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: \( r(\bar{u}) = 0 \).
  - Turbulence model: \( r'(\bar{u}, \theta) = 0 \).
  - Many turbulence models are available.

- 1st source of uncertainty: parametric uncertainty
Problem statement

- 2nd source of uncertainty: Error due to approximate physical turbulence models: $r'(\cdot)$
  - Many terms: model inadequacy, structural uncertainty, model uncertainty.
  - General term, **model error**, different methods.
  - Example: **model-inadequacy term** of Kennedy and O’Hagan [4]; $z = y + \eta + e$.

- 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 → posterior pdf \( p(\theta|z) \) of model parameters \( \theta \) conditioned on data \( z \)

\[
p(\theta|z) = \frac{p(z|\theta)p(\theta)}{p(z)} \tag{1}
\]

- \( p(z|\theta) \) is known as the **likelihood**
- \( p(\theta) \) is the **prior** uncertainty: a belief about \( \theta \).
- Equation (5) is a statistical calibration.
Tools: Calibration phase

- Statistical model from Cheung et. al. [1] specifies $p(z | \theta)$:
  - The RANS output as a function of the uncertain closure coefficients $\theta_k$: $u^+(y_k^+, t_k; \theta_k)$.
  - Model error term: $\eta_k(y_k^+; \gamma_k)$.
  - Experimental error term: $e_k$.

- $t_k$: are the non-random parameters and $k = 1, 2, \cdots, K$ is the flow-case index.

\[
z_k = \zeta_k(y_k^+) + e_k, \quad (2) \\
\zeta_k(y_k^+) = \eta_k(y_k^+; \gamma_k) \cdot u^+(y_k^+, t_k; \theta_k), \quad (3)
\]

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

- We used a fast boundary-layer code → Markov-Chain Monte-Carlo method to draw samples from $p(\theta_k | z_k)$.
- Used these samples of $\theta_k$ to construct approximate pdfs using a kernel-density estimation.
- Which of these pdfs can be informed from the experimental data? → 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

- $15 \times 4 \ p(\theta \mid z)$.
- 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 $\eta$ 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 $\theta^i_k$: multiple $p(\theta^i_k | z_k)$.
- Use posterior $\theta_k$ 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 $\Delta$ is then [2]:

$$E(\Delta | \mathcal{Z}) = \sum_{i=1}^{I} \sum_{k=1}^{K} E(\Delta | M_i, S_k, z_k) \cdot pr(M_i | S_k, z_k) \cdot pr(S_k)$$

(4)

- The scenario of $\Delta$ does not have to be in the set $S$.

- Each individual expectation in (4) is weighted by
  - The posterior model probability $pr(M_i | S_k, z_k)$.
  - The prior scenario probability $pr(S_k)$.
Validation Tools: BMA

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

$$\text{var} [\Delta | \mathcal{Z}] = \sum_{i=1}^{I} \sum_{k=1}^{K} \text{var} [\Delta | M_i, S_k, z_k] \text{pr} (M_i | S_k, z_k) \text{pr} (S_k) +$$

$$\sum_{i=1}^{I} \sum_{k=1}^{K} (E [\Delta | M_i, S_k, z_k] - E [\Delta | S_k, z_k])^2 \text{pr} (M_i | S_k, z_k) \text{pr} (S_k) +$$

$$\sum_{k=1}^{K} (E [\Delta | S_k, z_k] - E [\Delta | z_k])^2 \text{pr} (S_k)$$

- In-model in scenario variance (parametric uncertainty of each $p (\theta_k | z_k)$).
- 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 $p(C_{\varepsilon 2} | z_k)$ of the $k - \varepsilon$ model, high $S_u$. 
Results: Posterior distributions

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

Figure: Some $p(\kappa \mid z)$ of the $k - \varepsilon$ model, high $S_{\mu}$. 
Results: Posterior distributions

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

Figure: Some $p(C_{\mu} | z)$ of the $k - \varepsilon$ model, low $S_\nu$. 
Results: HPD intervals

$C_{\varepsilon_2}$ HPD intervals of the $k - \varepsilon$ model.
Results: HPD intervals

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 $M$ for a given $S_k$ using samples from $p(\theta_k \mid z_k)$.
- Can be considered a measure of consistency of calibrated model $M_i$ with data $z_k$. 
Results: BMA prediction

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

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

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

DynFluid

ANR

TU Delft

July 4, 2013 22 / 28
Results: BMA prediction

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

$$\mathcal{E}_k = \sum_{i=1}^I \| \mathbb{E} [\Delta | M_i, S_k, z_k] - \mathbb{E} [\Delta | S_k, z_k] \|_2 \text{pr} (M_i | S_k, z_k)$$

$$\| \mathbb{E} [\Delta | S_k, z_k] \|_2$$

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

$$\text{pr} (S_k) = \frac{\mathcal{E}_{k}^{-p}}{\sum_{k=1}^{K} \mathcal{E}_{k}^{-p}}, \quad k = 1, 2, \ldots, K$$

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

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

![Graph showing pr(S_k) over time](image)
Results: BMA prediction

(e) BMA predictions.

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

- Standard deviation reduced by approximately 30%.
- Prediction is brought closer to the validation data $z_V$. 
(g) BMA predictions.

(h) BMA std. dev., updated $\Pr(S_k)$.

- Uniform $\Pr(S_k)$ 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 → **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.
Part of this work can be found in: *Bayesian estimates of the parameter variability in the k – ε 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]

![Graphs showing boundary layer profiles and turbulence intensity](image)

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 → probability density function (pdf)
- Uses a set of observational data to infer a pdf of the closure coefficients → 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 $p(\theta|z)$ of model parameters $\theta$ conditioned on data $z$

$$p(\theta|z) = \frac{p(z|\theta)p(\theta)}{p(z)}$$  \hspace{1cm} (5)

- A framework able to incorporate multiple sources of uncertainty.
- The experimental observations $z$ also possess (measurement) uncertainties → data pdf $p(z)$
- $p(z|\theta)$ is known as the **likelihood**, i.e. the probability that the model will predict $z$ given $\theta$. $p(\theta)$ is the **prior** uncertainty in the model parameters. It represents a belief about $\theta$.
- Equation (5) is a statistical calibration, it infers the posterior pdf of the parameters $\theta$ that fits the model to the observations $z$. 
This the model-inadequacy term $\eta$ from Kennedy and O’Hagan, which is a means to represent the model error. Specify the statistical term for $\eta$ as $\eta \sim \text{GP}(1, c_\eta)$ with covariance function:

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

This implies a statistical model for the true process as:

$$\zeta | \theta, \gamma \sim \text{GP}(\mu_\zeta, c_\zeta)$$

$$\mu_\zeta(y^+ | \theta) = u^+(y^+, t; \theta)$$

$$c_\zeta(y^+, y^{+\prime} | \theta, \gamma) = u^+(y^+, t; \theta) \cdot c_\eta(y^+, y^{+\prime} | \gamma) \cdot u^+(y^{+\prime}, t; \theta).$$
The model-inadequacy term $\eta$ 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 $\eta$ 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 $\alpha$ and the magnitude by $\sigma$. 
Model-inadequacy term

- This becomes clear by drawing samples from $\eta$:
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 $p(\theta_u | z)$.
- The coefficients with low $S_u$ do not differ much from the uniform prior distribution $p(\theta_u)$. 
**HPD intervals**

HPD intervals of the $k - \omega$ model.
Some results

$\kappa$: HPD intervals of the Spalart-Allmaras model.
References

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.

A. Gelman, JB Carlin, HS Stern, and DB Rubin.
Bayesian data analysis. 1995.

Bayesian calibration of computer models.

G.B. Schubauer and PS Klebanoff.
Investigation of separation of the turbulent boundary layer.

B. Sudret.
Global sensitivity analysis using polynomial chaos expansions.

G. TANG, MS ELDRED, and L.P. SWILER.
Global sensitivity analysis for stochastic collocation expansion.