Estimating Space-Dependent Coefficients for 1D Transport using Gaussian Processes as State Estimator in the Frequency Domain

Ricky J. R. van Kampen[1,2], Matthijs van Berkel[1], and Hans Zwart[3,4]

***Abstract*— This letter presents a method to estimate the space-dependent transport coefficients (diffusion, convection, reaction, and source/sink) for a generic scalar transport model, e.g. heat or mass. As the problem is solved in the frequency domain, the complex valued state as a function of the spatial variable is estimated using Gaussian process regression. The resulting probability density function of the state, together with a semi-discretization of the model, and a linear parameterization of the coefficients are used to determine the maximum likelihood solution for these space-dependent coefficients. The proposed method is illustrated by simulations.**

I. INTRODUCTION

Scalar transport (e.g. heat or mass) plays an important role in many different fields. For example, the efficiency of a nuclear fusion reactor is mainly determined by how much heat and particles the core plasma loses to the reactor wall [1]. Another example is given in the field of hydrology, where the goal is to identify hot spots for contaminants and nutrients in stream beds as a result of vertical ground water fluxes [2]. Therefore, researchers require models which can be used for simulation, analysis, prediction, diagnosis and control of generic scalar transport. For most physical systems, these transport models are obtained using first principles, however the (exact) parameters to these models are unknown. Hence, data-driven estimation of the unknown physical parameters is necessary to complete the model, which is also known as an inverse problem or grey-box modeling.

The standard method to estimate the unknown parameters (diffusion, convection, reaction, and source/sink) in these typically infinite-dimensional models that describe the physical quantity, e.g. the temperature or density, is by minimizing the output error criterion, i.e. taking the (weighted) sum of the squared error between the measured data and the model output of a finite-dimensional approximation [3], [4]. As this approach does not impose constraints on the state's variation in-between sensors, the state is allowed to oscillate (spatial aliasing) while the observed (output) error is zero. Even though these oscillations are small in the state, the errors in the estimated parameters are often significant, e.g., if they are based on the second derivative of the state (diffusion). This is generally resolved by regularizing the unknown parameters [3], [4], [5]. As pointed out in [6], [7], this regularization is often artificial as there is usually no *a priori* information on how the unknown parameters change as function of the spatial variable. Therefore, another option to add regularization is to first estimate the state as a function of the spatial variable using the measurements, and then perform the parameter estimation process [3], [6], [7], [8]. This has the following advantages: (i) for an increasing number of samples the finite approximations used in these methods will converge towards the true solution, and so will their estimates; (ii) the availability of prior information on the state can be embedded in the state estimation process (such as smoothness due to the underlying model); (iii) the state estimate can be visually inspected for correctness and validated by taking additional (spatial) measurements; (iv) separating the state and parameter estimation process allows one to write the parameter estimation problem as a, often linear, regression problem [3], [6], [7]. Taking a frequency domain approach, [6] performs the state estimation via spline interpolation and exploits the linearity of the problem by deriving a closed-form solution for the global optimum using the ordinary least squares criterion. However, for uncertain measurements this ordinary least squares solution is biased as both, the output and the regressor can contain error terms. Especially for the heterogeneous case, i.e. space-dependent parameters, very little noise already results in poor estimates. Under those circumstance this problem requires an errors-in-variables approach. Therefore, the two main contributions of this letter are extending the methodology proposed in [6] by (i) determining the probability density function of the state as a function of the spatial variable using Gaussian process regression (GPR), and (ii) deriving a maximum likelihood solution such that it can deal with uncertainty in an optimal way when estimating the unknown parameters.

In contrast to other methods that use a Gaussian process (GP) to estimate the state, e.g. [9], [10], [11], we perform the GPR in the frequency domain instead of the time domain. In this way, our method does not need to determine the time derivative, which is often hard to estimate under

[1]DIFFER - Dutch Institute for Fundamental Energy Research, Energy Systems and Control group, De Zaale 20, 5612 AJ Eindhoven, The Netherlands {R.J.R.vanKampen,M.vanBerkel}@differ.nl

[2]Eindhoven University of Technology, Dept. of Mechanical Engineering, Control Systems Technology group, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

[3]University of Twente, Mathematical Systems Theory, P.O. Box 217, 7500 AE Enschede, The Netherlands h.j.zwart@utwente.nl

[4]Eindhoven University of Technology, Dept. of Mechanical Engineering, Dynamics and Control group, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Acknowledgment: DIFFER is part of the institutes organisation of NWO. This work has been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Programme (Grant Agreement No 101052200 - EUROfusion). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union not the European Commission can be held responsible for them. We want to thank Aaron Ho for his valuable comments regarding Gaussian process regression.

noisy conditions; and we can work with smaller data sets as in perturbation experiments only a limited number of frequencies are informative, i.e. excited and above the noise floor.

II. PROBLEM FORMULATION

The problem setup is specified as follows.

a) Model-Class: The estimation of the space-dependent physical coefficients is based on a one-dimensional linear parabolic partial differential equation (PDE) that describes the state, e.g. temperature or mass concentration, around an equilibrium point [12]:

$$\partial_t z = -\nabla \cdot (Vz - D\nabla z) + Kz + P\phi, \tag{1}$$

where $z : \mathbb{X} \times \mathbb{T} \to \mathbb{R}$ is a multi-variable function of a bounded space $x \in \mathbb{X} := [x_b, x_e] \subset \mathbb{R}$ and time $t \in \mathbb{T} := [t_0, t_e] \subseteq \mathbb{R}_{\geq_0}$. The state z is understood point-wise in x and t evaluated as $z(x,t)$. The physical transport coefficients are the diffusion $D : \mathbb{X} \to \mathbb{R}_{>0}$, convection $V : \mathbb{X} \to \mathbb{R}$, reaction $K : \mathbb{X} \to \mathbb{R}$, and the fixed spatially distributed source or sink $P : \mathbb{X} \to \mathbb{R}$. The source/sink is manipulated in time only by the external input $\phi : \mathbb{T} \to \mathbb{R}$. For well-posedness, the PDE is constrained at x_b and x_e by two (arbitrary) linear boundary conditions. Moreover, the initial condition $z(\cdot, t_0)$ is assumed to be compatible with the model and its boundaries. Due to linearity of the model, (1) can be considered in the frequency domain without loss of information [13]. Hence, the generic scalar transport equation in the frequency domain is

$$i\omega Z = -\nabla \cdot (VZ - D\nabla Z) + KZ + P\Phi, \tag{2}$$

with the Fourier transformed state $Z = \mathcal{F}(z) : \mathbb{X} \times \Omega \to \mathbb{C}$, input $\Phi = \mathcal{F}(\phi) : \Omega \to \mathbb{C}$, $i^2 = -1$, and angular frequency $\omega \in \Omega \subset \mathbb{R}$.

b) Measurement data: The measurement signals $\mathbf{y}(t) := \operatorname{col}(y_1(t), \ldots, y_M(t))$ are assumed to be band-limited measurements of the scalar transport state z at $M > 2$ fixed locations given by the set $\mathbb{X}_M := \{\check{x}_1, \ldots, \check{x}_M\} \subseteq \mathbb{X}$ and are disturbed by noise $\varepsilon(t)$, i.e. $y_m = z(\check{x}_m, t) + \varepsilon_m(t)$. As we consider the problem in the frequency domain, we assume the discrete Fourier transform (DFT) spectra $\mathbf{Y}(k) = \mathcal{F}(\mathbf{y})$ and $\Phi(k)$ to be processed such that it only contains the forced response. This means that transients (non-steady-state behavior, e.g. due to the initial condition) have been removed by either waiting until the transient terms are diminished, or compensating for it using (advanced) signal processing techniques such as the local polynomial method [13], [14] and hence, does not need to be considered further.

Additionally, we assume that the noise contributions are circular complex normally distributed in the frequency domain, which imposes weak assumptions on the time domain noise distribution [13]. As a result, the uncertainty of the (processed) Fourier spectrum is also circular complex normally distributed, i.e. $\mathbb{E}\{(\mathbf{Y} - \mathbb{E}\{\mathbf{Y}\})(\mathbf{Y} - \mathbb{E}\{\mathbf{Y}\})^{\mathsf{T}}\} = 0$ and $\mathbb{E}\{(\mathbf{Y} - \mathbb{E}\{Y\})(\mathbf{Y} - \mathbb{E}\{\mathbf{Y}\})^{\mathsf{H}}\} = C_{\mathbf{Y}}$, with the Hermitian transpose denoted as $^{\mathsf{H}}$. Here, $C_{\mathbf{Y}}$ is either known or can be estimated [13]. Moreover, due to linearity of the model, each (excited) angular frequency ω_k, with frequency bin $k \in \mathbb{K}$, is independent and only a finite number of the discrete angular frequencies are informative, i.e., those which are present in the (boundary) input and are above the noise level [13]. Therefore, we only consider those bins to be in $\mathbb{K}$, and excluded the DC and Nyquist frequency from the spectrum. All these signal and noise assumptions are standard (weak) assumptions for frequency domain analysis [13].

c) Problem definition: The goal is to estimate the set of space-dependent coefficients, i.e. functions, $\Gamma := \{D, V, K, P\}$ in (2), related to the physical quantities diffusion, convection, etc., using the (processed) spectra $\mathbf{Y}(k)$ and $\Phi(k)$. The (exact) boundary conditions that constrain the generic scalar transport are often unknown as they may depend on the equilibrium or unknown space-dependent parameters. Therefore, the extremum measurements are used as boundary inputs [15], [16], which reduces the domain on which the space-dependent coefficients are estimated to $\mathbb{X}_E := [\check{x}_1, \check{x}_M]$. However, other linear boundary conditions are allowed. Hence, the formal problem definition is

Problem 1. *Given the processed data-set*

$$\mathbb{D} := \{\mathbf{Y}(k), C_{\mathbf{Y}}(k), \Phi(k) \mid k \in \mathbb{K}, \check{x}_m \in \mathbb{X}_M\},$$

estimate the unknown functions $\Gamma = \{D, V, K, P\}$ *by minimizing a cost function* $\mathcal{V}(\mathbb{D}, Z(x, k; \Gamma))$ *over* Γ *such that the solution* $Z(x, k; \Gamma)$ *satisfies the model (2) subject to the boundary conditions*

$$Z(\check{x}_1, k) = Y_1(k), \quad Z(\check{x}_M, k) = Y_M(k), \quad k \in \mathbb{K}.$$

III. STATE ESTIMATION BY GAUSSIAN PROCESS REGRESSION

As [6], [8] shows, to have a unique solution for the coefficients, a continuous state Z is required. However, in general, the state is only measured at a limited number of spatial locations with uncertainty. To find the state Z as a continuous function of the spatial variable, we resort to GPR, because it takes the uncertainty into account and provides a probability density function of the state estimate which can then be used to determine the maximum likelihood solution of the unknown coefficients.

A. Gaussian process regression

A GP is defined as a collection of random variables, any finite number of which have a joint Gaussian distribution [17]. Therefore, a GP $f(x)$ is completely defined by its mean function $\mu(x)$ and covariance function $\kappa(x, \hat{x})$, and is denoted as

$$f(x) \sim \mathcal{GP}(\mu(x), \kappa(x, \hat{x})). \tag{3}$$

If the process is observed under some Gaussian distributed noise ε with a known covariance C_ε, i.e. $\mathbf{y} = \operatorname{col}(f(\check{x}_1), \ldots, f(\check{x}_M)) + \varepsilon$, the joint prior distribution of the measured values $\mathbf{y}$ at $\check{\mathbf{x}}$ and the prior predictive distribution $\hat{\mathbf{f}}$ at $\hat{\mathbf{x}}$ is given by

$$\begin{bmatrix} \mathbf{y} \\ \hat{\mathbf{f}} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}(\check{\mathbf{x}}) \\ \boldsymbol{\mu}(\hat{\mathbf{x}}) \end{bmatrix}, \begin{bmatrix} K(\check{\mathbf{x}}, \check{\mathbf{x}}) + C_\varepsilon & K(\check{\mathbf{x}}, \hat{\mathbf{x}}) \\ K(\hat{\mathbf{x}}, \check{\mathbf{x}}) & K(\hat{\mathbf{x}}, \hat{\mathbf{x}}) \end{bmatrix} \right), \tag{4}$$

with the mean vector $\boldsymbol{\mu}$ and covariance matrix $K(\mathbf{x},\mathbf{x})$ for which the v, w-th element has value $\kappa(x_v, x_w)$. The posterior predictive distribution after conditioning with the prior joint distribution is given by,

$$\hat{\mathbf{f}} \mid \check{\mathbf{x}}, \mathbf{y}, \hat{\mathbf{x}} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\hat{\mathbf{f}}}, C_{\hat{\mathbf{f}}}\right), \tag{5}$$

$$\begin{aligned}\boldsymbol{\mu}_{\hat{\mathbf{f}}} &= \boldsymbol{\mu}(\hat{\mathbf{x}}) \\ &\quad + K(\hat{\mathbf{x}},\check{\mathbf{x}})[K(\check{\mathbf{x}},\check{\mathbf{x}}) + C_\varepsilon]^{-1}\left(\mathbf{y} - \boldsymbol{\mu}(\check{\mathbf{x}})\right),\end{aligned} \tag{6}$$

$$\begin{aligned}C_{\hat{\mathbf{f}}} &= K(\hat{\mathbf{x}},\hat{\mathbf{x}}) \\ &\quad - K(\hat{\mathbf{x}},\check{\mathbf{x}})[K(\check{\mathbf{x}},\check{\mathbf{x}}) + C_\varepsilon]^{-1} K(\check{\mathbf{x}},\hat{\mathbf{x}}),\end{aligned} \tag{7}$$

which are the key equations for GPR [17]. Typically, the covariance function, also called kernel, will have some free parameters α, called hyperparameters, and are tuned by maximizing the marginal likelihood, i.e. the likelihood that the given prior has generated the observed data. This is equal to minimizing the following cost function

$$\begin{aligned}\mathcal{V}_{\mathcal{GP}}(\alpha) = &-\left(\mathbf{y} - \boldsymbol{\mu}(\check{\mathbf{x}})\right)^{\mathsf{T}}\left[K(\check{\mathbf{x}},\check{\mathbf{x}},\alpha) + C_\varepsilon\right]^{-1}\left(\mathbf{y} - \boldsymbol{\mu}(\check{\mathbf{x}})\right) \\ &- \log\left(\det\left(K(\check{\mathbf{x}},\check{\mathbf{x}},\alpha) + C_\varepsilon\right)\right).\end{aligned} \tag{8}$$

B. Complex valued Gaussian process regression

As the measurements are complex valued, the GPR should be adopted accordingly. For the sake of simplicity, we consider the state Z at each frequency separately and assume it to be a zero mean circular complex GP which is a standard assumption [18], [19]. However, we pose the additional constraint that the covariance between the real and imaginary part is zero. This allows us to use standard kernels for real valued GPs while only limiting the information that is shared between the real and imaginary part. The complex valued GP is then given by the two joint real valued GPs

$$\begin{bmatrix}\mathrm{Re}\left(Z(x,k)\right)\\ \mathrm{Im}\left(Z(x,k)\right)\end{bmatrix} \sim \mathcal{GP}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\kappa(x,\hat{x},\alpha) & 0\\ 0 & \kappa(x,\hat{x},\alpha)\end{bmatrix}\right), \tag{9}$$

denoted as $Z_{\mathrm{Re}} \sim \mathcal{GP}\left(0, \kappa_{\mathrm{Re}}(x,\hat{x})\right)$. The assumption of a zero mean is not a drastic restriction as it does not confine the mean of the posterior process to be zero [17]. However, in the future, a mean could be included to increase interpretability of the model or add prior information. The following transformation is used to go from complex valued data to a real valued vector $\mathbf{Y}_{\mathrm{Re}}$ and covariance matrix $C_{\mathbf{Y}_{\mathrm{Re}}}$

$$\mathbf{Y}_{\mathrm{Re}} = \frac{1}{2}\begin{bmatrix}I & I\\ -iI & iI\end{bmatrix}\begin{bmatrix}\mathbf{Y}\\ \overline{\mathbf{Y}}\end{bmatrix}, \tag{10}$$

$$C_{\mathbf{Y}_{\mathrm{Re}}} = \frac{1}{4}\begin{bmatrix}I & I\\ -iI & iI\end{bmatrix}\begin{bmatrix}C_{\mathbf{Y}} & 0\\ 0 & \overline{C_{\mathbf{Y}}}\end{bmatrix}\begin{bmatrix}I & I\\ -iI & iI\end{bmatrix}^{\mathsf{H}} \tag{11}$$

where $\overline{\mathbf{Y}}$ and $\overline{C_{\mathbf{Y}}}$ denote the complex conjugates. Now, the joint distribution of the real valued measurements and the real valued state prediction $\hat{\mathbf{Z}}_{\mathrm{Re}}$ follow from (4) while (5)-(7) are used to determine the (posterior) prediction $\hat{\mathbf{Z}}_{\mathrm{Re}}$ with mean $\boldsymbol{\mu}_{\hat{\mathbf{Z}}_{\mathrm{Re}}}$ and covariance matrix $C_{\hat{\mathbf{Z}}_{\mathrm{Re}}}$. To complete the state estimation process and go back to complex valued data, the following inverse transformation is used

$$\boldsymbol{\mu}_{\hat{\mathbf{Z}}} = \begin{bmatrix}I & iI\end{bmatrix}\boldsymbol{\mu}_{\hat{\mathbf{Z}}_{\mathrm{Re}}} \tag{12}$$

$$C_{\hat{\mathbf{Z}}} = \begin{bmatrix}I & iI\end{bmatrix} C_{\hat{\mathbf{Z}}_{\mathrm{Re}}} \begin{bmatrix}I & iI\end{bmatrix}^{\mathsf{H}}. \tag{13}$$

The kernel defines nearness or similarity between data points and can be used to embed prior information [17]. As we can use standard kernels for real valued GPs, we choose the Matérn covariance function with $\nu = \frac{5}{2}$ as it enforces the GP, i.e. the state Z, to be twice (mean square) differentiable with respect to x. This corresponds to the solution of the underlying model (2) that is at least twice differentiable. In this way, we incorporate valuable prior knowledge of our system into the state estimation procedure via GPR. For completeness, the kernel is given by

$$\kappa(x,\hat{x},\alpha) = \sigma^2\left(1 + \frac{\sqrt{5}|x-\hat{x}|}{\ell} + \frac{5(x-\hat{x})^2}{3\ell^2}\right)e^{-\frac{\sqrt{5}|x-\hat{x}|}{\ell}}, \tag{14}$$

with hyperparameters $\alpha = \mathrm{col}\left(\sigma, \ell\right)$ [17].

IV. FINITE-DIMENSIONAL PROBLEM FORMULATION

To create a finite-dimensional problem, we take the same approach as in [6], [16], starting with the linear parametrization of the unknown coefficients, followed by a finite difference scheme to approximate the spatial derivatives, resulting in a linear matrix equality. Therefore, the model (2) is reformulated such that it is linear in the state

$$i\omega_k Z = DZ'' + (D' - V)Z' + (K - V')Z + P\Phi, \tag{15}$$

where the prime ($'$) denotes the spatial derivative(s).

A. Parameterization of the unknown functions

To estimate the function $\gamma \in \Gamma = \{D, V, K, P\}$, we assume that each function can be described by a finite sum of basis functions B_r^γ weighted by $\theta^\gamma = \mathrm{col}\left(\theta_1^\gamma, \ldots, \theta_{R^\gamma}^\gamma\right)$,

$$\gamma(x,\theta^\gamma) = \sum_{r=1}^{R^\gamma} B_r^\gamma(x)\theta_r^\gamma. \tag{16}$$

Now, estimating the unknown coefficients is reduced to estimating $\theta = \mathrm{col}\left(\theta^D, \theta^V, \theta^K, \theta^P\right) \in \mathbb{R}^R$, with $R = R^D + R^V + R^K + R^P$. The basis functions B_r^γ should satisfy the model, which means that the derivative of B_r^D and B_r^V must exist. As the user chooses these basis functions, we consider these derivatives to be known.

B. Discretization procedure

The regression model requires spatial derivatives of the state. These can be obtained from the GPs [20]. Alternatively, by approximating them using a numerical scheme, we can bypass the GPR by directly using the measurement data ($\hat{\mathbf{Z}} = \mathbf{Y}$). This allows to study the effect of the state estimation procedure on the estimated parameters and can help verifying the outcome. The spatial derivatives are approximated by a central finite difference scheme for non-equidistant grids [21], as it is known that the finite-dimensional model

will converge to the true infinite-dimensional model for an increasing number of points. For this we consider the discrete state vector $\mathbf{Z}(k) := \mathrm{col}\,(Z(x_1,k),\dots,Z(x_N,k))$ that contains N spatial sample points $x_j \in \mathbb{X}_d \subseteq \mathbb{X}_E$, $j \in \{1,\dots,N\}$, $x_1 = \check{x}_1$ and $x_N = \check{x}_M$. The derivatives for Z'', Z' and Z can be written using the matrices L_2, L_1, L_0, e.g. $\mathbf{Z}' = L_1\mathbf{Z}$. By excluding the derivatives on the boundaries, i.e. at x_1 and x_N, L_2, L_1 and L_0 are of size $N-2\times N$. Moreover, in this way, the boundaries are directly included as Dirichlet boundary conditions. Hence, (15) can be expressed as

$$i\omega_k L_0\mathbf{Z} = \hat{A}(\theta)\mathbf{Z} + \hat{B}(\theta)\Phi, \tag{17}$$

with

$$\begin{aligned}\hat{A}(\theta) = &\sum_{r=1}^{R^D}(\tilde{B}_r^D L_2 + \tilde{B}_r^{D\prime} L_1)\theta_r^D \\ &- \sum_{r=1}^{R^V}(\tilde{B}_r^V L_1 + \tilde{B}_r^{V\prime} L_0)\theta_r^V + \sum_{r=1}^{R^K}\tilde{B}_r^K L_0\theta_r^K\end{aligned} \tag{18}$$

$$\hat{B}(\theta) = \sum_{r=1}^{R^P}(\tilde{B}_r^P\mathbf{1})\theta_r^P, \tag{19}$$

with $\tilde{B}_r^\gamma = \mathrm{diag}\,(B_r^\gamma(x_2),\dots,B_r^\gamma(x_{N-1})))$ as the diagonal matrix for each basis function $\gamma = D,V,K$ or P, and the column vector of ones denoted by $\mathbf{1}$. Hence, the problem is now written as a matrix equality that is bilinear in the unknown parameters θ and the discrete state vector $\mathbf{Z}$. Furthermore, note that other linear boundary conditions can be included by adapting $\hat{A}$ and $\hat{B}$ accordingly.

V. MAXIMUM LIKELIHOOD SOLUTION

For the maximum likelihood solution we assume that the state estimate or measurement $\hat{\mathbf{Z}}$ is circular complex normally distributed with a mean $\mu_{\hat{\mathbf{Z}}}$ and a covariance matrix $C_{\hat{\mathbf{Z}}}$. The cost function for the maximum likelihood is then given by [13] as

$$\mathcal{V}(\theta,\hat{\mathbf{Z}}) = \sum_{k\in\mathbb{K}}\mathbf{e}(\theta,\mu_{\hat{\mathbf{Z}}},k)^{\mathsf{H}}[C_{\mathbf{e}}(\theta,C_{\hat{\mathbf{Z}}},k)]^{-1}\mathbf{e}(\theta,\mu_{\hat{\mathbf{Z}}},k), \tag{20}$$

where the constant $F\ln(\pi)+\ln\left(\det\left(C_{\hat{\mathbf{Z}}}\right)\right)$ has been omitted as it does not affect the optima. The error vector from the model (17) yields,

$$\mathbf{e}(\theta,\mu_{\hat{\mathbf{Z}}},k) = (i\omega_k L_0 - \hat{A}(\theta))\mu_{\hat{\mathbf{Z}}}(k) - \hat{B}(\theta)\Phi(k) \tag{21}$$

with the corresponding covariance matrix

$$C_{\mathbf{e}}(\theta,C_{\hat{\mathbf{Z}}},k) = (i\omega_k L_0 - \hat{A}(\theta))C_{\hat{\mathbf{Z}}}(k)(i\omega_k L_0 - \hat{A}(\theta))^{\mathsf{H}}. \tag{22}$$

The maximum likelihood solution is found by minimizing the cost function

$$\hat{\theta} = \arg\min_\theta \mathcal{V}(\theta,\hat{\mathbf{Z}}). \tag{23}$$

A. Minimizing the cost function

As the cost function is non-convex, we choose to solve it using iterative optimization methods such as Gauss-Newton or Levenberg-Marquardts [22]. For this we resort to the pseudo-Jacobian matrix which generally gives faster convergence [23]. The parameter update $\Delta\theta$ for the iterative algorithms is found by solving the overdetermined set of equations

$$J_+(\theta,\hat{\mathbf{Z}},k)\Delta\theta = \left[C_{\mathbf{e}}(\theta,C_{\hat{\mathbf{Z}}},k)\right]^{-\frac{1}{2}}\mathbf{e}(\theta,\mu_{\hat{\mathbf{Z}}},k), \tag{24}$$

for all $k \in \mathbb{K}$. The pseudo-Jacobian J_+ is given by

$$J_+(\theta,\hat{\mathbf{Z}},k) = \begin{bmatrix}J_+^1(\theta,\hat{\mathbf{Z}},k) & \dots & J_+^R(\theta,\hat{\mathbf{Z}},k)\end{bmatrix}, \tag{25}$$

$$\begin{aligned}J_+^r(\theta,\hat{\mathbf{Z}},k) = &\left[C_{\mathbf{e}}(\theta,C_{\hat{\mathbf{Z}}},k)\right]^{-\frac{1}{2}}\big(\partial_{\theta_r}\mathbf{e}(\theta,\mu_{\hat{\mathbf{Z}}},k) \\ &- \frac{1}{2}\partial_{\theta_r}C_{\mathbf{e}}(\theta,C_{\hat{\mathbf{Z}}},k)\left[C_{\mathbf{e}}(\theta,C_{\hat{\mathbf{Z}}},k)\right]^{-1}\mathbf{e}(\theta,\mu_{\hat{\mathbf{Z}}},k)\big),\end{aligned} \tag{26}$$

where ∂_{θ_r} denotes the partial derivative to the weight θ_r and $C^{\frac{1}{2}}$ denotes the square root of the matrix [13], [23]. Note that these derivatives are easily obtained as the problem is formulated such that it is linear in the weights θ.

B. Calculation of the confidence intervals

The covariance matrix of the estimated weights $\hat{\theta}$ is given by [13] as

$$C_{\hat{\theta}} = \left[2\mathrm{Re}\left(\sum_{k\in\mathbb{K}}J_+(\hat{\theta},\hat{\mathbf{Z}},k)^{\mathsf{H}}J_+(\hat{\theta},\hat{\mathbf{Z}},k)\right)\right]^{-1}. \tag{27}$$

As the transformation from $\hat{\theta}_\gamma$ to γ is linear, the $\mathfrak{p}$ confidence interval for the estimated functions $\gamma \in \Gamma$ is given by

$$C_\gamma(x,\hat{\theta}^\gamma,\mathfrak{p}) = \gamma(x,\hat{\theta}^\gamma) \pm \sqrt{2\sigma_\gamma^2(x)}\mathrm{erf}^{-1}(\mathfrak{p}), \tag{28}$$

where the variance of the function γ is determined using propagation of uncertainty

$$\sigma_\gamma^2 = \begin{bmatrix}B_1^\gamma(x) & \dots & B_{R^\gamma}^\gamma(x)\end{bmatrix}C_{\hat{\theta}^\gamma}\begin{bmatrix}B_1^\gamma(x) & \dots & B_{R^\gamma}^\gamma(x)\end{bmatrix}^{\mathsf{T}}, \tag{29}$$

where $C_{\hat{\theta}^\gamma}$ is a submatrix of the total covariance matrix $C_{\hat{\theta}}$.

VI. SIMULATION RESULTS

The merit of the proposed methodology is demonstrated by generating a noisy data set and estimating the transport coefficients using our early work [6] and the newly derived maximum likelihood estimator with and without the novel state estimation via GPR, including a noiseless data set to focus on the differences GPR brings.

A. Data generation and state estimation

The simulation example is inspired by perturbative experiments in the field of nuclear fusion [24], [25]. The heat transport, (1), is generally analyzed on the normalized domain $\mathbb{X} = [x_b, x_e] = [0,1]$ of the minor plasma radius. Here x_b is at the center and x_e is at the edge of the plasma. The corresponding boundary conditions in the simulation are $Z'(x_b,\cdot) = 0$ due to (axi)symmetry and Dirichlet boundary

condition $Z(x_e, \cdot) = 0$ due to a significant temperature difference between core plasma ∼170 million °C and edge plasma ∼1 million °C. Typical parameter functions used in nuclear fusion are $D^{\text{sim}}(x) = 5x^3 - 0.005x + 5$, $V^{\text{sim}}(x) = -15x^2 + 0.005$, $K^{\text{sim}}(x) = 0$ and

$$P^{\text{sim}}(x) = \left(2 + \frac{7}{\sqrt{\pi}} e^{\frac{-(x-0.35)^2}{(0.1)^2}} + \frac{5.6}{\sqrt{\pi}} e^{\frac{-(x-0.6)^2}{(0.1)^2}}\right) \cdot 10^4,$$

which are equal to the functions in [6], except we set $K^{\text{sim}}(x) = 0$ as the reaction coefficient is intrinsically hard to estimate under noisy conditions compared to the other coefficients [26]. Furthermore, we increased the power deposition with four orders of magnitude to be more realistic [25]. For the perturbation of the plasma temperature, a microwave source is used, where the excitation signal $\Phi(\omega)$ is a block-wave of $\omega_0 = 50\pi$ with a 70% duty cycle. Here, only the first five harmonics $\omega_k = k\omega_0$, $k = 1, \ldots, 5$ are informative and used for the estimation. The temperature data is generated by simulation with a central finite difference grid of 1001 sample points where the state is measured at $M = 16$ spatial sensors locations, positioned at $\check{x}_m = \Delta x \cdot m$, with $\Delta x = 0.058$ and $m = 1, \ldots, M$. Noise from a circular normal distribution with covariance matrix $C_{\mathbf{Y}} = \sigma_\varepsilon(k) I$ is added to the measurements, where ε is chosen such that the maximum signal-to-noise ratio, $\text{SNR} = \frac{|Y_m(k)|}{\sigma_\varepsilon(k)}$, is 40 dB and lowers with 2.5 dB for each harmonic (due to transport). Hence, overall the SNRs range from 5 to 40 dB at the different data points. For the first two harmonics, the simulated and measured state with its $\mathfrak{p} = 0.95$ confidence interval are shown in Fig. 1. Based on the measurements and their uncertainty, the hyperparameters of the GP are tuned by optimizing (8) for each frequency separately. Then, the state estimate and covariance matrix is determined at $N = 76$ points given by $x_n = 0.058 + 0.0116(n-1)$, with $n = 1, \ldots, N$. The estimated state and its $\mathfrak{p} = 0.95$ confidence interval is also shown in Fig. 1.

B. Parameter estimation

For the estimation procedure, we consider that there is some prior knowledge on the shape of the coefficients, i.e. that D and V are polynomials and P Gaussian. Therefore, the basis functions to estimate D and V are monomials $B_r^\gamma(x) = x^{(r-1)}$ with $R^D = 7$ and $R^V = 6$, respectively. The orders are significantly higher than the actual order such that it is possible to find an exact (noiseless) description of D and V. The basis functions for the source are B-splines, as it is linear in the unknown parameters and can describe smooth complex shapes. The B-splines are designed using the De Boor's algorithm [27] with $R^P = 14$ control points and degree 3. Thus, in total, there are $R = 27$ free parameters. For a discussion on the effect of different basis functions, see [6].

The iterative optimization algorithm needs an initial starting point, which is set to $\theta = \mathbf{1}$. The result of the newly developed method with and without GPR is shown in Fig. 2 along with the simulated coefficients. To show the improvement, the estimates using the methodology in [6] is also shown in Fig. 2. Although the methodology in [6] comes with a closed-form solution of the global optimum, the estimates are poor. This is a result of the used ordinary least squares criterion that results in an biased estimator for the given problem formulation where the measurement uncertainty enters the regression matrix. The newly derived maximum likelihood estimators takes this into account resulting in better estimates. Moreover, in combination with the GPR as state estimator, spatial resolution and the quality of the estimated coefficients is improved. Although we used the GPs to apply regularization, there are still small oscillations in the estimated coefficients. This is a result of the small mismatch between the simulated and estimated state and the many degrees of freedom of the coefficients. For comparison, estimates based on noiseless observations are shown along the other estimates in Fig. 2. The coefficients are estimated accurately except at the boundaries of the domain. Especially when using GPR. This is the result of small errors made in the state estimation using the Gaussian process as the amplitude of the signal quickly decreases due to the Dirichlet boundary condition and there is less information available at the boundaries for the GP.

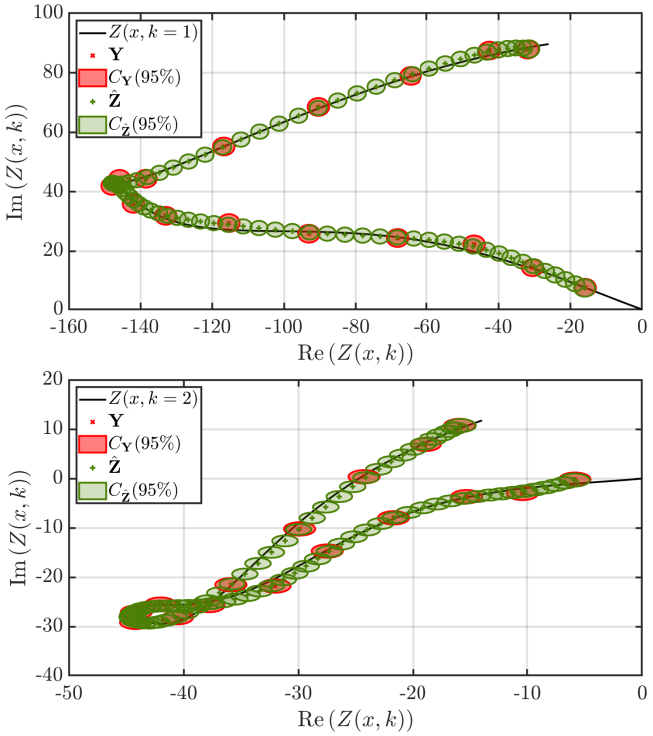

Fig. 1. The simulated Z, measured $\mathbf{Y}$ and estimated state $\hat{\mathbf{Z}}$ shown in the complex plane for the first harmonic $(k = 1)$ at $\omega = 50$ (top) and the second harmonic $(k = 2)$ at $\omega = 100$ (bottom), including their $\mathfrak{p} = 0.95$ confidence interval.

VII. CONCLUSION AND DISCUSSION

This letter presents a novel method to estimate the unknown space-dependent transport coefficients for 1D generic scalar transport from noisy measurements by first determining the probability density function of the state and then

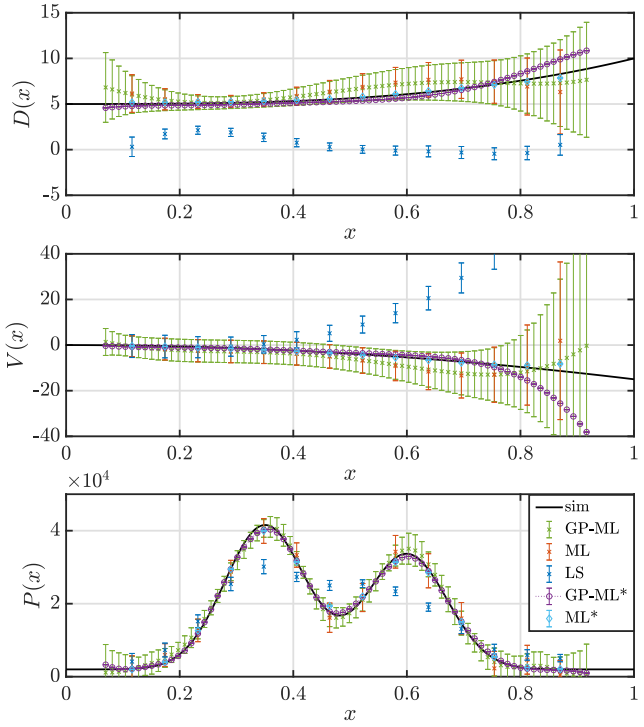

Fig. 2. The estimated transport parameters: diffusion (top), convection (middle), source location (bottom) and their $\mathfrak{p} = 0.95$ confidence interval, estimated by the earlier presented ordinary least squares method (LS), the novel maximum likelihood method (ML) and the maximum likelihood method where the state and its uncertainty is estimated using Gaussian process regression (GP-ML) where the * indicates the noiseless scenario.

using this information to estimate the unknown parameters with a maximum likelihood estimator. By separating the state and parameter estimation problem, we avoid the artificial regularization of the unknown transport coefficients, but apply regularization on the state via GPR. This approach has shown to be successful, although we must note that it has a reduced performance at the boundaries of the domain. However, we expect that this can be improved by including more (prior) knowledge, e.g. by forcing the solution of the GP to satisfy the model [9], [11], considering multiple frequencies simultaneously and embedding the stability of the transfer function into the kernel [18], [19], and simultaneously estimating the state and parameters by considering the coefficients as hyperparameters of the GP [9]. Nonetheless, in the current state, the proposed methodology shows to be a significant improvement over the earlier linear least squares method that results in biased estimates when considering noise. Moreover, the novel derived maximum likelihood estimator can also work with only measurement data in case the GPR is distrusted and help verifying the correctness or influence of the GPR results as we showed for the noiseless scenario.

References

[1] G. M. D. Hogeweij, N. J. L. Cardozo, M. R. de Baar, and A. M. R. Schilham, "A model for electron transport barriers in tokamaks, tested against experimental data from RTP," *Nucl. Fusion*, vol. 38, no. 12, pp. 1881–1891, dec 1998.

[2] F. Boano *et al.*, "Hyporheic flow and transport processes: Mechanisms, models, and biogeochemical implications," *Rev. Geophys.*, vol. 52, no. 4, pp. 603–679, oct 2014.

[3] H. T. Banks and K. Kunisch, *Estimation Techniques for Distributed Parameter Systems*. Boston, MA, USA: Birkhäuser, 1989.

[4] C. Kravaris and J. H. Seinfeld, "Identification of parameters in distributed parameter systems by regularization," *SIAM J. Control Optim.*, vol. 23, no. 2, pp. 217–241, mar 1985.

[5] K. Ito and K. Kunisch, *Lagrange Multiplier Approach to Variational Problems and Applications*. Philadelphia, PA, USA: SIAM, Jan. 2008.

[6] R. J. R. van Kampen, A. Das, S. Weiland, and M. van Berkel, "A closed-form solution to estimate spatially varying parameters in heat and mass transport," *IEEE Control Syst. Lett.*, vol. 5, no. 5, pp. 1681–1686, nov 2021.

[7] D. F. Escande and F. Sattin, "Calculation of transport coefficient profiles in modulation experiments as an inverse problem," *Phys. Rev. Lett.*, vol. 108, no. 12, p. 125007, mar 2012.

[8] H. Banks and P. Lamm, "Estimation of variable coefficients in parabolic distributed systems," *IEEE Trans. Automat. Contr.*, vol. 30, no. 4, pp. 386–398, apr 1985.

[9] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Machine learning of linear differential equations using Gaussian processes," *Journal of Computational Physics*, vol. 348, pp. 683–693, 2017.

[10] P. K. Rai and S. Tripathi, "Gaussian process for estimating parameters of partial differential equations and its application to the Richards equation," *Stochastic Environmental Research and Risk Assessment*, vol. 33, pp. 1629–1649, Jul. 2019.

[11] C. Jidling, N. Wahlström, A. Wills, and T. B. Schön, "Linearly constrained Gaussian processes," in *31st Conference on Neural Information Processing Systems*, 2017.

[12] C. E. Baukal, V. Y. Gershtein, and X. Li, *Computational Fluid Dynamics in Industrial Combustion*. CRC Press, oct 2000.

[13] R. Pintelon and J. Schoukens, *System Identification: A Frequency Domain Approach*, 2nd ed. Hoboken, NJ, USA: Wiley, mar 2012.

[14] M. van Berkel *et al.*, "Correcting for non-periodic behaviour in perturbative experiments: application to heat pulse propagation and modulated gas-puff experiments," *Plasma Phys. Contr. Fusion*, vol. 62, no. 9, p. 094001, jul 2020.

[15] U. Schneidewind *et al.*, "LPMLE3: A novel 1-D approach to study water flow in streambeds using heat as a tracer," *Water Resour. Res.*, vol. 52, no. 8, pp. 6596–6610, aug 2016.

[16] A. Das, S. Weiland, and M. van Berkel, "Frequency domain estimation of spatially varying parameters in heat and mass transport," in *2019 American Control Conference (ACC)*. IEEE, jul 2019.

[17] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2005.

[18] J. Lataire and T. Chen, "Transfer function and transient estimation by Gaussian process regression in the frequency domain," *Automatica*, vol. 72, pp. 217–229, oct 2016.

[19] N. Hallemans, R. Pintelon, D. Peumans, and J. Lataire, "Improved frequency response function estimation by Gaussian process regression with prior knowledge," *IFAC*, vol. 54, no. 7, pp. 559–564, 2021.

[20] A. Ho *et al.*, "Application of gaussian process regression to plasma turbulent transport model validation via integrated modelling," *Nucl. Fusion*, vol. 59, no. 5, p. 056007, mar 2019.

[21] H. Sundqvist and G. Veronis, "A simple finite-difference grid with non-constant intervals," *Tellus*, vol. 22, no. 1, pp. 26–31, jan 1970.

[22] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York, NY, USA: Springer, 1999.

[23] P. Guillaume and R. Pintelon, "A Gauss-Newton-like optimization algorithm for "weighted" nonlinear least-squares problems," *IEEE Trans. Signal Process.*, vol. 44, no. 9, pp. 2222–2228, 1996.

[24] M. van Berkel, G. Vandersteen, E. Geerardyn, R. Pintelon, H. Zwart, and M. de Baar, "Frequency domain sample maximum likelihood estimation for spatially dependent parameter estimation in PDEs," *Automatica*, vol. 50, no. 8, pp. 2113–2119, aug 2014.

[25] J. H. Slief, R. J. R. van Kampen, M. W. Brookman, and M. van Berkel, "Extension of the flux fit method for estimating power deposition profiles," *Phys. Plasmas*, vol. 29, no. 1, p. 010703, jan 2022.

[26] M. van Berkel, A. de Cock, T. Ravensbergen, G. M. D. Hogeweij, H. J. Zwart, and G. Vandersteen, "A systematic approach to optimize excitations for perturbative transport experiments," *Phys. Plasmas*, vol. 25, no. 8, p. 082510, aug 2018.

[27] C. de Boor, *A Practical Guide to Splines*. New York, NY, USA: Springer, 1978.