Received ; | Revised ; | Accepted

DOI: xxx/xxxx

ARTICLE TYPE

Bayesian MAP-estimation of κ turbulence model parameters using Algorithmic Differentiation in SOLPS-ITER

Stefano Carli* | Wouter Dekeyser | Maarten Blommaert | Reinart Coosemans | Wim Van Uytven | Martine Baelmans

Department of Mechanical Engineering, KU Leuven, Leuven, Belgium

Correspondence
*Stefano Carli, KU Leuven, Department of Mechanical Engineering, Celestijnenlaan 300 box 2421, 3001 Heverlee, Belgium
Email: stefano.carli@kuleuven.be

Radial anomalous diffusion coefficients are among the largest sources of uncertainty in mean-field plasma-edge codes such as SOLPS-ITER. These coefficients are machine, scenario, and space dependent, thus hampering the predictive and interpretive capability of these codes. In fact, modelers usually adjust these coefficients manually, based on expert judgment or on large, computationally expensive, parameter scans. Matching data from various diagnostics, each with their own experimental uncertainties, additionally complicates the problem of finding a good parameter set. In addition, standard nonlinear regression techniques have shown to become prohibitive for expensive plasma edge simulations with many unknown parameters, as gradient calculation was based on finite differences. In this paper, we apply Algorithmic Differentiation (AD) as an efficient and more accurate alternative for gradient calculation in large, continuously developed codes as SOLPS-ITER. Additionally, the lack of uncertainty information and danger of data overfitting, key limitations of regression techniques, are overcome by combining data and uncertainties from different diagnostics in a Bayesian framework. We implement for the first time such a Bayesian inference framework into SOLPS-ITER, using gradient-based optimization with gradients obtained through tangent AD to find the Maximum a Posteriori (MAP) values of the parameters. The recently developed κ turbulence model is employed to limit the number of unknown parameters compared to full spatial profiles of diffusion coefficients. The Bayesian MAP estimation is compared to standard regression techniques on a small-scale tokamak. We adopt fictitious experimental data obtained from a reference SOLPS-ITER solution with artificial measurement noise.

KEYWORDS:
parameter estimation, Bayesian inference, Algorithmic Differentiation, plasma-edge, SOLPS-ITER

1 | INTRODUCTION

Modeling of plasma-edge transport through mean-field codes such as SOLPS-ITER[1,2] is essential for interpretation of current day experimental devices. To constrain the simulations, certain model inputs can be directly measured from experiment, for example plasma density and temperature at the outer mid-plane (OMP), input power or gas puff strength. Other required inputs instead are a consequence of modeling approximations or lack of knowledge on the real physical process, and are not related

to a measurable quantity, thus introducing large uncertainties in modeling results. Examples of such inputs are the ad-hoc diffusion coefficients for the perpendicular anomalous transport, or boundary conditions at the artificial interface between plasma and void regions. Furthermore, these inputs differ for each device, operational regime and spatial location. It is through expert knowledge of the modeler, and more often parameter scans, that these inputs are manually adjusted in order to match experimental profiles[3–5]. Because plasma-edge simulations are expensive, these procedures easily become cumbersome. In addition, it is the modeler's arbitrary decision to stop the procedure when visual inspection of the agreement between simulated and experimental profiles is satisfying.

To alleviate these problems, a nonlinear least-squares regression technique was implemented earlier in a previous version of SOLPS, to automatically adjust spatially constant anomalous diffusion coefficients and boundary conditions to match experimental data[6,7], thereby introducing an objective metric to compare simulation results and experimental data. This is already a significant step forward compared to manual tuning of the inputs together with visual inspection of results, but the procedure incurs three main drawbacks: first, the high computational cost of finite differences (FD) used for the gradient calculation in Ref. 6; second, the lack of a consistent integration of uncertainties in parameters, model, and measurements in basic nonlinear least-squares regression techniques as the one implemented in Ref. 6; and third, the risk of overfitting inherent in such techniques.

Regarding the first drawback, FD computation of the full gradient comes at a cost proportional to the number of fitted parameters, as each has to be perturbed separately, thus limiting the application to small parameter sets. In fact, in Ref. 7 a three-fold increase in computational time was found when two additional parameters were introduced to improve fit results. An efficient and more accurate alternative for derivative calculation is the adjoint method, already applied in divertor design and parameter estimation based on edge codes[8–10], allowing gradient computation at a cost independent of the number of unknown parameters. For complex, well established and continuously developed codes such as SOLPS-ITER, an analytical derivation and manual implementation of the adjoint equations is unfeasible and prone to errors. Therefore, we resort to Algorithmic Differentiation (AD)[11], allowing efficient and accurate semi-automatic gradient computation either in tangent or adjoint mode. Applications of AD spans different domains, from aerodynamic shape optimization[12], over ice sheet modeling[13], to stellarator coils optimization[14].

The second drawback of regression techniques, the absence of any uncertainty information, can effectively be overcome with Bayesian inference methods[15], which in the past decades gained more and more popularity in several research fields, from fluid dynamics[16,17] to nuclear fusion[18–20]. In fact, Bayesian inference inherently accounts for experimental, modeling, and numerical uncertainties, providing a detailed description in terms of probability distribution of estimated parameters and their correlations, as well as uncertainty on simulation results, indispensable for model validation.

Moreover, the risk of overfitting, a relatively common problem for regression techniques, can also be avoided with Bayesian analysis. In fact, it is generally found that introducing more parameters allows reducing the fitting error, even if such parameters are physically irrelevant. To this end, Bayesian evidence provides a formal measure to compare competing models and assess their capability in describing experimental data, automatically penalizing for model complexity.

In this paper, we apply for the first time Bayesian inference methods in plasma-edge mean field codes, for the generally common case of estimating anomalous transport coefficients. We adopt the tangent mode of AD for an accurate gradient computation. To keep the parameter set limited, we do not estimate such coefficients directly, but we employ the recently developed κ-model for interchange turbulence[21], which allows retrieving spatially-dependent transport coefficients using fewer modeling parameters[22].

In Section 2 we describe the parameter estimation framework developed, while in Section 3 we give a brief overview of the κ-model equations. Afterwards, Section 4 provides information on the test case used for demonstrating the inference method, whose results are then discussed in Section 5. Finally, Section 6 closes the paper with conclusions and perspective for future work.

2 | PARAMETER ESTIMATION FRAMEWORK

Parameter estimation, also known as *model calibration*, can be cast into an optimization problem, where a scalar performance indicator $\mathcal{J}(\theta, q)$, referred to as cost function, is minimized. This cost function depends on the uncertain parameters θ, and on the state variables q. The uncertain parameters are a subset of the model input parameters, such as anomalous radial transport coefficients or boundary conditions. The state variables are instead the set of plasma and neutral quantities, such as density, temperature, and velocity, solved from the set of plasma transport equations and their boundary conditions, indicated as *state*

equations $B(\theta, q) = 0$. Formally, the cost function is subject to the condition that state equations are satisfied, therefore the problem can be formulated as a PDE-constrained optimization problem:

$$\begin{aligned} &\min_{\theta,q} \mathcal{J}(\theta, q) \\ &\text{s.t. } B(\theta, q) = 0. \end{aligned} \tag{1}$$

Since the state equations uniquely determine the state $q(\theta)$ for a fixed set of uncertain parameters through the solution of the state equations, $B(\theta, q(\theta)) \equiv 0$, the problem can be expressed in terms of a *reduced* cost functional $\hat{\mathcal{J}}(\theta) \equiv \mathcal{J}(\theta, q(\theta))$. The last expression implicitly contains the dependence on state variables, and the optimal solution is now the value of θ minimizing $\hat{\mathcal{J}}$. Moreover, further constraints can be included in the optimization, involving conditions on state or uncertain parameters. For the purpose of this work, only parameter constraints are relevant, for example imposing positiveness of diffusion coefficients. The final problem is thus a bound-constraint optimization:

$$\begin{aligned} &\min_{\theta} \hat{\mathcal{J}}(\theta) \\ &\text{s.t. } \theta_L \leq \theta \leq \theta_U, \end{aligned} \tag{2}$$

where θ_L and θ_U are some lower and upper bounds respectively.

Gradient-based optimization algorithms with line search can solve the problem in Eq. 2, converging to a local optimum[23]. The remaining missing information is therefore a proper definition of the cost function and a way of calculating its gradient with respect to the uncertain parameters.

2.1 | Regression cost function

The simplest way of defining the cost function for parameter estimation is using an L2-norm of the differences between code results and experimental data[6,10], thus effectively performing a (weighted) least squares minimization. The reduced cost function can therefore be written in a general form as:

$$\hat{\mathcal{J}}(\theta) = \sum_l \sum_q \frac{1}{\Omega_l} \int_{\Omega_l} \left(\frac{w_{l,q}}{\bar{q}_l^{\,2}} \frac{\left(q(\theta) - q^{\mathrm{EXP}}\right)^2}{2} \right) \mathrm{d}\Omega. \tag{3}$$

In Eq. 3 summation over l indicates different spacial locations, for example OMP or target, while summation over q indicates different state variables, for example plasma density n_e or electron temperature T_e. The superscript *EXP* indicates the experimental measurement. Ω_l is the domain over which the integration is performed and $w_{l,q}$ are weighting factors. Finally, $\bar{q}_l = 1/\Omega_l \int_{\Omega_l} q^{\mathrm{EXP}} \mathrm{d}\Omega$ is the average experimental value of q over domain Ω_l used to normalize the terms in the cost function.

2.2 | Bayesian MAP cost function

Bayesian analysis is a complex topic and for brevity, we introduce here only the necessary concepts needed for this work, with a deeper analysis carried out for example in Ref. 15.

The key result of Bayesian inference is the posterior probability distribution $\pi(\theta|\mathcal{D}, \mathcal{M})$ over the parameter space, defined from Bayes' theorem as

$$\pi(\theta|\mathcal{D}, \mathcal{M}) = \frac{\mathcal{L}(\mathcal{D}|\theta, \mathcal{M})\pi(\theta|\mathcal{M})}{\pi(\mathcal{D}|\mathcal{M})}, \tag{4}$$

where θ is the set of uncertain parameters as described above and $\mathcal{D}$ is the experimental data set available for the analysis, of which q^{EXP} is a specific measurement. Furthermore, $\mathcal{M}$ is the model considered for the analysis, which for this work is the SOLPS-ITER physics model described in Section 4 that includes the interchange turbulence model described in Section 3. In this expression $\mathcal{L}(\mathcal{D}|\theta, \mathcal{M})$ is called the likelihood function, $\pi(\theta|\mathcal{M})$ the prior distribution and $\pi(\mathcal{D}|\mathcal{M})$ the evidence. We describe now each of them separately, omitting the symbol $\mathcal{M}$ for simplicity.

The likelihood function $\mathcal{L}(\mathcal{D}|\theta)$ is defined as the probability of observing the data given the model and parameter set. In other words, it is a measure of how good the model $\mathcal{M}$ with parameters θ explains the data set $\mathcal{D}$. The great advantage of Bayesian analysis is that the likelihood explicitly accounts for the prediction error ϵ due to measurement, modeling and numerical uncertainties. In this work, we assume an additive error of the form

$$q^{\mathrm{EXP}} = q(\theta) + \epsilon. \tag{5}$$

While any error distribution is allowed in the Bayesian framework, the standard assumption for ϵ is that of a Gaussian distribution with mean μ and standard deviation σ_i, where i indicates a single element of the error vector ϵ. When considering N_D data points, the likelihood takes the following notation:

$$\mathcal{L}(\mathcal{D}|\theta) = \frac{1}{\sqrt{(2\pi)^{N_D}\det\Sigma}} \exp\left(-\frac{1}{2} z^T \Sigma^{-1} z\right), \tag{6}$$

where $z = \epsilon - \mu$ and Σ is the covariance matrix. The likelihood form in Eq. 6 allows to account for error correlation, in which case off-diagonal elements of the covariance matrix will take non-zero values. In this work however, we limit ourselves to the case of uncorrelated error, thus Σ will be a diagonal matrix with $\Sigma_{ii} = \sigma^2$. Additionally, one can opt for a multiplicative error form instead of the additive one in Eq. 5. In this case, the likelihood models a relative, instead of an absolute error, and the diagonal elements of Σ take the form $\Sigma_{ii} = \sigma_i^2 = (\sigma q_i^{\mathrm{EXP}})^2$, with σ a constant standard deviation[17]. It should be noted that the error parameters μ and σ_i might be known in advance, for example when the only error is an experimental measurement uncertainty that is well characterized. In case they are unknown, or if model uncertainty is expected on top of the experimental one, these parameters should be included in the vector θ and estimated in the inference procedure. Since they do not belong to the set of model inputs, they are called *nuisance* parameters.

The prior distribution $\pi(\theta)$ describes the knowledge on parameters in the absence of observed data. It therefore relies on the researcher's knowledge and on physical arguments, for example positiveness of diffusion coefficients. If little or no information is available, the prior should be an uninformative uniform distribution, while if some information is available, it can take the form of a Gaussian distribution, with a standard deviation representing the confidence in such information.

Finally, the evidence $\pi(\mathcal{D})$ describes how likely a model $\mathcal{M}$ explains the data $\mathcal{D}$, regardless of the parameter θ. The evidence therefore allows comparison of different competing models, the one with higher evidence being preferred. By applying the marginalization rule on the denominator in Eq. 4, we obtain an expression for the evidence, that depends on likelihood and prior:

$$\pi(\mathcal{D}) = \int \mathcal{L}(\mathcal{D}|\theta)\pi(\theta)\mathrm{d}\theta. \tag{7}$$

It is clear that, by integrating over the entire parameter space, the evidence is simply a normalization constant in Eq. 4.

Evaluation of the full posterior distribution is a complex and computationally demanding task, especially the evidence integral. However, it can be greatly simplified if we only search for the posterior mode $\hat{\theta}$, or maximum *a posteriori* (MAP), namely the θ value maximizing the posterior distribution:

$$\hat{\theta} = \arg\max_{\theta} \pi(\theta|\mathcal{D}). \tag{8}$$

Given that the evidence $\pi(\mathcal{D})$ is a constant with respect to θ, it can be left out of the MAP computation, which is therefore obtained by minimizing a reduced cost function of the form

$$\hat{J}(\theta) = -\log\left[\mathcal{L}(\mathcal{D}|\theta)\pi(\theta)\right], \tag{9}$$

where we employ the logarithm to reduce numerical errors that may arise in the likelihood computation.

MAP estimations are effectively simple but may be inaccurate if the posterior is multimodal. For a large number of independent data points instead, the posterior mean tends towards a normal distribution, as stated by the central limit theorem. In this case, the MAP is well defined and one can additionally apply the Laplace approximation to approximate the full posterior as a Gaussian distribution around the MAP[24].

2.3 | Gradient computation

As mentioned in the introduction, AD is an efficient and more accurate alternative to FD for computing the cost function gradient. The main characteristics and advantages of AD are briefly introduced hereafter, while more details can be found in Ref. 11.

The starting principle of AD is that any complex function or computational model F can be split into elementary operations f_i (sum, multiplication, power, etc.), whose derivative is simple and well-known. Formulated in mathematical form, we have the function $F : \mathbb{R}^n \to \mathbb{R}^m$, which evaluates the output vector $Y \in \mathbb{R}^m$ from the input vector $X \in \mathbb{R}^n$ as a composition of p elementary operations

$$Y = F(X) = f_p \circ f_{p-1} \circ \cdots \circ f_1(X). \tag{10}$$

It is then straightforward to evaluate the Jacobian J of F applying the chain rule

$$J = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} = \left[f'_p \circ f_{p-1} \circ \cdots \circ f_1(X) \right] \times \cdots \times f'_1(X) = f'_p(X_{p-1}) \times \cdots \times f'_1(X). \tag{11}$$

In Eq.11, f'_i are Jacobian matrices of the elementary operations and $X_i = f_i(X_{i-1})$ are the intermediate variables obtained at the elementary operation i. Note also that each operation $\times$ is an expensive matrix-matrix multiplication. However, if we multiply Eq. 11 to its right by a perturbed input vector $\dot{X} = e_j$, with all zeros except for a one at element j, we obtain the directional derivative $\dot{Y}$:

$$\dot{Y} = J \times \dot{X} = f'_p(X_{p-1}) \times f'_{p-1}(X_{p-2}) \times \cdots \times f'_1(X) \times \dot{X} = \begin{bmatrix} \frac{\partial y_1}{\partial x_j} & \frac{\partial y_2}{\partial x_j} & \cdots & \frac{\partial y_m}{\partial x_j} \end{bmatrix}^T. \tag{12}$$

Now all the $\times$ operations are cheaper matrix-vector multiplications, going from right to left, and the resulting vector $\dot{Y}$ is an entire column of the Jacobian, *i.e.* the sensitivity of each output with respect to the perturbed input j. This is effectively what the *tangent* or *forward* mode of AD produces.

Alternatively, if we transpose Eq. 11 and multiply to its right by a perturbed output vector $\bar{Y} = e_j$, we obtain the adjoint directional derivative $\bar{X}$:

$$\bar{X} = J^T \times \bar{Y} = f'_1(X)^T \times f'_2(X)^T \times \cdots \times f'_p(X_{p-1})^T \times \bar{Y} = \begin{bmatrix} \frac{\partial y_j}{\partial x_1} & \frac{\partial y_j}{\partial x_2} & \cdots & \frac{\partial y_j}{\partial x_n} \end{bmatrix}. \tag{13}$$

Again, the matrix-matrix multiplications of Eq.11 have been substituted by cheaper matrix-vector multiplications, and the result $\bar{X}$ is an entire row of the Jacobian, *i.e.* the sensitivity of a single output with respect to all inputs, also known as adjoint gradient. This second mode of AD is called *adjoint* or *reverse* mode. It must be underlined that none of the above expressions provide an analytical derivative of F, but only local values of the directional derivative at a given input vector.

Clearly, the two AD modes have different pros and cons, the best choice depending on the application. Tangent AD is preferable when the number of outputs m is large compared to that of the inputs n. In this case, the full Jacobian computation is performed by repeating n-times Eq.12 , perturbing each time a different element of the input vector, thus evaluating the Jacobian column-by-column. Since the computational cost of Eq.12 is roughly the same as of evaluating F, the total cost for the Jacobian calculation is $CPU(J)_{TGT} \sim n \cdot CPU(F)$, scaling proportionally to the number of input parameters. This is about the same cost as required by FD, with the major advantage that AD does not suffer from truncation errors and is exact to machine precision. A simple example demonstrating this is shown in Table1 for the function $F(x) = \cos(x)+x^2$ (we omit the capital letter for input and output vectors). The first column displays a possible computational model for $F(x)$, each row corresponding to an elementary operation involving one or more intermediate variables. The second column shows instead the tangent AD code, starting from the input vector perturbation $\dot{x}_0 = 1.0$. By substituting the dotted variables in the last row, one obtains $\dot{y} = -\sin(x_0)\dot{x}_0 + 2x_0\dot{x}_0$, which exactly equals the analytical directional derivative. The same argument holds for adjoint AD, shown in the third column of Table 1 , which starts with the output vector perturbation and, substituting the bar variables in the last row, yields $\bar{x}_0 = 2x_0\bar{y} - \sin(x_0)\bar{y}$. In contrast, using a finite, non-zero step size h to compute the derivative using forward FD introduces an approximation of the analytical derivative $F'(x) \approx \frac{F(x+h)-F(x)}{h}$, which suffers from a truncation error. Furthermore, it is also evident that the tangent AD code performs roughly the same number of operations as the original code, justifying the computational cost mentioned above.

TABLE 1 Example illustrating the use of AD to differentiate a function calculating $F(x) = \cos(x) + x^2$

Original code	**Tangent AD code**	**Adjoint AD code**
$x_0 = x$	$\dot{x}_0 = 1.0$	$\bar{y} = 1.0$
$x_1 = \cos(x_0)$	$\dot{x}_1 = -\sin(x_0)\dot{x}_0$	$\bar{x}_1 = \bar{y}$
$x_2 = x_0^2$	$\dot{x}_2 = 2x_0\dot{x}_0$	$\bar{x}_2 = \bar{y}$
$y = x_1 + x_2$	$\dot{y} = \dot{x}_1 + \dot{x}_2$	$\bar{x}_0 = 2x_0\bar{x}_2$
		$\bar{x}_0 = \bar{x}_0 - \sin(x_0)\bar{x}_1$

On the other hand, adjoint AD is more efficient when many inputs and few outputs are present, $n \gg m$. In fact, the full Jacobian can be evaluated row-by-row by performing Eq. 13 m-times, each one perturbing a different element of the output vector. The computational cost of Eq. 13 depends on the AD tool, and partially on the complexity of F and user choices, ranging typically from 3 to 10 times the cost of F. Therefore the full Jacobian is retrieved at a cost $CPU(J)_{ADJ} \sim 3 - 10 \cdot m \cdot CPU(F)$, and is roughly independent of the number of input parameters.

In the context of optimization, the number of outputs is essentially one, namely the cost functional, and for the present work only few uncertain parameters will be considered, see Section 4 below. Therefore, tangent AD mode is still more efficient than adjoint AD for calculating the gradient. To do so, we employ the open source AD tool TAPENADE[25], whose applicability on SOLPS-ITER was recently proven[26].

3 | INTERCHANGE TURBULENCE MODEL

The interchange turbulence model derived in Ref. 21 has been adapted for SOLPS-ITER by time and Favre averaging of the Braginskii equations. The full derivation of the model equations for SOLPS-ITER is not within the scope of this work and can be found in these proceedings[27]. Here, we only give a brief overview.

The model foresees a new PDE, describing the rate of change of the turbulent kinetic energy κ, solved together with the other plasma transport equations:

$$\frac{\partial n_i \kappa}{\partial t} + \nabla \cdot (\mathbf{\Gamma}\kappa - D_\kappa \nabla \kappa) = G_\kappa - S_{\kappa,L}. \tag{14}$$

The transport of κ is due to both convection with the mean-field particle flux $\mathbf{\Gamma}$ and turbulent transport, assumed diffusive and driven by κ gradients, with diffusion coefficient $D_\kappa = C_{D_\kappa} n_i D_\perp$ (n_i is the ion density, $D_\perp$ is the radial particle diffusion coefficient and C_{D_κ} is a multiplier). We also assume a diffusive anomalous flux in the mean-field particle flux $\mathbf{\Gamma} = nV_\parallel \mathbf{b} - D_\perp \nabla n$, where $D_\perp$ is made of two components, one related to turbulent transport $D_{E\times B}$ and the other acting as a lower limit set by classical transport D_{min}:

$$D_\perp = D_{E\times B} + D_{min} = C_D \rho_L \sqrt{\frac{\kappa}{m_i}} + D_{min}. \tag{15}$$

In the expression above we already made use of the same relationship as adopted in Ref. 22, which links $D_{E\times B}$ to the local value of κ, automatically retrieving spatially dependent diffusion coefficients. The remaining factors are the global gyro radius ρ_L (set to the OMP separatrix value), the ion mass m_i and a constant multiplier C_D.

The source terms on the right hand side of Eq. 14 describe turbulence interchange production G_κ and turbulence (linear) dissipation $S_{\kappa,L}$:

$$G_\kappa = \left[(T_i + T_e) D_{E\times B} \nabla n_i + \frac{2}{3} n_e \chi_{e,E\times B} \nabla T_e + \frac{2}{3} n_i \chi_{i,E\times B} \nabla T_i\right] \cdot \nabla(\ln(B^2)), \tag{16}$$

$$S_{\kappa,L} = C_L n_i \kappa \frac{c_s}{L_{conn}}. \tag{17}$$

In the expressions above c_s is the plasma sound speed, L_{conn} is the connection length and C_L is a constant multiplier. Moreover, the anomalous electron and ion heat diffusivities are assumed proportional to the turbulent particle diffusion coefficient $\chi_{e/i,E\times B} = C_{\chi,e/i} D_{E\times B}$, where $C_{\chi,e/i}$ are constant multipliers[28]. The production term in Eq. 16 inherently accounts for self-saturation and ballooning, typical of plasma turbulence, as recently shown for COMPASS[22]. In contrast to the model used in Ref. 22 however, we now have additionally included anisothermal effects in the interchange drive based on recent progress in the κ-model development using the TOKAM2D turbulence model[28]. The dissipation term is unchanged compared to Ref. 22. However, in Ref. 27 the modeling of this sink term is further refined by reformulation as a fast parallel transport term of κ, which improves consistency with observations from 3D turbulence simulations.

Additionally, in this work we cut-off at zero the lower limit of range of values of the net source $G_\kappa - S_{\kappa,L}$. While theoretically the net source should be allowed to become negative, we found that this leads to large residuals in the high field side (HFS) region, where κ is reduced to extremely low values, and convergence cannot be guaranteed. We argue that such behavior can be overcome when the sink term, always negative for this setup, is reformulated as parallel transport term, becoming positive at the HFS[27]. With the imposed cut-off, solutions up to machine precision are obtained, with some exceptions detailed in the results Section 5.

In summary, this model self-consistently accounts for the effect of plasma interchange turbulence on radial transport, employing only a limited set of new parameters, C_D, C_L, C_{D_κ}, $C_{\chi,e/i}$, and the solution of an additional PDE, to retrieve 2D distributions

of diffusion coefficients. In contrast, the standard approach requires full specification of radial profiles, that is, several space coordinate-coefficient pairs, together with ballooning parameters. In terms of degrees of freedom, the new κ-model is clearly advantageous.

4 | PARAMETER ESTIMATION CASE STUDY

To prove the working principle of the developed framework, we take a reference SOLPS-ITER solution obtained with the κ-model and a fluid neutral approximation, which we use to produce fictitious experimental data. We then verify how well the estimation procedure is able to match these data and retrieve the reference parameters.

The SOLPS-ITER version employed in this study features the recently implemented unstructured solver, which allows flexible vessel geometry description and grids extending up to the first wall[29]. This version inherently accounts for grid misalignment, an important requirement to obtain reliable fluid neutral simulations[30]. Furthermore, the most up-to-date Advanced Fluid Neutral (AFN) models available[31] are employed here, which provide similar accuracy as computationally expensive Monte Carlo kinetic neutrals in highly collisional regimes, without the need for additional parameters in contrast to other fluid models, and avoiding statistical noise which hampers gradient accuracy.

The reference solution is obtained on a COMPASS single-null configuration on a 48×24 grid, with pure deuterium plasma and without drifts, similarly to Ref. 22. At the core, we impose a fixed density boundary condition (BC) for D^+ ($n_i = 3.25 \times 10^{19}$ m^{-3}) and zero particle flux for D^0; additionally, we fix the ion and electron power ($P_i = P_e = 172$ kW), the turbulent kinetic energy value ($\kappa = 1.84$ eV), and zero parallel velocity. At the targets, a standard Bohm-Chodura BC is applied, while at the last flux surface and private flux region leakage BCs are applied with factor 3×10^{-3} for D^+ and κ, 1×10^{-2} for D^0 and 1.5×10^{-2} for heat, together with zero-gradient BC for parallel velocity. Surfaces are assumed 100% recycling apart from a 1% absorption set at the private flux boundary to account for pumping. Additional parameters of the κ-model take the following values, similar to the COMPASS application of Ref. 22: $C_D = 0.1$, $C_L = 1.6$, $C_{\chi,e} = C_{\chi,i} = 10$, $C_{D_\kappa} = D_{min} = 0.01$. Note that for simplicity we employ in this work a single parameter C_L for the whole domain, without distinguishing between closed and open field line regions as was instead done in the COMPASS study[22].

We compare the regression and MAP cost functions for a varying number of available measurement points N_D, and different levels of measurement noise σ_{ref}. These fictitious data points are obtained from the OMP plasma density profile of the reference solution, to which we add Gaussian noise in an absolute way $\mathcal{N}(0, \sigma_{ref})$. Five levels of absolute noise are envisaged with $\sigma_{ref} = [0.4, 0.8, 1.6, 3.2, 6.4]\ 10^{18}\mathrm{m}^{-3}$. For each σ_{ref}, five different numbers of data points are used $N_D = [5, 10, 20, 40, 80]$, for a total of 25 cases.

The optimization is then carried out for each of these cases considering the set of uncertain parameters $\theta = [\kappa_{core}, C_L, C_D, \sigma]^T$, where κ_{core} is the turbulent kinetic energy value imposed at the core boundary. The initial guess for the optimization is arbitrarily chosen $\theta_0 = [3, 3, 0.8, 0.1]^T$, while the lower bound $\theta_L = [0.1, 0.1, 0.05, 0.01]^T$ and the upper bound $\theta_U = [5, 5, 1, 2]^T$ are educated guesses for the parameter ranges. Note that only for the MAP estimation we consider the nuisance parameter σ, which is modeled as an additive error. For all parameters we suppose a uniform prior within the imposed range.

For the optimization we employ the BFGS algorithm from the PETSc/TAO library[32–34], setting tolerance 10^{-4} for the gradient norm ($||\nabla\hat{J}||$), gradient relative norm ($||\nabla\hat{J}||\backslash|\hat{J}|$) and gradient reduction ($||\nabla\hat{J}||\backslash||\nabla\hat{J}_0||$) stopping criteria. Additionally, we limit the iterations to a maximum of 50, after which the optimization is terminated. Note that each time the optimizer computes the cost function and/or its gradient, it will converge the SOLPS-ITER solution for 3000 iterations. While this was sufficient to decrease the residuals of the reference case to $\sim 10^{-3}$, it may not be the most efficient or accurate solution. An alternative we foresee to implement in the future is the possibility to impose a threshold for the residuals, which can be gradually decreased during the optimization.

5 | RESULTS AND DISCUSSION

Before comparing in detail nonlinear regression and MAP estimation results, we give a brief overview of the optimization process. We notice that in general nonlinear regression convergence is faster than MAP, probably due to the increased parameter set of the latter. Indeed, as a verification test, we also performed a MAP estimation adopting a fixed value of σ together with uniform priors and no error correlation in the covariance matrix, effectively performing a least-squares optimization. If we

compare this with nonlinear regression the results are very close to each other, the small remaining difference being due to the volume-weight in Eq. 3 and the different scaling of terms in the cost function and gradient.

Moreover, in some cases both nonlinear regression and MAP estimation failed in the line search, despite providing a good match in the density profile. This can sometimes be attributed to the semi-smooth nature of the simulation code combined with enforcing strong Wolfe conditions for the line search, which we plan to relax to only using the Armijo condition of sufficient decrease [23]. These line search failures are most of the times accompanied by numerical instabilities of the κ-model, with large residuals in the HFS region or close to the X-point, linked to larger values of the parameter C_D (this despite the already imposed cut-off for the net κ source). While the solution at the OMP is not affected by these instabilities, they can nevertheless compromise the gradient computation and point towards a reduction of upper bound θ_U for C_D or limits of the κ-model itself. Cases in which the line search failed are discarded, thus less than 25 cases are available for both the analysis of the nonlinear regression and MAP.

Finally, in a few cases the optimization was trapped in a clearly unsatisfactory local optimum from the point of view of the OMP density profile. In such cases, selecting a different initial point $\theta_0 = [1, 3, 0.3, 0.1]^T$ and upper bound $\theta_U = [3, 5, 0.5, 2]^T$ with smaller values for κ_{core} and C_D, in line with the argument above, was sufficient to avoid such trivial local optimum.

The results of the estimation process carried out are shown for a few selected cases in Fig. 1 . In the figure, plots on the same row have equal number of data points N_D and different level of noise σ_{ref}, and plots on the same column have thus equal level of noise and different number of data points. It can be noticed that both regression and MAP estimation provide a reasonable fit of the data points available, the two profiles being close to each other for the different cases.

An important difference between MAP estimation and regression analysis is underlined by the green shaded area, indicating the $\pm 2\sigma$ uncertainty region estimated by the MAP, which captures almost all data points. Regression analysis does not give any information on the general goodness-of-fit apart from the mean squared error, while the MAP provides a quantitative measure, essential for model validation purposes. However, the prediction error information is available *only* for the plasma density profile, without distinguishing between modeling and experimental uncertainty. To assess the uncertainty on other plasma quantities, we should first evaluate the posterior distribution of the model parameters and then propagate their uncertainties, namely perform a backward and forward uncertainty quantification. As explained in Section 2.2, the Laplace approximation can be employed for this purpose and we plan to apply this methodology in future work.

Interesting insights can be drawn from the scatter plots of the estimated parameters shown in Fig. 2 , in which each point corresponds to an estimation with a specific noise level and number of data points.

From these plots, it is clear that there is no unique solution to the estimation procedure, as very similar OMP density profiles are obtained with different combinations of parameters. This is logical, as we include in the cost function only the OMP density profile and, to retrieve the same radial gradient, one needs the same diffusion coefficient. Given that for our model $D_{E\times B} \sim C_D\sqrt{\kappa}$ in Eq. 15, it is straightforward to see that the same diffusion coefficient can be obtained in infinitely many combinations of C_D and κ, neglecting the dependency on Larmor radius ρ_L and D_{min}. The same argument explains also the evident $C_D \sim 1/\sqrt{\kappa_{core}}$ correlation in Fig. 2 (a). Two clear outliers from the expected trend are also visible, possibly due to a locally optimal solution.

Less straightforward is the role played by the parameter C_L in the dissipation term $S_{\kappa,L}$, see Eq. 17. We would expect that a larger source of κ imposed at the core requires more dissipation to attain the same turbulent kinetic energy profile, namely a larger C_L. However, the estimation results in Fig. 2 (b) show that C_L is lower for high κ_{core} values. This behavior is caused again by correlations among parameters, which can be derived by assuming a local balance between turbulence production and dissipation $G_\kappa = S_{\kappa,L}$. Since the production term scales with the turbulent diffusion coefficient $D_{E\times B}$, see Eq. 16, it follows from our argument above that it should be almost constant. In this case also the dissipation term in Eq. 17 is constant, so that $C_L \sim 1/\kappa_{core}$. However, as can be seen from the trendline in Fig. 2 (b), the dependence is much stronger than $1/\kappa_{core}$, possibly due to non perfect local balance among turbulence production and dissipation or additional nonlinear effects. For the same reason the correlation among $C_D - C_L$ observed from the data in Fig. 2 (c) also deviates from the theoretical $C_D \sim \sqrt{C_L}$, obtained by combining $C_D \sim 1/\sqrt{\kappa}$ and $C_L \sim 1/\kappa$, again in the assumption of perfect local balance of κ.

These strong correlations among parameters with non-unique solutions indicate an ill-conditioned, multi-modal optimization problem, showing the need to reduce the degrees of parametrization. Similar parameter correlations in a simplified version of the κ and $\kappa - \zeta$ models were also encountered in Ref. 19 when using insufficient data in the cost functions (the enstrophy ζ is a quantity related to viscous dissipation of turbulent kinetic energy). There it was shown that the ambiguity could be resolved by adding also measurements of κ and ζ in the cost function for parameter estimation, which would unambiguously determine not only the value of $D_{E\times B}$, but also those of the underlying turbulence characteristics. A similar approach could be used here by

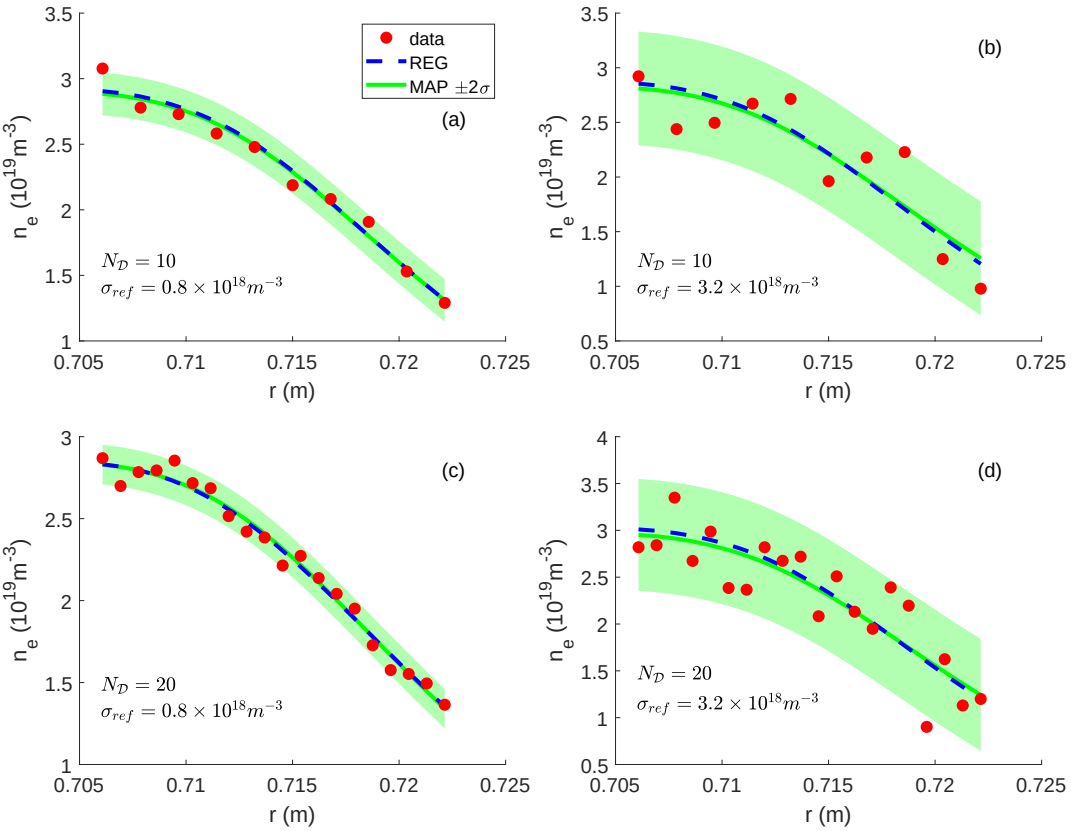

FIGURE 1 Optimized plasma density profile at the OMP using regression (blue dashed line) and MAP (green solid line) cost functions. The shaded green area indicates the 2σ range estimated with the MAP while fictitious experimental data points are indicated with red circles. Plots on the same row have equal number of data points N_D while plots on the same column have equal level of noise σ_{ref}.

adding synthetic data on κ to the cost function, and is left for future work. A more general solution is to obtain information on such parameters or the level of turbulent kinetic energy either from turbulence codes or from data fluctuations in experiments.

Finally, Fig. 2 (d) shows the estimated absolute standard deviation σ for different number of data points and level of noise. As can be seen, the reference value is very well captured by the MAP estimation, with improving accuracy if more data points are included in the procedure, as expected. Note that the reference noise levels do not correspond exactly to the five levels $\sigma_{ref} = [0.4, 0.8, 1.6, 3.2, 6.4]\ 10^{18}\mathrm{m}^{-3}$ mentioned in Section 4, as the actual data points are randomly generated. In fact, the plotted reference noise level is an effective standard deviation of the generated data with respect to the reference solution.

6 | CONCLUSIONS

We developed a new parameter estimation framework in SOLPS-ITER, relying on gradient-based optimization to minimize a cost function, which provides an objective metric to compare simulation results and experimental data. The gradient information required for the optimization is obtained through tangent AD with the TAPENADE tool. Depending on the specific estimation case, AD provides machine-precision accurate sensitivities more efficiently with its tangent mode (few uncertain parameters) or adjoint mode (many uncertain parameters). This already improves a previously developed estimation framework employing nonlinear regression, in which gradient information was based on finite differences, suffering from truncation errors. In addition, we employed the principles of Bayesian inference to implement a MAP estimation accounting for numerical, modeling, and experimental uncertainties, which regression techniques do not consider. The new estimation framework has been tested

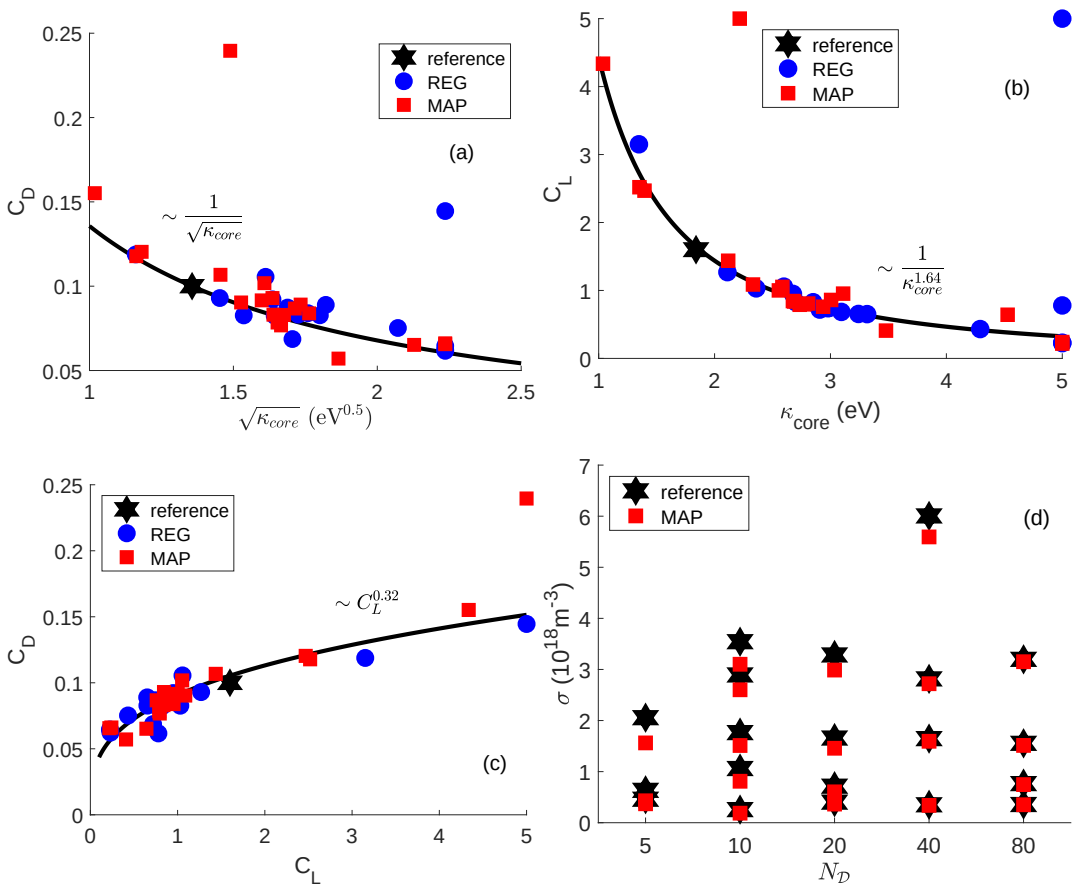

FIGURE 2 Scatter plots of the parameters estimated with regression (blue circles) and MAP (red squares) cost functions. The parameters adopted in the reference simulation are indicated with a black star. The solid black lines indicate the fitting curve defined by the equation in each plot.

on a simple case adopting fictitious OMP plasma density profile measurements obtained from a reference SOLPS-ITER simulation using the recently implemented κ turbulence model. We then compared nonlinear regression and MAP estimation for different number of data points available and introducing different levels of measurement noise. Both estimation methods gave similar results adopting an absolute prediction error in the likelihood, with the additional MAP-estimated uncertainty correctly encompassing the noisy fictitious data. The three κ-model parameters assumed as unknowns in the estimation showed strong correlations among each other and non-unique solutions for the same density profile were obtained. This is likely because the cost functions used in this study insufficiently constrains the κ-model, and points towards the need of including also information on turbulence levels in the parameter estimation procedure to correctly determine the model constants.

We plan on resolving the non-uniqueness of the solutions by additional constraints in the optimization, introducing information on the turbulent kinetic energy level coming from experimentally measured fluctuations or turbulence code results. This can be beneficial for the κ-model itself, avoiding numerical instabilities encountered with certain parameter values.

Regarding the optimization method, several paths are open for future investigations to improve convergence. We will explore the impact of the choice of optimization algorithm on speed and robustness of convergence. In this perspective, we will also assess the effect of employing the Armijo instead of the strong Wolfe condition in the line search methods, but also consider alternatives such as trust-region or interior-point methods. On the longer term we will resort to one-shot methods, which have proven faster convergence for PDE-constrained optimizations [8].

Furthermore, we are in the process of deploying adjoint AD in SOLPS-ITER, allowing estimations on more complex and realistic cases with several uncertain parameters, and the extension of the developed optimization framework to divertor design studies.

We also foresee to expand the Bayesian setting implemented in this work, both in terms of improvement and application. Considering the first, spatial error correlation should be included in the covariance matrix of the likelihood function, together with possible additional correlations among different diagnostics, for which close collaboration with diagnosticians is of foremost importance. Concerning application, we plan on performing an extended estimation study on real experimental cases, together with full uncertainty quantification exploiting the Laplace approximation[17,24]. This will also allow to formally and quantitatively compare the two competing models for the radial turbulent transport in SOLPS-ITER, namely the standard model with ad-hoc anomalous diffusion coefficients, against the newly developed κ-model.

ACKNOWLEDGMENTS

This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014–2018 and 2019-2020 under grant agreement No. 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission. The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation Flanders (FWO) and the Flemish Government – department EWI. R. Coosemans is funded by a strategic basic research grant of the Research Foundation Flanders (FWO), file number 48697.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

[1] S. Wiesen, *et al.*, *Journal of Nuclear Materials* **2015**, *463*, 480–484.

[2] X. Bonnin, *et al.*, *Plasma and Fusion Research* **2016**, *11*, 1403102.

[3] A. Xuereb, *et al.*, *Journal of Nuclear Materials* **2010**, *396*, 228–233.

[4] L. Aho-Mantila, *et al.*, *Nuclear Fusion* **2012**, *52* (10), 103006.

[5] F. Reimold, *et al.*, *Journal of Nuclear Materials* **2015**, *463*, 128–134.

[6] D.P. Coster, *et al.*, *Contributions to Plasma Physics* **2000**, *40*, 334–339.

[7] J.-W. Kim, *et al.*, *Journal of Nuclear Materials* **2001**, *290-293*, 644–647.

[8] W. Dekeyser, *et al.*, *Journal of Computational Physics* **2014**, *278*, 117–132.

[9] M. Blommaert, *et al.*, *Contributions to Plasma Physics* **2016**, *56*, 796–801.

[10] M. Baelmans, *et al.*, *Plasma Physics and Controlled Fusion* **2014**, *56*, 114009.

[11] A. Griewank, A. Walther, *Evaluating Derivatives*, SIAM, **2008**.

[12] F. Alauzet, *et al.*, *European Journal for Computational Mechanics* **2008**, *17* (1-2), 245–269.

[13] A. Höuck, *et al.*, *Optimization Methods and Software* **2017**, DOI:10.1080/10556788.2017.1396602.

[14] F. McGreivy, *et al.*, *Nuclear Fusion* **2021**, *61*, 026020.

[15] V. Dose, *Reports on Progress in Physics* **2003**, *66*, 1421.

[16] S.H. Cheung, *et al.*, *Reliability Engineering and System Safety* **2011**, *96*, 1137–1149.

[17] D.I. Papadimitriou, *et al.*, *Computer and Fluids* **2015**, *120*, 82–97.

[18] R. Fischer, *et al.*, *Fusion Science and Technology* **2010**, *58*, 675–684.

[19] R. De Wolf, *et al.*, *Nuclear Fusion* **2021**, *61*, 046048.

[20] C. Bowman, *et al.*, *Plasma Physics and Controlled Fusion* **2020**, *62*, 045014.

[21] R. Coosemans, *et al.*, *Physics of Plasmas* **2021**, *28*, 012302.

[22] S. Carli, *et al.*, *Contributions to Plasma Physics* **2020**, *60*, e201900155.

[23] J. Nocedal, S.J. Wright, *Numerical Optimization*, Springer, **2006**.

[24] L. Tierney, B.J. Kadane, *Journal of the American Statistical Association* **1986**, *81*, 82–86.

[25] L. Hascoet, V. Pascual, *ACM Transactions on Mathematical Software* **2013**, *39* (3), 20.

[26] S. Carli, *et al.*, *Nuclear Materials and Energy* **2019**, *18*, 6–11.

[27] W. Dekeyser, *et al.*, *These proceedings*.

[28] R. Coosemans, *et al.*, *These proceedings*.

[29] W. Dekeyser, *et al.*, *Nuclear Materials and Energy* **2021**, *27*, 100999.

[30] W. Dekeyser, *et al.*, *Nuclear Materials and Energy* **2019**, *18*, 125–130.

[31] W. Van Uytven, *et al.*, *Contributions to Plasma Physics* **2020**, *60*, e201900147.

[32] S. Balay, S. Abhyankar, M.F. Adams, S. Benson, J. Brown, P. Brune, K. Buschelman, E.M. Constantinescu, L. Dalcin, A. Dener, V. Eijkhout, W.D. Gropp, V. Hapla, T. Isaac, P. Jolivet, D. Karpeev, D. Kaushik, M.G. Knepley, F. Kong, S. Kruger, D.A. May, L.C. McInnes, R.T. Mills, L. Mitchell, T. Munson, J.E. Roman, K. Rupp, P. Sanan, J. Sarich, B.F Smith, S. Zampini, H. Zhang, H. Zhang, J. Zhang, *PETSc Web page*, https://petsc.org/, **2021**. https://petsc.org/.

[33] S. Balay, S. Abhyankar, M.F. Adams, S. Benson, J. Brown, P. Brune, K. Buschelman, E.M. Constantinescu, L. Dalcin, A. Dener, V. Eijkhout, W.D. Gropp, V. Hapla, T. Isaac, P. Jolivet, D. Karpeev, D. Kaushik, M.G. Knepley, F. Kong, S. Kruger, D.A. May, L.C. McInnes, R.T. Mills, L. Mitchell, T. Munson, J.E. Roman, K. Rupp, P. Sanan, J. Sarich, B.F Smith, S. Zampini, H. Zhang, H. Zhang, J. Zhang, PETSc/TAO Users Manual, *ANL-21/39 - Revision 3.16*, Argonne National Laboratory, **2021**.

[34] S. Balay, W.D. Gropp, L.C. McInnes, B.F. Smith, in *Modern Software Tools in Scientific Computing*, (Eds: E. Arge, A. M. Bruaset, H. P. Langtangen), Birkhäuser Press, **1997**, pp. 163–202.

How to cite this article: Carli S. *et al.* (20XX), Bayesian MAP-estimation of κ turbulence model parameters using Algorithmic Differentiation in SOLPS-ITER, *Contributions to Plasma Physics*, *20XX;00:0–0*.