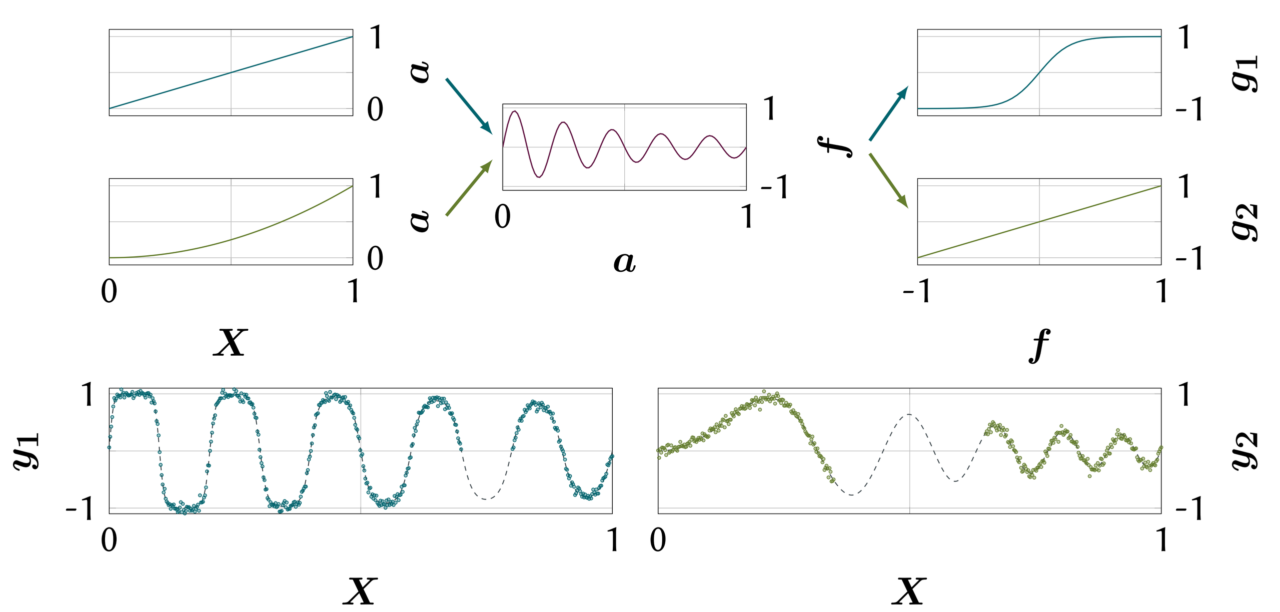
Bayesian alignments of warped multi-output Gaussian processes

Markus Kaiser [1,2] Clemens Otte [1] Thomas Runkler [1,2] Carl Henrik Ek [3]

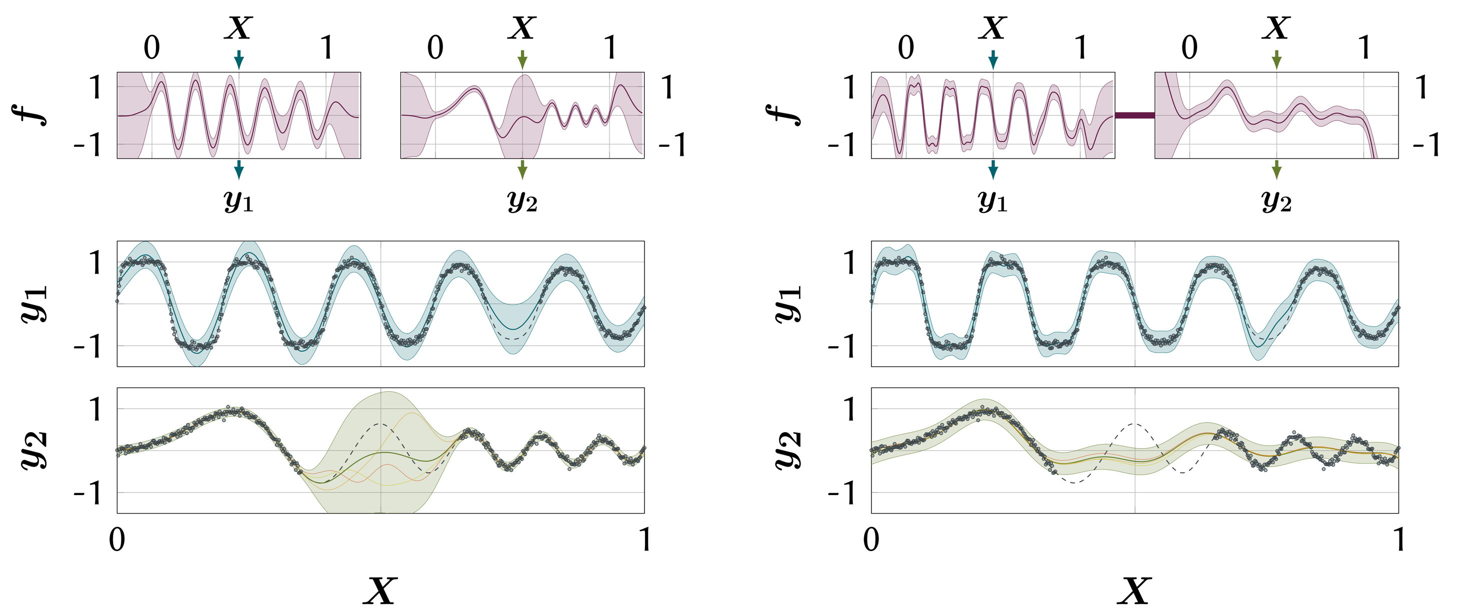
[1]Siemens AG [2]Technical University of Munich [3]University of Bristol

Artificial data set



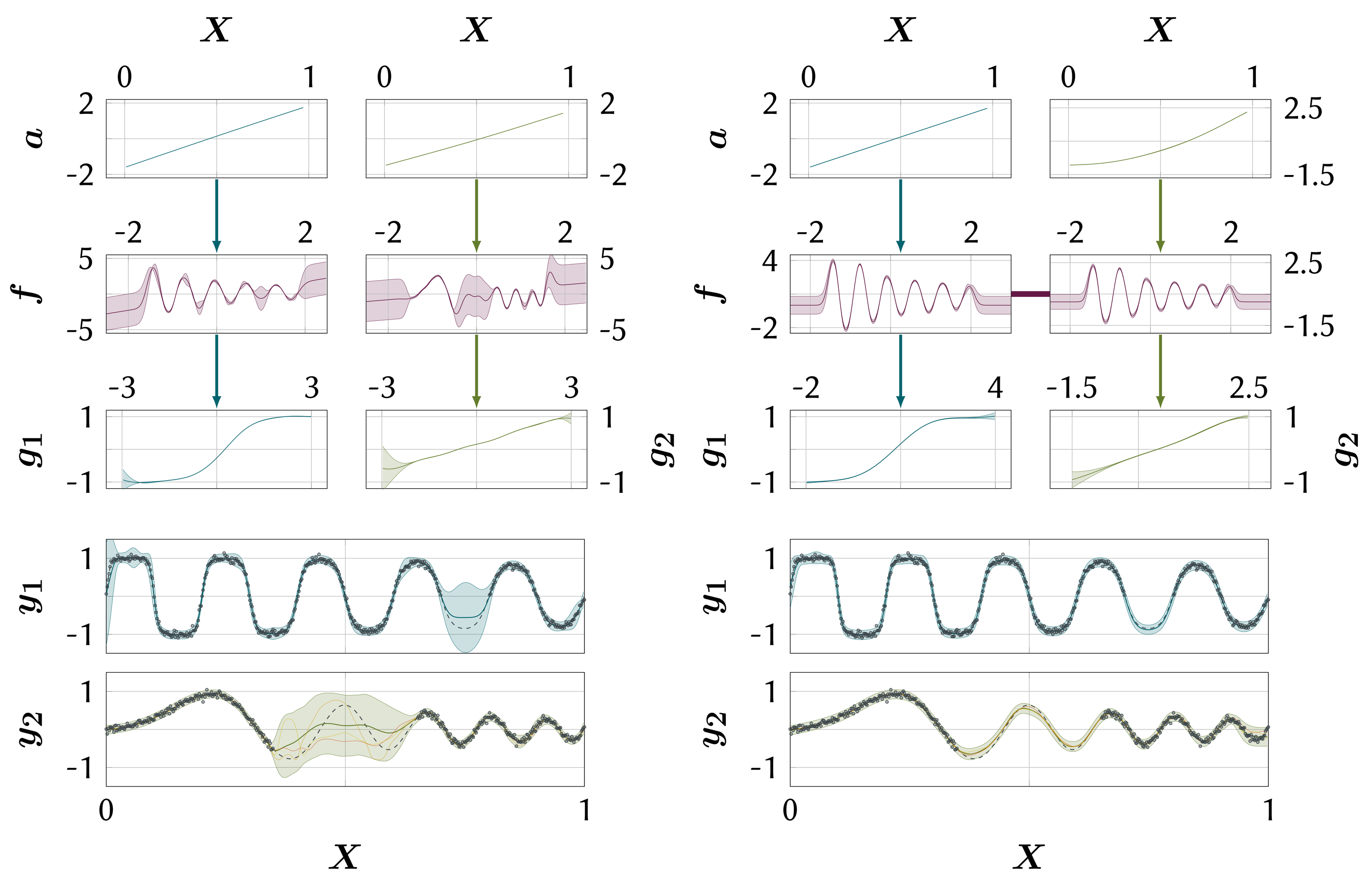
- Two time series are generated via a shared dampened sine function
- The shared function is never observed directly
- A non-linear warping and a non-linear alignment are applied to the time series
- The task is to recover this hierarchy and predict missing intervals

Model comparison



(a) Shallow GP with RBF kernel

(b) Multi-Output GP with dependent RBF kernel

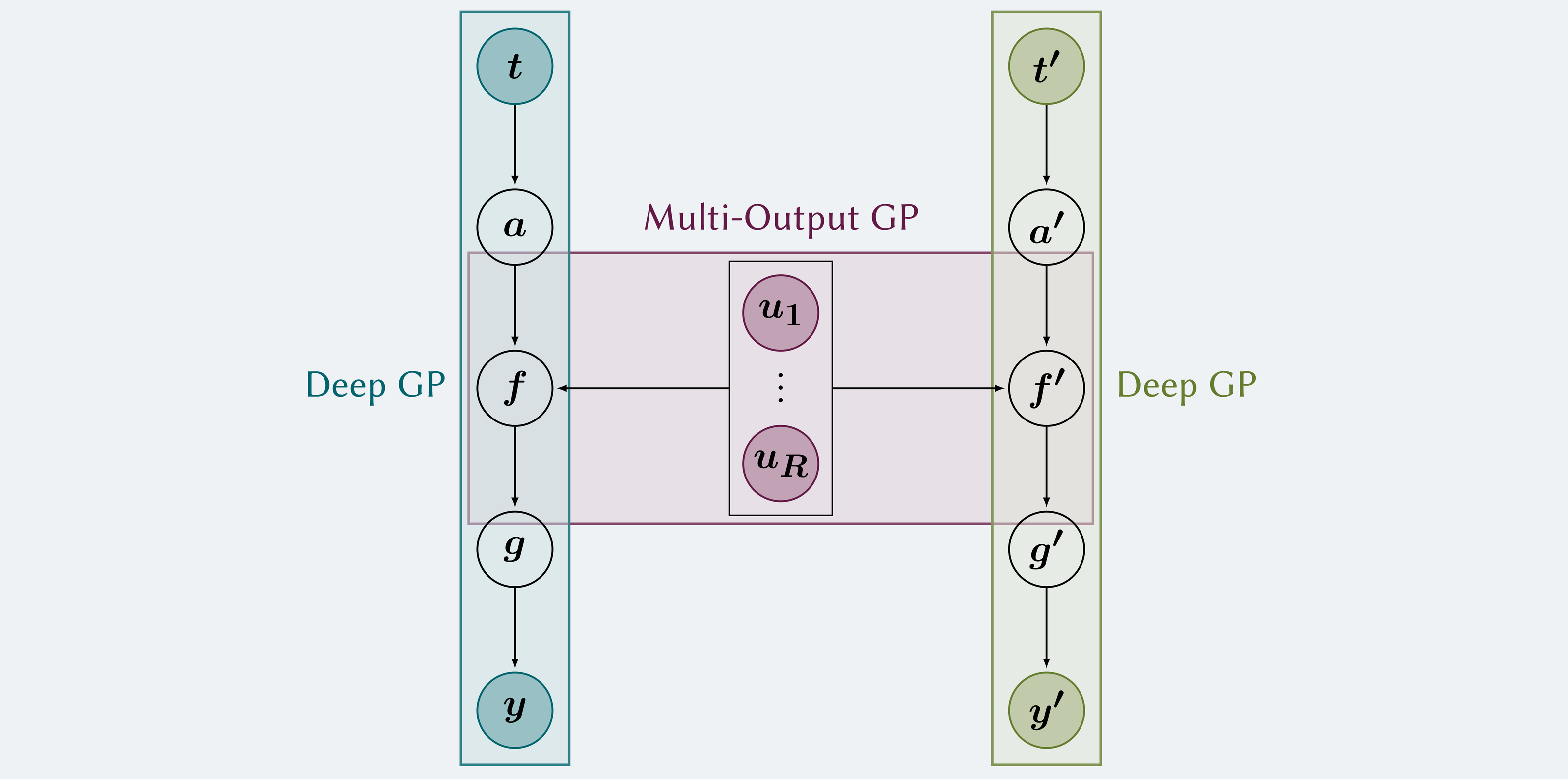


(c) Deep GP with RBF kernels

(d) **AMO-GP with (dependent) RBF kernels**

- AMO-GP correctly recovers the latent shared function, warping and alignment

Graphical model of AMO-GP



Contributions

- AMO-GP connects multiple deep GPs via a shared layer which is a multi-output GP
- Application to non-linear time series alignment with very noisy observations
- Learning scheme based on nested variational compression by Hensman and Lawrence (2014)
  - Add variational inducing variables $\mathrm{q}(\boldsymbol{u} \mid \boldsymbol{Z}) \sim \mathcal{N}(\boldsymbol{u} \mid \boldsymbol{m}, \boldsymbol{S})$
  - Compress uncertainties after propagating through a GP and apply recursively
  - Joint variational bound given by

$$
\begin{aligned}
\log \mathrm{p}(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{u}) \geq & \sum_{d=1}^{D} \log \mathcal{N}\left(\boldsymbol{y_d} \,\middle|\, \boldsymbol{\Psi_{g,d}} \boldsymbol{K}^{-1}_{\boldsymbol{u_{g,d}u_{g,d}}} \boldsymbol{m_{g,d}}, \sigma^2_{y,d}\mathbf{I}\right) - \sum_{d=1}^{D} \frac{1}{2\sigma^2_{a,d}} \operatorname{tr}(\boldsymbol{\Sigma_{a,d}}) \\
& - \frac{1}{2\sigma_f^2}\left(\psi_f - \operatorname{tr}\left(\boldsymbol{\Phi_f} \boldsymbol{K}^{-1}_{\boldsymbol{u_f u_f}}\right)\right) - \sum_{d=1}^{D} \frac{1}{2\sigma^2_{y,d}} \left(\psi_{g,d} - \operatorname{tr}\left(\boldsymbol{\Phi_{g,d}} \boldsymbol{K}^{-1}_{\boldsymbol{u_{g,d}u_{g,d}}}\right)\right) \\
& - \sum_{d=1}^{D} \mathrm{KL}(\mathrm{q}(\boldsymbol{u_{a,d}}) \parallel \mathrm{p}(\boldsymbol{u_{a,d}})) - \mathrm{KL}(\mathrm{q}(\boldsymbol{u_f}) \parallel \mathrm{p}(\boldsymbol{u_f})) - \sum_{d=1}^{D} \mathrm{KL}(\mathrm{q}(\boldsymbol{u_{y,d}}) \parallel \mathrm{p}(\boldsymbol{u_{y,d}})) \\
& - \frac{1}{2\sigma_f^2} \operatorname{tr}\left(\left(\boldsymbol{\Phi_f} - \boldsymbol{\Psi}^{\mathsf{T}}_{\boldsymbol{f}} \boldsymbol{\Psi_f}\right) \boldsymbol{K}^{-1}_{\boldsymbol{u_f u_f}} \left(\boldsymbol{m_f} \boldsymbol{m}^{\mathsf{T}}_{\boldsymbol{f}} + \boldsymbol{S_f}\right) \boldsymbol{K}^{-1}_{\boldsymbol{u_f u_f}}\right) \\
& - \sum_{d=1}^{D} \frac{1}{2\sigma^2_{y,d}} \operatorname{tr}\left(\left(\boldsymbol{\Phi_{g,d}} - \boldsymbol{\Psi}^{\mathsf{T}}_{\boldsymbol{g,d}} \boldsymbol{\Psi_{g,d}}\right) \boldsymbol{K}^{-1}_{\boldsymbol{u_{g,d}u_{g,d}}} \left(\boldsymbol{m_{g,d}} \boldsymbol{m}^{\mathsf{T}}_{\boldsymbol{g,d}} + \boldsymbol{S_{g,d}}\right) \boldsymbol{K}^{-1}_{\boldsymbol{u_{g,d}u_{g,d}}}\right)
\end{aligned}
$$

- Ψ-statistics for MO-GP based on dependent GPs by Boyle and Frean (2004)

$$
\psi_f = \mathbb{E}_{\mathrm{q}(\boldsymbol{a})}\left[\operatorname{tr}\left(\boldsymbol{K_{ff}}\right)\right] = \sum_{n=1}^{N} \hat{\sigma}^2_{nn}
$$

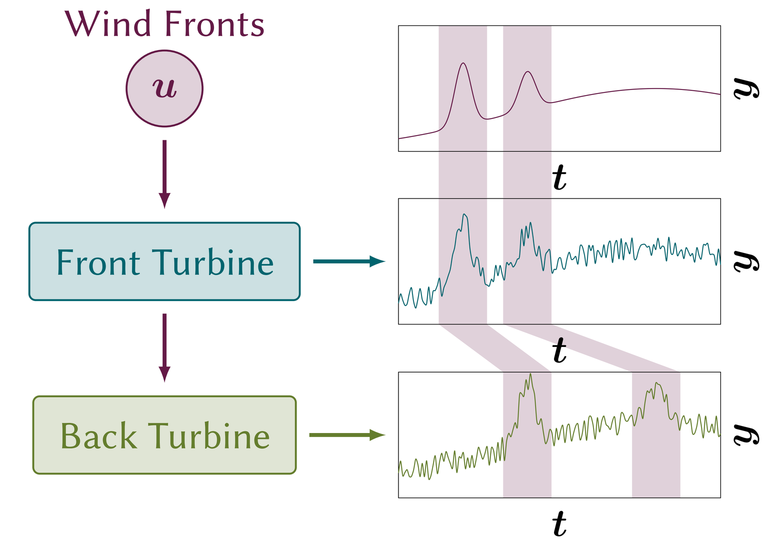
$$
\boldsymbol{\Psi_f} = \mathbb{E}_{\mathrm{q}(\boldsymbol{a})}\left[\boldsymbol{K_{fu}}\right], \text{ with}
$$

$$
\left(\boldsymbol{\Psi_f}\right)_{ni} = \hat{\sigma}^2_{ni} \sqrt{\frac{(\boldsymbol{\Sigma_a})^{-1}_{nn}}{\hat{\ell}_{ni} + (\boldsymbol{\Sigma_a})^{-1}_{nn}}} \exp\left(-\frac{1}{2} \frac{(\boldsymbol{\Sigma_a})^{-1}_{nn} \hat{\ell}_{ni}}{(\boldsymbol{\Sigma_a})^{-1}_{nn} + \hat{\ell}_{ni}} ((\boldsymbol{\mu_a})_n - \boldsymbol{Z_i})^2\right)
$$

$$
\boldsymbol{\Phi_f} = \mathbb{E}_{\mathrm{q}(\boldsymbol{a})}\left[\boldsymbol{K_{uf}} \boldsymbol{K_{fu}}\right], \text{ with}
$$

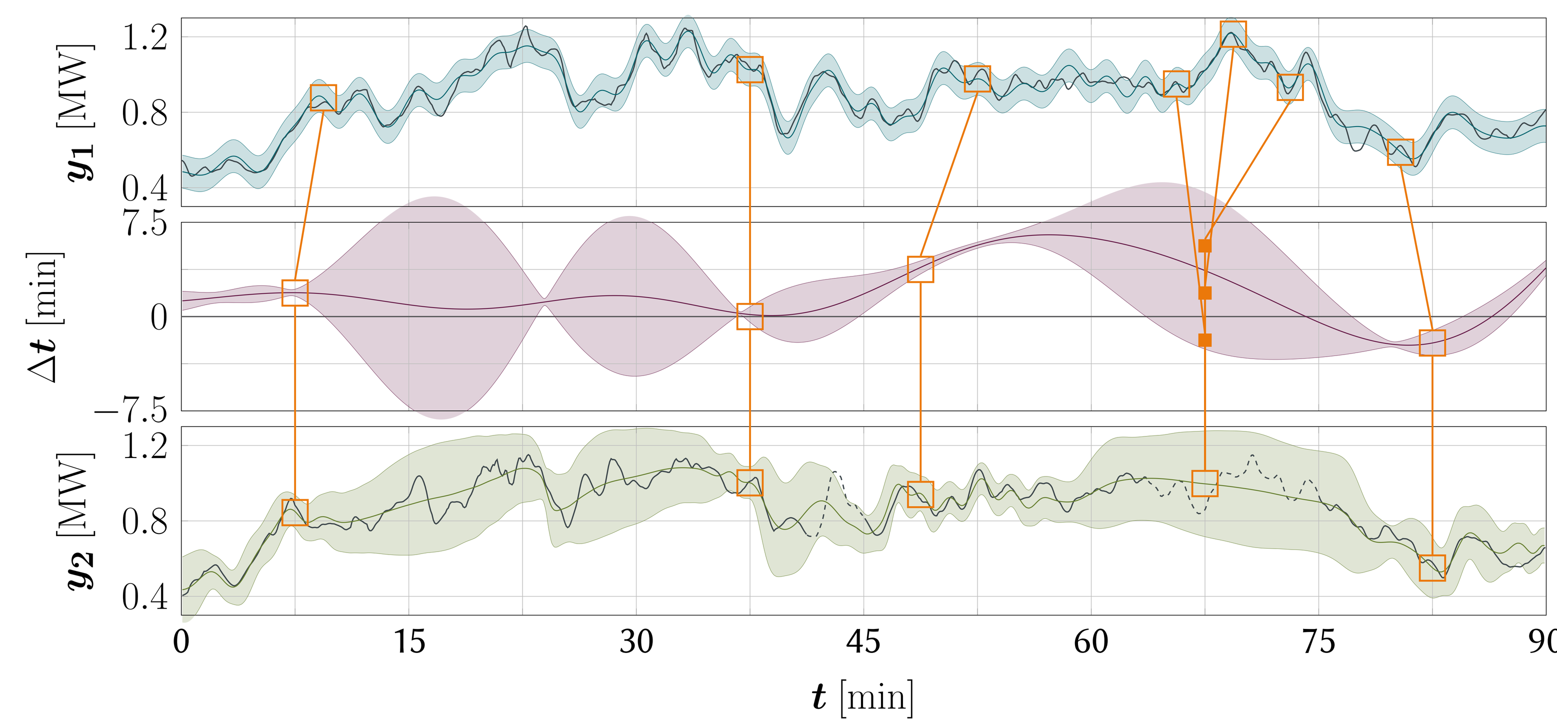
$$
\begin{aligned}
\left(\boldsymbol{\Phi_f}\right)_{ij} = \sum_{n=1}^{N} \hat{\sigma}^2_{ni} \hat{\sigma}^2_{nj} \sqrt{\frac{(\boldsymbol{\Sigma_a})^{-1}_{nn}}{\hat{\ell}_{ni} + \hat{\ell}_{nj} + (\boldsymbol{\Sigma_a})^{-1}_{nn}}} \exp & \left(-\frac{1}{2} \frac{\hat{\ell}_{ni} \hat{\ell}_{nj}}{\hat{\ell}_{ni} + \hat{\ell}_{nj}} (\boldsymbol{Z_i} - \boldsymbol{Z_j})^2 \right. \\
& \left. - \frac{1}{2} \frac{(\boldsymbol{\Sigma_a})^{-1}_{nn} (\hat{\ell}_{ni} + \hat{\ell}_{nj})}{(\boldsymbol{\Sigma_a})^{-1}_{nn} + \hat{\ell}_{ni} + \hat{\ell}_{nj}} \left((\boldsymbol{\mu_a})_n - \frac{\hat{\ell}_{ni} \boldsymbol{Z_i} + \hat{\ell}_{nj} \boldsymbol{Z_j}}{\hat{\ell}_{ni} + \hat{\ell}_{nj}}\right)^2\right)
\end{aligned}
$$

Real-world application



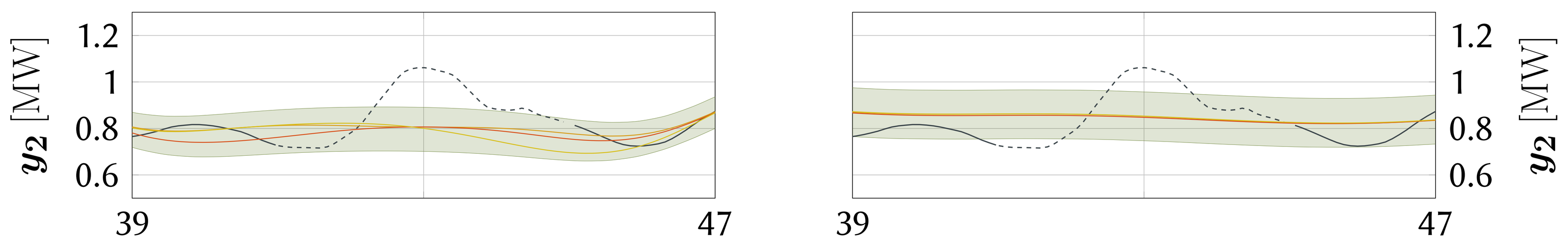
- The goal is to find shared latent wind conditions given the power production of wind turbines
- Alignment of time series unknown and dependent on changing propagation times
- Observations are very noisy due to local turbulence
- Principled uncertainties are important as no ground-truth is available

Wind propagation through a wind farm



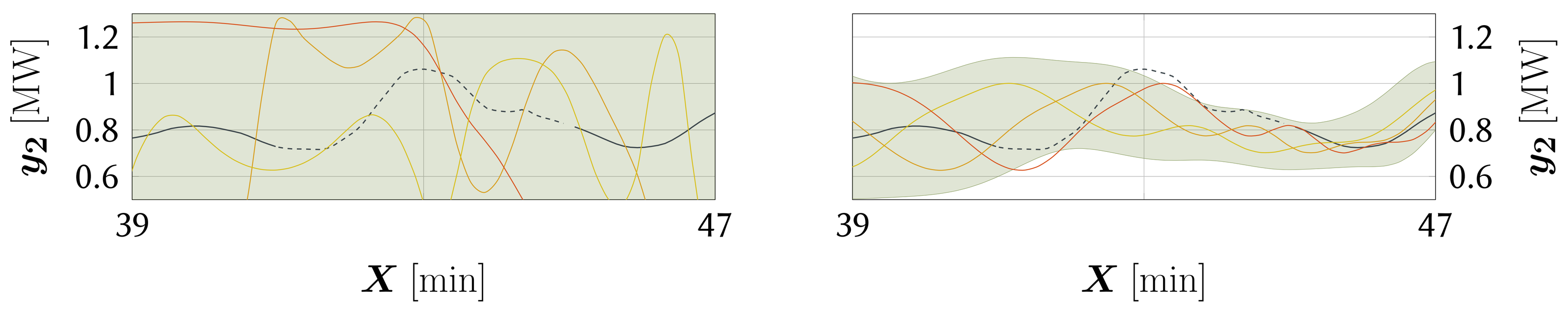
- AMO-GP recovers unique alignments with high certainty …
- …and places high uncertainty where noise dominates

Composite uncertainties



(a) Samples from a GP

(b) Samples from a MO-GP



(c) Samples from a DGP

(d) **Samples from AMO-GP**

- Samples from AMO-GP show rich structure while the other models are uninformative