequencies.

On the other hand, numerical simulation gives accurate results at large scale and in every location [4, 5]. Nevertheless, it needs a precise antenna description (location, power, radiation pattern ...). This information is sometimes accessible in the public domain, as in [6] managed by the ANFR

Received 4 February 2020, Accepted 20 May 2020, Scheduled 8 June 2020
* Corresponding author: Pierre Combeau (pierre.combeau@univ-poitiers.fr).
[1] CNRS, Université de Poitiers, XLIM, UMR 7252, F-86000 Poitiers, France. [2] CSTB, Lighting and Electromagnetism Division, 44000 Nantes, France. [3] Geomod, 69003 Lyon, France.

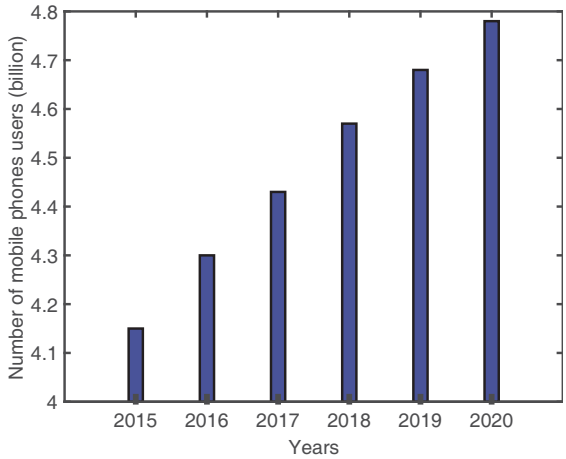


Figure 1. Evolution of the number of mobile phones users in the world.

(FRequency Nationnal Agency) in France. Unfortunately, it is very partial and inaccurate, and all the needed parameters for a numerical modeling are not furnished.

In this paper, the proposed idea consists in using smartphones as cheap but ubiquitous power sensors, in order to enhance the accessible data concerning BTS' (Base Transceiver Station) emitters from georeferenced information such as received power levels provided by the smartphones deployed in the environment. The simultaneous and complementary use of smartphone data and numerical modeling tool should give, via optimization and statistical process, the real and full characteristics of the antennas. Thus, it would become possible to produce continuous updated exposure maps at urban scale.

This article is structured as follows. Section 2 presents the general principle of the proposed monitoring system. Then Section 3 deals with the definition of several criteria used in the literature to compare measured and simulated data, and concludes on the proposed ones in this study. Section 4 is dedicated to the presentation of the optimization algorithm, used to determine the unknown input data (antenna location, ...) in order to optimize the criteria presented in Section 3. The theoretical validation of the monitoring system will be demonstrated in Section 5 on degraded pure numerical cases. Then, real data measured on smartphones are used in Section 6 to test the performance and the robustness of the proposed system in realistic conditions. Finally, conclusions and perspectives are provided in Section 7.

2. PRINCIPLE

As shown in Figure 2, the proposed monitoring system is based on three tools:

- Vigiphone: a smartphone application developed on Android by the CSTB (Scientific and Technical Center for Building) to get and store georeferenced information of received power level;
- an optimization algorithm to estimate the real characteristics of the antennas;
- MithraREM: a software to predict radiowave propagation in realistic environment, based on Geometric Optic and Uniform Theory of Diffraction.

The inputs of the system are the 3D geometrical data of the propagation environment and the partially known antenna parameters, such as the location, height, electric and mechanic tilts, radiation pattern, and emitted power. At the same time, a set of measurements are realized from the smartphones present in the environment. Then the measured and simulated data feed the optimization algorithm which tries to determine the unknown antenna parameters in a way that simulated data fit the measured ones as well as , whereas new measured data from smartphones are continuously injected. By this way, the objective is to compensate the relatively bad quality of each node (smartphone) with their large number.

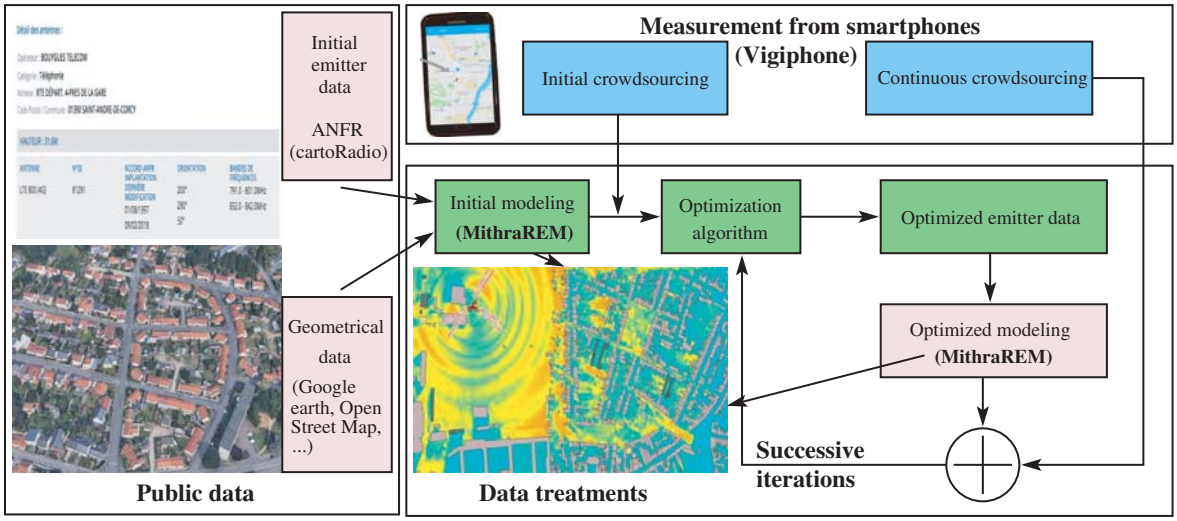


Figure 2. Principle of the monitoring system.

As a proof of concept [7], Figure 3 shows the comparison between measured RSSI from a 2G antenna and the corresponding simulated result, obtained in the case of a perfect knowledge of the antenna characteristics. A very good agreement can be observed.

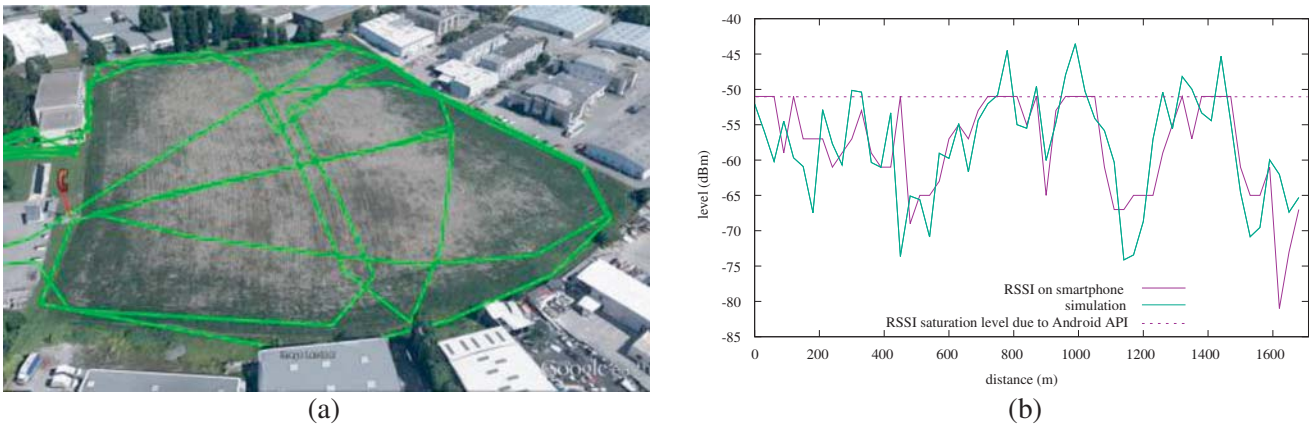


Figure 3. Proof of concept: (a) Measurement route with transmitter in red, (b) RSSI versus simulated power.

The optimization algorithm needs a criterion in order to estimate the level of similarity between measured and simulated signals. In the next section we present a quick state of the art of the different criteria used in the literature for comparing measured and simulated signal strength. Then we present those used for the proposed monitoring system of the EMF exposure.

3. COMPARISON CRITERIA

To establish a pertinent criterion of comparison between simulation and measurement, in any exposition configuration, is not a trivial thing. Indeed, the evolution of a single parameter can impact on different criteria in a contradictory way, an objective conclusion being difficult to take in this case.

In literature, the different criteria of similarity measurement are divided into two families, i.e., error type criteria and link type criteria.

3.1. Error Type Criteria

These criteria are based on the distance between two signals, i.e., in the concerned application simulation and measurement. From here we denote X and Y as two data vectors (measured and simulated) of N length. Table 1 presents the three main error type criteria used in the literature.

Table 1. Main error type criteria.

ME	***SDE***	***RMSE***
$\frac{1}{N}\sum_{i=1}^{N} X_i - Y_i$	$\sigma_X - \sigma_Y$ where $\sigma_X = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^2}$ and $\overline{X}$ is the mean value of X	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - Y_i)^2}$

First, the mean error criterion *ME* is the mean of the difference between X and Y. Second, the standard deviation error criterion *SDE* measures the difference between the standard deviations of the two signals. Finally, the root mean square error criterion *RMSE* is the root square of the square error between X and Y.

3.2. Link Type Criteria

These criteria, contrary to the precedents, are not interested in the absolute values of the signals but in the shape relation between them. They are usually based on a correlation measurement. Table 2 shows the two main ones.

Table 2. Main link type criteria.

Pearson correlation	**Spearman correlation**
$P = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$ with $Cov(X,Y) = \frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})$	$S = 1 - \frac{6\sum_{i=1}^{N}(r(X_i) - r(Y_i))^2}{N^3 - N}$ where $r(X_i)$ is the rank of the ith value of signal X in its sorted distribution

The Pearson's correlation coefficient P between two signals X and Y corresponds to the ratio of their covariance and the product of their standard deviation. It is a measure of the linear relation between X and Y. However, depending on the variability of the data, the use of this criterion can lead to wrong conclusions, because it is more dedicated to normal distributed data with no outliers. This criterion is not adapted to every exposure configurations, because of the strong variability of the electric field. Indeed, the multipath phenomena can lead to observe very strong and weak values at a few distance between two observation points.

The Spearman's correlation is an alternative to this problem. The Spearman's coefficient S is the Pearson's correlation applied to the samples' rank. It analyses the relation between the samples rank, allowing to detect the existence of monotonic relations whatever their shape. Unlike Pearson coefficient, it is not interested in the value of the sample but in its rank, which avoids errors due to outliers. Thus, it can also be used with signals presenting nonlinear relations.

3.3. Proposed Criteria

It is risky to only trust a single quality indicator to estimate a model. It is the reason why, in the literature, most of the methods to compare measurements and simulations are based at least on two criteria.

In [8], authors used *ME* and *RMSE* to compare measurements and simulation in 2G signals context. In [9], authors used *SD* and *P* in addition of *ME* and *RMSE* to study the potential interference between

LTE and WiFi signals. In [10], the SD and S are used to evaluate simulations according to measurements in the context of exposure to telephony base stations.

For this set of applications, at least two criteria have been used to evaluate performances of simulated results. However, in some cases concluding on a comparison may become difficult, because the two criteria may give opposite indications. For instance the standard deviation error may decrease while the correlation decreases. This highlights the importance of well choosing the used criteria, and have driven us to consider, in addition to classical criteria previously presented, new ones combining their respective avantages.

In this study, we consider the following criteria:

- Standard deviation error SDE: as a reference one because of its classical use in existing studies. This criteria has to be minimized by the optimization algorithm;
- Pearson's correlation coefficient P: it avoids the impact of a change in the power radiated by the base station. This last one can be found in post-treatment by minimizing the standard deviation between mesurements and simulations. This criteria has to be maximized;
- Spearman's correlation coefficient S: it avoids the problem of outliers. This criteria has to be maximized;
- Hybrid criterion P_H defined as:
$$P_H = (1 - P)RMSE \tag{1}$$
The idea is to add the respective advantages of the error type and link type criteria. But as it was previously shown, the evolution of a parameter set can drive, at the same time, to both correlation and error increase. So this particular criterion consists in jointly optimizing the two indicators. It has to be minimized;
- Hybrid criterion S_H defined as:
$$S_H = (1 - S)RMSE \tag{2}$$
This is the same approach as P_H but with the Spearman correlation. This criterion has to be minimized;

The next section presents the proposed optimization algorithm used to optimize the previous criteria.

4. OPTIMIZATION ALGORITHM

An optimization method aims at minimizing/maximizing an objective function, as those presented in Subsection 3.3. In our case it consists in finding, from the badly known public data, the real characteristics of a base station antenna, namely:

- its 2D horizontal location (longitude, latitude);
- its height;
- its mechanical tilt;
- its radiation pattern (including electric tilt);

In the literature, there are a lot of optimization methods. A first family is based on the knowledge of an analytical formulation of the objective function. The idea is then to found the local optimum with a gradient descend method, as for conjugate gradient method such as Marquadt, by choosing the descent successive directions and an optimal step in each direction. These methods are iterative and need the numerical value of the objective function at a given point, but also of the gradient, or even the hessian matrix. This information is computed independently of the algorithm. To converge these methods need to be well initialized, i.e., not too far from the global optimum, otherwise it presents the drawback to converge to a local optimum.

In our application, we do not have any analytical model which provides the objective function value from a set of input parameters. Furthermore, the purpose is not to find a local optimum, but the global one. The satisfactory resolution of an optimization problem which contains a large number of sub-optimal solutions, often justifies the use of the second family of methods, the meta-heuristic.

A meta-heuristic is an optimization algorithm to solve complex problems for which there is no efficient conventional solution. It is usually at least partially stochastic, allowing to explore all the solution' space, and doesn't need the derivative of the objective function. The three main meta-heuristics methods are the simulated annealing, the swarm of particles and the genetic algorithms. The simulated annealing is efficient to solve complex problem when there is no local optimum. In our case with numerous local optima it is too dependent on the initial parameter set. The main drawback of the swarm particles is its premature convergence, which also potentially leads to a stagnation around local optimum.

So the retained method is based on a genetic algorithm, able to converge to the global solution. Genetic algorithms are stochastic optimization methods reproducing natural selective mechanisms which have already shown their efficiency in similar application contexts as in [11]. Figure 4 shows the synoptic of the proposed genetic algorithm.

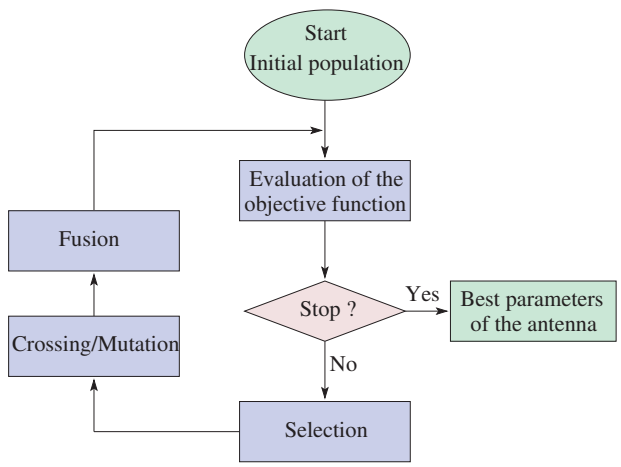


Figure 4. Synoptic of the proposed genetic algorithm.

The algorithm is randomly initialized with a population of potential solutions. The performances of these solutions are then evaluated by the objective function, and on this basis, a new population is generated by using evolutional operators: selection, crossing, mutation and fusion. This cycle is iterated until an acceptable solution is found or the maximum iterations number is achieved.

4.1. Parameter Coding and Search Domain

4.1.1. 3D Location and Orientation of the Antenna

The three first parameters to optimize (2D location, height, and mechanical tilt) are coded as numerical variables, either continuous (2D location and height) or discrete (mechanic tilt, with a one degree sampling). Their search intervals are directly derived from the uncertainty of their corresponding public data provided by [6], which are the following:

- location of the base station : 30m;
- base station height: 2m;
- mechanical tilt: ranging from 0 to 12°.

The azimuth of the antenna is considered as a known input, as it is available in the public databases.

4.1.2. Radiation Pattern (Including Electric Tilt)

Concerning the radiation pattern, a database of the different antennas (and their occurrence) used by the French telecom operators has been built from previous global exposure simulations in large areas. From this (anonymized) input, statistical parameters of antennas are extracted: horizontal aperture, vertical aperture, and electrical tilt. Table 3 shows the 2G & 3G antennas data in the Paris area:

Table 3. Statistical characteristics of 2G-3G antennas in Paris, France.

Variable	Average	RMS
Electrical tilt	−5.6	3.1
Vertical aperture	11.0	11.0
Horizontal aperture	64.2	22.5

Gaussian laws are created from these statistical parameters. In order to randomly select a radiation pattern, electrical tilt, horizontal and vertical apertures are randomly drawn using these normal laws. Then the diagram with the closest parameters to the drawn ones is selected. If the selected diagram parameters are too diverging from the drawn parameters, it is rejected and another draw is done.

The "distance" between the horizontal $\Delta\varphi$ and vertical $\Delta\theta$ drawn aperture parameters and those of diagram in the database $\Delta\varphi_i$ and $\Delta\theta_i$ is given by the formula:

$$2\left|\Delta\theta - \Delta\theta_i\right| + \frac{1}{2}\left|\Delta\varphi - \Delta\varphi_i\right| \tag{3}$$

The error on the vertical aperture is amplified compared to the error on the horizontal aperture to account for more directional main beam in the vertical plane for outdoor BTS. If this distance exceeds 5 (in degrees) or if the electrical tilt does not exist in the database, the sample is rejected. This diagram selection is done in order to have real diagrams (with secondary lobes) from existing real antennas. It could not be achieved by creating numerical diagrams with the given parameters.

4.2. Initialization

An iso-surface stratified sampling is used to draw the 2D location of the antenna. This is better than choosing purely random locations in the full search domain around the theoretical one (the one provided by ANFR) because it ensures a quasi homogenous distribution of the potential locations within this search domain. This is illustrated in Figure 5.

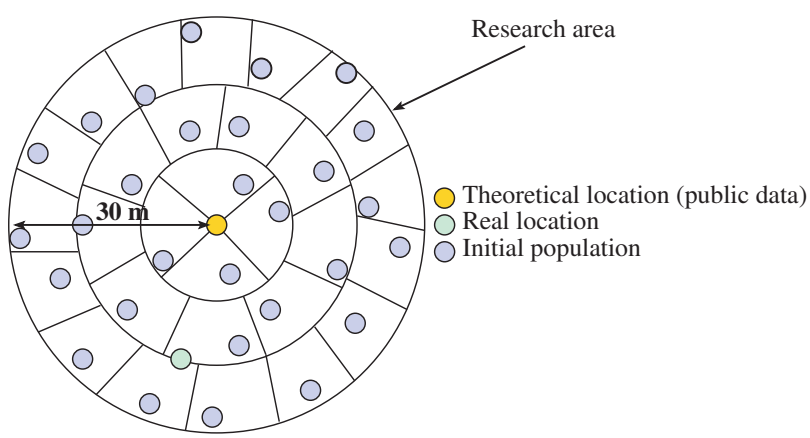


Figure 5. Iso-surface stratified initialization.

The radius of the search domain is first divided in several sub-radii. Each sub-radius is then divided in several angular sectors of equal area inside of which the initial population of potential locations is randomly chosen.

Then height and tilt are uniformly initialized in their respective ranges of uncertainties around their theoretical values, i.e., $[-1\,\text{m}, 1\,\text{m}]$ and $[0, +12°]$.

Finally, radiation patterns are initialized with the method presented in Subsection 4.1.

4.3. Evaluation of the Objective Function

In our application, the objective function evaluation corresponds to the simulation of the different links between the current population (a set of potential base station antennas) and the set of receivers corresponding to the smartphone measurements. Its evaluation leads to the computation of a scalar value, called *score* in the following, corresponding to one criterion among those presented in Subsection 3.3. Each simulation is performed using Mithra-REM tool which is a ray-based simulator relying on Geometrical Optics (GO) and the Uniform Theory of Diffraction (UTD). It takes into account all diffractions on top of the buildings from the emitter to the receiver and a given number of reflections on building walls (typically 1 to 2) to ensure good path loss estimation.

4.4. Selection

The selection step consists in choosing some elements in the current population which will be used to generate the new population at the next iteration. There are mainly four selection techniques:

- elitist selection: it consists in selecting the best elements, i.e., the ones which provides the best *score* values, so the best values of the objective function. Having a lot of local optimums, this method is not a good one because it quickly leads to a stalled algorithm around a local optimum;
- Russian roulette: each element of the population has a probability to be selected, proportional to its *score* value. Thus the best elements are always privileged. As for the elitist selection, this leads to quickly converge to a local optimum without exploring all the search domain;
- uniform selection: all the elements have the same probability to be selected. As a consequence there is a non-null probability at each iteration to select only bad (low *score* value) or good (high *score* value) elements. Both cases would mislead the algorithm since an heterogenous population is a mandatory condition to ensure a good exploration of the search domain.
- simulated annealing: this technique gives a significant probability of selecting bad elements, specifically in the first iterations, to ensure good exploration and avoid quick convergence to a local minimum. But as the number of iterations increases, the probability of selecting bad elements lowers (the name is an analogy to annealing in metallurgy, when the cooling of a material progressively reduces the possibility to change its internal structure).
  Be S_{n_i} the normalized *score* of the ith element, defined by:

$$S_{n_i} = \frac{score(i)}{\sum_{j=1}^{N_{pop}} score(j)} \tag{4}$$

  with N_{pop} is the number of elements in the population. The ith element will be selected if $e^{-\delta/T} > \xi_i$, with ξ_i an uniformly distributed value between 0 and 1, and $\delta = 1 - S_{n_i}$. The lower the T (the "temperature") value, the less likely the bad elements are to be selected, as it is illustrated in Figure 6(a).
  The principle of simulated annealing is to decrease the T parameter according to the iterations to start from a quasi uniform selection ($T \gg 1$) to a quasi elitist one ($T \ll 1$). From iteration i to iteration $i+1$, T decreases with the following law: $T_{i+1} = rate \times T_i$ with $0 < rate < 1$. If *rate* is close to 1, there is a slow decrease in the panning of the solution space, i.e., bad elements keep a significant probability to be chosen. In this case the algorithm behavior remains close to the one of a uniform selection for a long time. If *rate* is close to 0, the exploration capacity decreases rapidly and the algorithm behavior quickly becomes close to the elitist selection one. Initial T and *rate* values are key parameters which have to be adjusted according to the specific configuration under consideration.

In the proposed algorithm, we consider a total population of 150 elements from which 66% (i.e., 100) are selected at each iteration by considering the simulated annealing approach with an initial T (T_{init}) and *rate* values experimentally fixed to 1 and 0.99 respectively. Furthermore, the relatively high *rate* value is compensated by the fact that at each iteration, 5% of the selected elements are automatically

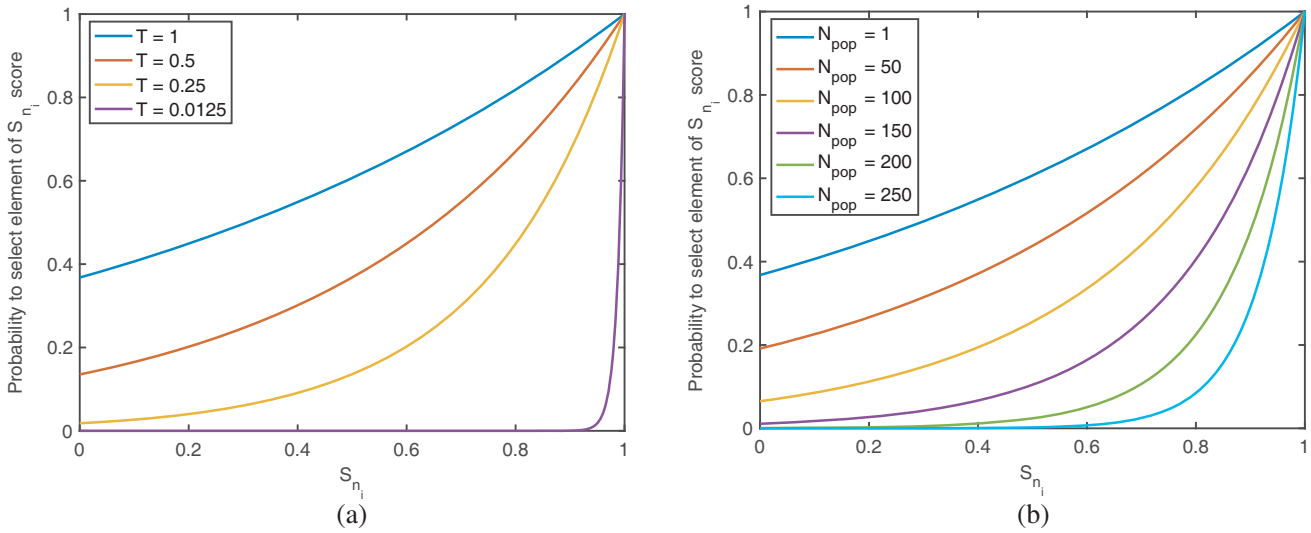


Figure 6. Probability to choose element with S_{n_i} *score* according to (a) T, (b) N_{pop} for $T_{init} = 1$ and $rate = 0.99$.

chosen and corresponds to the best elements of the population. In this way a good variety of the population is ensured by mixing best elements with bad ones.

4.5. Crossing/Mutation strategies

The diversity of the population must be maintained over the generations in order to explore as widely as possible the space of the solutions. This role is mainly that of crossing and mutation operators [12, 13], which are used to generate new elements (children) from randomly chosen couple of elements into the selected ones (parents). The proposed algorithm has three crossing/mutation options that alternate over iterations. The probability to cross/mutate a specific parameter of a couple of candidate base station antennas, following a uniform distribution between 0 and 1, differs according to each option as:

- Option 1:
  - 2D location (longitude, latitude): 50%;
  - Height: 16.66%;
  - Radiation pattern: 16.66%;
  - Vertical tilt: 16.66%.
- Option 2:
  - 2D location (longitude, latitude): 50%;
  - Height: 25%;
  - Radiation pattern: 25%.
- Option 3:
  - 2D location (longitude, latitude): 100%.

In all cases, the probability to cross/mutate the 2D location is the highest one, firstly because of the high sensitivity of the objective function to this parameter [4], and secondly of its search domain which is the highest one. Whatever the option, the treatments applied according to chosen parameters are:

- 2D location: be loc_i a uniformly distributed random value between 0 and 1.
  - if $loc_i < 0.5$, the latitude parameters of the two parents are crossed;
  - if $loc_i \geq 0.5$, the longitude parameters of the two parents are mutated.
- height: be h_i a uniformly distributed value between 0 and 1.
  - if $h_i < 0.5$, the heights of the two parents are crossed;

  - if $h_i \geq 0.5$, the heights of the two parents are mutated.
- vertical tilt: be t_i a uniformly distributed value between 0 and 1.
  - if $t_i < 0.5$, the tilts of both parents are permuted;
  - if $t_i \geq 0.5$, new tilts are chosen for both parents, from a Gaussian law with mean and standard values corresponding to the known antenna data bases (*cf.* Section 4.1.2).
- radiation pattern: be rp_i a uniformly distributed value between 0 and 1.
  - if $rp_i < 0.5$, the radiation patterns of both parents are permutated;
  - if $rp_i \geq 0.5$, new radiation patterns are chosen for both parents, as presented in Section 4.1.2.

4.5.1. Crossing of 2D Location and Height

An example of crossing between two parents is illustrated in Figure 7, in the case of a chosen longitude parameter. The child is built by incorporating the selected parameter (longitude in this example) of the first parent into the second parent set of parameters instead of its initial value. It is important to note that for the crossing step, the 2D search domain is divided into several circular cells of 1 meter radius. Then each crossing between two parent elements is achieved into the same cell, in order to avoid a quick convergence to a local extremum and so to ensure a good exploration of the search domain.

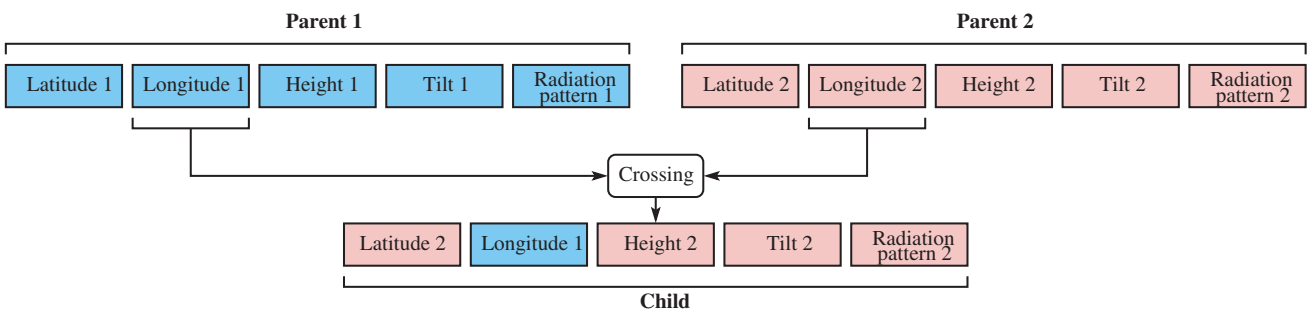


Figure 7. Example of 2D location crossing.

4.5.2. Mutation of 2D Location and Height

Mutation operator gives to genetic algorithms the ergodicity property of space travel [14, 15]. Following the strategy proposed by Michalewiz [16], it consists in perturbing the value of the selected parameter (latitude, longitude or height), not with a simple Gaussian noise, but by decreasing the level of perturbation along the iterations, still in order to well explore the search domain in the beginning of the algorithm and then to focus only on the best elements. Let x be the initial value of the parameter to mutate and η be a uniformly distributed value between 0 and 1. The mutated value x' of x is defined as follows:

$$x' = \begin{cases} x + \Delta(\max(x) - x) & \text{si} \quad \eta < 0.5 \\ x - \Delta(x - \min(x)) & \text{si} \quad \eta \geq 0.5 \end{cases} \tag{5}$$

where $\max(x)$ and $\min(x)$ are given in Subsection 4.1, and

$$\Delta(y) = r.y.\left(1 - \frac{t}{N_{MaxIter}}\right)^b \text{ with} \begin{cases} r & : \text{random value uniformely distributed between 0 and 1.} \\ t & : \text{current iteration number.} \\ N_{MaxIter} & : \text{maximal number of iterations.} \\ b & : \text{exposant empirically fixed to 5.} \end{cases} \tag{6}$$

Figure 8 illustrates the mutation step in the case where height parameter was selected.

4.6. Fusion

This step consists in choosing, among the child elements generated by the selection and crossing/mutation steps, the ones which will be introduced into the population at the next iteration of the algorithm, instead of their parents. Two approaches are followed along the iterations:

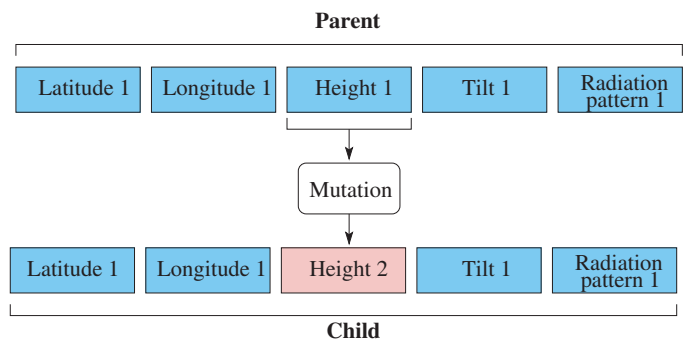


Figure 8. Example of height mutation.

- During the first one-third of the iterations, the child parameter set replaces the parent one if the two following conditions are fulfilled:
  - the value of the objective function (S_{n_i}) is better than the one of the parent;
  - the euclidian 2D distance between the parent's and child's locations is less than 50 cm.
- During the last third of the iterations, the parent which has an S_{n_i} value less than its child one is automatically replaced by its child, whatever the distance between them is. Thus the algorithm ends by focusing on the most interesting regions of the search domain.

4.7. Stop Criterion

Genetic algorithms have usually three stop criteria. The first one is based on an accuracy threshold, i.e., the algorithm stops when an element reaches a target *score* value. The second is to stop the algorithm when the population does not move from one generation to the next, whereas the last one is to stop the algorithm when a fixed maximum number of iterations is reached.

In this article, the last criterion is used, and the maximum number of iterations $N_{MaxIter}$ is fixed to 229 in order to have a final T value of 0.1.

5. THEORETICAL VALIDATION

5.1. Overview

In order to validate the whole concept, there is a need for a reference solution to compare results. The first step is to perform a pure virtual validation. There are many advantages in performing such a validation:

- the reference solution (both as exposure maps of electric field and as antenna parameters),
- there is no uncertainty in the geometrical model of the environment (ground and buildings) and in the propagation modelling,
- this allows the optimization algorithm parameters to be adjusted,
- this allows to quantify the number, distribution, and quality of RSSI data needed for convergence.

The process is illustrated in Figure 9.

The reference electric field map is first computed from an "Ar" (reference state of antennas). Virtual RSSI level and GPS locations are then randomly extracted from this map (outdoor environment only, 1.5 m above the ground). Finally, both the reference state of antennas and these "perfect" virtual measurements are degraded.

As far as antennas are concerned, positions are randomly moved using search parameters of the optimization algorithm (i.e., 30 m radius in the horizontal plane, and 2 m in the vertical dimension), and diagrams are replaced by randomly chosen ones from the database. For virtual measurements, locations are perturbed using horizontal GPS error (10 m radius, points moved inside buildings are removed). The RSSI measurement is degraded so as to fit real RSSI information (1 or 2 dB step, clamped to lower and upper limits).

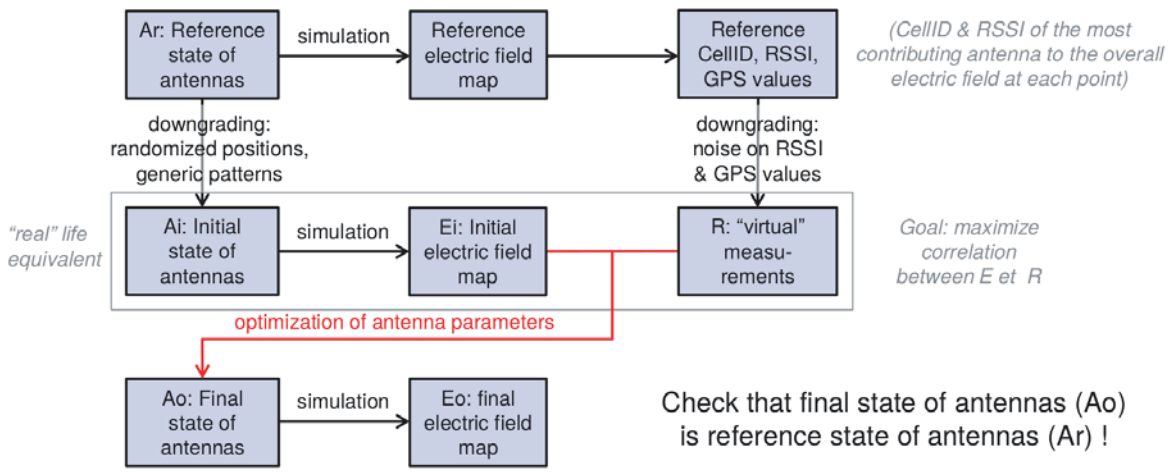


Figure 9. Theoretical validation of the monitoring system.

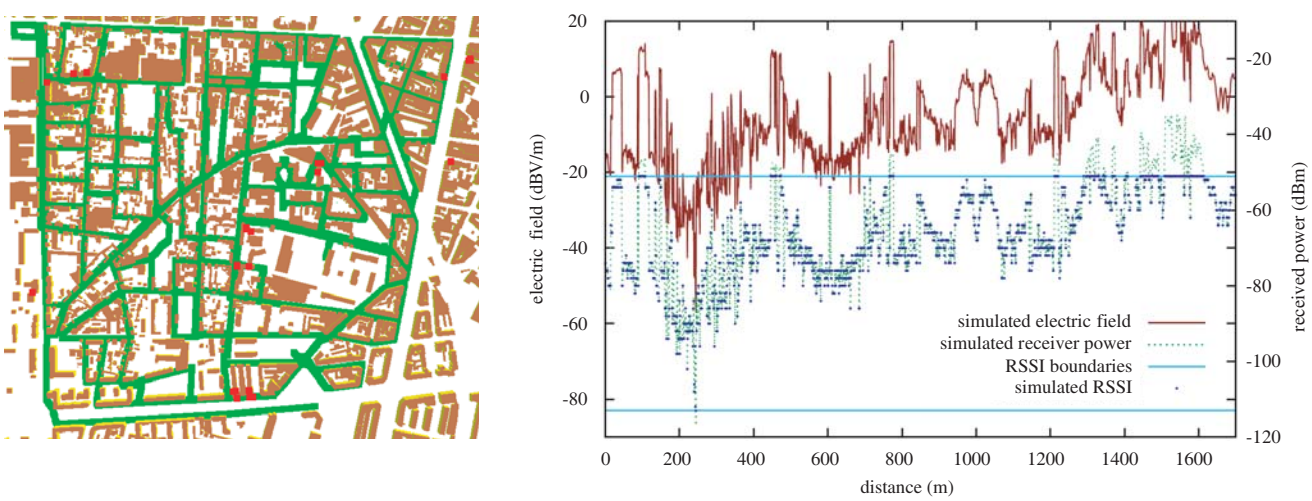
Once this degradation is done, a "real life" equivalence of the reference state is obtained, called the initial state. Then the optimization algorithm is performed, to get the final antenna state "Ao". This final state is compared to the reference state to assert that the original parameters are found.

5.2. Results

The theoretical validation has been conducted in an urban area in the city of Grenoble (France) with the following characteristics:

- a 2 km × 2 km area (*cf.* Figure 10(a)),
- 34 GSM emitters,
- 9000 receivers (ground and facade map).

The electric field is simulated using MithraREM with 1 reflection on building (vertical) frontages, and as much as needed reflections on the ground and diffractions on the buildings tops.



(a) (b)

Figure 10. (a) Area of Grenoble used for numerical validation of the monitoring system: receiver maps (green) and BTS locations (red points), (b) electric field to RSSI transformation.

First, degradation of reference state of antennas (from Ar to Ai) is performed:

- the position of each emitter is randomly moved within a 5 m × 5 m square horizontally and ±1 m vertically,
- the mechanic tilt is set to a random value (between 1° and 9°), whatever the real value is,
- the radiation diagrams are left untouched.

The CellID information of each antenna is kept unchanged. The downgrading of the receivers data is then performed (from Ei to R):

- the position of each receiver is randomy moved within a 5 m x 5 m square horizontally only (receivers inside buildings are removed), to account for GNSS inaccuracy,
- the simulated electric field at each receiver (contributions from all antennas at full power) is transformed into an RSSI (BCCH† power only), keeping for each receiver the most contributing antenna only, then discretized and clamped to the 32 possible dBm values (ignoring saturated values), as illustrated in Figure 10(b).

Using the optimization algorithm and the Pearson correlation criteria (between pseudo-measured RSSI and simulated one), antenna parameters are modified until they converge to stable values. Results show that the degraded antenna parameters (Ai) always converge toward the original ones (Ar). That is to say the mechanic tilt is the original one, and the 3D position of the antenna is within 50 cm from the original one. These results are very good but were expected. Since numerical simulation for the virtual RSSI measurements on the same "perfect" geometrical model are used, all uncertainties related to local variation of the electric field (moving cars, ...) and errors between the real environment (heterogenous and detailed buildings) were not taken into account in the simulation.

6. REAL TEST CASE

A real test case was conducted in Nantes (France), using a dedicated measurement trolley and controlled drive tests. The selected area (*cf.* Figure 11(a)) is a 1.6 km × 1.1 km area with four base stations, with different characteristics (on a mast, on a tall building, on a lower building, some in open environment and some in dense urban area), with all four French operators and all technologies (2G, 3G, 4G).

6.1. Measurements

In order to cover all the selected area, routes corresponding to a total length of about 13 km have been planned prior to the measurements (*cf.* Figure 11(a)).

(a) (b)

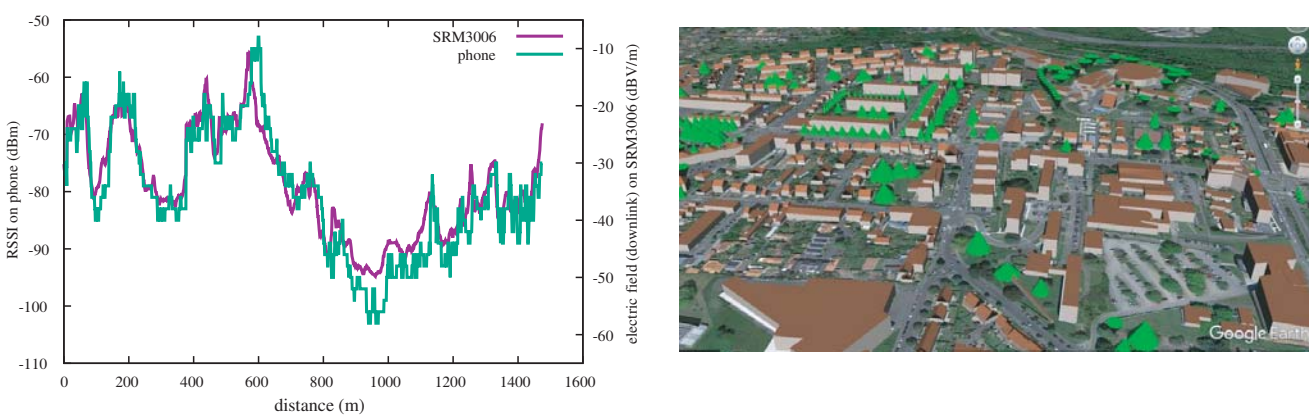
Figure 11. (a) Measurement area with BTS locations and trolley walks as colored lines, (b) measurement setup.

† Broadcast Control CHannel

A trolley dedicated to the measurements was developed (*cf.* Figure 11(b)). This trolley carries:

- two smartphones (four others were handheld by two researchers/students following the trolley),
- a Narda SRM3006 Spectrum analyzer, mesuring the electric field in each downlink band of each band of each operator,
- a fisheye camera, taking pictures (1 Hz frequency) at the front,
- a differential GPS (DGPS) chipset,
- a laptop to control both DGPS and camera.

The phones were configured so as to receive only one technology (2G, 3G, or 4G) so as to handle all emitters available in the area. The SRM3006 was used to identify the frequency band associated with a given CellID and to check that the measured RSSI on the phone was correctly correlated to the electric field, with just an offset, as illustrated in Figure 12(a). The fisheye camera was used in a post-processing stage, in areas where measured and simulated signals diverge. This helped identify missing buildings in the numerical model, as well as the fact that vegetation, such as trees, is needed to be taken into account. The DGPS chipset gives a more accurate location (around 2 m precision) than basic GPS smartphone chipset. As a consequence, measurements were located using DGPS instead of GPS (this should not be a problem in the future when GNSS with Galileo is widely available, with a higher precision). The laptop is here to control both the DGPS and the camera and to synchronize everything. The results of the measurements are DGPS located RSSI (or equivalent) values, for several CellIDs.



(a) (b)

Figure 12. (a) Comparison of smartphone and spectrum analyzer measurements, for a given CellId, (b) numerical model of Nantes.

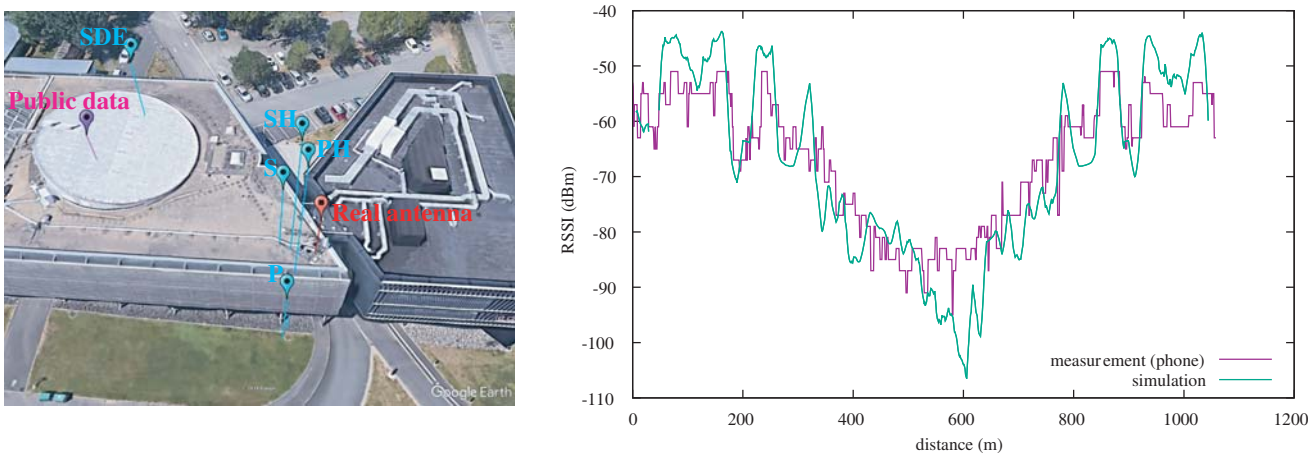
6.2. Numerical Model

The numerical model (illustrated in Figure 12(b)) was derived from GIS information of the IGN (Institut Géographique National), but was improved to account for slanted roofs and large trees. We noticed that using slanted roofs (instead of flat roofs at the top of the roof) improved correlation between measurements and simulation. Adding trees (modeled with International Telecommunication Union (ITU) attenuation models [17]) also largely improved correlation in open areas with no buildings.

6.3. Results

Results are presented here for two of the three BTSs in the experimental area (Orange and Bouygues Telecom, see Figure 11(a)). A unique antenna has been singled out for each BTS, the one with the azimuth heading towards the measurement area.

Concerning the Bouygues Telecom BTS, the antenna of interest corresponds to the LTE 1800 technology. It is deployed on the roof of a medium height building, and its precise location is referenced



(a) (b)

Figure 13. Results for a LTE 1800 emitter on a medium height building: (a) optimized BTS locations, (b) comparison between smartphone measurements and numerical simulations from optimized (S_H) BTS parameters.

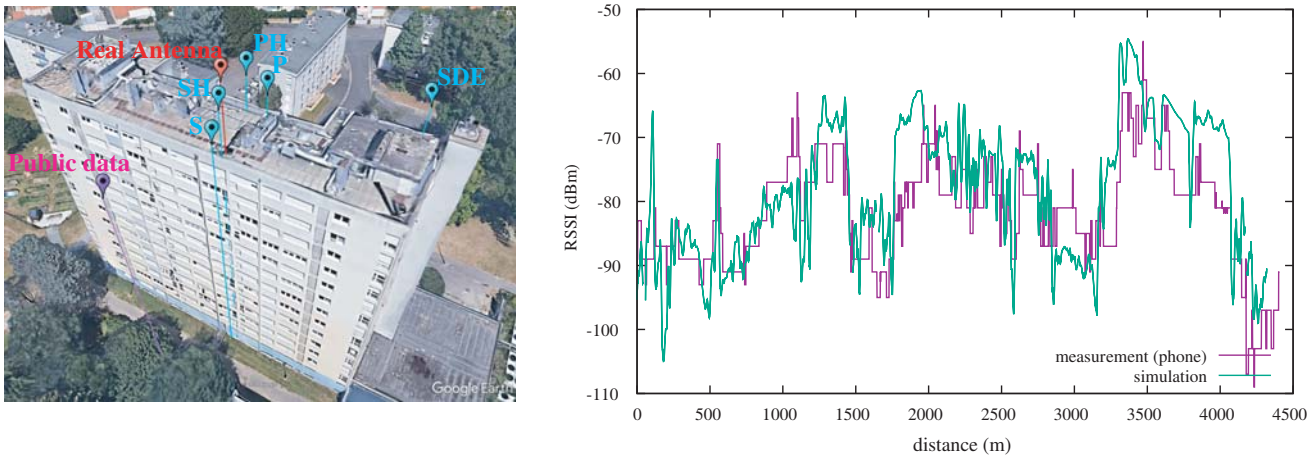
as "Real antenna" in Figure 13(a). Its location given by public data is referenced as "Public data" in the same figure. From this theoretical location, the initial population is first stochastically initialized, and the proposed iterative algorithm is then processed through the different steps described from Subsections 4.3 to 4.7.

Table 4 presents the results for the different antenna parameters according to each considered criterion. One can note the very bad location of the BTS stored in public data, with a horizontal (δl) and a vertical (δz) errors of about 28 and 2 m, respectively. The classical error type criterion *SDE* gives bad results with a horizontal location error of about 26 m, whereas overall the four other considered criteria all give good results, allowing us to identify the BTS with a maximum horizontal error of about 5 m. A more detailed analysis of these results shows that the link type criteria based on correlation (S and P) are more effective, which is explained by the fact that the correlation criteria are less sensitive to exceptional values than error type ones. Among the correlation criteria, the Spearman correlation S gives the best results with a location error less than 2 m in the horizontal plane. Finally, using the two hybrid criteria S_H and P_H increases the accuracy of the unknown parameters. As S is better than P, S_H is better than P_H and locates the BTS with an horizontal error of 1.03 m and a vertical error of 0.46 m. The final obtained locations according to each criterion are illustrated in Figure 13(a). Figure 13(b) shows the simulated RSSI obtained from the S_H optimized BTS parameters. It is very close to the RSSI measured on the smartphone, reinforcing the behavior of the genetic algorithm.

Figure 14 illustrates the second scenario concerning the Orange BTS (see Figure 11(a)) located on a tall building and corresponding to the UMTS 2100 technology. Table 5 presents the optimized BTS

Table 4. Optimized BTS parameters according to considered criteria: medium height building scenario.

	Latitude (°)	**Longitude** (°)	**x** (m)	**y** (m)	**z** (m)	**δl** (m)	**δz** (m)
Real antenna	47.250786	−1.55428	−167.2	97.5	44.19	0	0
Public data	47.250556	−1.554444	−181	73	42	28.12	−2.19
SDE parameters	47.250652	−1.554565	−189.58	83.86	43.20	26.21	−0.99
P parameters	47.250769	−1.554213	−162.22	95.34	45	5.43	0.81
S parameters	47.250771	−1.554282	−167.45	95.78	44.71	1.74	0.52
P_H parameters	47.250778	−1.554232	−163.65	96.43	45.18	3.71	0.99
S_H parameters	47.259780	−1.554269	−166.45	96.8	44.65	1.03	0.46



(a) (b)

Figure 14. Results for a UMTS 2100 emitter on a tall building: (a) optimized BTS locations, (b) comparison between smartphone measurements and numerical simulations from optimized (S_H) BTS parameters.

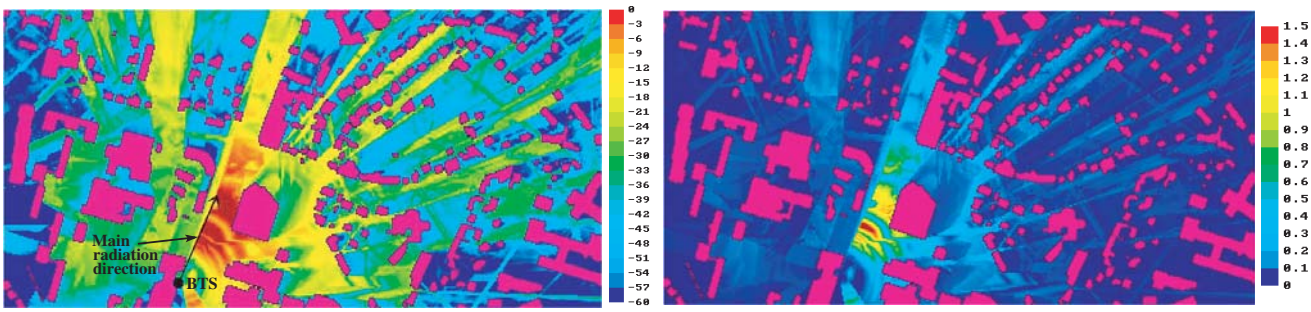
Table 5. Optimized BTS parameters according to considered criteria: tall building scenario.

	Latitude (°)	**Longitude** (°)	**x** (m)	**y** (m)	**z** (m)	**δl** (m)	**δz** (m)
Real antenna	47.254554	−1.567739	−1159.26	574.12	67.7	0	0
Public data	47.254444	−1.567778	−1162.91	562.06	67.7	12.60	0
SDE parameters	47.254658	−1.567543	−1143.80	584.77	68.08	18.73	0.38
P parameters	47.254607	−1.567730	−1158.21	579.9	67.36	5.87	−0.34
S parameters	47.254541	−1.567744	−1159.74	572.65	67.35	1.55	−0.35
P_H parameters	47.254615	−1.567775	−1161.56	580.98	67.56	7.24	−0.14
S_H parameters	47.254545	−1.567733	−1158.86	573.03	67.30	1.16	−0.40

parameters according to the different considered criteria. A similar analysis to the first scenario can be made. The best results are again obtained by considering the hybrid criterion S_H, which provides the location of the BTS with horizontal and vertical errors of respectively 1.16 and −0.40 m. *SDE* criterion gives the worst results with a horizontal location error of about 19 m.

Finally, Figure 15(a) shows the exposure map simulated in the whole area with optimized BTS parameters obtained from the proposed method, whereas Figure 15(b) presents the estimation error between this map and the one obtained with BTS parameters given in public data. Before the optimization, the overall level of electric field is clearly under-estimated, particularly in the region of high electric levels, near the BTS. This is emphasized by the statistical parameters of the exposure maps corresponding to the two BTS of Figures 13 and 14, given in Table 6. This table shows the average, median, maximum levels, and 99% levels (i.e., 99% of the levels are lower than this value) of the exposure maps. We can first note that the differences between the average and median values obtained from public and optimized BTS data are weak and not very significant. On the contrary, the differences between the maximal values and the 99% levels obtained from public and optimized BTS data, which are of great interest for exposure evaluation, are much more important. It is particularly true in the case of the LTE 1800 BTS for which public data locate the BTS on the middle of the building roof top, leading to a strong shadowing effect. This highlights the interest of the proposed monitoring system to evaluate in a realistic way the EMF exposure due to mobile telephony base station.

As far as computation time is concerned, the optimization process for an antenna takes between 5 mn and 30 mn depending on the complexity of the environment on a 10 core CPU (Intel Xeon ES-2660,



(a) (b)

Figure 15. (a) Simulated exposure map (in dBV/m) for the LTE 1800 emitter of Figure 13, (b) map of estimation error (in V/m) between exposure evaluated from public and optimized BTS data.

Table 6. Exposure level indicators for the LTE 1800 and the UMTS 2100 BTS.

		Average value	**Median value**	**Maximum Value**	**99% level**
LTE 1800	Public data	< 0.001 V/m	0.005 V/m	0.06 V/m	0.01 V/m
	Optimized BTS	0.021 V/m	0.006 V/m	2.67 V/m	0.61 V/m
UMTS 2100	Public data	0.087 V/m	0.009 V/m	2.41 V/m	1.01 V/m
	Optimized BTS	0.066 V/m	0.008 V/m	1.92 V/m	0.63 V/m

2.6 GHz). For the Nantes test case, this is 20 mn for a total about 23,000 evaluations of the objective function.

7. CONCLUSION

This article presents a new numerical monitoring system for estimating the exposure to EMF radiated by telephony base station. An originality of this work consists in proposing a method combining measurements and simulations.

Indeed, it is based on the use of smartphones deployed in the environment as poor quality but numerous probes to measure a partial view of the BTS coverage zones. Furthermore, it uses numerical simulations of radio wave propagation (Ray-Tracing + UTD tool) to compute the received levels from each known BTS. Finally, a genetic algorithm is proposed to iteratively adapt unknown BTS parameters to the real ones by trying to fit simulation to the smartphone measurements.

A set of different criteria has been investigated to compare measurement and simulation results. Two types of families based on error and correlation have been considered, and two new hybrid criteria combining their respective advantages were proposed. We have shown that they increase the accuracy of the estimated BTS parameters compared with conventional criteria.

All steps of the genetic algorithm were precisely described so that the reader can attempt to reproduce the results presented. A theoretical validation of the overall concept was tested based on virtual measurements produced from degraded simulated data. Two different real test cases corresponding to two different technologies and telephony operators were treated in a dense urban area in the Nantes down town in France.

The proposed method significantly corrected the BTS parameters given in public data, by reducing the location error to about 1 m. Furthermore, the proposed method is flexible and can dynamically adapt to changes in telephony networks, such as the modification of an existing antenna (electric tilt...) or the deployment of new antennas/BTS. Thus, it is a useful tool to estimate in real time the EMF exposure at urban scale.

As far as future generation antennas (Multi Users-Massive MIMO antennas with beamforming in future 5G networks for instance) the method should be adapted in order to account for fast varying radiation diagrams.

ACKNOWLEDGMENT

The authors would like to thank the Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (Anses) for supporting this work through the EST-2015 project titled "Système autonome de caractérisation de l'exposition aux champs électromagnétiques radiofréquences issus des stations de base de téléphonie mobile, hybridant acquisition sur smartphones et simulation numérique (VigiExpo)".

REFERENCES

1. Huang, Y., A. Krayni, A. Hadjem, J. Wiart, C. Person, and N. Varsier, "Comparison of the average global exposure of a population induced by a macro 3G network in urban, suburban and rural areas," *Radio Science Conference (URSI AT-RASC), 2015 1st URSI Atlantic*, 1-1, May 2015.
2. Conil, E., Y. Corre, N. Varsier, A. Hadjem, G. Vermeeren, W. Joseph, S. Aerts, D. Plets, L. Martens, L. M. Correia, and J. Wiart, "Exposure index of EU project lexnet: principles and simulation-based computation," *Proceedings of the 8th European Conference on Antennas and Propagation*, 3029–3032, IEEE, 2014.
3. Lo-Ndiaye, M., N. Noé, P. Combeau, F. Gaudaire, and Y. Pousset, "Analysis of electromagnetic waves spatio-temporal variability in the context of exposure to mobile telephony base station," *Progress In Electromagnetics Research C*, Vol. 88, 179–194, 2018.
4. Noé, N., F. Gaudaire, and M. D. B. L. Ndiaye, "Estimating and reducing uncertainties in raytracing techniques for electromagnetic field exposure in urban areas," *2013 IEEE-APS Topical Conference on Antennas and Propagation in Wireless Communications (APWC)*, 652–655, Sept. 2013.
5. Alwajeeh, T., P. Combeau, R. Vauzelle, and A. Bounceur, "A high-speed 2.5D ray-tracing propagation model for microcellular systems, application: Smart cities," *IEEE European Conference on Antennas and Propagation (EUCAP)*, Paris, France, Jan. 2017.
6. ANFR, Cartoradio, http://www.cartoradio.fr, 2004.
7. Noé, N., F. Gaudaire, M. Lo-Ndiaye, and P. Combeau, "Toward a stand-alone monitoring system for mobile telephony base stations exposure using simulations and smartphones crowdsourcing," *First URSI Atlantic Radio Science Conference*, Gran Canaria, Canary Islands, Spain, May 18–22, 2015.
8. COST, "Digital mobile radio towards future generations systems," http://www.lx.it.pt/cost-231/final report.htm, 1999.
9. Corre, Y. and Y. Lostanlen, "Three-dimensional urban em wave propagation model for radio network planning and optimization over large areas," *IEEE Transactions on Vehicular Technology*, Vol. 58, No. 7, 3112–3123, Sept. 2009.
10. Beekhuizen, J., R. Vermeulen, H. Kromhout, A. BuRgi, and A. Huss, "Geospatial modelling of electromagnetic fields from mobile phone base stations," *Journal of The Total Environment*, 445–446, 202–209, Feb. 2013.
11. Infantolino, J. M. K., M. J. Barney, and R. L. Haupt, "Using a genetic algorithm to determine an optimal position for an antenna mounted on a platform," *IEEE Military Communications Conference*, IEEE, Boston, MA, USA, Oct. 2009.
12. Kaya, Y., M. Uyar, and R. Tekin, "A novel crossover operator for genetic algorithms: Ring crossover," *CoRR*, abs/1105.0355, 2011.
13. Vekaria, K. and C. Clack, "Selective crossover in genetic algorithms: An empirical study," A. E. Eiben, T. Bäck, M. Schoenauer, and H.-P. Schwefel, editors, *Parallel Problem Solving from Nature — PPSN V*, 438–447, Springer Berlin Heidelberg, Berlin, Heidelberg, 1998.
14. Kalyanmoy, D. and D. Debayan, "Analysing mutation schemes for real-parameter genetic algorithms," *Int. J. Artif. Intell. Soft Comput.*, Vol. 4, No. 1, 1–28, Feb. 2014.
15. Soni, N. and T. Kumar, "Study of various mutation operators in genetic algorithms," *International Journal of Computer Science ans Information Technologies*, Vol. 5, No. 3, 4519–4521, 2014.

16. Michalewicz, Z., *Genetic Algorithms + Data Structures = Evolution Programs*, Springer Berlin Heidelberg, 1992.
17. ITU, Attenuation in vegetation, https://www.itu.int/dms pubrec/itu-r/rec/p/R-REC-P, 833-9-201609-I!!PDF-E.pdf, 2016.