Self-Diagnosing Low Coverage and High Interference in 3G/4G Radio Access Networks based on Automatic RF Measurement Extraction

M. Sousa[2,3], A. Martins[1,3] and P. Vieira[1,2]
[1]*Instituto de Telecomunicações (IT), Lisbon, Portugal*
[2]*Instituto Superior de Engenharia de Lisboa (ISEL), ADEETC, Lisbon, Portugal*
[3]*CELFINET, Consultoria em Telecomunicações Lda., Lisbon, Portugal*

Keywords: Wireless Communications, SON, Self-Diagnosis, Coverage Detection, Interference Control.

Abstract: This paper presents a new approach for automatic detection of low coverage and high interference scenarios (overshooting and pilot pollution) in Universal Mobile Telecommunications System (UMTS) /Long Term Evolution (LTE) networks. These algorithms, based on periodically extracted Drive Test (DT) measurements (or network trace information), identify the problematic cluster locations and compute harshness metrics, at cluster and cell level, quantifying the extent of the problem. Future work is in motion by adding self-optimization capabilities to the algorithms, which will automatically suggest physical and parameter optimization actions, based on the already developed harshness metrics. The proposed algorithms were validated for a live network urban scenario. 830 3^{rd} Generation (3G) cells were self-diagnosed and performance metrics were computed. The most negative detected behaviors regards high interference control and not coverage verification.

1 INTRODUCTION

The increasing network complexity, in terms of number of monitored parameters and parallel operation of 2^{nd} Generation (2G), 3G and 4^{th} Generation (4G), is increasing dramatically, besides the ever growing traffic volume and service diversity. In the third quarter of 2015, an average monthly data traffic of 4,700 petaBytes was registered, with an increase of 65% compared with the third quarter of 2014 (Ericsson, 2015). This increases the network Operating Expense (OpEx), forcing mobile operators to pursue strategies for reducing it. Self-Organizing Networks (SON) algorithms have been seen as the solution, and a way to automatically operate the current and beyond mobile networks.

This paper focus on the automatic-diagnosing algorithms of low coverage and high interference scenarios, used to trigger optimization processes in self-optimizing functions (within SON). Problem detection is based on periodically extracted DT measurements or geo-positioned network traces (Vieira et al., 2014). The DT data provides measurements of received signal strength and quality for the different pilot or reference signals that reach a certain location. Moreover, specific filtering is applied to identify the data that denotes either, coverage issues or interference problems, such as overshooting or pilot pollution.

SON and self-diagnosing is a hot research topic and recent work has been done leading to the current research stage (Duarte et al., 2015),(Sousa et al., 2015), (Sallent et al., 2011). The paper contribution is incremental to previous works. Firstly, the used cell's service area approximation based on propagation modelling is more accurate when compared with existent research. Secondly, a new cluster partition approach using the auto-correlation distances for shadow fading (Kysti et al., 2007), to limit cluster size, results in a more coherent data analysis. Finally, this work address the DT minimization target of 3^{rd} Generation Partnership Project (3GPP) by allowing trace based inputs.

The aim of the present work is to be the cornerstone in a fully self-optimization algorithm. This work corresponds to the network data analysis phase, which will report the detected areas with low performance issues and all relevant data for the further optimization algorithm. Additionally, harshness metrics are already computed, accessing the cell's detected problems.

This paper is organized as follows. Section 2 defines the scope embedding the present work. In section 3 the concept of DT reliability is presented. Sec-

Sousa, M., Martins, A. and Vieira, P.
Self-Diagnosing Low Coverage and High Interference in 3G/4G Radio Access Networks based on Automatic RF Measurement Extraction.
DOI: 10.5220/0005958300310039
In *Proceedings of the 13th International Joint Conference on e-Business and Telecommunications (ICETE 2016) - Volume 6: WINSYS*, pages 31-39
ISBN: 978-989-758-196-0

tion 4 presents the detailed self-diagnosing process for each algorithm. The results are shown in section 5 and finally, in section 6, conclusions are drawn.

2 SELF-OPTIMIZATION

Self-Optimization, aims to maintain network quality and performance with a minimum of manual intervention. It monitors and analyses, either Key Performance Indicator (KPI)s, DT measurements, traces or other sources of information, triggering automated actions in the considered network elements. The self-optimization process for a network under performance enhancement is shown in Figure 1. This work researches the Self-Diagnosis process block, hence future work will complete the self-optimization function by implementing the optimization and Operational Support System (OSS) script blocks.

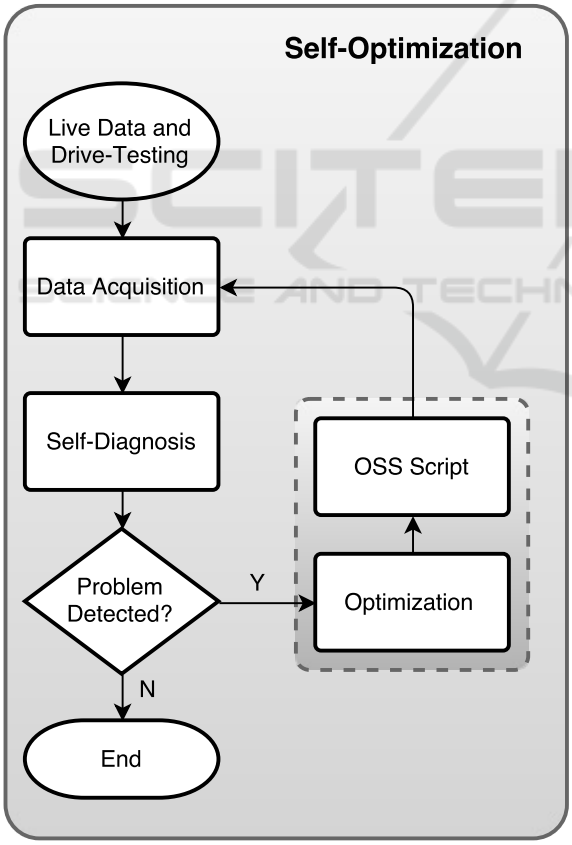

Figure 1: Global self-optimization flowchart.

The self-diagnosis module uses either DT measurements or trace data as in Figure 2. Thus, with an introduction of a reliability index R for the input data, it accomplishes not only a control mechanism for the self-optimization process, but also increases quality in the self-diagnosis output. Only when the available data, is significant to access the cell's performance, reflected by a value of R higher than a certain threshold, min_value, the self-diagnosis is executed. Otherwise, the available information is not sufficient to retain any strong conclusion. Furthermore, regarding the trace reliability block, it is under development.

The self-diagnosis module is composed by three different algorithms, coverage holes, overshooting and pilot pollution, see Figure 2. For all of them, the concerning data in the network optimization scenarios is grouped in clusters. Moreover, a harshness index H_{cell} is calculated accessing the problem's severity in the cell.

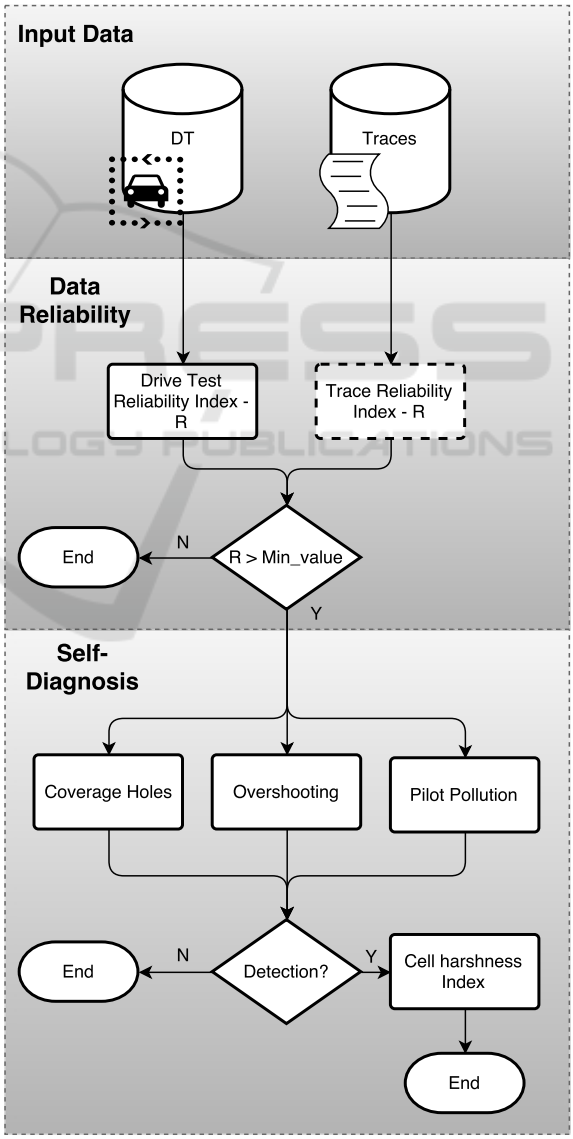

Figure 2: Self-diagnosis flowchart.

3 DRIVE-TEST DATA CLASSIFICATION

There is an undeniable correlation in the assertiveness of the DT based algorithm results and the completeness of a DT campaign. As well, the harshness metrics will be as accurate as the DT is complete or significant. This highlights the importance of a quality metric associated with the DT used in any algorithm. Hence, an assessment of the DT quality is executed before the use of DT data itself, giving origin to a DT reliability index R (Sousa et al., 2015). It evaluates the data collected in a cell's service area, taking into account the percentage of measurements collected and its spatial distribution.

3.1 DT Spatial Distribution

The quadrant method is a statistical spatial analysis method used as a mean to test a population point pattern (random, clustered and dispersed). In this context, the method is not used to classify the DT measurements on its geographical distribution pattern, but as a relative measurement of the dispersion level. This allows to quantify how well distributed are the DT measurements in the cell's service area.

For that purpose, the service area is divided into quadrants and the number of DT measurements on each accounted. Figure 3 shows a cell's service area divided into equal areas, before applying the method.

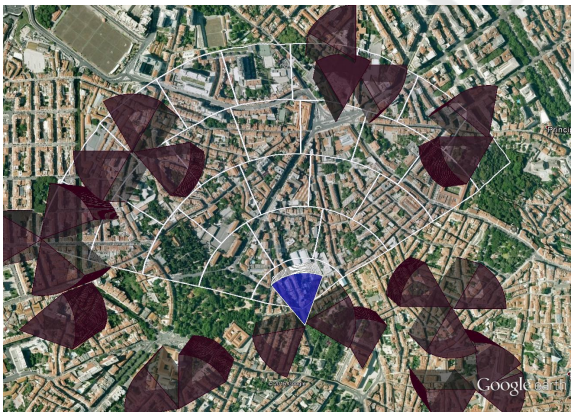

Figure 3: Cell's geographical area.

Considering the quadrant method data,

$$[Y(A_m)]_P = [x_1, ..., x_m] \tag{1}$$

where x_m is the number of measurements in the m quadrant, for all A_m quadrants covering the service area P. Using this data set, a parameter that quantifies the DT data pattern is calculated.

The parameter Variance to Mean Ratio (VMR) allows to identify the distribution pattern of a data set and it is given by,

$$VRM = \frac{s^2}{\bar{x}} \tag{2}$$

where s^2 is the variance of the number of measurements and $\bar{x}$ the average measurement number in each quadrant. When the VMR value is higher than one, the pattern is clustered. Below the unit is dispersed and if it's equal to one, the pattern is random. Due to the direct correlation between DT measurements and the own existence of roads, the pattern can't be random and hardly will ever be dispersed. So, the relevant information is how much clustered might be.

This knowledge is reflected in the Dispersion Index D_i for a cell i. Using Equation (2) over the data set given by Equation (1) the DT data dispersion in the cell's service area is calculated (VRM). The Dispersion Index D_i, results from the normalization of the DT data VMR, given by,

$$D_i = 1 - \frac{VRM}{VRM_{max}} \tag{3}$$

where VRM_{max} is calculated using Equation 2 in the case of all measurements being in one single area A, thus resulting in a relative value of the data distribution.

3.2 Road Filling Ratio for DT Measurements

The aim is to calculate the road filling ratio, P_i, which represents the percentage of road/street covered in the service area of cell i. In order to proceed, the road network information is fetched using an Application Programming Interface (API). The obtained points are linearly interpolated, for resolution purposes, resulting as shown in Figure 4.

An important piece of information, is the treatment that the original geospatial positioned measurements, enroll. They are subsequently aggregated on geographical areas of ten by ten meters, called bins. This term refers to several measurements from different cells aggregated in the same area, and will be mentioned throughout this paper. To refer to a single cell measurement contained in a bin, we refer as a cell sample. Bearing in mind this information, the same procedure must be applied to the retrieved road points. Only then P_i is calculates as,

$$P_i = \frac{M}{Max_m} \tag{4}$$

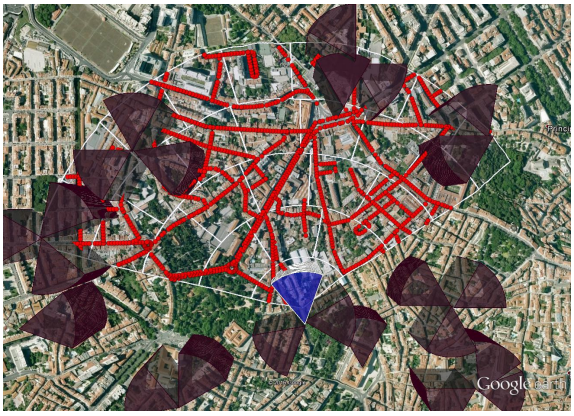

Figure 4: Street data information.

where M is the number of collected bins and Max_m the equivalent of bins for all the road/street extension in the service area of a cell i.

3.3 Calculating the DT Reliability Index

The DT Reliability Index R_i, gets values in the zero to one range, and it merges the previous two metrics, in the following way,

$$R_i = \alpha_d D_i + \alpha_p P_i \quad (5)$$

where α_d is the weight of the dispersion index D_i, for cell i and α_p is the weight of the percentage of road covered P_i, in cell i. Regarding the weights of both factors (α_d and α_p) they resulted from empirical knowledge, through a statistical analyze of fifty DT classification inquiries, performed to Celfinet's experienced radio engineers (Sousa et al., 2015).

4 THE SELF-DIAGNOSIS PROCESS

The underlying process for the identification of either coverage holes, overshooting or pilot pollution scenarios has a common ground. The goal is always, regarding the specific algorithm, to diagnose prevalent under performance situations in the form of clusters. Furthermore, to attribute a harshness metric based on a statistical analysis to each cluster, and a harshness level of the cell, due to all clusters found, as presented in Figure 5.

4.1 Self-Diagnosis Modules

As stated before, in this work three detection algorithms are presented. This subsection characterizes each algorithm, in terms of the data identification process.

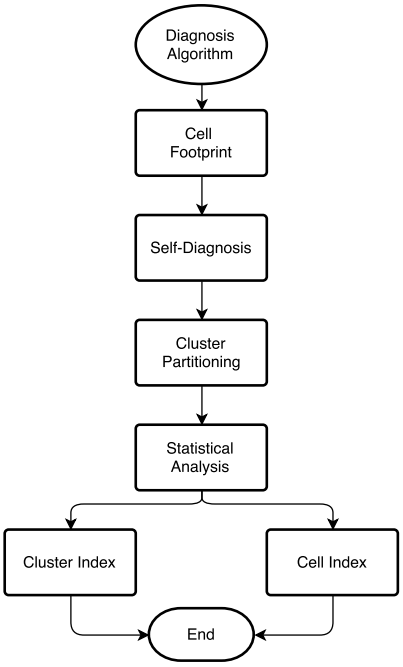

Figure 5: General detection flowchart.

4.1.1 Detecting Coverage Holes

A coverage hole is defined as an area where the pilot (or reference) signal power is in between the lowest network access threshold and the lowest value required for assigning full coverage. Users in this area tend to suffer poor voice or data user experience and possibly dropped calls or high latency.

From the cell's footprint, a coverage hole sample must accomplish two conditions. Firstly, it must be best server or within a small power interval to the best server of the respective bin. To these, we will refer as serving samples. Note that best server is the measurement that corresponds to the best cell, and that should serve the mobiles at that point. Secondly, the best server measurement from the respective serving sample bin, must be below the coverage threshold,

$$BS_{power} < Thr_{CH} \quad (6)$$

where BS_{power} is the best server value and Thr_{CH} is the minimum serving power value, which is tecnhology (3G 4G) and use case dependent.

4.1.2 Detecting Overshooting

An overshooting situation occurs when the cell's coverage reaches beyond what is planned. Generally, oc-

curs as an "island" of coverage in another cell's service area. Overshooted areas may also suffer from call drops and bad quality of experience.

From the cell's DT measurements, the ones considered in overshooting, must be located beyond the cell's service area. Then, to be considered overshooting, it must comply with the following conditions, regarding the power value measured,

$$\begin{cases} CS_{pwr} > Thr_{OS}, & \text{if } CS \text{ is best} \\ CS_{pwr} > Best_{pwr} - \Delta_{pwr}, & \text{if } CS \text{ is not best} \end{cases} \quad (7)$$

where CS_{pwr} is the cell sample power value, Thr_{OS} is the minimum power value to be considered overshooting, $Best_{pwr}$ is the power value of the bin best server cell and Δ_{pwr} is used to define a power range to consider overshooting. Two conditions are presented, for the case when the corresponding sample of the cell in analysis is the best server cell and for the opposite case.

Also, the quality of the cell's measurements is evaluated, using the same conditions as Equation 7 but regarding the quality measurments,

$$\begin{cases} CS_{qual} > QThr_{OS}, & \text{if } CS \text{ is best} \\ CS_{qual} > Best_{qual} - \Delta_{qual}, & \text{if } CS \text{ is not best} \end{cases} \quad (8)$$

Not all overshooting situations are necessarily damaging to the network or non-intended. Even though effectively being far from the normal service area, it might happen that, due to terrain profile, it still might be the cell in best condition to serve in that area. In that sense, in case of the majority of the overshooting bins, containing the best server as the analyzed cell and a delta value of power and quality to the second best server, these will also be marked. They will still be identified as overshooters and continue the process, but containing an observation that optimizing this overshooting area might reduce the overall network performance.

4.1.3 Detecting Pilot Pollution

Pilot pollution remarks a scenario where too many pilots (or reference signals in the case of LTE) are received in one area. Besides the excess of pilots, it lacks a dominant one. These areas are highly interfered, resulting in a poorer user experience.

The cell's measurements inducing pilot pollution, are in a serving ranking bellow the ideal maximum number of cells serving in that area. To the cell's footprint where the previous is verified, the following conditions, must be complied to be in a pilot pollution situation,

$$\begin{cases} CS_{pwr} > Best_{pwr} - \Delta_{pwr} \\ CS_{qual} > Best_{qual} - \Delta_{qual} \end{cases} \quad (9)$$

where CS_{pwr} is the cell sample power value, $Best_{pwr}$ is the power value of the bin best server cell and Δ_{pwr} is used to define a pilot pollution power range. The second condition is equal to the first, but regarding quality measurements.

4.2 Cluster Partitioning

The radio mobile channel is uncertain due mainly to the effects of fading and multipath. So, the values caught on DT measurements, which are reported at one single instance of time, may not correspond to the average behavior of the radio channel in that point. In order to mitigate this variability, the detection process is executed at the cluster level and not at bin level. This enables to detect prevailing under performance areas and not simply variations, normal and non-correlated with network issues. Therefore, using the auto-correlation distances for shadow fading (Kysti et al., 2007), to limit cluster size, this gives more assertiveness in the detected results.

Regarding the cluster division process itself, it is accomplished using a dendrogram structure (Izenman, 2008). It is a tree diagram that, in this application, translates the distance relation between all DT measurements detected, as can be seen in Figure 6. The tree diagram building process, aggregates succes-

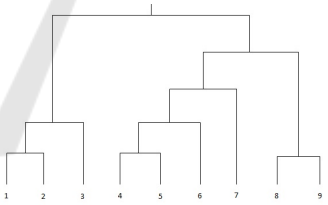

Figure 6: Dendrogram tree structure.

sively the closest bin/cluster pair until all of the bins form a unique cluster. At each aggregation, is constructed a different cluster division possibility.

The next step is to define which cluster arrangement is best. Several algorithms accomplish this purpose, and throw different metrics and approaches. But overall what they evaluate is how well concentrated are the points in clusters, when the distance similarity (Zheng and Xue, 2009) is high. The approach used was the silhouette method (Witten and Frank, 2005). It provides a metric $s(i)$,

$$s_i = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \quad (10)$$

where $a(i)$ is the average dissimilarity to the other points of the cluster, $b(i)$ the lowest average dissimilarity of i to any other cluster. It evaluates how well any given point lies within its cluster. Thus, an $s(i)$ close to one means that the bin is appropriately clustered. This approach is only applied to the cluster division possibilities that respect the maximum cluster size due to the correlation distance.

4.3 Cluster Statistical Analysis

A statistical analysis is conducted to each detected cluster. The purpose is to classify the harshness (severity of the problem) and rank by it. The result is a harshness index, $H_{cluster}$, per cluster, given by,

$$H_{cluster} = \frac{\sum_{i=1}^{N} \beta_i U(c(i))}{\sum_{i=1}^{N} \beta_i}, 0 \leq H_{cluster} \leq 1 \quad (11)$$

where β_i is the weight for condition i and the $U(c(i))$,

$$U(c(i)) = \begin{cases} 1, & \text{if } c(i) \text{ is full field} \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

, are the evaluated conditions. Once again these conditions are dependent on the behavior that is being analyzed.

The coverage hole algorithm calculates it's $H_{cluster}$ index (11) using the following conditions:

$$C(1) : Prob(Power_s \leq Power_{cov}) \geq Thr_{CH} \quad (13)$$

$$C(2) : Prob(Power_n \leq Power_{cov1}) \geq Thr_{CH} \quad (14)$$

where $Power_s$ is the cell's power measurement value, $Power_{cov}$ is a coverage threshold and Thr_{CH} is the minimum percentage of data samples that are below the coverage value, so that the condition is fulfilled. The variable $Power_n$ is the power measurement value of a neighbor cell, who also serves that area. The $Power_{cov1}$ is just another power value to be compared with. These conditions allow to classify the harshness of a coverage hole in two forms. The condition from Equation (13) concerns only to the power value of the source cell. Using different $Power_{cov}$ values and different percentages Thr_{Power} a cluster will be classified more or less harsh. The second condition, in Equation (14), evaluates the existence of other fallback cell in the cluster.

The coverage hole harshness index H_{cell} represents a percentage of clusters in coverage hole versus the clusters of the cell footprint.This metric might be devious, especially if the DT data are low in number. In that case, the metric will exceed the true percentage value. That's why the DT reliability index R is so important in terms of interpreting the results.

The overshooting cluster harshness evaluation proceeds with the following conditions:

$$C(1) : Prob(Power_s \geq Power_{OS}) \geq Thr_{OS} \quad (15)$$

$$C(2) : Prob(Power_s \geq Power_{OS1}) \geq Thr_{OS} \quad (16)$$

$$C(3) : Prob(Qual_{deg} \geq Qual_{degOS}) \geq Thr_{OS} \quad (17)$$

where $Power_{OS}$ and $Power_{OS1}$ are different power values to be compared with. The conditions from Equation (15) and Equation (16) are exactly the same, but using different $Power_{OS}$ values, evaluating the power level of the overshooting, allowing more resolution in distinguishing overshooting clusters in severe terms. The condition from Equation 17, evaluates the quality degradation caused by the existence of an overshooting cell (Sanchez-Gonzalez et al., 2013), in which $Qual_{deg}$ is the quality degradation caused and $Qual_{degOS}$ is a quality degradation threshold.

The overshooting harshness index H_{cell} represents the average percentage of overshooting clusters against the victim cells footprint, divided also in clusters. Once again, the importance of the DT reliability index must be highlighted.

Concerning the pilot pollution algorithm, the conditions to evaluate the harshness level are the following:

$$C(1) : Prob(Power_{best} \geq Power_{PP}) \geq Thr_{PP} \quad (18)$$

$$C(2) : Prob(Qual_{deg} \geq Qual_{degPP}) \geq Thr_{PP} \quad (19)$$

where $Power_{best}$ is the signal power of the best server measurement and $Power_{PP}$ is a threshold value to be compared with. The condition presented in Equation (18), reflects that, a pilot pollution scenario is as damaging to the network as higher is the power of the correspondent best servers. The second condition (in Equation (19)), as seen in the overshooting, gauge the quality degradation that the source cell induces on the best servers.

In case of the pilot pollution algorithm, being a scenario where cells affect another cell performance, the approach for the pilot pollution harshness index H_{cell} is the same as described in the overshooting module.

5 APPLYING THE ALGORITHM TO A LIVE NETWORK

The developed algorithms were applied in a live network, and for an urban scenario. 830 3G cells operating at different frequencies were self-diagnosed and performance metrics were computed. Extensive DT data corresponding to this area was used.

5.1 Overview Details

Regarding to the thresholds used in section 4.1) to identify the intended network behaviors, they are extremely dependent on mobile operator policies, requirements, type of service, etc., so a set of default parameters was used, as detailed in Table 1. The overall results are shown in Table 2. The columns ”CH”, ”OS” and ”PP” refer to coverage holes, overshooting and pilot pollution algorithms, respectively.

Table 1: Algorithms thresholds and inputs.

	CH	OS	PP
RSCP threshold [dBm]	-100	-95	-95
RSCP delta [dB]	3	6	6
Min EcN0 threshold [dB]	N/A	-10	N/A
EcN0 delta [dB]	N/A	6	4
Number of bins / cluster	10	10	10
Max cluster radius [m]	57,5	57,5	57,5
Pilot pollution window	N/A	N/A	3

The results on Table 2, reveal a considerable number of cells with performance malfunctions, which are currently under more detailed evaluation by radio optimization teams. In terms of the coverage hole index H_{cell}, it is the highest. The DT reliability index R is

Table 2: Self-diagnosis detected scenarios.

	CH	OS	PP
Analyzed cells	830		
Detected cells	13	35	64
Average clusters [#]	3	2	3
Average index H_{cell} [%]	38	11	11
Average index R [%]	37	45	55

the lowest, though. Which admits that the DT did not retrieve data from all cell service area, leading to an overestimated H_{cell} value. With regard to the interference detection algorithms, the average cell index H_{cell} was 11% with more quality of the respective DT data.

5.2 Coverage Holes

One of the detected coverage holes is illustrated in Figure 7. The blue cell is the diagnosed cell, with the red points corresponding to the cell's measurements in coverage hole, grouped in one cluster.

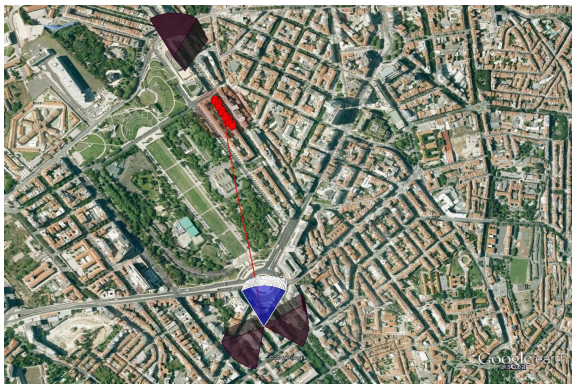

Figure 7: Coverage hole scenario.

The detailed Radio Frequency (RF) metrics are displayed in Table 3. It can be observed the -103 dBm

Table 3: Coverage hole measurements details.

	Number of bins	Average RSCP [dBm]	Average Ec/No [dB]
Cluster 1	14	-103	-9

low signal power level, on average, for the 14 detected measurements.

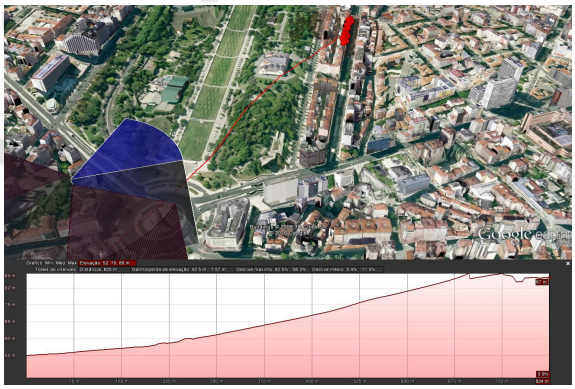

Figure 8: Coverage hole cluster terrain profile.

Using software with elevation profiling and 3D building modulation, see Figure 8, it can be asserted that the coverage hole area is in Non-Line-of-Sight (NLoS) due to building obstruction and terrain elevation.

In relation to the index H_{cell}, for this occurrence, it was 25%, meaning that 25% of the clusters are in coverage hole. Nevertheless, as the DT reliability index R for the cell was only 23%, it may indicate that the index S be off the real percentage.

5.3 Overshooting

Concerning the overshooting, Figure 9, illustrates an overshooting in an overlap area between two cells.

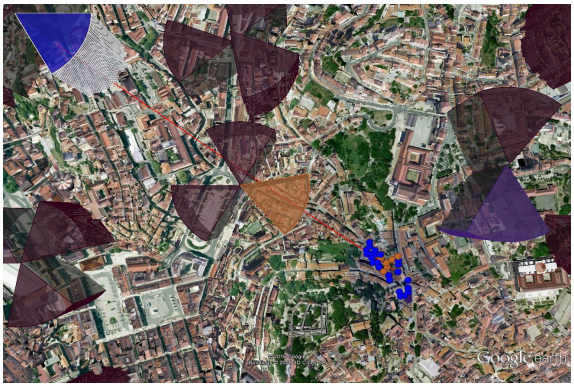

Figure 9: Overshooting scenario.

Again, the blue cell is the diagnosed cell, and in this case, it was detected an overshooting cluster. Furthermore, the cells with orange and purple, were identified as the victim cells, in the sense that, the overshooting cluster is located in their service areas. The measurement stats are shown in Table 4. The blue cell is reaching the other two cell service areas with a high power, causing interference.

Just for additional information, see Figure 10, where the terrain profile between the source cell and the respective overshooting cluster is illustrated, confirming the "high pass".

Table 4: Overshooting measurements details.

	Number of bins	Average RSCP [dBm]	Average Ec/No [dB]
Cluster 1	31	-69	-9

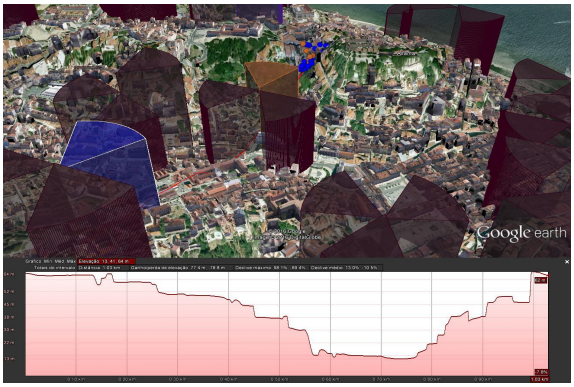

Figure 10: Overshooting cluster terrain profile.

It exists Line-of-Sight (LoS) between the source cell and the overshooting cluster which enables the overshooting scenario. In regard to the overshooting cell index, H_{cell}, it was obtained a value of 9%. This represents, in this particular case, that the overshooting clusters affected, on average, 9% of the victim cells footprint divided in clusters. Concerning the DT reliability index R, it was obtained an average value of 41% for the two victim cells.

5.4 Pilot Pollution

Regarding the pilot pollution detection algorithm, an example is illustrated in Figure 11. The diagnosed cell (in blue) is reaching an area, within the pilot pollution conditions (4.1.3), where the red, green and light purple cells are the serving (best) cells in that area. The detected pilot pollution bins, were arranged in one valid cluster.

Figure 11: Pilot pollution scenario.

The pilot pollution average metrics are presented in Table 5. Even for the high average power detected, the average quality level is very low, exhibiting how interfered is this area.

Table 5: Pilot pollution measurements details.

	Number of bins	Average RSCP [dBm]	Average Ec/No [dB]
Cluster 1	17	-77	-17

It can be seen in Figure 12 that the NLoS scenario is not enough to reduce the source power received. Due, mainly, to the relative small distance between the cluster and the source cell, besides that, this cell operates in the 900 MHz band.

In terms of the cell severity index H_{cell}, for this case the reported value, was 14%. It shows that the pilot pollution clusters are in average 14% of the detected serving cells footprint. In the matter of the DT reliability index of the victim cells, was retrieved an average value of 62%, which empowers greatly this

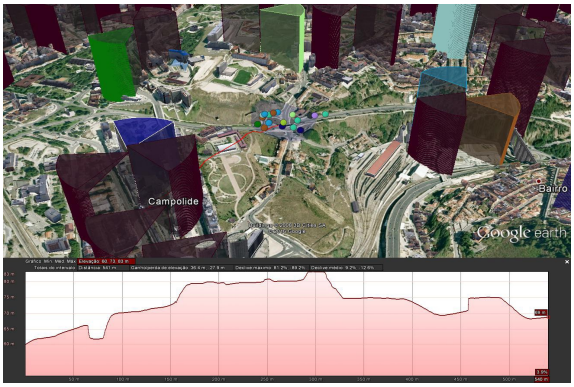

Figure 12: Pilot pollution cluster terrain profile.

pilot pollution H_{cell} index, compared to the overshooting harshness index.

By definition, pilot pollution areas display an excessive number of pilot signals. Sometimes, this number is high enough, that enables the algorithm to suggest which cell(s) are being victimized. In this scenario, the detection is done, but insufficient information retains the algorithm from the harshness evaluation and the affected cell identification.

6 CONCLUSIONS

This paper presented a new approach for automatic detection of low coverage and high interference scenarios (overshooting and pilot pollution) in UMTS /LTE networks. These algorithms, based on periodically extracted DT measurements, identify the problematic cluster locations and compute harshness metrics, at cluster and cell level, quantifying the extent of the problem.

The proposed algorithms were validated for a live network urban scenario. 830 3G cells were self-diagnosed and performance metrics were computed.

The results showed that for an urban area and with a high site density, the main optimization efforts rely on the interference mitigation and not coverage optimization. Moreover, the pilot pollution scenarios are the most prevalent.

The cluster division, with a RF correlation distance limitation, provides a simplification for the antenna physical parameter optimization algorithms, in the sense that, the optimization process can be reduced to the cluster's centroids.

Future work is in motion by adding self-optimization capabilities to the algorithms, which will automatically suggest physical and parameter optimization actions, based on the already developed harshness metrics.

ACKNOWLEDGEMENTS

This work was supported by the Instituto de Telecomunicações (IT) and the Portuguese Foundation for Science and Technology (FCT) under project PEst-OE/EEI/LA0008/2013.

REFERENCES

Duarte, D., Vieira, P., Rodrigues, A. J., and Silva, N. (2015). A new approach for crossed sector detection in live mobile networks based on radio measurements. In *Wireless Personal Multimedia Communications Symp. - WPMC*, volume 1.

Ericsson (2015). Ericsson mobility report. Technical report, ERiCSSON.

Izenman, A. J. (2008). *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. Springer Publishing Company, Incorporated, 1 edition.

Kysti, P., Meinil, J., Hentil, L., Zhao, X., Jms, T., Schneider, C., Narandzi, M., Milojevi, M., Hong, A., Ylitalo, J., Holappa, V., Alatossava, M., Bultitude, R., Jong, Y., and Rautiainen, T. (2007). Ist-4-027756 winner ii d1.1.2 v1.2. Technical report, EBITG, TUI, UOULU, CU/CRC, NOKIA.

Sallent, O., Perez-Romero, J., Sanchez-Gonzalez, J., Agusti, R., Diaz-Guerra, M., Henche, D., and Paul, D. (2011). Automatic detection of sub-optimal performance in umts networks based on drive-test measurements. In *Network and Service Management (CNSM), 2011 7th International Conference on*, pages 1–4.

Sanchez-Gonzalez, J., Sallent, O., Pérez-Romero, J., and Agustí, R. (2013). A multi-cell multi-objective self-optimisation methodology based on genetic algorithms for wireless cellular networks. *Int. Journal of Network Management*, 23(4):287–307.

Sousa, M., Martins, A., Vieira, P., Oliveira, N., and Rodrigues, A. (2015). Caracterizacao da fiabilidade de medidas rádio em larga escala para redes auto-otimizadas. In *9. Congresso do Comité Português da URSI - "5G e a Internet do futuro"*.

Vieira, P., Silva, N., Fernandes, N., Rodrigues, A. J., and Varela, L. (2014). Improving accuracy for otd based 3G geolocation in real urban/suburban environments. In *Wireless Personal Multimedia Communications Symp. - WPMC*, volume 1.

Witten, I. H. and Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Zheng, N. and Xue, J. (2009). *Statistical Learning and Pattern Analysis for Image and Video Processing*. Advances in Pattern Recognition. Springer.