

---

# Automatic Summarization of Composite Nonparametric Time-Series Models

---

**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

To complement automatic model-building and search methods, we demonstrate an automatic model-summarization procedure. Given a composite model, our method automatically constructs a report which visualizes and explains in words the meaning and relevance of each component of the model. These reports enable human model-checking, understanding of complex modeling assumptions, and understanding of complex structure in data.

## 1 Introduction

Parametric models such as linear regression usually have simple interpretations, and well-established methods for model-checking. In contrast, complicated non-parametric models may be reasonably viewed with suspicion, since their assumptions may be difficult to check, and their predictive implications difficult to explain.

Advances in nonparametric and compositional modeling have allowed fully automatic model search and construction of complex, structured, nonparametric regression models [1]. Such automatically-constructed models can be used for extrapolation without further human intervention, but if model-checking is desired, or human understanding is the goal, then simple extrapolations may be insufficient.

This paper describes a procedure for automatically generating a human-readable report, examining and explaining the properties of a given nonparametric model on a given dataset. The supplementary material to this paper is a report completely automatically generated by our method. This report summarizes and describes a dataset after it has been modeled by a Gaussian process with a complicated kernel. The various components of the model correspond to different features of the dataset, and the report discusses in detail the properties and significance of each component.

Section 2 gives an overview of the structure search method which produces the models that our procedure summarizes. Section 3 describes our method for generating these reports, and section 4 examines in detail the report included in the supplementary material.

## 2 Gaussian Process Structure Search

Gaussian process models[2] use a kernel to define the covariance between any two function values:  $\text{Cov}(y, y') = k(x, x')$ . The covariance function, or *kernel*, specifies which structures are likely under the GP prior, which in turn determines the generalization properties of the model.

Different kernels can express wide varieties of statistical structures, such as local smoothness, periodicity, additivity, and changepoints. Compositions of kernels can express even richer structure, such as functions which are periodic only locally, or functions whose amplitude grows over time. The Gaussian process structure search (GPSS) procedure simply searches over such compositions, in order to find complex structure in time-series or multidimensional functions [1].

When given a dataset, GPSS produces a complex kernel which can usually be expressed as a sum of simpler kernels. One property of Gaussian process models is that a sum of kernels is equivalent to a sum of functions. This means that the models produced by GPSS can be interpreted as decomposing the function being modeled into a sum of parts.

## 3 Automatic Report Generation

Our procedure takes as its starting point a dataset and a composite kernel, which together define a joint probability distribution over a sum of functions. The function of our procedure is to illustrate the properties of this complex distribution to the user.

### 3.1 Design goals

- **Intelligibility** The procedure was designed to produce reports intelligible to non-experts.
- **Illustrate Model Assumptions** One of the main design criteria when designing our procedure was to make clear what assumptions the model is making, and what these assumptions imply in terms of extrapolation. Even when simple Gaussian process models are used, it is often unclear what structure was captured by the model and what was not.
- **Illustrate Posterior Uncertainty** One of the selling points of Bayesian modeling over point estimates is that they produce a rich posterior distribution over possible explanations of the data. However, this posterior distribution is often quite a complex object, and is usually difficult to summarize.
- **Enable Model-checking** One of the most important reasons to examine a model is to discover flawed assumptions, or structure not captured by the model. Simply examining residuals, cross-validation or marginal likelihood scores is often not sufficient to notice when the model is failing to capture.

### 3.2 Report structure

Our procedure produces a report with three sections:

1. **Executive summary** The first section of each report describes which sorts of structure were discovered in the model, and the relevant importance of the different components in explaining the data.
2. **Discussion of additive components** The second section contains a detailed discussion of each component in turn. Every component is plotted, and various properties of the statistical structure represent are described. Because components with small variance are often not meaningful when plotted on their own, we also include plots of each component combined with all components of higher marginal variance.
3. **Extrapolation** The third section shows extrapolations into the future, as well as posterior samples from each component of the model. These samples help to characterize the uncertainty captured by the model, and the extent to which different components contribute to predicting the future behavior of the time series.

## 4 Example: Summarizing 400 Years of Solar Activity

To give an example of the capabilities of our procedure, here we show excerpts from the report automatically generated when run on annual solar irradiation data from 1610 to 2011. This dataset has two obvious features: First, a roughly 11-year cycle of solar activity, and second, a period from 1645 to 1715 with much smaller variance than the rest of the dataset. This flat region corresponds to the Maunder minimum, a period in which sunspots were extremely rare [3].

### 4.1 Executive Summary

#### 1 Executive summary

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

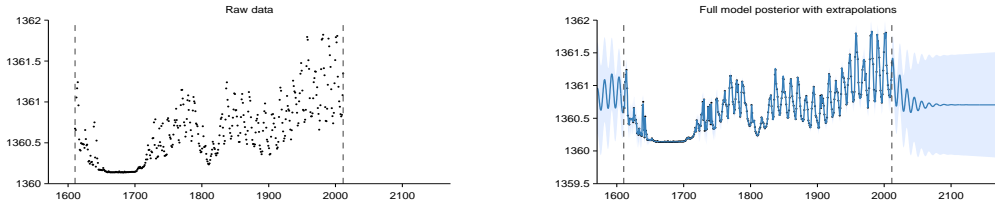


Figure 1: Raw data (left) and model posterior with extrapolation (right)

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.

Figure 1: An example of an automatically-generated summary of a dataset. The dataset is broken into a set of diverse types of structure, and each structure is explained in simple terms.

Figure 1 shows the reported summary of the solar dataset. The model discovers 9 separate components which explain the data, and reports that the first 4 components explain 90% of the variance in the data. This might seem incongruous with the observation that there are two main features of the data, but if we examine the first four components, we see that the first component is explaining the

#	$R^2$ (%)	$\Delta R^2$ (%)	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.

Figure 2: A table summarizing the relative contribution of the 9 different components of the model in terms of predictive performance.

mean of the dataset, the second explains the Maunder minimum, the third describes the long-term trend, and the fourth describes the 11-year period.

#### 4.1.1 Signal versus Noise

One design challenge we encountered was separating the recovered structure into signal and noise. Originally, the model always included had a term corresponding to i.i.d. additive Gaussian noise. However, philosophically and in practice, the distinction between signal and noise is unclear for two reasons. First, a component which varies arbitrarily quickly in time becomes indistinguishable from noise. Second, the variance of the noise may change over time (called heteroskedasticity), and this sort of pattern may be considered part of the signal.

Because of the blurry distinction between signal and noise, we include a table which summarizes the relative contribution of each component in terms of predictive power. To do this, we order the components in terms of how much each one improves predictive performance in a 10-fold cross-validation procedure. The intuition for this metric is that noise-like components do not contribute much to the extrapolation performance of the model, but that signal-like components do. Figure 2 show an example of this table on the solar dataset.

Because the user may be interested in local or noisy components, we report all components to the user. An interactive version of our procedure could allow users to specify which components are of interest, and group the remaining components into a single noise component.

## 4.2 Decomposition plots

Because the individual components produced by our procedure are produced by an open-ended grammar, it is impossible to write special code to handle every type of component that GPSS might produce. In addition, the posterior distribution on functions implied by the model depends on the data. Hence the part of the procedure which summarizes the individual components must be flexible, and also must present enough information that users can see the meaning of each component themselves. We summarize each component both in writing and in plots.

**Automatic English Summarization** For the written portion, a first paragraph reports the type of component found, and the meaning of all kernel parameters in that component. This is possible even in an open-ended grammar of kernels because of the compositional nature of kernel structures. Each type of kernel in an expression modifies the form of the posterior in a different way. Table 1 gives some examples.

Kernel	Description
SE	A smoothly varying function
SE $\times$ PER	A smoothly varying periodic function
SE $\times$ PER $\times$ LIN	A smoothly varying periodic function with growing amplitude

Table 1: A demonstration of how it is possible to write an open-ended summarizer. Different kernel structures modify the overall statistical structure of each component in independent ways.

A second paragraph explains the improvement in predictive performance gained by including this component in the model. This is the same informatino as included in the executive summary.

**Automatic Plotting** In the component plots, each component’s posterior is described in two ways. First, the posterior mean and variance of each component is plotted on its own. Second, the posterior mean and variance of all components shown so far is plotted against the data. By contrasting each of these plot with plots of earlier components, we can see qualitatively how the current component is contributing to an overall explanation of the data.

## 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.

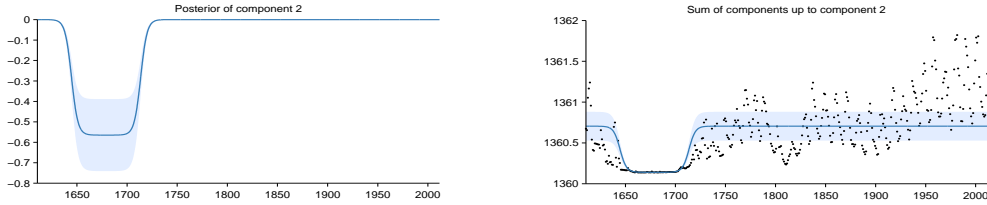


Figure 3: Posterior of component 2 (left) and the posterior of the cumulative sum of components with data (right)

Figure 3: Discovering the Maunder minimum. The kernel found by GPSS contained a pair of changepoints bracketing the period of low solar activity.

Figure 3 is a good example of a meaningful component discovered by GPSS, whose meaning would be unclear without an individual plot. In the GPSS grammar, we allow any kernel expression to be multiplied by a changepoint kernel, which smoothly modulates the whole expression over time. Two changepoints combined can encode a local change in covariance.

In the history of solar activity, the Maunder minimum is a good example of a local change in covariance. Specifically, from about 1645 to 1715, solar activity decreased, and very few sunspots were observed [3]. In this example, GPSS captured this structure with a pair of changepoint kernels, and the plotted summary shows exactly which sort of structure was recovered by this component.

**Isolating the smoothly-varying component** Examining the dataset by eye, overall solar activity seems to change slowly over decades. However, this intuition seems difficult to formalize. Linear or quadratic regression is clearly inappropriate, and methods based on local smoothing would need to control for the periodic component. Luckily, the GPSS procedure does exactly this, allowing us to isolate the slowly-varying component of the data, without having to forecast either the Maunder minimum or the periodic variation. Figure 4 shows the automatically-generated summary of this component.

### 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.

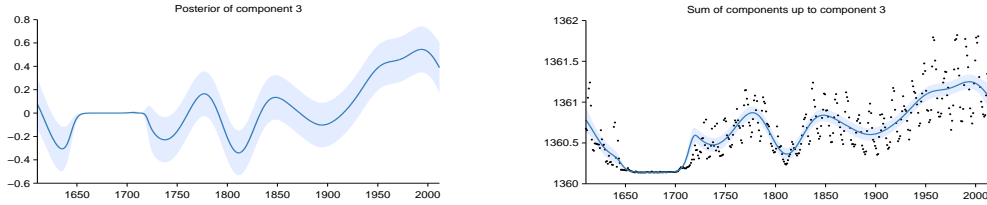


Figure 4: Posterior of component 3 (left) and the posterior of the cumulative sum of components with data (right)

Figure 4: Characterizing the medium-term smoothness of solar activity levels. By allowing other components to explain the periodicity, noise, and the Maunder minimum, we can isolate the part of the signal that's best explained by a slowly-varying trend.

### 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.

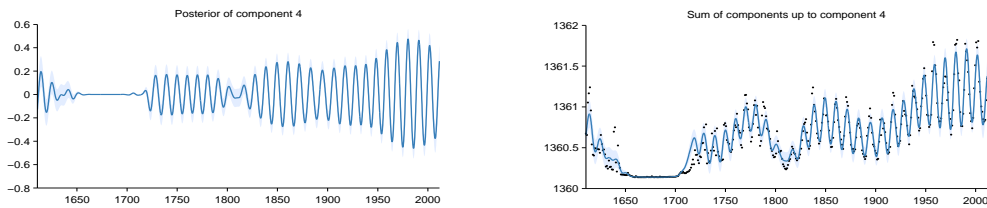


Figure 5: Posterior of component 4 (left) and the posterior of the cumulative sum of components with data (right)

Figure 5: Isolating the periodic component of the dataset. By isolating this aspect of the statistical structure, we can easily observe additional features, such as the shape of the peaks and troughs, or the fact that the amplitude changes over time.

Figure 5 shows the isolated periodic component of the signal. Here we see one of the main benefits of isolating individual components: we can now see, by eye, extra structure that was not captured by the model. Specifically, we can see that the amplitude of the periodic component varies over time, and by comparing with figure 4, we can see that it varies roughly in proportion to the overall

magnitude of the signal. This pattern suggests that some sort of log-transform might be appropriate for this dataset, or that the model should be extended in some way to capture this structure.

### 4.3 Extrapolation plots

The predictive mean and variance of the signals shown in the summary plots are useful, but do not capture the joint correlation structure in the posterior. Showing posterior samples are a simple and universal way to demonstrate the statistical structure captured by a model.

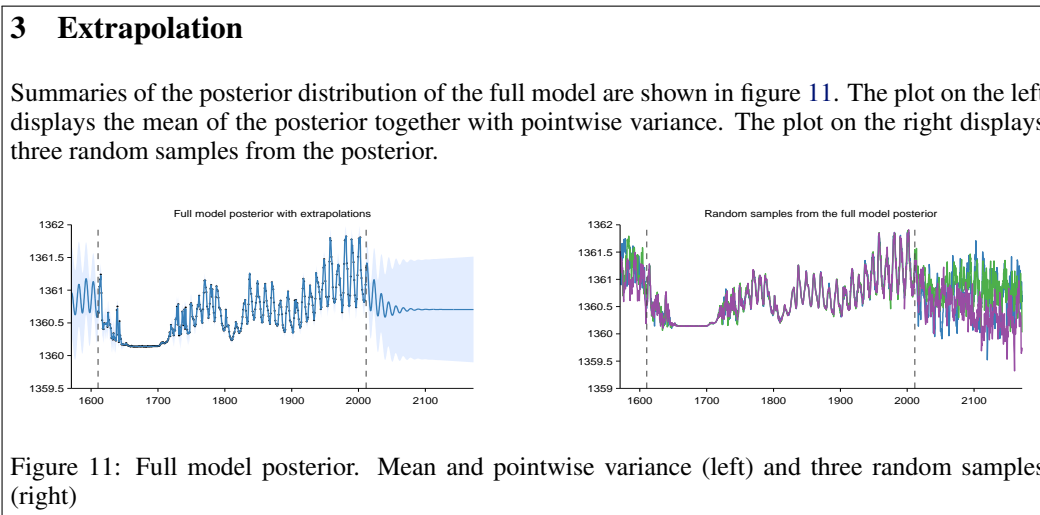


Figure 6: Sampling from the posterior. These samples help show not just the predictive mean and variance, but also the predictive covariance structure. Note, for example, that the predictive mean (left) doesn’t exhibit periodicity, but the samples (right) do.

For example, the left-hand plot in figure 6 shows the predictive mean and variance given the entire model. It is not clear from this plot whether or not the periodicity of the dataset is expected to continue into the future. However, from the samples on the right-hand side, we can see that this is the case.

**Extrapolating individual components** We can also examine the model’s expectations about the future behavior of individual components through sampling. For example, in figure 7, we are shown samples of the predictive distribution of the smooth component of variation. This plot indicates that the model considers a wide range of future average intensities to be plausible, but that it always expects those average intensities to vary smoothly.

## 5 Related Work

There exists a vast literature on both model visualization and model checking.

## 6 Discussion

“The availability of ‘user-friendly’ statistical software has caused authors to become increasingly careless about the logic of interpreting their results, and to rely uncritically on computer output, often using the ‘default option’ when something a little different (usually, but not always, a little more complicated) is correct, or at least more appropriate.”

G. Dyke, 1997

In this paper, we demonstrated a method for automatically summarizing a compositional Gaussian process model. These summaries can enable human experts and non-experts to understand the implications of a model, check its plausibility, and notice structure not yet discovered by the model.

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

Some discussion about extrapolation.

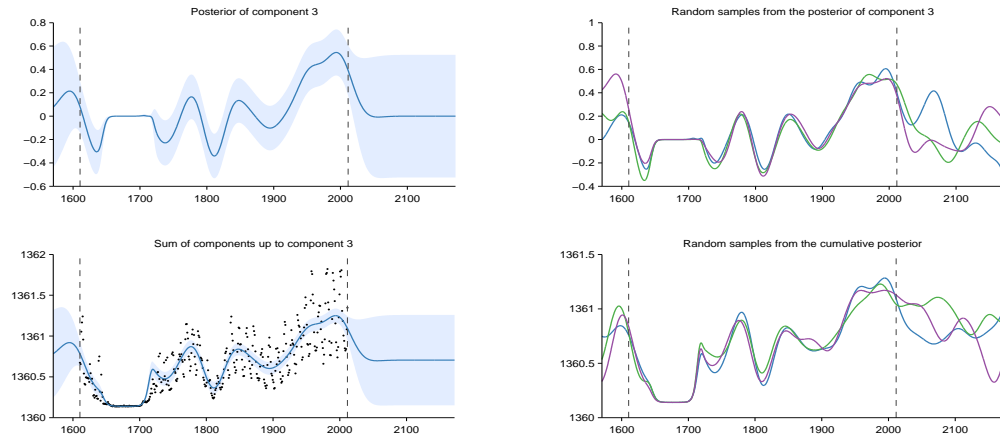


Figure 14: Posterior of component 3. Mean and pointwise variance (left) and three random samples from this distribution (right)

Figure 7: Extrapolating a single component of the model. Because our model class allows us to isolate individual components, we can show the extent to which the model expects different trends to continue. We also observe that the posterior samples are quite different from the posterior mean, away from the data.

## References

- [1] David Duvenaud, James Robert Lloyd, Roger Grosse, Joshua B. Tenenbaum, and Zoubin Ghahramani. Structure discovery in nonparametric regression through compositional kernel search. In *Proceedings of the 30th International Conference on Machine Learning*, June 2013.
- [2] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, MA, USA, 2006.
- [3] J. Lean, J. Beer, and R. Bradley. Reconstruction of solar irradiance since 1610: Implications for climate change. *Geophysical Research Letters*, 22(23):3195–3198, 1995.
- [4] George Dyke. How to avoid bad statistics. *Field crops research*, 51(3):165–187, 1997.