INL/CON-23-75545-Revision-1

Wireless Heatmap Prediction with Deep Gaussian Process Models

May 2024

Yanyu Hu, Xiang Zhang, Amitabh Mishra, Imtiaz Nasim, Joshua E Daw, Shannon Leigh Eggers, Mingyue Ji, Vivek Agarwal

Changing the World's Energy Future

INL is a U.S. Department of Energy National Laboratory operated by Battelle Energy Alliance, LLC

INL/CON-23-75545-Revision-1

Wireless Heatmap Prediction with Deep Gaussian Process Models

Yanyu Hu, Xiang Zhang, Amitabh Mishra, Imtiaz Nasim, Joshua E Daw, Shannon Leigh Eggers, Mingyue Ji, Vivek Agarwal

May 2024

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

A Bayesian Learning Approach to Wireless Outdoor Heatmap Construction using Deep Gaussian Process

Xiang Zhang*[1], Yanyu Hu*[1], Imtiaz Nasim†, Amitabh Mishra†, Joshua Daw†,
Shannon Eggers†, Vivek Agarwal†, Arupjyoti Bhuyan†, Sneha Kasera Kumar§ and Mingyue Ji*
Department of Electrical & Computer Engineering, University of Utah*
Idaho National Laboratory†
Kahlert School of Computing, University of Utah§
Email: *{xiang.zhang, yanyu.hu, mingyue.ji}@utah.edu,
†{imtiaz.nasim, amitabh.mishra, joshua.daw, shannon.eggers, vivek.agarwal, arupjyoti. bhuyan}@inl.gov,
§kasera@cs.utah.edu

***Abstract*—We present a novel Bayesian learning approach to outdoor radio heatmap construction utilizing deep Gaussian process (GP). The proposed approach employs a two-layer hierarchy which consists of two cascaded Gaussian processes that are capable of modeling more complex input-output relations than standard single-layer Gaussian processes. Since deriving the exact model likelihood is challenging, a lower bound is optimized instead so that gradient descent-based methods can be performed to find out the optimal model parameters. Typically, inducing points are used in GPs to facilitate low-rank approximation of covariance (kernel) matrices for computation speedup. However, the inaccuracy induced by inducing points can accumulate when stacking multiple layers of GP which may hinder the performance of deep GP. Moreover, since inducing points need to be learned, having them at all layers of deep GP also incurs computational burden. To overcome the above challenges, in contrast to the canonical deep GP model, we use a modified architecture where a full standard GP resides in the first layer and inducing points are only introduced for the second layer. This modified architecture strikes a balance between model accuracy and training complexity. In the proposed model, the noise parameter of the first GP layer is also eliminated to improve the training efficiency as the noise parameter at the output of the second layer suffices to model the uncertainty in the output. The proposed approach is evaluated on real-world datasets, in the form of location-Received Signal Strength (RSS) pairs, collected from the Platform for Open Wireless Data-driven Experimental Research (POWDER) located at the campus of the University of Utah. Experiment results show that the proposed approach can achieve smaller prediction errors on various training and testing data configurations than DNN-based and GP-based methods.**

I. Introduction

Wireless radio heatmap [1]–[3] is an intuitive tool that helps to analyze radio signal strength in both indoor and outdoor environments. The heatmap is usually constructed using received signal strength (RSS) samples which offers real-time and comprehensive monitoring of the wireless signal strength distribution and radio device activities. Conventionally, explicit path loss models are used for RSS prediction [4]–[7]. For example, the dual-slope model is used in [4]. Due to the complexity of channel modeling, these approaches face challenges in achieving satisfactory levels of RSS prediction accuracy given the variety of wireless propagation environments. More recently, machine learning (ML)-based approaches are explored in RSS prediction due to their ability to adapt to complex radio propagation environments [8]–[11]. In these approaches, available RSS measurements are used to train a relevant ML model which is then used for RSS prediction other than the measurement locations. For example, a deep neural network (DNN)-based approach is used by Zhang *et al.* [9] to model the intricate input-output relation in the RSS data. In addition, convolutional neural network (CNN)-based approach are also leveraged by Imai *et al.* [10] and Levie *et al.* [11] respectively for RSS prediction.

ML-based, especially DNN-based approaches, demonstrate great potential when there is adequate training data. However, in reality, wireless heatmap construction is constrained by limited data availability since RSS measurements can be costly in many scenarios. With limited data, DNNs are known to suffer from overfitting [12], hindering the RSS prediction performance. Moreover, RSS prediction poses special challenges to DNN-based approaches as DNNs do not have an explicit mechanism to model the uncertainty caused by noise (i.e., channel and measurement noise) in the RSS measurements. To overcome the above issues, the Gaussian process (GP) model [13] is proposed, which is a type of Bayesian learning method to wireless heatmap prediction [14]–[17]. Gaussian process regression [18] is a nonparametric probabilistic learning paradigm where the GP output (i.e., observation) is modeled as a multivariate Gaussian distribution governed by a covariance (kernel) matrix that is determined by the inputs to the GP. A set of i.i.d. Gaussian noise is also included in the GP model which simulates the uncertainty in the observation. In the context of heatmap prediction, geological locations and RSS measurements respectively serve as the input to and observation from the GP. For example, in the pioneering work of Schwaighofer *et al.* [14], GPs with Martérn kernels are used to conduct cellular user positioning using RSS measurements generated by base stations. Moreover, Miyagusuku *et al.* [15] propose to enhance the performance of GPs with path loss models as a prior, which is experimentally verified to achieve better model variance prediction.

[1]Equal contribution.

Compared to standard GPs, deep Gaussian process [19] is a more powerful model by stacking multiple GPs together that can learn more complex location-RSS mappings. As a result, deep GP is also employed for heatmap prediction [20], [21]. For example, Wang *et al.* [20] use a canonical two-layer deep GP model from [19] to carry out an indoor localization task based on RSS measurements from WiFi access points. In addition, a deep convolutional GP is used for millimeter-Wave outdoor localization based on beamforming images [21].

In this paper, we propose a novel deep GP-based wireless outdoor heatmap construction approach using a *customized* two-layer deep GP with several adjustments over the canonical model architecture used in [20] in order to achieve a better trade-off between model accuracy and complexity. First, we use a full Gaussian process for the first layer and inducing points (IPs) are only introduced for the second layer. This is in contrast to the canonical model [20] where (the same number of) IPs are used at both layers to reduce training complexity through low-rank approximation of the kernel matrices, at the cost of an increased inaccuracy in model likelihood estimation. Preliminary experimentation shows that using a full GP at layer 1 helps to improve model performance at a slightly higher training complexity. Second, the Gaussian noise parameter in the first layer is eliminated (need not to learn this parameter) in the proposed model to improve training efficiency. This elimination causes no visible degradation to the model capability because the noise parameter at the second layer suffices to model the uncertainty in the RSS measurements. We then derive a corresponding evidence lower bound on the logarithm of the true model likelihood which can be optimized using gradient descent-based methods. In addition, we conduct extensive experiments using real-world RSS data, in the form of location-RSS pairs, collected from the Platform for Open Wireless Data-driven Experimental Research (POWDER) [22], a campus-scale wireless research platform located at the campus of the University of Utah. The proposed approach is evaluated on various training and testing data configurations representing different levels of training data availability. Experiment results show that the proposed approach can outperform GP and DNN-based baseline schemes, demonstrating its capability in dealing with scarce data.

II. Problem Description and Proposed Approach

We aim to learn the mapping between the measurement locations and the RSS values using the proposed deep Gaussian process model. In particular, the input to the deep GP model is the 2-dimensional coordinates of the measurement locations and the output/observation of the model is the corresponding RSS values. The RSS dataset is stored in the form of location-RSS value pairs and will be divided into a training and a testing subset for the purpose of model training and prediction respectively. Before proceeding to the description of the proposed approach, we first provide a brief overview of GP and deep GP regression models.

A. An Overview of GP and Deep GP

In standard single-layer GP [23], the input $Z \in \mathbb{R}^{N\times Q_Z}$ and observation $\mathbf{y} \in \mathbb{R}^N$ relation is governed by a Gaussian process where each input $\mathbf{z}_n \in \mathbb{R}^{Q_Z}, n = 1, ..., N$ is mapped to an output function value $f(\mathbf{z}_n) \sim \mathcal{GP}(0, k(\mathbf{z}_n, \mathbf{z}_n))$ where $k(\mathbf{z}_n, \mathbf{z}_{n'})$ is the covariance function that measures the correlation between the output function values $f(\mathbf{z}_n)$ and $f(\mathbf{z}_{n'})$. Usually, the automatic relevance determination (ARD) kernel $k(\mathbf{z}_n, \mathbf{z}_{n'}) = \eta \exp\left(-\frac{1}{2}\sum_{q=1}^{Q_Z} \alpha_q (z_{nq} - z_{n'q})^2\right)$ is used where $\eta > 0$ is called the output-scale, which determines the scale of the covariance, and α_q's are called inverse length-scales that reflect the importance of each input dimension. One advantage of the ARD kernel is that it enforces automatic dimension reduction by setting some α_q's to be zero. In GP, the output is usually corrupted with random Gaussian noise, i.e., each observation $y_n \in \mathbb{R}, n = 1, ..., N$ is given by $y_n = f(\mathbf{z}_n) + \varepsilon_n, \forall n$, where $\varepsilon_n \sim \mathcal{N}(0, \beta^{-1})$. The observation follows a multivariate Gaussian distribution

$$p(\mathbf{y}|Z) = \mathcal{N}\left(\mathbf{y}|\mathbf{0}, K_{ZZ} + \beta^{-1} I_N\right) \tag{1}$$

where $K_{ZZ} \in \mathbb{R}^{N\times N}$ is the covariance matrix and $(K_{ZZ})_{n,n'} = k(\mathbf{z}_n, \mathbf{z}_{n'})$. GPs can be trained by maximizing the log-likelihood $\log p(\mathbf{y}|Z)$ to the optimal kernel parameters η^*, $\{\alpha_q^*\}_{q=1}^{Q_Z}$ and noise parameter β^*. Once done with training, the prediction at a new unseen input $\mathbf{z}^*$ can be carried out by training input Z_{train} and training observation $\mathbf{y}_{\text{train}}$ as follows. The posterior predictive distribution of the output function value $f(\mathbf{z}^*)$ is given by

$$p\left(f(\mathbf{z}^*)|Z_{\text{train}}, \mathbf{y}_{\text{train}}, \mathbf{z}^*\right) = \mathcal{N}\left(f(\mathbf{z}^*)|\mu^*, \Sigma^*\right) \tag{2}$$

where $\mu^* = K_{\mathbf{z}^* Z_{\text{train}}}(K_{ZZ} + (\beta^*)^{-1} I_1)^{-1} \mathbf{y}_{\text{train}}$ and $\Sigma^* = K_{\mathbf{z}^*\mathbf{z}^*} - K_{\mathbf{z}^* Z_{\text{train}}}(K_{ZZ} + (\beta^*)^{-1} I_1)^{-1} K^T_{\mathbf{z}^* Z_{\text{train}}}$. The final predictive distribution of observation y^* is obtained by $p(y^*|Z_{\text{train}}, \mathbf{y}_{\text{train}}, \mathbf{z}^*) = \mathcal{N}\left(y^*|\mu^*, \Sigma^* + (\beta^*)^{-1} I_1\right)$. μ^* is used as the prediction at the new testing point.

Due to the limitations of GP and advantages of deep GP that will be described in III-A1, let us consider a simple supervised deep GP consisting of two layers. The input $Z \in \mathbb{R}^{N\times Q_Z}$ passes through the first GP mapping f^X and produces an (unobserved) output $X \in \mathbb{R}^{N\times Q_X}$ which then passes through the second GP mapping f^Y and produces the final observation $Y \in \mathbb{R}^{N\times D}$. Given the input $Z = \{\mathbf{z}_n\}_{n=1}^N$ and observation $Y = \{\mathbf{y}_n\}_{n=1}^N$, the deep GP can be trained under a supervised learning setting in order to find the optimal kernel parameters $\eta_X^*, \{\alpha_{X,q}^*\}_{q=1}^{Q_Z}$ and $\eta_Y^*, \{\alpha_{Y,q}^*\}_{q=1}^{Q_X}$ associated with the GP mappings f^X and f^Y respectively. More specifically, the input-output relation in the considered two-layer deep GP is

$$x_{nq} = f_q^X(\mathbf{z}_n) + \varepsilon_{nq}^X,\ q = 1, \cdots, Q_X \tag{3}$$

$$y_{nd} = f_d^Y(\mathbf{x}_n) + \varepsilon_{nd}^Y,\ d = 1, \cdots, D \tag{4}$$

where $f^X = [f_q^X(\mathbf{z}_1), \cdots, f_q^X(\mathbf{z}_N)]$ and $f^Y = [f_d^Y(\mathbf{x}_1), \cdots, f_d^Y(\mathbf{x}_N)]$. Note that $\mathbf{x}_n = \{x_{nq}\}_{q=1}^{Q_X} \in \mathbb{R}^{Q_X}$ is latent variable which cannot be observed. The prediction of the function value and observation at unseen input is similar to the standard GP.

B. Proposed Approach

We present the proposed deep GP-based wireless heatmap prediction approach in this section. A two-layer deep GP model is used where two GP mappings f^X and f^Y need to be learned. The input Z to our model is simply the measurement location coordinates in an area where the radio heatmap needs to be modeled, i.e., we have $Z \in \mathbb{R}^{N\times Q_Z}$, $Q_Z = 2$. The observation of the deep GP is defined as the received signal strength (RSS) measured at the corresponding locations. In this model, we have $Y = [\mathbf{y}_1, \cdots, \mathbf{y}_D] \in \mathbb{R}^{N\times D}$, where the observation dimension represents the number of RSS values at each location, i.e., $D = 1$. In the deep GP, the probability $p(Y|Z)$ is not analytically tractable. Therefore, direct optimization of the probability is not feasible. Instead, a variational approach [24], [25] from Bayesian learning is used to obtain an analytically tractable lower bound $F_v \leq \log p(Y|Z)$. This approach provides a tractable training procedure for deep GPs and has been shown to obtain promising experimental performance [19]. In this section, we present a novel deep GP architecture to derive new lower bounds that are well-suited for wireless radio heatmap prediction with a small dataset.

1) Model Training: In this supervised training stage, we use the measurement location coordinates Z and observed RSS value Y to maximize the lower bound of log-likelihood.

Let $F^X \triangleq [f^X(\mathbf{z}_1), \cdots, f^X(\mathbf{z}_N)]^T \in \mathbb{R}^{N\times Q_X}$ and $F^Y \triangleq [f^Y(\mathbf{x}_1), \cdots, f^Y(\mathbf{x}_N)]^T \in \mathbb{R}^{N\times D}$ denote the output function values at the two layers respectively. In our proposed deep GP model, the noise variance β_X^{-1} in layer 1 will be treated as a positive constant parameter while the noise variance β_Y^{-1} in layer 2 is learnable parameter. There are three purposes for this setup. First, the learnable noise at layer 2 is adequate to model the random Gaussian noise added to the received RSS. Second, the proposed deep GP model maintains its ability to model large geometric structures influenced by fading, shading, multipath, etc. Third, the constant noise in layer 1 can improve the stability of the values due to the constant diagonal shift to the ARD kernel. The probability $p(Y|Z)$ can be written as $p(Y|Z) = \int_{\mathcal{M}} p(Y, F^Y, F^X, X|Z)$ where $\mathcal{M} \triangleq \{F^Y, F^X, X\}$. Let $Q(\mathcal{M})$ be a variational distribution over $\mathcal{M}$, we can write the probability as $\log p(Y|Z) = \log \int_{\mathcal{M}} Q(\mathcal{M}) \frac{p(Y,F^Y,F^X,X|Z)}{Q(\mathcal{M})} = \log\left(\mathbb{E}_{\mathcal{M}\sim Q(\mathcal{M})}\left[p(Y, F^Y, F^X, X|Z)/Q(\mathcal{M})\right]\right)$. Since log is a concave function, by using Jensen's inequality, the lower bound can be written as

$$\log\left(\mathbb{E}_{\mathcal{M}\sim Q(\mathcal{M})}\left[p(Y, F^Y, F^X, X|Z)/Q(\mathcal{M})\right]\right) \geq \mathbb{E}_{\mathcal{M}\sim Q(\mathcal{M})} \log\left(p(Y, F^Y, F^X, X|Z)/Q(\mathcal{M})\right). \quad (5)$$

We factorize the joint probability as $p(Y, F^Y, F^X, X|Z) = p(Y|F^Y)p(F^Y|X)p(X|F^X)p(F^X|Z)$. It can be seen that (5) is intractable due to the $\int_X p(F^Y|X)$ term where X is involved in $p(F^Y|X)$ in a nonlinear way. To overcome this issue, a set of K IP $\widetilde{X} \in \mathbb{R}^{K\times Q_X}$ is introduced for latent variables X. The purpose of the IPs in layer 2 (not layer 1) is two-fold: First, IPs expand the joint probability space of (5) such that the intractable terms can be canceled; Second, the training dataset we used is small, so it is unnecessary to use a smaller set of IPs to replace the training data in layer 1 to speed up the training process rather than maintain the accuracy. Let $U^Y = [\mathbf{u}_1^Y, \cdots, \mathbf{u}_D^Y] \in \mathbb{R}^{K\times D}$ be the function values corresponding to the IPs $\widetilde{X}$. We can then expand the joint probability space as $p(Y, F^Y, F^X, X, U^Y, \widetilde{X}|Z)$ which can be factorized as

$$p(Y, F^Y, F^X, X, U^Y, \widetilde{X}|Z) = p(Y|F^Y)p(F^Y|U^Y, X) \times p(U^X|\widetilde{X})p(X|F^X)p(F^X|Z), \quad (6)$$

where note that $\tilde{X}$ is not a random variable which will be omitted in what follows. We can generate a variational distribution

$$Q(\tilde{\mathcal{M}}) = p(F^Y|U^Y, X)q(U^Y)q(X)p(F^X|Z), \quad (7)$$

where $\tilde{\mathcal{M}} \triangleq \{F^Y, F^X, X, U^Y\}$. $q(X)$ is chosen as a factorized Gaussian distribution $q(X) = \prod_{q=1}^{Q_X} \mathcal{N}\left(X(:,q)|\boldsymbol{\mu}_q^X, S_q^X\right)$. $q(U^Y)$ is a variational distribution over the IP value U^Y. It can be factorized as $q(U^Y) = \prod_{d=1}^{D} q(\mathbf{u}_d^Y)$, and the optimal form is $q(\mathbf{u}_d^Y) \propto e^{\mathbb{E}_{X\sim q(X)}\left[\log \mathcal{N}(\mathbf{y}_d|K_{X\tilde{X}}K_{\tilde{X}\tilde{X}}^{-1}\mathbf{u}_d^Y, \beta^{-1}I_N)\right]} p(\mathbf{u}_d^Y)$. We can then plug (7) and (6) into (5) so that the intractable terms can be eliminated. In particular, we obtain the lower bound

$$F_v = \int_{\tilde{\mathcal{M}}} Q \log\left(\frac{p(Y|F^Y)p(U^Y)p(X|F^X)}{q(U^Y)q(X)}\right), \quad (8)$$

which can be analytically calculated as

$$F_v = g + r + H(q(X)), \quad (9)$$

where $H(q(X)) = -\int q(X)\log q(X)dX$ is the differential entropy of $q(X)$ which is equal to $H(q(X)) = \sum_{q=1}^{Q_X}(\frac{N}{2}(\ln 2\pi + 1) + \frac{1}{2}\ln|S_q^X|)$. The first term g is related to the second layer of the considered deep GP model and is equal to $g = \mathbb{E}_{p(F^Y|U^Y,X)q(U^Y)q(X)}\left[\log p(Y|F^Y) + \log\frac{p(U^Y)}{q(U^Y)}\right]$. We can further factorize g as $g = \sum_{d=1}^{D} g_d$, where

$$g_d \geq \log\left[\frac{(\beta_Y)^{\frac{N}{2}}|K_{\tilde{X}\tilde{X}}|^{\frac{1}{2}}}{(2\pi)^{\frac{N}{2}}|\beta_Y\Psi_2 + K_{\tilde{X}\tilde{X}}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\mathbf{y}_d^T W \mathbf{y}_d\right)\right] - \frac{\beta_Y}{2}\psi_0 + \frac{\beta_Y}{2}\mathrm{Tr}\left(K_{\tilde{X}\tilde{X}}^{-1}\Psi_2\right), \quad (10)$$

where $W = \beta_Y I_N - \beta_Y^2\Psi_1\left(\beta_Y\Psi_2 + K_{\tilde{X}\tilde{X}}\right)^{-1}\Psi_1^T$, and $\psi_0 = N\eta_Y$. Ψ_1 is a $N \times K$ matrix with $(\Psi_1)_{nk} = \eta_Y\prod_{q=1}^{Q_X}\exp\left(-\frac{\alpha_{Y,q}\left(\mu_{nq}^X - \tilde{x}_{kq}\right)^2}{2\left(\alpha_{Y,q}S_{nq}^X + 1\right)}\right)/\left(\alpha_{Y,q}S_{nq}^X + 1\right)^{\frac{1}{2}}$. $\Psi_2 = \sum_{n=1}^{N}\Psi_2^n$ is a $K \times K$ matrix where $(\Psi_2^n)_{kk'} = \eta_Y^2\prod_{q=1}^{Q_X}\frac{\exp\left(-\frac{\alpha_{Y,q}\left(\tilde{x}_{kq} - \tilde{x}_{k'q}\right)^2}{4} - \frac{\alpha_{Y,q}\left(\mu_{nq}^X - \bar{x}_q\right)^2}{2\alpha_{Y,q}S_{nq}^X + 1}\right)}{\left(2\alpha_{Y,q}S_{nq}^X + 1\right)^{\frac{1}{2}}}$, and $\bar{x}_q = \frac{1}{2}\left(\tilde{x}_{kq} + \tilde{x}_{k'q}\right)$. The second term r is related to the first layer of deep GP and is equal to $r = \mathbb{E}_{p(F^X|Z)q(X)}\left[\log p(X|F^X)\right]$ which is factorized as $r = \sum_{q=1}^{Q_X} r_q$, where $r_q = -\frac{N}{2}\left(\log 2\pi - \log|\Sigma_{xq}| - \left(\boldsymbol{\mu}_q^X\right)^T\Sigma_{xq}^{-1}\boldsymbol{\mu}_q^X - \mathrm{Tr}(\Sigma_{xq}^{-1}S_q^X)\right)$ and $\Sigma_{xq} = K_{ZZ} + \beta_X^{-1}I_N$. Now, the variational lower bound (9)

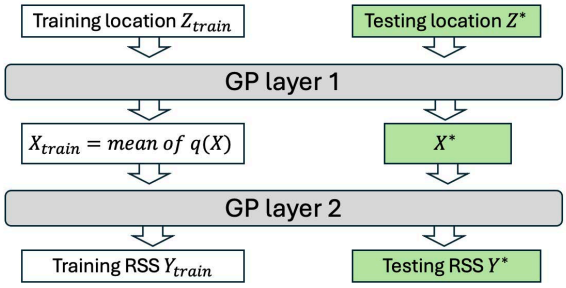

Fig. 1. Prediction in the proposed two-layer deep GP model. Prediction for layer 1 uses the optimal mean of $q(X)$ as the output corresponding to Z_{train}. Prediction for layer 2 uses predicted output X^* of layer 1 as test input to layer 2.

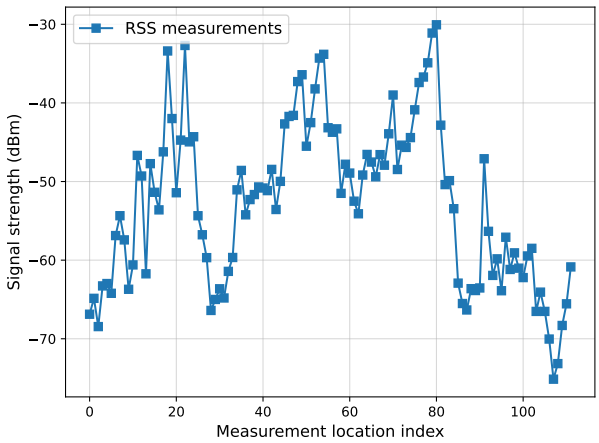

Fig. 2. RSS measurements (dBm) at all 112 measurement locations.

can be analytically calculated so that gradient descent-based methods can be used to find the optimal kernel and variational parameters.

2) Prediction: The deep GP prediction (Fig. 1) is the cascaded GP predictions. We use the optimal mean of variational distribution $q(X)$ as an estimate X_{train} for the training output of layer 1 and the training input of layer 2. The estimated RSS value at the new locations $Z^* \in \mathbb{R}^{N^* \times Q_Z}$ can be carried out by training location Z_{train} and training RSS value Y_{train} in a similar way to the standard GP. Thus, the distribution of layer 1's output $X^* \in \mathbb{R}^{N^* \times Q_X}$ is

$$p\left(X^* | Z_{\text{train}}, X_{\text{train}}, Z^*\right) = \mathcal{N}\left(X^* | M_X^*, \Sigma_X^* + \beta_X^{-1} I_{N^*}\right), \tag{11}$$

where $M_X^* = K_{Z^* Z_{\text{train}}}(K_{ZZ} + \beta_X^{-1} I_{N^*})^{-1} X_{\text{train}}$ and $\Sigma_X^* = K_{Z^* Z^*} - K_{Z^* Z_{\text{train}}}(K_{ZZ} + \beta_X^{-1} I_{N^*})^{-1} K_{Z^* Z_{\text{train}}}^T$. M_X^* is used as the predicted testing output of layer 1 at new testing location coordinates, which in turn becomes the test input for layer 2. The distribution of observed RSS value $Y^* \in \mathbb{R}^{N^* \times D}$ is

$$p\left(Y^* | X_{\text{train}}, Y_{\text{train}}, X^*\right) = \mathcal{N}\left(Y^* | M_Y^*, \Sigma_Y^* + (\beta_Y^*)^{-1} I_{N^*}\right), \tag{12}$$

where $M_Y^* = K_{X^* X_{\text{train}}}(K_{XX} + (\beta_Y^*)^{-1} I_{N^*})^{-1} Y_{\text{train}}$ and $\Sigma_Y^* = K_{X^* X^*} - K_{X^* X_{\text{train}}}(K_{XX} + (\beta_Y^*)^{-1} I_{N^*})^{-1} K_{X^* X_{\text{train}}}^T$. M_Y^* is the predicted RSS at the testing location.

In summary, in the RSS training and prediction stage, the proposed two-layer deep GP model utilizes location coordinates as inputs and corresponding RSS values as observations to estimate their relationship. Then, the optimal model and variational parameters are derived, enabling the generation of radio heatmap predictions at testing locations along the measurement route.

III. Numerical Evaluation

A. Experiment Setup

1) Data Processing: The proposed DGP model is evaluated on the real dataset collected from the POWDER Platform [22], which is for open wireless experimental research at the University of Utah. Four dense nodes (Mario, Wastch, Moran, and Guesthouse) and one rooftop node (USTAR) are used as BSs, and we use mobile USRP B210 radio to collect 112 sets of the I/Q (In-phase and Quadrature-phase) samples within an open outdoor area of size 700×700 square meters in the University of Utah campus, and a GPS device to collect the corresponding location data. Each of these BSs uniformly emits a continuous-wave signal in the CBRS Band (bandwith $=$ 27.5 kHz, $f_{\text{c, USTAR}} = 3533.9175$ MHz, $f_{\text{c, Guesthouse}} = 3533.95875$ MHz, $f_{\text{c, Mario}} = 3534$ MHz, $f_{\text{c, Moran}} = 3534.04125$ MHz, $f_{\text{c, Wasatch}} = 3534.0825$ MHz). The locations are converted from longitude and latitude in degrees to meters, and an origin is selected to obtain the relative coordinates of all the measurement locations. The number of received I/Q samples for one location is 131,072. The processing of the I/Q samples consisted of first separating the data into in-phase components and quadrature components and then calculating the power values using the I/Q samples. The data for each location is represented by the average power of all samples collected at that location, thus obtaining 1-dimensional observation data. Afterward, the power is converted to RSS by converting the Watts unit to dBm units. The conversion is used due to the fact that most RSS measurements are extremely small values near zero. It leads to numerical instability in deep GP training, as well as baseline training. Therefore, the utilization of dBm units ensures that RSS values fall within a suitable numerical range. Fig. 2 and Fig. 3 show the RSS values and the corresponding measurement locations (112 in total). We will divide the dataset into different numbers of training and testing sets to test the performance of deep GP.

2) Experiment setup: In our model, the dimension of latent variable X and the dimension of observed RSS values Y are both 1. The input locations Z are scaled by a factor of $1/700$ to enhance the training efficiency of deep GP. The mean of the variational distribution $q(X)$ and IPs in layer 2 are initialized by applying the principal component analysis (PCA) [26] method to the observed RSS values. The variance of the variational distribution is initialized with product LL^T, where L is an upper triangular matrix. Since we have 112 points in total which is small, the number of IPs in layer 2 is equal to the number of training inputs, $N = K$. First, random initialization

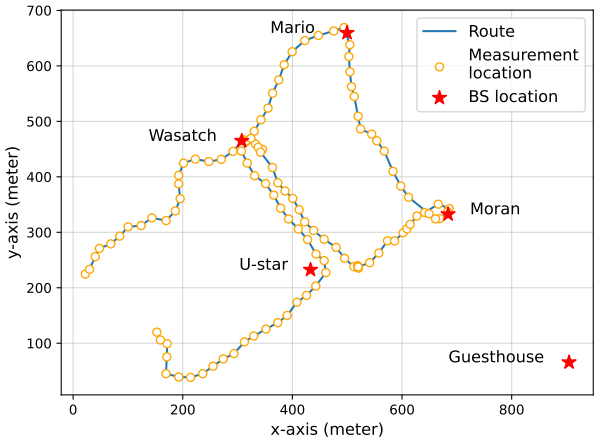

Fig. 3. Demonstration of the RSS measurement locations along the traffic route. BS locations are marked by red stars (five in total).

of length-scale, output-scale and noise variance are set for prediction to evaluate the normalized root mean square error (NRMSE). Subsequently, the initialization values are selected based on the obtained NRME value. Then, we optimize the model hyperparameters and variational parameters together. We use the Adam optimizer with a learning rate of 0.001 and 50,000 iterations to ensure the convergence of deep GP. The proposed approach is implemented with Python and uses PyTorch as the auto-differentiation engine for model training.

3) Baselines: For comparison, we implement and evaluate two baseline schemes. The first one is the standard GP model. The experiments maintain the same dataset and the initialization method of parameters and optimizer, only altering the different model architectures. The second baseline is the DNN, with four hidden layers and 200 neurons for each layer. The activation function is the rectified linear unit function, and the loss function is the mean square error (MSE) loss function.

B. Result

The experiment data and results are presented in this section.

1) Training set and testing set: The experiment employs three different configurations to divide the dataset into training and testing subsets with different numbers of data points, as shown in the table below. Fig. 4 shows the specific distribu-

Configuration	1	2	3
# of training/testing locations	38/74	56/56	94/18
Training percentage	34%	50%	84%

tions of the training locations (denoted by blue crosses) and testing locations (denoted by orange circles) under different configurations. BS locations are marked by red stars. The training set is generated along the measurements route by taking one training point at every three locations, one training point at every alternate location, and five training points at every alternate location for three different training set configurations. The test sets include RSS values from the remaining locations.

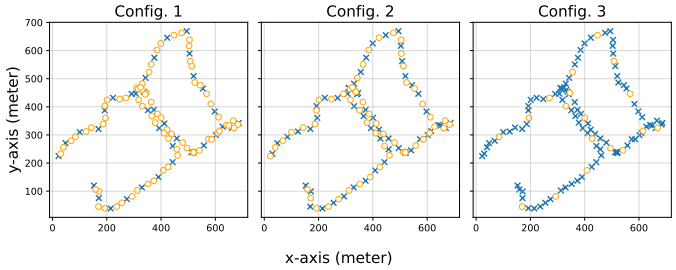

Fig. 4. Training location (blue cross) and testing location (orange circle) distribution under different configurations.

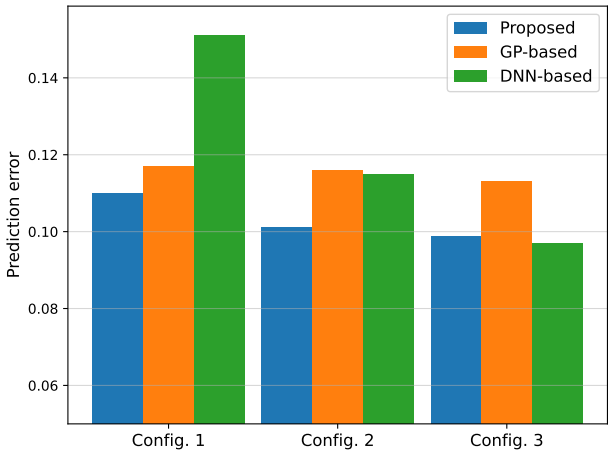

Fig. 5. Comparison of prediction errors with the baselines under different configurations.

The purpose of having multiple configurations is to evaluate the performance of the proposed approach under scarce and relatively adequate training data conditions.

2) Experiment result: Fig. 5 and Fig. 6 show the experimental results. In particular, Fig. 5 compares the prediction NRMSE of deep GP and the two baselines in different con-

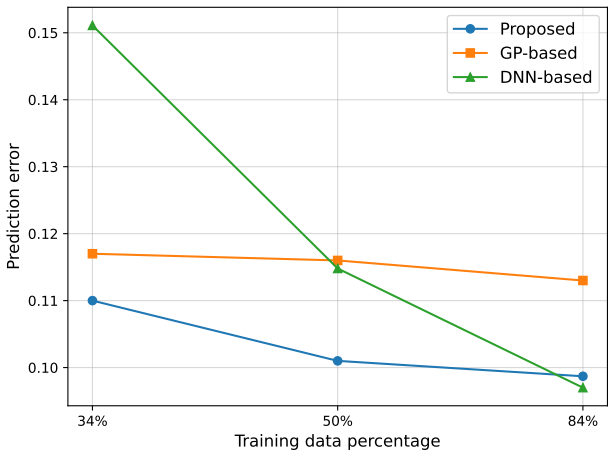

Fig. 6. Prediction error versus training data percentage.

figurations. All three schemes are able to regress the ground truth RSS value when the percentage of training data is 84%, which is relatively large. We can see the prediction error of deep GP is comparable to but slightly higher than the error of DNN. This is because the DNN model we used a more complex and deeper structure than our proposed deep GP, i.e., the structure of DNN is 4 hidden layers and 200 neurons while the deep GP is a two-layer model. Furthermore, the error for the standard GP is much higher than that of the other two methods. When gradually reducing the number of training points, such as in configurations 1 and 2, the prediction errors of the three methods show an increasing trend of different degrees, as shown in Fig. 6. DNN is the most affected, with its NRMSE increasing significantly from 0.097 to 0.151. The errors for both deep GP and standard GP are slightly higher than the results for large dataset, but the predicted RSS accuracy of deep GP is still better than that of standard GP. Overall, compared with standard GP, deep GP has better accuracy regardless of the size of the training set because it can model more complex data structures than standard GP, especially in capturing the nonlinear relationship between input and observation. Deep GP achieves higher accuracy and is more robust than the DNN baseline scheme for small training datasets. In contrast, the performance of DNN can be affected by overfitting problems to the training data, which induces large generalization (prediction) errors. Even with a larger dataset, deep GP achieves comparable predictive performance to DNN with more complex and deeper structures. Therefore, we attest that deep GP has a good performance among these three prediction methods with a small training dataset, and it is robust enough to make the wireless outdoor heatmap prediction.

IV. Conclusion

In this paper, we proposed a Bayesian learning approach to outdoor heatamp construction using a two-layer deep Gaussian process. The proposed deep GP model replaces the standard deep GP architecture with a full GP in the first layer and only uses inducing point for the second layer, which strikes a balance in model approximation accuracy and complexity. A corresponding optimizable lower bound on the model likelihood was also derived. In addition, the proposed approach was evaluated using practical RSS datasets under a multi-BS scenario from the POWDER platform. Experiment results demonstrated the superior performance of the deep GP-based approach over the baseline schemes including GP and DNN-based regression models.

References

[1] S.-S. Jan, S.-J. Yeh, and Y.-W. Liu, "Received signal strength database interpolation by kriging for a wi-fi indoor positioning system," *Sensors*, vol. 15, no. 9, pp. 21 377–21 393, 2015.

[2] P. Dhere, P. Chilveri, R. Vatti, V. Iyer, and K. Jagdale, "Wireless signal strength analysis in a home network," in *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*. IEEE, 2018, pp. 1–5.

[3] Z. Utkovski, P. Agostini, M. Frey, I. Bjelakovic, and S. Stanczak, "Learning radio maps for physical-layer security in the radio access," in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2019, pp. 1–5.

[4] C. B. Andrade and R. P. F. Hoefel, "Ieee 802.11 wlans: A comparison on indoor coverage models," in *CCECE 2010*. IEEE, 2010, pp. 1–6.

[5] A. Borrelli, C. Monti, M. Vari, and F. Mazzenga, "Channel models for ieee 802.11 b indoor system design," in *2004 IEEE International Conference on Communications (IEEE Cat. No. 04CH37577)*, vol. 6. IEEE, 2004, pp. 3701–3705.

[6] G. Ding, Z. Tan, J. Zhang, and L. Zhang, "Regional propagation model based fingerprinting localization in indoor environments," in *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. IEEE, 2013, pp. 291–295.

[7] W. Sangkusolwong and A. Apavatjrut, "Indoor wifi signal prediction using modelized heatmap generator tool," in *2017 21st International Computer Science and Engineering Conference (ICSEC)*. IEEE, 2017, pp. 1–5.

[8] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1067–1080, 2007.

[9] W. Zhang, K. Liu, W. Zhang, Y. Zhang, and J. Gu, "Deep neural networks for wireless localization in indoor and outdoor environments," *Neurocomputing*, vol. 194, pp. 279–287, 2016.

[10] T. Imai, K. Kitao, and M. Inomata, "Radio propagation prediction model using convolutional neural networks by deep learning," in *2019 13th European Conference on Antennas and Propagation (EuCAP)*. IEEE, 2019, pp. 1–5.

[11] R. Levie, Ç. Yapar, G. Kutyniok, and G. Caire, "Radiounet: Fast radio map estimation with convolutional neural networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 6, pp. 4001–4015, 2021.

[12] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[13] M. Seeger, "Gaussian processes for machine learning," *International journal of neural systems*, vol. 14, no. 02, pp. 69–106, 2004.

[14] A. Schwaighofer, M. Grigoras, V. Tresp, and C. Hoffmann, "Gpps: A gaussian process positioning system for cellular networks," *Advances in Neural Information Processing Systems*, vol. 16, 2003.

[15] R. Miyagusuku, A. Yamashita, and H. Asama, "Improving gaussian processes based mapping of wireless signals using path loss models," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 4610–4615.

[16] F. Seco, C. Plagemann, A. R. Jiménez, and W. Burgard, "Improving rfid-based indoor positioning accuracy using gaussian processes," in *2010 International Conference on Indoor Positioning and Indoor Navigation*. IEEE, 2010, pp. 1–8.

[17] A. Bekkali, T. Masuo, T. Tominaga, N. Nakamoto, and H. Ban, "Gaussian processes for learning-based indoor localization," in *2011 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*. IEEE, 2011, pp. 1–6.

[18] E. Schulz, M. Speekenbrink, and A. Krause, "A tutorial on gaussian process regression: Modelling, exploring, and exploiting functions," *Journal of Mathematical Psychology*, vol. 85, pp. 1–16, 2018.

[19] A. Damianou and N. D. Lawrence, "Deep gaussian processes," in *Artificial intelligence and statistics*. PMLR, 2013, pp. 207–215.

[20] X. Wang, X. Wang, S. Mao, J. Zhang, S. C. Periaswamy, and J. Patton, "Indoor radio map construction and localization with deep gaussian processes," *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 11 238–11 249, 2020.

[21] X. Wang, M. Patil, C. Yang, S. Mao, and P. A. Patel, "Deep convolutional gaussian processes for mmwave outdoor localization," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 8323–8327.

[22] J. Breen, A. Buffmire, J. Duerig, K. Dutt, E. Eide, M. Hibler, D. Johnson, S. K. Kasera, E. Lewis, D. Maas *et al.*, "Powder: Platform for open wireless data-driven experimental research," in *Proceedings of the 14th International Workshop on Wireless Network Testbeds, Experimental evaluation & Characterization*, 2020, pp. 17–24.

[23] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006, vol. 2, no. 3.

[24] M. K. Titsias, "Variational model selection for sparse gaussian process regression," *Report, University of Manchester, UK*, 2009.
[25] M. Titsias and N. D. Lawrence, "Bayesian gaussian process latent variable model," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 2010, pp. 844–851.
[26] I. T. Jolliffe, *Principal component analysis for special types of data*. Springer, 2002.