
An automatic report for the dataset : 01-airline

James Robert Lloyd
University of Cambridge
Department of Engineering
jrl44@cam.ac.uk

David Duvenaud
University of Cambridge
Department of Engineering
dkd23@cam.ac.uk

Roger Grosse
M.I.T.
Brain and Cognitive Sciences
rgrosse@mit.edu

Joshua B. Tenenbaum
M.I.T.
Brain and Cognitive Sciences
jbt@mit.edu

Zoubin Ghahramani
University of Cambridge
Department of Engineering
zoubin@eng.cam.ac.uk

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.

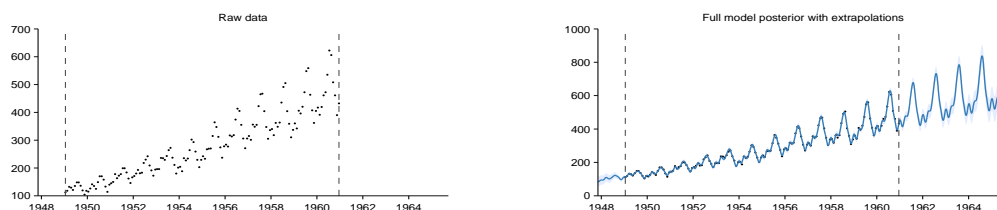


Figure 1: Raw data (left) and model posterior with extrapolation (right)

The structure search algorithm has identified four additive components in the data. The first 2 additive components explain 99.3% of the variation in the data as shown by the coefficient of determination (R^2) values in table 1. After the first 3 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 very smooth monotonically increasing function.
- An approximately periodic function with a period of 1.0 years and with approximately linearly increasing amplitude.
- A smooth function.
- Uncorrelated noise with linearly increasing standard deviation.

#	R^2 (%)	ΔR^2 (%)	Residual R^2 (%)	Cross validated MAE	Reduction in MAE (%)
-	-	-	-	280.30	-
1	86.2	86.2	86.2	32.42	88.4
2	99.3	13.1	95.1	9.57	70.5
3	99.8	0.5	68.1	7.54	21.2
4	100.0	0.2	100.0	7.54	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 very smooth monotonically increasing function

This function is very smooth and monotonically increasing.

This component explains 86.2% of the total variance. The addition of this component reduces the cross validated MAE by 88.4% from 280.3 to 32.4.

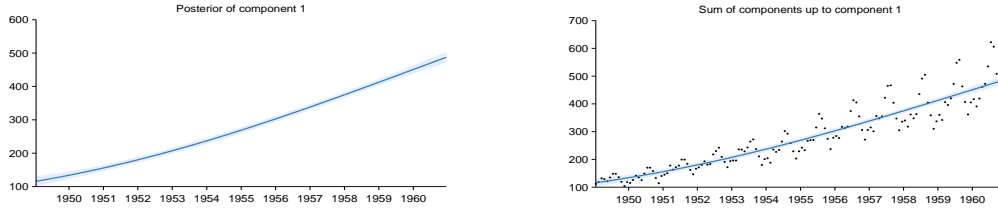


Figure 2: Posterior of component 1 (left) and the posterior of the cumulative sum of components with data (right)

2.2 Component 2 : An approximately periodic function with a period of 1.0 years and with approximately linearly increasing amplitude

This component is approximately periodic with a period of 1.0 years and varying amplitude. Across periods the shape of this function varies very smoothly. The amplitude of the function increases approximately linearly. The shape of this function within each period has a typical lengthscale of 6.1 weeks.

This component explains 95.1% of the residual variance; this increases the total variance explained from 86.2% to 99.3%. The addition of this component reduces the cross validated MAE by 70.47% from 32.42 to 9.57.

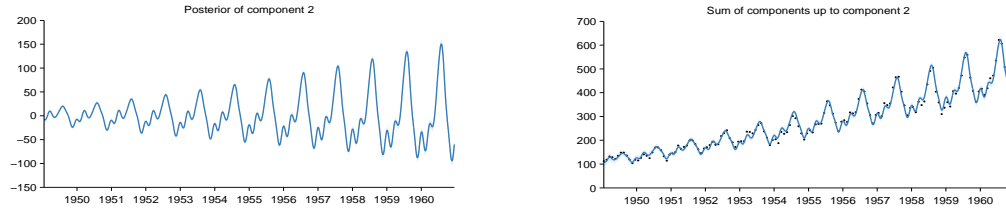


Figure 3: Posterior of component 2 (left) and the posterior of the cumulative sum of components with data (right)

2.3 Component 3 : A smooth function

This component is a smooth function with a typical lengthscale of 6.7 months.

This component explains 68.1% of the residual variance; this increases the total variance explained from 99.3% to 99.8%. The addition of this component reduces the cross validated MAE by 21.22% from 9.57 to 7.54.

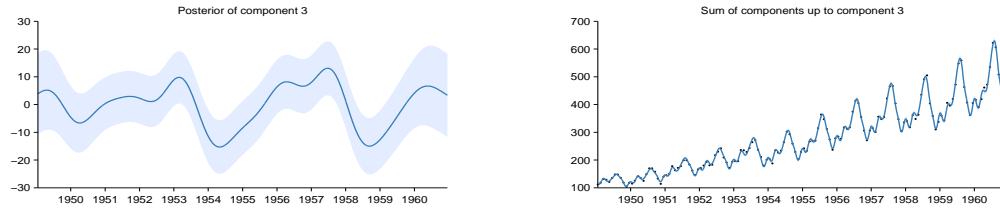


Figure 4: Posterior of component 3 (left) and the posterior of the cumulative sum of components with data (right)

2.4 Component 4 : Uncorrelated noise with linearly increasing standard deviation

This component models uncorrelated noise. The standard deviation of the noise increases linearly.

This component explains 100.0% of the residual variance; this increases the total variance explained from 99.8% to 100.0%. The addition of this component reduces the cross validated MAE by 0.00% from 7.54 to 7.54. 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.

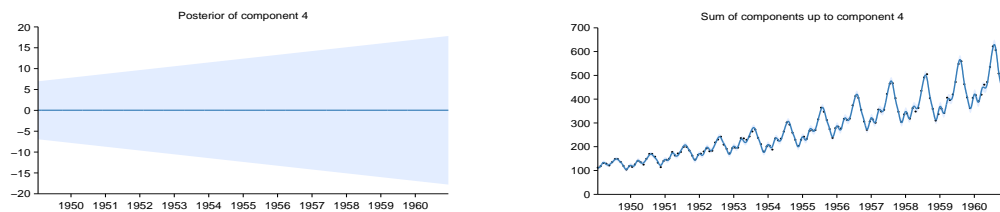


Figure 5: Posterior of component 4 (left) and the posterior of the cumulative sum of components with data (right)

3 Extrapolation

Summaries of the posterior distribution of the full model are shown in figure 6. 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.

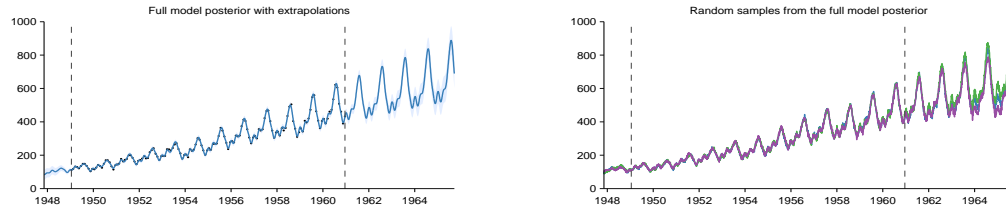


Figure 6: Full model posterior. Mean and pointwise variance (left) and three random samples (right)

3.1 Component 1 : A very smooth monotonically increasing function

Some discussion about extrapolation.

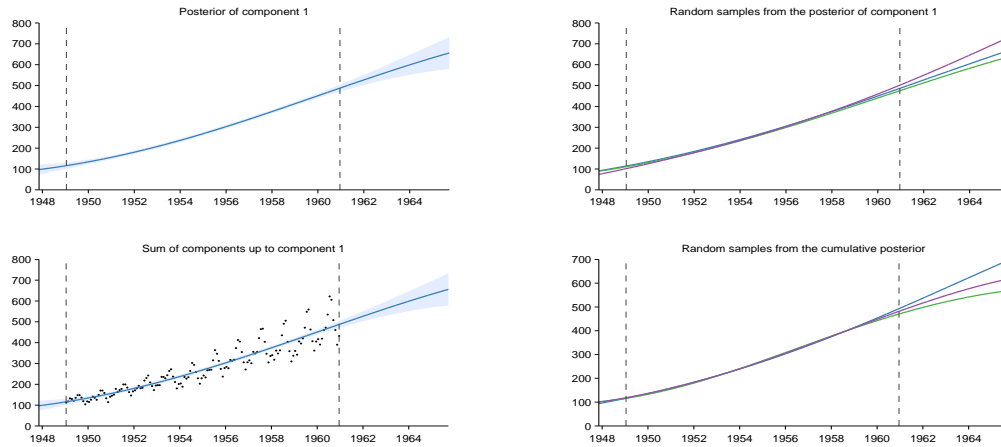


Figure 7: Posterior of component 1. Mean and pointwise variance (left) and three random samples from this distribution (right)

3.2 Component 2 : An approximately periodic function with a period of 1.0 years and with approximately linearly increasing amplitude

Some discussion about extrapolation.

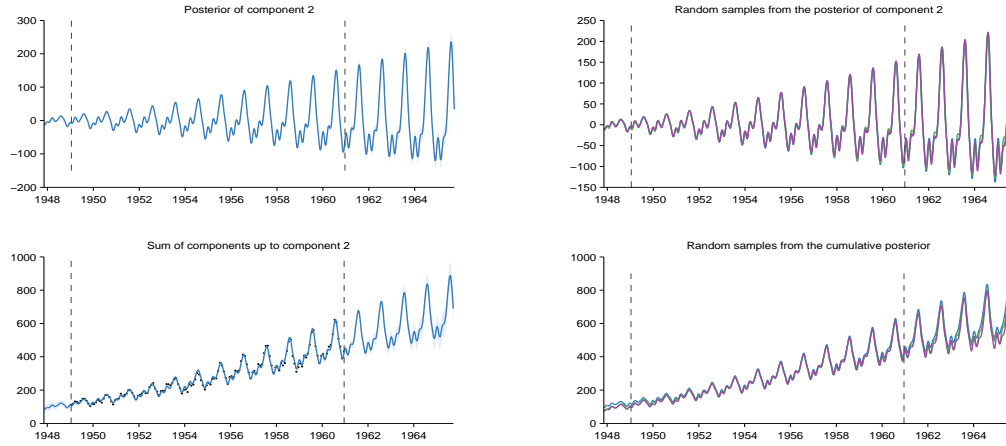


Figure 8: Posterior of component 2. Mean and pointwise variance (left) and three random samples from this distribution (right)

3.3 Component 3 : A smooth function

Some discussion about extrapolation.

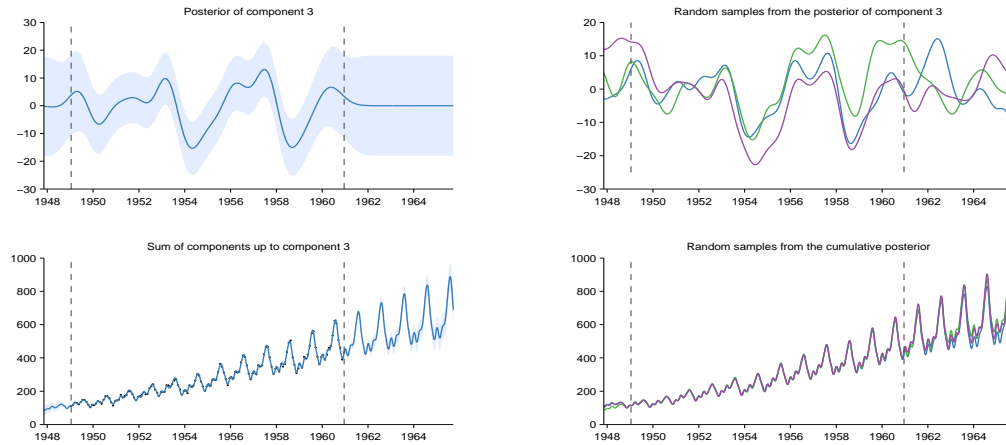


Figure 9: Posterior of component 3. Mean and pointwise variance (left) and three random samples from this distribution (right)

3.4 Component 4 : Uncorrelated noise with linearly increasing standard deviation

Some discussion about extrapolation.

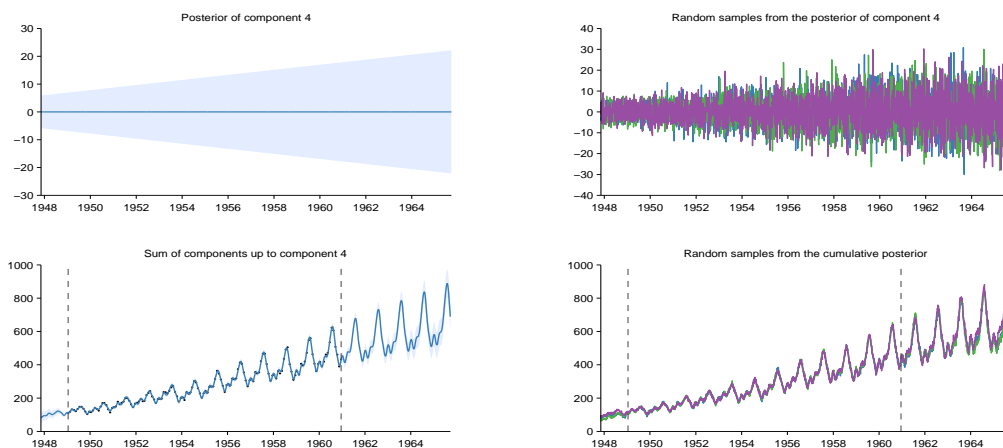


Figure 10: Posterior of component 4. Mean and pointwise variance (left) and three random samples from this distribution (right)