tu dortmund university

computational statistics

Simulation-Based Prior Knowledge Elicitation for Parametric Bayesian Models

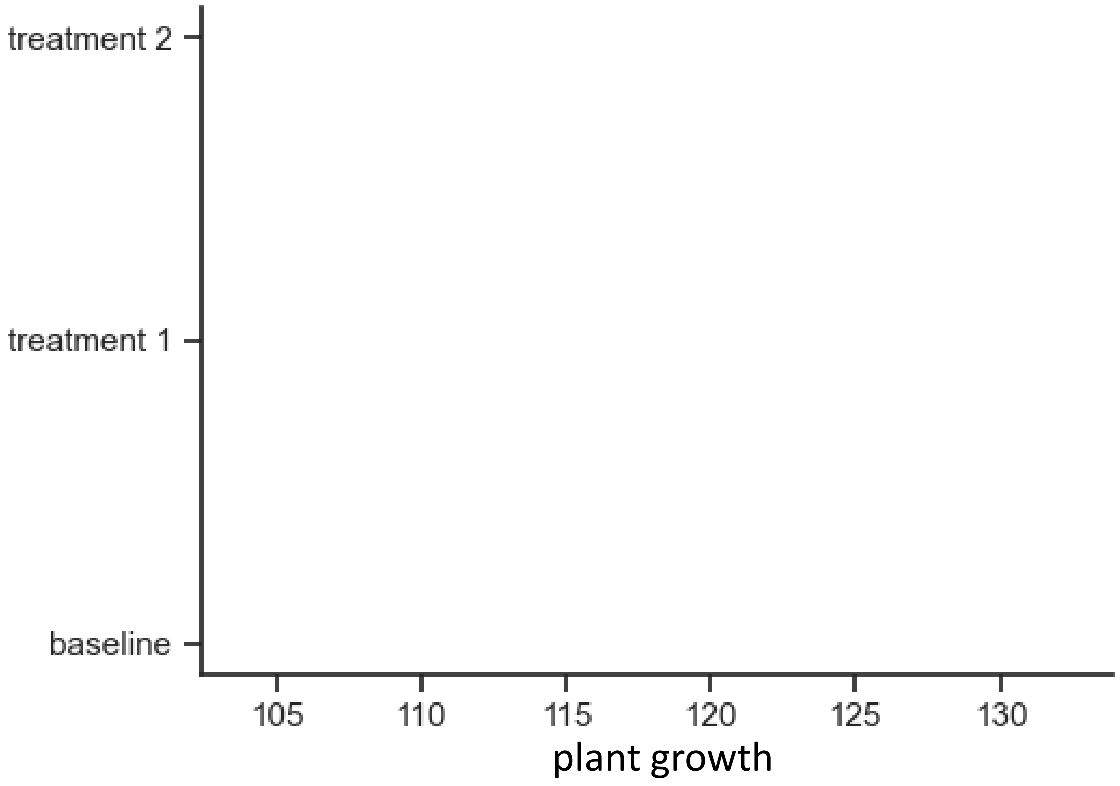
Florence Bockting
Stefan T. Radev
Paul-Christian Bürkner

ISBA 2024

Introductory example

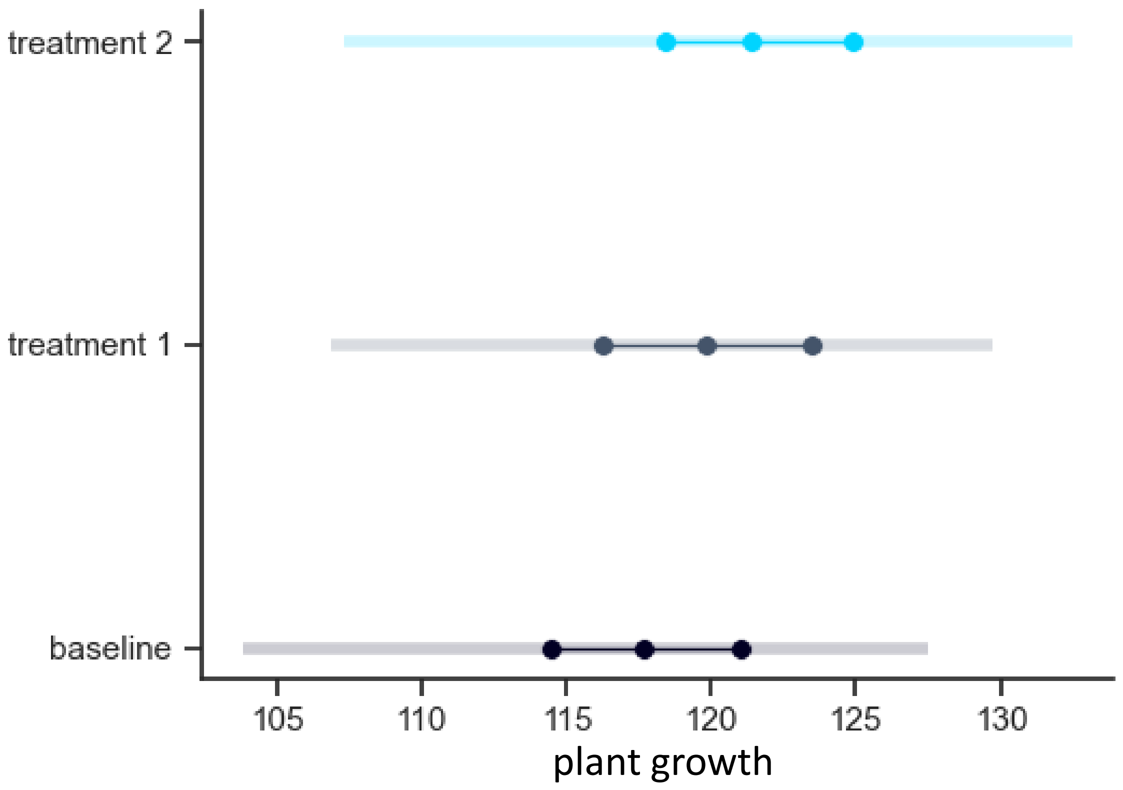
One factorial design with three levels

- Investigate treatment-effect on some dependent variable (e.g. plant growth)
  - treatment 1
  - treatment 2
  - control



Introductory example
Expert expectations

- Investigate treatment-effect on some dependent variable (e.g. plant growth)
  - treatment 1
  - treatment 2
  - control



Expert assumptions:

- treatment 1,2 $\geq$ control
- treatment 1 $\leq$ treatment 2

florence.bockting@tu-dortmund.de

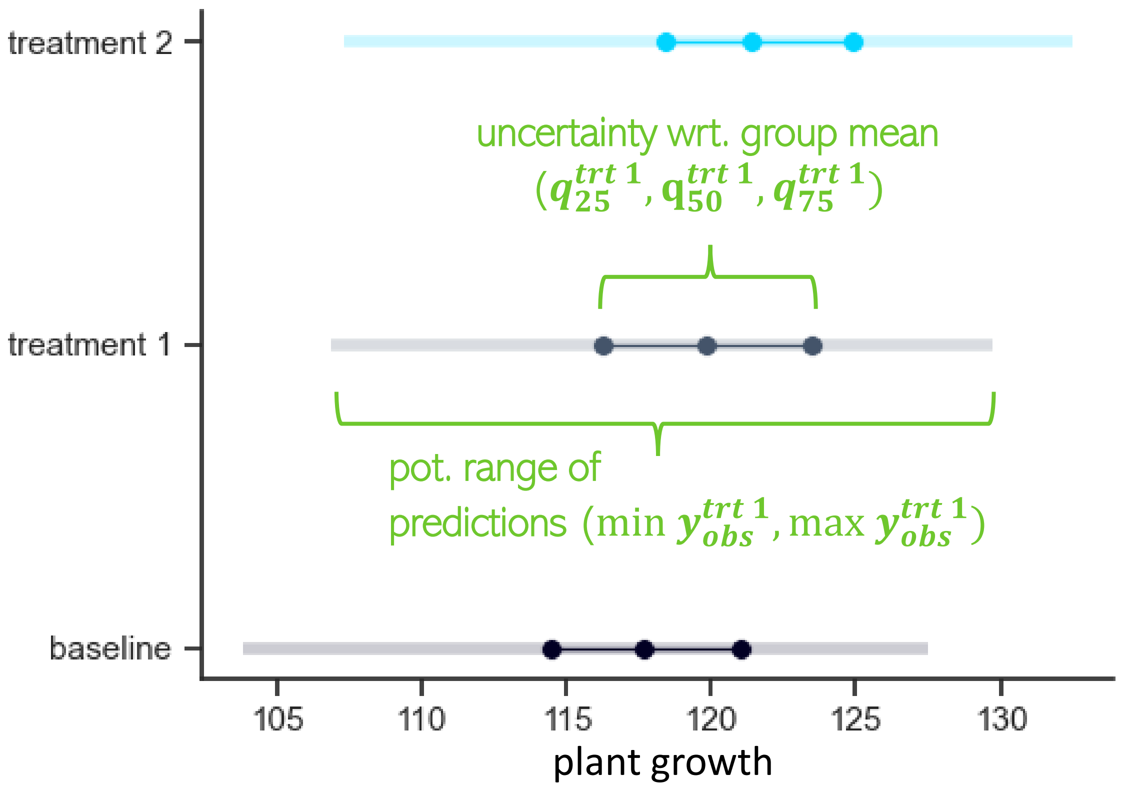
Introductory example

Expert expectations

- Investigate treatment-effect on some dependent variable (e.g. plant growth)
  - treatment 1
  - treatment 2
  - control

Expert assumptions:

- treatment 1,2 $\geq$ control
- treatment 1 $\leq$ treatment 2



florence.bockting@tu-dortmund.de

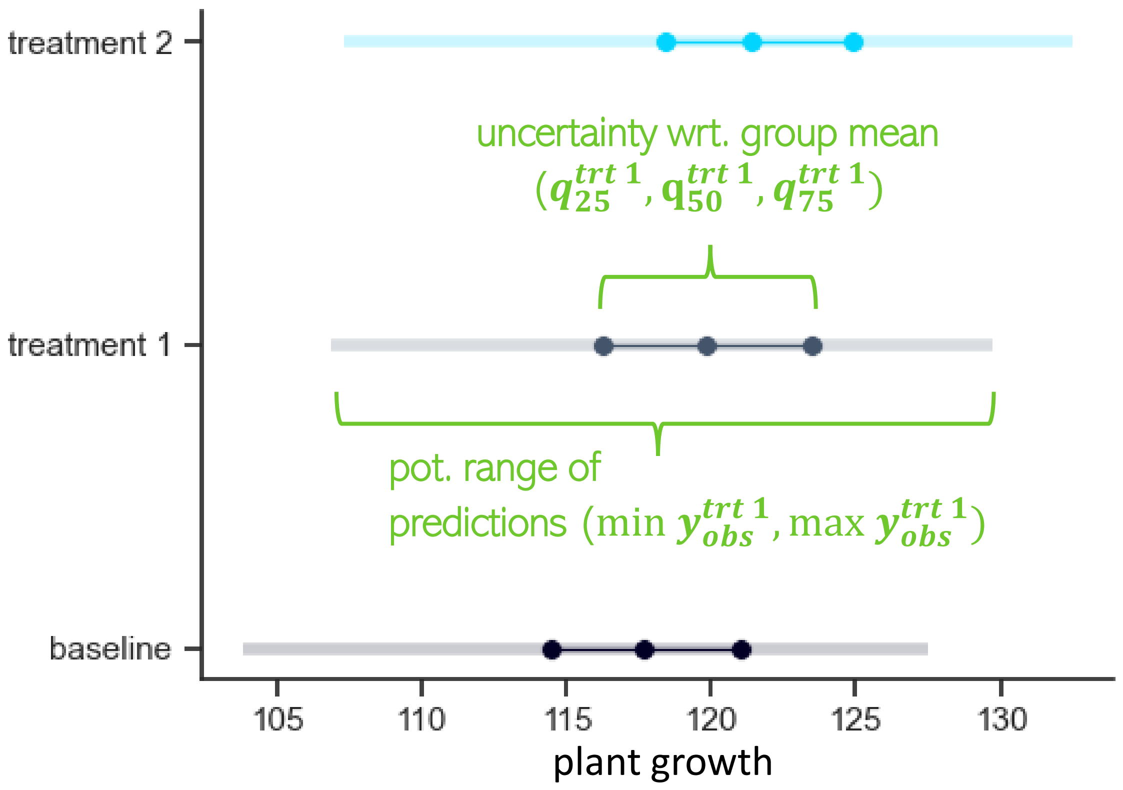
Introductory example

Statistical model

$$\beta_0 \sim \text{Normal}(\mu_0, \sigma_0)$$
$$\beta_1 \sim \text{Normal}(\mu_1, \sigma_1)$$
$$\beta_2 \sim \text{Normal}(\mu_2, \sigma_2)$$
$$s \sim \text{Normal}^+(\sigma)$$

$$\theta_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i}$$

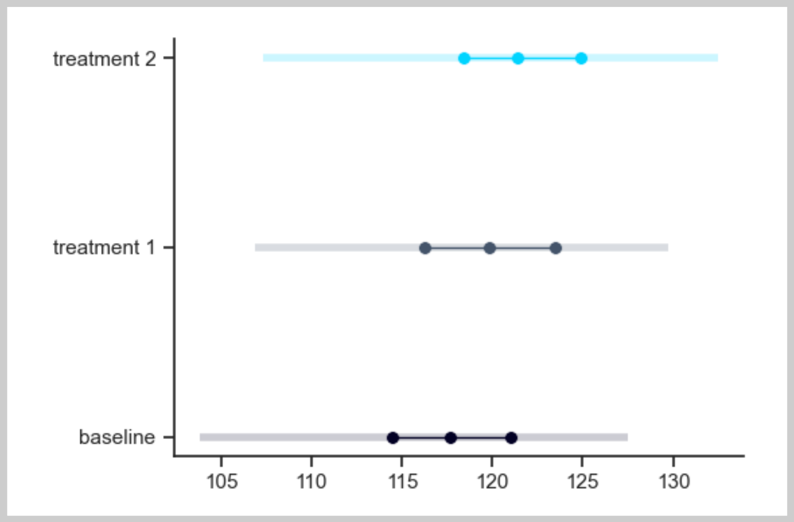
$$y_i \sim \text{Normal}(\theta_i, s)$$



florence.bockting@tu-dortmund.de

The problem

Translate expert beliefs into corresponding priors



$$\beta_0 \sim \text{Normal}(\mu_0, \sigma_0)$$
$$\beta_1 \sim \text{Normal}(\mu_1, \sigma_1)$$
$$\beta_2 \sim \text{Normal}(\mu_2, \sigma_2)$$
$$s \sim \text{Normal}^+(\sigma)$$

$$\theta_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i}$$

$$y_i \sim \text{Normal}(\theta_i, s)$$

?

The problem is actually not new

Expert prior elicitation has a long history

- Recent review: Mikkola et al. (2023)
- Historically, methods focused on model parameters
- Recent shift to methods that focus on prior predictive distribution
  - e.g., da Silva et al. (2019); Hartmann et al. (2020); Manderson & Goudie (2023)

query information from expert	translate →	prior distribution $\boldsymbol{\beta} \sim \boldsymbol{p}_{\boldsymbol{\lambda}}$
Parameter space		Parameter space
Observable space		Parameter space
Parameter/ Observable space		Parameter space (our method)

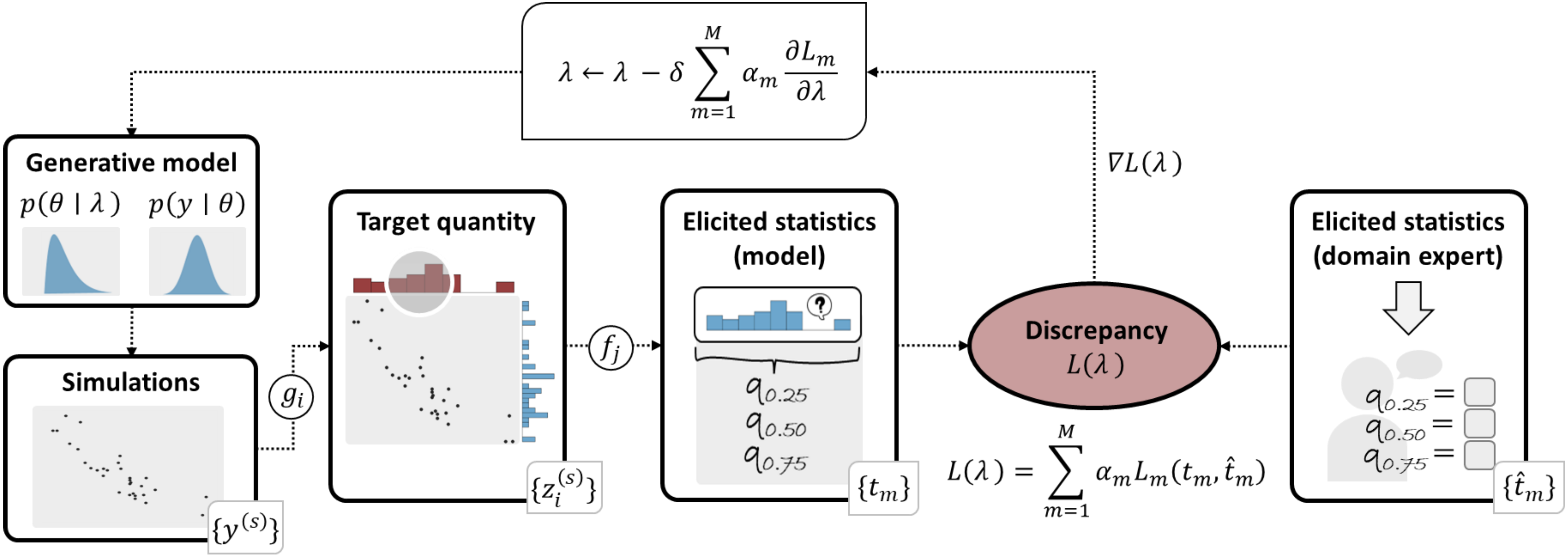
florence.bockting@tu-dortmund.de

Our contribution to the problem

Overview of our prior elicitation method

dortmund university computational statistics

Bockting, F., Radev, S. T., & Bürkner, P. C. (2024). Simulation-Based Prior Knowledge Elicitation for Parametric Bayesian Models. *arXiv preprint arXiv:2308.11672*.

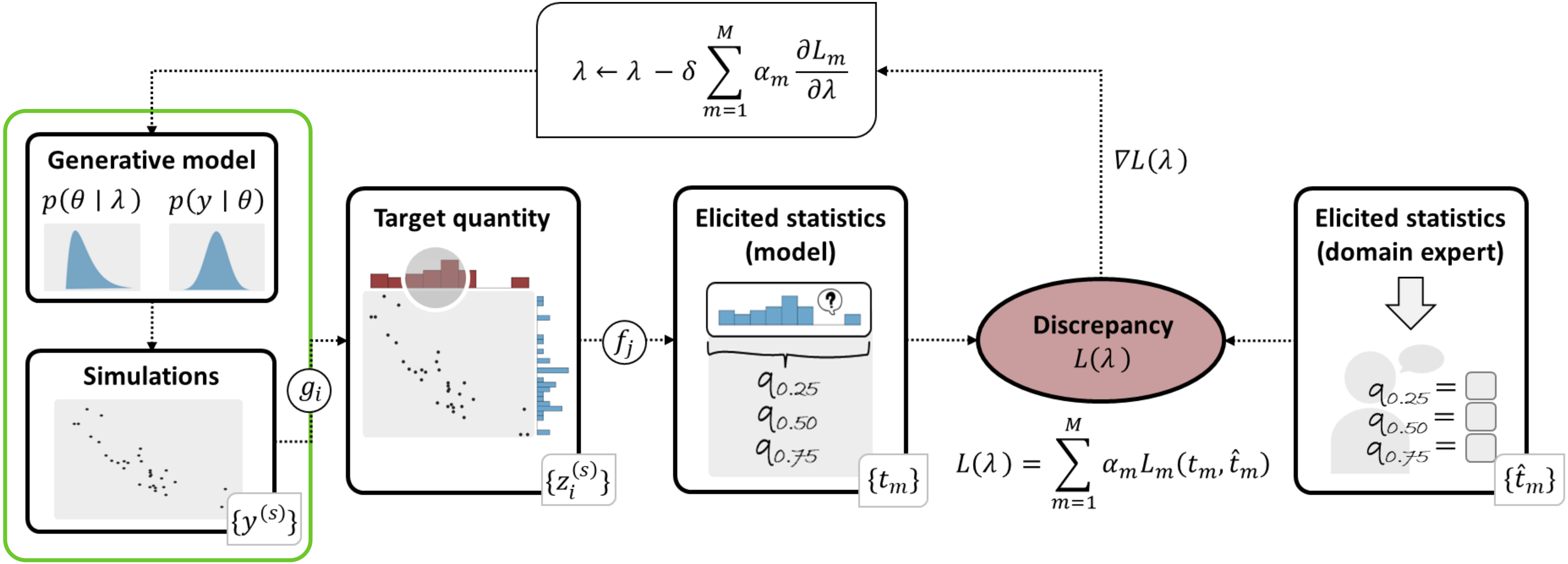


florence.bockting@tu-dortmund.de

Our contribution to the problem

Overview of our prior elicitation method

Bockting, F., Radev, S. T., & Bürkner, P. C. (2024). Simulation-Based Prior Knowledge Elicitation for Parametric Bayesian Models. *arXiv preprint arXiv:2308.11672*.



A closer look into our method

Initialize hyperparameter vector

tu dortmund university
computational statistics

$$\beta_0 \sim \text{Normal}(\mu_0, \sigma_0)$$
$$\beta_1 \sim \text{Normal}(\mu_1, \sigma_1)$$
$$\beta_2 \sim \text{Normal}(\mu_2, \sigma_2)$$
$$s \sim \text{Normal}^+(\sigma)$$

$$\theta_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i}$$

$$y_i \sim \text{Normal}(\theta_i, s)$$

florence.bockting@tu-dortmund.de

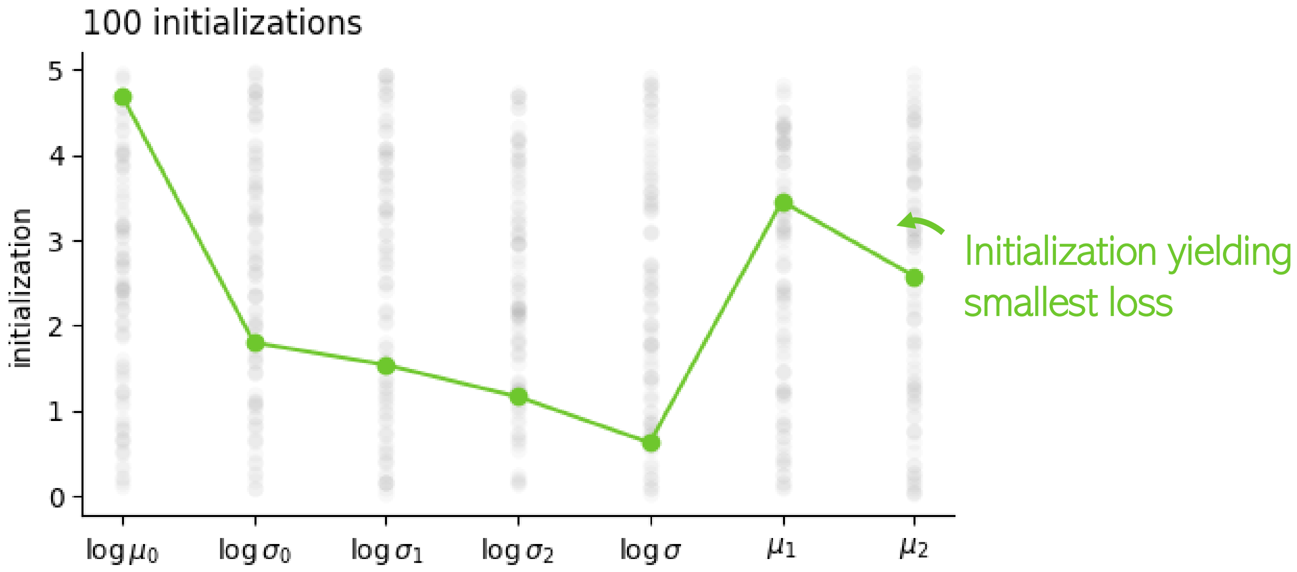
A closer look into our method

Initialize hyperparameter vector

$$\beta_0 \sim \text{Normal}(\mu_0, \sigma_0)$$
$$\beta_1 \sim \text{Normal}(\mu_1, \sigma_1)$$
$$\beta_2 \sim \text{Normal}(\mu_2, \sigma_2)$$
$$s \sim \text{Normal}^+(\sigma)$$

$$\theta_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i}$$

$$y_i \sim \text{Normal}(\theta_i, s)$$



florence.bockting@tu-dortmund.de

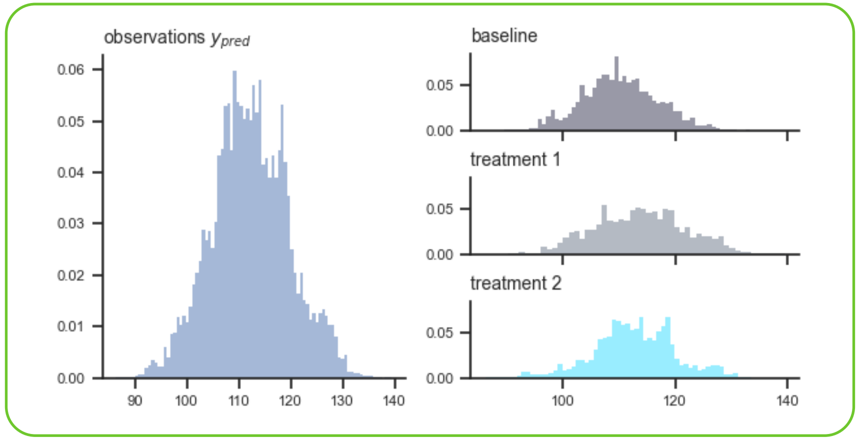
A closer look into our method

Initialize hyperparameter vector

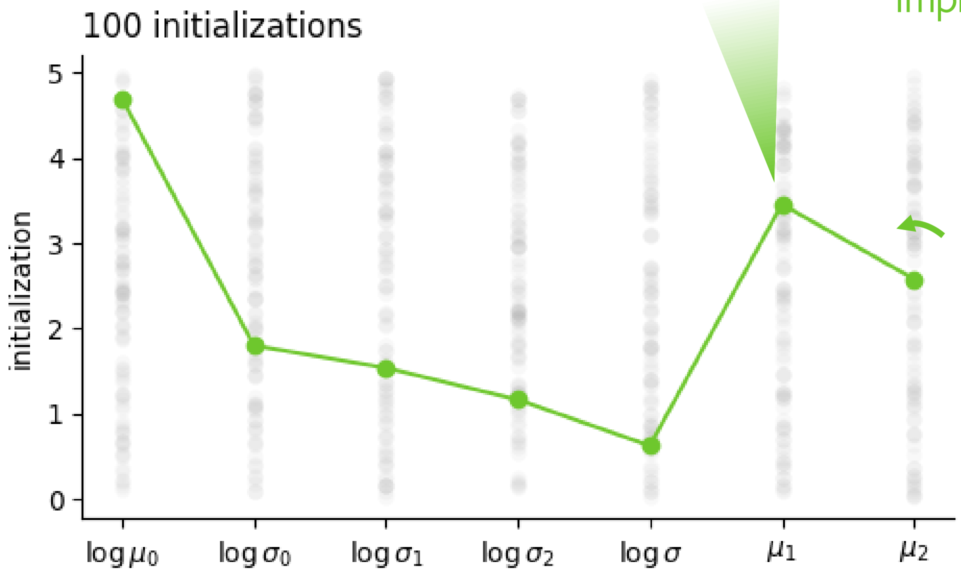
$$\beta_0 \sim \text{Normal}(\mu_0, \sigma_0)$$
$$\beta_1 \sim \text{Normal}(\mu_1, \sigma_1)$$
$$\beta_2 \sim \text{Normal}(\mu_2, \sigma_2)$$
$$s \sim \text{Normal}^+(\sigma)$$

$$\theta_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i}$$

$$y_i \sim \text{Normal}(\theta_i, s)$$



Implied prior predictions



Initialization yielding smallest loss

florence.bockting@tu-dortmund.de

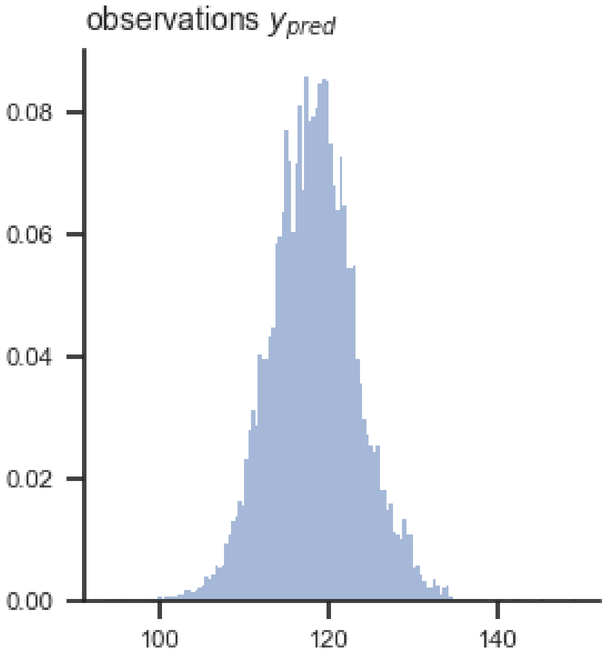
A closer look into our method

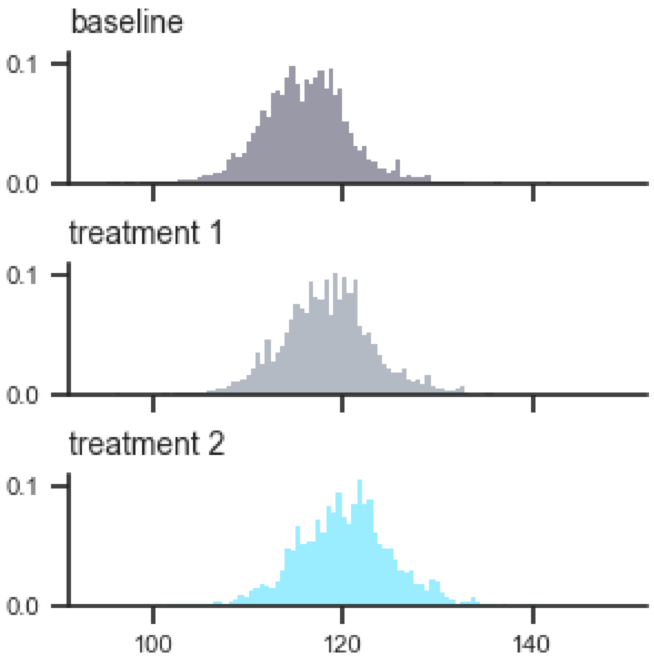
Simulate from the generative model …

$$\beta_0 \sim \text{Normal}\left(\mu_0^{ini}, \sigma_0^{ini}\right)$$
$$\beta_1 \sim \text{Normal}\left(\mu_1^{ini}, \sigma_1^{ini}\right)$$
$$\beta_2 \sim \text{Normal}\left(\mu_2^{ini}, \sigma_2^{ini}\right)$$
$$s \sim \text{Normal}^+\left(\sigma^{ini}\right)$$

$$\theta_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i}$$

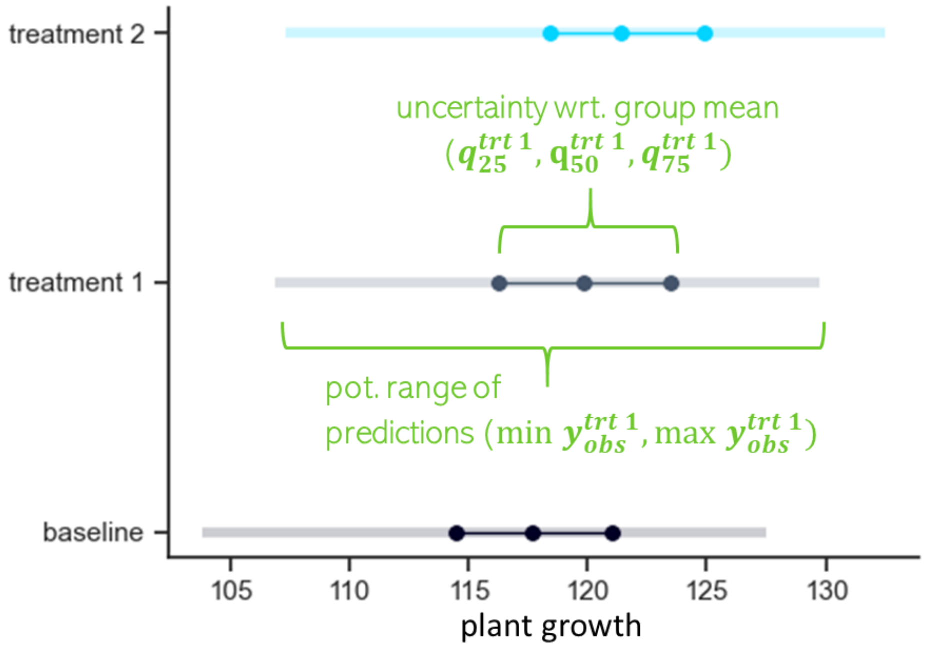
$$y_i \sim \text{Normal}(\theta_i, s)$$

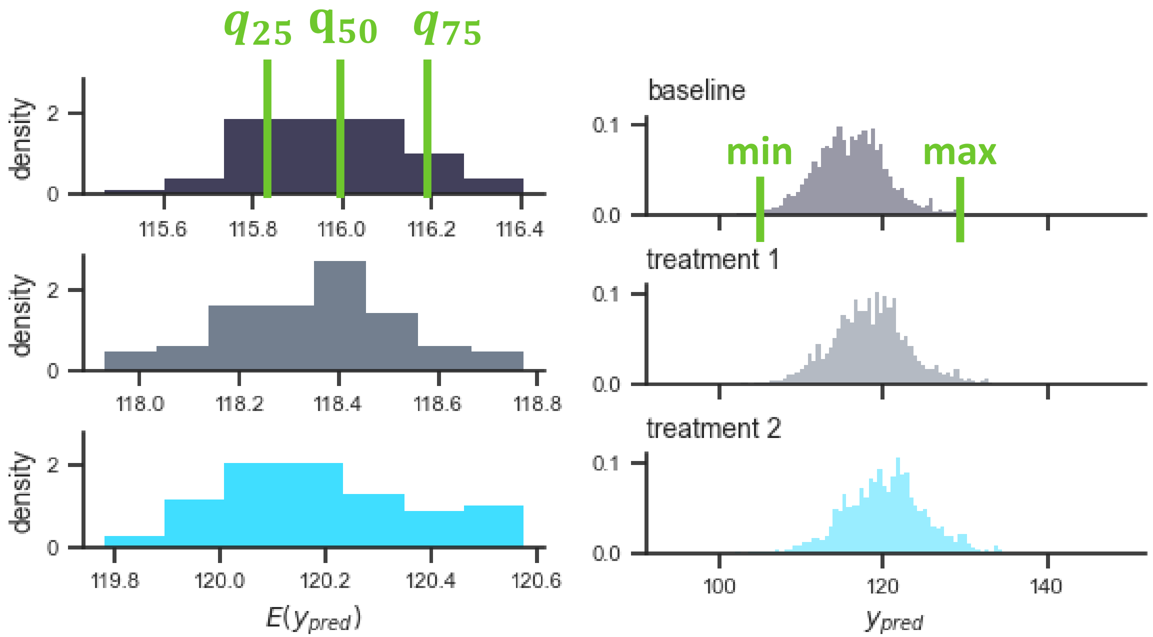




A closer look into our method

… and compute the elicited statistics





florence.bockting@tu-dortmund.de

A closer look into our method

Learn hyperparameter vector via batch SGD

- Compute loss based on simulated data and expert expectations

$$L(\lambda) = \alpha_1 L_1\left(\min y_{pred}^{trt\,1}, \min \hat{y}_{pred}^{trt\,1}\right) + \alpha_2 L_2\left(q_p^{trt\,1}, \hat{q}_p^{trt\,1}\right) + \cdots + \alpha_6 L_6\left(q_p^{crt}, \hat{q}_p^{crt}\right)$$

florence.bockting@tu-dortmund.de

A closer look into our method

Learn hyperparameter values via batch SGD

- Compute loss based on simulated data and expert expectations

$$L(\lambda) = \alpha_1 L_1\left(\min y_{pred}^{trt\,1}, \min \hat{y}_{pred}^{trt\,1}\right) + \alpha_2 L_2\left(q_p^{trt\,1}, \hat{q}_p^{trt\,1}\right) + \cdots + \alpha_6 L_6\left(q_p^{crt}, \hat{q}_p^{crt}\right)$$

- Compute gradient of loss w.r.t. λ and adjust λ in the opposite direction of the gradient

$$\lambda \leftarrow \lambda - \delta \sum_{m=1}^{M} \alpha_m \frac{\partial L_m}{\partial \lambda}$$

florence.bockting@tu-dortmund.de

A closer look into our method

Learn hyperparameter values via batch SGD

- Compute loss based on simulated data and expert expectations

$$L(\lambda) = \alpha_1 L_1\left(\min y_{pred}^{trt\,1}, \min \hat{y}_{pred}^{trt\,1}\right) + \alpha_2 L_2\left(q_p^{trt\,1}, \hat{q}_p^{trt\,1}\right) + \cdots + \alpha_6 L_6\left(q_p^{crt}, \hat{q}_p^{crt}\right)$$

- Compute gradient of loss w.r.t. λ and adjust λ in the opposite direction of the gradient

$$\lambda \leftarrow \lambda - \delta \sum_{m=1}^{M} \alpha_m \frac{\partial L_m}{\partial \lambda}$$

- Repeat until max. number of epochs

$$\text{update}\left(\lambda^{init}\right) \mapsto \lambda^{t_1}$$

...

$$\text{update}\left(\lambda^{t_{\max-1}}\right) \mapsto \lambda^{t_{\max}}$$

florence.bockting@tu-dortmund.de

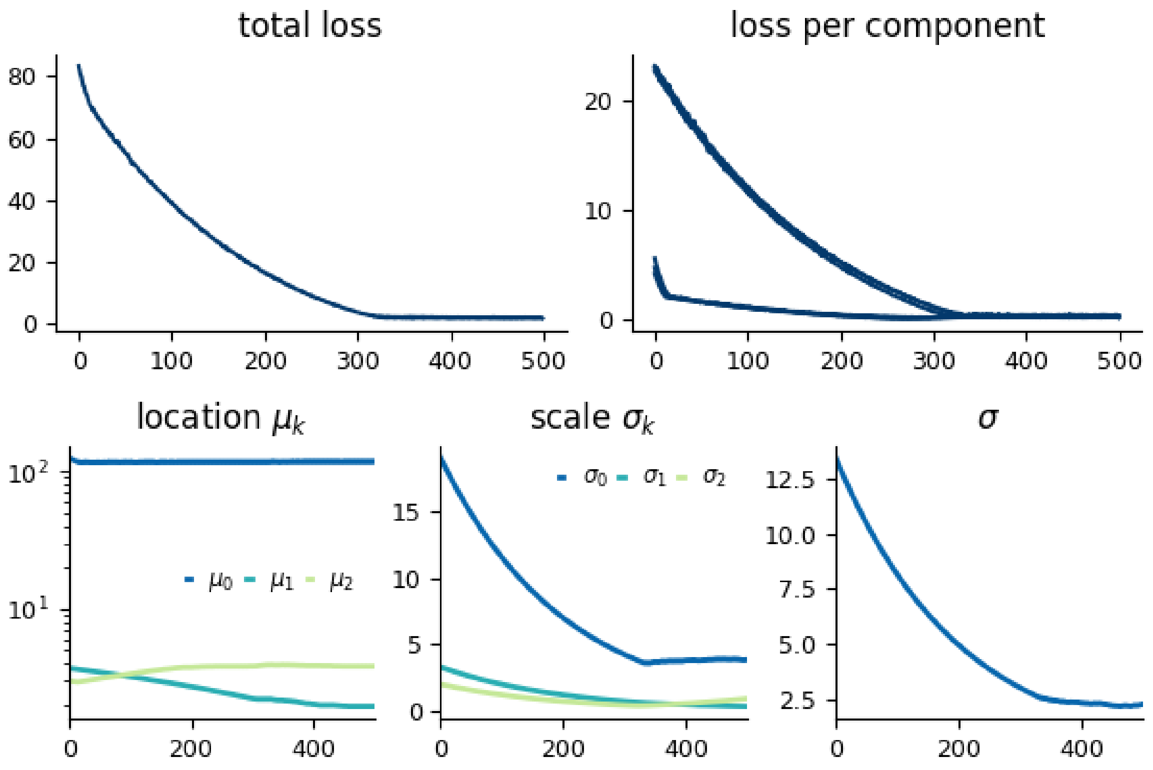
A closer look into our method

Convergence diagnostics

$$\beta_0 \sim \text{Normal}(\mu_0, \sigma_0)$$
$$\beta_1 \sim \text{Normal}(\mu_1, \sigma_1)$$
$$\beta_2 \sim \text{Normal}(\mu_2, \sigma_2)$$
$$s \sim \text{Normal}^+(\sigma)$$

$$\theta_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i}$$

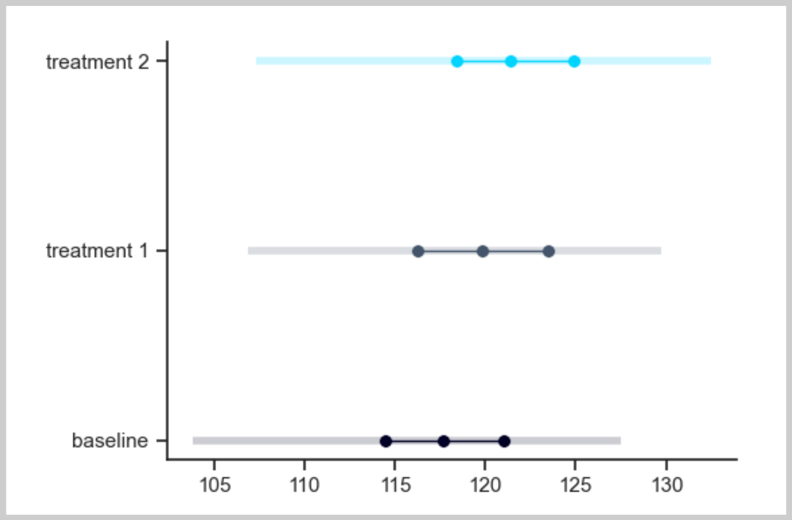
$$y_i \sim \text{Normal}(\theta_i, s)$$



florence.bockting@tu-dortmund.de

Reminder: The problem

Translate expert beliefs into corresponding priors



$$\beta_0 \sim \text{Normal}(\mu_0, \sigma_0)$$
$$\beta_1 \sim \text{Normal}(\mu_1, \sigma_1)$$
$$\beta_2 \sim \text{Normal}(\mu_2, \sigma_2)$$
$$s \sim \text{Normal}^+(\sigma)$$

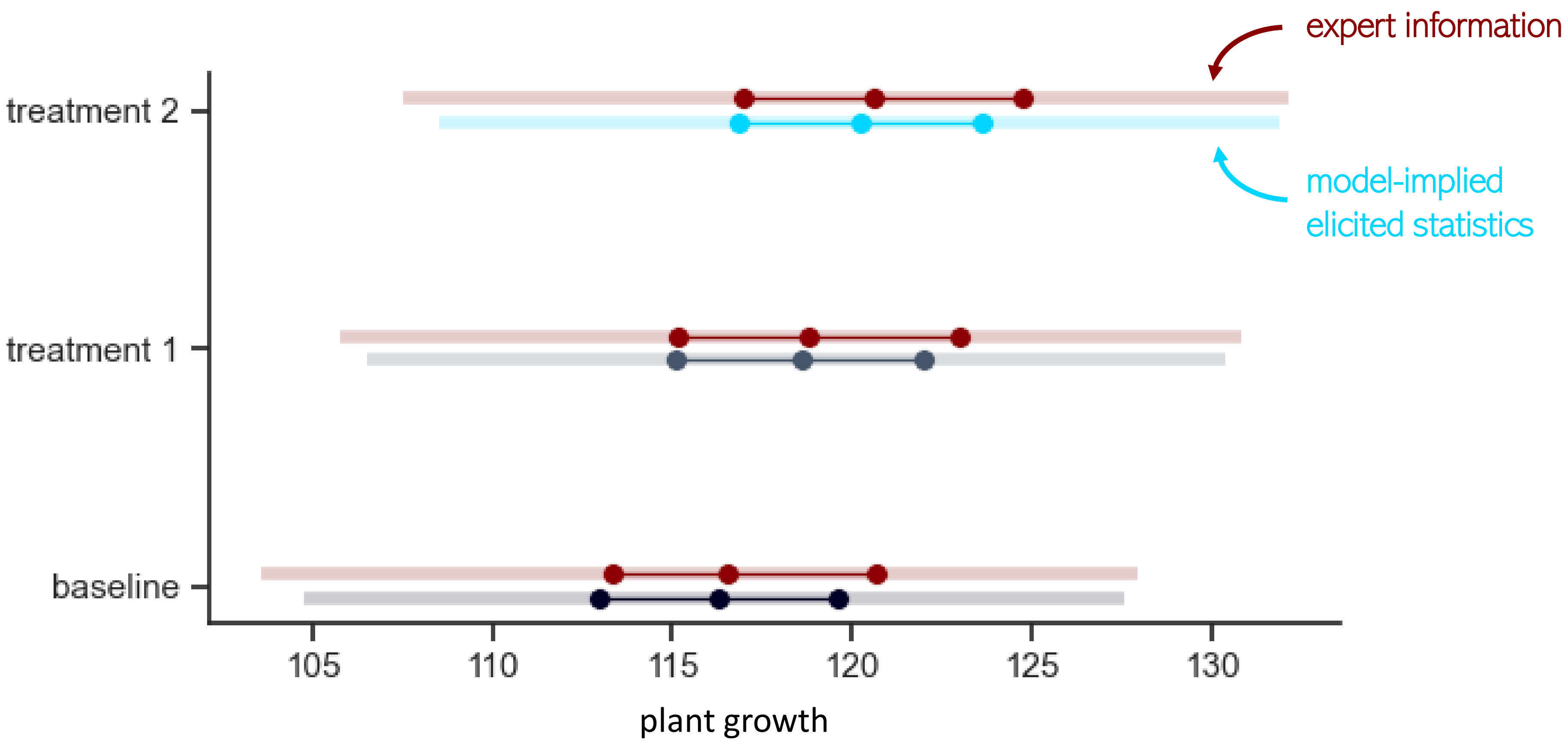
$$\theta_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i}$$

$$y_i \sim \text{Normal}(\theta_i, s)$$

?

A closer look into our method

Results: learned vs. expert-elicited statistics



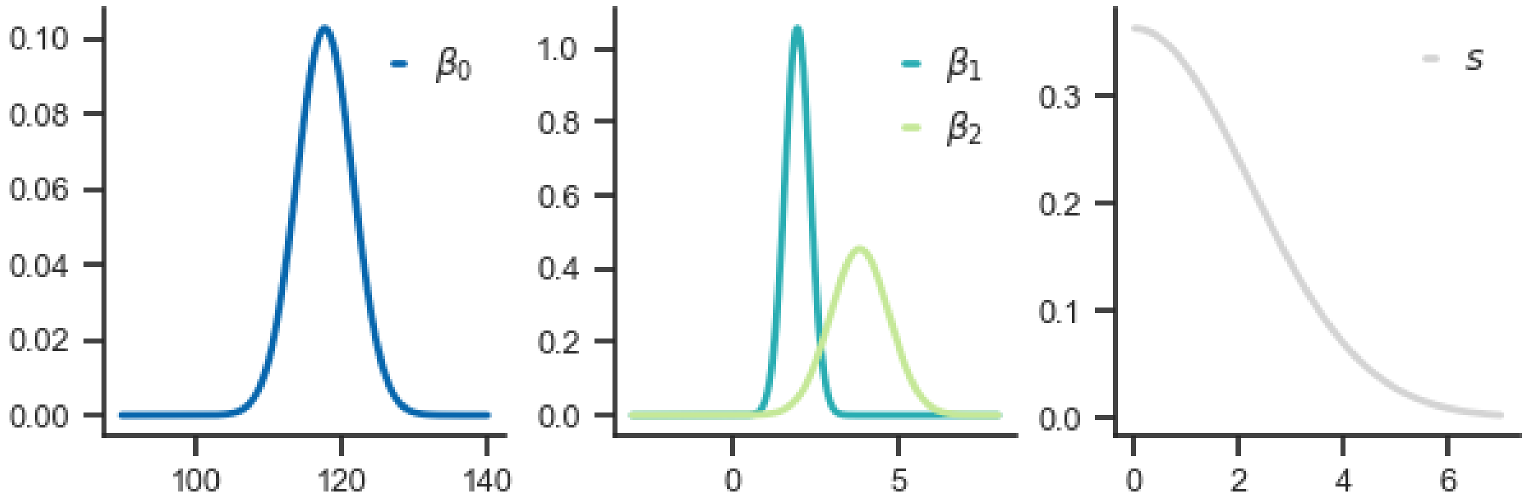
A closer look into our method

Results: Learned prior distributions

$$\beta_0 \sim \text{Normal}(\mu_0, \sigma_0)$$
$$\beta_1 \sim \text{Normal}(\mu_1, \sigma_1)$$
$$\beta_2 \sim \text{Normal}(\mu_2, \sigma_2)$$
$$s \sim \text{Normal}^+(\sigma)$$

$$\theta_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i}$$

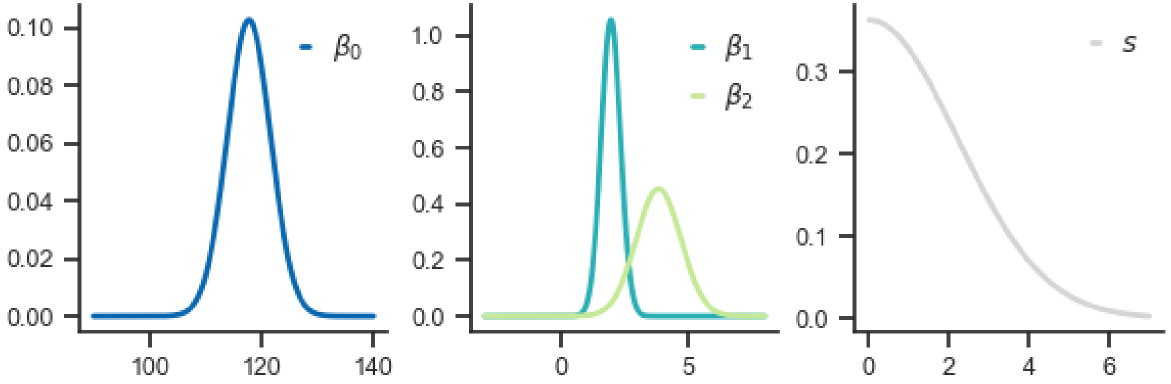
$$y_i \sim \text{Normal}(\theta_i, s)$$



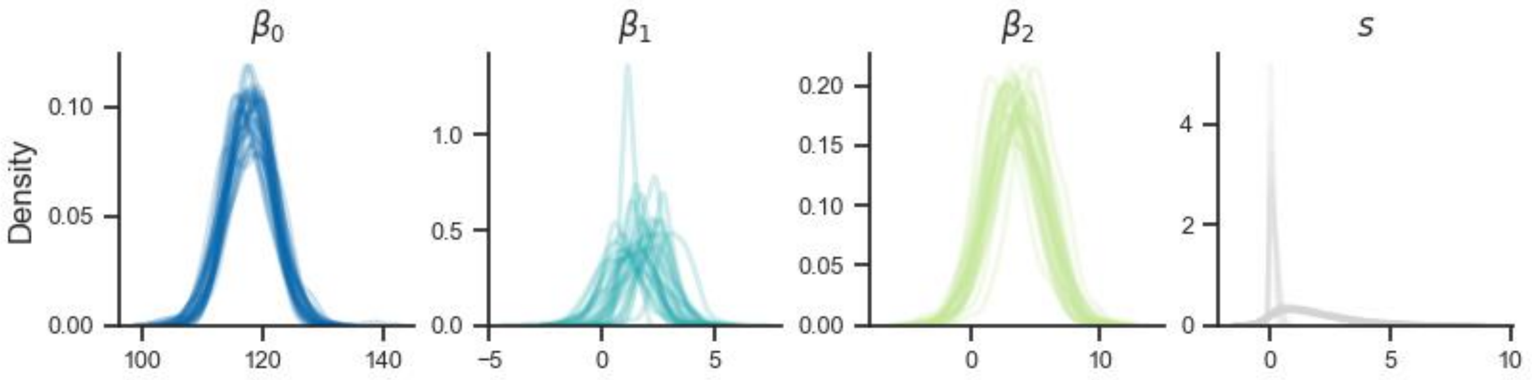
florence.bockting@tu-dortmund.de

Sensitivity analysis

Sensitivity of learned prior distributions



Sensitivity analysis (30 replications with varying seed)



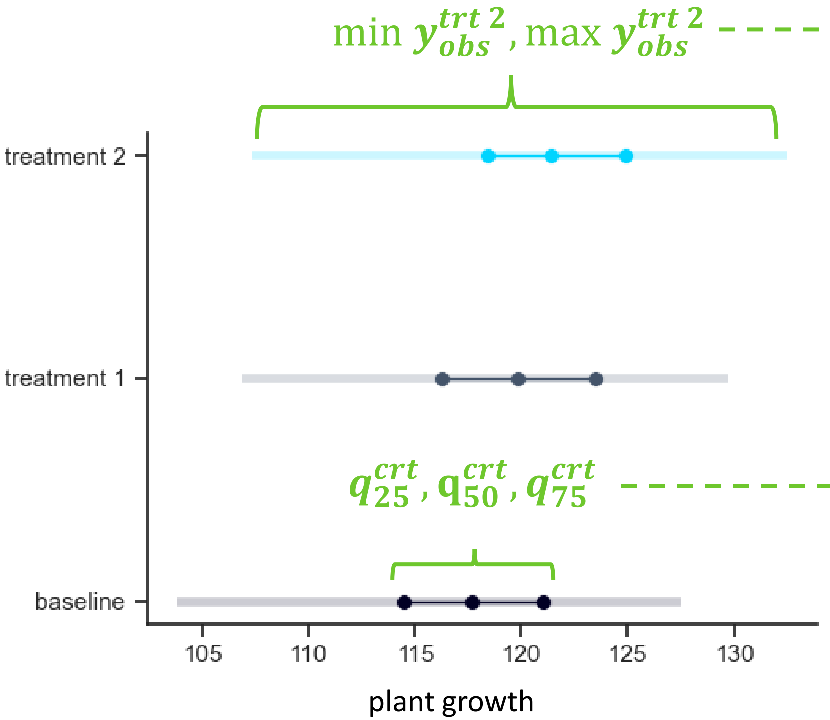
florence.bockting@tu-dortmund.de

Sensitivity analysis

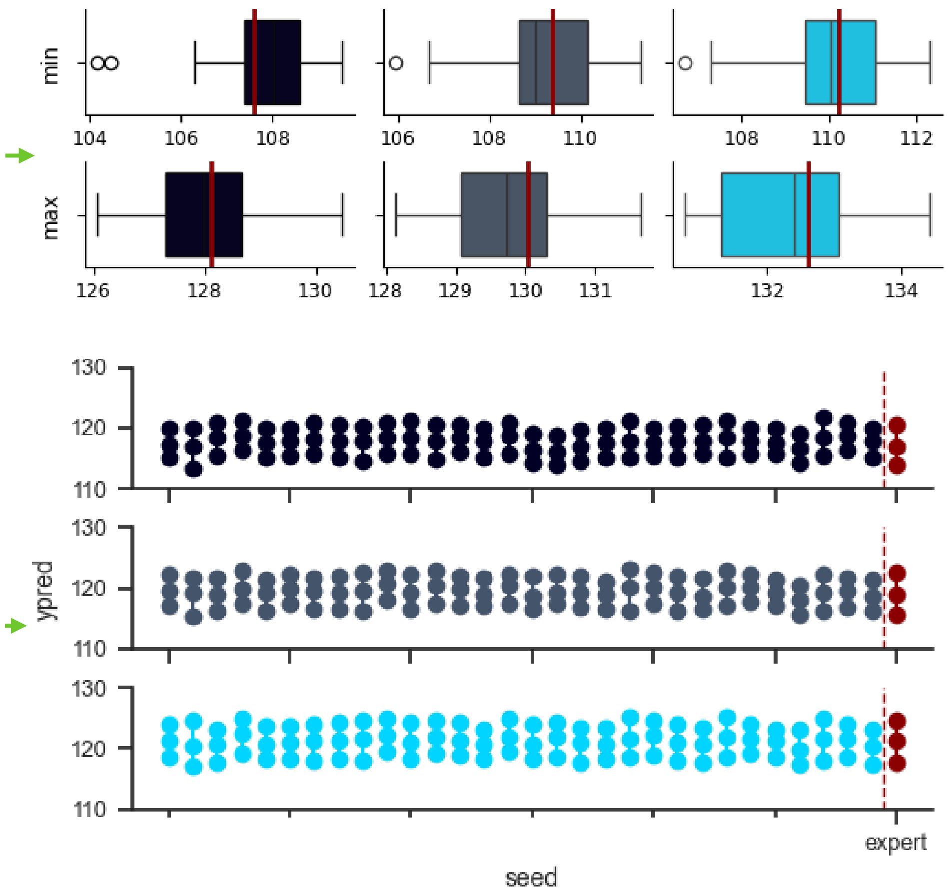
Sensitivity of learned elicited statistics

tu dortmund university
computational statistics

Expert information (reminder)



Sensitivity analysis (30 replications with varying seed)

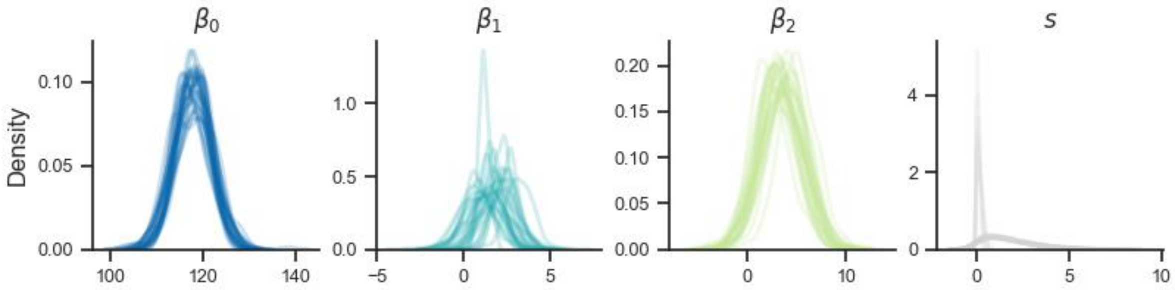


florence.bockting@tu-dortmund.de

Sensitivity analysis

Approaches for dealing with sensitive results

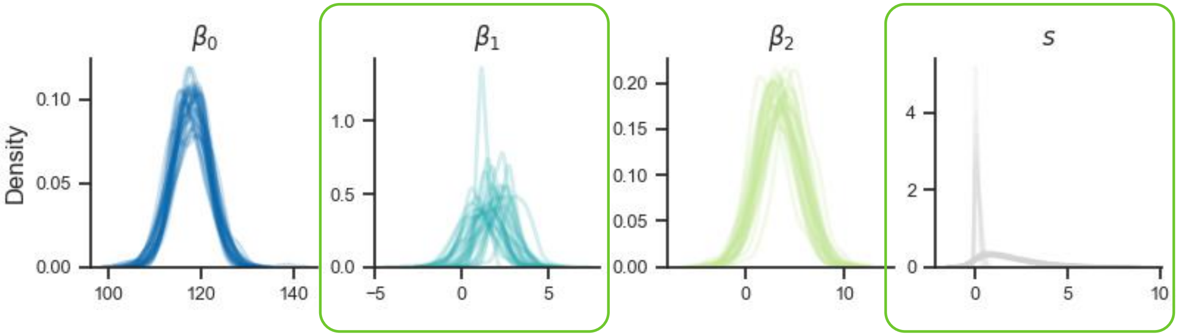
- **Elicit additional expert information and incorporate it in the learning algorithm**

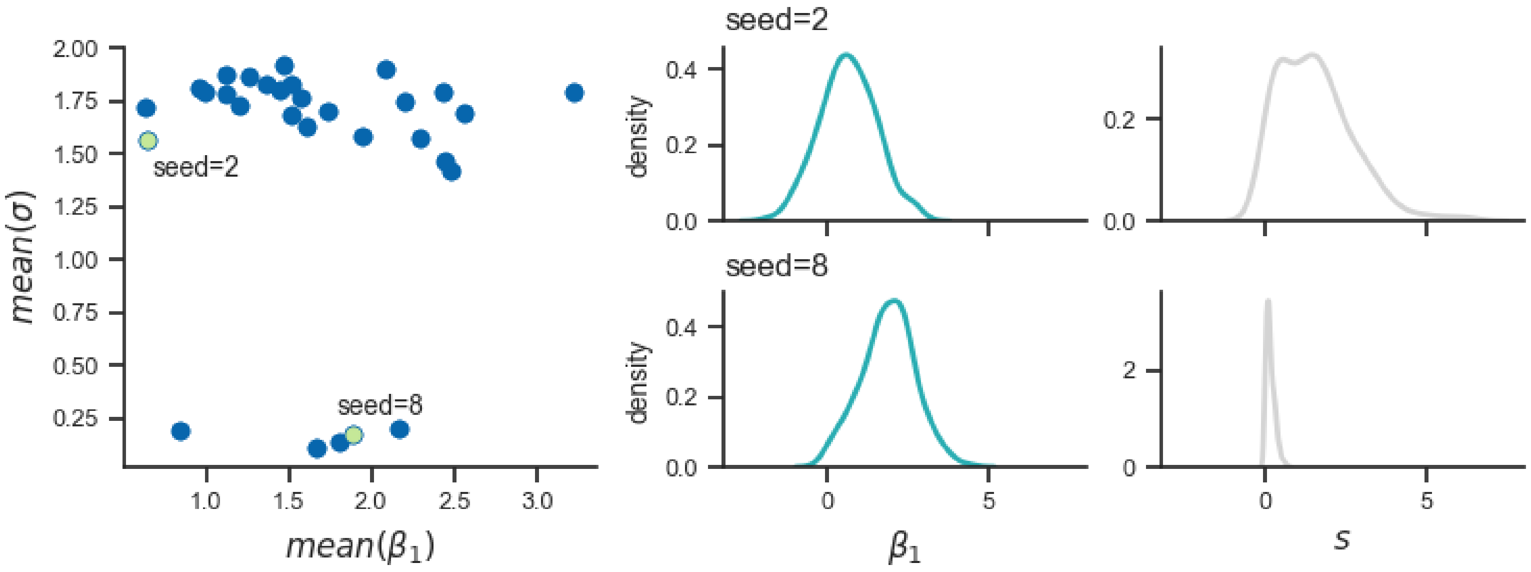


Sensitivity analysis

Approaches for dealing with sensitive results

- **Elicit additional expert information and incorporate it in the learning algorithm**
- **Select plausible prior distributions among learned hyperparameter values**



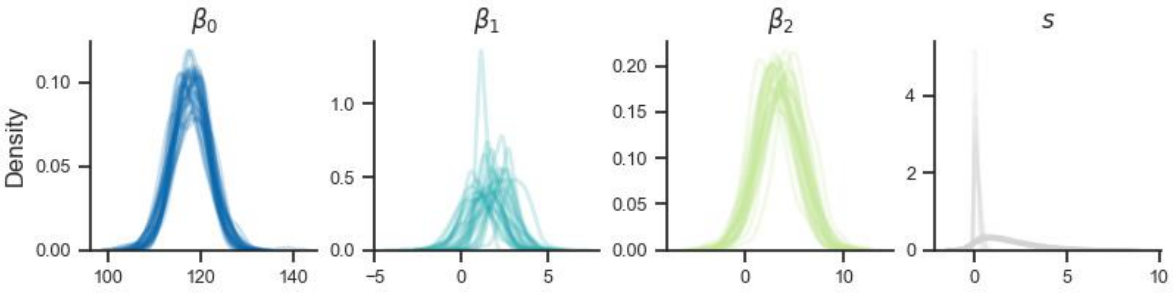


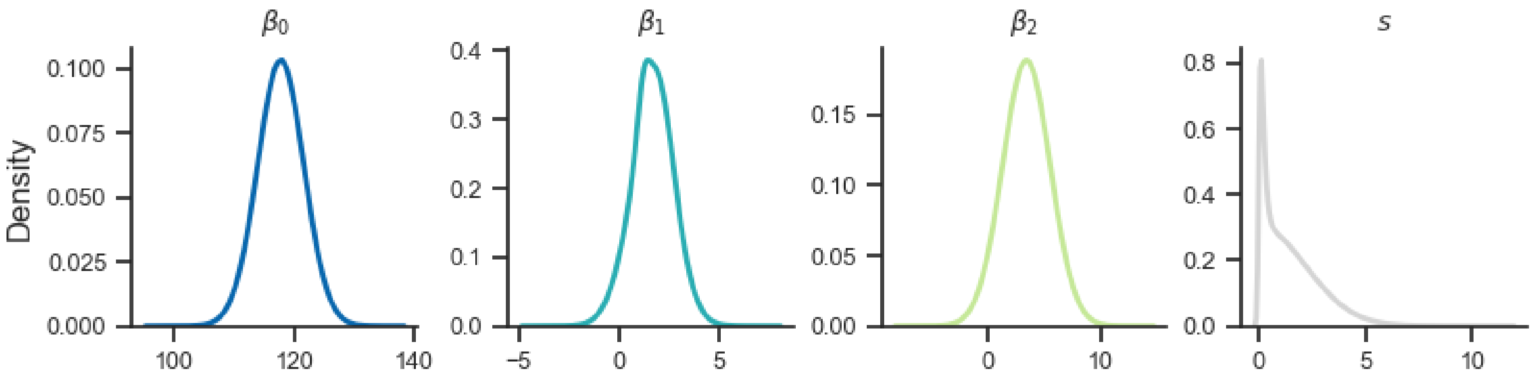
florence.bockting@tu-dortmund.de

Sensitivity analysis

Approaches for dealing with sensitive results

- **Elicit additional expert information and incorporate it in the learning algorithm**
- Select plausible prior distributions among learned hyperparameter values
- Model averaging





florence.bockting@tu-dortmund.de

Outlook

Where to go from here?

- Make the method actually Bayesian …
  - Explicitly represent uncertainty about the elicitation process and learn a posterior distribution of the hyperparameter values
- Instead of learning the hyperparameters of a prespecified family – learn the whole joint distribution on the model parameters
  - Work in progress – preprint is coming soon
- Approaches that deal with multiple expert beliefs
- Work out helpful diagnostics

florence.bockting@tu-dortmund.de

Outlook

Where to go from here?

- Interface to R/Stan (current implementation is in Python TensorFlow)
- Tutorial paper for practitioners
- Applications
  - I am looking for collaborators who have an application (+ an expert) and are willing to try out the method.

florence.bockting@tu-dortmund.de

dortmund university
computational statistics

Thank you for your attention.

Contact:



Florence Bockting
TU Dortmund University, GER

florence.bockting@tu-dortmund.de

Stefan T. Radev
Rensselaer Polytechnic Institute, NY, USA

radevs@rpi.edu

Paul-Christian Bürkner
TU Dortmund University, GER

https://paul-buerkner.github.io/

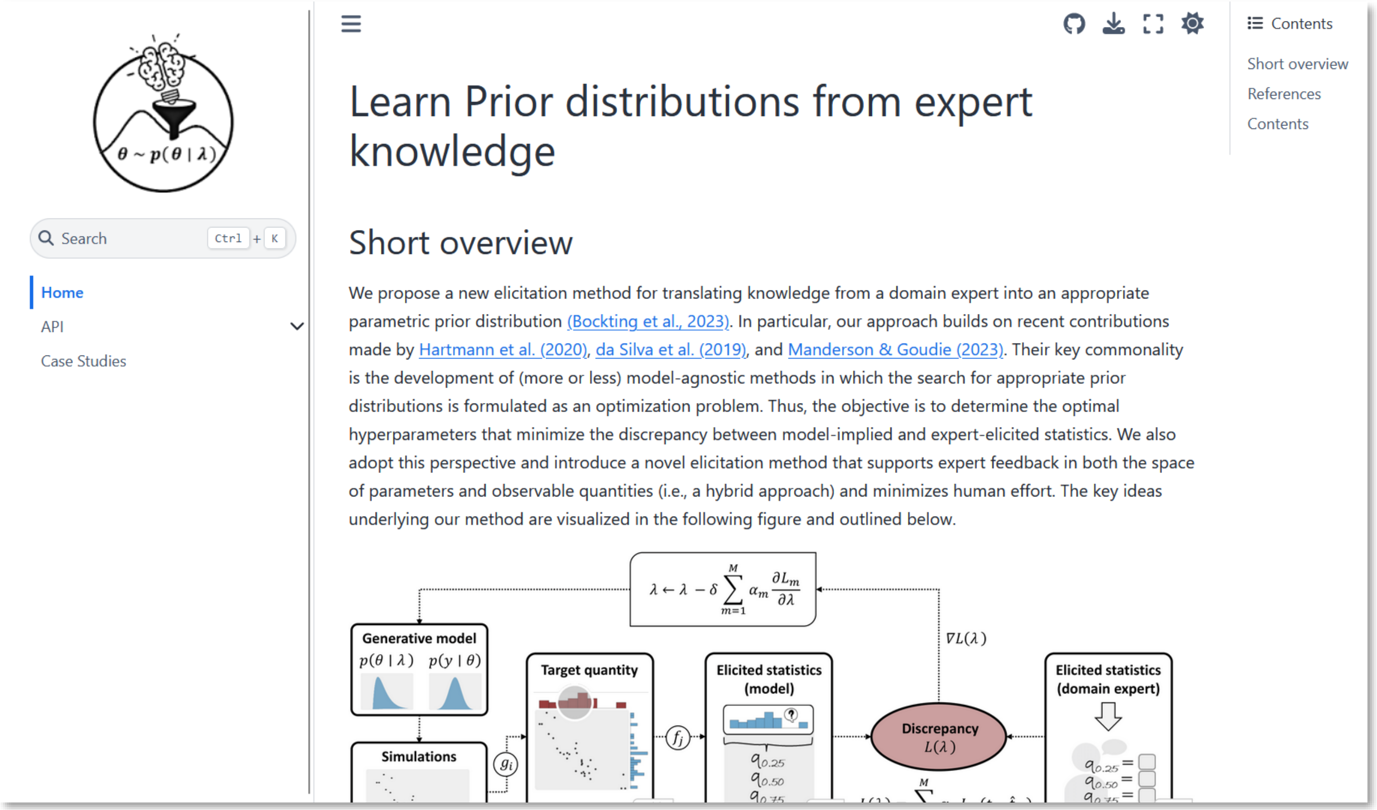
dortmund university

computational statistics

Thank you for your attention.

Project website: (under construction)

https://florence-bockting.github.io/PriorLearning/index.html



Search Ctrl + K

Home

API

Case Studies

Contents

Short overview

References

Contents

Learn Prior distributions from expert knowledge

Short overview

We propose a new elicitation method for translating knowledge from a domain expert into an appropriate parametric prior distribution (Bockting et al., 2023). In particular, our approach builds on recent contributions made by Hartmann et al. (2020), da Silva et al. (2019), and Manderson & Goudie (2023). Their key commonality is the development of (more or less) model-agnostic methods in which the search for appropriate prior distributions is formulated as an optimization problem. Thus, the objective is to determine the optimal hyperparameters that minimize the discrepancy between model-implied and expert-elicited statistics. We also adopt this perspective and introduce a novel elicitation method that supports expert feedback in both the space of parameters and observable quantities (i.e., a hybrid approach) and minimizes human effort. The key ideas underlying our method are visualized in the following figure and outlined below.

ISBA 2024

References

- Albert, I., Donnet, S., Guihenneuc-Jouyaux, C., Low-Choy, S., Mengersen, K., & Rousseau, J. (2012). Combining Expert Opinions in Prior Elicitation. *Bayesian Analysis*, 7(3), 503-532.
- da Silva, E. D. S., Kuśmierczyk, T., Hartmann, M., & Klami, A. (2023). Prior Specification for Bayesian Matrix Factorization via Prior Predictive Matching. *Journal of Machine Learning Research*, 24(67), 1-51.
- Hartmann, M., Agiashvili, G., Bürkner, P., & Klami, A. (2020). Flexible prior elicitation via the prior predictive distribution. In Conference on Uncertainty in Artificial Intelligence (pp. 1129-1138). PMLR.
- Manderson, A. A., & Goudie, R. J. (2023). Translating predictive distributions into informative priors. ArXiv preprint.
- Mikkola, P., Martin, O. A., Chandramouli, S., Hartmann, M., Pla, O. A., Thomas, O., … & Klami, A. (2021). Prior knowledge elicitation: The past, present, and future. ArXiv preprint.

florence.bockting@tu-dortmund.de