

# Transferring Knowledge for Tilt-Dependent Radio Map Prediction

Claudia Parera\*, Alessandro E. C. Redondi\*, Matteo Cesana\*, Qi Liao<sup>†</sup>, Lutz Ewe<sup>†</sup> and Cristian Tatino<sup>‡</sup>

\*DEIB, Politecnico di Milano, Milan, Italy

<sup>†</sup>Nokia Bell Labs, Stuttgart, Germany

<sup>‡</sup>Linköping University, Linköping, Sweden

E-Mails: {claudia.parera, alessandroenrico.redondi, matteo.cesana}@polimi.it  
{qi.liao, lutz.ewe}@nokia-bell-labs.com, cristian.tatino@liu.se

**Abstract**—Fifth generation wireless networks (5G) will face key challenges caused by diverse patterns of traffic demands and massive deployment of heterogeneous access points. In order to handle this complexity, machine learning techniques are expected to play a major role. However, due to the large space of parameters related to network optimization, collecting data to train models for all possible network configurations can be prohibitive. In this paper, we analyze the possibility of performing a knowledge transfer, in which a machine learning model trained on a particular network configuration is used to predict a quantity of interest in a new, unknown setting. We focus on the tilt-dependent received signal strength maps as quantities of interest and we analyze two cases where the knowledge acquired for a particular antenna tilt setting is transferred to (i) a different tilt configuration of the same antenna or (ii) a different antenna with the same tilt configuration. Promising results supporting knowledge transfer are obtained through extensive experiments conducted using different machine learning models on a real dataset.

**Index Terms**—Radio map prediction, antenna tilt, machine learning, knowledge transfer

## I. INTRODUCTION

The fifth generation wireless networks (5G) are expected to support a series of unique features compared to current architectures, including more users, higher data rates, reduced latency and improved energy efficiency. To reach these goals, 5G will leverage the availability of dense and heterogeneous deployments coupled with novel technologies able to be dynamically managed both in a centralized or distributed manner. To cope with such a complex scenario, it is envisioned that machine learning tools will play a major role in enabling the transition from current mobile networks to the future 5G architecture [1]. By exploiting the constantly increasing availability of data from both network devices and user terminals, such tools will assist network operators in facing the increasing complexity in the setting of the control parameters for network optimization, forming the basis for automatic and smart network management techniques.

Among the many parameters that can be configured at the base station (BS), one of the most important is the antenna

tilt, that is the angle formed by the vertical direction of the antenna with the horizon. Antenna tilt can be realized either mechanically (by physically inclining the antenna up or down) or electrically (relying on beamforming techniques that steer the main beam of the antenna towards a desired vertical direction), or even as a combination of the two. The antenna tilt directly impacts on many performance measures of the cell served by a BS, such as coverage, signal strength, and inter-cell interference, which in turn determine the quality of service experienced by the end users.

From an operator perspective, being able to predict such performance measures without performing extensive trials or measurement campaigns is of key importance for at least two reasons: first, extensive measurement campaigns such as test driving are time consuming and costly. Secondly, even assuming that measurements can be obtained in a cheap way (e.g., directly from user terminals through crowdsourcing), this requires that all possible antenna tilts are tested, which can easily cause issues on the cell performance and decrease the perceived quality of service. Therefore, such tests could be realistically performed only for very short periods and during off-peak traffic hours, making the idea of obtaining data measurements from users unfeasible.

Given such difficulties, an option which may be particularly appealing to network operators is to transfer the knowledge acquired through a single measurement campaign (e.g., for a given antenna tilt setting) to a new domain (e.g., a new tilt setting) without needing to acquire a complete set of additional measurements. Such a learning paradigm, in which the data distributions of the training (source) and testing (target) sets are significantly different, can be seen as an example of *transfer learning* and has received increasing attention in the last few years [2].

In this paper, we study the possibility of performing such a knowledge transfer for the task of predicting the radio signal strength map of a particular BS cell. We start from a dataset of signal strength measurements obtained from real-life long term evolution (LTE) base stations and analyze the performance of several machine learning algorithms in two different scenarios where the knowledge acquired for a particular antenna tilt setting is transferred to (i) a different tilt configuration of the same antenna or (ii) a different antenna with the same tilt

This work is funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 643002.

configuration.

We show through extensive experiments that transferring knowledge allows for promising prediction performance and that the domain similarity (the difference between training and testing data distributions) plays a role in selecting the machine learning technique that obtains the best performance.

The rest of this paper is organized as it follows: Section II reviews related works in the area of network planning and optimization, with particular focus on those works dealing with antenna tilt. Section III describes in details the scenario under consideration, while Section IV focuses on the machine learning tools used for this work. Experiments and discussion on the obtained results are reported in Section V. Finally, Section VI concludes the paper and describes future work directions.

## II. RELATED WORK

Recent works have studied the importance of antenna tilt in the context of network planning and optimization problems.

The literature can be broadly categorized in two areas: (i) network optimization through antenna tilt adjustment and (ii) prediction of tilt-dependent radio maps. With reference to the former area, in [3], a method for finding the optimal antenna tilts in a heterogeneous network is proposed. The method is built on a reinforcement learning algorithm which adapts the antenna tilts to the specific network load conditions in order to maximize the user throughput fairness and the overall energy efficiency. Results on simulated data demonstrate that the proposed method outperforms a fixed strategy for the antenna tilts. A similar approach is taken by the authors of [4], where reinforcement learning is again used to optimize the antenna tilts, this time with the objective of maximizing the overall data rate of the network. Finally, the work in [5] proposes a general machine learning-based network planning tool. The flexibility of the approach is demonstrated with examples, including readjusting the antenna tilts to compensate for loss of service caused by faulty cells.

As for the second area of works, the work in [6] proposes a geometrical-based extension to different traditional log-distance path loss models (Okumura-Hata, Walfisch-Ikegami) to take into account the antenna tilt during the prediction of the signal strength at a given distance from the base station. The proposed extension, named vertical gain correction (VGC), is calculated directly from the antenna patterns provided by the manufacturer and is added to the signal strength estimated by the path loss models to compensate for the antenna tilt. Experimental results on data collected from LTE base stations show that the VGC improves the signal strength prediction performance compared to traditional models. Such method is then used to compute predictions that are used in a following work [7] to optimize the antenna tilts in order to maximize the capacity and coverage of a simulated LTE network. The predictive performance are reported to be lower for locations close to the antenna. Similarly, the work in [8] investigates the effect of antenna tilt on radio maps, comparing the path loss models developed by 3GPP for different propagation environments [9]

with the results obtained by a ray tracing tool able to take into account antenna tilts. Results demonstrate that changing antenna tilt has a significant impact on the shadowing map, therefore calling for a rethinking of currently available 3GPP propagation models and assumptions, which apply identical shadowing map independently from the antenna tilt.

Our work shares the same research objectives of this second class with one fundamental novelty: in all the aforementioned works, the domain of the data used for predicting the signal strength is similar (if not the same) to the target domain; as an example, the signal strength radio map of an antenna under a given tilting configuration is interpolated or predicted out of available signal strength samples collected for the same antenna in the same tilting configuration. We also analyze the cases where the performance of target antenna configuration is predicted by using training dataset collected under different configurations or even referring to a totally different antenna.

## III. PROBLEM STATEMENT AND BACKGROUND

In this paper we address the following problem in mobile radio networks: *"how to predict the performance of a given network configuration by leveraging performance information of diverse network configurations"*. The performance measure that we target here is the received signal strength in the downlink and the network configuration domains include the tilting configuration of the emitting base stations.

Assume  $K$  base stations are deployed in the area under analysis. Each base station can work in  $H$  different tilt configurations, indexed by  $h = 1, \dots, H$ . Let  $s_{k,h}(\mathbf{x}_i)$ , be the measured signal strength received at location  $\mathbf{x}_i = \{y_i, z_i\}$ , from the  $k$ -th base station when running the  $h$ -th tilt configuration, where  $y_i$  and  $z_i$  indicate the latitude and the longitude of the  $i$ -th location, respectively. Let  $\mathcal{M}_k^h$  be the set of location indexes where measurements for base station  $k$  running configuration  $h$  have been taken.

The problem at hand can be defined as follows: given  $\{s_{k,h}(\mathbf{x}_i) : i \in \mathcal{M}_k^h\}$ , estimate the unknown signal strength  $\hat{s}_{m,n}(\mathbf{x}_j)$  at the same or different locations,  $\mathbf{x}_j$ , under diverse and different network configuration domains, that is  $\mathbf{x}_j, j \in \mathcal{M}_m^n$  with  $m \neq k$  and/or  $n \neq h$ .

The dataset used in this work is composed of reference signal received power (RSRP) outdoor measurements collected in Espoo, Finland in November 2016 out of two LTE commercial base stations with three different 120° sectors each and operating at 2.6 GHz (see Fig. 1 reporting the positions of the two antennas and the representation of the target area). Out of the two antennas, we focus here on three physical cell identifiers (PCIs) which will be referred to as PCI 1, 2 and 3. PCIs 1 and 2 refer to two different sectors of the same base stations, whereas PCI 3 is a sector of a totally different base station.

The RSRP measurements were collected using an Android device equipped with an application capable of storing the RSRP from all the received cells, the cell identifier, the GPS position of the device and the timestamp. Such measurements were carried out at a frequency of 1 Hz while walking along

PCI	total tilt setting (in degrees)
1	3,5,6,7,9
2	3,5,6,7,9
3	5,6,9

TABLE I: Available down-tilt configurations.



Fig. 1: Map showing the base stations position and the PCIs (sectors) in the reference dataset.

a route of about 8 km within each cell coverage area, with a minimum and maximum distance from the base station of 30 m and 900 m respectively. The testing paths were designed to include varied propagation conditions: university campus with two- or three-story buildings, residential areas, parking lots, lower density rural and open area with different types of roads (i.e. pedestrian, cycling and main roads). Each testing path was walked once for each electronic tilt setting. The different tilt configurations are shown in Table I. The receiver was placed at the height of 1.5 m and always kept with the same orientation. The weather conditions were stable and cloudy and the road was covered by snow for most of the measurement campaign. In total, about  $3 \cdot 10^5$  RSRP measurements were obtained. Each observation contains the following fields: the measurement position, the RSRP value, the corresponding PCI and the timestamp. The raw dataset was pre-processed to remove outliers and corrupted samples: as an example, at the beginning of each experiment the GPS receiver takes some time to set up, recording incorrect position. Moreover, the RSRP values are averaged over grid segmentation of the map, with the grid size of  $20 \text{ m} \times 20 \text{ m}$ . After the pre-processing steps, the dataset was reduced to about  $3.5 \cdot 10^3$  observations per PCI and per tilt configuration, for a total of about  $5 \cdot 10^4$  measurements.

#### IV. PREDICTION APPROACHES

In this section, we describe two different prediction approaches. The first one leverages spatial information only (latitude and longitude) to predict the signal strength; conversely, the second approach adopts a larger set of features to capture path loss and propagation properties.

##### A. Location-Only Approaches

Based on the strong spatial correlation between signal strength and location, three different techniques have been used:

*Baseline:* The *Baseline (B)* method takes as the predicted value of signal strength for a target position  $\mathbf{x}$  the signal strength of the closest point  $\mathbf{x}_i$  available in the training dataset  $\mathcal{M}$ . This method is essentially  $k$ -nearest neighbors ( $k$ -NN) with  $k = 1$ . Formally,

$$\hat{s}(\mathbf{x}) = s(\mathbf{x}_i), \quad \mathbf{x}_i = \arg \min_{\mathbf{x}_i, i \in \mathcal{M}} d(\mathbf{x}, \mathbf{x}_i). \quad (1)$$

*Adjusted Baseline:* As proposed in [6], the *Baseline* method can be extended by leveraging *a priori* information on the radiation pattern of the reference antenna. Formally, the *Adjusted Baseline (AB)* predicts the value of the signal strength as:

$$\hat{s}(\mathbf{x}) = s(\mathbf{x}_i) + \Delta_H(\mathbf{x}, \mathbf{x}_i) + \Delta_V(\mathbf{x}, \mathbf{x}_i), \quad \mathbf{x}_i = \arg \min_{\mathbf{x}_i, i \in \mathcal{M}} d(\mathbf{x}, \mathbf{x}_i), \quad (2)$$

being  $\Delta_H(\mathbf{x}, \mathbf{x}_i)$  and  $\Delta_V(\mathbf{x}, \mathbf{x}_i)$  the difference of the antenna horizontal and vertical gain respectively in the directions towards the target position  $\mathbf{x}$  and the closest known position  $\mathbf{x}_i$ . Formally,

$$\Delta_H(\mathbf{x}, \mathbf{x}_i) = \eta(\mathbf{x}) - \eta(\mathbf{x}_i) \quad \text{and} \quad \Delta_V(\mathbf{x}, \mathbf{x}_i) = \gamma(\mathbf{x}) - \gamma(\mathbf{x}_i),$$

where  $\eta(\cdot)$  and  $\gamma(\cdot)$  are the horizontal and vertical gain of the antenna respectively.

*k-Nearest Neighbors with Inverse Distance Weighting:* This technique extends the classical nearest neighbor approaches [10] and predicts the signal at an unknown target location as a weighted average of the signals at the  $k$  closest known locations:

$$\hat{s}(\mathbf{x}) = \sum_{i \in \mathcal{M}(\mathbf{x})} w_i s(\mathbf{x}_i). \quad (3)$$

The set  $\mathcal{M}(\mathbf{x})$  includes the indexes of the  $k$  locations which are geographically closest to the target unknown location  $\mathbf{x}$ . Weights  $w_i$  are chosen to be inversely proportional to the distance  $d(\mathbf{x}_i, \mathbf{x})$  and normalized to sum to unity, that is:

$$w_i = \frac{d(\mathbf{x}_i, \mathbf{x})^{-1}}{\sum_{j \in \mathcal{M}(\mathbf{x})} d(\mathbf{x}_j, \mathbf{x})^{-1}}. \quad (4)$$

##### B. Geometric-aware Approaches

Location-only approaches tend to produce good results under ideal conditions, when the available labeled dataset is rich enough and the reference area is sampled uniformly. On the other hand, when the available dataset is sparsely sampled in space, then the location-only approaches may fail. Therefore, we have resorted to more complex prediction models leveraging additional features besides location. Namely, being  $\mathbf{x}_A$  the antenna location and  $\mathbf{x}$  the target position, the following set of features is considered for the prediction task (see Figure 2):

- the physical *distance* between the antenna and the measurement position,  $d(\mathbf{x}, \mathbf{x}_A)$ ;
- the *relative elevation angle* between the down-tilt of the antenna and the vertical direction from the antenna emitting element to the measurement position, defined as:

$$\delta_1 = 90^\circ - (\alpha_A + \alpha_E), \quad (5)$$

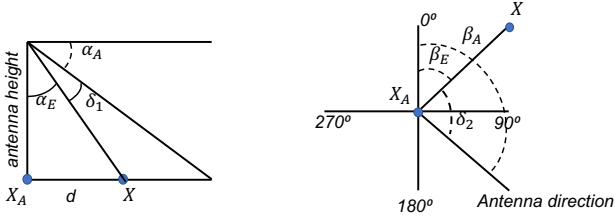


Fig. 2: Relative angles in the vertical (left) and horizontal (right) planes between the antenna pointing direction and the direction towards the test position  $\mathbf{x}$ .

being  $\alpha_A$  the antenna down-tilt (mechanical plus electrical) and  $\alpha_E$  the angle at which the antenna “sees” the target position;

- the *relative azimuth* between the horizontal orientation of the antenna and the horizontal direction to the measurement position defined as:

$$\delta_2 = \beta_A - \beta_E, \quad (6)$$

being  $\beta_A$  the horizontal orientation of the antenna and  $\beta_E$  the horizontal orientation of the target position with respect to the antenna position.

- the *tilt* configuration of the antenna,  $\alpha_A$ .

Each signal strength sample in the training dataset is therefore associated to a feature vector  $\mathcal{X} = \{1, d, \delta_1, \delta_2, \alpha_A\}^T$ , after applying the logarithm transformation to  $d$ , given that RSRP samples are measured in dBm.

*Multivariate Linear Regression:* A linear model can be trained to relate the measured signal strength in a position to the corresponding feature vector. Given a target position  $\mathbf{x}$ , the predicted signal strength is modeled as:

$$\hat{s}(\mathbf{x}) = \Theta^T \mathcal{X},$$

being  $\Theta$  the parameter vector of the linear model.

*Random Forest:* *Random forest (RF)* is one of the ensemble methods used for classification and regression purposes. The algorithm introduced by Ho [11] in 1995, and later extended by Breiman and Cutler [12], uses the idea of bagging to perform predictions. During the process several trees are grown independently using different bootstrapped samples of the data and majority voting or averaging are used for the final prediction. In contrast to traditional trees, the variable used to perform the split in each node is chosen randomly from a set of predictors [12]. RF is known to sometimes outperform other machine learning techniques such as Neural Networks due to its resistance to overfitting [13].

*XGBoost:* Boosting is a technique that can be used to improve the performance of a generic machine learning algorithm, by iteratively tuning a model each time giving more importance to mispredicted test samples [14]. Here we use a particular version of boosting called XGBoost [15], which unifies several ideas: Gradient Boosting, regularization to avoid over fitting, column sampling taken from RF and sparse data manipulation.

## V. EXPERIMENTS

Inspired by the concept of *transfer learning* [2], we evaluate here the performance of different prediction approaches when varying the *degree of similarity* of the datasets (domains) used for training and testing. The quantitative measure to capture domain similarity used in the analysis is the Kullback-Leibler divergence index (KL) [16] which measures the relative entropy of a given probability distribution with respect to another one. Namely, given two reference datasets, one used for training and one used for testing, we derive the KL divergence indexes of the probability distributions of the logarithm of the distance ( $d$ ), relative angle ( $\delta_1$ ) and relative azimuth ( $\delta_2$ ) of the two datasets. In details, the range of distances, relative angles and relative azimuth contained in the two datasets was uniformly quantized to  $k$  intervals, further deriving the related discrete probability distributions. For example, the symmetric KL divergence index of the distance probability distributions in one training and one testing dataset is given by:

$$D_{KL}(d) = \sum_{i=1}^k P_d^{(tr)}(i) \log \frac{P_d^{(tr)}(i)}{P_d^{(te)}(i)} + \sum_{i=1}^k P_d^{(te)}(i) \log \frac{P_d^{(te)}(i)}{P_d^{(tr)}(i)}, \quad (7)$$

where  $P_d^{(tr)}(i)$  and  $P_d^{(te)}(i)$  with  $i = 1 \dots k$  defining the discrete probability distributions of the distance in the training and testing dataset respectively. Similar definitions hold for the KL divergence indexes related to the relative angle,  $\delta_1$  and relative azimuth  $\delta_2$ . Finally, to give a more succinct representation of domain similarity, we introduce the domain distance measure (DD) by summing the three indexes together, that is:

$$DD = D_{KL}(d) + D_{KL}(\delta_1) + D_{KL}(\delta_2). \quad (8)$$

Prediction quality is assessed through the mean absolute error (MAE) and the mean absolute percentage error (MAPE), defined in Eq. (9), Eq. (10), respectively, where  $n$  is the number of target positions in the testing dataset.

$$\text{MAE}(s, \hat{s}) = \frac{1}{n} \sum_{i=0}^{n-1} |s_i - \hat{s}_i| \quad (9)$$

$$\text{MAPE} = \frac{100}{n} \sum_{i=0}^{n-1} \left| \frac{s_i - \hat{s}_i}{s_i} \right|. \quad (10)$$

### A. Tilt-to-Tilt Knowledge Transfer

In *Tilt-to-Tilt* knowledge transfer, we use a dataset obtained under a given tilt setting to predict the performance of the same antenna under a different tilt configuration. Table II reports the results obtained by the different prediction approaches described in Section IV when considering PCI 1 and different combinations for the training and testing tilt conditions. The text in the table header explains the knowledge transfer involved in the experiment: as an example, the first two columns (2→3) refer to the case where the training set was obtained with an antenna tilt of 2 degrees and the tilt configuration for the test set was 3 degrees.

	Tilt 2 ->Tilt 3		Tilt 4 ->Tilt 3		Tilt 6 ->Tilt 3	
	DD 0.21		DD 0.55		DD 2.58	
	MAE	MAPE%	MAE	MAPE%	MAE	MAPE%
<b>B</b>	4.02	4.03	5.27	5.23	7.15	7.31
<b>AB</b>	4.45	4.47	7.49	7.64	11.03	10.86
<b>k-NN</b>	3.9	3.92	5.26	5.19	6.69	6.83
<b>LR</b>	7.75	8.2	7.6	8.1	8.15	8.83
<b>RF</b>	3.89	4.11	5.41	5.38	6.06	6.02
<b>XGB</b>	4.33	4.41	4.68	4.89	6.25	6.59

TABLE II: Prediction performance of different algorithms under *tilt-to-tilt* knowledge transfer for reference PCI 1.

From the results in Table II, we can make the following observations.

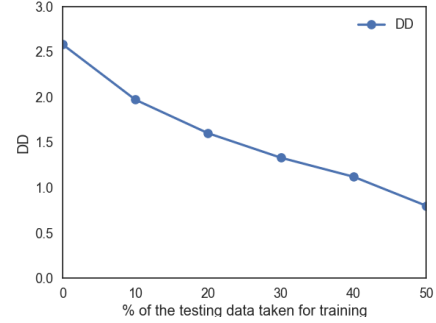
- The prediction error is impacted by the training and test domain distance. As the domain distance increases, the prediction performance decreases for all tested algorithms.
- Notably, location-only approaches (first three rows in Table II) and geometric-aware prediction approaches (last three rows) perform similarly when domain distance is small, whereas geometric-aware approaches have better performance when the domain distance increases. This is because, geometric-aware approaches better capture the physical properties of channel propagation, thus being more robust against the cross-domain missing information.
- RF generally provides a good prediction performance, regardless of the domain difference.

Similar results obtained for other PCIs are not reported here for the sake of brevity.

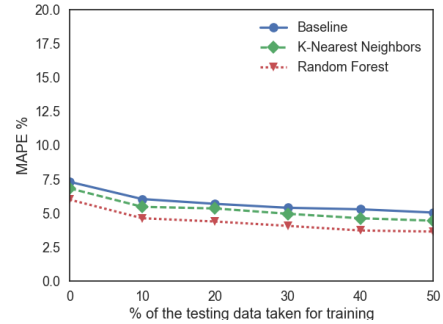
To further investigate the impact of the domain distance measure, we have considered the cases where the training dataset is integrated with samples taken from the test domain. Figure 3 refers to the cases where increasing percentages of the training dataset samples (tilt configuration 6) are replaced by samples from the testing domain (tilt configuration 3). Sampling is performed randomly and results are averaged over 10 different trials. Fig. 3a and Fig. 3b report the value of the Domain Distance measure and the MAPE when varying the percentage of testing samples added to the training datasets. All tested methods show decreasing prediction error as the distance between domains decreases.

### B. PCI-to-PCI Knowledge Transfer

In case of PCI-to-PCI knowledge transfer, a dataset referring to one antenna and a given tilt configuration is used to train the models to predict the signal strength of another antenna in the same tilt configuration. Table III reports the same information as Table II when training and test datasets refer to different antennas (PCIs). In general, the prediction error is higher with respect to the case of *Tilt-to-Tilt* knowledge transfer as training and test datasets are more dissimilar. As already observed in case of *Tilt-to-Tilt* knowledge transfer, the error is smaller when the domain difference is smaller for all the tested algorithms.



(a) Domain Distance (DD)



(b) MAPE

Fig. 3: Relationship between Domain Distance and MAPE (a), (b): training on tilt 6, testing on tilt 3 for PCI 1.

	PCI 1 ->PCI 2		PCI 3 ->PCI 2	
	DD 1.89		DD 1.76	
	MAE	MAPE%	MAE	MAPE%
<b>B</b>	8.41	8.63	10.27	11.09
<b>AB</b>	8.41	8.63	10.27	11.09
<b>k-NN</b>	8.2	8.43	10.27	11.08
<b>LR</b>	11.39	11.21	8.53	8.95
<b>RF</b>	11.37	11.2	8.58	8.85
<b>XGB</b>	11.06	10.97	8.88	9.02

TABLE III: Prediction performance of different algorithms under *PCI-to-PCI* knowledge transfer for reference tilt 2.

The same strategy of taking samples of the testing data to train the model can be applied when transferring knowledge from PCI to PCI. Figures 4a and 4b report the value of the Domain Distance measure and the average prediction error when varying the percentage of samples from the testing domain (PCI 2) added to the training datasets (PCI 1). Figures 4c and 4d report the same graphs for the case of training over PCI 3 and testing over PCI 2. In both cases the average error decreases sharply when adding 20% of the testing dataset to the training dataset. Notably, the prediction performance levels off and does not improve much when going beyond 20%.

## VI. CONCLUSIONS

In this paper we addressed the problem of predicting the signal strength in the downlink of a real LTE network where the antennas can be tuned to operate with different tilting antenna configurations. Different prediction approaches

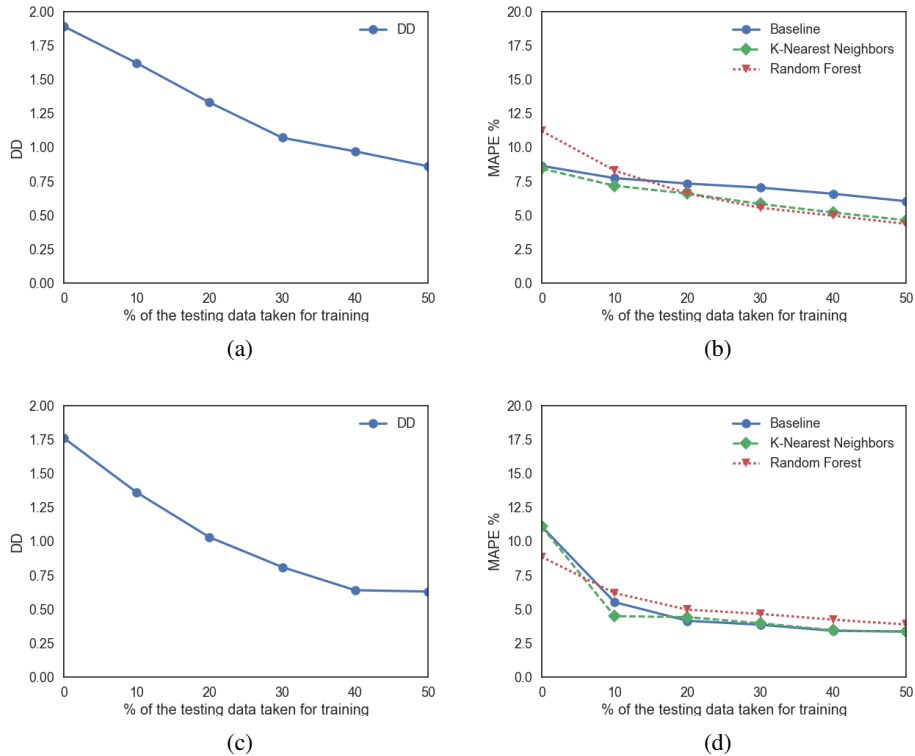


Fig. 4: DD vs MAPE. (a),(b): training on PCI 1, testing on PCI 2. (c),(d): training on PCI 3, testing on PCI 2 at tilt 2.

were considered with increasing complexity, starting from models/approaches only leveraging location information to predict the signal strength up to models/approaches based on more refined features related to propagation and antenna configuration. As opposed to other works in the field of radio map inference, we studied the quality of prediction of the aforementioned approaches when the datasets used for training and testing have different statistical characteristics. We observed that the performance of the predictive models is highly dependent on the difference between data distributions of the training and testing domain. Thus, analyzing domain similarity plays a crucial role in selecting the best performing model. Furthermore, two approaches are applied to further reduce the domain difference: (i) choosing the training set obtained from a tilt setting with higher similarity to the testing domain, and (ii) adding to the training set a limited number of samples from the testing domain. Future work will analyze possible strategies to achieve good tradeoff between improving prediction performance and reducing measuring effort for cross-domain data collection.

#### REFERENCES

- [1] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Communications*, vol. 24, no. 2, pp. 98–105, 2017.
- [2] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [3] W. Guo, S. Wang, Y. Wu, J. Rigelsford, X. Chu, and T. O'Farrell, "Spectral-and energy-efficient antenna tilting in a hetnet using reinforcement learning," in *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*. IEEE, 2013, pp. 767–772.
- [4] N. Dandanov, H. Al-Shatri, A. Klein, and V. Poulkov, "Dynamic self-optimization of the antenna tilt for best trade-off between coverage and capacity in mobile networks," *Wireless Personal Communications*, vol. 92, no. 1, pp. 251–278, 2017.
- [5] J. Moysen, L. Giupponi, and J. Mangues-Bafalluy, "A mobile network planning tool based on data analytics," *Mobile Information Systems*, vol. 2017, 2017.
- [6] I. Rodriguez, H. C. Nguyen, T. B. Sørensen, J. Elling, M. B. Gentsch, M. Sørensen, L. Kuru, and P. Mogensen, "A geometrical-based vertical gain correction for signal strength prediction of downtilted base station antennas in urban areas," in *Vehicular Technology Conference (VTC Fall), 2012 IEEE*. IEEE, 2012, pp. 1–5.
- [7] H. C. Nguyen, I. Rodriguez, T. B. Sorensen, J. Elling, M. B. Gentsch, M. Sorensen, and P. Mogensen, "Validation of tilt gain under realistic path loss model and network scenario," in *Vehicular Technology Conference (VTC Fall), 2013 IEEE 78th*. IEEE, 2013, pp. 1–5.
- [8] D. W. Kifle, B. Wegmann, I. Viering, and A. Klein, "Impact of antenna tilting on propagation shadowing model," in *Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th*. IEEE, 2013, pp. 1–5.
- [9] 3GPP, "Evolved universal terrestrial radio access (e-utra); further advancements for (e-utra) physical layer aspects," *TR 36.814, Technical report*, 2006.
- [10] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [11] T. K. Ho, "Random decision forests," in *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, vol. 1. IEEE, 1995, pp. 278–282.
- [12] A. Liaw, M. Wiener *et al.*, "Classification and regression by randomforest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.
- [13] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [14] —, "Arcing the edge," Technical Report 486, Statistics Department, University of California at Berkeley, Tech. Rep., 1997.
- [15] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. ACM, 2016, pp. 785–794.
- [16] J. M. Joyce, "Kullback-leibler divergence," in *International Encyclopedia of Statistical Science*. Springer, 2011, pp. 720–722.