Bayesian strategies for calibrating heteroskedastic static sensors with unknown model structures

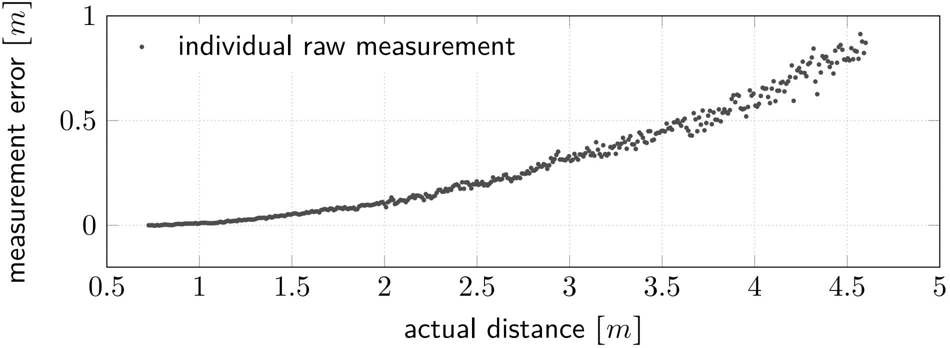
A. Alhashimi, S. Del Favero, D. Varagnolo, T. Gustafsson, G. Pillonetto

Luleå University of Technology

University of Padova

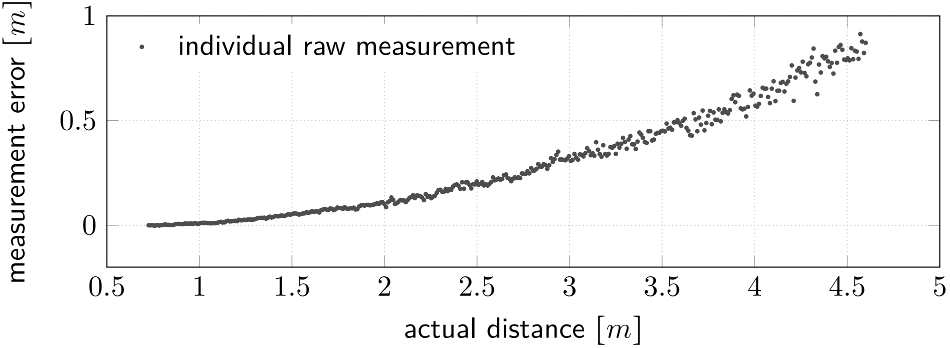
The problem in practice

How shall we calibrate a sensor that behaves in this way?



The problem in practice

How shall we calibrate a sensor that behaves in this way?



Rephrasing: how shall we account for:

- a systematic bias that smoothly depends on the measurand?
- a measurement noise whose variance also smoothly depends on the measurand?

The problem in practice – an illustrative example

The problem in formulas

$$y_i = f_{\mathrm{mean}}\left(x_i\right) + f_{\mathrm{noise}}\left(x_i\right) \tag{1}$$

The problem in formulas

$$y_i = f_{\text{mean}}\left(x_i\right) + f_{\text{noise}}\left(x_i\right) \tag{1}$$

$$f_{\text{mean}}\left(x_i\right) = \begin{bmatrix} 1 & x_i & x_i^2 & \dots & x_i^N \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} \tag{2}$$

The problem in formulas

$$y_i = f_{\text{mean}}\left(x_i\right) + f_{\text{noise}}\left(x_i\right) \tag{1}$$

$$f_{\text{mean}}\left(x_i\right) = \begin{bmatrix} 1 & x_i & x_i^2 & \ldots & x_i^N \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} \tag{2}$$

$$\boldsymbol{\alpha} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \Sigma_{\boldsymbol{\alpha}}\right) \qquad \boldsymbol{\mu}_{\boldsymbol{\alpha}} := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \end{bmatrix}^T \qquad \Sigma_{\boldsymbol{\alpha}} := \operatorname{diag}\left(\boldsymbol{\tau}_{\boldsymbol{\alpha}}^{-2}\right) \tag{3}$$

(assumption: $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$ and $\boldsymbol{\tau}_{\boldsymbol{\alpha}}$ known)

The problem in formulas – noise term

The problem in formulas – noise term

$$\begin{aligned}
\text{Case I:} \quad & f_{\text{noise}}(x_i) &= \sigma_\nu \\
\text{Case II:} \quad & f_{\text{noise}}(x_i) &= \sigma_\nu x_i^\rho \\
\text{Case III:} \quad & f_{\text{noise}}(x_i) &= \sigma_\nu f_{\text{mean}}(x_i)^\rho
\end{aligned} \tag{4}$$

The problem in formulas – noise term

$$
\begin{aligned}
\text{Case I:} \quad & f_{\text{noise}}(x_i) && = \sigma_\nu \\
\text{Case II:} \quad & f_{\text{noise}}(x_i) && = \sigma_\nu x_i^\rho \\
\text{Case III:} \quad & f_{\text{noise}}(x_i) && = \sigma_\nu f_{\text{mean}}(x_i)^\rho
\end{aligned} \tag{4}
$$

What do these models represent?

Case I: homoskedastic sensors

Case II: heteroskedasticity depending on the actual state

Case III: heteroskedasticity depending on the expected measurement

The problem in formulas – noise term

$$
\begin{aligned}
\text{Case I:} \quad & f_{\text{noise}}(x_i) &= \sigma_\nu \\
\text{Case II:} \quad & f_{\text{noise}}(x_i) &= \sigma_\nu x_i^\rho \\
\text{Case III:} \quad & f_{\text{noise}}(x_i) &= \sigma_\nu f_{\text{mean}}(x_i)^\rho
\end{aligned}
\tag{4}
$$

What do these models represent?

Case I: homoskedastic sensors
Case II: heteroskedasticity depending on the actual state
Case III: heteroskedasticity depending on the expected measurement

Assumed priors

- $\sigma_\nu^{-2} = \tau_\nu \sim \text{Gamma}(a_\nu, b_\nu)$
- $\rho \in \mathcal{N}^+(a_\rho, b_\rho)$

The problem in formulas – summary

Given

$$y_i = \begin{bmatrix} 1 & \dots & x_i^N \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_N \end{bmatrix} + \begin{cases} \sigma_\nu \\ \sigma_\nu x_i^\rho \\ \sigma_\nu \left(\begin{bmatrix} 1 & \dots & x_1^N \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_N \end{bmatrix} \right)^\rho \end{cases}$$

a dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^M$ and opportune priors, estimate

- $\boldsymbol{\alpha}$
- σ_ν
- ρ

literature review

Literature review

$$y_i = \begin{bmatrix} 1 & \dots & x_i^N \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_N \end{bmatrix} + \begin{cases} \sigma_\nu \\ \sigma_\nu x_i^\rho \\ \sigma_\nu \left(\begin{bmatrix} 1 & \dots & x_1^N \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_N \end{bmatrix} \right)^\rho \end{cases}$$

ordinary least squares $\implies$ unbiased estimate of the mean, biased estimate of the variance (that worsens with the degree of heteroskedasticity)

Box & Hill (1974)
Correcting inhomogeneity of variance with power transformation weighting
Technometrics

White (1980)
A heteroskedasticity-consistent cov. matrix estimator and a direct test for heteroskedasticity
Econometrica: Journal of the Econometric Society

Literature review

$$y_i = \begin{bmatrix} 1 & \dots & x_i^N \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_N \end{bmatrix} + \begin{cases} \sigma_\nu \\ \sigma_\nu x_i^\rho \\ \sigma_\nu \left(\begin{bmatrix} 1 & \dots & x_1^N \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_N \end{bmatrix} \right)^\rho \end{cases}$$

other schemes focusing on simplified models $\implies$ Gibbs samplers, MCMC schemes

Geweke (1993)
Bayesian treatment of the independent Student-t linear model
Journal of applied econometrics

Boscardin & Gelman (1994)
Bayesian computation for parametric models of heteroscedasticity in the linear model
TODO

Tanizaki & Zhang (2001)
Posterior analysis of the multiplicative heteroscedasticity model
TODO

Our contributions

- slightly more generic model *(unknown ρ)*
- use exact likelihoods instead of approximated ones
- create a stepping stone for schemes where also the x_i's are unknown

the calibration algorithms

disclaimer: the models (and associated calibration procedures) are meaningful only for static sensors

Case I: $f_{\mathrm{noise}}(x_i) = \sigma_\nu$

Case I: $f_{\text{noise}}(x_i) = \sigma_\nu$

problem: find the MAP for $\boldsymbol{\alpha}$, with the additional complexity that σ_ν is unknown

Case I: $f_{\text{noise}}(x_i) = \sigma_\nu$

problem: find the MAP for $\boldsymbol{\alpha}$, with the additional complexity that σ_ν is unknown

solution: Gibbs sampler, since we know the expressions of the conditional distributions & all the priors and likelihoods are conjugate

Algorithm for Case I: $f_{\text{noise}}(x_i) = \sigma_\nu$

1. initialization: $\boldsymbol{\alpha}^{(0)} = \boldsymbol{\mu}_{\boldsymbol{\alpha}}$ $\quad \tau_\nu^{(0)} \sim \text{Gamma}\,(a_\nu, b_\nu)$
2. for $k = 0, 1, \ldots$ up to convergence or k_{max}:
   1. update τ_ν and $\boldsymbol{\alpha}$ using Gibbs sampling:

$$
\begin{aligned}
\boldsymbol{\alpha}^{(k+1)} &\sim p\left(\boldsymbol{\alpha}^{(k)} \,\middle|\, \boldsymbol{x}, \tau_\nu^{(k)}\right) \\
\tau_\nu^{(k+1)} &\sim p\left(\tau_\nu^{(k)} \,\middle|\, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\alpha}^{(k+1)}\right)
\end{aligned} \tag{5}
$$

where:

$$
\begin{aligned}
&p\left(\boldsymbol{\alpha}^{(k)} \,\middle|\, \boldsymbol{x}, \boldsymbol{y}, \tau_\nu^{(k)}\right) \propto \mathcal{N}\left(B^{(k)} A^{(k)}, B^{(k)}\right) \\
&A^{(k)} = \tau_\nu^{(k)} G_{\boldsymbol{x}}^T \boldsymbol{y} - \Sigma_{\boldsymbol{\alpha}}^{-1} \boldsymbol{\mu}_{\boldsymbol{\alpha}} \\
&B^{(k)} = \left(\tau_\nu^{(k)} G_{\boldsymbol{x}}^T G_{\boldsymbol{x}} + \Sigma_{\boldsymbol{\alpha}}^{-1}\right)^{-1} \\
&p\left(\tau_\nu^{(k)} \,\middle|\, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\alpha}^{(k+1)}\right) \propto \text{Gamma}\left(a_\nu + \frac{M}{2}, \left(\frac{1}{b_\nu} + \frac{1}{2} {C^{(k+1)}}^T C^{(k+1)}\right)^{-1}\right) \\
&C^{(k+1)} = \left(\boldsymbol{y} - G_{\boldsymbol{x}} \boldsymbol{\alpha}^{(k+1)}\right)
\end{aligned} \tag{6}
$$

Case II: $f_{\text{noise}}(x_i) = \sigma_\nu x_i^\rho$

example: *MAP* $\Longrightarrow$ $$\arg \max_{\boldsymbol{\alpha}\in\mathbb{R}^N \; \sigma_\nu^2\in\mathbb{R}_+ \; \rho\in\mathbb{R}_+} p\left(\boldsymbol{\alpha}, \sigma_\nu^2, \rho \,|\, \boldsymbol{x}, \boldsymbol{y}\right) \tag{7}$$

Case II: $f_{\text{noise}}(x_i) = \sigma_\nu x_i^\rho$

example: *MAP* $\Longrightarrow$ $$\arg \max_{\boldsymbol{\alpha} \in \mathbb{R}^N \; \sigma_\nu^2 \in \mathbb{R}_+ \; \rho \in \mathbb{R}_+} p\left(\boldsymbol{\alpha}, \sigma_\nu^2, \rho \,|\, \boldsymbol{x}, \boldsymbol{y}\right) \tag{7}$$

problem: now both σ_ν and ρ are unknown *(implying that also* $p(\rho \,|\, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\alpha}, \tau_\nu)$ *is unknown)*

Case II: $f_{\text{noise}}(x_i) = \sigma_\nu x_i^\rho$

example: *MAP* $\Longrightarrow$ $$\arg \max_{\boldsymbol{\alpha} \in \mathbb{R}^N \ \sigma_\nu^2 \in \mathbb{R}_+ \ \rho \in \mathbb{R}_+} p\left(\boldsymbol{\alpha}, \sigma_\nu^2, \rho \,|\, \boldsymbol{x}, \boldsymbol{y}\right) \tag{7}$$

problem: now both σ_ν and ρ are unknown *(implying that also $p\left(\rho \,|\, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\alpha}, \tau_\nu\right)$ is unknown)*

solution: Single-Component Metropolis-Hastings scheme

Algorithm for Case II: $f_{\text{noise}}(x_i) = \sigma_\nu x_i^\rho$

1. initialization: $\boldsymbol{\alpha}^{(0)} = \boldsymbol{\mu}_{\boldsymbol{\alpha}}$ $\qquad \tau_\nu^{(0)} \sim \text{Gamma}(a_\nu, b_\nu)$ $\qquad \rho^{(0)} = 0$
2. for $k = 0, 1, \ldots$ up to convergence or k_{max}:
   1. update τ_ν and $\boldsymbol{\alpha}$ using the Gibbs sampler:
   $$\begin{aligned} \boldsymbol{\alpha}^{(k+1)} &\sim p\left(\boldsymbol{\alpha}^{(k)} \middle| \boldsymbol{x}, \tau_\nu^{(k)}, \rho^{(k)}\right) \\ \tau_\nu^{(k+1)} &\sim p\left(\tau_\nu^{(k)} \middle| \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\alpha}^{(k+1)}, \rho^{(k)}\right) \end{aligned} \tag{8}$$
   2. generate a new proposal:
   $$\rho^{(k+1)} \sim \mathcal{N}\left(\rho^{(k)}, \beta\right) \tag{9}$$
   3. calculate the acceptance probability:
   $$\gamma = \min\left[1, \frac{p\left(\boldsymbol{y} \middle| \boldsymbol{x}, \rho^{(k+1)}, \boldsymbol{\alpha}^{(k+1)}, \tau_\nu^{(k+1)}\right)}{p\left(\boldsymbol{y} \middle| \boldsymbol{x}, \rho^{(k)}, \boldsymbol{\alpha}^{(k+1)}, \tau_\nu^{(k+1)}\right)} \frac{p\left(\rho^{(k+1)}\right)}{p\left(\rho^{(k)}\right)}\right] \tag{10}$$
   4. accept the new proposal if $\gamma > \mathcal{U}[0,1]$ and $0 \le \rho \le 10$

Case III: $f_{\text{noise}}(x_i) = \sigma_\nu f_{\text{mean}}(x_i)^\rho$

$$f_{\text{mean}}(x_i) = \begin{bmatrix} 1 & \dots & x_i^N \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_N \end{bmatrix} \qquad \Sigma_\nu := \sigma_\nu^2 \operatorname{diag}\Big(f_{\text{mean}}(x_1)^{2\rho}, \dots, f_{\text{mean}}(x_M)^{2\rho} \Big)$$

Case III: $f_{\text{noise}}(x_i) = \sigma_\nu f_{\text{mean}}(x_i)^\rho$

$$f_{\text{mean}}(x_i) = \begin{bmatrix} 1 & \dots & x_i^N \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_N \end{bmatrix} \qquad \Sigma_\nu := \sigma_\nu^2 \operatorname{diag}\Big(f_{\text{mean}}(x_1)^{2\rho}, \dots, f_{\text{mean}}(x_M)^{2\rho} \Big)$$

problem: now not only $p(\rho \,|\, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\alpha}, \tau_\nu)$, but also $p(\boldsymbol{\alpha} \,|\, \boldsymbol{x}, \boldsymbol{y}, \tau_\nu, \rho)$ is unknown

Case III: $f_{\text{noise}}(x_i) = \sigma_\nu f_{\text{mean}}(x_i)^\rho$

$$f_{\text{mean}}(x_i) = \begin{bmatrix} 1 & \dots & x_i^N \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_N \end{bmatrix} \qquad \Sigma_\nu := \sigma_\nu^2 \operatorname{diag}\Big(f_{\text{mean}}(x_1)^{2\rho}, \dots, f_{\text{mean}}(x_M)^{2\rho} \Big)$$

problem: now not only $p(\rho \,|\, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\alpha}, \tau_\nu)$, but also $p(\boldsymbol{\alpha} \,|\, \boldsymbol{x}, \boldsymbol{y}, \tau_\nu, \rho)$ is unknown

solution: use acceptance/rejection mechanisms also for $\boldsymbol{\alpha}$

Algorithm for Case III: $f_{\text{noise}}(x_i) = \sigma_\nu f_{\text{mean}}(x_i)^\rho$

1. initialization: $\boldsymbol{\alpha}^{(0)} = \boldsymbol{\mu}_{\boldsymbol{\alpha}}$ $\quad \tau_\nu^{(0)} \sim \text{Gamma}(a_\nu, b_\nu)$ $\quad \rho^{(0)} = 0$
2. for $k = 0, 1, \ldots$ up to convergence or k_{max}:
   1. update τ_ν using the Gibbs sampler:
   $$\tau_\nu^{(k+1)} \sim p\left(\tau_\nu^{(k)} \middle| \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\alpha}^{(k)}\right) \tag{11}$$
   2. generate the new proposals:
   $$\boldsymbol{\alpha}^{(k+1)} \sim \mathcal{N}\left(\boldsymbol{\alpha}^{(k)}, \beta\right) \qquad \rho^{(k+1)} \sim \mathcal{N}\left(\rho^{(k)}, \beta'\right) \tag{12}$$
   3. calculate the acceptance probability:
   $$\gamma = \min\left[1, \frac{p\left(\boldsymbol{y} \middle| \boldsymbol{x}, \rho^{(k+1)}, \boldsymbol{\alpha}^{(k+1)}, \tau_\nu^{(k+1)}\right)}{p\left(\boldsymbol{y} \middle| \boldsymbol{x}, \rho^{(k)}, \boldsymbol{\alpha}^{(k)}, \tau_\nu^{(k+1)}\right)} \frac{p\left(\rho^{(k+1)}\right) p\left(\boldsymbol{\alpha}^{(k+1)}\right)}{p\left(\rho^{(k)}\right) p\left(\boldsymbol{\alpha}^{(k)}\right)}\right] \tag{13}$$
   4. accept the new proposal if $\gamma > \mathcal{U}[0, 1]$ and $0 \le \rho \le 10$

Recap

- $f_{\text{noise}}(x_i) = \sigma_\nu \implies$ we know all the conditional distributions $\implies$ we can use Gibbs samplers for $\boldsymbol{\alpha}$ and τ_ν

Recap

- $f_{\text{noise}}(x_i) = \sigma_\nu \implies$ we know all the conditional distributions $\implies$ we can use Gibbs samplers for $\boldsymbol{\alpha}$ and τ_ν

- $f_{\text{noise}}(x_i) = \sigma_\nu x_i^\rho \implies$ we don't know the conditional distribution for $\rho \implies$ we shall use Gibbs samplers for $\boldsymbol{\alpha}$ and τ_ν, but a MH sampler for ρ

Recap

- $f_{\text{noise}}(x_i) = \sigma_\nu \implies$ we know all the conditional distributions $\implies$ we can use Gibbs samplers for $\boldsymbol{\alpha}$ and τ_ν

- $f_{\text{noise}}(x_i) = \sigma_\nu x_i^\rho \implies$ we don't know the conditional distribution for $\rho \implies$ we shall use Gibbs samplers for $\boldsymbol{\alpha}$ and τ_ν, but a MH sampler for ρ

- $f_{\text{noise}}(x_i) = \sigma_\nu f_{\text{mean}}(x_i)^\rho \implies$ we don't know the conditional distributions for $\boldsymbol{\alpha}$ and $\rho \implies$ we shall use a Gibbs sampler for τ_ν, and MH samplers for $\boldsymbol{\alpha}$ and ρ

Test case: artificial setup

$$f_{\text{mean}}(x_i) = \sum_{n=0}^{3} \alpha_n x_i^n \qquad f_{\text{noise}}(x_i) = \sigma_\nu f_{\text{mean}}(x_i)^\rho$$

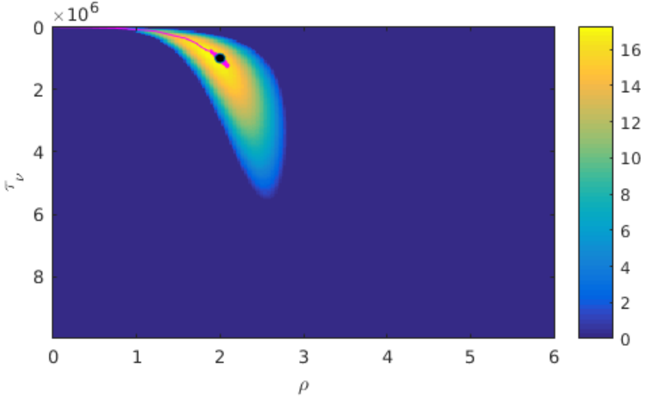
Test case: artificial setup

$$f_{\text{mean}}(x_i) = \sum_{n=0}^{3} \alpha_n x_i^n \qquad f_{\text{noise}}(x_i) = \sigma_\nu f_{\text{mean}}(x_i)^\rho$$

how does $p(\rho, \tau_\nu | \boldsymbol{x}, \boldsymbol{y})$ look like?

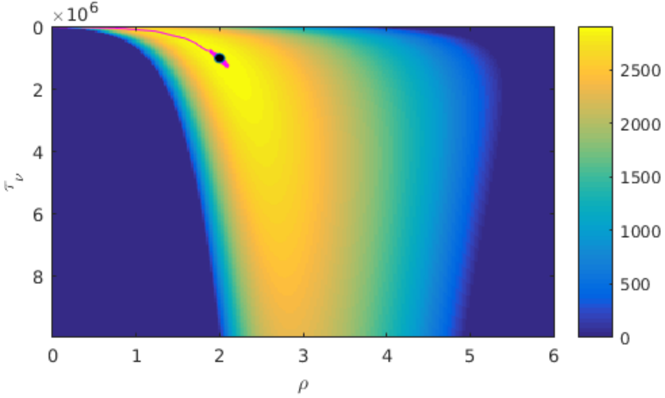
Test case: artificial setup - how does $p\left(\rho, \tau_\nu \middle| \boldsymbol{x}, \boldsymbol{y}\right)$ look like? ($M = 50$)

$$f_{\text{mean}}(x_i) = \sum_{n=0}^{3} \alpha_n x_i^n \qquad f_{\text{noise}}(x_i) = \sigma_\nu f_{\text{mean}}(x_i)^\rho$$

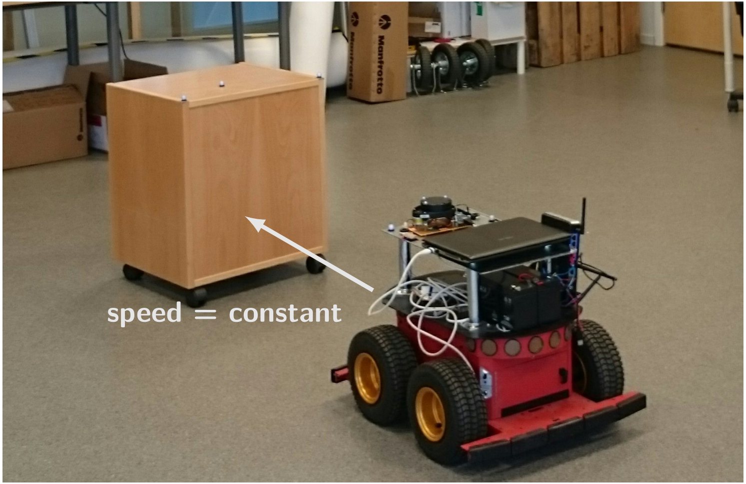


Test case: artificial setup - how does $p(\rho, \tau_\nu | \boldsymbol{x}, \boldsymbol{y})$ look like? ($M = 900$)

$$f_{\text{mean}}(x_i) = \sum_{n=0}^{3} \alpha_n x_i^n \qquad f_{\text{noise}}(x_i) = \sigma_\nu f_{\text{mean}}(x_i)^\rho$$

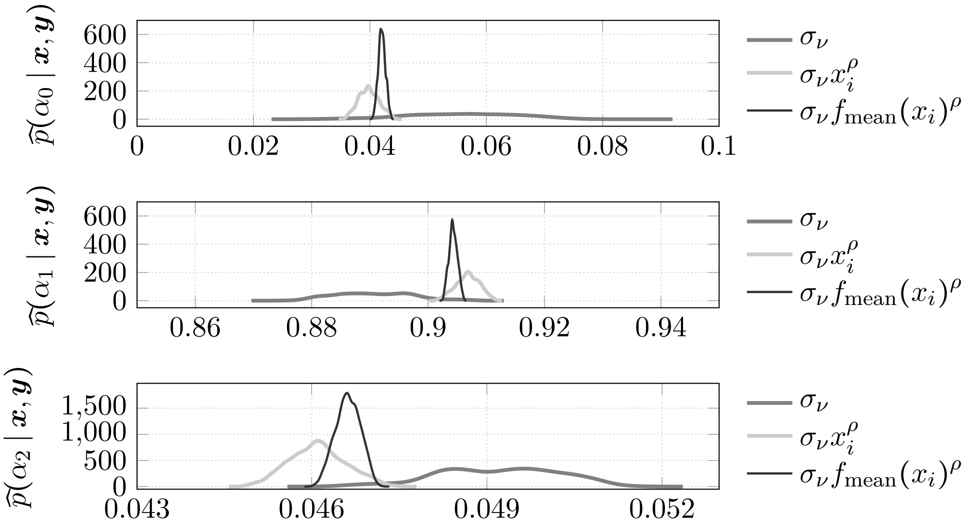


Test case: experimental setup



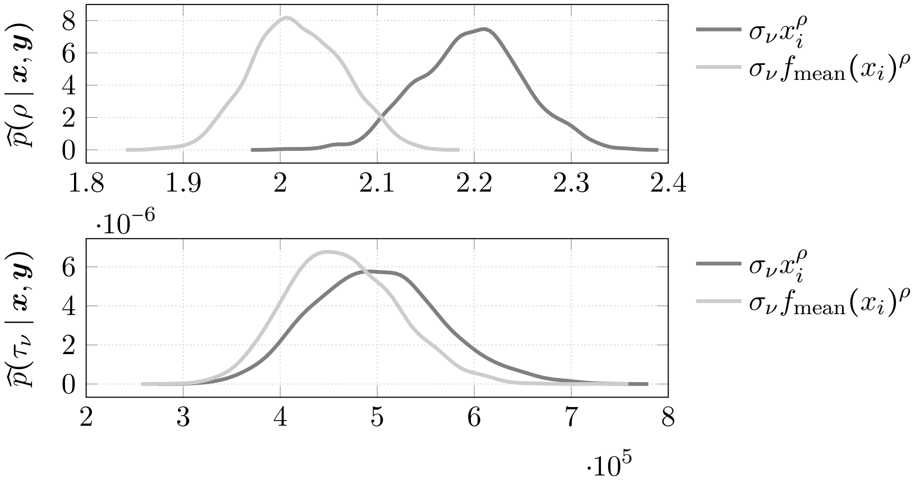
Test case: experimental results

Posteriors for α_0, α_1 and α_2



Test case: experimental results

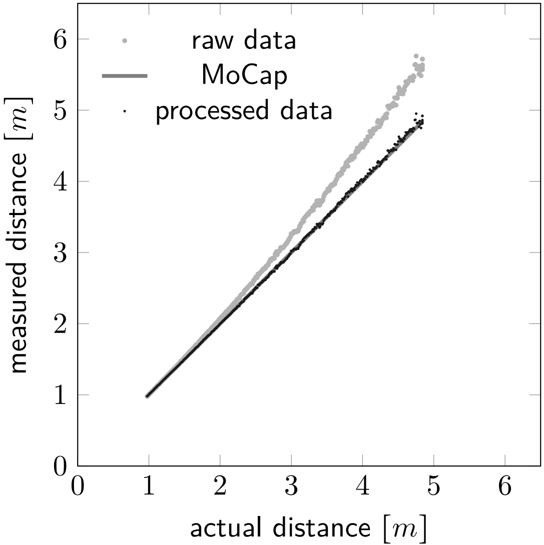
Posteriors for ρ and τ_ν



How shall we use the estimates?

$$\text{e.g.,} \quad \widehat{x}_i = \arg\max_{x_k \in \mathcal{X}} p\left(x_i \,|y_i, \boldsymbol{\alpha}, \sigma_\nu, \rho\right) \tag{14}$$

Test case: experimental results



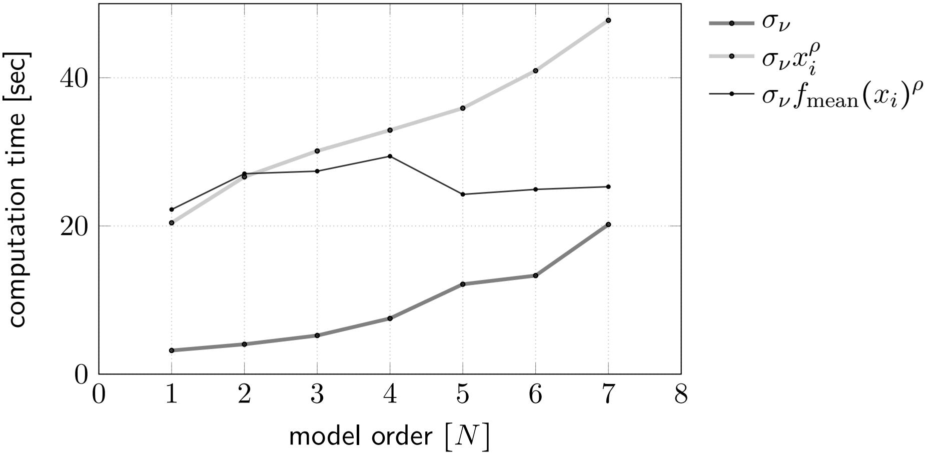
Test case: experimental results

MSE performance on the test set

N	σ_ν	$\sigma_\nu x_i^\rho$	$\sigma_\nu f_{\text{mean}}(x_i)^\rho$
1	1397.59261	50.14214	3220.53529
2	3.15795	0.27043	0.02243
3	0.49000	0.00507	**0.00185**
4	0.48642	0.00404	**0.00088**
5	0.48714	0.00220	**0.00092**
6	0.48675	0.00229	0.01049
7	0.48754	0.00285	0.45820

For completeness: computational times for estimating the models

Matlab on a standard laptop (Intel quad core i7-2640 CPUs 2.8GHz)



Conclusions

- heteroskedastic measurement noise + polynomial bias $\Longrightarrow$ great flexibility
- price: need for "advanced" estimation schemes
- Bayesian approach enables exploiting prior information
- meaningful results on both synthetic and field usecases

Conclusions

- heteroskedastic measurement noise + polynomial bias $\implies$ great flexibility
- price: need for "advanced" estimation schemes
- Bayesian approach enables exploiting prior information
- meaningful results on both synthetic and field usecases

Next (ongoing) step: what if the x_i's are unknown?

$$y_i = \begin{bmatrix} 1 & \dots & x_i^N \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_N \end{bmatrix} + \begin{cases} \sigma_\nu \\ \sigma_\nu x_i^\rho \\ \sigma_\nu \left(\begin{bmatrix} 1 & \dots & x_1^N \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_N \end{bmatrix} \right)^\rho \end{cases}$$

Bayesian strategies for calibrating heteroskedastic static sensors with unknown model structures

A. Alhashimi, S. Del Favero, D. Varagnolo, T. Gustafsson, G. Pillonetto

Luleå University of Technology

University of Padova