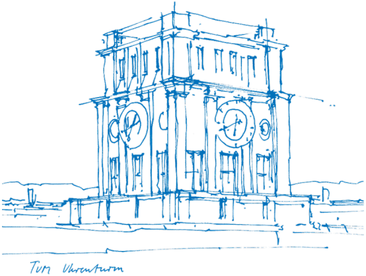
Lehrstuhl Treffen

Ivana Jovanovic Buha
Chair of Scientific Computing in Computer Science
TUM School of Computation, Information and Technology



Outline

Uncertainty Quantification (UQ) and(vs.) Sensitivity Analysis (SA)

UQEF & UQEF-Dynamic

- Two software tools for efficient UQ developed at the chair
- Case study: HBV Hydrology model

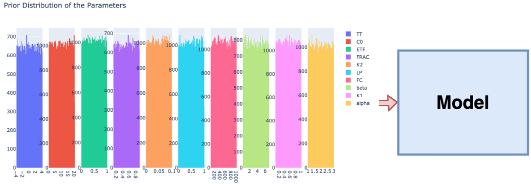
UQ when QoIs are time series

- Research question: Building a gPCE for time-dependent responses?

UQ and Sparse Grids

- Research question: Building an adaptive SG surrogate for time-dependent responses?

UQ and/vs. SA



UQ:

$$\mathbb{E}[\mathcal{O}(f)] := \int_{\Gamma} \mathcal{O}(f)(t, \boldsymbol{\theta})\rho(\boldsymbol{\theta})d\boldsymbol{\theta}$$

$$Var[\mathcal{O}(f)] := \mathbb{E}[\mathcal{O}(f)^2] - (\mathbb{E}[\mathcal{O}(f)])^2$$

Variance-based SA [Sobol 2001]:

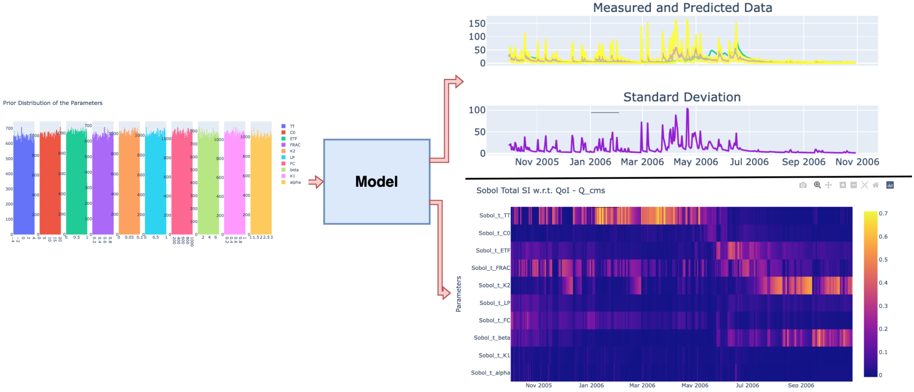
$$S_j := \frac{Var[f] - \mathbb{E}_{\theta_j}[Var_{\boldsymbol{\theta}_{\sim j}}[f|\theta_j]]}{Var[f]}$$

UQ and/vs. SA

UQ:

$$\mathbb{E}[\mathcal{O}(f)] := \int_{\Gamma} \mathcal{O}(f)(t,\theta)\rho(\theta)d\theta$$

$$Var[\mathcal{O}(f)] := \mathbb{E}[\mathcal{O}(f)^2] - (\mathbb{E}[\mathcal{O}(f)])^2$$



Variance-based SA [Sobol 2001]:

$$S_j := \frac{Var[f] - \mathbb{E}_{\theta_j}[Var_{\theta_{\sim j}}[f|\theta_j]]}{Var[f]}$$

UQ & SA: Monte Carlo Methods

Stochastic quadrature algorithm with uniform weights and (pseudo) random node

- Pro: scales well to higher dim.; simple to implement; embarrassingly parallel
- Cons: slow convergence rate $RMSE[\hat{\mathbb{E}}] \in \mathcal{O}(\sqrt{\frac{Var[f]}{N}})$

UQ:

$$\mathbb{E}[\mathcal{O}(f)] := \int_{\Gamma} \mathcal{O}(f)(t,\boldsymbol{\theta})\rho(\boldsymbol{\theta})d\boldsymbol{\theta}$$

$$Var[\mathcal{O}(f)] := \mathbb{E}[\mathcal{O}(f)^2] - (\mathbb{E}[\mathcal{O}(f)])^2$$

UQ - MC:

$$\hat{\mathbb{E}}[f] = \bar{f} = \frac{1}{N}\sum_{k=1}^{N} f_k$$

$$\hat{Var}[f] = \frac{1}{N-1}\sum_{k=1}^{N}(f_k - \bar{f})^2$$

Variance-based SA [[12]]:

$$S_j := \frac{Var[f] - \mathbb{E}_{\theta_j}[Var_{\boldsymbol{\theta}_{\sim j}}[f|\theta_j]]}{Var[f]}$$

SA - MC: [11]

- generate triplet of matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{A}_{\boldsymbol{B}}^{(j)}$
- needed number of model evaluations $N(d+2)$

$$S_j^T = \frac{1}{2N}\sum_{k=1}^{N}(f(\boldsymbol{A})_k - f(\boldsymbol{A}_{\boldsymbol{B}}^{(j)})_k)^2/\hat{Var}[f]$$

- good news: there are new approaches that only require N model runs [5]

UQ & SA: gPCE surrogate

UQ:

$$\mathbb{E}[\mathcal{O}(f)] := \int_{\Gamma} \mathcal{O}(f)(t,\boldsymbol{\theta})\rho(\boldsymbol{\theta})d\boldsymbol{\theta}$$

$$Var[\mathcal{O}(f)] := \mathbb{E}[\mathcal{O}(f)^2] - (\mathbb{E}[\mathcal{O}(f)])^2$$

UQ - gPCE:

$$\mathbb{E}[f_N(t,\boldsymbol{\theta})] = c_0(t)$$

$$Var[f_N(t,\boldsymbol{\theta})] = \sum_{position(\boldsymbol{p})=1}^{N-1} c_{\boldsymbol{p}}^2(t)$$

$$f(t,\boldsymbol{\theta}) \approx f_N(t,\boldsymbol{\theta}) := \sum_{\boldsymbol{p}} c_{\boldsymbol{p}}(t)\boldsymbol{\Phi}_{\boldsymbol{p}}(\boldsymbol{\theta}) := \sum_{\boldsymbol{p}} < f(t,\boldsymbol{\theta}), \boldsymbol{\Phi}_{\boldsymbol{p}}(\boldsymbol{\theta}) >_{\rho(\boldsymbol{\theta})} \boldsymbol{\Phi}_{\boldsymbol{p}}(\boldsymbol{\theta})$$

$$\hat{c}_{\boldsymbol{p}}(t) := \sum_{q=1}^{Q} f(t,\boldsymbol{\theta}_q)\boldsymbol{\Phi}_{\boldsymbol{p}}(\boldsymbol{\theta}_q)\omega$$

SA:

$$S_j := \frac{Var[f] - \mathbb{E}_{\theta_j}[Var_{\boldsymbol{\theta}_{\sim j}}[f|\theta_j]]}{Var[f]}$$

SA - gPCE: [[13]]

$$S_j^T = \frac{\sum_{\boldsymbol{p}\in A_j} c_{\boldsymbol{p}}^2(t)}{Var[f_N(t,\boldsymbol{\theta})]}$$

$$A_j = \{\boldsymbol{p} \in \mathcal{P}_P \wedge \forall k \neq j, \boldsymbol{p}_k = 0\}$$

UQEF & UQEF-Dynamic Software Tools

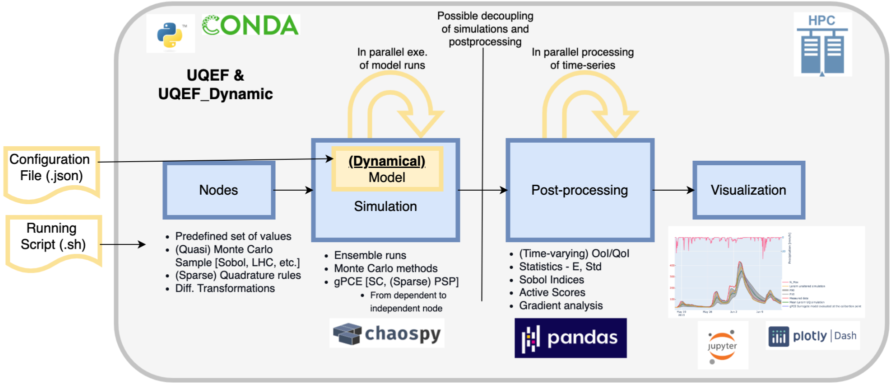
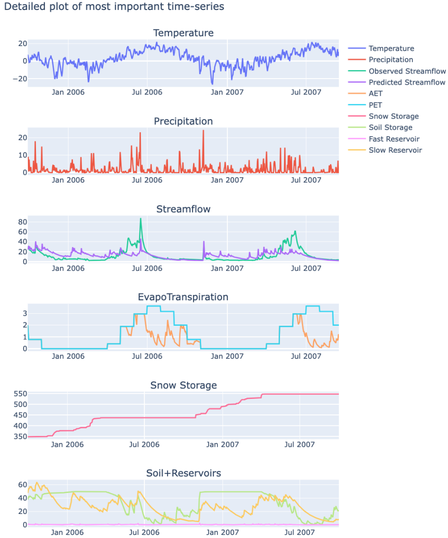
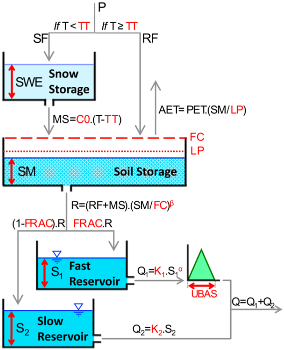


Figure: Schematic representation of the UQEF-Dynamic toolbox for efficient time-varying UQ & SA with different functionalities and important aspects

UQEF-Dynamic - Case study: HBV-SASK [6]

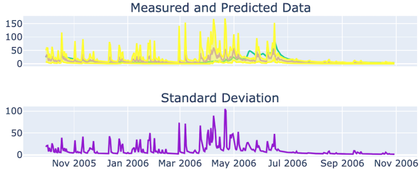


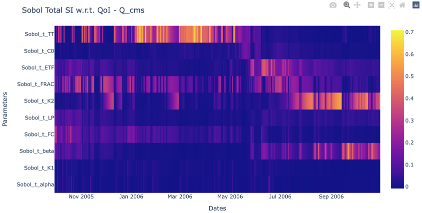


UQEF-Dynamic - Functionalities

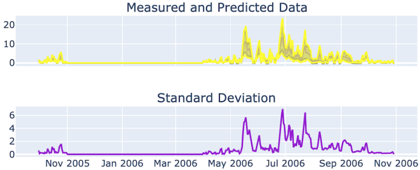
Facilitates analysis with multiple quantities of interest (QoI)

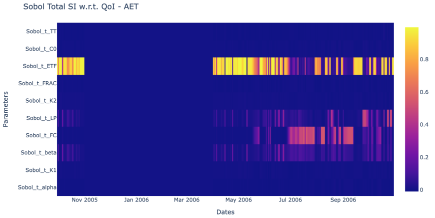
- Important aspect when designing UQ and SA and interpreting the results (of SA)
- QoI - Streamflow





- QoI - Actual Evapotranspiration

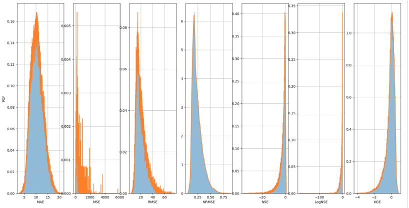




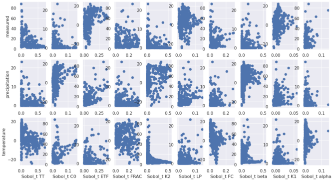
UQEF-Dynamic - Functionalities

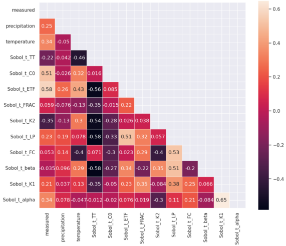
Facilitates different analyses, providing diverse insights into the complex model

- Total period likelihood functions analysis



- Analysis of SI over different hydrometeorological conditions





Time-varying gPCE Surrogates

Single timestamp -> single gPCE surrogate

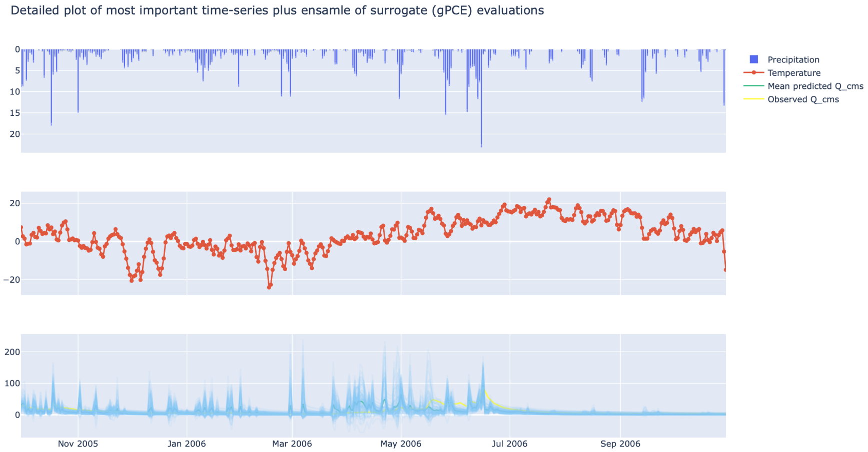
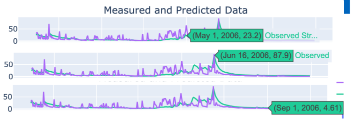


Figure: Evaluation of the time-vise gPCE surrogate

Time-varying gPCE Surrogates

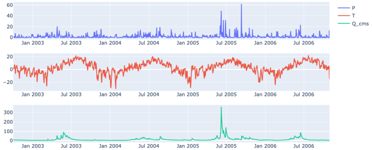
Single timestamp -> single gPCE surrogate



Surrogates built in this way are so-called frozen-it-time (i.e., can hardly be used as general surrogates for different forcing/internal-state conditions)

Figure: Inspection of learned gPCE coeff. over time (i.e., for different hydrometeorological conditions)

gPCE for Time-dependent Responses



1. Moving-window approach

Heuristic approaches

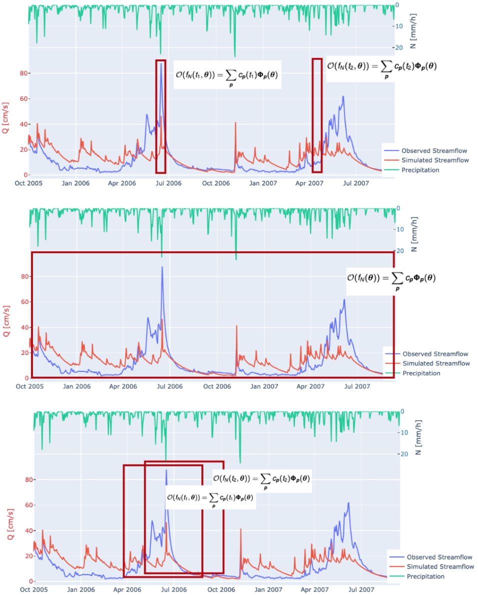
2. Autoregressive model of the first order

3. Filtering technique followed by construction of the gPCE surrogate

4. System identification technique followed by construction of the gPCE surrogate

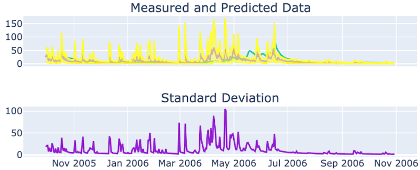
UQ of Dynamical Models
Sliding-window I

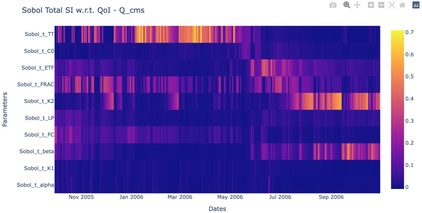
- Time-varying analysis
  (i.e., single timestamp -> single gPCE)
- Total period time-aggregate analysis
  - QoI: mean of the mode output, data-misfit function, etc.
  - convenient for identifying annual variability
  - or building surrogate for inversio
- Sliding-window analysis
  - QoI: mean of the mode output, data-misfit function, etc.



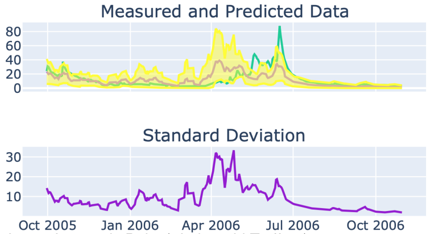
UQ of Dynamical Models Sliding-window II

- Time-varying analysis

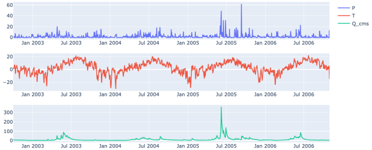




- Sliding-window analysis



gPCE for Time-dependent Responses

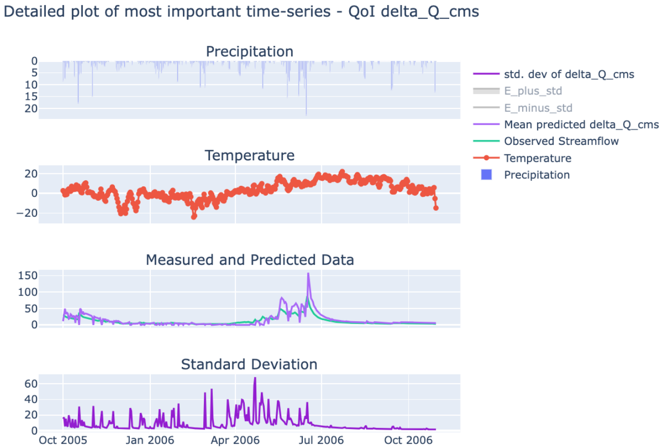


Heuristic approaches

1. Moving-window approach
2. **Autoregressive model of the first order**
3. Filtering technique followed by construction of the gPCE surrogate
4. System identification technique followed by construction of the gPCE surrogate

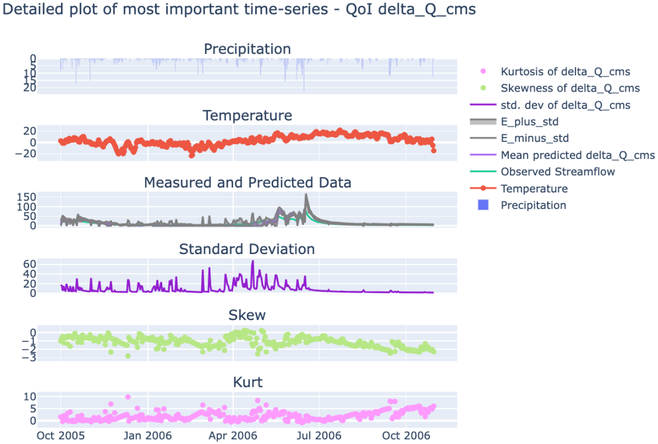
UQ of Dynamical Models
Autoregressive-like model

- Another heuristic approach - Building surrogate of the “delta” of QoI:
  $e_t = \alpha \cdot e_{t-1} + g_t; \; g_t \sim \mathcal{N}(0, \sigma^2) \;\Rightarrow\; \mathrm{f}_t = f_{t-1} + \sum_p c_{tp} \boldsymbol{\Phi}_p(\boldsymbol{\theta})$

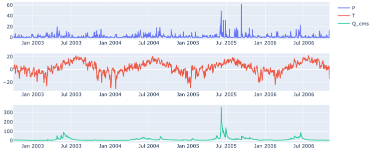


UQ of Dynamical Models
Autoregressive-like model

- Another heuristic approach - Building surrogate of the “delta” of QoI:
  $e_t = \alpha \cdot e_{t-1} + g_t;\ g_t \sim \mathcal{N}(0, \sigma^2) \ \Rightarrow\ \mathrm{f}_t = f_{t-1} + \sum_p c_{tp} \mathbf{\Phi}_p(\boldsymbol{\theta})$



gPCE for Time-dependent Responses



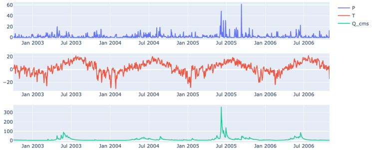
1. Moving-window approach

2. Autoregressive model of the first order

3. **Filtering technique (e.g Particle Filter) followed by construction of the gPCE surrogate [3]**
    - Idea:
        - 3.1 explicitly incorporate model states and sequentially import observations to update parameter dist.
        - 3.2 for each timestamp - feed the current parameter posterior dist. to the gPCE model
    - Produce “more accurate” UQ analysis
    - Work-in-progress...

4. System Identification technique followed by construction of the gPCE surrogate

gPCE for Time-dependent Responses



1. Moving-window approach

2. Autoregressive model of the first order

3. Filtering technique (e.g Particle Filter) followed by construction of the gPCE

4. **System Identification technique followed by construction of the gPCE surrogate [7]**
   - Idea:
     - **4.1** build a surrogate (e.g., nonlinear autoregressive exogenous model - NARX) - explicitly incorporate past and current forcing data
     - **4.2** conduct UQ in the parameter space of the surrogate (i.e., build gPCE surrogate of the NARX surrogate)
   - Work-in-progress...

Efficient UQ & SA using (adaptive) Sparse Grids

Coupling UQEF-Dynamic and SparseSpACE [9]
(tool for dimension-wise spatially adaptive SG)

UQ & SG: Multiple ways how to combine the gPCE and (Adaptive) SG

Var 1: Sparse Quadrature

- Approximate all the weighted integrals of f via some (adaptive) sparse interpolatory quadrature scheme
$\hat{c}_{\boldsymbol{p}}(t) = \sum_m \dots \sum_m f(t, \Theta_m^1, \dots, \Theta_m^d)\boldsymbol{\Phi}_{\boldsymbol{p}}(\Theta_m^1, \dots, \Theta_m^d)\omega_m^1 \dots \omega_m^d$

Var 2: SG Interpolation Surrogate + gPCE

- SG Interpolation of $f(x, \boldsymbol{\theta})$
- Use SG model surrogate to compute the gPCE coefficients [4]

$$\hat{c}_{\boldsymbol{p}}(t) = \int_{\Gamma} f_{\mathcal{SGI}}(t, \boldsymbol{\theta})\boldsymbol{\Phi}_{\boldsymbol{p}}(\boldsymbol{\theta})\boldsymbol{\rho}(\boldsymbol{\theta})d\boldsymbol{\theta}$$

Var 3: Sparse PSP [1, 2]

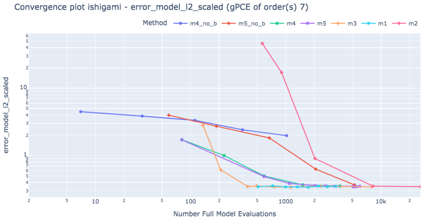
- Rely on the connection between projection and interpolation
- Use the resulting Combination Technique from SG-based interpolation of the function
- Compute gPCE coeff. on smaller anisotropic grids and then combine them according to the interpolatory CT scheme

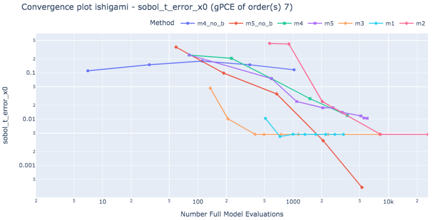
UQ & SG: Benchmark Examples

Step 1: Benchmark Convergence of different methods

- **UQ & SA:** Ishigami function (3D) - analytical values for SI available
- Convergence results as expected
- For simple cases, building an intermediate SG surrogate is not beneficial time-wise; wip - implementation and analysis of Var 3

Variant	Method	Interpolation method (SGI)	quadrature method	gPCE
Var 1	m1	no	Full Gauss-Legendre	yes
	m2	no	(Sparse) Clenshaw-Curtis	yes
	m3	no	(Sparse) delayed Kronrod-Patterson[10]	yes
Var 2	m4	(piecewise linear) standard CT	Gauss-Legendre (high order) or analytical computation	yes
	m5	(piecewise linear) spatially adaptive CT	Gauss-Legendre (high order) or analytical computation	yes





UQ & SG: Coupling UQEF-Dynamic and SparseSpACE tools

Step 2: Efficient UQ & SA of Hydrological Model (6D-10D) with SG surrogate

- Time-aggregated - Var 2: Building a single spatially adaptive SG approximation of the RMSE surface (relying on SparseSpACE [8])
- **Again, the same question - How to build an SG surrogate over time?**
  One idea: building multiple surrogates for different regimes

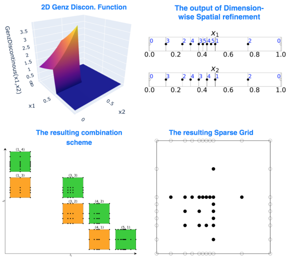


Figure: Dimension-wise spatial refinements of Gen function

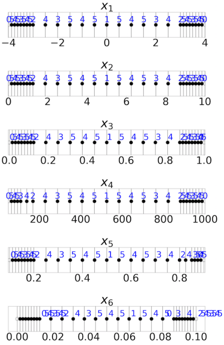


Figure: Dimension-wise spatial refinements of stochastic parameters of the HBV model

To conclude...

UQEF & UQEF-Dynamic Software tools

When Dynamics enters the game

- **Work-in-progress** - filtering and system identification approaches

Other paths of my research

- Sparse Grids for UQ - optimizing the number of necessary model runs
- **Work-in-progress** - how to build spatially adaptive SG surrogates for dynamical models
- Using learned surrogate models for efficient calibration under uncertainty

MISC

- PyApprox: A newly published python-based tool that implements some of the methods in the domain of UQ
- UM-Bridge: tool intended for coupling advanced models to existing methods/frameworks

Thank you for your attention!
Looking forward to fruitful discussions! :-D

References I

[1] Patrick R Conrad and Youssef M Marzouk. ADAPTIVE SMOLYAK PSEUDOSPECTRAL APPROXIMATIONS. Tech. rep. 2013. arXiv: 1209.1406v2.

[2] Paul G Constantine, Michael S Eldred, and Eric T Phipps. “Sparse pseudospectral approximation method”. In: Computer Methods in Applied Mechanics and Engineering 229 (2012), pp. 1–12.

[3] Yurui Fan et al. “Parameter uncertainty and temporal dynamics of sensitivity for hydrologic models: A hybrid sequential data assimilation and probabilistic collocation method”. In: Environmental Modelling & Software 86 (Dec. 2016). DOI: 10.1016/j.envsoft.2016.09.012.

[4] IonuÈ›-Gabriel FarcaÈ™ et al. “Nonintrusive uncertainty analysis of fluid-structure interaction with spatially adaptive sparse grids and polynomial chaos expansion”. In: SIAM Journal on Scientific Computing 40.2 (2018), B457–B482.

References II

[5] Fabrice Gamboa et al.
Global Sensitivity Analysis: a new generation of mighty estimators based on rank st
2020. arXiv: 2003.01772 [math.ST].

[6] Hoshin V. Gupta and Saman Razavi. “Revisiting the Basis of Sensitivity Analysis for Dynamical Earth System Models”. In: Water Resources Research 54.11 (2018), pp. 8692–8717. ISSN: 19447973. DOI: 10.1029/2018WR022668. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR022668.

[7] C. V. Mai et al.
Surrogate modelling for stochastic dynamical systems by combining NARX models
2016. arXiv: 1604.07627 [stat.ME].

[8] Michael Obersteiner.
SparseSpACE - The Sparse Grid Spatially Adaptive Combination Environment.
https://github.com/obersteiner/sparseSpACE. 2020.

References III

[9] Michael Obersteiner and Hans Joachim Bungartz. "A generalized spatially adaptive sparse grid combination technique with dimension-wise refinement". In: SIAM Journal on Scientific Computing 43.4 (2021), A2381–A2403. ISSN: 10957197. DOI: 10.1137/20M1325885.

[10] Knut Petras. "Smolyak cubature of given polynomial degree with few nodes for increasing dimension". In: Numerische Mathematik 93 (Feb. 2003). DOI: 10.1007/s002110200401.

[11] Andrea Saltelli et al. "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index". In: Computer Physics Communications 181.2 (2010), pp. 259–270. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2009.09.018. URL: https://www.sciencedirect.com/science/article/pii/S0010465509003087.

[12] I M Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo 2001, pp. 271–280.

References IV

[13] Bruno Sudret. “Global sensitivity analysis using polynomial chaos expansions”. In: Reliability Engineering and System Safety 93 (2008), pp. 964–979. ISSN: 09518320. DOI: 10.1016/j.ress.2007.04.002.

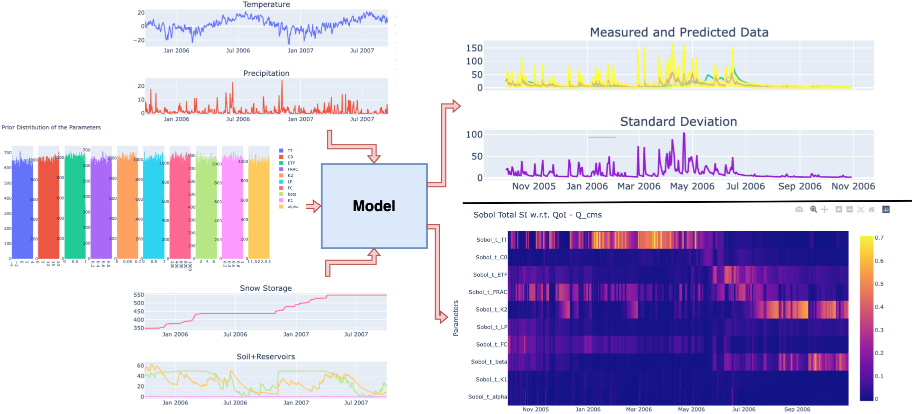
Extra Slides

UQ and/vs. SA

UQ:

$$\mathbb{E}[\mathcal{O}(f)] := \int_{\Gamma} \mathcal{O}(f)(t,\theta)\rho(\theta)d\theta$$

$$Var[\mathcal{O}(f)] := \mathbb{E}[\mathcal{O}(f)^2] - (\mathbb{E}[\mathcal{O}(f)])^2$$



Variance-based SA [Sobol 2001]:

$$S_j := \frac{Var[f] - \mathbb{E}_{\theta_j}[Var_{\theta_{\sim j}}[f|\theta_j]]}{Var[f]}$$

General Polynomial Chaos Expansion (gPCE) with Pseudo-spectral projection (PSP)

$$f(t,\boldsymbol{\theta}) \approx f_N(t,\boldsymbol{\theta}) := \sum_{\boldsymbol{p}} c_{\boldsymbol{p}}(t)\boldsymbol{\Phi}_{\boldsymbol{p}}(\boldsymbol{\theta}) := \sum_{\boldsymbol{p}} < f(t,\boldsymbol{\theta}), \boldsymbol{\Phi}_{\boldsymbol{p}}(\boldsymbol{\theta}) >_{\rho(\boldsymbol{\theta})} \boldsymbol{\Phi}_{\boldsymbol{p}}(\boldsymbol{\theta})$$

- $\boldsymbol{p} = (p_1, \dots, p_d)$ is a multi-index in $\mathcal{P}_P = \{\boldsymbol{p} \in \mathbb{N}^d : \sum_{j=1}^d p_j \leq P\}$
- $\boldsymbol{\Phi}_{\boldsymbol{p}}(\boldsymbol{\theta}) := \Phi_{p_1}(\theta_1) \cdot \ldots \cdot \Phi_{p_d}(\theta_d)$ such that $< \Phi_{p_{j_k}}(\theta_j), \Phi_{p_{j_m}}(\theta_j) >_{\rho_j(\theta_j)} = \delta_{km}$

$$\hat{c}_{\boldsymbol{p}}(t) := \sum_{\boldsymbol{q}=1}^{\boldsymbol{Q}} f(t,\boldsymbol{\theta_q})\boldsymbol{\Phi}_{\boldsymbol{p}}(\boldsymbol{\theta_q})\boldsymbol{\omega}$$

- total number of coefficients. $N = \binom{P+d}{d}$
- total number of model evaluations $\boldsymbol{Q} = \prod_{j=1}^d Q_j$
- has to hold - $p_j = floor(DE_{(Q_j)}/2)$ [1]

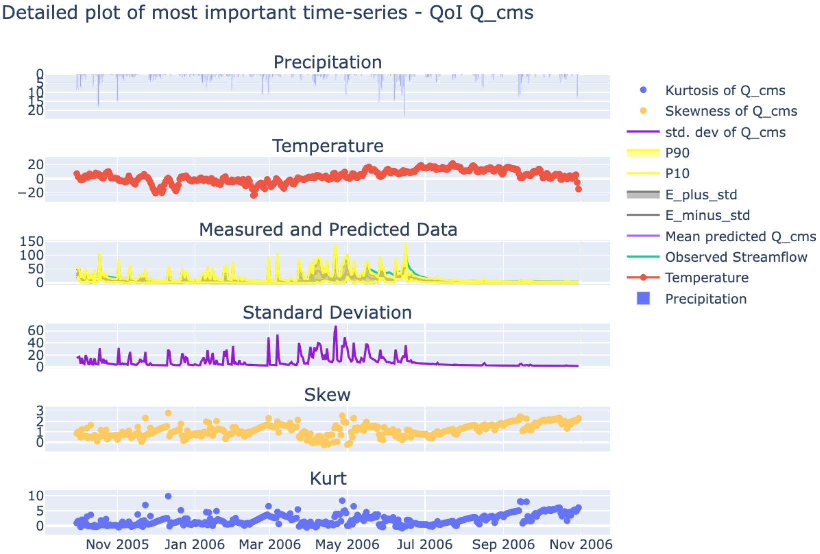
UQEF & UQEF-Dynamic Software Tools

The tools implement various methods for efficient inspection of high-dimensional parameter spaces with

- **in-parallel** execution of (expensive) simulation models
- **efficient collection, processing, storing, and visualizing** the results of the UQ and SA
- with the goal of facilitating **simulation-aided knowledge discovery, prediction, and design**.

Tools should be easy to install, configure, deploy to the computer clusters, and extend

Time-varying UQ analysis



Time-varying UQ analysis