Chapter 1

Gaussian Conditionals

A standard result shows how to condition on knowing a subset of the dimensions $\mathbf{y}_B$ of a vector $\mathbf{y}$ having a multivariate Gaussian distribution. If

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_A \\ \mathbf{y}_B \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AB} \\ \boldsymbol{\Sigma}_{BA} & \boldsymbol{\Sigma}_{BB} \end{bmatrix} \right) \tag{1.1}$$

then

$$\mathbf{y}_A | \mathbf{y}_B \sim \mathcal{N}(\boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1} (\mathbf{x}_B - \boldsymbol{\mu}_B), \boldsymbol{\Sigma}_{AA} - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\boldsymbol{\Sigma}_{BA}). \tag{1.2}$$

This result can be used in the context of Gaussian process regression, where $\mathbf{y}_B = [f(\mathbf{x}_1), f(\mathbf{x}_2), \ldots, f(\mathbf{x}_N)]$ represents a set of function values observed at some subset of locations $[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$, while $\mathbf{y}_A = [f(\mathbf{x}_1\star), f(\mathbf{x}_2\star), \ldots, f(\mathbf{x}_N\star)]$ represents test points whose predictive distribution we'd like to know. In this case, the necessary covariance matrices are given by:

$$\boldsymbol{\Sigma}_{AA} = k(\mathbf{X}, \mathbf{X}) \tag{1.3}$$

$$\boldsymbol{\Sigma}_{AB} = k(\mathbf{X}, \mathbf{X}^\star) \tag{1.4}$$

$$\boldsymbol{\Sigma}_{BA} = k(\mathbf{X}^\star, \mathbf{X}) \tag{1.5}$$

$$\boldsymbol{\Sigma}_{BB} = k(\mathbf{X}^\star, \mathbf{X}^\star) \tag{1.6}$$

and similarily for the mean vectors.

Chapter 2

Kernel Definitions

Here we give the formulas for all one-dimensional base kernels mentioned in the thesis. Each of these formulas is multiplied by a scale factor σ_f^2, which we omit for clarity.

$$\mathrm{C}(x, x') = 1 \tag{2.1}$$

$$\mathrm{WN}(x, x') = \delta(x - x') \tag{2.2}$$

$$\mathrm{Lin}(x, x') = (x - c)(x' - c) \tag{2.3}$$

$$\mathrm{SE}(x, x') = \exp\left(-\frac{(x - x')^2}{2\ell^2}\right) \tag{2.4}$$

$$\mathrm{RQ}(x, x') = \left(1 + \frac{(x - x')^2}{2\alpha\ell^2}\right)^{-\alpha} \tag{2.5}$$

$$\mathrm{Per}(x, x') = \sigma_f^2 \frac{\exp\left(\frac{1}{\ell^2}\cos 2\pi\frac{(x-x')}{p}\right) - I_0\left(\frac{1}{\ell^2}\right)}{\exp\left(\frac{1}{\ell^2}\right) - I_0\left(\frac{1}{\ell^2}\right)} \tag{2.6}$$

$$\cos(x, x') = \cos\left(\frac{2\pi(x - x')}{p}\right) \tag{2.7}$$

$$\mathrm{CP}(k_1, k_2)(x, x') = \sigma(x)k_1(x, x')\sigma(x') + (1 - \sigma(x))k_2(x, x')(1 - \sigma(x')) \tag{2.8}$$

$$\boldsymbol{\sigma} = \sigma(x)\sigma(x') \tag{2.9}$$

$$\bar{\boldsymbol{\sigma}} = (1 - \sigma(x))(1 - \sigma(x')) \tag{2.10}$$

where $\delta_{x,x'}$ is the Kronecker delta function, I_0 is the modified Bessel function of the first kind of order zero, and other symbols are kernel parameters. Equations (2.3), (2.4) and (2.6) are plotted in **??**, and equations (2.2), (2.5) and (2.7) are plotted in **??**. Draws from GP priors with changepoint kernels are shown in **??**.

The Generalized Periodic Kernel

Lloyd (2013) showed that the standard periodic kernel due to **?** can be decomposed into a periodic and a constant component. He derived the equivalent periodic kernel without any constant component, shown in equation (2.6). He further showed that its limit as the lengthscale grows is the cosine kernel:

$$\lim_{\ell \to \infty} \mathrm{Per}(x, x') = \cos\left(\frac{2\pi(x - x')}{p}\right). \tag{2.11}$$

Separating out the constant component allows us to express negative prior covariance, as well as increasing the interpretability of the resulting models.

Chapter 3

Search Operators

The model construction phase of ABCD starts with the noise kernel, WN. New kernel expressions are generated by applying search operators to the current kernel, which replace some part of the existing kernel expression with a new kernel expression.

The search used in the multidimensional regression experiments in **????** used only the following search operators:

$$\mathcal{S} \rightarrow \mathcal{S} + \mathcal{B} \tag{3.1}$$

$$\mathcal{S} \rightarrow \mathcal{S} \times \mathcal{B} \tag{3.2}$$

$$\mathcal{B} \rightarrow \mathcal{B}' \tag{3.3}$$

where $\mathcal{S}$ represents any kernel subexpression and $\mathcal{B}$ is any base kernel within a kernel expression. These search operators represent addition, multiplication and replacement. When the multiplication operator is applied to a subexpression which includes a sum of subexpressions, parentheses () are introduced. For instance, if rule (3.2) is applied to the subexpression $k_1 + k_2$, the resulting expression is $(k_1 + k_2) \times \mathcal{B}$.

Afterwards, we added several more search operators in order to speed up the search. These new operators do not change the set of possible models.

To accommodate changepoints and changewindows, we introduced the following additional operators to our search:

$$\mathcal{S} \rightarrow \mathrm{CP}(\mathcal{S}, \mathcal{S}) \tag{3.4}$$

$$\mathcal{S} \rightarrow \mathrm{CW}(\mathcal{S}, \mathcal{S}) \tag{3.5}$$

$$\mathcal{S} \rightarrow \mathrm{CW}(\mathcal{S}, \mathrm{C}) \tag{3.6}$$

$$\mathcal{S} \rightarrow \mathrm{CW}(\mathrm{C}, \mathcal{S}) \tag{3.7}$$

where C is the constant kernel. The last two operators result in a kernel only applying outside, or within, a certain region.

To allow the search to simplify existing expressions, we introduced the following operators:

$$\mathcal{S} \rightarrow \mathcal{B} \tag{3.8}$$

$$\mathcal{S} + \mathcal{S}' \rightarrow \mathcal{S} \tag{3.9}$$

$$\mathcal{S} \times \mathcal{S}' \rightarrow \mathcal{S} \tag{3.10}$$

where $\mathcal{S}'$ represents any other kernel expression. We also introduced the operator

$$\mathcal{S} \rightarrow \mathcal{S} \times (\mathcal{B} + \mathrm{C}) \tag{3.11}$$

Which allows a new base kernel to be added along with the constant kernel, for cases when multiplying by a base kernel by itself would restrict the model too much.

Chapter 4

Example Report

The following pages of this appendix contain an entire automatically-generated report, run on a dataset measuring annual solar irradiation data from 1610 to 2011. This dataset was previously analyzed by **?**.

Other example reports can be found at `mlg.eng.cam.ac.uk/Lloyd/abcdoutput/`

1 Executive summary

The raw data and full model posterior with extrapolations are shown in figure 1.

Figure 1: Raw data (left) and model posterior with extrapolation (right)

The structure search algorithm has identified nine additive components in the data. The first 4 additive components explain 92.3% of the variation in the data as shown by the coefficient of determination (R^2) values in table 1. The first 8 additive components explain 99.2% of the variation in the data. After the first 5 components the cross validated mean absolute error (MAE) does not decrease by more than 0.1%. This suggests that subsequent terms are modelling very short term trends, uncorrelated noise or are artefacts of the model or search procedure. Short summaries of the additive components are as follows:

- A constant.
- A constant. This function applies from 1644 until 1713.
- A smooth function. This function applies until 1644 and from 1719 onwards.
- An approximately periodic function with a period of 10.8 years. This function applies until 1644 and from 1719 onwards.
- A rapidly varying smooth function. This function applies until 1644 and from 1719 onwards.
- Uncorrelated noise.
- A rapidly varying smooth function with marginal standard deviation increasing linearly away from 1843. This function applies from 1751 onwards.
- A rapidly varying smooth function. This function applies until 1644 and from 1719 until 1751.
- A constant. This function applies from 1713 until 1719.

#	R^2 (%)	ΔR^2 (%)	Residual R^2 (%)	Cross validated MAE	Reduction in MAE (%)
-	-	-	-	1360.65	-
1	0.0	0.0	0.0	0.33	100.0
2	35.3	35.3	35.3	0.23	29.4
3	72.5	37.2	57.5	0.18	20.7
4	92.3	19.9	72.2	0.15	16.4
5	97.8	5.5	71.4	0.15	0.4
6	97.8	0.0	0.2	0.15	0.0
7	98.4	0.5	24.8	0.15	-0.0
8	99.2	0.8	50.7	0.15	-0.0
9	100.0	0.8	100.0	0.15	-0.0

Table 1: Summary statistics for cumulative additive fits to the data. The residual coefficient of determination (R^2) values are computed using the residuals from the previous fit as the target values; this measures how much of the residual variance is explained by each new component. The mean absolute error (MAE) is calculated using 10 fold cross validation with a contiguous block design; this measures the ability of the model to interpolate and extrapolate over moderate distances. The model is fit using the full data so the MAE values cannot be used reliably as an estimate of out-of-sample predictive performance.

2 Detailed discussion of additive components

2.1 Component 1 : A constant

This component is constant.

This component explains 0.0% of the total variance. The addition of this component reduces the cross validated MAE by 100.0% from 1360.6 to 0.3.

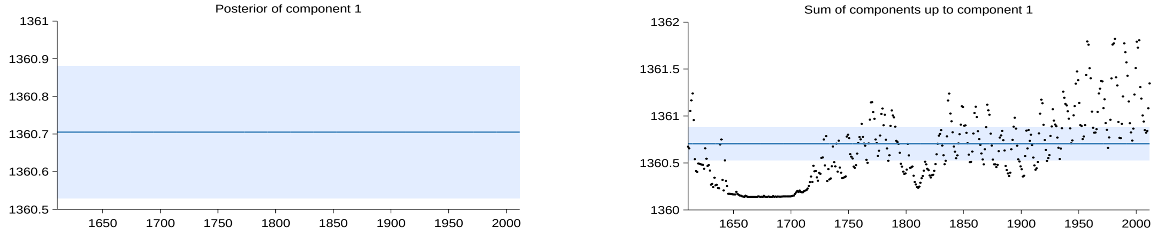


Figure 2: Posterior of component 1 (left) and the posterior of the cumulative sum of components with data (right)

2.2 Component 2 : A constant. This function applies from 1644 until 1713

This component is constant. This component applies from 1644 until 1713.

This component explains 35.3% of the residual variance; this increases the total variance explained from 0.0% to 35.3%. The addition of this component reduces the cross validated MAE by 29.42% from 0.33 to 0.23.

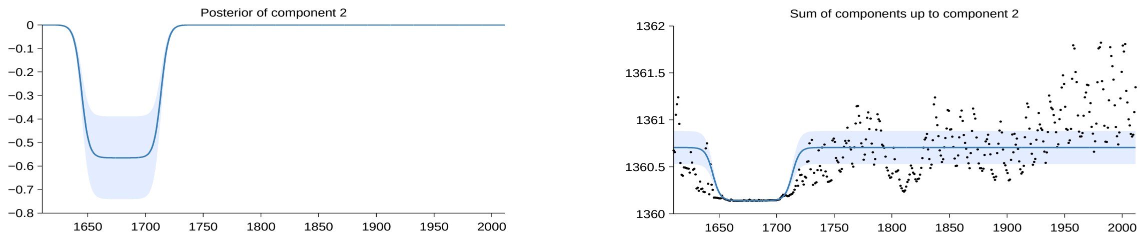


Figure 3: Posterior of component 2 (left) and the posterior of the cumulative sum of components with data (right)

2.3 Component 3 : A smooth function. This function applies until 1644 and from 1719 onwards

This component is a smooth function with a typical lengthscale of 21.9 years. This component applies until 1644 and from 1719 onwards.

This component explains 57.5% of the residual variance; this increases the total variance explained from 35.3% to 72.5%. The addition of this component reduces the cross validated MAE by 20.66% from 0.23 to 0.18.

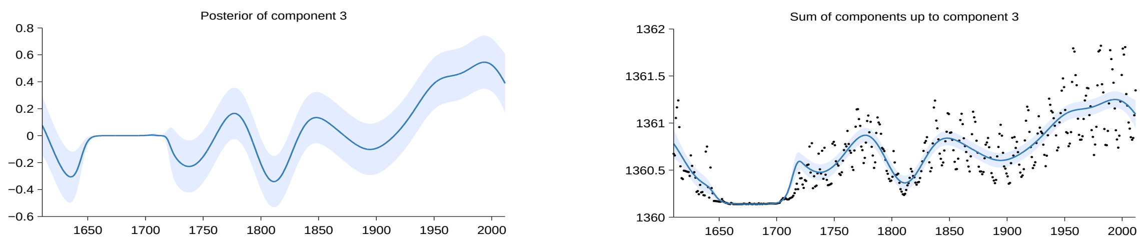


Figure 4: Posterior of component 3 (left) and the posterior of the cumulative sum of components with data (right)

2.4 Component 4 : An approximately periodic function with a period of 10.8 years. This function applies until 1644 and from 1719 onwards

This component is approximately periodic with a period of 10.8 years. Across periods the shape of the function varies smoothly with a typical lengthscale of 33.2 years. The shape of the function within each period has a typical lengthscale of 12.6 years. This component applies until 1644 and from 1719 onwards.

This component explains 72.2% of the residual variance; this increases the total variance explained from 72.5% to 92.3%. The addition of this component reduces the cross validated MAE by 16.42% from 0.18 to 0.15.

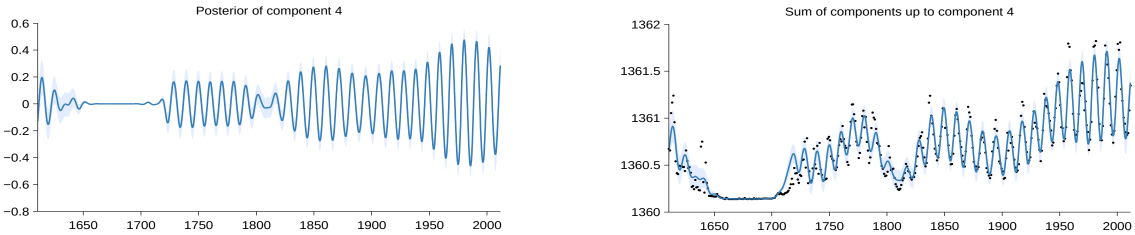


Figure 5: Posterior of component 4 (left) and the posterior of the cumulative sum of components with data (right)

2.5 Component 5 : A rapidly varying smooth function. This function applies until 1644 and from 1719 onwards

This function is a rapidly varying but smooth function with a typical lengthscale of 1.2 years. This component applies until 1644 and from 1719 onwards.

This component explains 71.4% of the residual variance; this increases the total variance explained from 92.3% to 97.8%. The addition of this component reduces the cross validated MAE by 0.41% from 0.15 to 0.15.

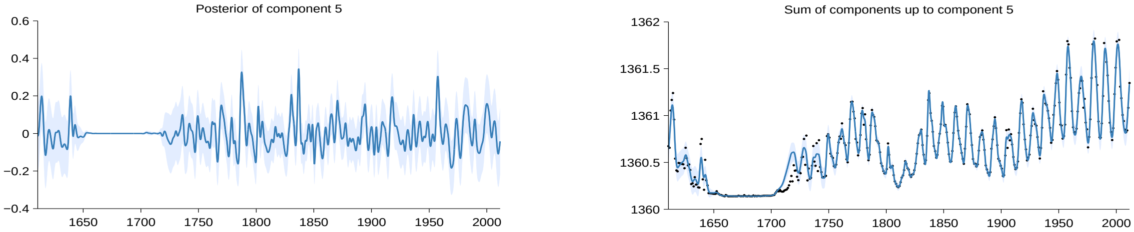


Figure 6: Posterior of component 5 (left) and the posterior of the cumulative sum of components with data (right)

2.6 Component 6 : Uncorrelated noise

This component models uncorrelated noise.

This component explains 0.2% of the residual variance; this increases the total variance explained from 97.8% to 97.8%. The addition of this component reduces the cross validated MAE by 0.00% from 0.15 to 0.15. This component explains residual variance but does not improve MAE which suggests that this component describes very short term patterns, uncorrelated noise or is an artefact of the model or search procedure.

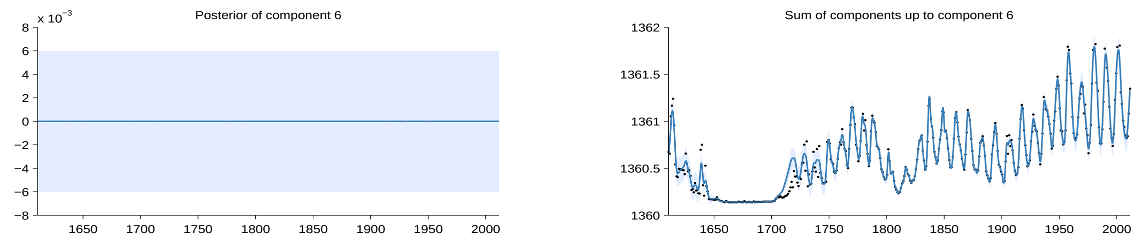


Figure 7: Posterior of component 6 (left) and the posterior of the cumulative sum of components with data (right)

2.7 Component 7 : A rapidly varying smooth function with marginal standard deviation increasing linearly away from 1843. This function applies from 1751 onwards

This function is a rapidly varying but smooth function with a typical lengthscale of 3.1 months. The marginal standard deviation of the function increases linearly away from 1843. This component applies from 1751 onwards.

This component explains 24.8% of the residual variance; this increases the total variance explained from 97.8% to 98.4%. The addition of this component increases the cross validated MAE by 0.00% from 0.15 to 0.15. This component explains residual variance but does not improve MAE which suggests that this component describes very short term patterns, uncorrelated noise or is an artefact of the model or search procedure.

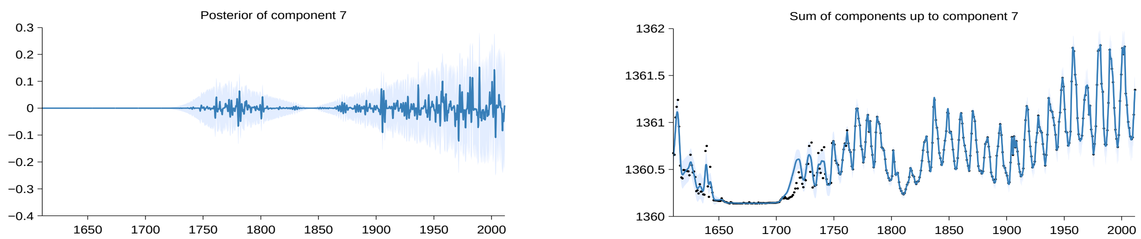


Figure 8: Posterior of component 7 (left) and the posterior of the cumulative sum of components with data (right)

2.8 Component 8 : A rapidly varying smooth function. This function applies until 1644 and from 1719 until 1751

This function is a rapidly varying but smooth function with a typical lengthscale of 3.1 months. This component applies until 1644 and from 1719 until 1751.

This component explains 50.7% of the residual variance; this increases the total variance explained from 98.4% to 99.2%. The addition of this component increases the cross validated MAE by 0.00% from 0.15 to 0.15. This component explains residual variance but does not improve MAE which suggests that this component describes very short term patterns, uncorrelated noise or is an artefact of the model or search procedure.

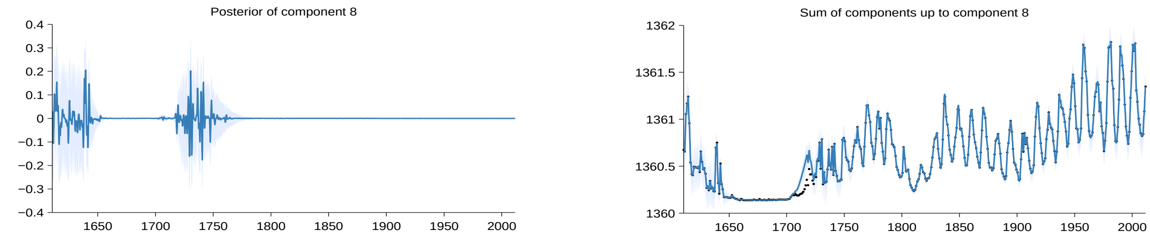


Figure 9: Posterior of component 8 (left) and the posterior of the cumulative sum of components with data (right)

2.9 Component 9 : A constant. This function applies from 1713 until 1719

This component is constant. This component applies from 1713 until 1719.

This component explains 100.0% of the residual variance; this increases the total variance explained from 99.2% to 100.0%. The addition of this component increases the cross validated MAE by 0.01% from 0.15 to 0.15. This component explains residual variance but does not improve MAE which suggests that this component describes very short term patterns, uncorrelated noise or is an artefact of the model or search procedure.

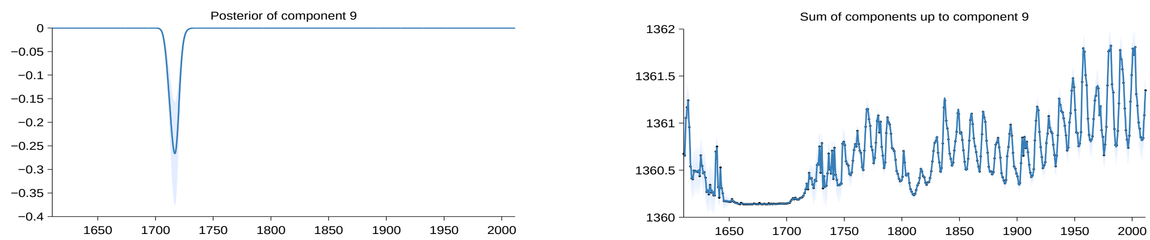
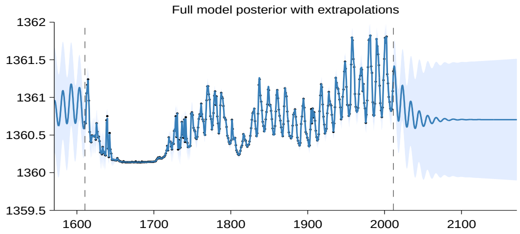


Figure 10: Posterior of component 9 (left) and the posterior of the cumulative sum of components with data (right)

3 Extrapolation

Summaries of the posterior distribution of the full model are shown in figure 11. The plot on the left displays the mean of the posterior together with pointwise variance. The plot on the right displays three random samples from the posterior.



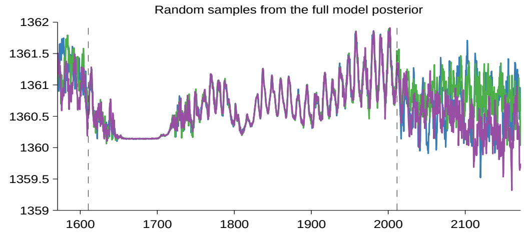


Figure 11: Full model posterior. Mean and pointwise variance (left) and three random samples (right)

References

James Robert Lloyd. personal communication, 2013. (page 3)

Nutonian. Eureqa, 2011. URL http://www.nutonian.com/.