Iterative Bayesian Optimization of an Implicit LES Method for Underresolved Simulations of Incompressible Flows

Josef M. Winter
Institute of Aerodynamics and Fluid Mechanics
Technische Universität München
Garching, Germany
josef.winter@tum.de

Felix S. Schranner
Institute of Aerodynamics and Fluid Mechanics
Technische Universität München
Garching, Germany
schranne@tum.de

Nikolaus A. Adams
Institute of Aerodynamics and Fluid Mechanics
Technische Universität München
Garching, Germany
Nikolaus.Adams@tum.de

ABSTRACT

In the numerical simulation of turbulent flows, resolution is often low. The solution in those under-resolved regions is strongly affected by the truncation error of the underlying numerical schemes. Although, the truncation error can be used to model the evolution of otherwise resolved scales - it acts as a physically consistent subgrid-scale (SGS) model. In particular, the truncation error of high-order WENO-based schemes can function as an implicit SGS model. The sixth-order adaptive central-upwind weighted essentially non-oscillatory scheme with implicit scale-separation, denoted as WENO-CU6-M1, allows for physically consistent implicit SGS modeling when its parameters are chosen properly.

Schranner *et al.* (2016) determined an optimal combination of the WENO-CU6-M1 modeling parameters by means of design optimization (DO). Optimizing the WENO-CU6-M1 parameters is computationally expensive. This work proposes a general iterative optimization algorithm which is based on hierarchical Bayesian modeling. It allows to reduce the computational costs for optimizing WENO-CU6-M1 parameters.

Optimization implies the procedure of identifying input-output relations of a system. Thereby, the minimum or the maximum output of a system, as well as the input for which it occurs, may be sought. The proposed optimization algorithm relies on Gaussian Processes (GP). Methods to fit a GP are proposed. Based thereupon, an optimization algorithm which iteratively improves the quality of the GP is introduced. To complete this work, the algorithm is used to identify an optimal parameter set for the WENO-CU6-M1 scheme.

INTRODUCTION

Modified differential equation analysis (MDEA) (Margolin & Rider (2002)) has shown that the truncation error of nonlinear discretization schemes can be constructed to represent an implicit subgrid-scale (SGS) model for turbulent flows (Adams *et al.* (2004)). The nonlinear regularization mechanism of high-order finite volume schemes with shock-capturing capabilities can be employed for implicit large-eddy simulations (ILES). For a review refer to Grinstein *et al.* (2007). On the basis of a spectral extension of the MDEA, the truncation error of a nonlinear scheme has been designed to recover the theoretical spectral eddy viscosity when the flow is turbulent and under-resolved. Such a situation, where the non-negligible local truncation error of a numerical scheme recovers correct physical SGS behavior, is called physically consistent behavior (Balsara & Shu (2000); Hickel *et al.* (2006)).

Hu *et al.* (2010) proposed a weighted, essentially non-oscillatory (WENO) scheme combining the advantages of an upwind scheme, e.g., the fifth-order WENO scheme (Jiang & Shu (1996)), and a sixth-order central scheme. It adaptively alters its biasing between central and upwind by evaluation of the smoothness indicators of the optimal higher-order stencil and lower-order upwind stencils. Thereby, it decreases numerical dissipation in smooth flow regions and permits a numerically stable solution in non-smooth flow regions while preserving shock-capturing capabilities. It is denoted as WENO-CU6. Based thereupon, a central-upwind WENO scheme with implicit SGS modeling capabilities has been developed (Hu & Adams (2011)); it is denoted as WENO-CU6-M1.

Schranner *et al.* (2013) identified that WENO-CU6-M1 offers a set of free parameters, enabling implicit subgrid-scale modeling by controlling scale separation of resolved and non-resolved scales for compressible as well as incompressible flows. In combination with the material modeling, i.e., equation of state (EOS), and an appropriate Riemann solver, an adjustment of the model permits recovery of self-similar isotropic turbulence when physical viscosity diminishes.

Schranner *et al.* (2016) determined an optimal combination of the WENO-CU6-M1 modeling parameters by means of design optimization (DO). The notion of being optimal implies that the transition is predicted physically consistently and inertial subrange scaling, which is characteristic of isotropic turbulence, is recovered as most optimally when dissipating energy only within the SGS.

In this work, the applicability of an iterative stochastic optimization approach is investigated. The application to find an optimal combination of the WENO-CU6-M1 modeling parameters is sought. In the following, the formulation of the optimization problem and an appropriate optimization algorithm are layed out.

MODEL FORMULATION

Weakly compressible and barotropic fluids can be modelled by Tait's equation of state (EOS). Tait's EOS decouples the energy equation from the continuity and momentum equations. Thus, the flow is governed by equations for the conservation of mass and momentum, only. In one dimension (for simplicity), $\mathbf{U} = (\rho, \rho u)$ is the solution of

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial}{\partial x}\mathbf{F}(\mathbf{U}) = 0 \tag{1}$$

In a discrete space-time-domain, the discrete conservation equation

$$\widehat{\mathbf{U}}_{[i]}^{n+1} = \widehat{\mathbf{U}}_{[i]}^{n} + \frac{\Delta t}{\triangle x_{[i]}}\left(\overline{\mathbf{F}}_{[i-\frac{1}{2}]} - \overline{\mathbf{F}}_{[i+\frac{1}{2}]}\right) \tag{2}$$

for the cell-averaged solution $\widehat{\mathbf{U}}_{[i]} = \frac{1}{\Delta x_{[i]}} \int\limits_{\Delta x_{[i]}} \mathbf{U}_{[i]}^{n} dx$ requires approximations of the cell-face fluxes

$$\overline{\mathbf{F}}_{[i\pm\frac{1}{2}]} = \frac{1}{\Delta t}\int\limits_{t^n}^{t^{n+1}} \mathbf{F}_{[i\pm\frac{1}{2}]}\, dt = \frac{1}{\Delta t}\int\limits_{t^n}^{t^{n+1}} \mathbf{F}\left(\mathbf{U}_{[i\pm\frac{1}{2}]}\right)\, dt \tag{3}$$

where Δt is sufficiently small.

A low-dissipation advective flux approximation is due to the Roe (1981) approximate Riemann solver, which is used within the scope of this work. The key idea of these is to use the linearized local flux Jacobian $\widetilde{\mathbf{A}} = \widetilde{\mathbf{A}}(\overline{U}^{(l)}, \overline{U}^{(r)})$. (l) and (r) denote the high-order reconstructed conservative states at the left and right side of cell-face $[i \pm \frac{1}{2}]$. The eigenvalues of $\widetilde{\mathbf{A}}$, $\widetilde{\lambda}^{(j)}(\overline{U}^{(l)}, \overline{U}^{(r)})$ and right eigenvectors $\widetilde{\mathbf{K}}^{(j)}(\overline{U}^{(l)}, \overline{U}^{(r)})$ are determined, so that the Roe numerical flux can be computed as

$$\overline{\mathbf{F}}_{[i+\frac{1}{2}]} = \frac{1}{2}\left(\overline{\mathbf{F}}(\overline{\mathbf{U}}^{(l)}) + \overline{\mathbf{F}}(\overline{\mathbf{U}}^{(r)})\right) - \frac{1}{2}\sum_{j=1}^{m} \widetilde{\delta}^{(j)} \left|\widetilde{\lambda}^{(j)}\right| \widetilde{\mathbf{K}}^{(j)} \qquad (4)$$

where $\widetilde{\delta}^{(j)}$ denote the wave strengths.

The conservation equations (2) are integrated explicitly in time with a 3^{rd} order TVD Runge-Kutta scheme Shu & Osher (1988).

In order to obtain $\overline{\mathbf{U}}^{(\alpha)}_{[i\pm\frac{1}{2}]}$, where α is either l or r, weighted essentially non-oscillatory (WENO) schemes define m reconstruction polynomials for non-averaged conservative cell-face vector candidates

$$\mathbf{u}^{(\alpha,\gamma)}_{i+\frac{1}{2}} = \sum_{j=0}^{m-1} c_{\gamma,j}\widehat{\mathbf{U}}_{i-\gamma+j} \qquad \gamma = 0,...,m-1 \qquad (5)$$

on m candidate stencils $S_\gamma[i] \equiv \left\{\widehat{\mathbf{U}}_{[i-\gamma]},...,\widehat{\mathbf{U}}_{[i]},...,\widehat{\mathbf{U}}_{[i-\gamma+m-1]}\right\}$ in the viscinity of the cell-face and combine these convexly according to

$$\overline{\mathbf{U}}^{(\alpha)}_{[i\pm\frac{1}{2}]} = \sum_{\gamma=0}^{m-1} \omega^{(\alpha)}_\gamma \mathbf{u}^{(\alpha,\gamma)}_{i+\frac{1}{2}} \qquad (6)$$

(Shu (1998)). Hereby, the set of nonlinear weights $\left\{\omega^{(\alpha)}_\gamma\right\}$, satisfying $\omega^{(\alpha)}_\gamma \geq 0$, $\sum_{\gamma=0}^{m-1}\omega^{(\alpha)}_\gamma = 1$, ensures stability and consistency. Jiang & Shu (1996) have formulated computationally efficient weights such that these are C^∞, i.e. smooth functions of the involved cell averages:

$$\omega^{(\alpha)}_\gamma = \frac{\alpha^{(\alpha)}_\gamma}{\sum_{s=0}^{m-1}\alpha^{(\alpha)}_s} \qquad \alpha^{(\alpha)}_\gamma = f\left(d_\gamma, \beta^{(\alpha)}_\gamma\right) \qquad (7)$$

where d_β and $\beta^{(\alpha)}_\gamma$ are the ideal weights and *smoothness indicators*, respectively.

"The smoothness indicators diminish with increasing smoothness of the solution on a stencil" (Schranner *et al.* (2013)). In defining α_γ, the core idea is to consider each of the $\mathbf{u}^{(\gamma)}_{i+\frac{1}{2}}$ according to their smoothness by weighting them appropriately. Thereby, $\omega^{(\alpha)}_\gamma$ approximates d_β. Yet, if $\mathbf{u}(x)$ was to contain a discontinuity in at least one of the stencils $S_\gamma[i]$, leading to $\beta^{(\alpha)}_\gamma = \mathcal{O}(1)$, the corresponding weights $\omega^{(\alpha)}_\gamma$ needs to diminish to exclude the approximation $\mathbf{u}^{(\gamma)}_{i+\frac{1}{2}}$ and thereby keep the overall non-oscillatory behaviour. The WENO weighting factors $\alpha^{(\alpha)}_\gamma$ of Jiang & Shu (1996) fulfill these requirements:

$$\alpha^{(\alpha)}_\gamma = \frac{d_\gamma}{\left(\varepsilon + \beta^{(\alpha)}_\gamma\right)^q} \qquad (8)$$

The WENO-CU6-M1 weights, a further-development of the original WENO-weighting presented in equation (8), remedy excessive dissipation of the underlying WENO-CU6 scheme while preserving its shock-capturing properties and thus allow to recover physical consistency for both, the solenoidal and the dilatational components of the velocity field, without the need to explicitely distinguish these (Hu & Adams (2011)). The WENO weighting factors of the WENO-CU6-M1 scheme are:

$$\alpha^{(\alpha)}_\gamma = d_\gamma\left(C_q + \frac{\tau^{(\alpha)}_6}{\varepsilon + \beta^{(\alpha)}_\gamma}\right)^q \qquad \gamma = 0,...,3 \qquad (9)$$

where the reference smoothness indicator is $\tau^{(\alpha)}_6 = \beta^{(\alpha)}_3 - \frac{1}{6}\left(\beta^{(\alpha)}_0 + 4\beta^{(\alpha)}_1 + \beta^{(\alpha)}_2\right)$, according to Hu *et al.* (2010) and Borges *et al.* (2008). $\mathbf{u}^{(\alpha,\gamma)}_{i+\frac{1}{2}}$, d_γ, $\beta^{(\alpha)}_\gamma$, with $\gamma = 0,1,2$ are identical to the ones of the 5^{th} order WENO scheme, which can be found in Shu (1998). $\mathbf{u}^{(\alpha,3)}_{i+\frac{1}{2}}$, d_3, $\beta^{(\alpha)}_3$, as well as the four ideal weights may be found in Schranner *et al.* (2016).

OPTIMIZATION

Statement of the Problem

In the formulation of the WENO-CU6-M1 scheme two parameters can be identified. Those are the integer power exponent q, which controls the amount of nonlinear dissipation (Henrick *et al.* (2005)), and the linear weight bias C_q, which reduces numerical dissipation for higher C_q (Hu & Adams (2011)). They define the parameter set $\mathbf{x} = (C_q; q)$.

For a given material modeling and Riemann solver, an optimal parameter set is sought so that the thereby stated implicit LES model is fulfilling two design requirements best. Firstly, transitioning of the incompressible, physically inviscid, originally two-dimensional Taylor-Green vortex (TGV) (Taylor & Green (1937)) to three-dimensional statistically isotropic turbulence (Fauconnier *et al.* (2009)) must occur. Secondly, Kolmogorov scaling inertial subrange, i.e. $E(k) = C\varepsilon^{2/3}k^{-5/3}$, ought to be recovered most optimally as a consequence of proper transition. C and ε refer to the Kolmogorov constant and the the dissipation rate, respectively.

The global optimum is expected in the parameter domain $\mathbf{X} = 1{,}000 \leq C_q \leq 20{,}000 \times 1 \leq q \leq 20$ (Hu & Adams (2011); Schranner *et al.* (2013)). Schranner *et al.* (2016) define the quality of a sample parameter set - and thus the value of the target function $z(\mathbf{x})$ - as the total least-squares difference between the numerically simulated Kolmogorov scaling inertial subrange $E[k_i]$ and the original Kolmogorov scaling inertial subrange

$$z(\mathbf{x}) = \sum_{i=n}^{m}\left(E[k_i] - Ak_i^{-5/3}\right)^2 \qquad (10)$$

where n and m mark the first and last wave number of the inertial subrange. $A = C\varepsilon^{2/3}$ summarizes the dependence of the theoretical Kolmogorov scaling inertial subrange on the Kolmogorov constant C, and the dissipation rate ε. Evaluation of this function requires knowledge of A, which is different for each sample. Linear

regression analysis presented in Schranner *et al.* (2016) is used to determine A. Since one is only interested in the deviation from the $-5/3$-law, we use

$$z(\mathbf{x}) = \sum_{i=n}^{m} \left(\frac{E[k_i]}{A} - k_i^{-5/3} \right)^2 \tag{11}$$

as target function in this work. Therein, the numerically simulated spectra are normalized with A.

Finally, the optimization problem is to find the parameter set $\mathbf{x} = (C_q, q)$, also called input vector, where the global $\min(z(\mathbf{x}))$ of the target function $z(\mathbf{x})$ occurs. The target function is evaluated after $t = 20s$. Completion of transition is expected at $t \approx 9s$ (Schranner *et al.* (2016)). Simulating a TGV evolution up to $t = 20s$ on a domain with 64 cells in each of the three spatial directions requires approximately 80 minutes on 64 processors on dual-socket Intel SandyBridge-EP Xeon E5-2670 nodes.

Iterative Bayesian Optimization Algorithm

The response surface approach (Schranner *et al.* (2016); Jones *et al.* (1998)) reduces the complexity of otherwise expensive target functions. Response surfaces, also denoted as surrogate models, mimic the behavior of the target function $z(\mathbf{x})$ by means of an approximate $f(\mathbf{x})$. Typically, they are simpler to evaluate and can be optimized by standard optimization techniques. The surrogate model is constructed based on a training set $\mathscr{D} = \{(\mathbf{x}_i, y_i) \mid i = 1, \ldots, n\}$ of n observations, with input vectors $\mathbf{x}_i$ and corresponding observations y_i. For notational simplicity, the matrix $\mathbf{X}$ which comprises the training points $\mathbf{x}_i$ is introduced. Observations are associated to the target function by $y = f(\mathbf{x}) + \varepsilon$, with Gaussian measurement noise ε. For computer experiments $\varepsilon = 0$ holds, i.e. observations accord to target function values. In the scope of this work, a Gaussian Process (GP) serves as surrogate model.

Gaussian Process A Gaussian Process is a non-parametric stochastic surrogate model. Non-parametric implies that the structure of the model is variable and free of parameters. In general, non-parametric models are based on fewer assumptions than parametric ones and thus are more flexible. These are often applied when little is known about the functional dependencies of the system of interest. To define the prior distribution of a GP, the mean $m(\mathbf{x}; \theta_m)$ and covariance function $k(\mathbf{x}, \mathbf{x}'; \theta_k)$ of the distribution over functions have to be specified. The covariance function provides the covariance between two locations $\mathbf{x}$ and $\mathbf{x}'$ in the input space. θ_m and θ_k are a priori unknown free hyperparameters. Hyper emphasizes that these are parameters of the prior distribution of a non-parametric model. Hyperparameters describe properties, such as characteristic length-scales or smoothness of the mean, respective, covariance function. Thus, prior knowledge of the target function can be incorporated by the choice of the mean and covariance function and their hyperparameters.

At unobserved points $\mathbf{x}_*$, a GP predicts the function value $f_* = f(\mathbf{x}_*)$ to be normally distributed with mean $\bar{f}_*(\mathbf{x}_*)$ and covariance $\mathrm{Cov}(f_*(\mathbf{x}_*))$, i.e. $f_* \sim \mathscr{N}(\bar{f}_*(\mathbf{x}_*), \mathrm{Cov}(f_*(\mathbf{x}_*)))$. Analytical expressions for the posterior predictive mean and covariance in the absence of measurement noise $\varepsilon = 0$ are given by

$$\begin{aligned} \bar{f}_* &= \mathbf{m}(\mathbf{x}_*; \theta_m) \\ &\quad + \mathbf{k}(\mathbf{x}_*, \mathbf{X}; \theta_k)^T K^{-1} (\mathbf{y} - \mathbf{m}(\mathbf{X}; \theta_m)) \end{aligned} \tag{12}$$

and

$$\begin{aligned} \mathrm{Cov}(f_*) &= k(\mathbf{x}_*, \mathbf{x}_*; \theta_k) \\ &\quad - k(\mathbf{x}_*, \mathbf{X}; \theta_k)^T K^{-1} k(\mathbf{x}_*, \mathbf{X}; \theta_k) \end{aligned} \tag{13}$$

Therein, K denotes the covariance matrix. Its entries are defined as $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j; \theta_k)$. The covariances between the test input $\mathbf{x}_*$ and training points $\mathbf{x}_i$ are comprised in the vector $k(\mathbf{x}_*, \mathbf{X}; \theta_k)$ with entries $k_i = k(\mathbf{x}_*, \mathbf{x}_i; \theta_k)$. Note, that the predictive mean of a GP interpolates the training data $\mathscr{D}$ in the absence of measurement noise $\varepsilon = 0$, which is beneficial when predicting the outcome of computer experiments. The predictive variance describes the uncertainty of the function $f(\mathbf{x})$, knowing $\mathscr{D}$. It is high in regions where the population provides few data, compared to regions with higher data density. It can easily be shown, that the prediction for a test input according to a training input yields a zero predictive covariance (Jones *et al.* (1998)).

The mean function of the Gaussian process is often chosen to be zero, i.e. $m(\mathbf{x}) = 0$. This approach is pursued throughout this work. Accordingly, the covariance function $k(\mathbf{x}, \mathbf{x}'; \theta_k)$ is the remaining part to define the surrogate model. For this purpose the squared-exponential covariance function

$$k(\mathbf{x}, \mathbf{x}'; \theta_k) = \exp\left(\frac{-r^2(\mathbf{x}, \mathbf{x}'; \theta_k)}{2} \right) \tag{14}$$

is chosen (Rasmussen (2006)). Automatic relevance determination is achieved by

$$r^2(\mathbf{x}, \mathbf{x}'; \theta_k) = \left| (\mathbf{x} - \mathbf{x}')^T M (\mathbf{x} - \mathbf{x}') \right| \tag{15}$$

with the matrix $M = \mathrm{diag}(\mathbf{l})^2$, where $\mathbf{l}$ is the vector of positive characteristic length-scales for each dimension of the input space. $\mathbf{l}$ states the vector of hyperparameters θ_k of the covariance function. The hyperparameters have to be adapted when fitting the model to the training set $\mathscr{D}$. This process is called inference and is based on hierarchical Bayesian modeling. Note, that non-zero mean functions could also depend on hyperparameters which would have to be fitted.

Inference of unknown quantities is possible by inverse probability. This is based on the Bayes' rule. Without further derivation, details can be found in Rasmussen (2006), the Bayes' rule

$$p(\theta | \mathbf{y}, X) = \frac{p(\mathbf{y} | X, \theta) p(\theta)}{p(\mathbf{y} | X)} \tag{16}$$

on which the inference of the hyperparameters is based is given. It states an expression for the posterior distribution of θ based on $\mathscr{D}$. Inferring knowledge about the hyperparameters is based on the data of the training set. One approach to determine point-estimates of the hyperparameters is to maximize the marginal likelihood with respect to the hyperparameters. This approach is called Maximum Likelihood Estimation (MLE). The marginal likelihood is the first term of the enumerator in the above equation and is analytically tractable. Nevertheless, in a fully Bayesian approach a prior distribution (second term in the enumerator of the above equation) has to be specified for the hyperparameters. Specific values for these are obtained by averaging over the posterior distribution. This can be

realized by a Markov Chain Monte Carlo (MCMC) method (Neal (1997))

$$\mathrm{E}[\theta] = \int \theta p(\theta|\mathbf{y},X)d\theta \approx \frac{1}{N}\sum_{i=1}^{N}\theta_i \tag{17}$$

To approximate the integral sufficiently precise, a large number N of samples $\{\theta_i\}_{i=1}^{N}$ from the posterior distribution $p(\theta|\mathbf{y},X)$ is drawn. To sample from the posterior distribution, the Metropolis algorithm (Metropolis *et al.* (1953)) as implemented in Caldwell *et al.* (2009) is used. Convergence status is monitored as suggested by Gelman & Rubin (1992).

Algorithm Based on GP surrogate models an iterative algorithm for global optimization can be formulated. The algorithm closely follows the Efficient Global Optimization (EGO) algorithm proposed by Jones *et al.* (1998). In a first step, an initial training set $\mathscr{D}_{initial}$ is determined. The input vectors $\mathbf{x}_i$ of $\mathscr{D}_{initial}$ can be obtained by standard Design of Experiment techniques. Full Factorial Design and Latin Hypercube sampling can be mentioned as representatives. Both methods require to split each input dimension in the same number r of regions. For the former the number of required samples can be calculated by r^D, where D denotes the dimension of the input vectors. This is not feasible for expensive target function. Latin Hypercube sampling only requires r samples in total. Thus, in this work Distributed Hypercube Sampling (Manteufel (2001); Beachkofski & Grandhi (2002)), a variant of the Latin Hypercube Sampling (McKay *et al.* (2000)), is used.

In a second step, a Gaussian Process is fitted to the initial training set. This can be realized with the inference techniques presented above.

Subsequently, the iterative part of the algorithm is executed. It aims to iteratively improve the quality of the surrogate model at places where a global minimum is suspected by evaluating samples in those regions. To identify such interesting regions, a quantity which describes a trade-off between exploiting (sampling where the minimum of the target function is expected) and improving (sampling where approximation error may be high) the approximation is required (Jones *et al.* (1998)). The expected improvement (EI) unifies those needs (Kleijnen *et al.* (2012); Jones *et al.* (1998); Jones (2001)). The EI is defined as

$$\begin{aligned}\mathrm{EI}[\mathbf{x}] = (f_{min} - \mathrm{E}(f(\mathbf{x})))\,\Phi\left(\frac{(f_{min} - \mathrm{E}(f(\mathbf{x})))}{\sqrt{\mathrm{Cov}(f(\mathbf{x}))}}\right) \\ + \sqrt{\mathrm{Cov}(f(\mathbf{x}))}\,\phi\left(\frac{(f_{min} - \mathrm{E}(f(\mathbf{x})))}{\sqrt{\mathrm{Cov}(f(\mathbf{x}))}}\right)\end{aligned} \tag{18}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the probability density function respectively the cumulative distribution function of the standard normal distribution. f_{min} denotes the minimum of the posterior predictive mean of the GP and thus of the surrogate surface. Accordingly, $\mathrm{E}(f(\mathbf{x}))$ and $\sqrt{\mathrm{Cov}(f(\mathbf{x}))}$ are the posterior predictive mean and standard deviation of the GP at point $\mathbf{x}$. It can be identified, that the first term in equation (18) accounts for exploiting the surrogate surface, whereas the second term accounts for improving the approximation. Maximizing the EI delivers the next sample for which the target function is evaluated. The data set is extended with the new sample and its corresponding target function value. Based on the new data the Gaussian process is refitted and the next iteration is performed. In general, it is not necessary to infer the hyperparameters for each iteration. Instead, the hyperparameters inferred from the initial population can be used.

Table 1. Supposed location $(C_{q,min}, q_{min})$ of the minimum of the target function $z(\mathbf{x})$ after the i-th iteration of the optimization algorithm.

i	$C_{q,min}$	q_{min}	i	$C_{q,min}$	q_{min}	i	$C_{q,min}$	q_{min}
1	6563.1	17	2	5583.3	18	3	5579.3	18
4	5957.1	17	5	3808.2	9	6	3996.1	9
7	1992.1	7	8	3272.2	5	9	7.7435e-05	1
10	3693.9	8	11	3732.7	9	12	3732.8	9
13	1923.3	4	14	3155.5	12	15	3197.8	11
16	1912.1	4	17	1928.7	4	18	1925.5	4
19	1936.1	4	20	1941.3	4	21	1951.1	4
22	1950.3	4	23	1959.2	4	24	1960.2	4
25	2043	4	26	7355.4	8	27	3570.1	10
28	3564.2	10	29	3548.1	10	30	3532	10
31	3538	10	32	3538.3	10	33	3542.6	10
34	3542.4	10	35	9445.4	16	36	9403.9	16
37	2431.5	5	38	2416.3	5	39	9007.5	15
40	9002.5	15	41	9021.4	15	42	6174.2	13
43	8686.9	15	44	8686.2	15	45	8718	15
46	8755.7	15	47	8839.9	15	48	8838	15
49	8362.3	15	50	8355.8	15			

During one iteration step of the overall algorithm, two respectively three optimization problems have to be solved. Those are the maximization of the log marginal likelihood in the case of a MLE estimate for the hyperparameters, the minimization of the predictive mean of the Gaussian process, and the maximization of the EI. Since analytical expressions are available for those target functions, and they are cheap to evaluate, standard optimization techniques can be used. To infer θ the inverse of the covariance matrix has to be calculated for different θ. This is realized with a Cholesky decomposition. The computational effort increases marginally with the size of $\mathscr{D}$.

In the following, the algorithm is applied to determine an optimal set of the WENO-CU6-M1 parameters.

RESULTS

The optimization problem is formulated as introduced in the preceding. $\mathscr{D}_{initial}$ consists of $n = 20$ samples. The values of the hyperparameters $\theta_k = (l_{C_q}, l_q)$, which correspond to the characteristic length-scales in C_q respectively q direction, are determined by MCMC for $\mathscr{D}_{initial}$. An uniform prior has been chosen for both hyperparameters. They are reused for the subsequent 50 iterations of the optimization algorithm. This yields a total of 70 evaluations of the target function. Figures 1 and 2 depict the histograms of the marginalized posterior distributions of the hyperparameters based on 400000 samples of the converged Markov Chain for each hyperparameter. Mean values $l_{C_q} = 0.9122$ and $l_q = 0.6625$ result.

Figures 3 and 4 show the response surface $f(\mathbf{x})$ of the target function $z(\mathbf{x})$, its contour plot, and the training set $\mathscr{D}$ after completion of 50 iterations. In figure 4, the number besides the dots of $\mathscr{D}$ indicate the corresponding iteration these were acquired. Dots marked with 0 are members of the initial training set $\mathscr{D}_{initial}$. One finds that in the vicinity of a line with the slope $\Delta q/\Delta C_q = 19/10000$, passing through the origin $(C_q, q) = (0,0)$, the surro-

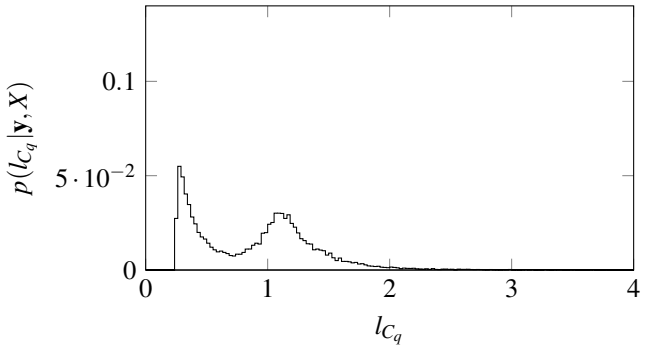

Figure 1. Histogram of the marginalized posterior distribution of the hyperparameter l_{C_q} obtained from MCMC.

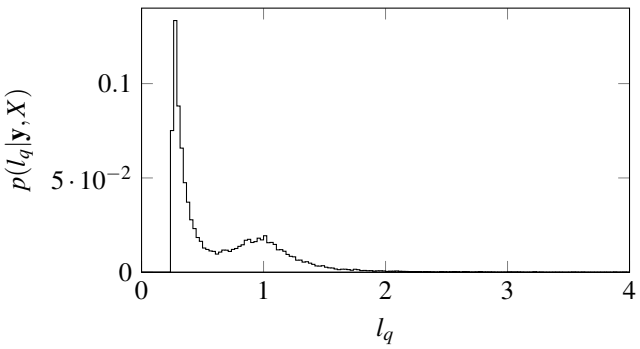

Figure 2. Histogram of the marginalized posterior distribution of the hyperparameter l_q obtained from MCMC.

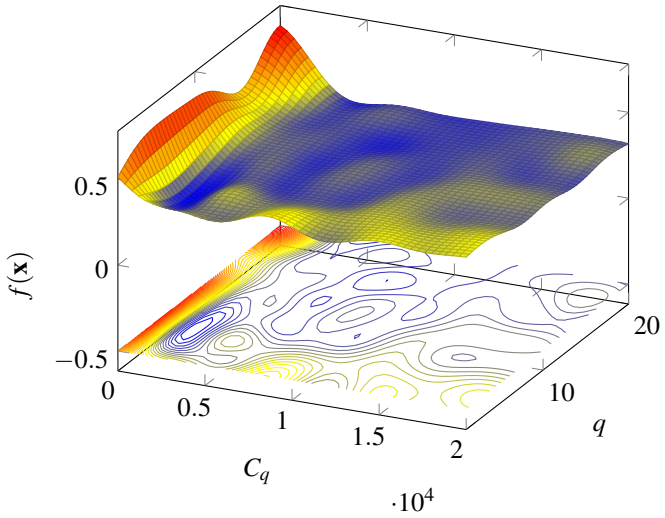

Figure 3. Surface and contour of the surrogate model after 50 iterations.

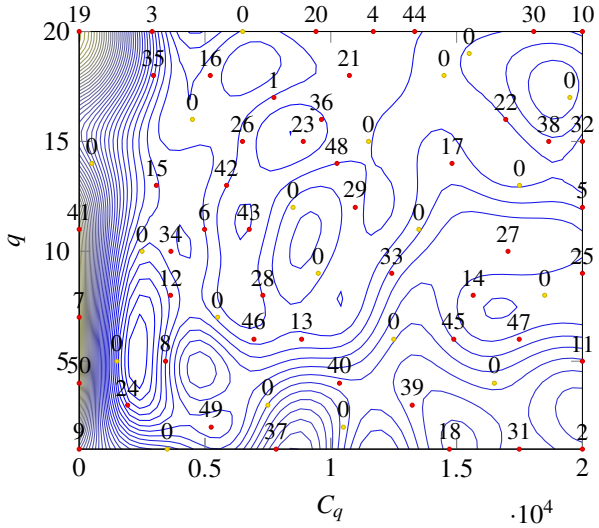

Figure 4. Contour of the surrogate model and input vectors of the training set $\mathscr{D}$ after 50 iterations.

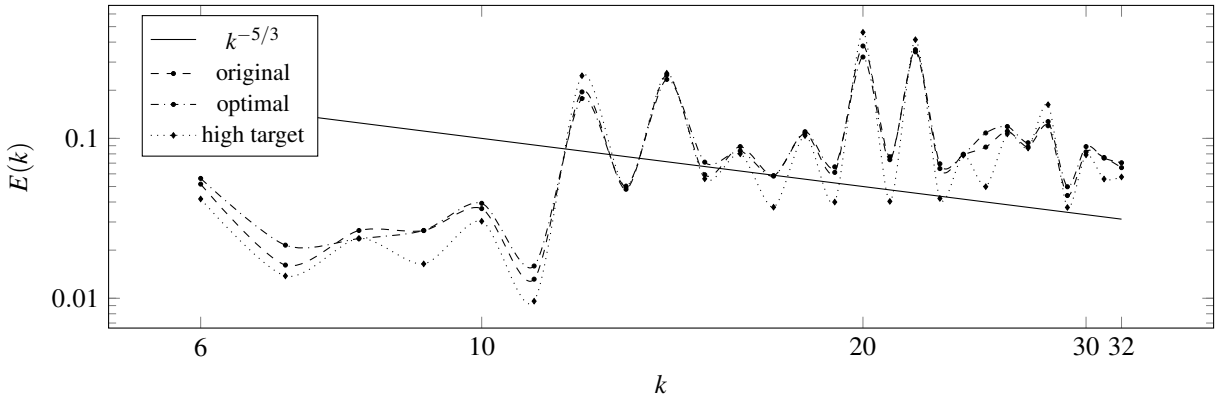

Figure 5. Comparison of $E(k)$ for the original $x_{original} = (1000,4)$, optimal $\mathbf{x}_{opt,c} = (8700,15)$, and high target value $\mathbf{x}_{high\ target} = (400,15)$ model to the theoretical $E \propto k^{-5/3}$ inertial subrange spectrum.

gate model value is minimal. Hence, from the target function, $f_{min} = f(\mathbf{x}_{min}) \sim z_{min} = z(\mathbf{x}_{min})$ is suspected. Indeed, for higher iterations, a region of high sample density develops around the line where f_{min} is suspected. Other samples are acquired where the initial sample density is low, i.e. $C_q > (10000/19)q$. This reflects the trade-off between exploiting and improving the approximation. Table 1 shows the computed location of the minimum after each iteration i. The minimum is suspected to be in the vicinity of $\mathbf{x}_{opt,a} = (1900,4)$, $\mathbf{x}_{opt,b} = (3550,10)$ or $\mathbf{x}_{opt,c} = (8700,15)$. The values of the target function at those points are $z(\mathbf{x}_{opt,a}) = 0.298667$, $z(\mathbf{x}_{opt,b}) = 0.350967$ and $z(\mathbf{x}_{opt,c}) = 0.298038$. The mean target function value of all sampled points is $\bar{z}(\mathbf{x}) = 0.353454$. The global optimum is suspected at $\mathbf{x}_{opt,c}$.

Figure 5 depicts the kinetic energy spectra at $t = 20$ for the WENO-CU6-M1 scheme based on $x_{original} = (1000,4)$ (Hu & Adams (2011)) with $z(x_{original}) = 0.329484$, $\mathbf{x}_{opt,c} = (8700,15)$,

and $\mathbf{x}_{high\ target} = (500, 14)$ with $z(\mathbf{x}_{high\ target}) = 0.460451$ where a high target value evaluates. They are compared to the theoretical inertial subrange spectrum $E(k) \propto k^{-5/3}$. Numerical tests suggest $n = 6$ and $m = 32$ as the first and last wave number of the inertial subrange spectrum. It can be detected, that oscillations are significantly lower for the optimal model.

CONCLUSION

The works Schranner *et al.* (2013) and Schranner *et al.* (2016) identified that a WENO-CU6-M1-based scheme can be shaped such that the energy cascade outscatter and backscatter is physically consistently, and that design optimization is applicable straightforwardly. In consequence an implicit subgrid-scale model has been formulated and the optimal scheme has demonstrated to be robust for a wide range of turbulent and non-turbulent flows.

In this work, an efficient iterative Bayesian optimization algorithm has been proposed and employed successfully to optimize the free parameters, the integer power exponent q and the linear weight bias C_q of the WENO-CU6-M1 scheme with regard to two requirements. Convergence of the algorithm for the highly nonlinear target function requires only half of the number of samples as the thin-plate RBF surrogate model optimization approach proposed in Schranner *et al.* (2016).

Generally, the design optimization algorithm identifies regions in the input domain in which the global minimum is suspected. Due to the fact that it is not ensured that the optimal parameter set is also valid for other flow configurations, identifying regions where the minimum is assumed can be used when optimizing the parameters to other flow configurations.

REFERENCES

Adams, NA, Hickel, S & Franz, S 2004 Implicit subgrid-scale modeling by adaptive deconvolution. *Journal of Computational Physics* **200** (2), 412–431.

Balsara, Dinshaw S & Shu, Chi-Wang 2000 Monotonicity preserving weighted essentially non-oscillatory schemes with increasingly high order of accuracy. *Journal of Computational Physics* **160** (2), 405–452.

Beachkofski, Brian K & Grandhi, Ramana V 2002 Improved distributed hypercube sampling. In *43 rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Denver, CO*.

Borges, Rafael, Carmona, Monique, Costa, Bruno & Don, Wai Sun 2008 An improved weighted essentially non-oscillatory scheme for hyperbolic conservation laws. *Journal of Computational Physics* **227** (6), 3191–3211.

Caldwell, Allen, Kollár, Daniel & Kröninger, Kevin 2009 {BAT} – the bayesian analysis toolkit. *Computer Physics Communications* **180** (11), 2197 – 2209.

Fauconnier, Dieter, De Langhe, Chris & Dick, Erik 2009 Construction of explicit and implicit dynamic finite difference schemes and application to the large-eddy simulation of the taylor–green vortex. *Journal of Computational Physics* **228** (21), 8053–8084.

Gelman, Andrew & Rubin, Donald B 1992 Inference from iterative simulation using multiple sequences. *Statistical science* pp. 457–472.

Grinstein, Fernando F, Margolin, Len G & Rider, William J 2007 *Implicit large eddy simulation: computing turbulent fluid dynamics*. Cambridge university press.

Henrick, Andrew K, Aslam, Tariq D & Powers, Joseph M 2005 Mapped weighted essentially non-oscillatory schemes: achieving optimal order near critical points. *Journal of Computational Physics* **207** (2), 542–567.

Hickel, Stefan, Adams, Nikolaus A & Domaradzki, J Andrzej 2006 An adaptive local deconvolution method for implicit les. *Journal of Computational Physics* **213** (1), 413–436.

Hu, XY & Adams, Nikolaus A 2011 Scale separation for implicit large eddy simulation. *Journal of Computational Physics* **230** (19), 7240–7249.

Hu, XY, Wang, Q & Adams, Nikolaus Andreas 2010 An adaptive central-upwind weighted essentially non-oscillatory scheme. *Journal of Computational Physics* **229** (23), 8952–8965.

Jiang, Guang-Shan & Shu, Chi-Wang 1996 Efficient implementation of weighted eno schemes. *Journal of computational physics* **126** (1), 202–228.

Jones, Donald R 2001 A taxonomy of global optimization methods based on response surfaces. *Journal of global optimization* **21** (4), 345–383.

Jones, Donald R, Schonlau, Matthias & Welch, William J 1998 Efficient global optimization of expensive black-box functions. *Journal of Global optimization* **13** (4), 455–492.

Kleijnen, Jack PC, van Beers, Wim & Van Nieuwenhuyse, Inneke 2012 Expected improvement in efficient global optimization through bootstrapped kriging. *Journal of global optimization* **54** (1), 59–73.

Manteufel, Randall D 2001 Distributed hypercube sampling algorithm. In *Third AIAA non-deterministic approaches forum paper AIAA-2001-1673, 42nd structures, structural dynamics, and materials conference*.

Margolin, Len G & Rider, William J 2002 A rationale for implicit turbulence modelling. *International Journal for Numerical Methods in Fluids* **39** (9), 821–841.

McKay, Michael D, Beckman, Richard J & Conover, William J 2000 A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **42** (1), 55–61.

Metropolis, Nicholas, Rosenbluth, Arianna W, Rosenbluth, Marshall N, Teller, Augusta H & Teller, Edward 1953 Equation of state calculations by fast computing machines. *The journal of chemical physics* **21** (6), 1087–1092.

Neal, Radford M 1997 Monte carlo implementation of gaussian process models for bayesian regression and classification. *arXiv preprint physics/9701026* .

Rasmussen, Carl Edward 2006 Gaussian processes for machine learning .

Roe, Philip L 1981 Approximate riemann solvers, parameter vectors, and difference schemes. *Journal of computational physics* **43** (2), 357–372.

Schranner, Felix S, Hu, Xiangyu Y & Adams, Nikolaus A 2013 A physically consistent weakly compressible high-resolution approach to underresolved simulations of incompressible flows. *Computers & Fluids* **86**, 109–124.

Schranner, Felix S, Rozov, Vladyslav & Adams, Nikolaus A 2016 Optimization of an implicit large-eddy simulation method for underresolved incompressible flow simulations. *AIAA Journal* pp. 1567–1577.

Shu, Chi-Wang 1998 Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws. In *Advanced numerical approximation of nonlinear hyperbolic equations*, pp. 325–432. Springer.

Shu, Chi-Wang & Osher, Stanley 1988 Efficient implementation of essentially non-oscillatory shock-capturing schemes. *Journal of Computational Physics* **77** (2), 439–471.

Taylor, GI & Green, AE 1937 Mechanism of the production of small eddies from large ones. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **158** (895), 499–521.