

Algorithmic Optimal Control - CO₂ Uptake of the Ocean

Junior Research Group A3

Surrogate-Based Optimization of Climate Model Parameters

Dipl. Phys. Malte Prieß* - mpr@informatik.uni-kiel.de

Prof. Slawomir Koziel** - koziel@ru.is

Prof. Thomas Slawig* - ts@informatik.uni-kiel.de

*

Institute for Computer Science
Cluster The Future Ocean
Christian-Albrechts Universität zu Kiel, Germany

**

Engineering Optimization & Modeling Center
School of Science and Engineering
Reykjavik University, Iceland



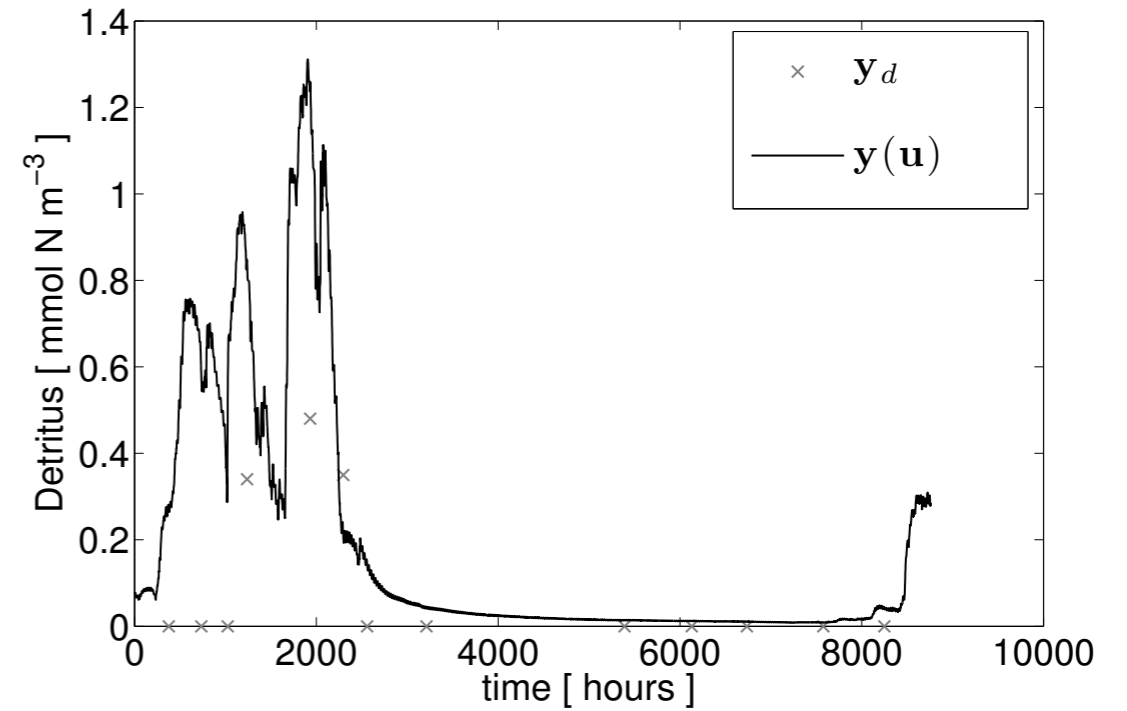
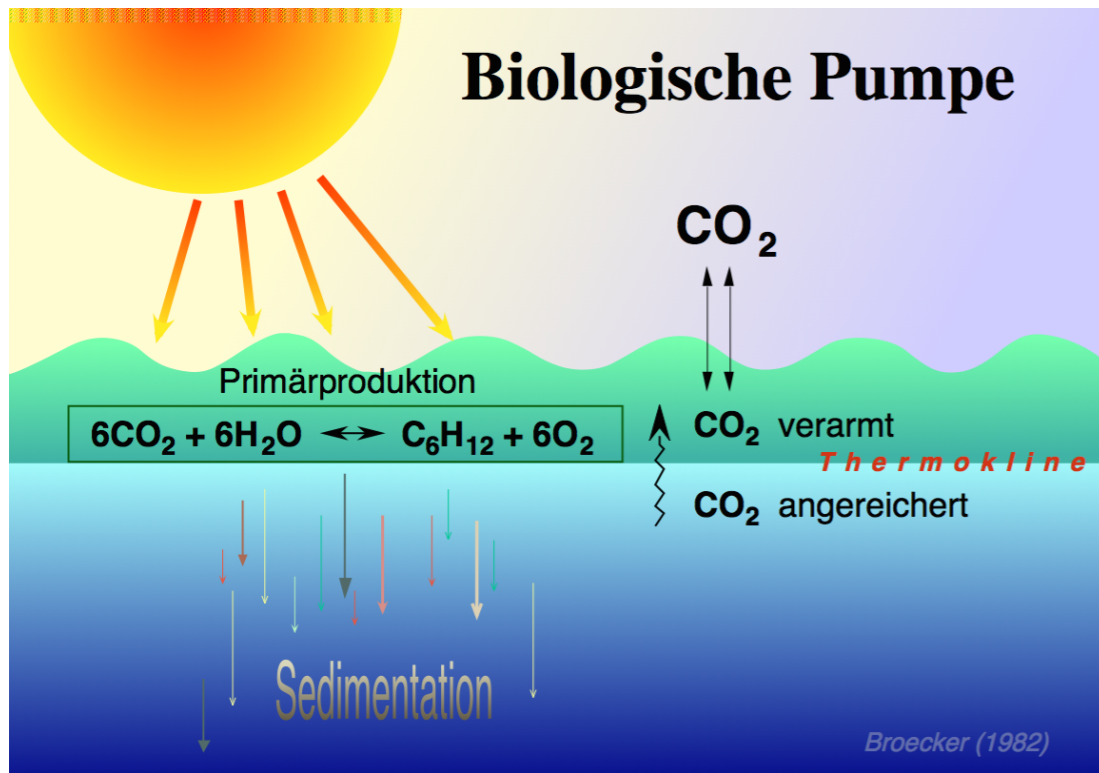
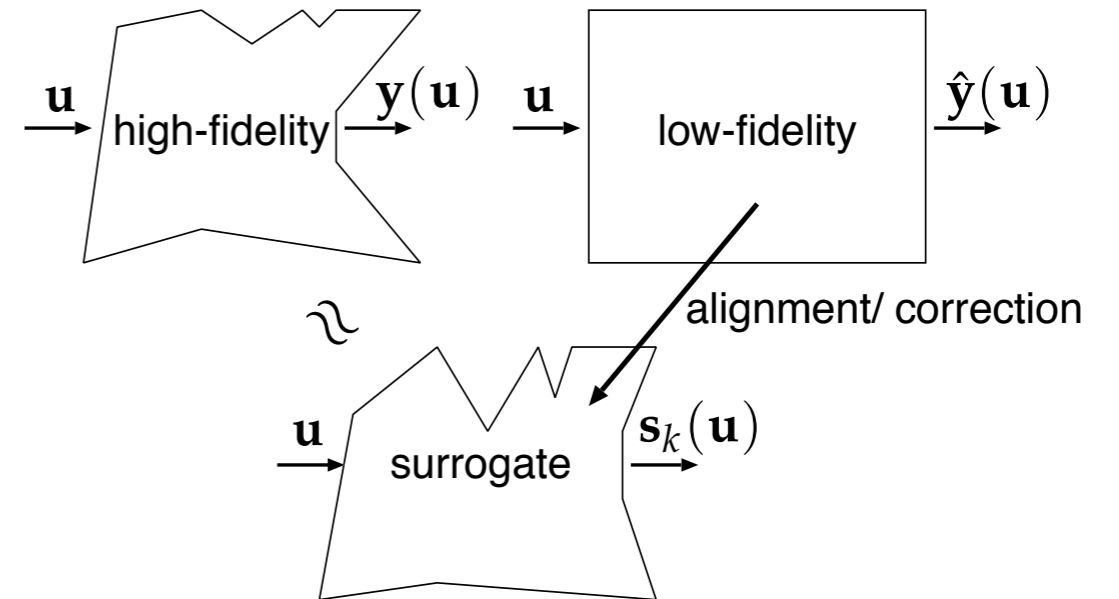
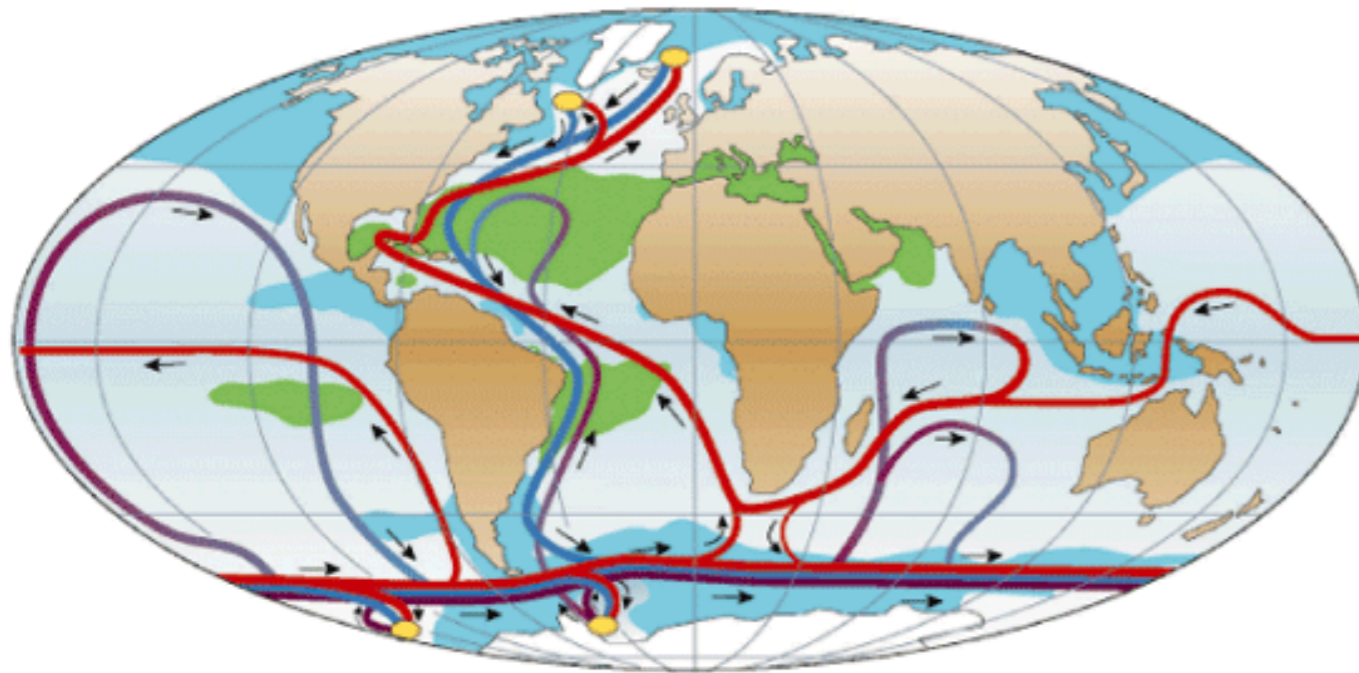


Figure 1: Model output $y^{(D)}$ (detritus) and target data y_d for one year at depth $z \approx -25$ m.



- ▶ Initial boundary value problem (*IBVP*) for a system of time-dependent partial differential or differential algebraic equations (*PDEs/DAEs*) of the following form:

$$E \frac{\partial y}{\partial t} = f(y, u) \quad \text{in } \Omega \times (0, T)$$

$$y(x, 0) = y_{init}(x) \quad \text{in } \Omega$$

$$y(x, t) = y_{bdr}(x, t) \quad \text{on } \partial\Omega \times (0, T)$$

- ▶ *Ocean circulation models (Navier-Stokes equations):*
 - ▶ y may consist for example of the *velocity field, pressure, temperature, salinity*

- ▶ Initial boundary value problem (*IBVP*) for a system of time-dependent partial differential or differential algebraic equations (*PDEs/DAEs*) of the following form:

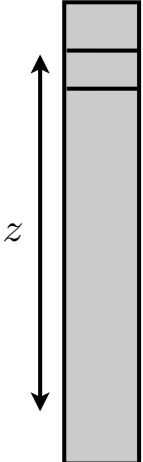
$$\begin{aligned} E \frac{\partial y}{\partial t} &= f(y, u) && \text{in } \Omega \times (0, T) \\ y(x, 0) &= y_{init}(x) && \text{in } \Omega \\ y(x, t) &= y_{bdr}(x, t) && \text{on } \partial\Omega \times (0, T) \end{aligned}$$

- ▶ *Ocean circulation models (Navier-Stokes equations):*
 - ▶ y may consist for example of the *velocity field, pressure, temperature, salinity*
- ▶ *Marine ecosystem model:*
 - ▶ The matrix E can be set to the identity and thus omitted
 - ▶ here, the *rhs* $f(y, u)$ contains
 - (a) the transport (*diffusion, advection*) and nonlinear coupling of so-called *biogeochemical tracers* such as phyto-/ zooplankton etc.
 - (b) the ocean model data: *precalculated („offline“)* or *obtained simultaneously („online“)*

- ▶ Although *one-dimensional*, the following example illustrates the general formulation of this type of models and *actually provides the basis for many marine ecosystem models (also 3D)*

- ▶ Model is of so-called *NPZD type*:

Concentrations of the tracers *dissolved inorganic nitrogen N*, *phytoplankton P*, *zooplankton Z*, and *detritus (i.e., dead material) D* are simulated in a water column, $y = (y^{(l)})_{l=N,P,Z,D}$



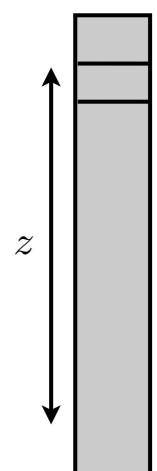
$$\frac{\partial y^{(l)}}{\partial t} = \frac{\partial}{\partial z} \left(\kappa \frac{\partial y^{(l)}}{\partial z} \right) + Q^{(l)}(y, u_2, \dots, u_n), \quad l = N, P, Z$$

$$\frac{\partial y^{(D)}}{\partial t} = \frac{\partial}{\partial z} \left(\kappa \frac{\partial y^{(D)}}{\partial z} \right) + Q^{(D)}(y, u_2, \dots, u_n) - \frac{\partial y^{(D)}}{\partial z} u_1, \quad l = D$$

- ▶ Although *one-dimensional*, the following example illustrates the general formulation of this type of models and *actually provides the basis for many marine ecosystem models (also 3D)*

- ▶ Model is of so-called *NPZD type*:

Concentrations of the tracers *dissolved inorganic nitrogen N*, *phytoplankton P*, *zooplankton Z*, and *detritus (i.e., dead material) D* are simulated in a water column, $y = (y^{(l)})_{l=N,P,Z,D}$



$$\frac{\partial y^{(l)}}{\partial t} = \frac{\partial}{\partial z} \left(\kappa \frac{\partial y^{(l)}}{\partial z} \right) + Q^{(l)}(y, u_2, \dots, u_n), \quad l = N, P, Z$$

$$\frac{\partial y^{(D)}}{\partial t} = \frac{\partial}{\partial z} \left(\kappa \frac{\partial y^{(D)}}{\partial z} \right) + Q^{(D)}(y, u_2, \dots, u_n) - \frac{\partial y^{(D)}}{\partial z} u_1, \quad l = D$$

- ▶ Here: *ocean model data* (the turbulent mixing coefficient $\kappa = \kappa(z,t)$ and temperature) is *precalculated* by one ocean model
- ▶ The terms $Q^{(l)}$ are the *biogeochemical coupling* (or source-minus-sink) terms for the four tracers and $\mathbf{u} = (u_1, \dots, u_n)$ is the *vector of unknown physical and biological parameters*

- ▶ Adjust/identify model parameters \mathbf{u} such that *given measurement data \mathbf{y}_d is matched by the model output $\mathbf{y}(\mathbf{u})$*
- ▶ The mathematical task thus can be classified as a *least-squares type optimization or inverse problem*
- ▶ The opt. process requires a *substantial number of (typically expensive) function evaluations*
- ▶ Methods that aim at *reducing the optimization cost* (e.g. surrogate-based optimization), are *highly desirable*

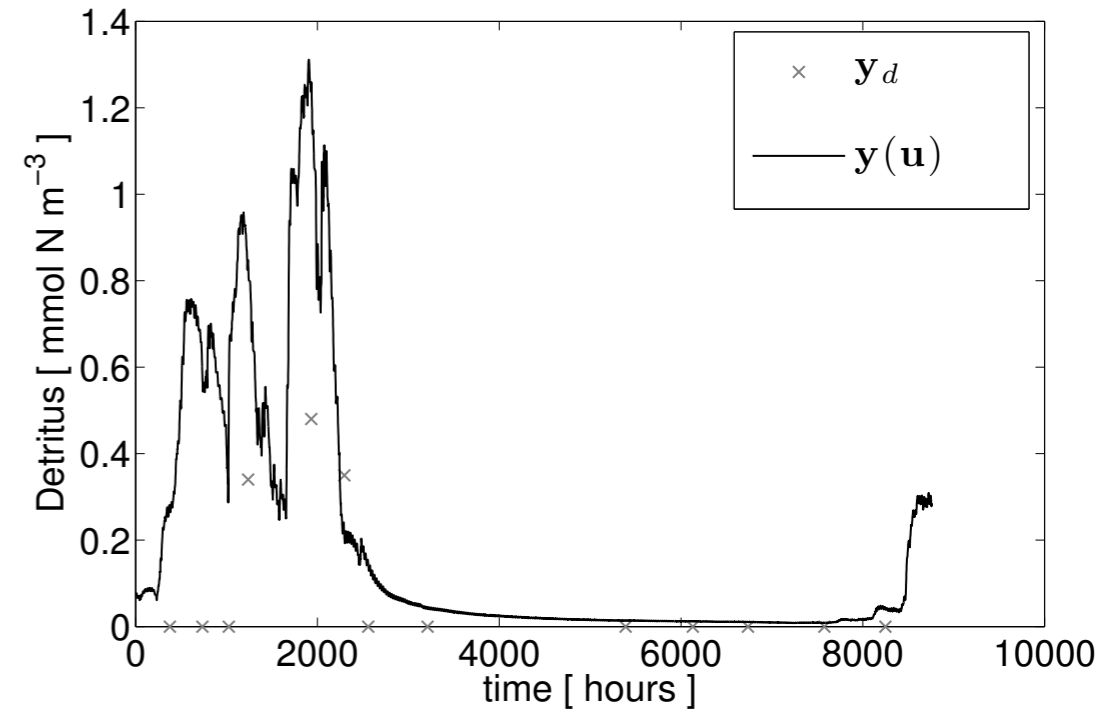


Figure 1: Model output $\mathbf{y}^{(D)}$ (detritus) and target data \mathbf{y}_d for one year at depth $z \approx -25$ m.

$$\min_{\mathbf{u} \in U_{ad}} J(\mathbf{y}(\mathbf{u})) \quad (1)$$

$$J(\mathbf{y}) := \frac{1}{2} \|\mathbf{y} - \mathbf{y}_d\|_Y^2, \quad U_{ad} := \{\mathbf{u} \in \mathbb{R}^n : \mathbf{b}_l \leq \mathbf{u} \leq \mathbf{b}_u\}, \quad \mathbf{b}_l, \mathbf{b}_u \in \mathbb{R}^n, \quad \mathbf{b}_l < \mathbf{b}_u.$$

- ▶ *Nonlinear optimization problem of the form*

$$\min_{\mathbf{u} \in U_{ad}} J(\mathbf{y}(\mathbf{u}))$$

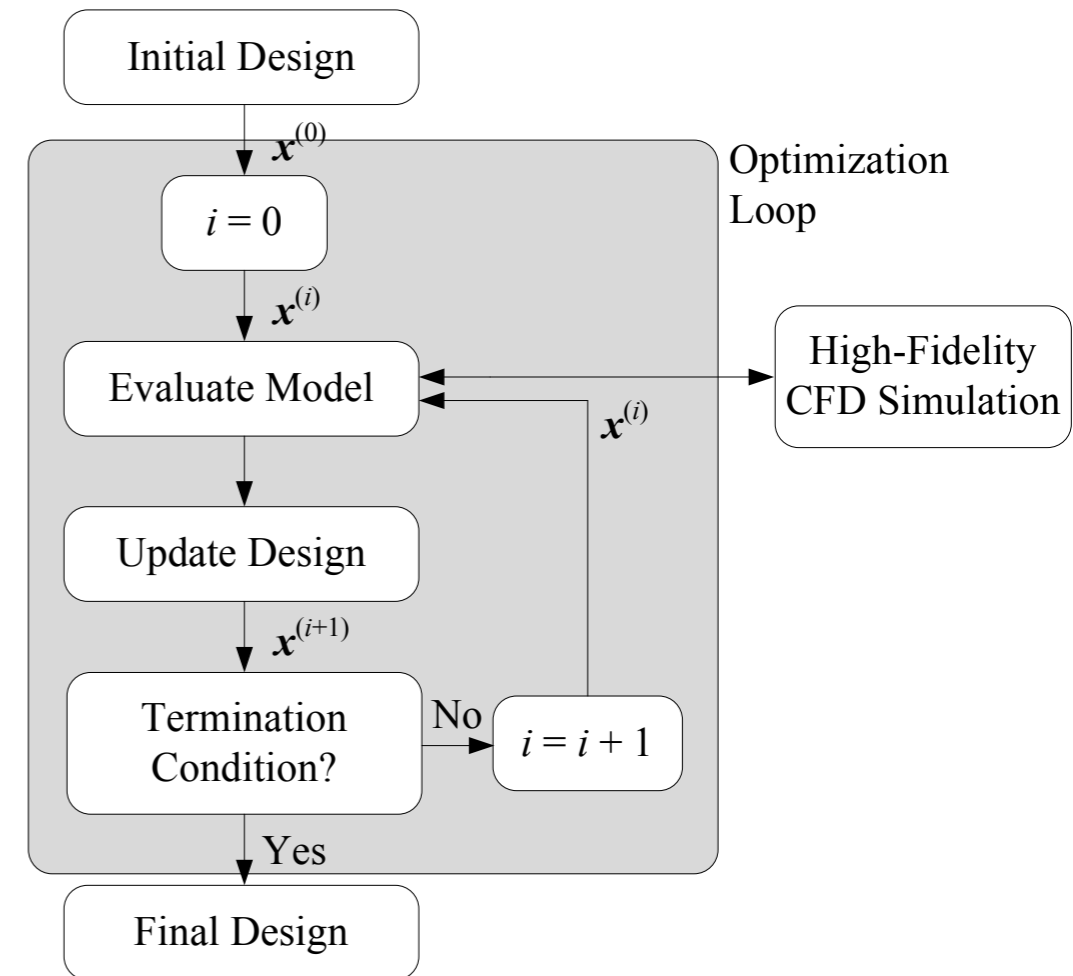
- ▶ *Complex (so-called high-fidelity) models often are computationally very expensive*

- ▶ 1D/2D: 30min to several hours
- ▶ 3D: days, weeks, months

- ▶ Lack of sensitivity information or sensitivity expensive to compute

- ▶ As a consequence, a *direct optimization* approach for such models is *often still beyond the capability of modern numerical algorithms* and computer power

Direct Optimization



Source:
L. Leifsson, S. Koziel, Reykjavik university

- ▶ Idea: exploit a surrogate, a *computationally cheap and yet reasonably accurate representation of the high-fidelity model*

- ▶ The *surrogate replaces the high-fidelity model* in the optimization process

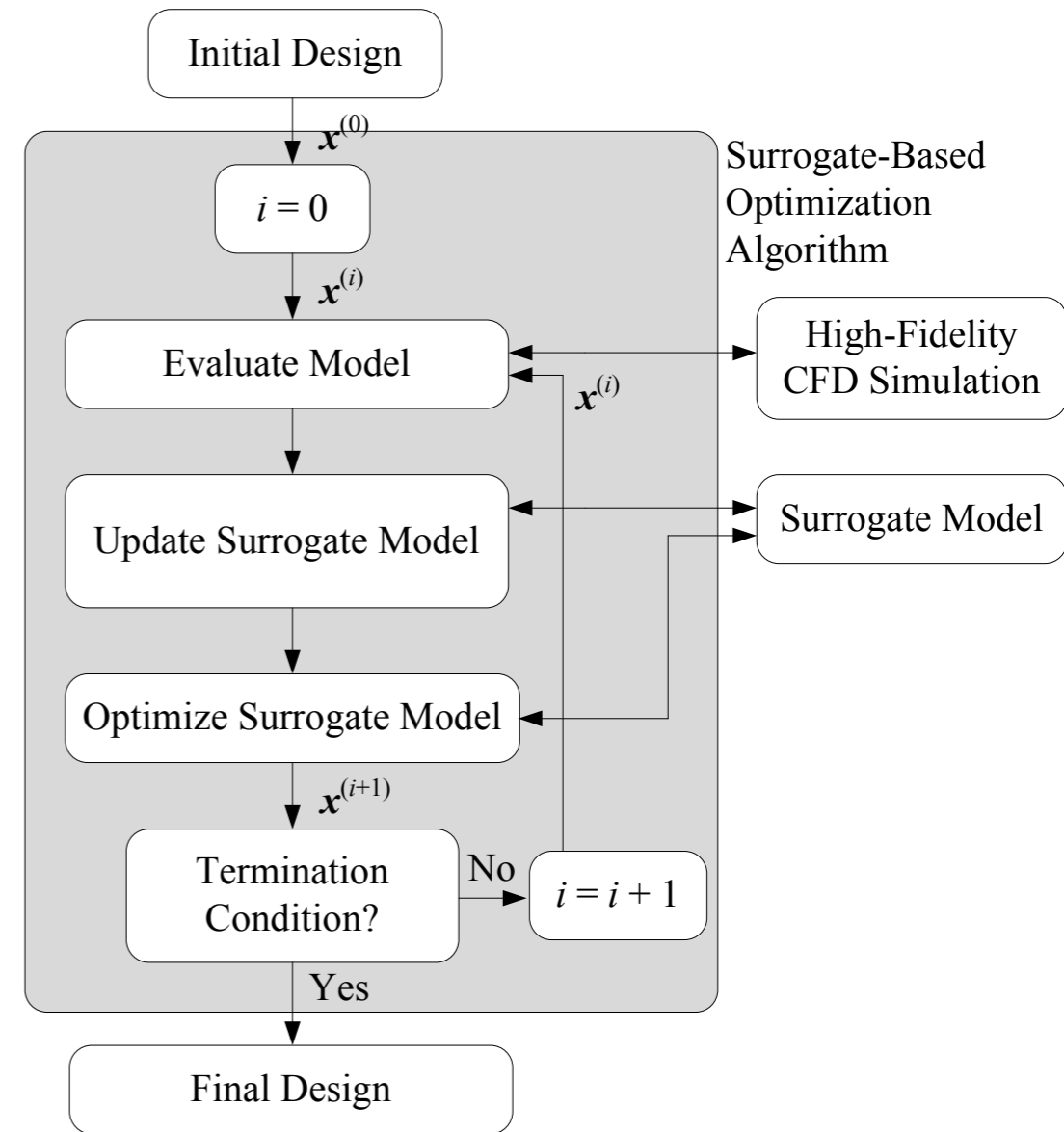
$$\mathbf{u}_{k+1} = \underset{\mathbf{u} \in U_{ad}, \|\mathbf{u} - \mathbf{u}_k\| \leq \delta_k}{\operatorname{argmin}} J(\mathbf{s}_k(\mathbf{u})). \quad (2)$$

- ▶ Also, it is *updated using the high-fidelity model data* accumulated during the process

- ▶ The *scheme (2) is normally iterated* in order to refine the search and to locate the high-fidelity model optimum as precisely as possible

- ▶ ... until some stopping criteria are satisfied (e.g. $\|\mathbf{u}_{k+1} - \mathbf{u}_k\| < \varepsilon$)

Surrogate-Based Optimization



Source:
L. Leifsson, S. Koziel, Reykjavik university

- ▶ *High-fidelity model evaluated only a few times* (preferably only once) per iteration
- ▶ Surrogate model should be *accurate (at least locally), cheap and smooth*
- ▶ Assuming *0- and 1st-order consistency conditions* are satisfied, i.e.,

$$\mathbf{s}_k(\mathbf{u}_k) = \mathbf{y}(\mathbf{u}_k) \quad , \quad \mathbf{s}'_k(\mathbf{u}_k) = \mathbf{y}'(\mathbf{u}_k)$$

- ▶ and provided that the *opt. step is restricted to some trust-region δ_k*
- ⇒ *(2) is provable convergent to at least a local minimum of our original problem (1)*

- ▶ Discretized model equation of our *high-fidelity model* (with state variable \mathbf{y}):

$$\underbrace{[I - \tau A_j^{\text{diff}}]}_{:=B_j^{\text{diff}}} \mathbf{y}_{j+1} = \underbrace{[I + \tau A^{\text{sink}}]}_{:=B^{\text{sink}}} B_j^Q \circ B_j^Q \circ B_j^Q \circ B_j^Q (\mathbf{y}_j),$$

$$B_j^Q(\mathbf{y}_j) := \left[\mathbf{y}_j + \frac{\tau}{4} Q_j(\mathbf{y}_j) \right] \quad \mathbf{y}_j = (y_{ji})_{i=1, \dots, I}, \quad j = 1, \dots, M$$

(M = # of discrete temporal points of the fine model, I = # of discrete spatial points)

- ▶ In the original discrete model (high-fidelity model) the time step τ is chosen as one hour
- ▶ The *low-fidelity model* (with state variable $\hat{\mathbf{y}}$) is obtained by using a *coarser time discretization* with

$$\hat{\tau} = \beta \tau$$

(with a *coarsening factor* $\beta \in \mathbb{N} \setminus \{0, 1\}$, while keeping the spatial discretization fixed)

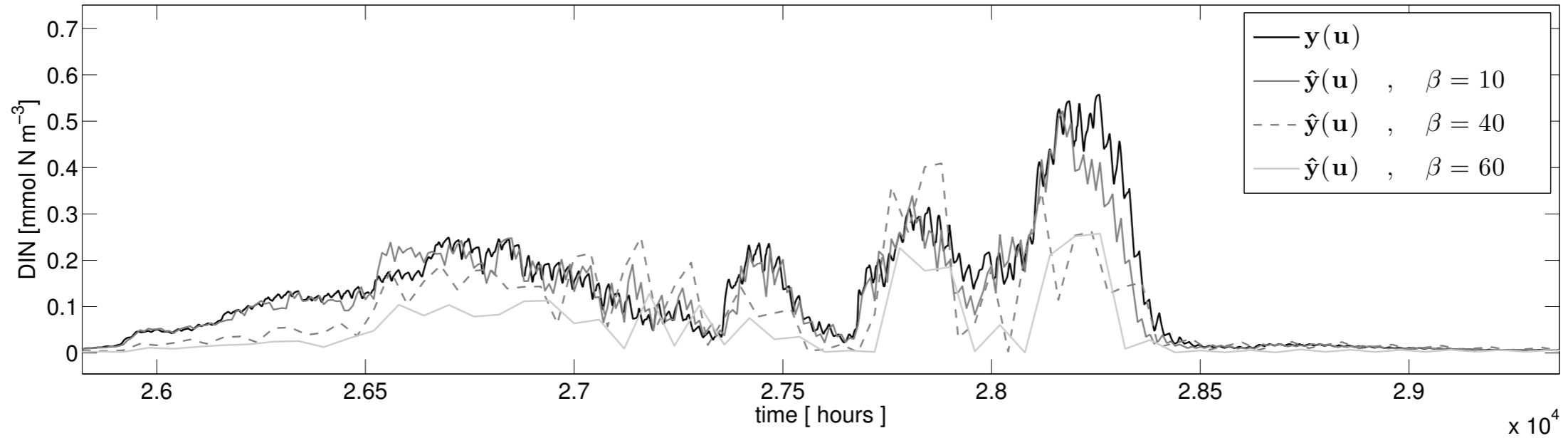


Figure 2: High- and low-fidelity model output y , \hat{y} , respectively, for the state dissolved inorganic nitrogen at depth $z \approx -2.68$ m for different values of the coarsening factor β and the same randomly chosen parameter vector u .

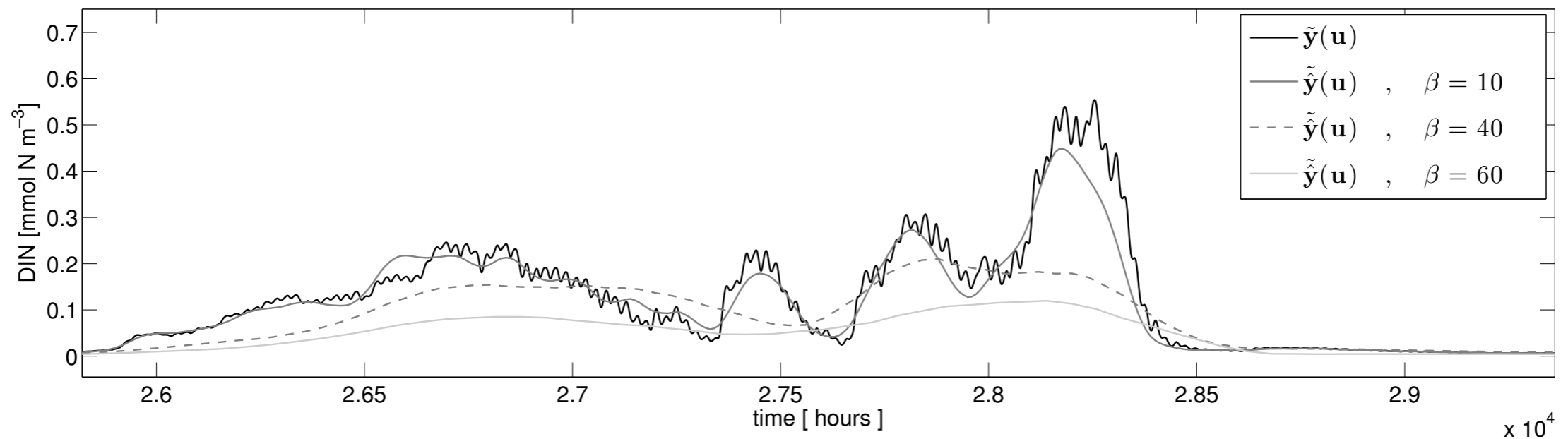


Figure 3: Same as in Figure 2 but now using smoothing for both the coarse and the fine model. Smoothing helps removing the numerical noise in the model outputs so that the optimization process is able to identify and track relevant changes of the traces of interest.

- ▶ *Elemental (multiplicative) response correction of (smoothed) coarse model at iteration k*

$$\begin{aligned}
 \mathbf{s}_k(\mathbf{u}) &:= A_k \circ \tilde{\mathbf{y}}(\mathbf{u}), \quad \mathbf{s}_k \in \mathbb{R}^{\hat{M}I} \\
 A_k &:= (A_{kji})_{j,i} \in \mathbb{R}^{\hat{M} \times I} \\
 A_{kji} &:= \frac{\tilde{y}_{ji}^\beta(\mathbf{u}_k)}{\tilde{y}_{ji}(\mathbf{u}_k)} \quad \left. \begin{array}{l} k = 1, 2, \dots \\ j = 1, \dots, \hat{M} \\ i = 1, \dots, I \\ \beta = M / \hat{M} \end{array} \right\} \quad (3)
 \end{aligned}$$

(\hat{M} = # of discrete temporal points of coarse model, M = # of discrete temporal points of fine model, I = # of discrete spatial points, β = grid coarsening factor, k = iteration index, y^β = down-sampled model response, \sim = smoothed model response)

- ▶ By definition, the surrogate satisfies *exact 0-order consistency*, i.e.,

$$\mathbf{s}_k(\mathbf{u}_k) = \tilde{\mathbf{y}}^\beta(\mathbf{u}_k)$$

- ▶ Note: we do not use sensitivity information from the fine model (*1st-order consistency condition cannot be satisfied exactly*)
- ▶ *Nevertheless*: this surrogate model *exhibits quite good generalization capability*

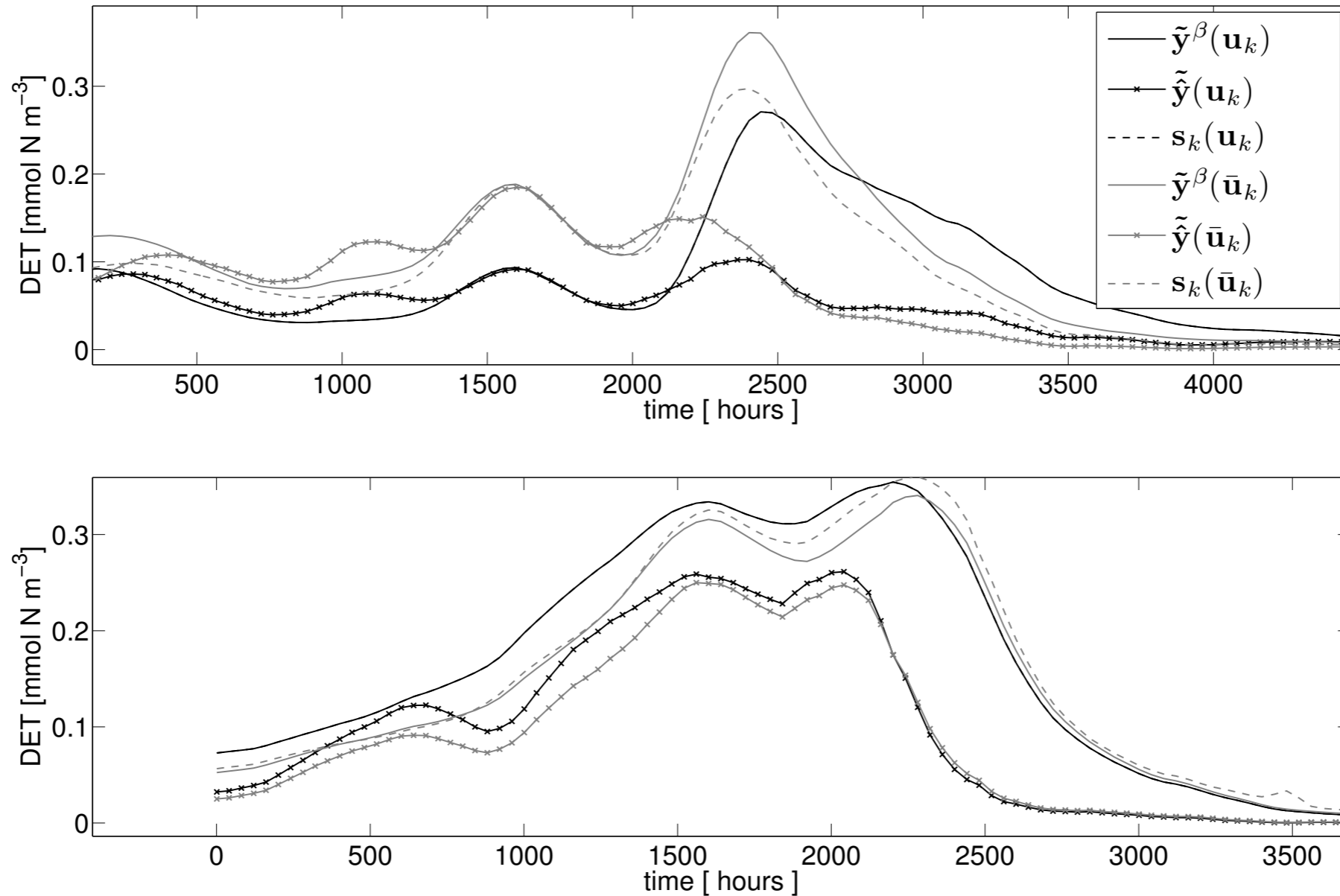
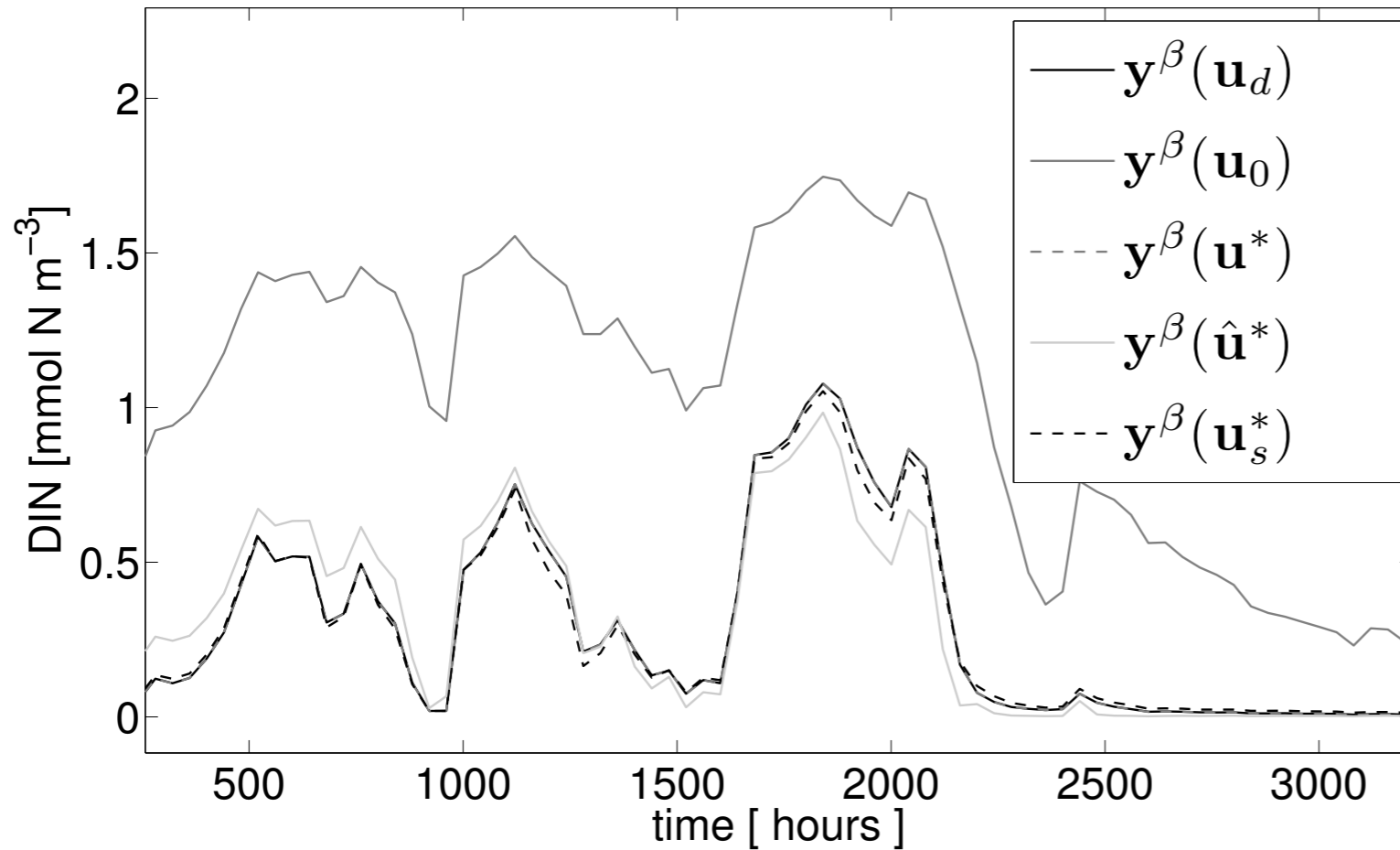


Figure 5: Surrogate's, fine and coarse model output (some time intervall) for the state detritus at depth $z \approx -2.68 \text{ m}$ and at two iterates \mathbf{u}_k and in a neighbourhood $\bar{\mathbf{u}}_k$. The surrogate obviously provides a reasonable approximation of the fine model at the point and in the neighborhood.



iterate	$J(\mathbf{y}^\beta(\mathbf{u}))$	C_i
\mathbf{u}_0	6.609e+04	
\mathbf{u}^*	1.267e-02	983
$\hat{\mathbf{u}}^*$	2.96e+03	11.275
\mathbf{u}_s^*	48.527	59.575
\mathbf{u}_d	~ 84% reduction	

Figure 6:

(left) Fine model output y^β for dissolved inorganic nitrogen at depth $z \approx 2.68$ m. Shown are (from top to bottom):

(i) Synthetic target data, i.e., fine model output at randomly chosen parameters u_d ,

(ii) fine model output at the initial value u_0 ,

(iii) at the result of the direct fine model optimization u^* ,

(iv) at the coarse model optimum \hat{u}^*

(v) and at the result u_s^* of an exemplary SBO run based on the response correction (3)

(right) Values of the cost function J , computational costs C_i (in terms of number of equivalent fine model evaluations)

Cost savings, when using SBO, are about 84% when compared to the direct fine model optimization.

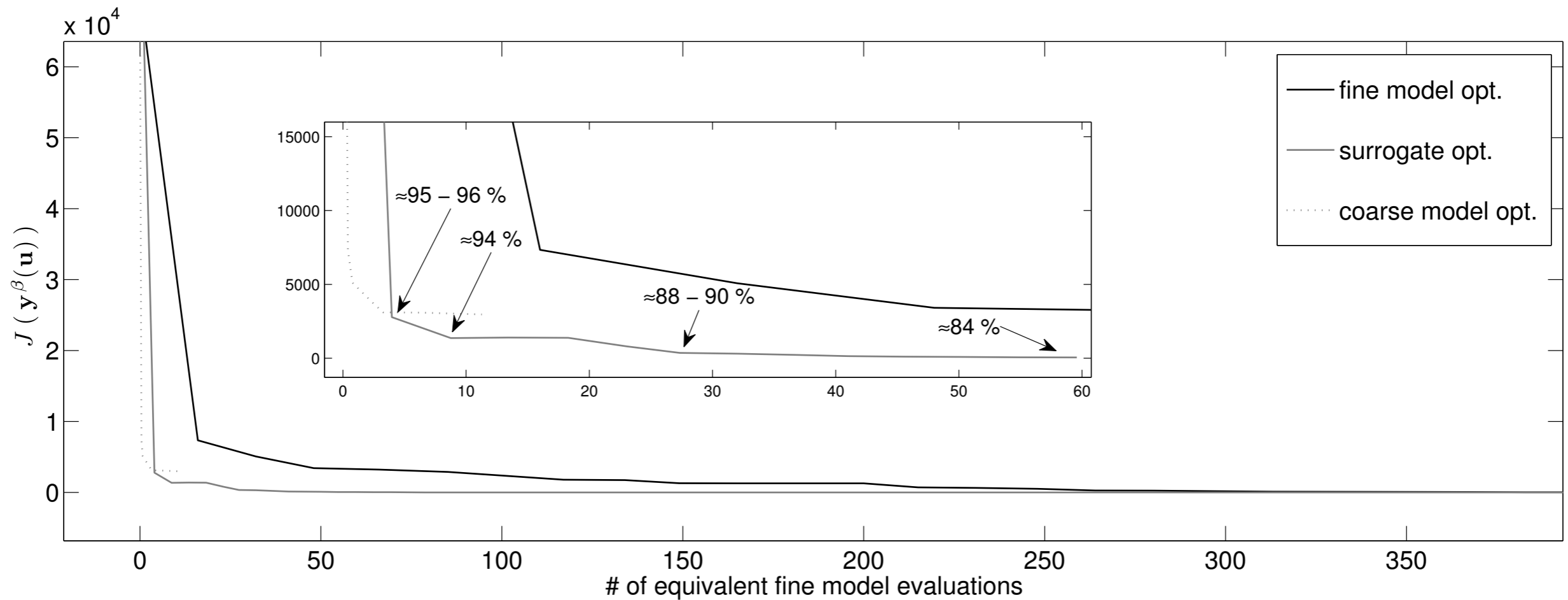


Figure 7: Values of the cost function J versus the equivalent number of fine model evaluations for the fine, coarse and the surrogate-based optimization run. Results of fine model and surrogate optimization given in Figure 6 (left) correspond to the point marked as $\sim 84\%$.

- ▶ *Aggressive Space Mapping* (firstly developed by John W. Bandler et., 1994) is based on:

$$\mathbf{s}_k(\mathbf{u}) := \hat{\mathbf{y}}(\mathbf{p}_k(\mathbf{u})), \quad \mathbf{p}_k(\mathbf{u}) = \mathbf{p}(\mathbf{u}_k) + \mathbf{p}'(\mathbf{u}_k)(\mathbf{u} - \mathbf{u}_k),$$

$$\hat{\mathbf{u}}_k = \mathbf{p}(\mathbf{u}_k) := \operatorname{argmin}_{\mathbf{u} \in U} \|\hat{\mathbf{y}}(\mathbf{u}) - \mathbf{y}(\mathbf{u}_k)\|_Y^2.$$

- ▶ If either the fine model nearly matches the data in an optimum or if *both models* are *similar near their respective optima* we obtain, using (5), so-called *perfect mapping*

$$\mathbf{p}(\mathbf{u}^*) = \operatorname{argmin}_{\mathbf{u} \in U} \|\hat{\mathbf{y}}(\mathbf{u}) - \mathbf{y}(\mathbf{u}^*)\|_Y^2 \approx \operatorname{argmin}_{\mathbf{u} \in U} \|\hat{\mathbf{y}}(\mathbf{u}) - \mathbf{y}_d\|_Y^2 = \hat{\mathbf{u}}^*.$$

- ▶ This motivates to solve for

$$\mathbf{F}(\bar{\mathbf{u}}) := \mathbf{p}(\bar{\mathbf{u}}) - \hat{\mathbf{u}}^* = 0. \quad \hat{\mathbf{u}}^* := \operatorname{argmin}_{\mathbf{u} \in U} J(\hat{\mathbf{y}}(\mathbf{u}))$$

- ▶ Under certain conditions ASM is equivalent to use the surrogate given above in a SBO algorithm

$$\bar{\mathbf{u}}_s = \operatorname{argmin}_{\mathbf{u} \in U} J(\hat{\mathbf{y}}(\mathbf{p}(\mathbf{u})))$$

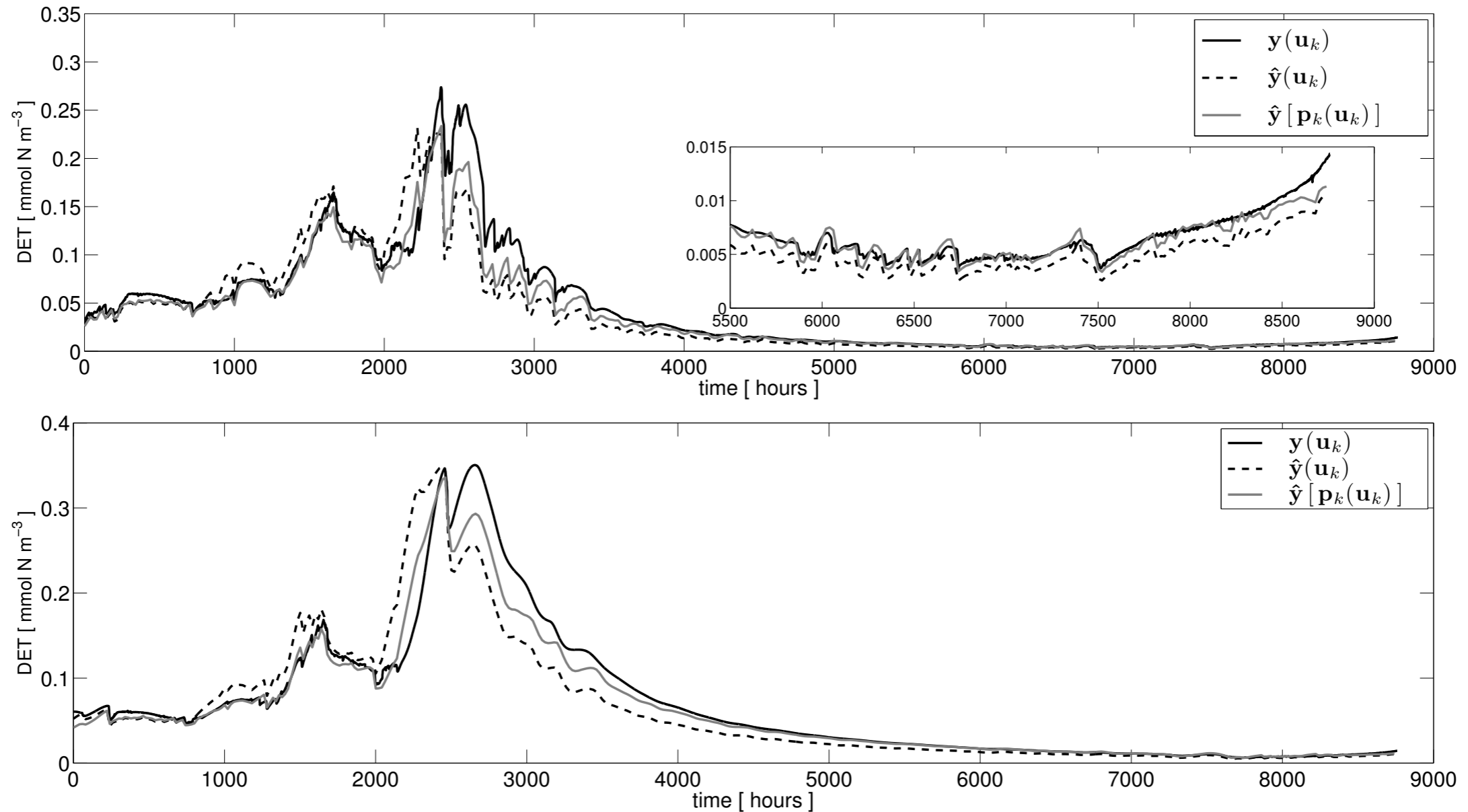
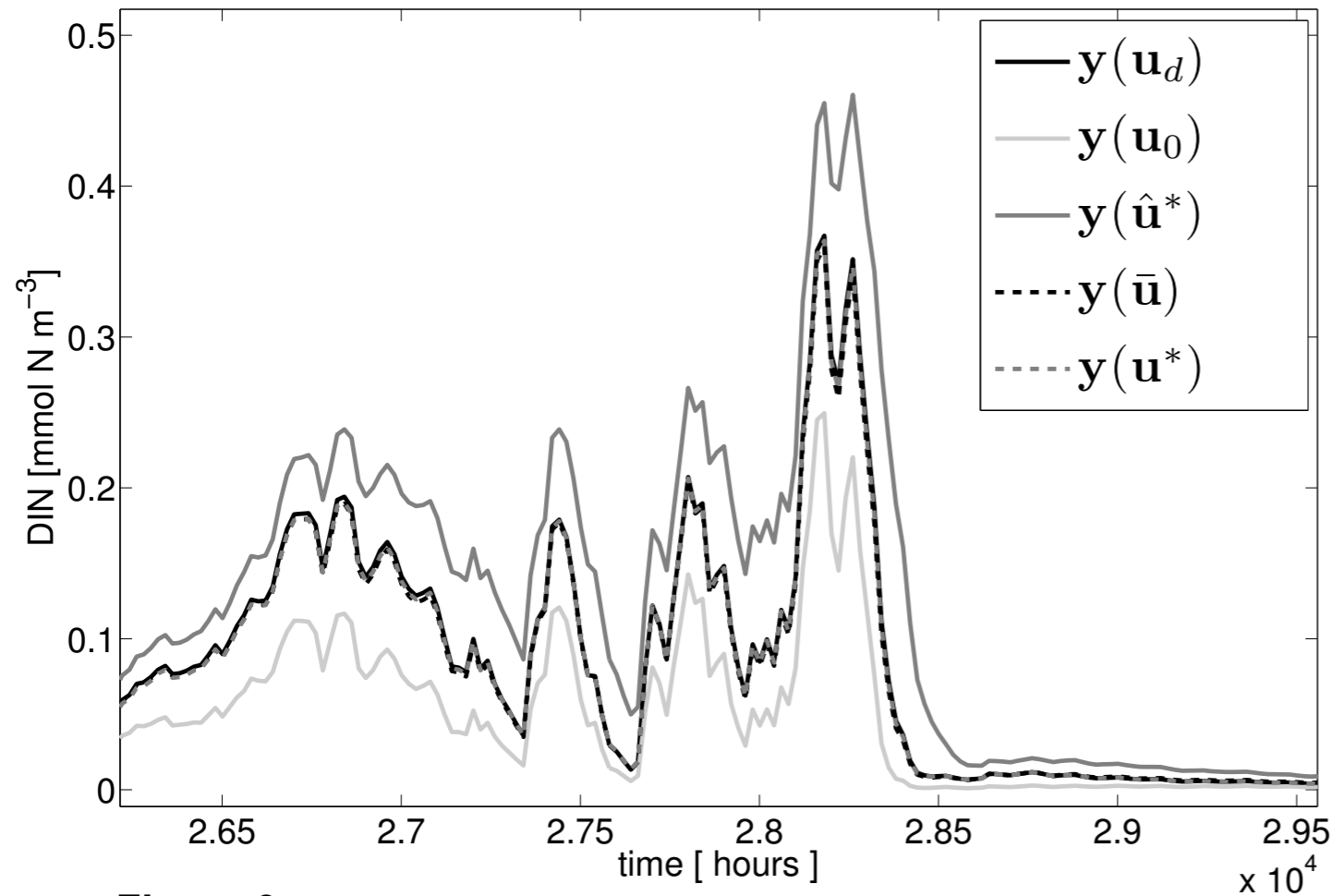


Figure 8: Fine and coarse model output y , \hat{y} as well as the aligned surrogate $s_k(\mathbf{u}_k) = \hat{y}(\mathbf{p}_k(\mathbf{u}_k))$ for the state detritus, at the same randomly chosen parameter vector \mathbf{u}_k , at depths $z \approx 25\text{m}$ (top) and $z \approx 60\text{m}$ (bottom).



	J	C_i
\mathbf{u}_0	5.9e-03	
\mathbf{u}^*	1.6e-05	281
$\hat{\mathbf{u}}^*$	1.8e-03	19.95
$\bar{\mathbf{u}}$	5.0e-05	80.25
\mathbf{u}_d	57.54% reduction	

Figure 9:

(left) Fine model output y^β for dissolved inorganic nitrogen at depth $z \approx 2.68$ m. Shown are (from top to bottom):

- (i) Synthetic target data, i.e., fine model output at randomly chosen parameters u_d ,
- (ii) fine model output at the initial value u_0 ,
- (iii) at the coarse model optimum \hat{u}^*
- (iv) at the result of the ASM algorithm \bar{u}
- (v) at the result of the direct fine model optimization u^* ,

(right) Values of the cost function J , computational costs C_i (in terms of number of equivalent fine model evaluations) Cost savings, when using ASM algorithm, are about 57% when compared to the direct fine model optimization.

- ▶ We presented *two efficient optimization methodologies* for the optimization of climate model parameters
- ▶ We use a *one-dimensional marine ecosystem model* as a representative of this class of models
- ▶ The presented approaches are *based on a coarser discretized low-fidelity model*
 - ▶ *Surrogate-Based Optimization* approach using a *multiplicative response correction*
 - ▶ The *Aggressive Space Mapping (ASM)*
- ▶ Both optimization processes yielded very reasonable solutions at a *cost of a few high-fidelity model evaluations only*
- ▶ *Cost savings are significant*, about *57% (ASM)* and *84% (SBO)* when compared to the high-fidelity model optimization

- ▶ **Prof. Slawomir Koziel** - *Engineering Optimization & Modeling Center, School of Science and Engineering, Reykjavik University* (koziel@ru.is)
- ▶ **Prof. Andreas Oschlies** - *IFM-GEOMAR, Kiel* (aoschlies@ifm-geomar.de)
- ▶ N. V. Queipo, R. T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, and P. K. Tucker, “Surrogate-based analysis and optimization,” *Prog. Aerosp. Sci.*, vol. 41, no. 1, pp. 1 – 28, 2005
- ▶ A. I. Forrester and A. J. Keane, “Recent advances in surrogate-based optimization,” *Prog. Aerosp. Sci.*, vol. 45, no. 1-3, pp. 50 – 79, 2009
- ▶ J. W. B, Q. S. Cheng, S. A. Dakrouy, A. S. Mohamed, M. H. Bakr, K. Madsen, and J. Søndergaard, “Space mapping: The state of the art,” *IEEE T. Microw. Theory.*, vol. 52, no. 1, 2004
- ▶ S. Koziel, J. Bandler, and Q. Cheng, “Robust trust-region space-mapping algorithms for microwave design optimization,” *IEEE T. Microw. Theory.*, vol. 58, pp. 2166 –2174, Aug. 2010.
- ▶ L. Leifsson and S. Koziel, “Multi-fidelity design optimization of transonic airfoils using physics-based surrogate modeling and shape-preserving response prediction,” *Journal of Computational Science*, vol. 1, no. 2, pp. 98 – 106, 2010
- ▶ M. Priess, T. Slawig, and S. Koziel, “Improved surrogate-based optimization of climate model parameters using response correction,” *Int. Conf. Simulation and Modeling Methodologies, Technologies and Appl., SIMULTECH 2011*, Noordwijkerhout, The Netherlands, July 29-31, pp. 449-457, 2011
- ▶ J. Sarmiento and N. Gruber, *Ocean Biogeochemical Dynamics*. Princeton University Press, 2006
- ▶ A. Oschlies and V. Garcon, “An eddy-permitting coupled physical-biological model of the north atlantic. 1. sensitivity to advection numerics and mixed layer physics,” *Global Biogeochem. Cy.*, vol. 13, pp. 135– 160, 1999.

Physically based:

Constructed from physical low-fidelity model (with suitable correction/alignment)

Pro:

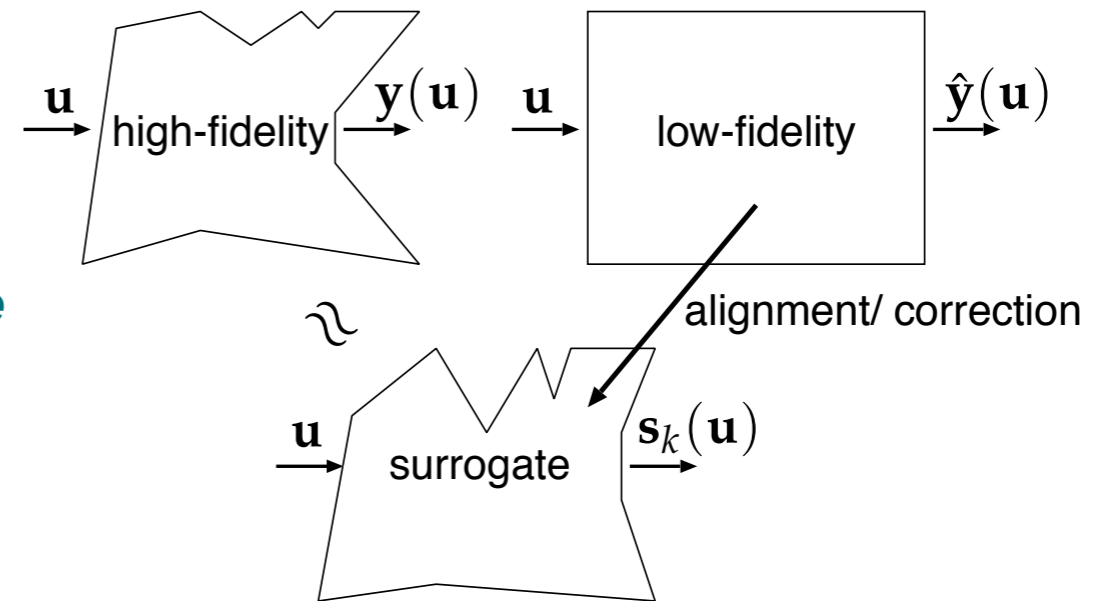
Inherits more characteristics of the system

Contra:

Dedicated (reuse is rare)

Typically more expensive

Low-fidelity model must be available



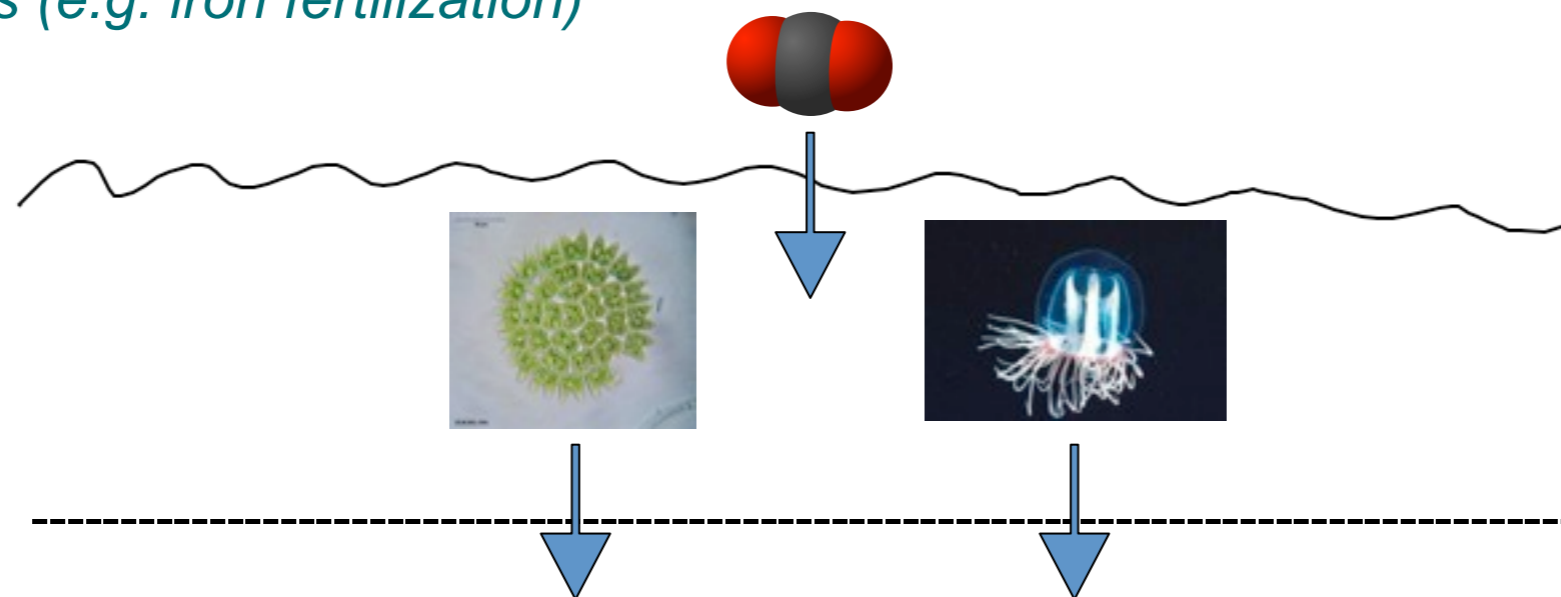
Popular techniques:

Response correction, Space Mapping

How to obtain the low-fidelity model?

- ▶ Using simplified physics (e.g., ignoring second order effects)
- ▶ *Coarse discretization*
- ▶ Using analytical formulas if available

- ▶ Used for example to compute and predict the *oceanic uptake of CO₂* as part of the *global carbon cycle*
- ▶ This uptake is determined by the solution of CO₂ in the water via the ocean surface and *physical and biogeochemical processes* in the water, i.e.
 - ▶ Ocean circulation
(-> *Ocean models*)
 - ▶ Photosynthesis, consumption by zooplankton, sinking of dead material
(-> *Marine ecosystem models*)
- ▶ Simulations based on those models can play a *key tool in CDR (Carbon Dioxide Reduction) approaches (e.g. iron fertilization)*



- ▶ Coarse model response might be close to zero (and maybe even negative due to approximation errors) and a few magnitudes smaller than the fine one
- ▶ This leads to *large (possibly negative) entries in the corresponding correction tensor A_k*
- ▶ Such a correction tensor still ensures zero-order consistency at the point where it was established (i.e., \mathbf{u}_k),
- ▶ But it *may lead to (locally) poor approximation in the vicinity of \mathbf{u}_k*
- ▶ Still, the *overall shape of the surrogate's response provides a reasonable approximation*

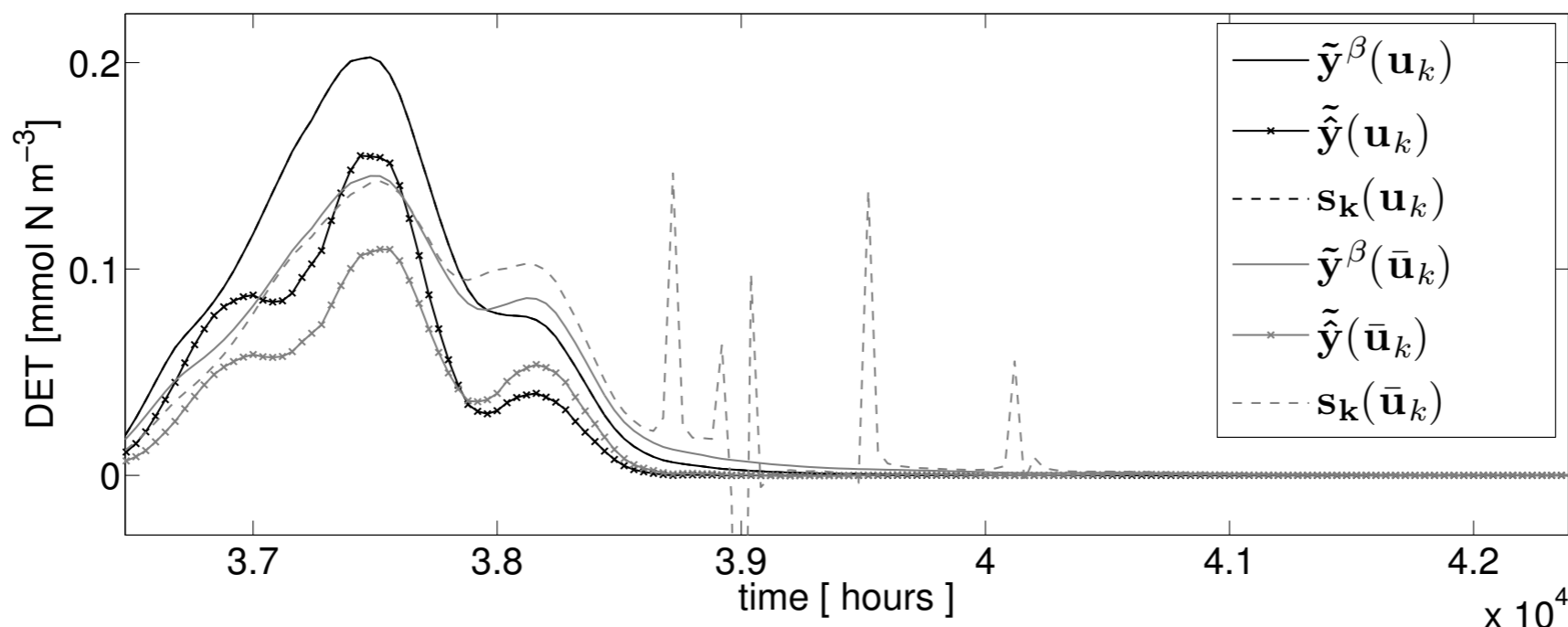


Figure 8: Surrogate's, fine and coarse model responses for the state detritus at depth $z \approx -2.68$ m, at one iterate \mathbf{u}_k and in a vicinity $\bar{\mathbf{u}}_k$.

- ▶ *A few simple means can address these issues* and further improve the accuracy of the surrogate's response as well as the performance of the optimization algorithm

$$\begin{aligned}
 (i) \quad & \hat{y}_{ji}(\mathbf{u}_k) = \max\{ \hat{y}_{ji}(\mathbf{u}_k), 1e-8 \} \\
 (ii) \quad & A_{kji} = \min\{ A_{kji}, 10 \} \\
 (iii) \quad & A_{kji} = 1 \text{ if } \left(\tilde{y}_{ji}^\beta(\mathbf{u}_k) \leq \varepsilon \text{ and } \tilde{y}_{ji}(\mathbf{u}_k) \leq \varepsilon \right)
 \end{aligned}
 \tag{4}$$

- ▶ *Large positive and negative peaks* present in the surrogate responses using the original correction scheme (3) *are removed*

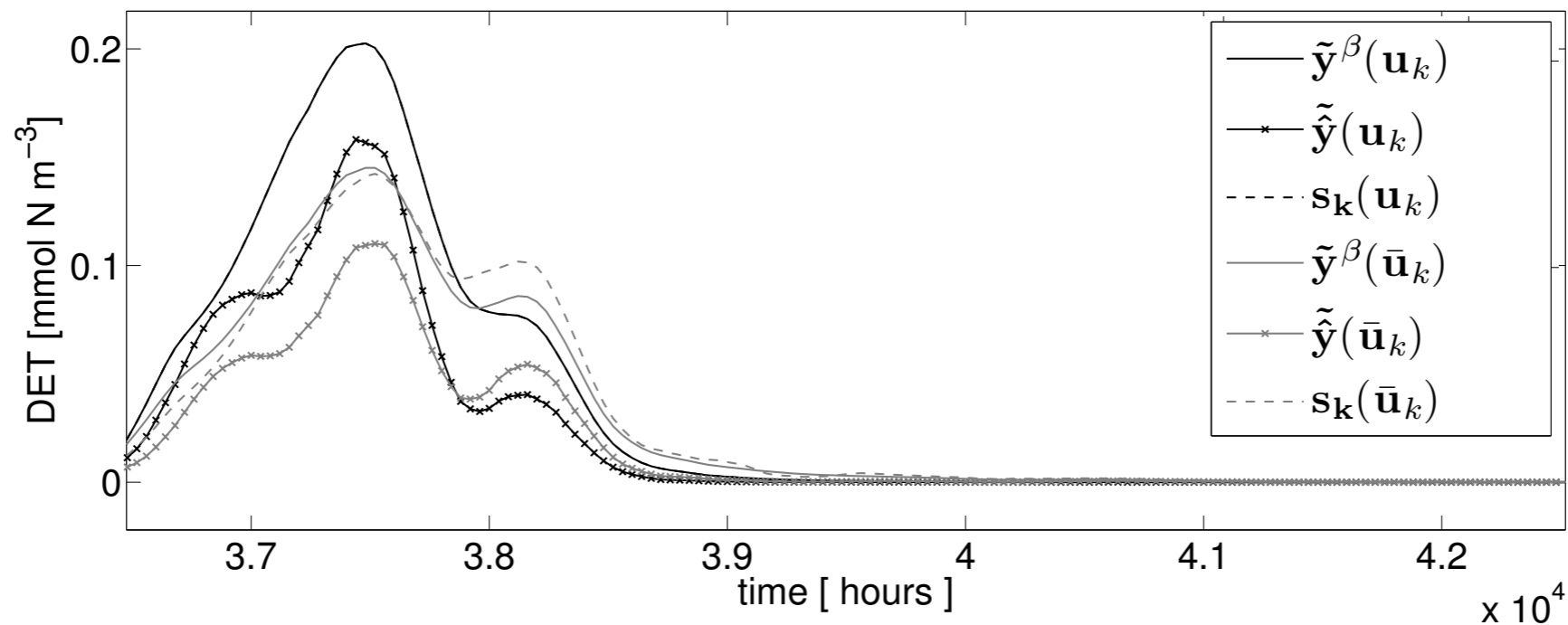


Figure 9: Same model responses as in Figure 8.

- ▶ Using the improved correction scheme allows us to *further improve the computational efficiency of the original SBO scheme*
- ▶ *The optimization cost is reduced three times when compared to the original technique*

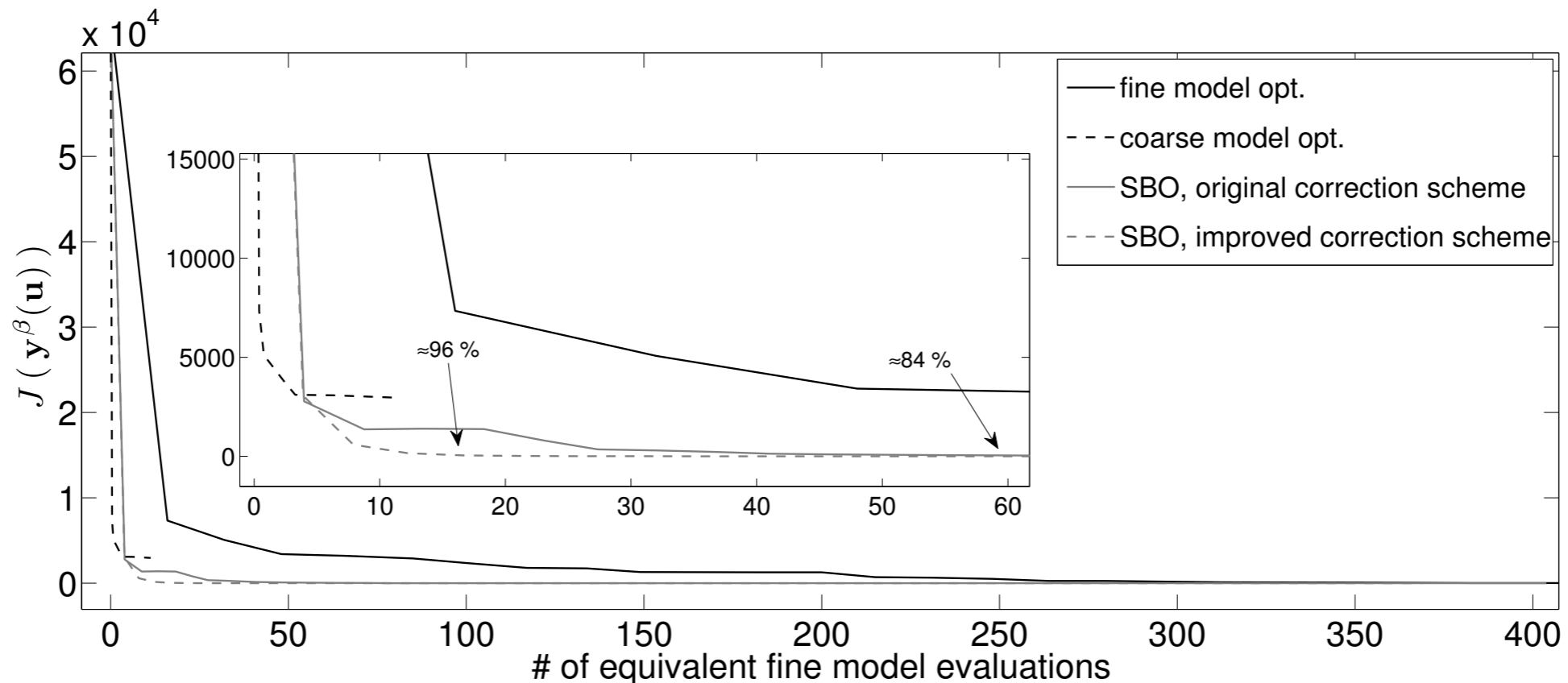


Figure 10: The values of the cost function J versus the equivalent number of fine model evaluations for the same SBO run using the surrogate model exploiting the original and the improved correction scheme, as well as for a fine and coarse model optimization run.

- ▶ Due to *numerical noise* (cf. Figure 2), it is *reasonable to smoothen* the coarse model output
- ▶ It was observed by visual inspection of the model outputs that this procedure allows us to *remove the numerical noise and identify the main characteristics* of the traces of interest
- ▶ For the smoothing we use a *walking average with span $\pm n$* given as:

$$\tilde{y}_{ji} := \frac{1}{2n+1} \sum_{m=j-n}^{j+n} \left[\frac{1}{2n+1} \sum_{p=m-n}^{m+n} \hat{y}_{pi} \right] \quad j = 1, \dots, \hat{M}, \quad i = 1, \dots, I$$

- ▶ It turns out, also by visual inspection, that a value of $n = 3$ and “double” smoothing are suitable for the considered problem

- ▶ It is important to keep in mind that *choosing β too large could lead to a numerically unstable scheme*
- ▶ The condition of stability is dependent on the ratio h / v and the nonlinear coupling term Q (h = spatial step-size, v = here, sinking velocity)

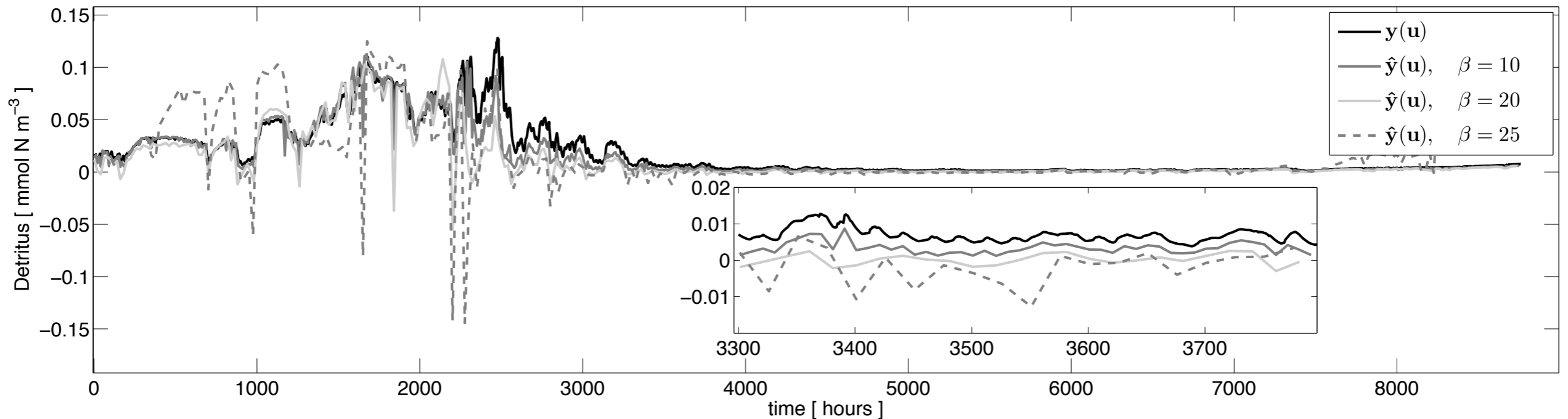


Figure 4: The figure shows one year of the fine model output $y(\mathbf{u})$ and of the coarse model output \hat{y} for the state detritus at depth $z \approx 25$ m for different values of the coarsening factor β and at some fixed parameters \mathbf{u} .