Heteroscedastic Gaussian process model for received signal strength based device-free localization

Ossi Kaltiokallio, Roland Hostettler, Jukka Talvitie, Mikko Valkama

DOI:

10.1109/PLANS61210.2025.11028551

Heteroscedastic Gaussian Process Model for Received Signal Strength Based Device-Free Localization

Ossi Kaltiokallio*, Roland Hostettler†, Jukka Talvitie*, and Mikko Valkama*
*Unit of Electrical Engineering, Tampere University, Tampere, Finland, e-mail:{firstname.lastname}@tuni.fi
†Department of Electrical Engineering, Uppsala University, Uppsala, Sweden, e-mail:{firstname.lastname}@angstrom.uu.se

***Abstract*—Received signal strength (RSS) based passive localization approaches measure human-induced changes in the electromagnetic field to localize and track people. Bayesian estimation methods have been widely utilized to solve the problem, mainly because of their convenience in representing uncertainties in the models and in modeling physical randomness. The localization performance is significantly influenced by the measurement model that describes the electromagnetic field changes as a function of the location of the target, and a wide variety of empirical and analytical models have been proposed. Common to these models is that the measurement noise is assumed homoscedastic, that is, the measurement noise is constant. In this paper, the measurement noise is assumed to depend on the location of the target, and a novel heteroscedastic Gaussian process model for RSS-based device-free localization and tracking (DFLT) is proposed. In addition, algorithms to train the model parameters and solve the RSS-based DFLT problem are presented. The models and tracking algorithms are evaluated using experiments conducted in an open-space indoor environment and in a fully furnished downtown residential apartment. The results imply that the proposed approach can decrease the localization error with respect to the benchmark RSS models and that real-time sub-decimeter tracking accuracy can be achieved in both environments.**

***Index Terms*—Gaussian process, heteroscedastic noise, propagation modeling, received signal strength, device-free localization and tracking, Bayesian estimation**

I. Introduction

Radio frequency (RF) sensing technologies utilize perturbations of the electromagnetic field for estimating physical quantities of interest such as presence of people, crowd density, activity, gestures, location and vital signs [1]. The attributes "device-free", "passive" and "sensorless" are typically used to highlight that the technology does not require the monitored subject, which we refer to as target from now on, to carry or wear any active or passive device. The technology leverages the fact that targets alter the propagation characteristics of radio signals, and these changes can be quantified at the receiver (RX) using the radio channel estimate of the radio module. Then, the physical quantity of interest can be inferred from the channel estimate(s) using for example a parametric statistical model [2] or a non-parametric machine learning model [3]. Over the past two decades, various wireless technologies have been used to demonstrate the capabilities of RF sensing. As an example, dedicated hardware combined with large bandwidths and large antenna arrays can be used for acquiring high resolution delay and angle estimates, enabling for example remote vital sign monitoring [4], multipath-enhanced passive localization [5], and radio-based simultaneous localization and mapping (SLAM) [6]. At the other end of the spectrum are commodity wireless devices that provide received signal strength (RSS) estimates. Even though the RSS is not as informative as delay or angle estimates, it still conveys useful information that can be used to realize various RF sensing applications such as breathing monitoring [7] and localization [8]. The main benefit of using inexpensive commodity wireless devices is that they can be deployed in numbers forming a dense mesh network and to perform *multistatic sensing*. In this paper, we consider a multistatic sensing system, composed of commodity narrowband wireless devices capable of measuring the RSS, for device-free localization and tracking (DFLT).

The performance of state-of-the-art DFLT systems is significantly influenced by the measurement model that describes the RSS as a function of target's location. Significant research efforts have been undertaken to model the RSS using, for example, first principles, and it has been argued that the target induced perturbations to the wireless channel are caused by shadowing [9], [10], reflection [11], and/or diffraction [12], [13]. In addition, a wide variety of empirical models have been proposed [2], [14], [15]. Common to all these models is that the largest RSS changes are measured physically in between the transmitter (TX) and RX, and that the influence decays as the bistatic range to the target increases. A major limitation of the analytical and empirical narrowband models is that they are unable to explain RSS changes when the target alters an existing multipath component or when the bistatic range is large. Another significant drawback is that the models typically assume *homoscedastic* measurement noise, that is, the RSS is assumed to be corrupted by independent and identically distributed (iid.) random noise [3], [14], [15]. However, both theoretical [12], [16] and empirical evidence [17], [18] suggest that the noise process is *heteroscedastic*, that is, the measurement noise is dependent on the target's location. Fingerprint-based machine learning approaches can capture the RSS changes as well as the heteroscedastic noise appropriately [19]. However, collecting the fingerprints is labo-

This work was partially supported by the Research Council of Finland under the grants #338224, #345654, and #352754, and by the Business Finland under the 6G-ISAC project.

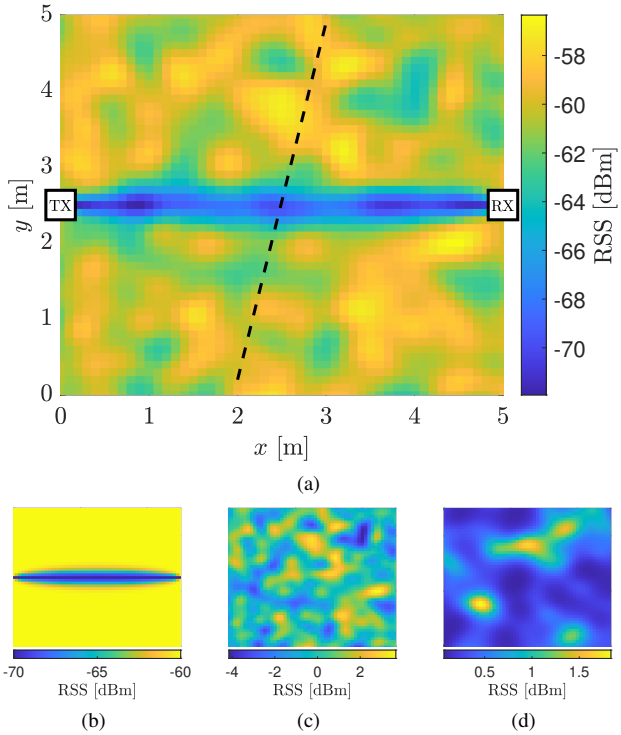


Fig. 1. An illustration of the proposed heteroscedastic GP model given in (16a)–(16c) for a link with $\mathbf{p}_{\text{TX}} = [0,\ 2.5]^\top$ and $\mathbf{p}_{\text{RX}} = [5,\ 2.5]^\top$. Example realization of the proposed RSS model shown in (a) and a decomposition of the model in (b)–(d). The model expresses that the RSS is close to a global linear model shown in (b), with the residuals (c) as well as the noise (d) being modeled by GPs.

rious and the models are non-continuous allowing estimation only at discrete fingerprint locations.

In this paper, we propose a heteroscedastic GP model for RSS-based DFLT and develop four different Bayesian localization and tracking algorithms to solve the problem. GPs provide a Bayesian, non-parametric approach for data driven modeling of smooth functions [20]. The main benefit of GPs is that they allow us to model highly nonlinear RSS variations that stem from multipath propagation [3], uncertainty can be correctly handled, the model is continuous allowing estimation at arbitrary locations, and prior knowledge and beliefs can be fused into the GP model [20]. The main contributions of the proposed GP model include: i) we relax the constraint of iid. random noise and model the heteroscedastic noise using a GP; ii) we explicitly model the mean function which allows us to integrate our prior knowledge on the propagation mechanisms into the GP model and which also improves interpretability of the model; and iii) we incorporate prior beliefs on model parameters by deriving the maximum a posteriori (MAP) estimate for the parameters. The proposed model, illustrated in Fig. 1, expresses that the RSS is close to a global linear in parameters model with the residuals as well as the noise being modeled by GPs. The mean function of the model, shown in Fig. 1b, captures the baseline signal strength and target induced RSS changes caused by line-of-sight (LOS) blockage. The GP model for the residuals, for which a realization is illustrated in Fig. 1c, captures other slow fading phenomena such as perturbations to existing non-line-of-sight (NLOS) components. Lastly, a realization of heteroscedastic noise is visualized in Fig. 1d, and the model captures hardware induced errors, interference from co-located wireless networks, as well as target location dependent fast fading.

The development efforts of the paper are demonstrated using commodity radios that operate on the 2.4 GHz ISM band and according to the IEEE 802.15.4 standard. However, it is important to note that the developed methods can be generalized to any device capable of measuring the signal strength including Wi-Fi, Bluetooth and RFID. The experiments are conducted in an open-space indoor environment and in a fully furnished downtown residential apartment and in both experiments, 20 sensors are deployed into the environment. Using the experimental data, it is shown that the proposed approach can decrease the localization error up to 80% with respect to a benchmark parametric model and up to 49% with respect to a benchmark GP model. Furthermore, real-time sub-decimeter tracking accuracy is achieved in both environments, using the proposed heteroscedastic GP model and a particle filter (PF) tracking algorithm.

The rest of the paper is organized as follows. In the following section, the problem is formulated and required background knowledge is briefly summarized. In Section III, the proposed heteroscedastic GP model and two simplifications of the model are presented. Thereafter, algorithms to solve the RSS-based DFLT problem are presented in Section IV. Section V describes the conducted experiments with Section VI presenting the results. Conclusions are drawn thereafter.

II. Background

A. Problem Formulation

Let us consider a monitored area $\mathcal{A} \subset \mathbb{R}^2$ and a target located within the area at $\mathbf{x}_i \in \mathcal{A}$. Let us also consider a TX which periodically transmits sounding signals and an RX measuring the RSS. The TX-RX pair forms a *bistatic sensing system*, referred to simply as a *link* from now on. Typically, the RSS of a link is modeled as a function $f(\cdot)$ plus additive noise ε_i such that $y_i = f(\mathbf{x}_i) + \varepsilon_i$. If $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$, the RSS can be expressed as

$$y_i \sim \mathcal{N}(f(\mathbf{x}_i), \sigma^2), \tag{1}$$

where $f(\mathbf{x}_i)$ is the unknown RSS model associated with target location $\mathbf{x}_i$ and σ^2 is the measurement noise variance.

Now, the considered problem is two-fold. The first problem is related to modeling the RSS and training the model parameters. Required background knowledge for the first problem is reviewed in the following and two benchmark solutions are presented. Thereafter, the RSS models proposed in this paper are presented in Section III. The second problem is solving the inverse problem, that is, estimating the target location using unseen data. In Section IV, four different methods to solve the inverse problem are presented.

B. Gaussian Process Regression

In this section, we briefly review GP regression and for a more thorough introduction to GPs, see [20]. GPs provide a Bayesian, non-parametric approach for data-driven modeling of smooth functions [20]. In particular, assume that the random function $f : \mathbb{R}^{\dim(\mathbf{x})} \mapsto \mathbb{R}$ is assigned a zero-mean GP prior $\mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$ with covariance function $k(\mathbf{x}, \mathbf{x}')$, that is, let

$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}')). \tag{2}$$

GP regression is usually formulated as predicting the unknown function value $f(\mathbf{x}_*)$ associated with a known test point $\mathbf{x}_*$ given a training data set $\mathcal{D} = \{(\mathbf{x}_i, y_i) \mid i = 1, \ldots, n\}$. Predicting the function value can be achieved by finding the predictive density [20]

$$p(f(\mathbf{x}_*) \mid \mathcal{D}) = \mathcal{N}(f(\mathbf{x}_*) \mid \mathrm{E}[f(\mathbf{x}_*)], \mathrm{V}[f(\mathbf{x}_*)]) \tag{3}$$

with predictive mean $\mathrm{E}[f(\mathbf{x}_*)]$ and variance $\mathrm{V}[f(\mathbf{x}_*)]$

$$\mathrm{E}[f(\mathbf{x}_\star)] = \mathbf{k}_*^\top (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{y}, \tag{4a}$$
$$\mathrm{V}[f(\mathbf{x}_\star)] = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^\top (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{k}_*. \tag{4b}$$

In the above, we have $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \ldots & y_n \end{bmatrix}^\top$, $\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ and $\mathbf{k}_*$ is an n-dimensional vector with the ith element being $k(\mathbf{x}_*, \mathbf{x}_i)$.

The covariance function is a measure of the similarity between different function values and its design and choice is key when employing GP models. One of the most commonly used covariance functions, which will also be frequently used in the following sections, is the squared exponential [20]

$$k_{\text{se}}(\mathbf{x}, \mathbf{x}') = \sigma_{\text{se}}^2 \exp\left(-\tfrac{1}{2\ell_{\text{se}}^2} \|\mathbf{x} - \mathbf{x}'\|^2\right). \tag{5}$$

The covariance function is parametrized by a set of hyperparameters $\boldsymbol{\phi}_{\text{se}} = [\sigma_{\text{se}}^2, \ell_{\text{se}}^2]$ in which σ_{se} and ℓ_{se} represent the magnitude scale and characteristic length-scale, respectively. The hyperparameters can be learned from data, for instance by maximizing the marginal likelihood in (11a).

C. Benchmark Models

1) Parametric Exponential Model: Let $\mathbf{p}_{\text{TX}}$ and $\mathbf{p}_{\text{RX}}$ denote the TX and RX positions of the link. Next, consider that the target is at range $d_{\text{TX}}(\mathbf{x}_i) = \|\mathbf{p}_{\text{TX}} - \mathbf{x}_i\|$ from the TX, range $d_{\text{RX}}(\mathbf{x}_i) = \|\mathbf{p}_{\text{RX}} - \mathbf{x}_i\|$ from the RX, and the TX and RX are distance $d = \|\mathbf{p}_{\text{TX}} - \mathbf{p}_{\text{RX}}\|$ apart. Now, the bistatic range can be expressed as

$$\Delta(\mathbf{x}_i) = d_{\text{TX}}(\mathbf{x}_i) + d_{\text{RX}}(\mathbf{x}_i) - d, \tag{6}$$

which defines an ellipse with TX and RX as the two foci. A de-facto standard model for the general RSS model in (1) is the exponential model (EM) [14], given by

$$\begin{aligned} f(\mathbf{x}_i) &\triangleq \mathbf{h}(\mathbf{x}_i)\boldsymbol{\theta}, \\ &\triangleq \begin{bmatrix} 1 & \exp(-\Delta(\mathbf{x}_i)/\ell_{\text{em}}) \end{bmatrix} \begin{bmatrix} \mu_{\text{em}} \\ \phi_{\text{em}} \end{bmatrix}, \end{aligned} \tag{7}$$

where $\boldsymbol{\theta}$ is the model parameter vector with mean parameter μ_{em} and magnitude parameter ϕ_{em}, and ℓ_{em} is the constant length-scale parameter of the model.

2) Non-parametric GP Model: Rather than using a parametric model or modeling the different components analytically using, for example, first principles, [3] proposed to use a GP model according to (1)–(2) instead. The main challenge in utilizing GPs is the choice of the covariance function and the regressors. Various choices were explored in [3] and the best overall performance was obtained using the model

$$f(\mathbf{x}) \sim \mathcal{GP}(0, \sigma_0^2 + k_{\text{se}}^{(1)}(\Delta(\mathbf{x}), \Delta(\mathbf{x}')) + k_{\text{se}}^{(2)}(\mathbf{x}, \mathbf{x}')), \tag{8}$$

in which the covariance function is a sum of a constant covariance function parameterized using σ_0^2, and two squared exponential covariance functions with different regressors. The first squared exponential uses bistatic range given in (6) as the regressor, whereas the second uses Cartesian coordinates of the target as the regressor. The hyperparameters of the benchmark Gaussian process (BGP) model are $\boldsymbol{\phi} = [\sigma_0^2, \boldsymbol{\phi}_{\text{se}}^{(1)}, \boldsymbol{\phi}_{\text{se}}^{(2)}]$.

III. Proposed GP Models

A. Homoscedastic GP Model

Instead of modeling the RSS using (1) and a zero-mean GP prior in (2), we can explicitly model the mean function to improve model interpretability and incorporate prior knowledge to the model. The proposed homoscedastic GP model is given by

$$f(\mathbf{x}) \sim \mathcal{GP}(\mathbf{h}(\mathbf{x})\boldsymbol{\theta}, k(\mathbf{x}, \mathbf{x}')), \tag{9}$$

which expresses that the RSS is close to the exponential model in (7) with the residuals being modeled by a GP. In this paper, $k(\mathbf{x}, \mathbf{x}')$ is the squared exponential in (5) parameterized by hyperparameters $\boldsymbol{\phi} = [\sigma_{\text{se}}^2, \ell_{\text{se}}^2]$. The predictive density is given in (3) and since the mean is modeled explicitly, the expression for the predictive mean changes, whereas the variance remains the same [20]. The predictive mean is given by

$$\mathrm{E}[f(\mathbf{x}_\star)] = \mathbf{h}(\mathbf{x}_\star)\boldsymbol{\theta} + \mathbf{k}_\star^\top (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} (\mathbf{y} - \mathbf{H}\boldsymbol{\theta}), \tag{10}$$

where $\mathbf{H} = \begin{bmatrix} \mathbf{h}(\mathbf{x}_1)^\top & \mathbf{h}(\mathbf{x}_2)^\top & \ldots & \mathbf{h}(\mathbf{x}_n)^\top \end{bmatrix}^\top$.

A common way to learn the model parameters is by maximizing the *marginal likelihood function* [20]. The log marginal likelihood, $\mathcal{L}_{\mathbf{y}} \triangleq \log p(\mathbf{y} \mid \boldsymbol{\Theta})$, and its derivatives are:

$$\mathcal{L}_{\mathbf{y}} = -\frac{1}{2}\left(\boldsymbol{\eta}^\top \mathbf{K}_{\mathbf{y}}^{-1} \boldsymbol{\eta} + \log|\mathbf{K}_{\mathbf{y}}| + n\log(2\pi)\right), \tag{11a}$$
$$\frac{\partial \mathcal{L}_{\mathbf{y}}}{\partial \boldsymbol{\theta}} = \mathbf{H}^\top \mathbf{K}_{\mathbf{y}}^{-1} \boldsymbol{\eta}, \tag{11b}$$
$$\frac{\partial \mathcal{L}_{\mathbf{y}}}{\partial \boldsymbol{\phi}_i} = \frac{1}{2}\boldsymbol{\eta}^\top \mathbf{K}_{\mathbf{y}}^{-1} \frac{\partial \mathbf{K}_{\mathbf{y}}}{\partial \boldsymbol{\phi}_i} \mathbf{K}_{\mathbf{y}}^{-1} \boldsymbol{\eta} - \frac{1}{2}\operatorname{Tr}\left(\mathbf{K}_{\mathbf{y}}^{-1} \frac{\partial \mathbf{K}_{\mathbf{y}}}{\partial \boldsymbol{\phi}_i}\right), \tag{11c}$$
$$\frac{\partial \mathcal{L}_{\mathbf{y}}}{\partial \sigma^2} = \frac{1}{2}\boldsymbol{\eta}^\top \mathbf{K}_{\mathbf{y}}^{-1} \mathbf{K}_{\mathbf{y}}^{-1} \boldsymbol{\eta} - \frac{1}{2}\operatorname{Tr}\left(\mathbf{K}_{\mathbf{y}}^{-1}\right), \tag{11d}$$

where $\mathbf{K}_{\mathbf{y}} = (\mathbf{K} + \sigma^2 \mathbf{I})$ and $\boldsymbol{\eta} = \mathbf{y} - \mathbf{H}\boldsymbol{\theta}$. Once the marginal likelihood and its derivatives are available, the model parameters can be trained using, for example, a gradient-based optimizer.

If we also assume that we have a prior over the parameters, that is, $\boldsymbol{\Theta} \sim p(\boldsymbol{\Theta})$ (where $\boldsymbol{\Theta}$ may be any of the model $\boldsymbol{\theta}$ or GP

hyperparameters ϕ), we may instead maximize the *marginal log-posterior* of the parameters, given by

$$\log p(\boldsymbol{\Theta} \mid \mathbf{y}) = \log p(\mathbf{y} \mid \boldsymbol{\Theta}) + \log p(\boldsymbol{\Theta}), \tag{12}$$

where the prior of the parameters can typically be factorized such that

$$\log p(\boldsymbol{\Theta}) = \log p(\boldsymbol{\theta}) + \log p(\sigma^2) + \log p(\sigma^2_{\text{se}}) + \log p(\ell^2_{\text{se}}). \tag{13}$$

In this paper, we use the approximation $p(\boldsymbol{\theta}) = \mathcal{N}(\mathbf{m}, \mathbf{P})$ since a Gaussian provides a good empirical fit to the EM parameters (see Section V-B). Respectively, the Inverse-Gamma distribution is the conjugate prior distribution for the variance of a Gaussian distribution and for this reason, using $p(\sigma^2) = \Gamma^{-1}(a, b)$ is common in Bayesian analysis [21]. We could assume a prior for the GP hyperparameters as well but we decided to solely rely on data to learn the hyperparameters such that $\log p(\boldsymbol{\Theta}) = \log p(\boldsymbol{\theta}) + \log p(\sigma^2)$. The logarithm of the prior, $\mathcal{L}_{\boldsymbol{\Theta}} \triangleq \log p(\boldsymbol{\Theta})$, is

$$\begin{aligned}\mathcal{L}_{\boldsymbol{\Theta}} = &-\tfrac{1}{2}\|\boldsymbol{\theta} - \mathbf{m}\|^2_{\mathbf{P}^{-1}} - \tfrac{1}{2}\log|\mathbf{P}| - \log(2\pi) \\ &+ \log\frac{b^a}{\Gamma(a)} - (a+1)\log(\sigma^2) - \frac{b}{\sigma^2},\end{aligned} \tag{14}$$

and its derivative with respect to $\boldsymbol{\theta}$ and σ^2 are:

$$\frac{\partial \mathcal{L}_{\boldsymbol{\Theta}}}{\partial \boldsymbol{\theta}} = -\mathbf{P}^{-1}(\boldsymbol{\theta} - \mathbf{m}), \tag{15a}$$

$$\frac{\partial \mathcal{L}_{\boldsymbol{\Theta}}}{\partial \sigma^2} = \frac{b}{\sigma^4} - \frac{a+1}{\sigma^2}. \tag{15b}$$

Now, the marginal log-posterior is the sum of (11a) and (14), and is given by $\mathcal{L}_{\boldsymbol{\Theta}|\mathbf{y}} = \mathcal{L}_{\mathbf{y}} + \mathcal{L}_{\boldsymbol{\Theta}}$. Since the derivative of the sum of two functions is the sum of the derivatives, gradients of the marginal log-posterior are straightforward to compute, for example $\partial \mathcal{L}_{\boldsymbol{\Theta}|\mathbf{y}}/\partial \boldsymbol{\theta} = \partial \mathcal{L}_{\mathbf{y}}/\partial \boldsymbol{\theta} + \partial \mathcal{L}_{\boldsymbol{\Theta}}/\partial \boldsymbol{\theta}$.

B. Heteroscedastic GP Model

Instead of assuming iid. measurement noise as in (1), in the following we assume input-dependent noise where the noise is assumed to be zero-mean Gaussian but its variance depends on $\mathbf{x}$. We follow the works in [22]–[25] and place a GP prior on the log noise levels to assure that the variance is positive. The considered heteroscedastic Gaussian process (HGP) model can be expressed as

$$y_i \sim \mathcal{N}(f(\mathbf{x}_i), \sigma^2(\mathbf{x}_i)), \tag{16a}$$

$$f(\mathbf{x}) \sim \mathcal{GP}(\mathbf{h}(\mathbf{x})\boldsymbol{\theta}, k_f(\mathbf{x}, \mathbf{x}')), \tag{16b}$$

$$\log(\sigma^2(\mathbf{x})) \sim \mathcal{GP}(\mu_0, k_{\sigma^2}(\mathbf{x}, \mathbf{x}')). \tag{16c}$$

Both GPs use the squared exponential covariance function in (5) such that the model is fully specified by the EM parameters $\boldsymbol{\theta}$, GP hyperparameters $\boldsymbol{\phi}_f$ and $\boldsymbol{\phi}_{\sigma^2}$ and the noise mean hyperparameter μ_0. Assuming a zero-mean GP prior for the noise would impose an arbitrary and high noise level for the RSS ($\exp(0) = 1\,\text{dB}^2$) and therefore, μ_0 is explicitly modeled to control the scale of the noise process [24].

Since the GP prior through the exponential function is not a conjugate prior for the Gaussian, the problem is no longer analytically tractable and approximations are required. The fully Bayesian treatment was considered in [22] using Markov chain Monte Carlo (MCMC), the work in [23] considered the computationally efficient MAP approach instead, and also approximations relying on variational Bayes [24] and expectation propagation [25] have been proposed. In this paper, we utilize the MAP approach introduced in [23]. By placing a GP prior on $f(\mathbf{x})$ and assuming a noise rate function $\sigma^2(\mathbf{x})$, the predictive density is given in (3) with predictive mean and variance, given by [22], [23]

$$\mathrm{E}[f(\mathbf{x}_*)] = \mathbf{h}(\mathbf{x}_*)\boldsymbol{\theta} + \mathbf{k}_*^\top(\mathbf{K} + \boldsymbol{\Sigma})^{-1}(\mathbf{y} - \mathbf{H}\boldsymbol{\theta}), \tag{17a}$$

$$\mathrm{V}[f(\mathbf{x}_*)] = k(\mathbf{x}_*, \mathbf{x}_*) + \sigma^2(\mathbf{x}_*) - \mathbf{k}_*^\top(\mathbf{K} + \boldsymbol{\Sigma})^{-1}\mathbf{k}_*, \tag{17b}$$

where $\boldsymbol{\Sigma} = \operatorname{diag}(\begin{bmatrix}\sigma^2(\mathbf{x}_1) & \sigma^2(\mathbf{x}_2) & \dots & \sigma^2(\mathbf{x}_n)\end{bmatrix})$ and the other parameters are as given in (4a)–(4b). It is to be noted that the expression for the predictive mean slightly differs from [22], [23] since the works assume a zero-mean GP prior on $f(\mathbf{x})$.

The hyperparameters of the model cannot be learned directly from data and we resort to an iterative expectation–maximization type algorithm proposed in [23] which we shortly summarize in the following:

1. Estimate a homoscedastic GP model, denoted by $\mathcal{GP}_1$, using $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$ and as presented in Section III-A.
2. Compute logarithm of the empirical noise levels, i.e., $z_i = \log(\mathrm{Var}(y_i, \mathcal{GP}_1))$ using $\mathcal{GP}_1$ (see [23] for details).
3. Form a new data set $\mathcal{D}' = \{(\mathbf{x}_i, z_i)\}_{i=1}^n$ and use $\mathcal{D}'$ to estimate a second GP, denoted by $\mathcal{GP}_2$.
4. Use $\mathcal{GP}_2$ to predict the logarithm of the noise levels (i.e., $\log(\sigma^2(\mathbf{x}))$) and thereafter, use $\mathcal{D}$ again to estimate the heteroscedastic GP model denoted by $\mathcal{GP}_3$.
5. If not converged, set $\mathcal{GP}_1 = \mathcal{GP}_3$ and return to step 2.

For more details on learning the hyperparameters, the reader is referred to [23].

C. Heteroscedastic Exponential Model

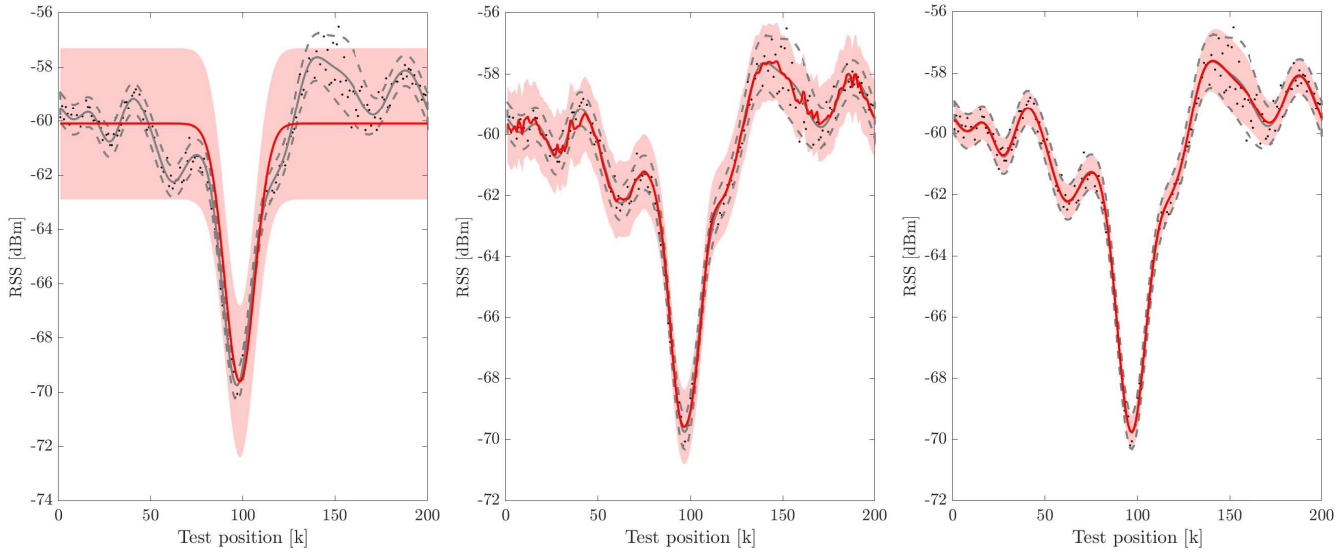
In the following, we model the noise variance using a GP according to (16c) but $f(\cdot)$ is modeled using the parametric model in (7). The proposed heteroscedastic exponential model (HEM) can be expressed using (16a) with $f(\mathbf{x}_i) = \mathbf{h}(\mathbf{x}_i)\boldsymbol{\theta}$ and $\sigma^2(\mathbf{x}_i)$ as in (16c). Since the mean is a deterministic function, the predictive mean and variance are given by

$$\mathrm{E}[f(\mathbf{x}_*)] = \mathbf{h}(\mathbf{x}_*)\boldsymbol{\theta}, \tag{18a}$$

$$\mathrm{V}[f(\mathbf{x}_*)] = \sigma^2(\mathbf{x}_*). \tag{18b}$$

The procedure to train the model parameters and the GP hyperparameters differs from the algorithm presented in Section III-B. In the first step, instead of estimating $\mathcal{GP}_1$ we estimate EM_1, which is a homoscedastic exponential model given in (1) and (7). If we assume that we have a prior over the parameters as in Section III-A and $\boldsymbol{\Theta} = [\boldsymbol{\theta}, \sigma^2]$, we can maximize the marginal log-posterior

$$\log p(\boldsymbol{\Theta} \mid \mathbf{y}) = \log p(\mathbf{y} \mid \boldsymbol{\Theta}) + \log p(\boldsymbol{\theta}) + \log p(\sigma^2), \tag{19}$$



(a) Benchmark EM (b) Benchmark GP (c) HGP

Fig. 2. Example model predictions using the benchmark EM, benchmark GP and the proposed HGP model. In the figures, the gray solid line illustrates the true RSS signal, the dashed gray lines visualize the 95% confidence interval, the black dots represent the measurements, the model prediction shown with the solid red line and the predicted 95% confidence interval illustrated using the transparent red region.

in closed-form. This is achieved by computing the gradients of $\log p(\boldsymbol{\Theta} \mid \mathbf{y})$ with respect to the parameters

$$\frac{\partial \mathcal{L}_{\boldsymbol{\Theta}|\mathbf{y}}}{\partial \boldsymbol{\theta}} = \frac{1}{2\sigma^2}\mathbf{H}^\top(\mathbf{y} - \mathbf{H}\boldsymbol{\theta}) - \frac{1}{2\sigma^2}\mathbf{P}^{-1}(\boldsymbol{\theta} - \mathbf{m}), \tag{20a}$$

$$\frac{\partial \mathcal{L}_{\boldsymbol{\Theta}|\mathbf{y}}}{\partial \sigma^2} = \frac{1}{2\sigma^4}\left(\|\mathbf{y} - \mathbf{H}\boldsymbol{\theta}\|^2 + \|\boldsymbol{\theta} - \mathbf{m}\|^2_{\mathbf{P}^{-1}} + 2b\right) - \frac{\tilde{n}}{2\sigma^2}, \tag{20b}$$

where $\tilde{n} = n + 2 + 2(a+1)$. Then, setting the gradients to zero and solving for the parameters independently yields the MAP estimates

$$\hat{\boldsymbol{\theta}} = \mathbf{A}^{-1}\mathbf{b}, \tag{21a}$$

$$\hat{\sigma}^2 = \frac{1}{\tilde{n}}\left(\mathbf{y}^\top\mathbf{y} - \mathbf{b}^\top\mathbf{A}^{-1}\mathbf{b} + \mathbf{m}^\top\mathbf{P}^{-1}\mathbf{m} + 2b\right), \tag{21b}$$

where $\mathbf{A} = \mathbf{H}^\top\mathbf{H} + \mathbf{P}^{-1}$ and $\mathbf{b} = \mathbf{H}^\top\mathbf{y} + \mathbf{P}^{-1}\mathbf{m}$. Note that the maximum likelihood estimate (MLE) proposed in [26] can be easily obtained from (21) using noninformative priors $\mathbf{P}^{-1} = \mathbf{0}$ and $a, b \to 0$, which yield $\hat{\boldsymbol{\theta}}_{\text{mle}} = \mathbf{A}^{-1}\mathbf{b}$ and $\hat{\sigma}^2_{\text{mle}} = \frac{1}{n}(\mathbf{y}^\top\mathbf{y} - \mathbf{b}^\top\mathbf{A}^{-1}\mathbf{b})$, where $\mathbf{A} = \mathbf{H}^\top\mathbf{H}$ and $\mathbf{b} = \mathbf{H}^\top\mathbf{y}$.

After computing $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}^2$, we can proceed to estimate hyperparameters of the noise process. The second and third step of the procedure are performed as explained in Section III-B with the exception that $\mathcal{GP}_1$ is replaced by EM_1. Thereafter, the algorithm is terminated and the obtained GP model can be used to compute the predictive variance in (18b).

D. Discussion

The proposed HGP model is visualized in Fig. 1a and the model expresses that the RSS is close to a global linear model with the residuals as well as the noise being modeled by GPs. The explicit mean function is illustrated in Fig. 1b and it captures the average received power and target induced perturbation when the bistatic range is small [27]. The covariance function in (16b), with an example realization illustrated in Fig. 1c, is a measure of the similarity between different residual values and it captures various propagation phenomena that cannot be explained by the exponential model. These include diffraction effects [13], [16], new multipath created by reflection from the target [11], and perturbation of existing multipath [3], [18]. Lastly, the input-dependent measurement noise is modeled using a second GP and an example realization is visualized in Fig. 1d. In many real-world applications, including the one considered in this paper, the assumption of constant variance can be unrealistic, and it is highly desirable to consider models with input-dependent variance, as we will display in Section VI.

To demonstrate properties of the proposed HGP and benchmark models, we create a synthetic RSS data set, train the models on that data, and compare the model predictions to the ground truth. In the following, RSS is modeled according to (16a)–(16c) with parameters $\boldsymbol{\theta} = [-60, -10]^\top$, $\boldsymbol{\phi}_f = [2, (2\lambda_c)^2]^\top$, $\mu_0 = -2$ and $\boldsymbol{\phi}_{\sigma^2} = [2, (4\lambda_c)^2]^\top$ in which $\lambda_c = 12.5\,\text{cm}$ is the wavelength. The ground truth RSS values are visualized Fig. 1a and the different models are trained using $\mathcal{D} = \{\mathbf{x}_i, y_i\}_{i=1}^n$ in which $\mathbf{x}_i$'s evenly cover the $5 \times 5\,\text{m}^2$ area at $0.1\,\text{m}$ intervals so that $n = 2601$. The equispaced test trajectory is illustrated in Fig. 1a using a black dashed line, and the RSS and model predictions along the trajectory are shown in Fig. 2. As illustrated, EM is neither capable of explaining the RSS nor the variance very accurately. The benchmark GP model clearly outperforms the parametric model but the model clearly overestimates the variance in some areas and underestimates it in others. The proposed HGP model accurately captures the signal statistics,

TABLE I
MODEL SUMMARY AND EQUATIONS TO COMPUTE THE PREDICTIVE MEAN $\mu_{l,k}$ AND PREDICTIVE VARIANCE $\sigma^2_{l,k}$

Name	Abbr.	Model	$\mu_{l,k}$	$\sigma^2_{l,k}$
Benchmark EM	EM	(1) and (7)	(7)	(1)
Benchmark GP	BGP	(1)–(2)	(4a)	(4b)†
Homoscedastic GP	GP	(1) and (9)	(10)	(4b)†
Heteroscedastic GP	HGP	(16a)–(16c)	(17a)	(17b)
Heteroscedastic EM	HEM	(16a), (7) and (16c)	(18a)	(18b)

†Predictive variance is computed as $\sigma^2_{l,k} = \mathrm{V}[f_l(\mathbf{x}_k)] + \sigma^2_l$, since we must include the noise variance σ^2_l as we are predicting noisy $y_{l,k}$ [20]

since the estimated mean and variance closely resemble the ground truth values. It is also interesting that the mean of the proposed model is smoother than the mean of BGP which is quite jagged—possibly due to the fact that BGP uses three covariance functions.

IV. LOCALIZATION AND TRACKING

In the previous sections, modeling the RSS and training the model parameters for a single link has been considered. However, to enable accurate inference, a single link may not be sufficient. Therefore, S wireless sensors are deployed in the monitored area, $\mathcal{P} = \{\mathbf{p}_j \in \mathcal{A} \mid j = 1, \ldots, S\}$, and each sensor can transmit and receive to other sensors of the network. Assuming full connectivity, the considered *multistatic sensing system* has $L = S \times (S-1)$ unique bistatic links, indexed by $l \in \{1, 2, \ldots, L\}$ in the following. Since the model parameters are unique for every link, training is performed independently for the links [3]. The presented localization and tracking methods require computing the predictive mean and variance for link l and time instant k, denoted from now on using $\mu_{l,k}$ and $\sigma^2_{l,k}$, respectively. The different models are summarized in Table I as well as the equations to compute $\mu_{l,k}$ and $\sigma^2_{l,k}$.

A. Localization

To localize a target (i.e., estimate its position) at time instant k given previously unseen data for L links $\mathbf{y}_k = \{y_{1,k}, y_{2,k}, \ldots, y_{L,k}\}$, we can maximize the predictive log marginal likelihood with respect to the unknown $\mathbf{x}_k$. The predictive marginal likelihood for L links is given by [20]

$$p(\mathbf{y}_k \mid \mathbf{x}_k) = \prod_{l=1}^{L} \mathcal{N}(y_{l,k}; \mu_{l,k}, \sigma^2_{l,k}), \tag{22}$$

where $\mu_{l,k}$ and $\sigma^2_{l,k}$ are given in Table I. Localization amounts then to maximizing the log marginal likelihood according to

$$\hat{\mathbf{x}}_k = \arg\max_{\mathbf{x}_k} \left(\log p(\mathbf{y}_k \mid \mathbf{x}_k) \right), \tag{23}$$

in which the predictive log marginal likelihood is given by

$$\log p(\mathbf{y}_k \mid \mathbf{x}_k) = -\frac{1}{2} \sum_{l=1}^{L} \log(2\pi\sigma^2_{l,k}) + \frac{(y_{l,k} - \mu_{l,k})^2}{\sigma^2_{l,k}}. \tag{24}$$

B. Tracking

Recursive Bayesian filtering computes the marginal posterior distribution, $p(\mathbf{x}_k \mid \mathbf{y}_{1:k})$, of the state $\mathbf{x}_k$ at time step k, given the sequence of measurements $\mathbf{y}_{1:k} = \{\mathbf{y}_1, \ldots, \mathbf{y}_k\}$ up to time instant k. Starting from the posterior at time step $k-1$, the recursive equations to compute the marginal posterior distribution is given by the following Bayesian filtering equations. The prediction step of the Bayesian filtering recursion can be computed using the Chapman-Kolmogorov equation [28]

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} \mid \mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1}, \tag{25}$$

in which $p(\mathbf{x}_k \mid \mathbf{x}_{k-1})$ is the transition density. Once measurement $\mathbf{y}_k$ is available at time k, the predictive distribution can be updated using the Bayes' rule [28]

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_k \mid \mathbf{x}_k) p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1})}{\int p(\mathbf{y}_k \mid \mathbf{x}_k) p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1}) d\mathbf{x}_k}, \tag{26}$$

where $p(\mathbf{y}_k \mid \mathbf{x}_k)$ is the likelihood function. The considered problem does not admit a closed-form solution to the Bayesian filtering recursion and therefore, three different approximations are presented in the following.

1) Grid-based Filter (GF): In the considered problem, the continuous state space can be decomposed into N cells $\{\mathbf{x}^i_k : 1, \ldots, N\}$ in which each cell represents a possible target location. Discretization of the state space allows us to approximate the posterior density using a grid-based filter (GF) as follows. Let us assume that the approximation of the posterior at time $k-1$ is given by [29]

$$p(\mathbf{x}_{k-1} \mid \mathbf{y}_{1:k-1}) \approx \sum_{i=1}^{N} w^i_{k-1|k-1} \delta(\mathbf{x}_{k-1} - \mathbf{x}^i_{k-1}), \tag{27}$$

where $w^i_{k-1|k-1} \triangleq \Pr(\mathbf{x}_{k-1} = \mathbf{x}^i_{k-1} \mid \mathbf{y}_{1:k-1})$ is the conditional probability of state $\mathbf{x}_{k-1}$ given measurement sequence $\mathbf{y}_{1:k-1} = \{\mathbf{y}_1, \ldots, \mathbf{y}_{k-1}\}$, and $\delta(\cdot)$ denotes the Dirac delta function. The prediction step of the GF is given by [29]

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1}) \approx \sum_{i=1}^{N} w^i_{k|k-1} \delta(\mathbf{x}_k - \mathbf{x}^i_k), \tag{28}$$
$$w^i_{k|k-1} \triangleq \sum_{j=1}^{N} w^j_{k-1|k-1} p(\mathbf{x}^i_k \mid \mathbf{x}^j_{k-1}),$$

where $p(\mathbf{x}^i_k \mid \mathbf{x}^j_{k-1})$ is the transition density. After observing $\mathbf{y}_k$, the GF posterior can be updated according to [29]

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) \approx \sum_{i=1}^{N} w^i_{k|k} \delta(\mathbf{x}_k - \mathbf{x}^i_k), \tag{29}$$
$$w^i_{k|k} \approx \frac{w^i_{k|k-1} p(\mathbf{y}_k \mid \mathbf{x}^i_k)}{\sum_{j=1}^{N} w^j_{k|k-1} p(\mathbf{y}_k \mid \mathbf{x}^j_k)}$$

where the likelihood, $p(\mathbf{y}_k \mid \mathbf{x}^i_k)$, is computed using (22). The cell in which the posterior has a global maximum is then used as the location estimate, mathematically expressed as

$$\hat{\mathbf{x}}_k = \arg\max_{\mathbf{x}_k} p(\mathbf{x}_k \mid \mathbf{y}_{1:k}). \tag{30}$$

In this paper, we use a Gaussian transition density

$$p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k \mid \mathbf{F}_{\mathrm{gf}} \mathbf{x}_{k-1}, \mathbf{Q}_{\mathrm{gf}}) \tag{31}$$

in which $\mathbf{F}_{\text{gf}}$ and $\mathbf{Q}_{\text{gf}}$ are the time-invariant transition matrix and process noise covariance, respectively.

2) Kalman Filter (KF): Next, we present a two-step approach for tracking the target. In the first step, (23) is computed at time step k in order to obtain a positioning measurement $\mathbf{z}_k \leftarrow$ (23). In the second step, $\mathbf{z}_k$ is used as a measurement input to a Kalman filter (KF) which recursively computes the marginal posterior distribution, $p(\mathbf{x}_k \mid \mathbf{z}_{1:k})$, of state $\mathbf{x}_k$.

The Bayesian filtering equations are recursively computed using the KF algorithm as presented in [28, Eq. (4.19)] which requires that the posterior, transition density and likelihood function are Gaussian

$$p(\mathbf{x}_k \mid \mathbf{z}_{1:k}) = \mathcal{N}(\mathbf{x}_k \mid \hat{\mathbf{x}}_k, \mathbf{P}_k), \tag{32}$$

$$p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k \mid \mathbf{F}_{\text{kf}}\mathbf{x}_{k-1}, \mathbf{Q}_{\text{kf}}), \tag{33}$$

$$p(\mathbf{z}_k \mid \mathbf{x}_k) = \mathcal{N}(\mathbf{z}_k \mid \mathbf{H}_{\text{kf}}\mathbf{x}_k, \mathbf{R}_{\text{kf}}). \tag{34}$$

In (32), $\hat{\mathbf{x}}_k$ and $\mathbf{P}_k$ denote the posterior mean and covariance, respectively; in (33), $\mathbf{F}_{\text{kf}}$ is the transition matrix and $\mathbf{Q}_{\text{kf}}$ the process noise covariance; and in (34), $\mathbf{H}_{\text{kf}}$ is the measurement model and $\mathbf{R}_{\text{kf}}$ the measurement noise covariance.

It is to be noted that even though the Kalman filter is the closed-form solution to the Bayesian filtering equations, the proposed two-step approach is expected to perform suboptimally. The reason is that the conversion from nonlinear RSS measurements to linear position measurements introduces a notable discretization error and also transforms the measurement noise so that $p(\mathbf{z}_k \mid \mathbf{x}_k)$ is unlikely Gaussian.

3) Particle Filter (PF): The PF uses a weighted set of N particles $\{(\mathbf{x}_k^i, w_k^i) : 1, \ldots, N\}$ for representing the posterior distribution [30]

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) \approx \sum_{i=1}^{N} w_k^i \delta(\mathbf{x}_k - \mathbf{x}_k^i), \tag{35}$$

and the weights satisfy the recursion

$$w_k^i \propto w_{k-1}^i \frac{p(\mathbf{y}_k \mid \mathbf{x}_k^i) p(\mathbf{x}_k^i \mid \mathbf{x}_{k-1}^i)}{q(\mathbf{x}_k^i \mid \mathbf{x}_{k-1}^i, \mathbf{y}_k)}. \tag{36}$$

In (36), $p(\mathbf{y}_k \mid \mathbf{x}_k^i)$ is the likelihood given in (22), $p(\mathbf{x}_k^i \mid \mathbf{x}_{k-1}^i)$ is the transition density, and $q(\mathbf{x}_k^i \mid \mathbf{x}_{k-1}^i, \mathbf{y}_k)$ is the importance distribution. The posterior distribution is approximated using the bootstrap PF [31], in which the dynamic model $p(\mathbf{x}_k^i \mid \mathbf{x}_{k-1}^i)$ is used as the importance density, and the algorithm at time instant k is implement as follows:

1) For every particle, sample from the transition density

$$\mathbf{x}_k^i \sim p(\mathbf{x}_k^i \mid \mathbf{x}_{k-1}^i), \quad i = 1, \ldots, N. \tag{37}$$

2) Compute the weights

$$w_k^i \propto w_{k-1}^i p(\mathbf{y}_k \mid \mathbf{x}_k^i), \quad i = 1, \ldots, N. \tag{38}$$

3) Normalize the weights $w_k^i = w_k^i / \sum_j^N w_k^j$ and resample if the effective number of particles is below threshold T_{eff}

$$\frac{1}{\sum_i^N (w_k^i)^2} < T_{\text{eff}}. \tag{39}$$

The PF is implemented using a Gaussian transition density

$$p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k \mid \mathbf{F}_{k-1}\mathbf{x}_{k-1}, \mathbf{Q}_{k-1}) \tag{40}$$

in which $\mathbf{F}_{k-1}$ and $\mathbf{Q}_{k-1}$ are the time-variant transition matrix and process noise covariance, respectively.

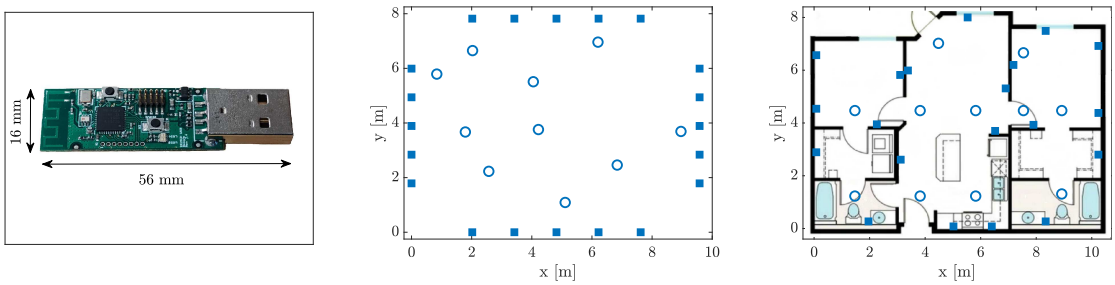
V. Experiments

A. Experimental Setup

We utilize a publicly available dataset [27] which we summarize in the following for the sake of completeness. The development efforts of this paper are validated in two different experimental environments and using 20 sensors. The used sensors are Texas Instruments CC2531 USB dongles [32] shown in Fig. 3a. The sensors communicate in a round-robin schedule in which one sensor broadcasts at a time while the other sensors listen, measure the RSS and append the RSS to the payload of the next packet they transmit. The sensors communicate on the 2.4 GHz ISM band ($\lambda_c \approx 12.5$ cm) using all 16 frequency channels defined by the IEEE 802.15.4 standard [33] and on average, the time interval between transmissions is $\tau \approx 2.9$ ms. A base station (BS) that listens to the ongoing transmissions extracts the RSS from the packets for centralized processing. The considered multistatic sensing system has $20 \cdot (20-1) \cdot 16 = 6080$ unique bistatic links but it is to be noted that full connectivity nor multi-channel communication is not mandatory. The used hardware and communication protocol are further explained in [34].

The first experimental environment is an open-space indoor environment in which each sensor has LOS propagation conditions to every other sensor, and multipath propagation is scarce. The sensors are set on top of podiums at a height of 90 cm and deployed around a 75 m^2 area as illustrated in Fig. 3b. The second experimental environment, shown in Fig. 3c, is a fully furnished two bedroom downtown apartment in which multipath propagation is severe, and NLOS propagation conditions prevail. The sensors are deployed evenly throughout the 82 m^2 apartment and installed by electric sockets of the apartment at various heights—most of the electric sockets locate slightly above floor level while some locate above bathroom and kitchen countertops.

Reference positions are measured and marked in both experimental environments (see Figs. 3b and 3c). During the experiment, the target stands still for a few seconds in one of the reference positions after which the target walks to a different reference position along a straight line and trying to maintain a constant speed. During the experiment the target visits the reference positions in no particular order and some reference positions are visited multiple times. In both environments, three different trials are conducted and each trial is approximately three minutes long. The target is carrying a video camera and the video is used in post-processing to define the departure and arrival times at every reference position. Subsequently, the ground truth trajectory is generated by interpolating between the reference position coordinates and annotated time instances. Since this method is subject to small errors, the localization and tracking accuracy



(a) Sensor (b) Open-space (c) Apartment

Fig. 3. In (a), the used TI CC2541 sensor. In (b) and (c), the experimental layouts in which the reference positions (○) and the sensors (■) are shown.

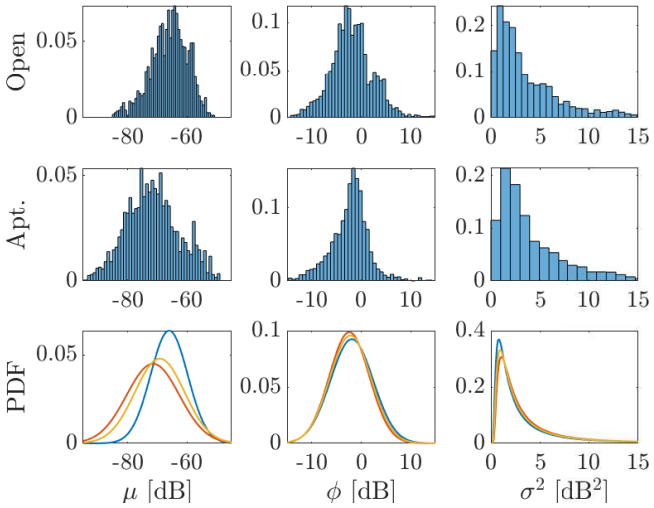


Fig. 4. Normalized empirical histograms of the estimated EM parameters (μ, ϕ, σ^2) shown on the top (open ex.) and middle row (apt. ex.). To the estimated model parameters, a suitable distribution is fitted and the result is shown on the bottom row of the figure. The legend is: open-space indoor experiment (——), apartment experiment (——) and open and apartment experiments combined (——).

are evaluated only when the target remains stationary to ensure the ground truth position is accurate.

B. Training

The data that is gathered when the target is moving is used for training the models, whereas the data collected when the target is stationary is used for validation. Since the exact departure and arrival times are hard to determine from the videos, the half a second time period before the target starts to move and after the target stops are removed from the validation data and appended to the training data. The aggregated training and validation time periods ranged between 101-111 and 63-77 seconds, respectively. To put this in terms of RSS measurements per link, the average number of training and validation measurements ranged between 106-113 and 64-83 samples, respectively.

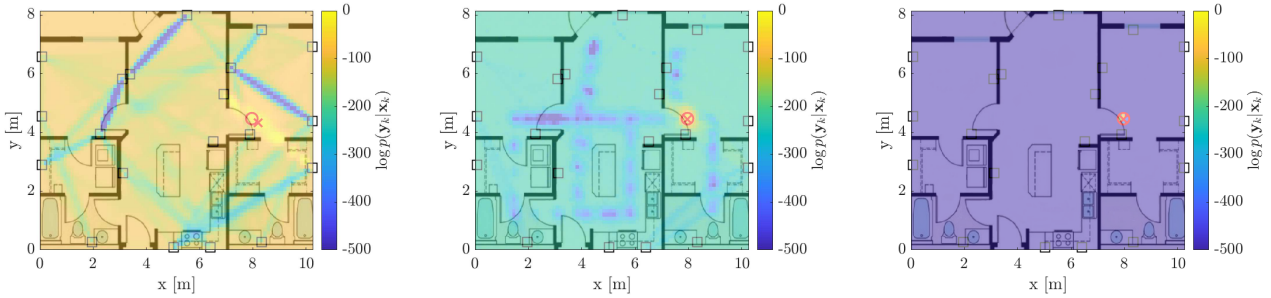
The model parameter priors, $p(\boldsymbol{\Theta}_l)$ in which $\boldsymbol{\Theta}_l = [\boldsymbol{\theta}_l, \sigma_l^2]$, $l \in \{1, \ldots, L\}$, are determined from experimental data as follows. We use another similar set of experiments but using 20 different TI CC2531 USB dongles that communicate on four frequency channels instead. The experimental environments and sensor locations are the same as described in Section V-A. The MLE of the parameters are computed using (21) with noninformative priors. Histograms of the estimated model parameters are visualized on the top and middle row of Fig. 4. Thereafter, the model parameter estimates from the open-space and apartment experiments are combined and a distribution is fitted to the combined data. Lastly, the fitted distribution is used as the prior when training the models. The model parameter priors are: $\mu \sim \mathcal{N}(-69.21, 69.63)$, $\phi \sim \mathcal{N}(-2.20, 17.36)$ and $\sigma^2 \sim \Gamma^{-1}(1.19, 2.02)$. It is to be noted that variances of the priors are relatively high such that the priors are not very informative.

During the development phase of the algorithms, we found out that the proposed HGP model performs better, if the predictive mean of the noise levels is used as the predictive variance, instead of using the expression given in (17b). We believe the problem is with the gradient-based optimizer which has a tendency of overfitting the model since the training data is sparse. In future research, this problem will be further investigated and other methods for parameter inference such as MCMC and Hamiltonian Monte Carlo (HMC) methods [35] will be tested.

C. Experimental Parameters

In the next section, we evaluate the performance of five different RSS models as well as four different localization and tracking algorithms. The parameters of the models and algorithms are defined as follows.

- *Model Parameters* – these are determined from training data except the length scale parameters of the EM and GP models. As in [3], the length scale parameters are fixed to $\ell_{\text{em}} = \lambda_c/4$ and $\ell_{\text{se}} = 2\lambda_c$.
- *Localization* – The predictive marginal likelihood in (22) is computed using $L = 380$ and the MLE in (23) is computed using a grid search with $\delta_{\mathbf{x}} = 12.5\,\text{cm}$ grid spacing. The number of links L corresponds to one round-robin cycle, that is, one transmission by every sensor meaning that the MLE is computed once every 58 ms.
- *Grid Filter (GF)* – As in *localization*, the likelihood in (22) is computed using $L = 380$, the grid spacing is $\delta_{\mathbf{x}} = 12.5\,\text{cm}$ and the sampling rate is $\tau_{\text{gf}} = 58\,\text{ms}$. The GF is implemented using a Gaussian random walk model as the transition density with the power spectral density of the process noise in (31) set to $q_{\text{gf}} = 0.01$ such that $\mathbf{Q}_{\text{gf}} = q_{\text{gf}}\tau_{\text{gf}}\mathbf{I}_2$ and $\mathbf{F}_{\text{gf}} = \mathbf{I}_2$.



(a) Benchmark EM (b) Benchmark GP (c) Proposed HGP model

Fig. 5. The predictive log marginal likelihoods of the three models in one example position. The ground truth location illustrated using (○) and the estimate with (×).

- *Kalman Filter (KF)* – The MLE in (23) is computed using the same parameters as in *localization* and the MLE is used as measurement input to the KF. The KF is implemented using a second-order kinematic model, given by [36]

$$\mathbf{F}_{\text{kf}} = \mathbf{I}_2 \otimes \begin{bmatrix} 1 & \tau_{\text{kf}} \\ 0 & 1 \end{bmatrix}, \quad \mathbf{Q}_{\text{kf}} = \mathbf{I}_2 \otimes q_{\text{kf}} \begin{bmatrix} \frac{1}{3}\tau_{\text{kf}}^3 & \frac{1}{2}\tau_{\text{kf}}^2 \\ \frac{1}{2}\tau_{\text{kf}}^2 & \tau_{\text{kf}} \end{bmatrix}, \tag{41}$$

  where $\otimes$ denotes the Kronecker product, $\tau_{\text{kf}} = 58\,\text{ms}$ is the sampling rate and $q_{\text{kf}} = 1.0$ is the power spectral density of the process noise. The measurement model matrix in (34) is $\mathbf{H}_{\text{kf}} = \mathbf{I}_2 \otimes \begin{bmatrix} 1 & 0 \end{bmatrix}$ and covariance of the measurement noise is $\mathbf{R}_{\text{kf}} = \sigma_{\text{kf}}^2 \mathbf{I}_2$ in which variance of the measurement noise is $\sigma_{\text{kf}}^2 = 0.1\,\text{m}^2$.
- *Particle Filter (PF)* – Non-linear filters enable the measurements to be processed sequentially according to their true transmission time instant allowing a more accurate description of the time evolution [26]. Thus, the likelihood in (22) is computed using $L = 19$ and the PF sampling rate is set equal to the transmission interval, that is, $\tau_{\text{pf}} \approx 2.9\,\text{ms}$. The PF is implemented using a Gaussian random walk model as the transition density with the spectral density of the white noise in (40) set to $q_{\text{pf}} = 0.1$ such that $\mathbf{Q}_{\text{pf}} = q_{\text{pf}}\tau_{\text{pf}}\mathbf{I}_2$ and $\mathbf{F}_{\text{pf}} = \mathbf{I}_2$. The number of used particles is $N = 10$ and the effective sample size threshold is $T_{\text{eff}} = N/2$. We also tested the second-order kinematic model with the PF, but the model did not provide any significant advantage over the random walk model while requiring significantly more particles because of the higher state dimension.

D. Evaluation Metrics

The following measures are computed for assessing the performance of different methods: the standardized mean squared error (SMSE) and the localization root mean squared error (RMSE). The SMSE measures the modeling accuracy and it is defined as:

$$e_{\mathbf{y}} = \frac{1}{KL} \sum_{k=1}^{K} \sum_{l=1}^{L} \frac{(y_{l,k} - \mu_{l,k})^2}{\text{Var}[\mathbf{y}_l]}, \tag{42}$$

where $\mu_{l,k}$ is the predictive mean (see Table I) for the test sample and $y_{l,k}$ is the actual RSS value. The training data variance for link l is denoted by $\text{Var}[\mathbf{y}_l]$. The localization and tracking accuracy is assessed using the RMSE and it is defined as:

$$e_{\mathbf{x}} = \left(\frac{1}{K} \sum_{k=1}^{K} \|\mathbf{x}_k - \hat{\mathbf{x}}_k\|_2^2 \right)^{\frac{1}{2}} \tag{43}$$

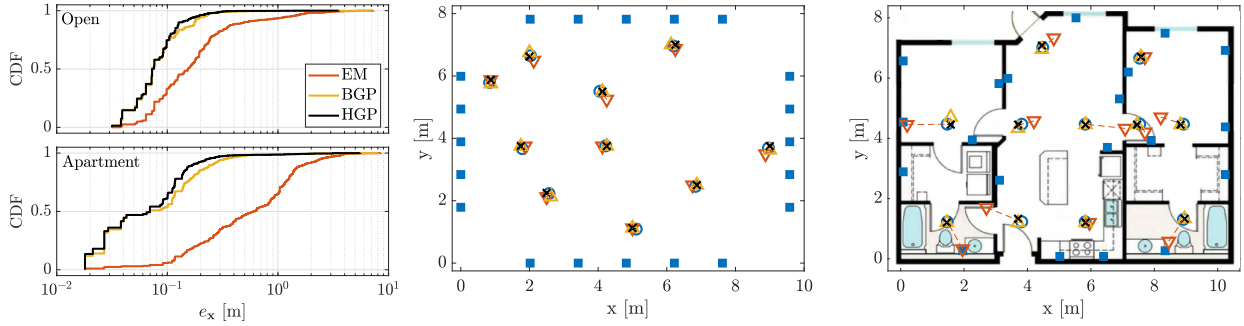
where $\mathbf{x}_k$ and $\hat{\mathbf{x}}_k$ denote the ground truth location and its estimate for test location k in respective order, and $\|\cdot\|_2^2$ denotes the square of the Euclidean norm.

VI. Experimental Results

A. Modeling Results

The parametric EM model is widely used in DFLT and the model is able to capture RSS changes when the target is blocking or in the vicinity of the link line. A downside of the model is that it is unable to explain, for example, single-bounce reflections and perturbations to existing multipath components. The BGP model can capture various signal propagation patterns, including link line blockage, single-bounce reflections, as well as perturbations of existing NLOS paths [3]. The proposed HGP model has similar expressiveness as BGP since the EM model is used as the mean function and the covariance function captures the more complex signal propagation patterns that cannot be explained with simple link line geometry. The modeling errors are tabulated in Table II and we can conclude that the GP models are superior with respect to the EM model. Moreover, the GP models yield similar performance, HGP slightly outperforming BGP. Since (42) is not a function of the predictive variance, the proposed heteroscedastic EM and homoscedastic GP models yield the same modeling accuracy as EM and HGP models, respectively.

Ideally, the predictive log marginal likelihood in (22) should be unimodal so that the peak coincides with the true target location, and elsewhere the value should approach $p(\mathbf{y}_k \mid \mathbf{x}_k) \to 0$ as the distance to the target grows. The predictive log marginal likelihoods of the three models in an example target location are illustrated in Fig. 5. Now, the benefit of the HGP model is apparent, since the predictive log marginal likelihood is very low everywhere except in the close vicinity of the true target location and the peak is in the same cell as the target. The predictive log marginal likelihood of the BGP and EM models are not as concentrated and typically the values



(a) CDFs (b) Open-space (c) Apartment

Fig. 6. In (a), localization error CDFs in the two experimental environments. In (b) and (c), example localization accuracy and the legend of the figures are: sensors (■), reference positions (○), EM model (▽), BGP model (△) and HGP model (×). RMSE of the example position in (b): 16.2 cm (EM), 8.5 cm (BGP), 7.5 cm (HGP) and in (c): 82.9 cm (EM), 11.4 cm (BGP), 8.3 cm (HGP).

TABLE II
SUMMARY OF LOCALIZATION AND MODELING ERRORS

Experiment / Model	Open		Apartment	
	$e_{\mathbf{x}}$ [cm]	e_y [dB]	$e_{\mathbf{x}}$ [cm]	e_y [dB]
EM	62.1 ± 54.6	1.00	120.6 ± 86.5	0.97
BGP	24.3 ± 22.0	0.78	38.4 ± 35.9	0.76
GP†	19.8 ± 16.7	**0.76**	34.3 ± 29.9	**0.74**
HGP†	$\mathbf{12.5 \pm 8.7}$	**0.76**	$\mathbf{29.4 \pm 27.3}$	**0.74**
HEM†	21.2 ± 17.0	1.00	47.9 ± 42.9	0.97

†Model proposed in this paper

are much higher than with the HGP model. Moreover, the predictive log marginal likelihood of the BGP and EM models is more frequently multimodal, which can have significant ramifications to localization which we will elaborate on in the next section.

B. Localization Results

The localization RMSE $\pm$ one standard deviation results are presented in Table II and the Cramér-Rao bound (CRB) for the discretized MLE in (23) is $\delta_{\mathbf{x}}/\sqrt{6} \approx 5.1$ cm [26]. Moreover, the localization error cumulative distribution functions (CDFs) are illustrated in Fig. 6a and example localization performance in the open-space and apartment experiments are illustrated in Fig. 6b and Fig. 6c, respectively. The results indicate that the EM model yields the least accurate localization performance, the proposed HGP model provides the best localization accuracy, and the performance of the other models is somewhere in between. The results suggest that, as the heteroscedastic EM model outperforms the homoscedastic EM model, it may be beneficial to treat the measurement noise as non-iid.. In addition, since the homoscedastic GP model slightly outperforms the homoscedastic BGP model, it is advantageous to explicitly model the mean function instead of using a zero-mean GP prior. The HGP model uses an explicit mean function and accounts for heteroscedastic noise resulting in superior accuracy—the proposed model decreases the localization RMSE by up to 80 % with respect to the EM model and by up to 49 % with respect to the BGP model.

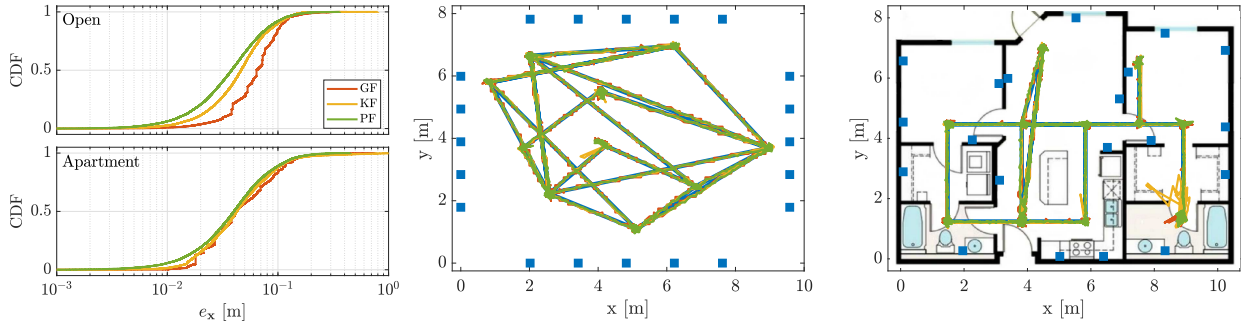
The localization error CDFs illustrated in Fig. 6a reveals the advantages and disadvantages of the models. Clearly, the GP models yield higher localization accuracy than the EM model and this notable advantage mainly comes from the GPs ability to model perturbations of existing multipath components. The difference between BGP and HGP is more subtle, and the main difference is that the error distribution of the BGP model is heavier-tailed. The reason for this behavior stems from the predictive log marginal likelihood of the BGP model which is frequently multimodal. Moreover, the peak closest to the ground truth target location is not always the one that has the highest value and as an outcome, the wrong local maxima is taken as the estimate which can be very inaccurate. The proposed HGP model is not immune to such unwanted behavior but suffers from it less frequently and therefore, the error distribution is not as heavy-tailed as with the BGP model.

C. Tracking Results

The results in the two experimental environments using different models and tracking filters are summarized in Table III and as expected, the accuracy improves when the time evolution is taken into account. The results indicate that the EM model yields the least accurate tracking performance and the proposed HGP model provides the best overall tracking accuracy. In the open-space experiment, the homoscedastic GP model yields the best accuracy using the GF, the HGP is the most accurate using the KF, and the BGP has the lowest RMSE when using the PF. In the apartment experiment, the proposed HGP model yields the lowest RMSE with all filters. It should be noted, however, that the differences between the GP models is rather small especially when considering that the tracked target is not actually a point in space. A more accurate representation of the target is a cylinder and the RMSE of the GP models is within or on par with the spatial extent of the cylinder.

In the following, properties of the different tracking filters are analyzed and for brevity, the analysis only considers the HGP model. The tracking error CDFs are illustrated in Fig. 7a and example tracking performance in the open-space and apartment experiments are illustrated in Fig. 7b and Fig. 7c, respectively. The main difference between the tracking filters is the discretization of the target state and how outlier measurements impact the filter update:

- *Discretization* – The GF divides the monitored area into equally spaced grid cells and the filter is sub-optimal



(a) CDFs (b) Open-space (c) Apartment

Fig. 7. In (a), tracking error CDFs in the two experimental environments computed using the full trajectory. In (b) and (c), example tracking performance and the legend of the figures are: sensors (■), ground truth trajectory (——), GF (——), KF (——) and PF (——). RMSE of the example trajectory in (b): 7.6 cm (GF), 6.7 cm (KF), 5.8 cm (PF) and in (c): 8.2 cm (GF), 15.6 cm (KF), 6.8 cm (PF).

TABLE III
SUMMARY OF TRACKING ERRORS (RMSE $\pm$ STD)

Expt.	Model	GF [cm]	KF [cm]	PF [cm]
Open	EM	12.8 ± 5.9	29.4 ± 20.8	12.3 ± 7.4
Open	BGP	8.9 ± 4.1	11.2 ± 8.6	$\mathbf{7.7 \pm 5.0}$
Open	GP†	$\mathbf{8.8 \pm 3.8}$	10.2 ± 7.2	8.8 ± 4.8
Open	HGP†	9.3 ± 4.4	$\mathbf{6.5 \pm 3.9}$	7.9 ± 4.3
Open	HEM†	11.0 ± 5.3	10.0 ± 6.9	10.6 ± 5.6
Apt.	EM	55.7 ± 40.6	65.4 ± 39.8	24.9 ± 17.0
Apt.	BGP	12.3 ± 8.5	17.4 ± 14.8	10.1 ± 7.0
Apt.	GP†	14.7 ± 10.9	17.4 ± 13.8	11.2 ± 7.0
Apt.	HGP†	$\mathbf{10.6 \pm 7.1}$	$\mathbf{16.3 \pm 14.1}$	$\mathbf{9.5 \pm 5.5}$
Apt.	HEM†	14.1 ± 9.0	28.4 ± 23.2	13.6 ± 7.2

†Model proposed in this paper

since the underlying state space is not truly discrete in nature. The state space of the KF is continuous but discretization also degrades the filter accuracy since the input measurements of the KF are quantized by the MLE in (23). The PF circumvents the discretization problem by representing the continuous state space using a set of particles and therefore, the posterior approximation of the PF is more accurate than with the GF and KF.

- *Outliers* – As already discussed in the previous sections, the predictive log marginal likelihood in (22) is occasionally multimodal and the maximum peak is not always the one closest to the ground truth state. Since the posterior of the GF is proportional to the product of the likelihood and the prior, the prior typically suppresses the peaks that are far away from the tracked target so that the maximum peak is the one closest to the ground truth state. The KF is sensitive to outlier measurements (see e.g. coordinate (9,1) m in Fig. 7c) and therefore, the error distribution of the KF is heavier-tailed than with the GF. The multimodal likelihood function typically does not impact the PF since the likelihood is only evaluated at the particle support points which are concentrated near the ground truth state as long as the prior is not biased.

Using a Matlab implementation on a Dell Precision 3591 computer equipped with a Intel Core Ultra 7 CPU and 64 GB of memory, the time to process a data sequence of 177.9 seconds takes 33.8 s, 19.9 s and 96.7 s with the GF, KF and PF algorithms, respectively. Thus, a real-time implementation of all the algorithms is possible. There exists a trade-off between computational complexity and accuracy since the computation time of the PF algorithm can be reduced using fewer particles, whereas the efficiency of the GF and KF algorithms can be improved by increasing the grid-spacing. As an example, if the grid spacing is doubled ($\delta_{\mathbf{x}} = 25$ cm), the RMSEs increase to 11.0 cm (GF) and 8.7 cm (KF) in the open-space experiments, whereas the processing times decrease down to 4.8 s (GF) and 3.5 s (KF). In the end, the preferred choice of the tracking filter depends on the application. If a low complexity algorithm is desired than a good choice is either the GF or KF, whereas the PF should be used in applications requiring best possible tracking accuracy.

VII. CONCLUSIONS

This paper presented a heteroscedastic GP model for RSS-based DFLT. By modeling the RSS with a normal distribution having a GP prior on both the mean and log noise levels, complex signal propagation patterns and target location dependent measurement noise can be modeled. The explicit basis function of the GP prior on the mean captures the baseline signal strength and target induced RSS changes caused by LOS blockage, whereas the covariance function captures other slow fading phenomena such as perturbations to existing NLOS components that cannot be explained with simple link line geometry. Respectively, the GP prior on the log noise levels captures not only hardware induced errors and interference from co-located wireless networks, but also target location dependent fast fading. In addition, we presented algorithms to train the model parameters as well as solve the RSS-based DFLT problem.

Analysis was carried out in an open-space indoor environment experiment and in a more complex residential apartment experiment in which NLOS propagation conditions prevailed. Results demonstrated that the proposed model improved the localization and tracking accuracy with respect to parametric and non-parametric benchmark models. Moreover, it was demonstrated that real-time sub-decimeter tracking accuracy can be achieved in both experimental environments using a PF and the proposed heteroscedastic GP model. The newly proposed model admits numerous possibilities of future research into

other heteroscedastic GP model approaches, improvements via alternate training procedures, or enhancements via more sophisticated tracking filters. For example, with sparse data the gradient-based optimizer has a tendency to over-fit the model resulting in overconfident predictions and therefore, alternative training procedures such as MCMC [35] will be investigated. Also, other heteroscedastic GP approaches will be explored including variational Bayes [24] and expectation propagation [25]. Lastly, sigma-point filtering [28], iterated posterior linearization [37], and random finite set based filters [8] are potential candidates to improve the accuracy and/or efficiency of tracking, and these methods will be studied in future research.

REFERENCES

[1] J. Xiao, H. Li, M. Wu, H. Jin, M. J. Deen, and J. Cao, "A survey on wireless device-free human sensing: Application scenarios, current solutions, and open issues," *ACM Computing Surveys*, vol. 55, no. 5, Dec. 2022.

[2] M. Schmidhammer, C. Gentner, S. Sand, and U.-C. Fiebig, "Multipath-enhanced device-free localization in wideband wireless networks," *IEEE Antennas and Wireless Propagation Letters*, vol. 20, no. 4, pp. 453–457, 2021.

[3] O. Kaltiokallio, R. Hostettler, J. Talvitie, and M. Valkama, "Gaussian processes for received signal strength based device-free localization," in *2024 18th European Conference on Antennas and Propagation*, 2024, pp. 1–5.

[4] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, "Smart homes that monitor breathing and heart rate," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015, p. 837–846.

[5] M. Schmidhammer, B. Siebler, C. Gentner, S. Sand, and U.-C. Fiebig, "Bayesian multipath-enhanced device-free localisation: Simulation- and measurement-based evaluation," *IET Microwaves, Antennas & Propagation*, vol. 16, no. 6, pp. 327–337, 2022.

[6] N. González-Prelcic, M. Furkan Keskin, O. Kaltiokallio, M. Valkama, D. Dardari, X. Shen, Y. Shen, M. Bayraktar, and H. Wymeersch, "The integrated sensing and communication revolution for 6G: Vision, techniques, and applications," *Proceedings of the IEEE*, vol. 112, no. 7, pp. 676–723, 2024.

[7] N. Patwari, L. Brewer, Q. Tate, O. Kaltiokallio, and M. Bocca, "Breathfinding: A wireless network that monitors and locates breathing in a home," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 1, pp. 30–42, 2014.

[8] O. Kaltiokallio, H. Yiğitler, J. Talvitie, and M. Valkama, "Random finite set approach to signal strength based passive localization and tracking," in *2023 IEEE/ION Position, Location and Navigation Symposium*, 2023, pp. 1215–1225.

[9] J. Wilson and N. Patwari, "Radio tomographic imaging with wireless networks," *IEEE Transactions on Mobile Computing*, vol. 9, no. 5, pp. 621–632, 2010.

[10] B. Wang, Y. Ma, X. Zhang, J. Zhang, and Z. Huang, "InterMap-NRTI: A novel radio tomographic imaging positioning model based on the intersection map of wireless links," *IEEE Transactions on Vehicular Technology*, pp. 1–12, 2025.

[11] H. Yiğitler, R. Jäntti, O. Kaltiokallio, and N. Patwari, "Detector based radio tomographic imaging," *IEEE Transactions on Mobile Computing*, vol. 17, no. 1, pp. 58–71, 2018.

[12] S. Savazzi, M. Nicoli, F. Carminati, and M. Riva, "A Bayesian approach to device-free localization: Modeling and experimental assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 1, pp. 16–29, 2014.

[13] M. Schmidhammer, M. Walter, C. Gentner, and S. Sand, "Physical modeling for device-free localization exploiting multipath propagation of mobile radio signals," in *2020 14th European Conference on Antennas and Propagation*, 2020, pp. 1–5.

[14] Y. Li, X. Chen, M. Coates, and B. Yang, "Sequential Monte Carlo radio-frequency tomographic tracking," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2011, pp. 3976–3979.

[15] Y. Guo, K. Huang, N. Jiang, X. Guo, Y. Li, and G. Wang, "An exponential-Rayleigh model for RSS-based device-free localization and tracking," *IEEE Transactions on Mobile Computing*, vol. 14, no. 3, pp. 484–494, 2015.

[16] V. Rampa, S. Savazzi, M. Nicoli, and M. D'Amico, "Physical modeling and performance bounds for device-free localization systems," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 1864–1868, 2015.

[17] N. Patwari and J. Wilson, "RF sensor networks for device-free localization: Measurements, models, and algorithms," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1961–1973, 2010.

[18] M. Schmidhammer, C. Gentner, M. Walter, S. Sand, B. Siebler, and U.-C. Fiebig, "Empirical fading model and Bayesian calibration for multipath-enhanced device-free localization," *IEEE Transactions on Wireless Communications*, vol. 23, no. 8, pp. 8168–8183, 2024.

[19] C. Xu, B. Firner, Y. Zhang, R. Howard, J. Li, and X. Lin, "Improving RF-based device-free passive localization in cluttered indoor environments through probabilistic classification methods," in *Proceedings of the 11th International Conference on Information Processing in Sensor Networks*, 2012, p. 209–220.

[20] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[21] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin, *Bayesian Data Analysis, Third Edition*. Taylor & Francis, 2013.

[22] P. Goldberg, C. Williams, and C. Bishop, "Regression with input-dependent noise: A Gaussian process treatment," in *Advances in Neural Information Processing Systems*, M. Jordan, M. Kearns, and S. Solla, Eds., vol. 10. MIT Press, 1997.

[23] K. Kersting, C. Plagemann, P. Pfaff, and W. Burgard, "Most likely heteroscedastic Gaussian process regression," in *Proceedings of the 24th International Conference on Machine Learning*, 2007, p. 393–400.

[24] M. Lázaro-Gredilla and M. K. Titsias, "Variational heteroscedastic Gaussian process regression," in *Proceedings of the 28th International Conference on International Conference on Machine Learning*, 2011, p. 841–848.

[25] V. Tolvanen, P. Jylänki, and A. Vehtari, "Expectation propagation for nonstationary heteroscedastic Gaussian process regression," in *2014 IEEE International Workshop on Machine Learning for Signal Processing*, 2014, pp. 1–6.

[26] O. Kaltiokallio, R. Hostettler, and N. Patwari, "A novel Bayesian filter for RSS-based device-free localization and tracking," *IEEE Transactions on Mobile Computing*, vol. 20, no. 3, pp. 780–795, 2021.

[27] O. Kaltiokallio, R. Hostettler, H. Yiğitler, and M. Valkama, "Unsupervised learning in RSS-based DFLT using an EM algorithm," *Sensors*, vol. 21, no. 16, 2021.

[28] S. Särkkä and L. Svensson, *Bayesian Filtering and Smoothing, Second Edition*. Cambridge University Press, 2023.

[29] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[30] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–2008, 2000.

[31] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, pp. 107–113, 1993.

[32] Texas Instruments CC2531 USB Dongle, [Online, accessed 20 December 2024]. Available: https://www.ti.com/tool/CC2531EMK.

[33] IEEE 802.15.4-2003 standard, [Online, accessed 20 December 2024]. Available: https://www.ieee802.org/15/pub/TG4Expert.html.

[34] M. Bocca, O. Kaltiokallio, and N. Patwari, "Radio tomographic imaging for ambient assisted living," in *Evaluating AAL Systems Through Competitive Benchmarking*, S. Chessa and S. Knauth, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 108–130.

[35] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, *Handbook of Markov Chain Monte Carlo (1st ed.)*. Chapman and Hall/CRC, 2011.

[36] Y. Bar-Shalom and X.-R. Li, *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, Inc., 2001.

[37] A. F. García-Fernández, L. Svensson, M. R. Morelande, and S. Särkkä, "Posterior linearization filter: Principles and implementation using sigma points," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5561–5573, October 2015.

TABLE IV
PERFORMANCE SUMMARY (RMSE $\pm$ STD)

Expt.	Model	MLE [cm]	GF [cm]	KF [cm]	PF [cm]
Open	EM	62.1 ± 54.6	12.8 ± 5.9	29.4 ± 20.8	12.3 ± 7.4
Open	BGP	24.3 ± 22.0	8.9 ± 4.1	11.2 ± 8.6	$\mathbf{7.7 \pm 5.0}$
Open	HGP	$\mathbf{12.5 \pm 8.7}$	9.3 ± 4.4	$\mathbf{6.5 \pm 3.9}$	7.9 ± 4.3
Apt.	EM	120.6 ± 86.5	55.7 ± 40.6	65.4 ± 39.8	24.9 ± 17.0
Apt.	BGP	38.4 ± 35.9	12.3 ± 8.5	17.4 ± 14.8	10.1 ± 7.0
Apt.	HGP	$\mathbf{29.4 \pm 27.3}$	$\mathbf{10.6 \pm 7.1}$	$\mathbf{16.3 \pm 14.1}$	$\mathbf{9.5 \pm 5.5}$