Gaussian process models for Combining GCM Output

Will Tebbutt

Machine Learning Group, University of Cambridge

06-03-2019

Overview

Section 1

A Motivating Problem

Motivating Problem

- Consider estimating some interesting function of climate e.g. heatwave frequency
- Try with single climate model.
- Try with multiple (average?) of climate models.
- What goes wrong?
- What if we try with corrected marginal statistics?
- What goes wrong?

Motivating Problem

Desiderata

A statistical spatio-temporal model which:

1. has the ability to utilise many sources of information
2. is grid agnostic / can do downscaling
3. is able to model multi-timestep events e.g. heatwaves
4. can handle uncertainty
5. is computationally tractable

Machine Learning

Assumptions + Data

Section 2

Types of Uncertainty

Types of Uncertainty
UQ

- Multiple sources of uncertainty
- Some more tractable than others

Types of Uncertainty
[Kennedy and O'Hagan, 2001]

- Parameter uncertainty (eg. unknown parametrisation settings)

Types of Uncertainty
[Kennedy and O'Hagan, 2001]

- Parameter uncertainty (eg. unknown parametrisation settings)
- Model inadequacy (eg. missing feedbacks)

Types of Uncertainty

[Kennedy and O'Hagan, 2001]

- Parameter uncertainty (eg. unknown parametrisation settings)
- Model inadequacy (eg. missing feedbacks)
- Residual variability (eg. all of weather, annual / decadal oscillations)

Types of Uncertainty

[Kennedy and O'Hagan, 2001]

- Parameter uncertainty (eg. unknown parametrisation settings)
- Model inadequacy (eg. missing feedbacks)
- Residual variability (eg. all of weather, annual / decadal oscillations)
- Parametric variability (eg. unknown forcing inputs)

Types of Uncertainty

[Kennedy and O'Hagan, 2001]

- Parameter uncertainty (eg. unknown parametrisation settings)
- Model inadequacy (eg. missing feedbacks)
- Residual variability (eg. all of weather, annual / decadal oscillations)
- Parametric variability (eg. unknown forcing inputs)
- Observation error

Types of Uncertainty
[Kennedy and O'Hagan, 2001]

- Parameter uncertainty (eg. unknown parametrisation settings)
- Model inadequacy (eg. missing feedbacks)
- Residual variability (eg. all of weather, annual / decadal oscillations)
- Parametric variability (eg. unknown forcing inputs)
- Observation error
- Code uncertainty (eg. unknown output at certain locations)

Section 3

Gaussian Processes

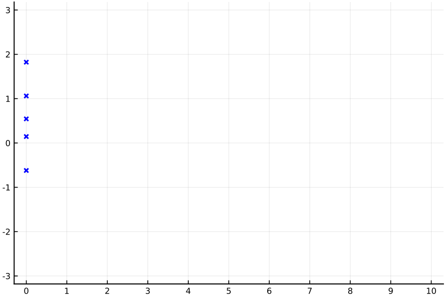
Gaussian processes

Interesting properties of GPs / why bother?

- Flexible, interpretable, uncertainty-aware, probabilistic models for functions
- Combine simple GPs to construct complicated GPs
- Natural data-efficient way to infer hyperparameters
- Exact Bayesian inference tractable for small-medium data sets
- Good / excellent approximations available for large data sets
- See GPML textbook [Rasmussen and Williams, 2006] for a thorough (ML-centric) introduction

Gaussian processes

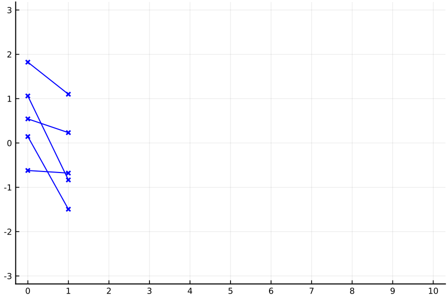
Multivariate Gaussians



$$\begin{bmatrix} 1.0 \end{bmatrix}$$

Gaussian processes

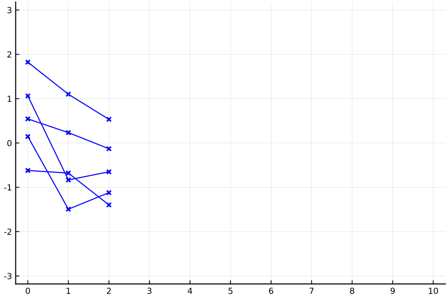
Multivariate Gaussians



$$\begin{bmatrix} 1.0 & 0.61 \\ 0.61 & 1.0 \end{bmatrix}$$

Gaussian processes

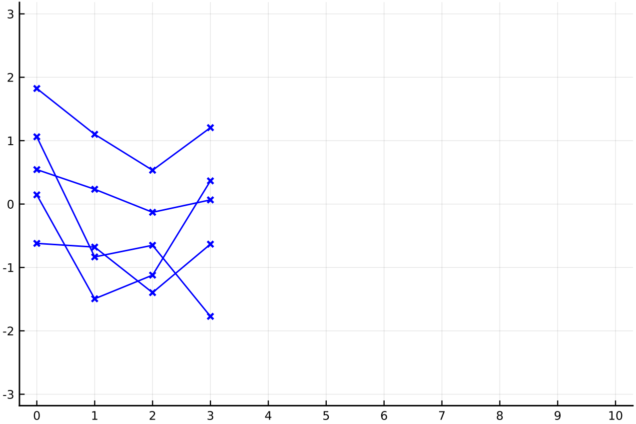
Multivariate Gaussians



$$\begin{bmatrix} 1.0 & 0.61 & 0.14 \\ 0.61 & 1.0 & 0.61 \\ 0.14 & 0.61 & 1.0 \end{bmatrix}$$

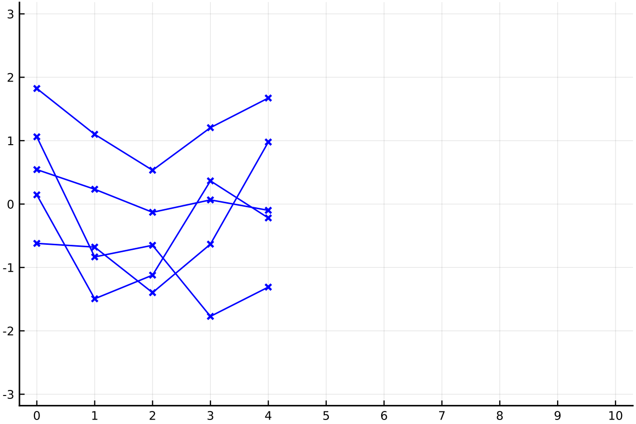
Gaussian processes

Multivariate Gaussians



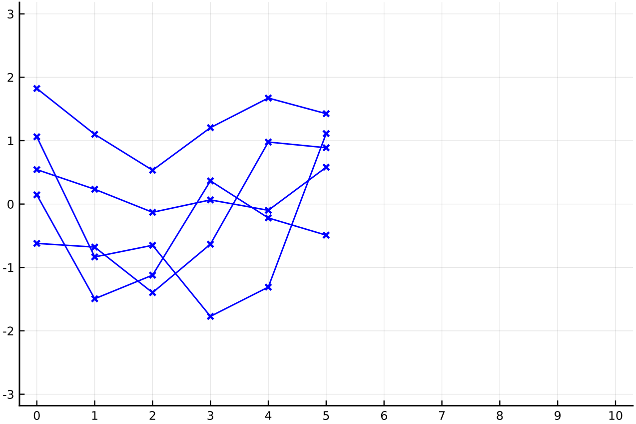
Gaussian processes

Multivariate Gaussians



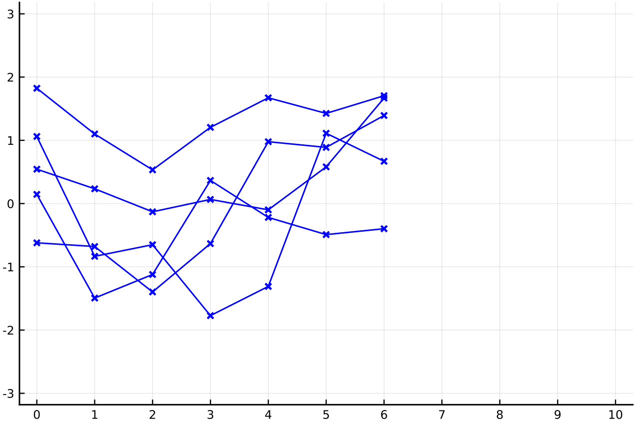
Gaussian processes

Multivariate Gaussians



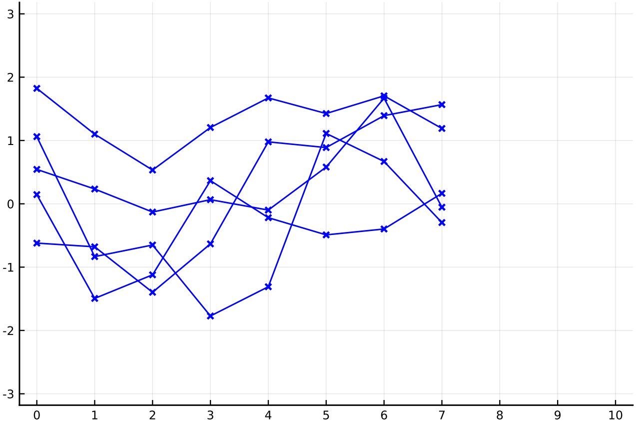
Gaussian processes

Multivariate Gaussians



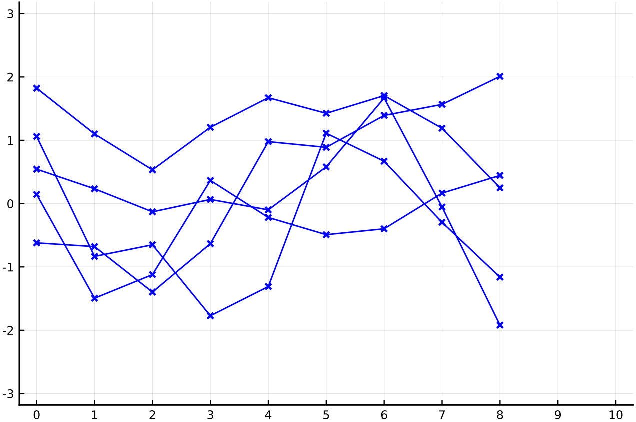
Gaussian processes

Multivariate Gaussians



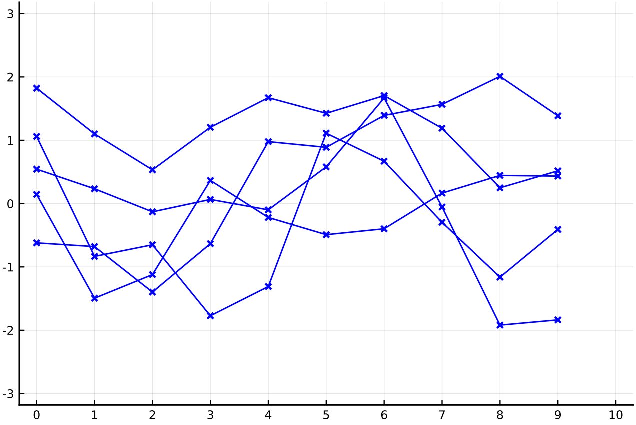
Gaussian processes

Multivariate Gaussians



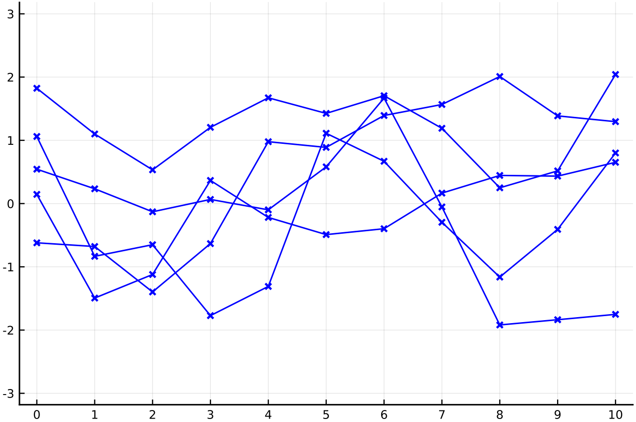
Gaussian processes

Multivariate Gaussians



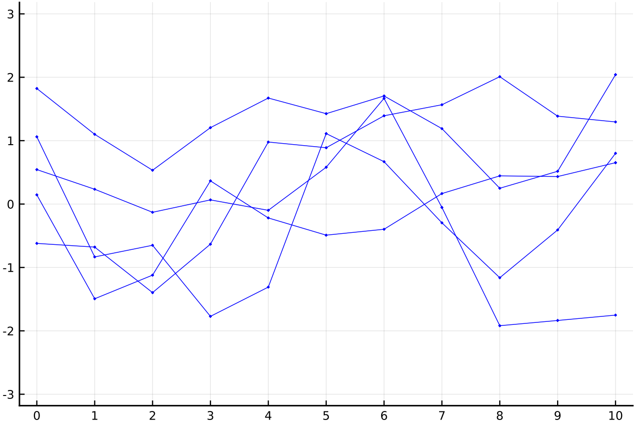
Gaussian processes

Multivariate Gaussians



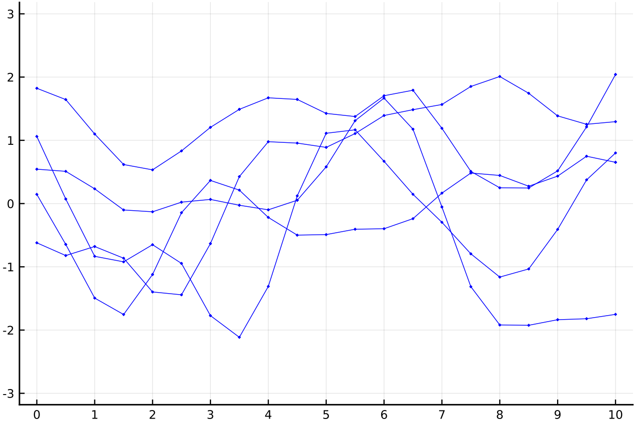
Gaussian processes

Multivariate Gaussians



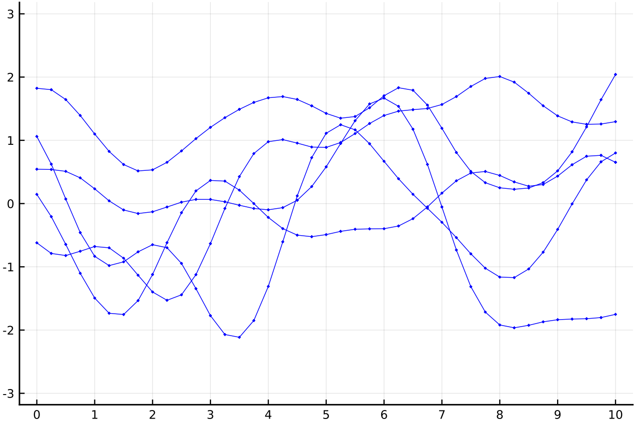
Gaussian processes

Multivariate Gaussians



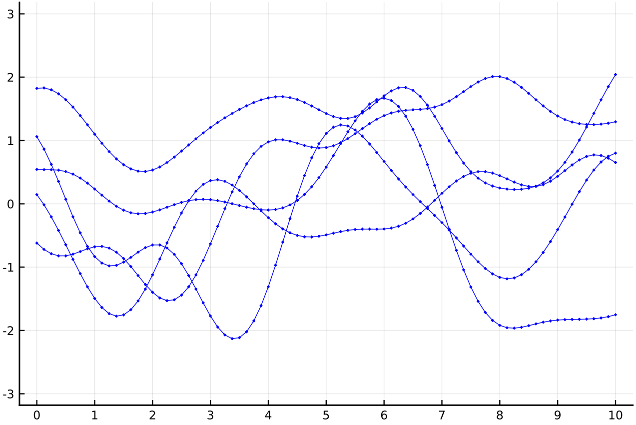
Gaussian processes

Multivariate Gaussians



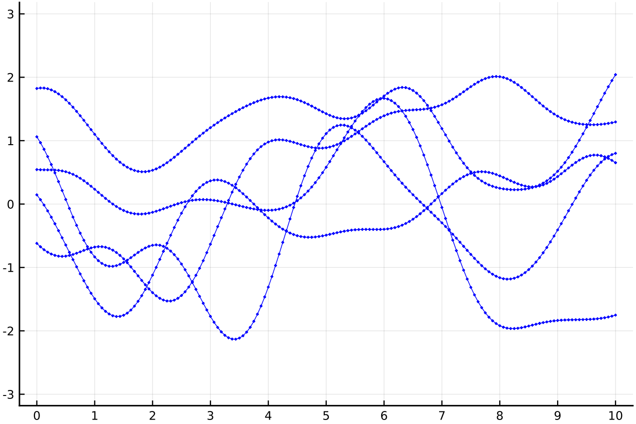
Gaussian processes

Multivariate Gaussians



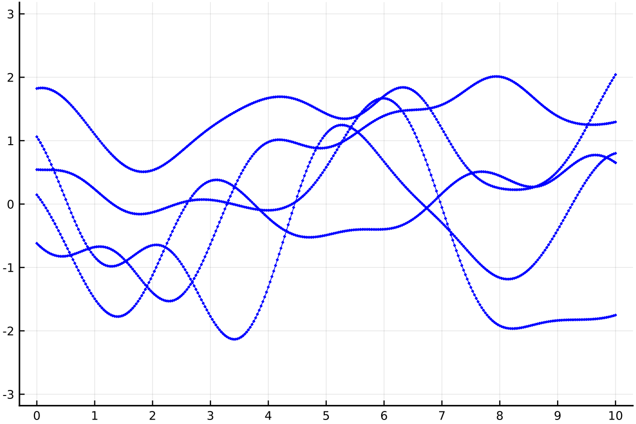
Gaussian processes

Multivariate Gaussians



Gaussian processes

Multivariate Gaussians



Gaussian processes

From Multivariate Gaussians to Gaussian Processes - Construction

Multivariate Gaussian

$$\mathsf{f} \sim \mathcal{N}(\mu, \Sigma)$$

$$\mu \in \mathbb{R}^D$$
$$\Sigma \in \mathbb{R}^{D \times D}$$

Gaussian Process

$$f \sim \mathcal{GP}(m, \kappa)$$

$$m : \mathbb{R} \to \mathbb{R}$$
$$\kappa : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$$

Gaussian processes

From Multivariate Gaussians to Gaussian Processes - Construction

Let $\mathbf{x} \in \mathbb{R}^N$ be a vector of input locations, then

$$f(\mathbf{x}) \sim \mathcal{N}(\mathbf{m}, \mathbf{C})$$

where

$$\begin{aligned} \mathbf{m}_n &:= m(\mathbf{x}_n) \\ \mathbf{C}_{nm} &:= \kappa(\mathbf{x}_n, \mathbf{x}_m) \end{aligned}$$

(Follows from the marginalisation property of Gaussians)

Gaussian processes

From Multivariate Gaussians to Gaussian Processes - Conditioning

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{g} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_{\mathbf{f}} \\ \mu_{\mathbf{g}} \end{bmatrix}, \begin{bmatrix} \Sigma_{\mathbf{ff}} & \Sigma_{\mathbf{fg}} \\ \Sigma_{\mathbf{gf}} & \Sigma_{\mathbf{gg}} \end{bmatrix} \right)$$

Gaussian processes

From Multivariate Gaussians to Gaussian Processes - Conditioning

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{g} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_{\mathbf{f}} \\ \mu_{\mathbf{g}} \end{bmatrix}, \begin{bmatrix} \Sigma_{\mathbf{ff}} & \Sigma_{\mathbf{fg}} \\ \Sigma_{\mathbf{gf}} & \Sigma_{\mathbf{gg}} \end{bmatrix} \right)$$

$$\implies \begin{bmatrix} \mathbf{f} \\ \mathbf{g} \end{bmatrix} | \mathbf{f} \sim \mathcal{N}(\mu', \Sigma')$$

Gaussian processes

From Multivariate Gaussians to Gaussian Processes - Conditioning

$$f \sim \mathcal{GP}(m, \kappa)$$

Gaussian processes

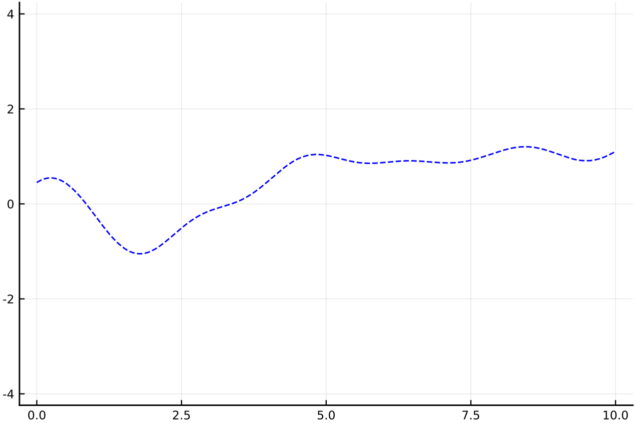
From Multivariate Gaussians to Gaussian Processes - Conditioning

$$f \sim \mathcal{GP}(m, \kappa)$$

$$\implies f | f(\mathbf{x}) \sim \mathcal{GP}\big(m', \kappa'\big)$$

Gaussian processes

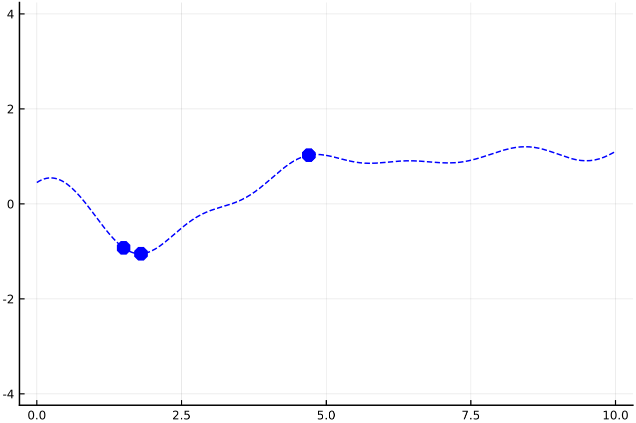
Non-Linear Regression



$$m(x) := 0, \quad \kappa\big(x, x'\big) := \exp\big(-(x - x')^2/2\big)$$

Gaussian processes

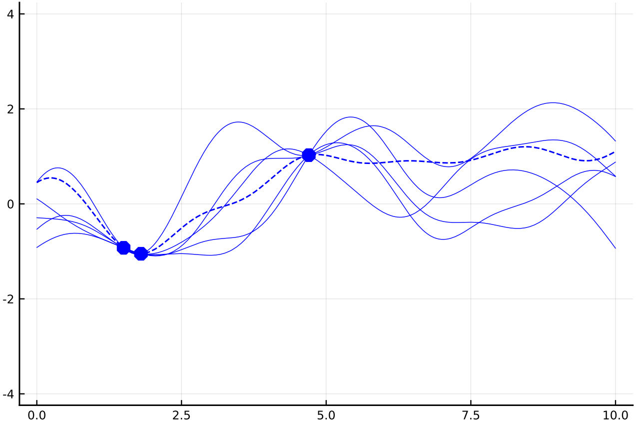
Non-Linear Regression



$$m(x) := 0, \quad \kappa\big(x, x'\big) := \exp\big(-(x - x')^2/2\big)$$

Gaussian processes

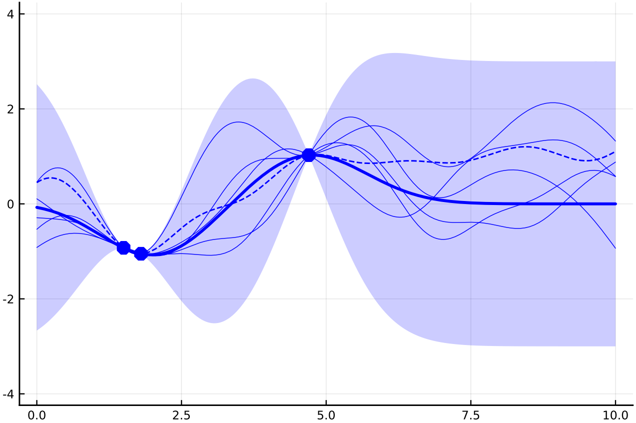
Non-Linear Regression



$$m(x) := 0, \quad \kappa\big(x, x'\big) := \exp\big(-(x - x')^2/2\big)$$

Gaussian processes

Non-Linear Regression



$$m(x) := 0, \quad \kappa\big(x, x'\big) := \exp\big(-(x - x')^2/2\big)$$

Transformations of GPs

Linear (and affine) transformations of GPs yield GPs e.g.

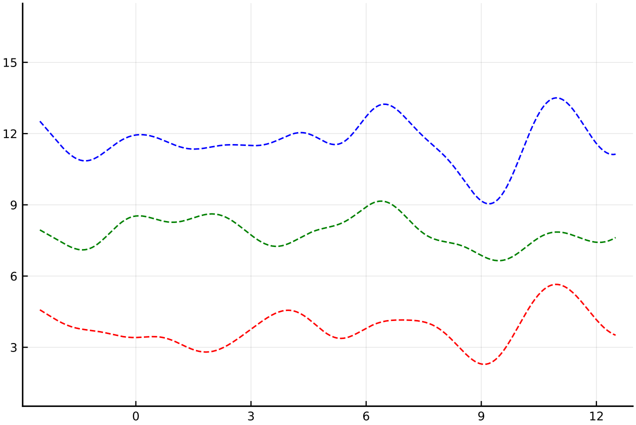
$$
\begin{aligned}
\text{addition:}\quad & f_3(x) := f_1(x) + f_2(x) \\
\text{scaling:}\quad & f_2(x) := a f_1(x) \\
\text{differentiation:}\quad & f_2(x) := \mathrm{d} f_1(x) / \mathrm{d}x \\
\text{integration:}\quad & f_2(x) := \int_l^x f_1(s) \, \mathrm{d}s
\end{aligned}
$$

Also conditioning, indexing, convolution, composition with deterministic functions, translation, etc

Transformations of GPs

Worked Example: Addition

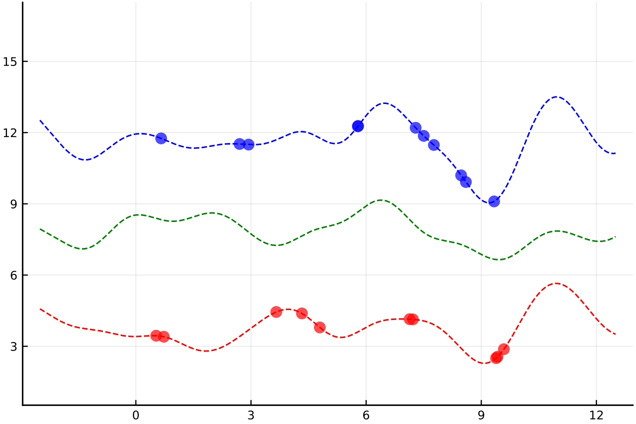
$$f_3 = f_1 + f_2$$



Transformations of GPs

Worked Example: Addition

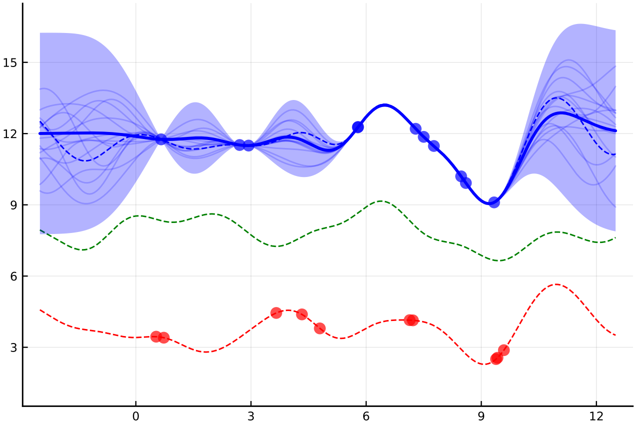
$$f_3 = f_1 + f_2$$



Transformations of GPs

Worked Example: Addition

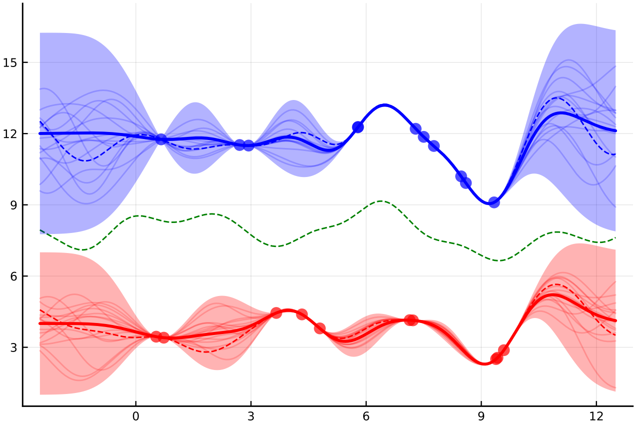
$$\boxed{f_3} = \boxed{f_1} + \boxed{f_2}$$



Transformations of GPs

Worked Example: Addition

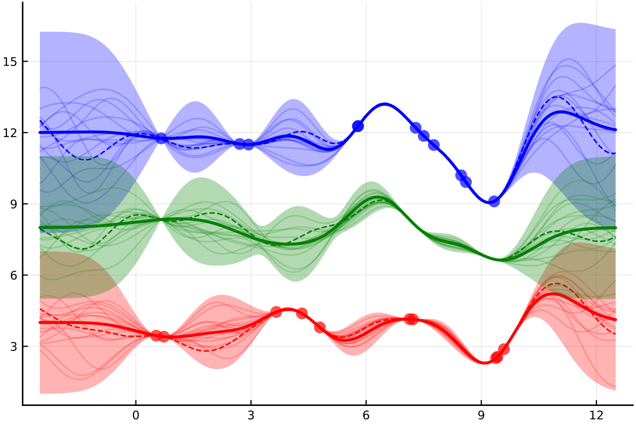
$$\boxed{f_3} = \boxed{f_1} + \boxed{f_2}$$



Transformations of GPs

Worked Example: Addition

$$\boxed{f_3} = \boxed{f_1} + \boxed{f_2}$$



Gaussian processes

- GPs have scalability issues if implemented naively
- (One) frontier: global-scale spatio-temporal phenomena

Section 4

A Model for GCM Combination

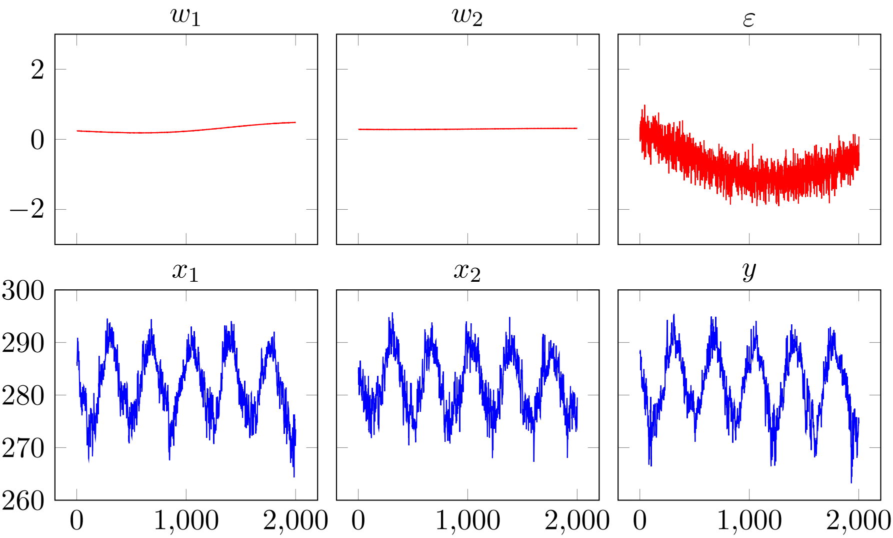
A Model for GCM Combination

Probabilistic Model

$$
\begin{aligned}
x_p &:= \text{Output of } p^{\text{th}} \text{ GCM} \\
w_p &\sim \mathcal{GP}\left(0, \kappa_p^{\mathrm{w}}\right), \quad p \in \{1, ..., P\} \\
f(t) &:= \sum_{p=1}^{P} w_p(t)\, x_p(t)
\end{aligned}
$$

$$
\begin{aligned}
\varepsilon &\sim \mathcal{GP}(0, \kappa^{\varepsilon}) \\
y(t) &:= f(t) + \varepsilon(t)
\end{aligned}
$$

A Model for GCM Combination



A Model for GCM Combination

Concrete Set Up

- Observe $\mathbf{x}_p := [x_p(t_1), ..., x_p(t_N)]^\top$
- Observe $\mathbf{y} := [y(t_1), ..., y(t_N)]^\top$
- Infer w_1, ..., w_P, and ε
- Learn kernel parameters

$$\underset{\theta}{\operatorname{argmax}} \quad \log p(\mathbf{y} \mid \mathbf{x}_{1:P}, \theta)$$

A Model for GCM Combination
Comments

- Correlated weather
- Allows for time-varying weights
- Jointly Gaussian, so inference tractable

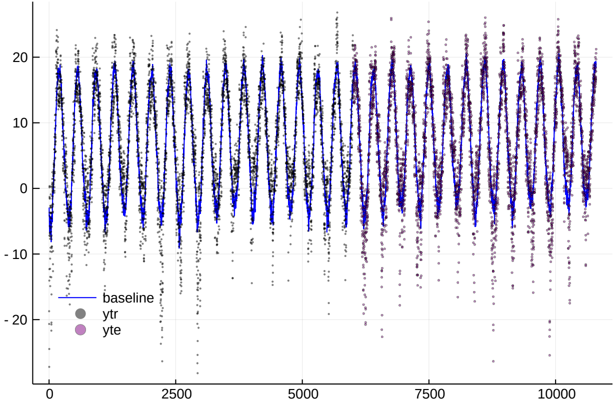
Section 5

Results

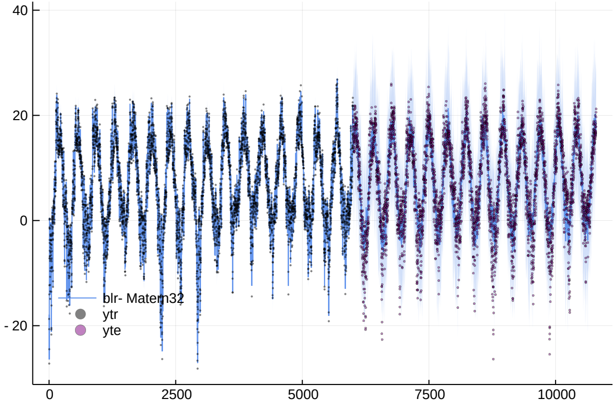
Results

- 28 GCMs, CMIP5, AMIP
- Era Interim
- Roughly 30 year's worth of data (10800 days)
- 6000 train, 4800 test

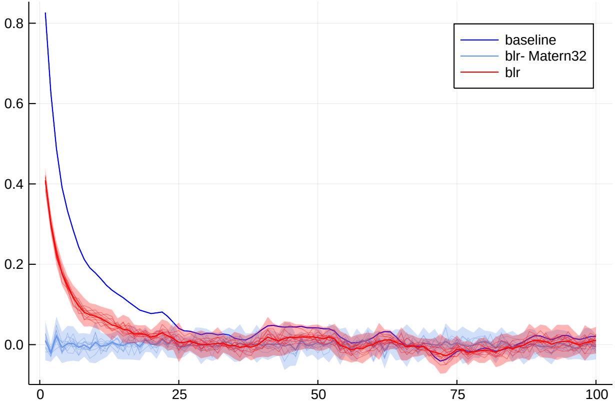
Results



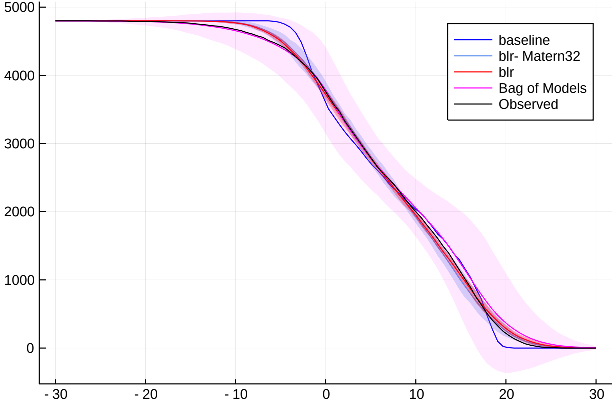
Results



Results

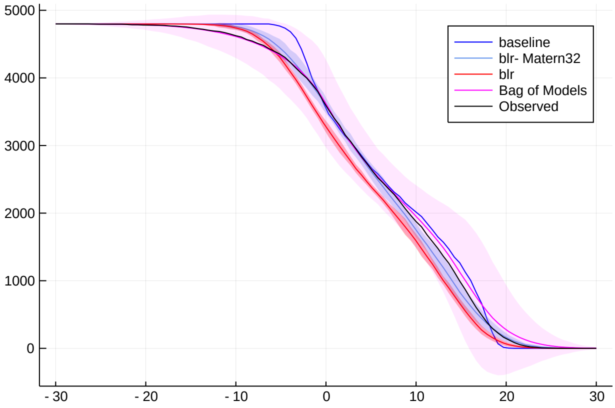


Results
1



5000
4000
3000
2000
1000
0
- 30
- 20
- 10
0
10
20
30
baseline
blr- Matern32
blr
Bag of Models
Observed

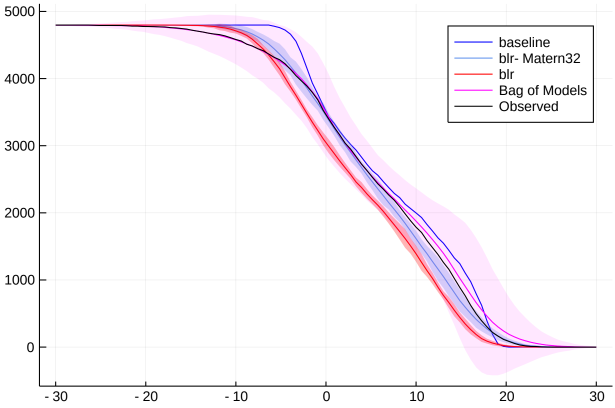
Results
2



5000
4000
3000
2000
1000
0
- 30
- 20
- 10
0
10
20
30
baseline
blr- Matern32
blr
Bag of Models
Observed

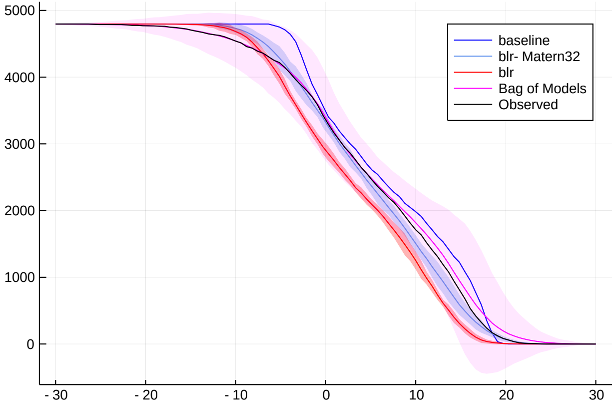
Results

3


5000
4000
3000
2000
1000
0
- 30
- 20
- 10
0
10
20
30
baseline
blr- Matern32
blr
Bag of Models
Observed

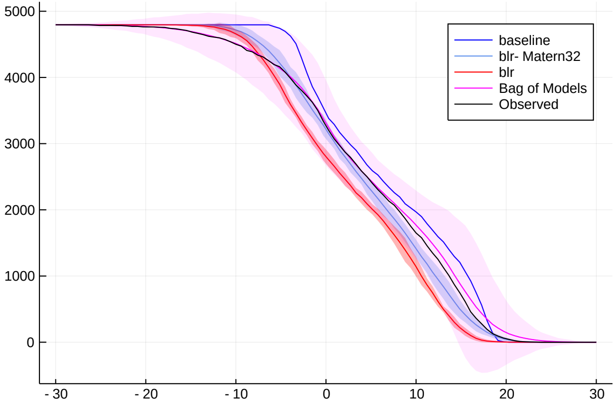
Results

4


5000
4000
3000
2000
1000
0
- 30
- 20
- 10
0
10
20
30
baseline
blr- Matern32
blr
Bag of Models
Observed

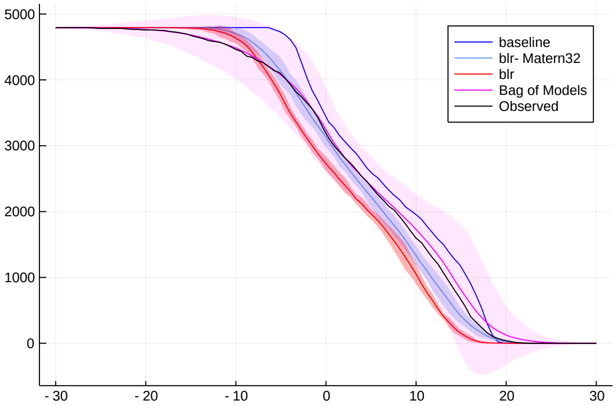
Results

5


5000
4000
3000
2000
1000
0
- 30
- 20
- 10
0
10
20
30
baseline
blr- Matern32
blr
Bag of Models
Observed

Results

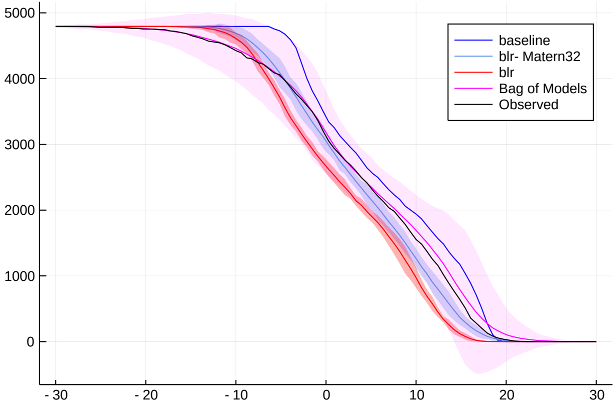
6



5000
4000
3000
2000
1000
0
- 30
- 20
- 10
0
10
20
30
baseline
blr- Matern32
blr
Bag of Models
Observed

Results

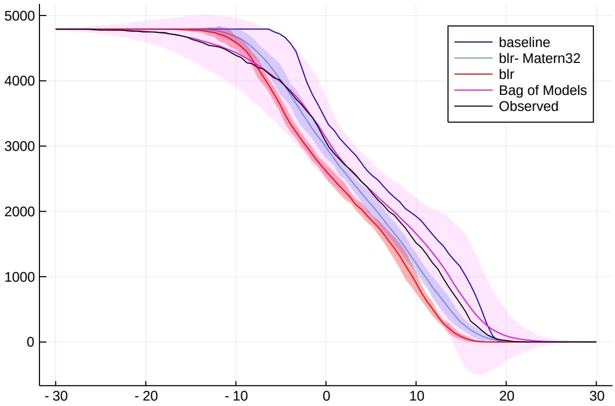
7



5000
4000
3000
2000
1000
0
- 30
- 20
- 10
0
10
20
30
baseline
blr- Matern32
blr
Bag of Models
Observed

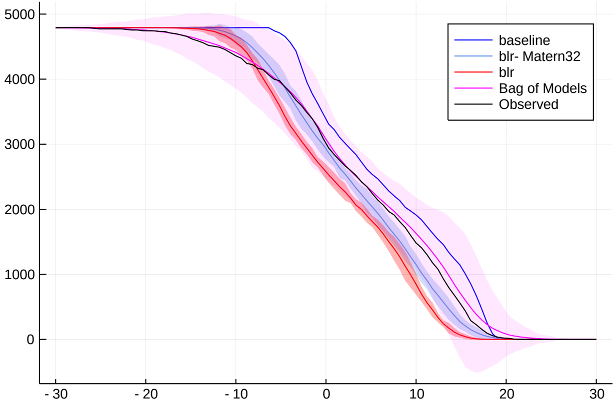
Results

8


5000
4000
3000
2000
1000
0
- 30
- 20
- 10
0
10
20
30
baseline
blr- Matern32
blr
Bag of Models
Observed

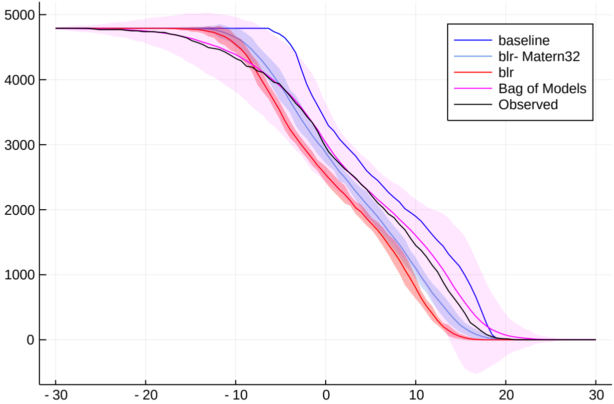
Results

9


5000
4000
3000
2000
1000
0
- 30
- 20
- 10
0
10
20
30
baseline
blr- Matern32
blr
Bag of Models
Observed

Results

10


5000
4000
3000
2000
1000
0
- 30
- 20
- 10
0
10
20
30
baseline
blr- Matern32
blr
Bag of Models
Observed

Section 6

Conclusion

Conclusion

- Certain aspects of GCM uncertainty can be addressed with machine learning / statistical techniques
- Gaussian processes are a possible candidate for temperature - fine-tuning required for optimal performance
- Quite efficient inference for time-series
- Open problem to scale to very large spatio-temporal problems
- Hierarchical models containing GPs needed for eg. precipitation
- Side information?

Conclusion

- Model ensembling
- Downscaling
- Statistical weather modelling
- Single statistical model for all three?

Conclusion

- Big up-tick in interest in climate science within ML

Bibliography I

Chandler, R. E. (2013).
Exploiting strength, discounting weakness: combining information from multiple climate simulators.
Phil. Trans. R. Soc. A, 371(1991):20120388.

Giorgi, F. and Mearns, L. O. (2002).
Calculation of average, uncertainty range, and reliability of regional climate changes from aogcm simulations via the "reliability ensemble averaging"(rea) method.
Journal of Climate, 15(10):1141–1158.

Hartikainen, J. et al. (2013).
Sequential inference for latent temporal gaussian process models.

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

Krishnamurti, T., Kishtawal, C., LaRow, T. E., Bachiochi, D. R., Zhang, Z., Williford, C. E., Gadgil, S., and Surendran, S. (1999).
Improved weather and seasonal climate forecasts from multimodel superensemble.
Science, 285(5433):1548–1550.

McKay, M. D., Beckman, R. J., and Conover, W. J. (1979).
Comparison of three methods for selecting values of input variables in the analysis of output from a computer code.
Technometrics, 21(2):239–245.

Bibliography II

McQuade, S. and Monteleoni, C. (2012).
Global climate model tracking using geospatial neighborhoods.
In *AAAI*.

Monteleoni, C., Schmidt, G. A., Saroha, S., and Asplund, E. (2011).
Tracking climate models.
Statistical Analysis and Data Mining: The ASA Data Science Journal, 4(4):372–392.

Nychka, D. and Tebaldi, C. (2003).
Comments on "calculation of average, uncertainty range, and reliability of regional climate changes from aogcm simulations via the 'reliability ensemble averaging'(rea) method".
Journal of Climate, 16(5):883–884.

Rasmussen, C. E. and Williams, C. K. (2006).
Gaussian processes for machine learning.
the MIT Press, 2(3):4.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989).
Design and analysis of computer experiments.
Statistical science, pages 409–423.

Särkkä, S. and Solin, A. (2019).
Applied stochastic differential equations, volume 10.
Cambridge University Press.

Solin, A. et al. (2016).
Stochastic differential equation methods for spatio-temporal gaussian process regression.

Bibliography III

Stainforth, D. A., Allen, M. R., Tredger, E. R., and Smith, L. A. (2007).
Confidence, uncertainty and decision-support relevance in climate predictions.
Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 365(1857):2145–2161.

Section 7

Bibliographic Notes

Bibliographic Notes

GPs as Linear SDEs

- Final chapter of [Särkkä and Solin, 2019]
- Arno's these [Solin et al., 2016]
- Jouni Hartikainen's thesis: [Hartikainen et al., 2013]

Bibliographic Notes

Combining GCM Predictions

- [Stainforth et al., 2007]: provides strong (not unreasonable) objections this entire line of work. [Chandler, 2013] provides a pragmatic alternative view. This pair of papers are the most important to read in my opinion. They elucidate all of the things that can go wrong, and that you therefore need to be aware of.
- Early work: [Krishnamurti et al., 1999], [Giorgi and Mearns, 2002], [Nychka and Tebaldi, 2003].
- [Monteleoni et al., 2011] - an interesting approach, the first (to my knowledge) with a time-varying combination of models. [McQuade and Monteleoni, 2012] extends to time + space varying.

Bibliographic Notes

Foundational Work on Simulators

Work on ways to "correct" the output of simulators has a long history. Some

- [McKay et al., 1979] - the earliest work I could find on the matter
- [Sacks et al., 1989] - very influential early work.
- [Kennedy and O'Hagan, 2001] - early uncertainty quantification work. A must read to get a good understanding of the issues.

Section 8

Gaussian processes for Time Series

GPs for Time Series

Preliminaries

GPs for Time Series

Preliminaries

$$f \sim \mathcal{GP}(0, \kappa)$$

GPs for Time Series

Preliminaries

$$f \sim \mathcal{GP}(0, \kappa)$$
$$\mathbf{f} := [f(t_1), ..., f(t_N)]^\top \sim \mathcal{N}(0, \mathbf{C})$$

GPs for Time Series

Preliminaries

$$f \sim \mathcal{GP}(0, \kappa)$$

$$\mathbf{f} := [f(t_1), ..., f(t_N)]^\top \sim \mathcal{N}(0, \mathbf{C})$$

$$\text{where } \mathbf{C}_{nm} := \kappa(t_n, t_m)$$

GPs for Time Series

Preliminaries

$$f \sim \mathcal{GP}(0, \kappa)$$

$$\mathbf{f} := [f(t_1), ..., f(t_N)]^\top \sim \mathcal{N}(0, \mathbf{C})$$

where $\mathbf{C}_{nm} := \kappa(t_n, t_m)$

$\varepsilon \sim \mathcal{N}(0, \Sigma)$, assume Σ is diagonal

GPs for Time Series

Preliminaries

$$f \sim \mathcal{GP}(0, \kappa)$$

$$\mathbf{f} := [f(t_1), ..., f(t_N)]^\top \sim \mathcal{N}(0, \mathbf{C})$$

where $\mathbf{C}_{nm} := \kappa(t_n, t_m)$

$\varepsilon \sim \mathcal{N}(0, \Sigma)$, assume Σ is diagonal

$$\mathbf{y} = \mathbf{f} + \varepsilon \sim \mathcal{N}(0, \mathbf{C} + \Sigma)$$

GPs for Time Series

Asymptotic Complexity

GPs for Time Series

Asymptotic Complexity

$\mathcal{O}(N^3)$ temporal complexity

GPs for Time Series

Asymptotic Complexity

$\mathcal{O}(N^3)$ temporal complexity
$\mathcal{O}(N^2)$ spatial complexity

GPs for Time Series

Ooooo that takes a while

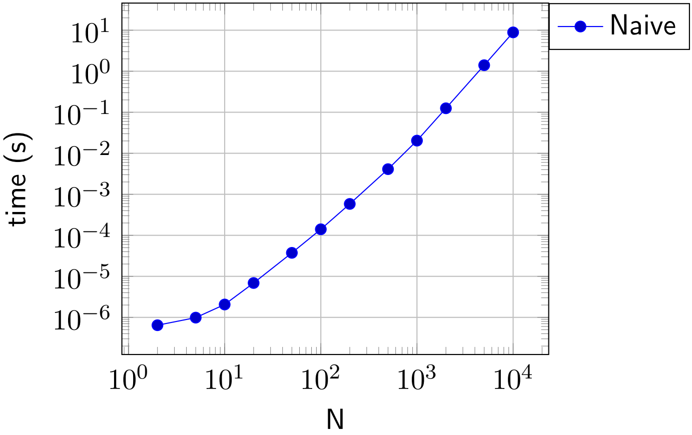


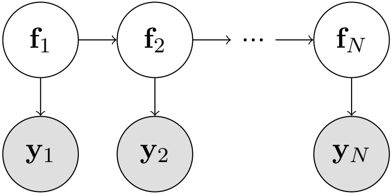
Figure 2: Naive log marginal likelihood computation requires $\mathcal{O}(N^3)$ time. Single thread.

Section 9

State Space Inference for Gaussian Processes

Linear Gaussian SSMs

What are they?

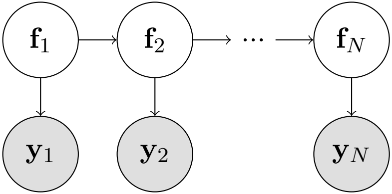


$$\mathbf{f}_n, \mathbf{q}_n \in \mathbb{R}^{D_{\text{lat}}}$$
$$\mathbf{y}_n, \mathbf{r}_n \in \mathbb{R}^{D_{\text{obs}}}$$

$$\mathbf{q}_n \sim \mathcal{N}(0, \mathbf{Q}_n)$$
$$\mathbf{f}_n = \mathbf{A}_n \mathbf{f}_{n-1} + \mathbf{q}_n$$
$$\mathbf{r}_n \sim \mathcal{N}(0, \mathbf{R}_n)$$
$$\mathbf{y}_n = \mathbf{H}_n \mathbf{f}_n + \mathbf{r}_n$$

Linear Gaussian SSMs

What are they?



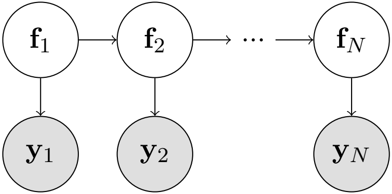
- Specified by $(\mathbf{A}_n, \mathbf{Q}_n, \mathbf{H}_n, \mathbf{R}_n)_{n=1}^{N}$

$$\mathbf{f}_n, \mathbf{q}_n \in \mathbb{R}^{D_{\text{lat}}}$$
$$\mathbf{y}_n, \mathbf{r}_n \in \mathbb{R}^{D_{\text{obs}}}$$

$$\mathbf{q}_n \sim \mathcal{N}(0, \mathbf{Q}_n)$$
$$\mathbf{f}_n = \mathbf{A}_n \mathbf{f}_{n-1} + \mathbf{q}_n$$
$$\mathbf{r}_n \sim \mathcal{N}(0, \mathbf{R}_n)$$
$$\mathbf{y}_n = \mathbf{H}_n \mathbf{f}_n + \mathbf{r}_n$$

Linear Gaussian SSMs

What are they?



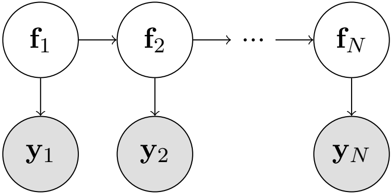
- Specified by $(\mathbf{A}_n, \mathbf{Q}_n, \mathbf{H}_n, \mathbf{R}_n)_{n=1}^N$
- $p(\mathbf{f}_{1:N}, \mathbf{y}_{1:N})$ jointly Gaussian

$$\mathbf{f}_n, \mathbf{q}_n \in \mathbb{R}^{D_{\text{lat}}}$$
$$\mathbf{y}_n, \mathbf{r}_n \in \mathbb{R}^{D_{\text{obs}}}$$

$$\mathbf{q}_n \sim \mathcal{N}(0, \mathbf{Q}_n)$$
$$\mathbf{f}_n = \mathbf{A}_n \mathbf{f}_{n-1} + \mathbf{q}_n$$
$$\mathbf{r}_n \sim \mathcal{N}(0, \mathbf{R}_n)$$
$$\mathbf{y}_n = \mathbf{H}_n \mathbf{f}_n + \mathbf{r}_n$$

Linear Gaussian SSMs

What are they?



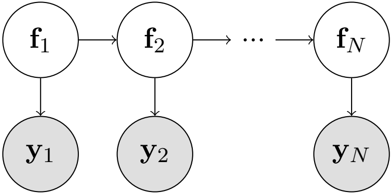
- Specified by $(\mathbf{A}_n, \mathbf{Q}_n, \mathbf{H}_n, \mathbf{R}_n)_{n=1}^N$
- $p(\mathbf{f}_{1:N}, \mathbf{y}_{1:N})$ jointly Gaussian
- Inference tractable, $\log p(\mathbf{y}_{1:N})$ has

$$\mathbf{f}_n, \mathbf{q}_n \in \mathbb{R}^{D_{\text{lat}}}$$
$$\mathbf{y}_n, \mathbf{r}_n \in \mathbb{R}^{D_{\text{obs}}}$$

$$\mathbf{q}_n \sim \mathcal{N}(0, \mathbf{Q}_n)$$
$$\mathbf{f}_n = \mathbf{A}_n \mathbf{f}_{n-1} + \mathbf{q}_n$$
$$\mathbf{r}_n \sim \mathcal{N}(0, \mathbf{R}_n)$$
$$\mathbf{y}_n = \mathbf{H}_n \mathbf{f}_n + \mathbf{r}_n$$

Linear Gaussian SSMs

What are they?



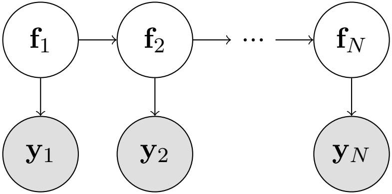
- Specified by $(\mathbf{A}_n, \mathbf{Q}_n, \mathbf{H}_n, \mathbf{R}_n)_{n=1}^N$
- $p(\mathbf{f}_{1:N}, \mathbf{y}_{1:N})$ jointly Gaussian
- Inference tractable, $\log p(\mathbf{y}_{1:N})$ has
  - $\mathcal{O}\left(N {D_{\text{obs}}}^3 {D_{\text{lat}}}^3\right)$ temporal complexity

$$\mathbf{f}_n, \mathbf{q}_n \in \mathbb{R}^{D_{\text{lat}}}$$
$$\mathbf{y}_n, \mathbf{r}_n \in \mathbb{R}^{D_{\text{obs}}}$$

$$\mathbf{q}_n \sim \mathcal{N}(0, \mathbf{Q}_n)$$
$$\mathbf{f}_n = \mathbf{A}_n \mathbf{f}_{n-1} + \mathbf{q}_n$$
$$\mathbf{r}_n \sim \mathcal{N}(0, \mathbf{R}_n)$$
$$\mathbf{y}_n = \mathbf{H}_n \mathbf{f}_n + \mathbf{r}_n$$

Linear Gaussian SSMs

What are they?



- Specified by $(\mathbf{A}_n, \mathbf{Q}_n, \mathbf{H}_n, \mathbf{R}_n)_{n=1}^N$
- $p(\mathbf{f}_{1:N}, \mathbf{y}_{1:N})$ jointly Gaussian
- Inference tractable, $\log p(\mathbf{y}_{1:N})$ has
  - $\mathcal{O}\left(N {D_{\text{obs}}}^3 {D_{\text{lat}}}^3\right)$ temporal complexity
  - $\mathcal{O}\left(N {D_{\text{obs}}}^2 {D_{\text{lat}}}^2\right)$ spatial complexity

$$\mathbf{f}_n, \mathbf{q}_n \in \mathbb{R}^{D_{\text{lat}}}$$
$$\mathbf{y}_n, \mathbf{r}_n \in \mathbb{R}^{D_{\text{obs}}}$$

$$\mathbf{q}_n \sim \mathcal{N}(0, \mathbf{Q}_n)$$
$$\mathbf{f}_n = \mathbf{A}_n \mathbf{f}_{n-1} + \mathbf{q}_n$$
$$\mathbf{r}_n \sim \mathcal{N}(0, \mathbf{R}_n)$$
$$\mathbf{y}_n = \mathbf{H}_n \mathbf{f}_n + \mathbf{r}_n$$

The Main Idea

- Convert GP f into a linear SDE
- Convert linear SDE into Linear Gaussian SSM at times $t_{1:N}$
- Do inference e.g. compute $\log p(\mathbf{y}_{1:N})$

See [Särkkä and Solin, 2019] for details

State Space Inference Techniques

Compute the log marginal likelihood

- Inputs:
  - f: a GP in terms of a kernel κ_θ
  - $t_{1:N}$: input locations
  - $\mathbf{y}_{1:N}$: corresponding observations
  - Σ: observation model covariance matrix
- Procedure:
  - Compute Linear Gaussian SSM $(\mathbf{A}_n, \mathbf{Q}_n, \mathbf{H}_n, \mathbf{R}_n)_{n=1}^N$ from κ_θ, $t_{1:N}$ and Σ
  - Compute $\log p(\mathbf{y}_{1:N})$ using Linear Gaussian SSM

Upshot

Properties

- Exact or almost exact inference
- $\mathcal{O}(ND^3)$ temporal complexity
- $\mathcal{O}(ND^2)$ spatial complexity
- D is reasonable in lots of interesting cases. Governed by smoothness of samples from the GP.

Upshot

Benchmarks

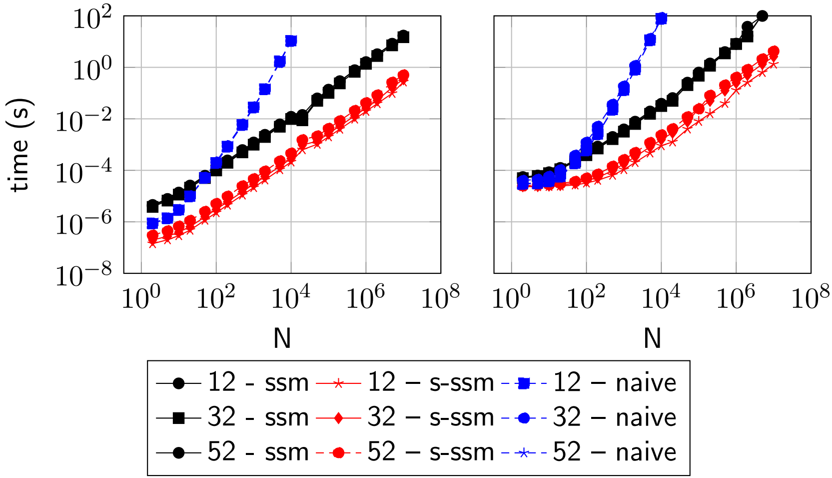
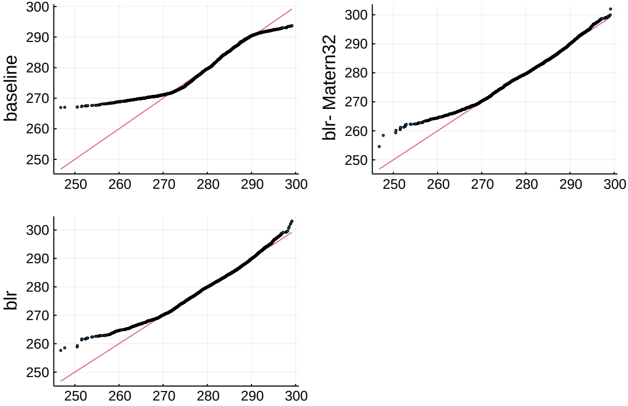


Figure 3: The total time to compute the log marginal likelihood (left) and log


baseline
blr- Matern32
blr
250
260
270
280
290
300