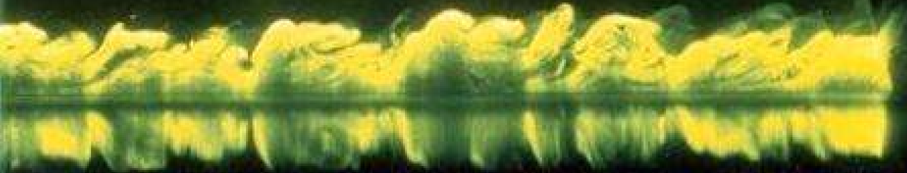
Bayesian estimates of the parameter variability in turbulence models

SAMO 2013



Wouter Edeling, Paola Cinnella, R. Dwight, H. Bijl

DynFluid ParisTech and TU Delft

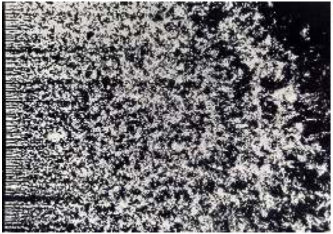
July 4, 2013

Outline

- Problem statement + our approach.
- Statistical tools.
- Results.

Problem statement

- Direct Numerical Simulation approach intractable.
- Reynolds-Averaged Navier-Stokes Equations:
  - Averaged governing equations: $r(\bar{\mathbf{u}}) = 0$.
  - Turbulence model: $r'(\bar{\mathbf{u}}, \boldsymbol{\theta}) = 0$.
  - Many turbulence models are available.



Simplified governing equations isolate a closure coefficient

⇩

$$C_{\epsilon 2} = \left(\frac{n+1}{n}\right)$$

⇩

Experimental data to estimate n:

$C_{\epsilon 2} = 2.00$, Jones – Launder

$C_{\epsilon 2} = 1.92$, Launder – Sharma

$C_{\epsilon 2} = 1.80$, Chien

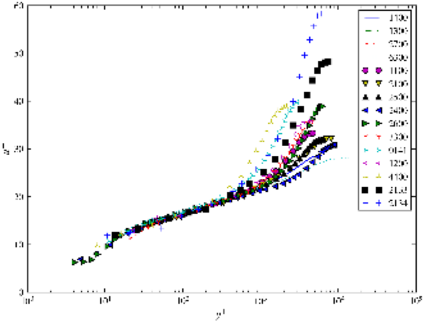
- 1st source of uncertainty: parametric uncertainty

Problem statement

- 2nd source of uncertainty: Error due to approximate physical turbulence models: $\underline{r}'(\cdot)$
  - Many terms: model inadequacy, structural uncertainty, model uncertainty.
  - General term, **model error**, different methods.
  - Example: **model-inadequacy term** of Kennedy and O'Hagan [4]; $\mathbf{z} = \mathbf{y} + \underline{\boldsymbol{\eta}} + \mathbf{e}$.
- Our overall goal is to compute estimates of the model error in turbulence models.

Our approach

1. Define a class of flows to be considered and choose models.
2. Collect experimental data (u^+) for multiple flow cases.
3. **Calibrate** using Bayesian model updating $\rightarrow$ multiple posterior distributions.
4. **Validate** the method $\rightarrow$ put error bars on the model output of a **predictive** flow case.



Tools: Bayesian inference

- Theoretical model: Bayes' theorem $\rightarrow$ posterior pdf $p(\boldsymbol{\theta}|\mathbf{z})$ of model parameters $\boldsymbol{\theta}$ conditioned on data $\mathbf{z}$

$$p(\boldsymbol{\theta}|\mathbf{z}) = \frac{p(\mathbf{z}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{z})} \tag{1}$$

- $p(\mathbf{z}|\boldsymbol{\theta})$ is known as the **likelihood**
- $p(\boldsymbol{\theta})$ is the **prior** uncertainty: a belief about $\boldsymbol{\theta}$.
- Equation (5) is a statistical calibration.

Tools: Calibration phase

- Statistical model from Cheung et. al. [1]→ specifies $p(\mathbf{z}|\boldsymbol{\theta})$:
  - The RANS output as a function of the uncertain closure coefficients $\boldsymbol{\theta}_k$: $u^+(\mathbf{y}_k^+, \mathbf{t}_k; \boldsymbol{\theta}_k)$.
  - Model error term: $\eta_k(\mathbf{y}_k^+; \boldsymbol{\gamma}_k)$.
  - Experimental error term: $\mathbf{e}_k$.
- $\mathbf{t}_k$: are the non-random parameters and $k = 1, 2, \cdots, K$ is the flow-case index.

$$\mathbf{z}_k = \zeta_k(\mathbf{y}_k^+) + \mathbf{e}_k, \tag{2}$$

$$\zeta_k(\mathbf{y}_k^+) = \eta_k(\mathbf{y}_k^+; \boldsymbol{\gamma}_k) \cdot u^+(\mathbf{y}_k^+, \mathbf{t}_k; \boldsymbol{\theta}_k), \tag{3}$$

Other model error forms (e.g. additive) are also possible.

Tools: Sampling

- We used a fast boundary-layer code $\rightarrow$ Markov-Chain Monte-Carlo method to draw samples from $p(\boldsymbol{\theta}_k \mid \mathbf{z}_k)$.
- Used these samples of $\boldsymbol{\theta}_k$ to construct approximate pdfs using a kernel-density estimation.
- Which of these pdfs can be informed from the experimental data? $\rightarrow$ ANOVA sensitivity analysis
- We computed the main Sobol indices of the velocity profiles using a Stochastic Collocation Expansion, based on the work of Sudret [6] and Tang [7]

Tools: Summarizing the posteriors

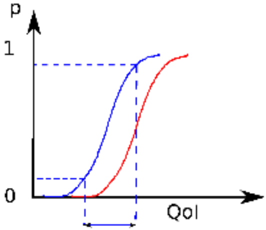
- $15 \times 4 \; p(\boldsymbol{\theta} \mid \mathbf{z})$.
- To summarize this large amount of information, we plot the Highest-Posterior Density (HPD) intervals.
- HPD intervals are credible intervals with the added properties:
  - The density of every point inside the interval is greater than that of every point outside the interval.
  - The $(1-\beta)$ HPD interval is of the smallest possible width.

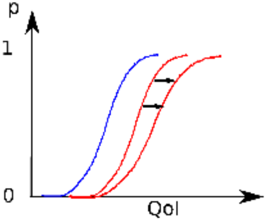
Validation Tools

- Since η requires calibration, it tells you something about the model error of that case alone.
- Build a more general model for the uncertainty present in the turbulence models, using the 15 posterior closure coefficients distributions.
- 1st attempt: construct a Probability box (p-box) for a flow **not** in the calibration set.
- 2nd attempt: Bayesian Model Averaging.

Validation Tools: P-boxes

- For a given θ_k^i: multiple $p(\theta_k^i \mid \mathbf{z}_k)$.
- Use posterior $\boldsymbol{\theta}_k$ samples to construct an empirical cdf (ecdf) of the Quantity-of-Interest (QoI).
- The envelope formed by this collection of ecdfs $\forall\, k$ is a p-box.
- Error bars due to: parametric uncertainty + spread of posteriors + (model error).





Validation Tools: BMA

- Let M_i be a turbulence model in set $\mathcal{M}$, S_k a $d\bar{p}/dx$ scenario in set $\mathcal{S}$ and $\mathcal{Z}$ be the set of all experimental calibration data.
- The BMA prediction of a QoI Δ is then [2]:

$$\mathrm{E}\left(\Delta \mid \mathcal{Z}\right) = \sum_{i=1}^{I}\sum_{k=1}^{K} \mathrm{E}\left(\Delta \mid M_i, S_k, \mathbf{z}_k\right) \mathrm{pr}\left(M_i \mid S_k, \mathbf{z}_k\right) \mathrm{pr}\left(S_k\right) \tag{4}$$

- The scenario of Δ does not have to be in the set $\mathcal{S}$.
- Each individual expectation in (4) is weighted by
  - The posterior model probability $\mathrm{pr}\left(M_i \mid S_k, \mathbf{z}_k\right)$.
  - The prior scenario probability $\mathrm{pr}\left(S_k\right)$.

Validation Tools: BMA

As a measure of uncertainty in the BMA prediction we calculate the variance [2]:

$$\mathrm{var}\left[\Delta \mid \mathcal{Z}\right] = \sum_{i=1}^{I}\sum_{k=1}^{K} \mathrm{var}\left[\Delta \mid M_i, S_k, \mathbf{z}_k\right] \mathrm{pr}\left(M_i \mid S_k, \mathbf{z}_k\right) \mathrm{pr}\left(S_k\right) +$$

$$\sum_{i=1}^{I}\sum_{k=1}^{K} \left(\mathrm{E}\left[\Delta \mid M_i, S_k, \mathbf{z}_k\right] - \mathrm{E}\left[\Delta \mid S_k, \mathbf{z}_k\right]\right)^2 \mathrm{pr}\left(M_i \mid S_k, \mathbf{z}_k\right) \mathrm{pr}\left(S_k\right) +$$

$$\sum_{k=1}^{K} \left(\mathrm{E}\left[\Delta \mid S_k, \mathbf{z}_k\right] - \mathrm{E}\left[\Delta \mid \mathbf{z}_k\right]\right)^2 \mathrm{pr}\left(S_k\right)$$

- In-model in scenario variance (parametric uncertainty of each $p\left(\boldsymbol{\theta}_k \mid \mathbf{z}_k\right)$).
- Between-model in scenario variance (model error).
- Between-scenario variance (spread).

Results: Posterior distributions

- Posterior distributions for $C_{\varepsilon 2}$ for a favorable, zero, mild, moderate and strongly adverse $d\bar{p}/dx$.

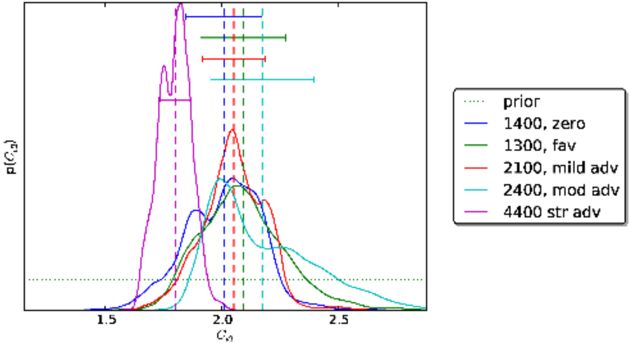


Figure: Some marginal $p\left(C_{\varepsilon 2} \mid \mathbf{z}_k\right)$ of the $k - \varepsilon$ model, high S_u.

Results: Posterior distributions

- Posterior distributions for κ for a favorable, zero, mild, moderate and strongly adverse $d\bar{p}/dx$.

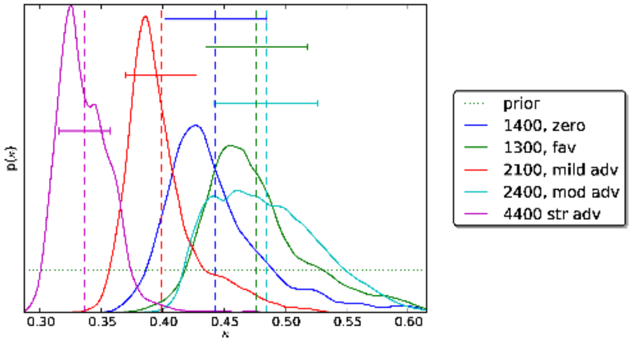


Figure: Some $p(\kappa \mid \mathbf{z})$ of the $k - \varepsilon$ model, high S_u.

Results: Posterior distributions

- Posterior distributions for C_μ for a favorable, zero, mild, moderate and strongly adverse $d\bar{p}/dx$.

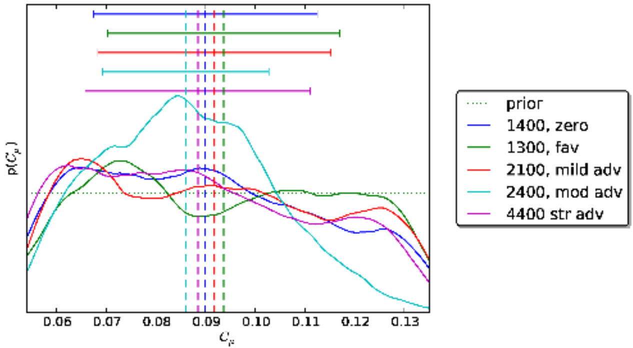
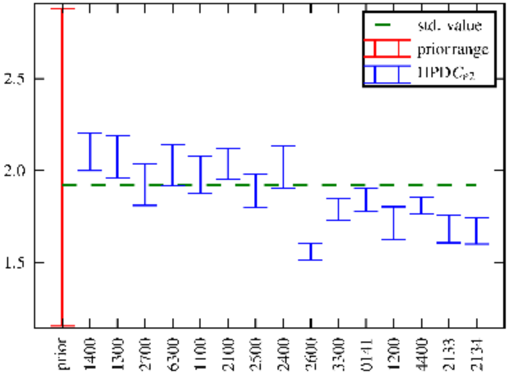


Figure: Some $p(C_\mu \mid \mathbf{z})$ of the $k-\varepsilon$ model, low S_u.

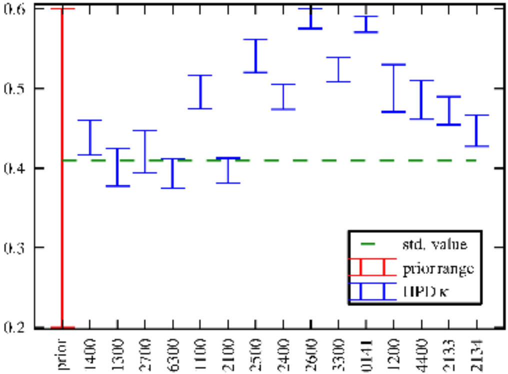
Results: HPD intervals

$C_{\varepsilon 2}$ HPD intervals of the $k - \varepsilon$ model.

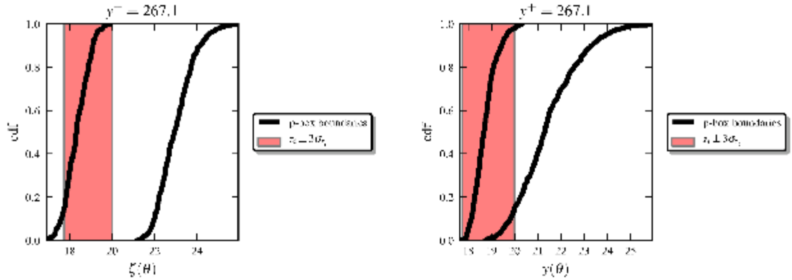


Results: HPD intervals

- κ: HPD intervals of the Baldwin-Lomax model: large spread.



Results: P-boxes

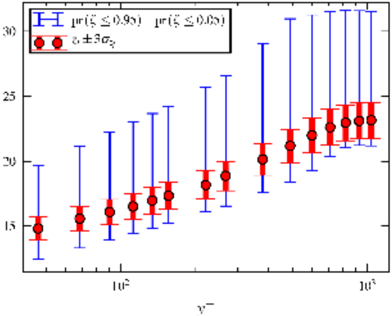


(a) A p-box from the $k-\varepsilon$ model (b) A p-box from the Baldwin-Lomax model

- Both are consistent with the experimental data.
- Both generate rather large error bars on the prediction.

Results: P-box error bars

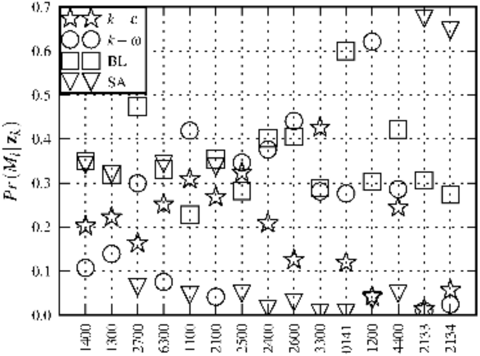
We extract (90 %) credible intervals from the p-boxes



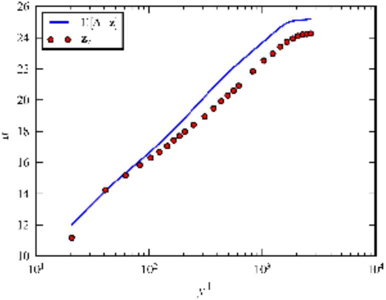
- The p-box overestimates the amount of uncertainty because in a pbox each $d\bar{p}/dx$ scenario is equally weighted.
- Bayesian Model Averaging does have weights.

Results: Posterior model probability

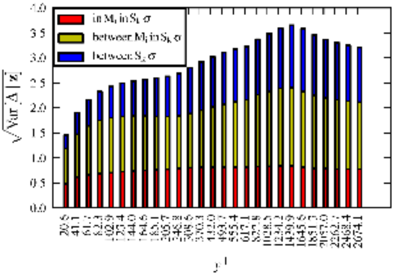
- Computed for all models in $\mathcal{M}$ for a given S_k using samples from $p(\boldsymbol{\theta}_k \mid \mathbf{z}_k)$.
- Can be considered a measure of consistency of calibrated model M_i with data $\mathbf{z}_k$.



Results: BMA prediction



(c) BMA prediction, uniform $\mathrm{pr}(S_k)$.



(d) BMA std. dev., uniform $\mathrm{pr}(S_k)$.

- Same situation as with the p-box, overestimation of uncertainty due to equally weighting scenarios.
- However, we are free to modify $\mathrm{pr}(S_k)$.

Results: BMA prediction

- Define a model-error measure $\forall k$ based on the between-model in-scenario variance:

$$\mathcal{E}_k = \frac{\sum_{i=1}^{I} \|\mathrm{E}\left[\Delta \mid M_i, S_k, \mathbf{z}_k\right] - \mathrm{E}\left[\Delta \mid S_k, \mathbf{z}_k\right]\|_2 \mathrm{pr}\left(M_i \mid S_k, \mathbf{z}_k\right)}{\|\mathrm{E}\left[\Delta \mid S_k, \mathbf{z}_k\right]\|_2}$$

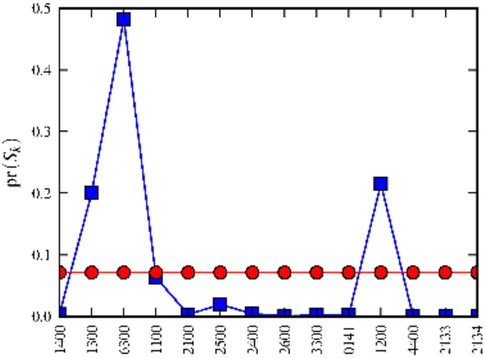
- Use $\mathcal{E}_k$ to inform $\mathrm{pr}\left(S_k\right)$:

$$\mathrm{pr}\left(S_k\right) = \frac{\mathcal{E}_k^{-p}}{\sum_{k=1}^{K} \mathcal{E}_k^{-p}}, \quad k = 1, 2, \cdots, K$$

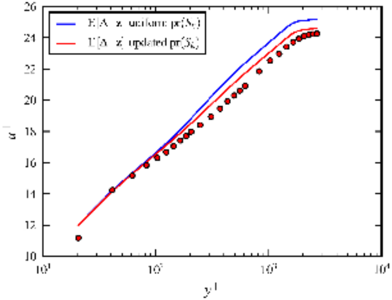
Here, p is used to penalize those scenarios that have a high $\mathcal{E}_k$.

Results: BMA prediction

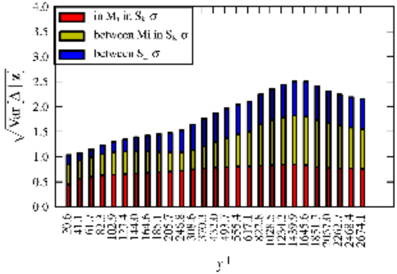
- The uniform and the updated $\mathrm{pr}(S_k)$ using $p = 8$:



Results: BMA prediction



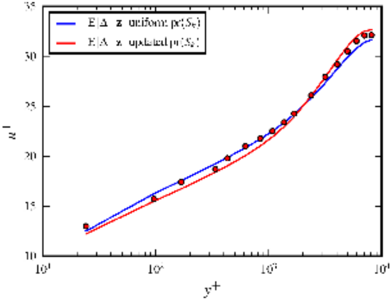
(e) BMA predictions.



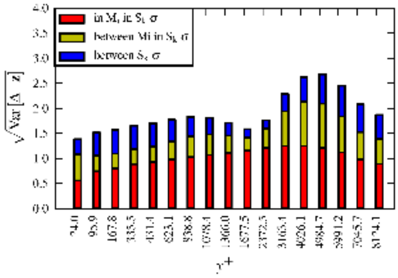
(f) BMA std. dev., updated $\mathrm{pr}(S_k)$.

- Standard deviation reduced by approximately 30%.
- Prediction is brought closer to the validation data $\mathbf{z}_v$.

Results: Another BMA prediction



(g) BMA predictions.



(h) BMA std. dev., updated $\mathrm{pr}(S_k)$.

- Uniform $\mathrm{pr}\left(S_k\right)$ prediction was already good, updated one does not deviate much.
- However, the standard deviation is reduced by roughly 40%.

Conclusion

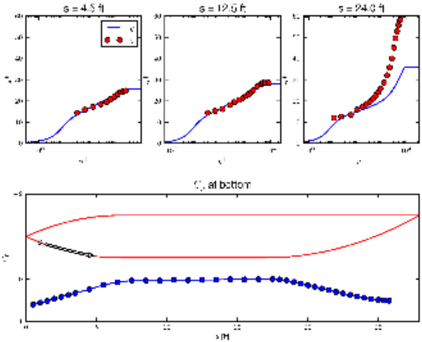
- The spread in most-likely closure coefficients due to different pressure gradients is significant for all considered models $\rightarrow$ **no 'true value' for the closure coefficients**.
- **Posterior model probabilities also vary a lot**.
- For validation cases: **BMA is more flexible than p-boxes**.
- So far we have tested the BMA approach on 15 boundary-layer validation flows with good results.
- Computational challenges will increase when we move to more interesting flows.

Paper

- Part of this work can be found in: *Bayesian estimates of the parameter variability in the $k - \varepsilon$ turbulence model*, W.N. Edeling, P. Cinnella, R.P. Dwight, H. Bijl (submitted).
- **Thank you for your attention.**

Problem dependent performance

Example: boundary layers over an airfoil shaped body, calculated with the standard $k - \varepsilon$ model. Data from Schubauer [5]



No universal turbulence model exists, (re)-calibration is required. Normally this is done in a **deterministic** way.

Bayesian inference

We quantify the uncertainties using a **stochastic** framework: Bayesian inference

Definition

Bayesian inference is the process of fitting a probability model to a set of data and summarizing the result by a probability distribution on the parameters of the model and on unobserved quantities such as predictions for new observations [3]

- Bayesian inference represents all types of uncertainty as probability $\rightarrow$ probability density function (pdf)
- Uses a set of observational data to infer a pdf of the closure coefficients $\rightarrow$ estimate $+$ measure of confidence in estimate

Our approach

- Advantage estimating model error by uncertainty and spread in closure coefficients:
  - Geometry independence.
  - Coefficients are related to underlying physics, and thus to some part of the model error.

Bayesian inference

- Theoretical model: Bayes' theorem → posterior pdf $p(\boldsymbol{\theta}|\mathbf{z})$ of model parameters $\boldsymbol{\theta}$ conditioned on data $\mathbf{z}$

$$p(\boldsymbol{\theta}|\mathbf{z}) = \frac{p(\mathbf{z}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{z})} \tag{5}$$

- A framework able to incorporate multiple sources of uncertainty.
- The experimental observations $\mathbf{z}$ also possess (measurement) uncertainties → data pdf $p(\mathbf{z})$
- $p(\mathbf{z}|\boldsymbol{\theta})$ is known as the **likelihood**, i.e. the probability that the model will predict $\mathbf{z}$ given $\boldsymbol{\theta}$. $p(\boldsymbol{\theta})$ is the **prior** uncertainty in the model parameters. It represents a belief about $\boldsymbol{\theta}$.
- Equation (5) is a statistical calibration, it infers the posterior pdf of the parameters $\boldsymbol{\theta}$ that fits the model to the observations $\mathbf{z}$.

Model-inadequacy term

- This the model-inadequacy term η from Kennedy and O'Hagan, which is a means to represent the model error.
- Specify the statistical term for η as $\eta \sim \mathrm{GP}(1, c_\eta)$ with covariance function:

$$c_\eta(y^+, y^{+\prime} \mid \gamma) := \sigma^2 \exp\left[-\left(\frac{y^+ - y^{+\prime}}{10^\alpha l}\right)^2\right],$$

- This implies a statistical model for the true process as:

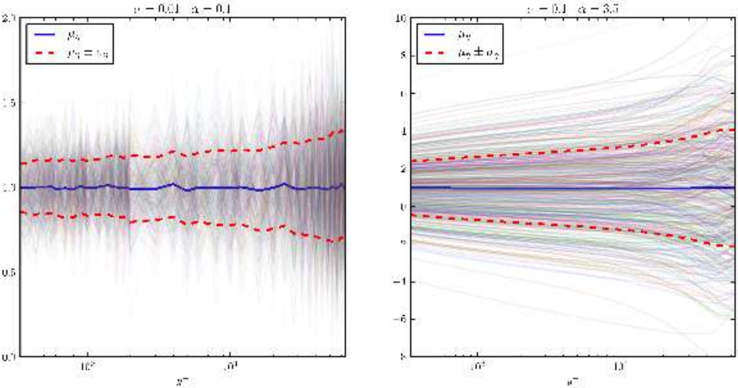
$$\begin{aligned} \zeta \mid \boldsymbol{\theta}, \boldsymbol{\gamma} &\sim \mathrm{GP}(\mu_\zeta, c_\zeta) \\ \mu_\zeta(y^+ \mid \boldsymbol{\theta}) &= u^+(y^+, \mathbf{t}; \boldsymbol{\theta}) \\ c_\zeta(y^+, y^{+\prime} \mid \boldsymbol{\theta}, \boldsymbol{\gamma}) &= u^+(y^+, \mathbf{t}; \boldsymbol{\theta}) \cdot c_\eta(y^+, y^{+\prime} \mid \boldsymbol{\gamma}) \cdot u^+(y^{+\prime}, \mathbf{t}; \boldsymbol{\theta}). \end{aligned} \tag{6}$$

Model-inadequacy term

- The model-inadequacy term η needs to be calibrated to fit a certain problem $\rightarrow$ (hyper) parameters $\gamma := [\sigma, \alpha]$ are calibrated along with the closure coefficients.
- The model-inadequacy term η implies a certain topology for the error.
- Namely, the structural error has some smoothness and it increases with increasing velocity.
- The smoothness of the model inadequacy term is controlled by α and the magnitude by σ.

Model-inadequacy term

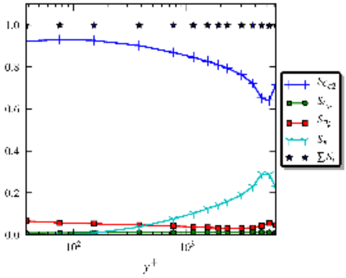
- This becomes clear by drawing samples from η:



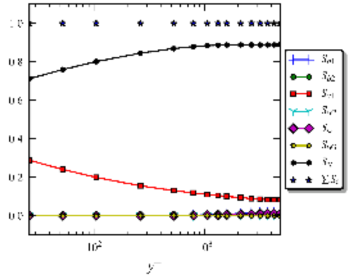
ANR

TUDelft

Sensitivity analysis



(i) Main S_u of $k - \varepsilon$ model

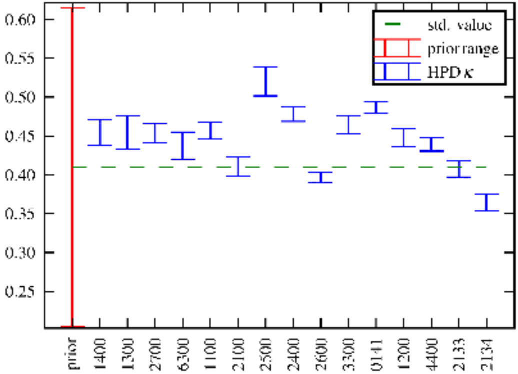


(j) Main S_u of SA model

- The coefficients with high S_u have informed posterior distributions $p(\theta_u \mid \mathbf{z})$.
- The coefficients with low S_u do not differ much from the uniform prior distribution $p(\theta_u)$.

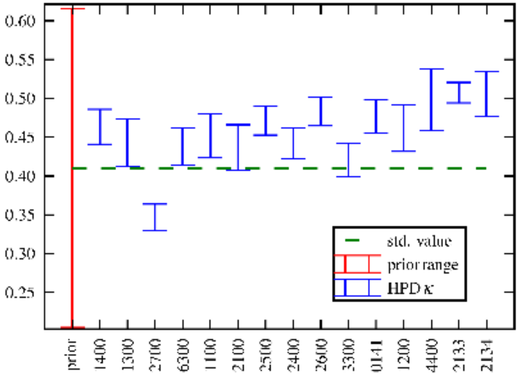
HPD intervals

- κ: HPD intervals of the $k - \omega$ model.



Some results

- κ: HPD intervals of the Spalart-Allmaras model.



References

S.H. Cheung, T.A. Oliver, E.E. Prudencio, S. Prudhomme, and R.D. Moser.
Bayesian uncertainty analysis with applications to turbulence modeling.
Reliability engineering and systems safety, 96(9):1137–1149, 2011.

David Draper.
Assessment and propagation of model uncertainty.
Journal of the Royal Statistical Society. Series B (Methodological), pages 45–97, 1995.

A. Gelman, JB Carlin, HS Stern, and DB Rubin.
Bayesian data analysis. 1995.
Chapman&Hall, London.

M.C. Kennedy and A. O'Hagan.
Bayesian calibration of computer models.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(3):425–464, 2001.

G.B. Schubauer and PS Klebanoff.
Investigation of separation of the turbulent boundary layer.
Technical report, DTIC Document, 1950.

B. Sudret.
Global sensitivity analysis using polynomial chaos expansions.
Reliability Engineering & System Safety, 93(7):964–979, 2008.

G. TANG, MS ELDRED, and L.P. SWILER.
Global sensitivity analysis for stochastic collocation expansion.
CSRI SUMMER PROCEEDINGS 2009, page 100, 2010.