
An automatic report for the dataset : 02-solar

The Gaussian Process Structure Search Summarizer
University of Cambridge, Massachusetts Institute of Technology

Abstract

This report was produced automatically by the Gaussian process structure search algorithm. See <http://arxiv.org/abs/1302.4922> for a preliminary paper and see <https://github.com/jamesrobertlloyd/gpss-research> for the latest source code.

1 Executive summary

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

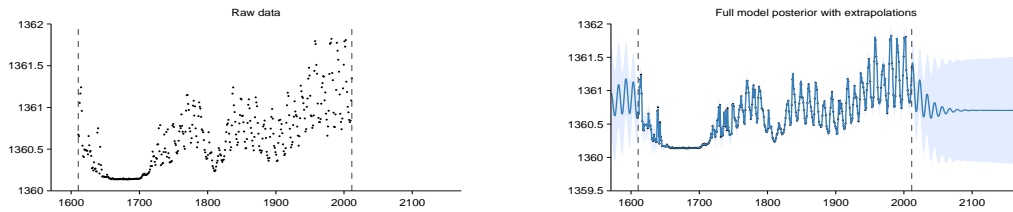


Figure 1: Raw data (left) and model posterior with extrapolation (right)

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

- A constant. The 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.

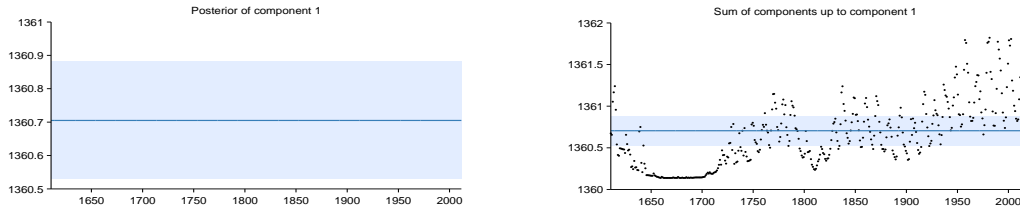


Figure 2: Posterior of component 1 (left) and the posterior of the cumulative sum of components with data (right)

2.2 Component 2 : A constant. The 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.

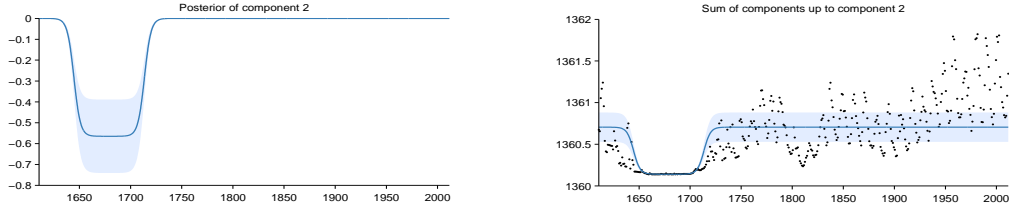


Figure 3: Posterior of component 2 (left) and the posterior of the cumulative sum of components with data (right)

2.3 Component 3 : A smooth function. The 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.

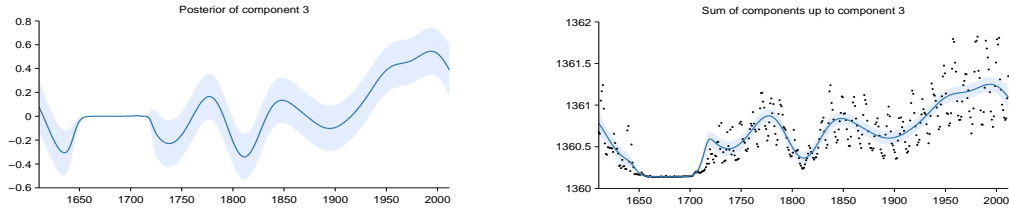


Figure 4: Posterior of component 3 (left) and the posterior of the cumulative sum of components with data (right)

2.4 Component 4 : An approximately periodic function with a period of 10.8 years. The 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.

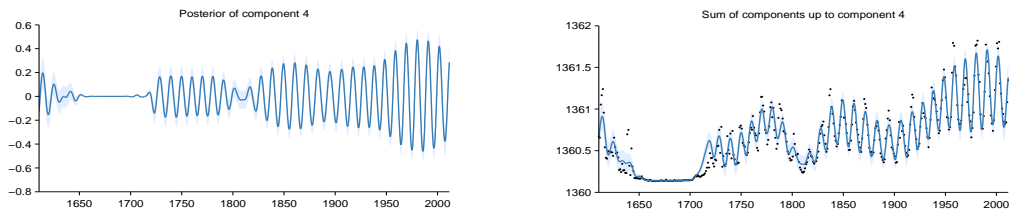


Figure 5: Posterior of component 4 (left) and the posterior of the cumulative sum of components with data (right)

2.5 Component 5 : A rapidly varying smooth function. The 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.

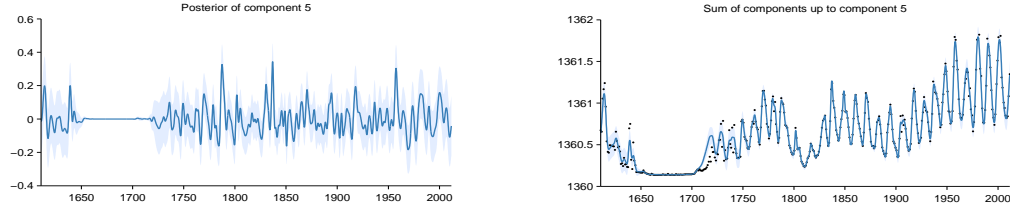


Figure 6: Posterior of component 5 (left) and the posterior of the cumulative sum of components with data (right)

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.

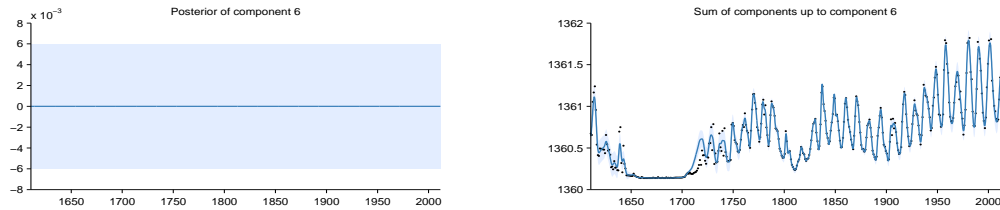


Figure 7: Posterior of component 6 (left) and the posterior of the cumulative sum of components with data (right)

2.7 Component 7 : A rapidly varying smooth function with marginal standard deviation increasing linearly away from 1843. The 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.

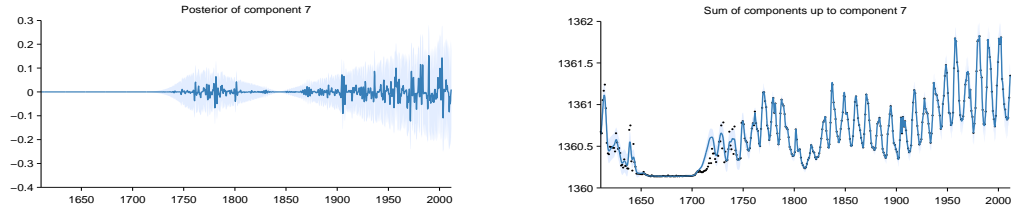


Figure 8: Posterior of component 7 (left) and the posterior of the cumulative sum of components with data (right)

2.8 Component 8 : A rapidly varying smooth function. The 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.

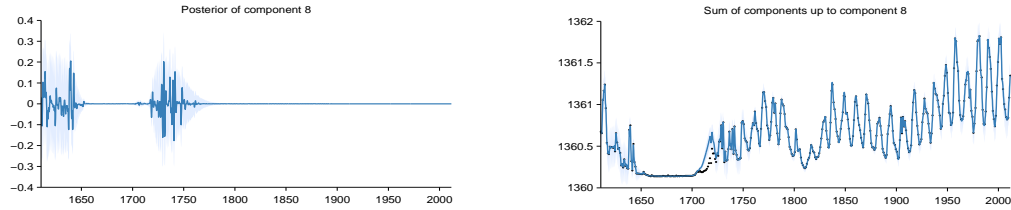


Figure 9: Posterior of component 8 (left) and the posterior of the cumulative sum of components with data (right)

2.9 Component 9 : A constant. The 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.

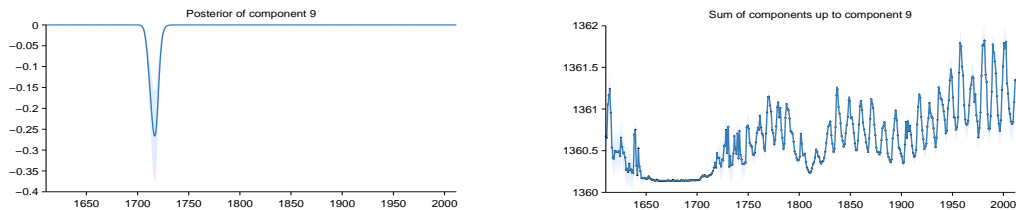


Figure 10: Posterior of component 9 (left) and the posterior of the cumulative sum of components with data (right)

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.

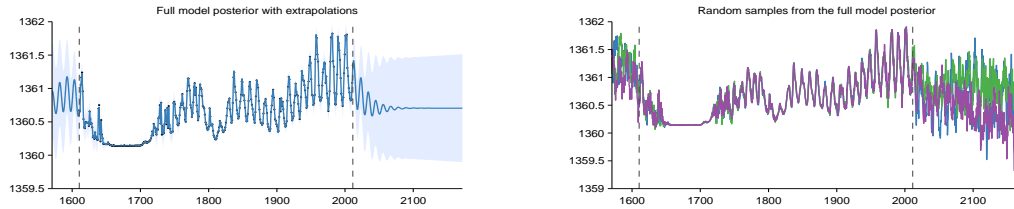


Figure 11: Full model posterior. Mean and pointwise variance (left) and three random samples (right)

3.1 Component 1 : A constant

Some discussion about extrapolation.

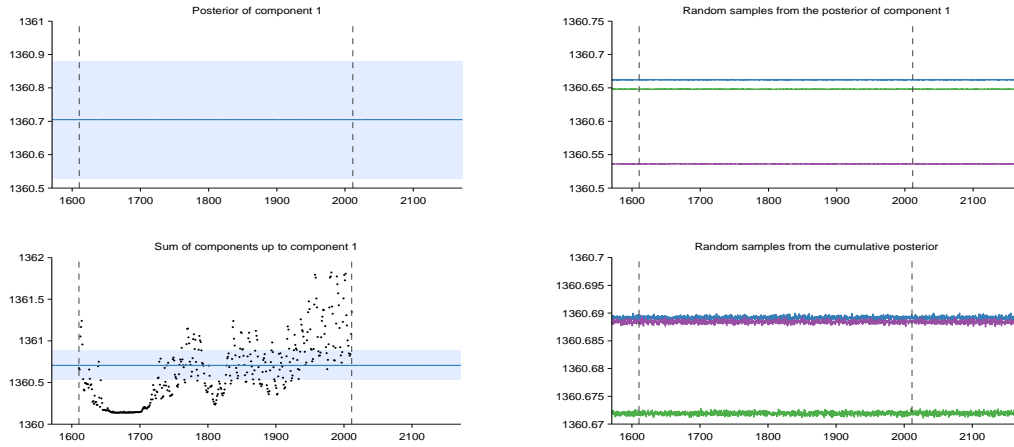


Figure 12: Posterior of component 1. Mean and pointwise variance (left) and three random samples from this distribution (right)

3.2 Component 2 : A constant. The function applies from 1644 until 1713

Some discussion about extrapolation.

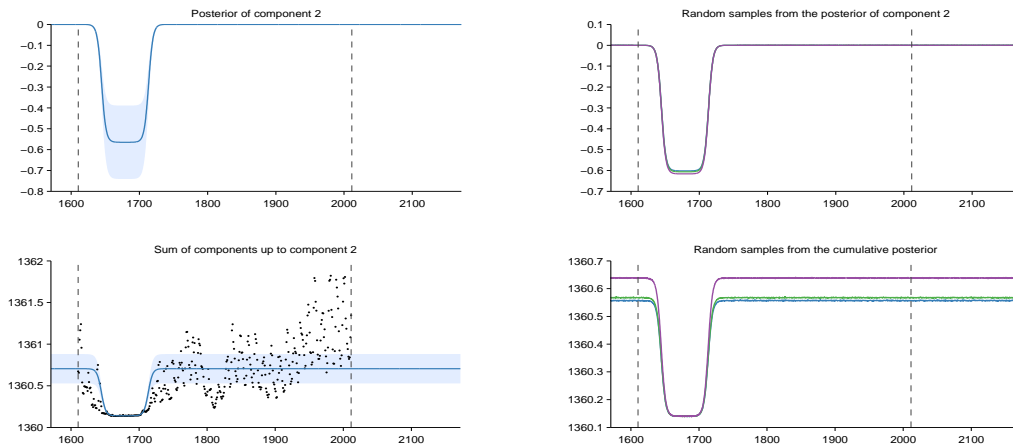


Figure 13: Posterior of component 2. Mean and pointwise variance (left) and three random samples from this distribution (right)

3.3 Component 3 : A smooth function. The function applies until 1644 and from 1719 onwards

Some discussion about extrapolation.

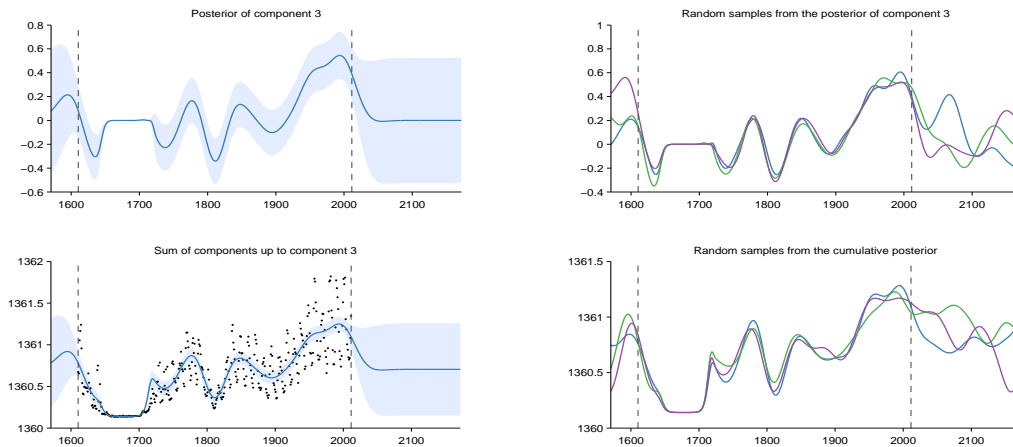


Figure 14: Posterior of component 3. Mean and pointwise variance (left) and three random samples from this distribution (right)

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

Some discussion about extrapolation.

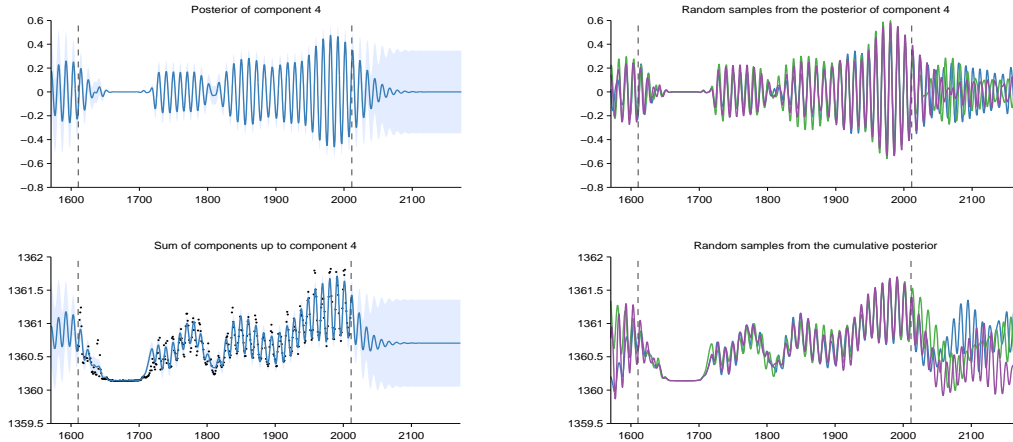


Figure 15: Posterior of component 4. Mean and pointwise variance (left) and three random samples from this distribution (right)

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

Some discussion about extrapolation.

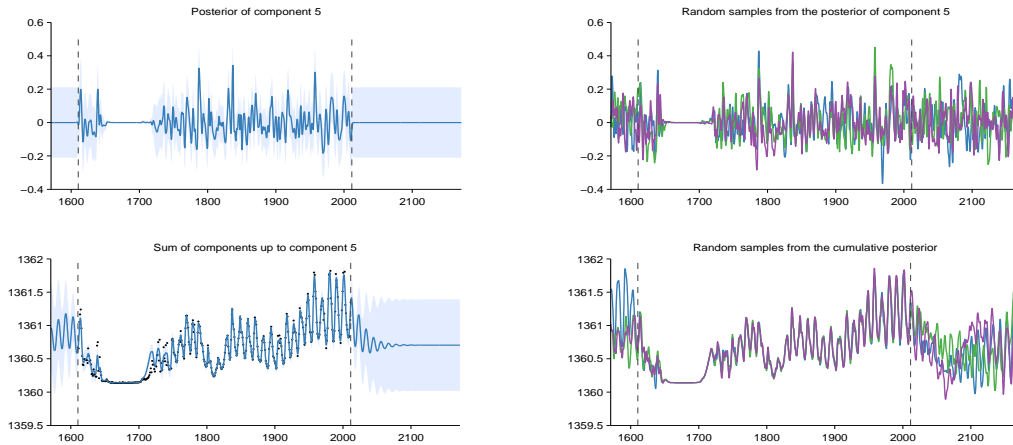


Figure 16: Posterior of component 5. Mean and pointwise variance (left) and three random samples from this distribution (right)

3.6 Component 6 : Uncorrelated noise

Some discussion about extrapolation.

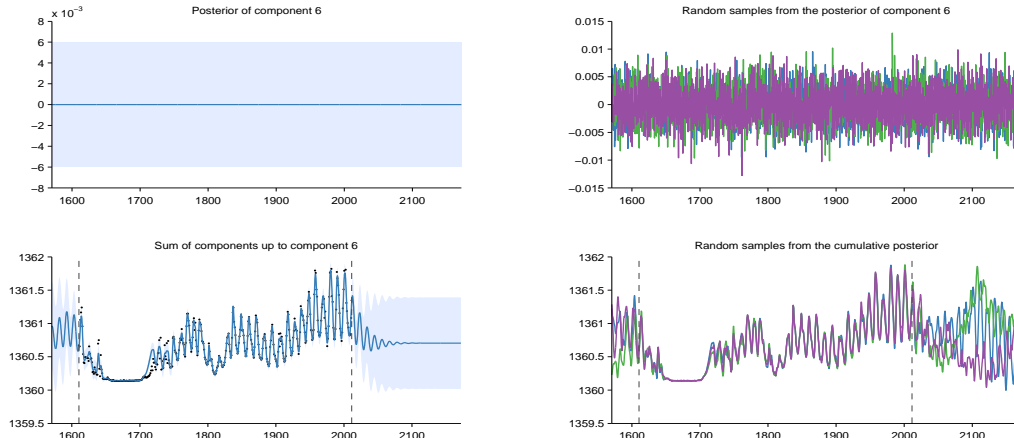


Figure 17: Posterior of component 6. Mean and pointwise variance (left) and three random samples from this distribution (right)

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

Some discussion about extrapolation.

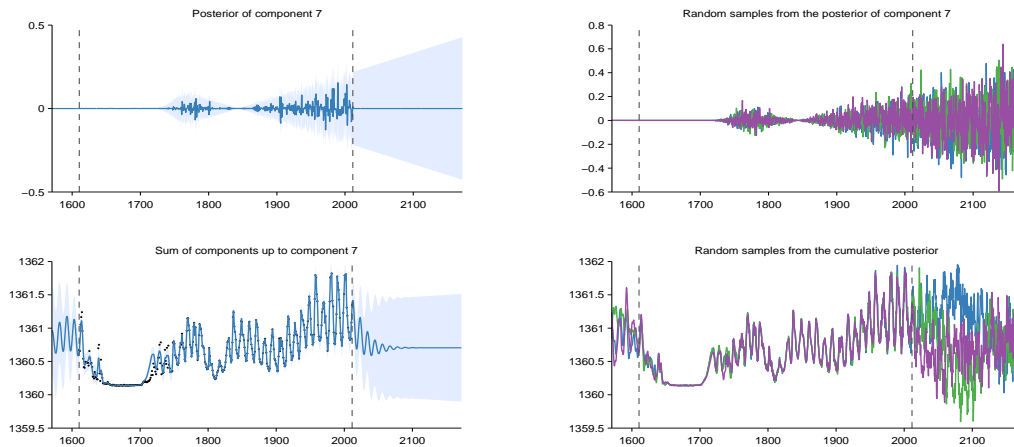


Figure 18: Posterior of component 7. Mean and pointwise variance (left) and three random samples from this distribution (right)

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

Some discussion about extrapolation.

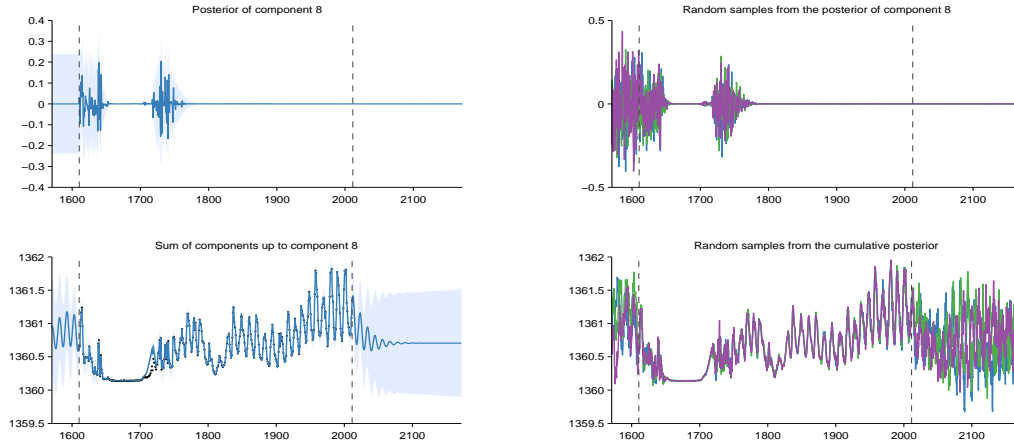


Figure 19: Posterior of component 8. Mean and pointwise variance (left) and three random samples from this distribution (right)

3.9 Component 9 : A constant. The function applies from 1713 until 1719

Some discussion about extrapolation.

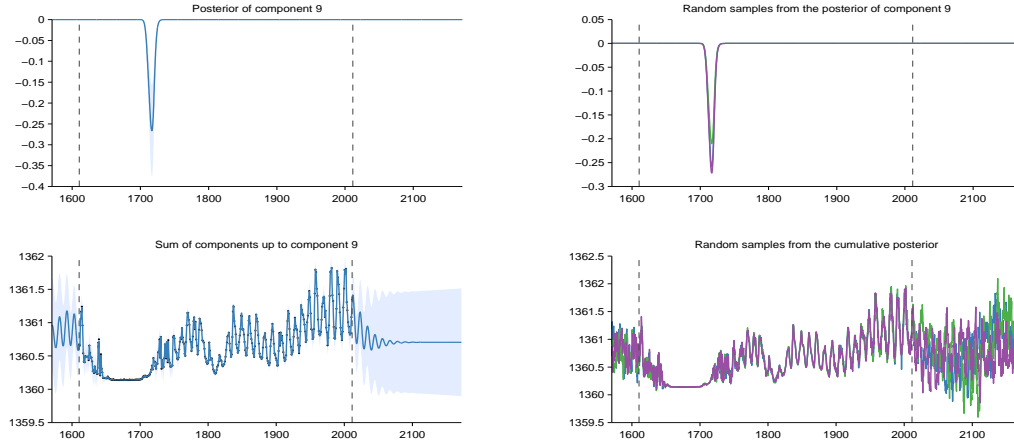


Figure 20: Posterior of component 9. Mean and pointwise variance (left) and three random samples from this distribution (right)