Automatic construction and description of nonparametric models

James Robert Lloyd[1], David Duvenaud[1], Roger Grosse[2], Joshua B. Tenenbaum[2], Zoubin Ghahramani[1]

1: Department of Engineering, University of Cambridge, UK 2: Massachusetts Institute of Technology, USA

This analysis was automatically generated

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

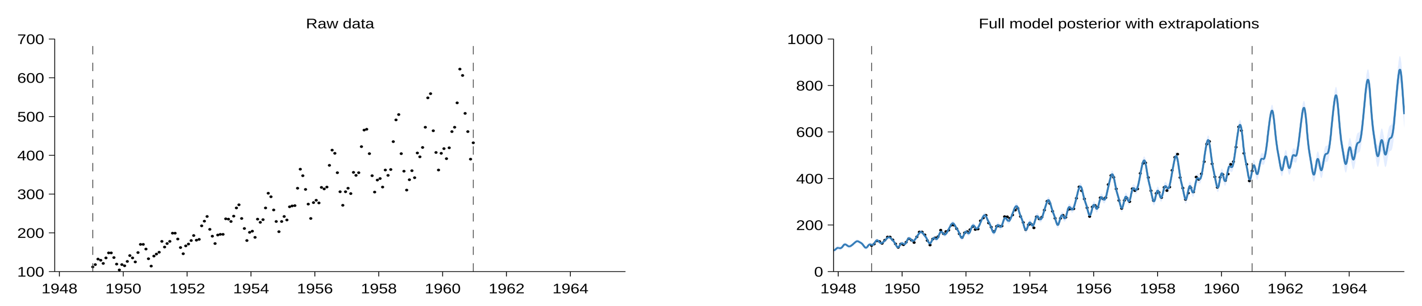


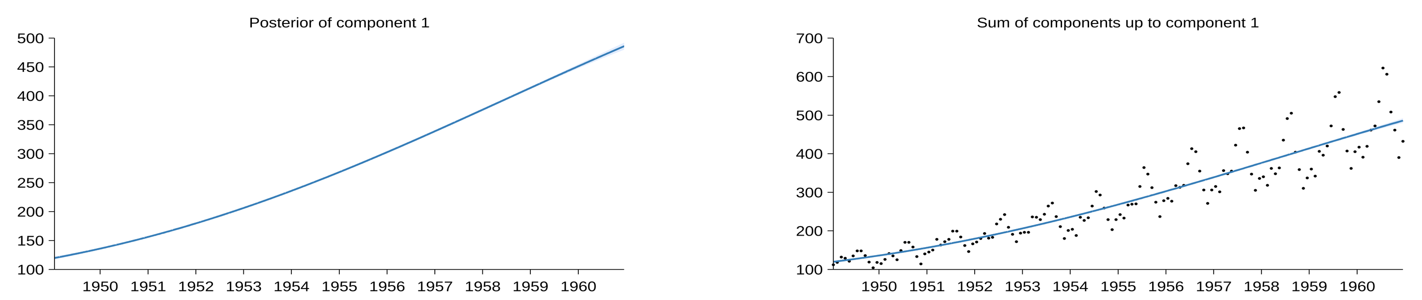
Figure 1: Raw data (left) and model posterior with extrapolation (right)

The structure search algorithm has identified four additive components in the data:

- A very smooth monotonically increasing function.
- An approximately periodic function with a period of 1.0 years and with approximately linearly increasing amplitude.
- An exactly periodic function with a period of 4.3 years but with linearly increasing amplitude.
- Uncorrelated noise with linearly increasing standard deviation.

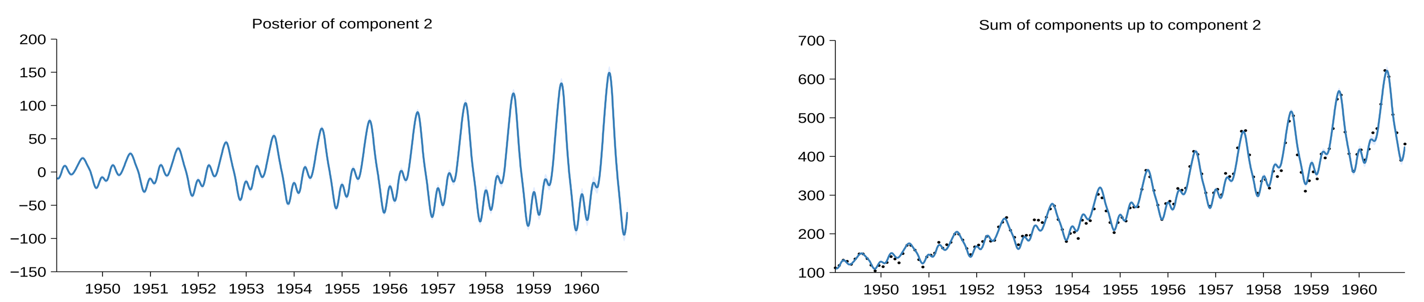
2.1 Component 1

This component is a very smooth and monotonically increasing function.



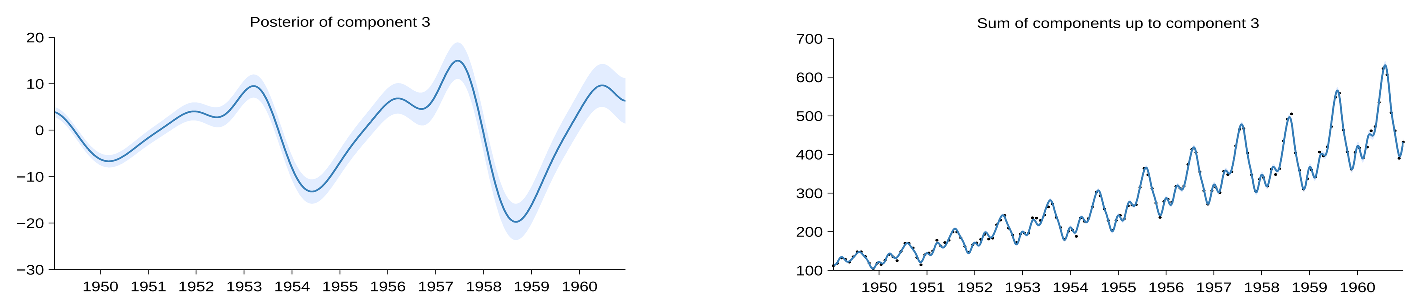
2.2 Component 2

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 6.0 weeks.



2.3 Component 3

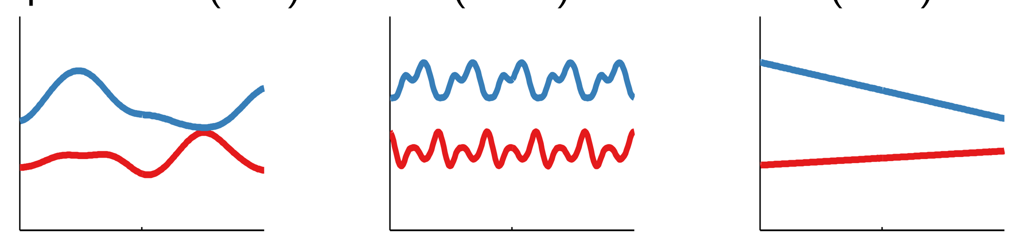
This component is exactly periodic with a period of 4.3 years but with varying amplitude. The amplitude of the function increases linearly. The shape of this function within each period has a typical lengthscale of 7.4 months.



Gaussian processes model structure through kernels

- For Gaussian process models, the prior — and hence, the pattern of generalisation — is determined by a kernel function. Common examples include:

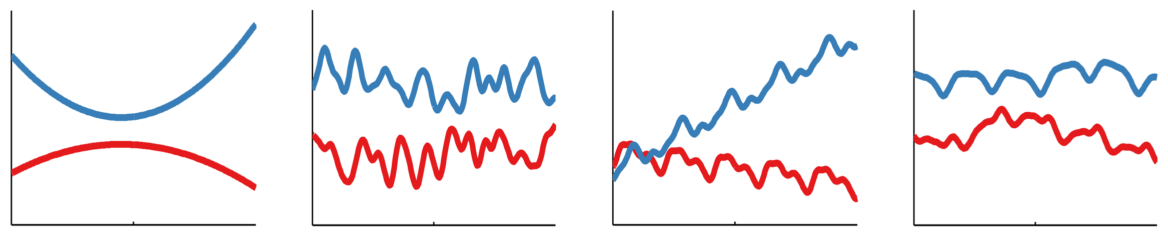
Squared Exponential (SE) | Periodic (PER) | Linear (LIN)



local variation | repeating structure | linear functions

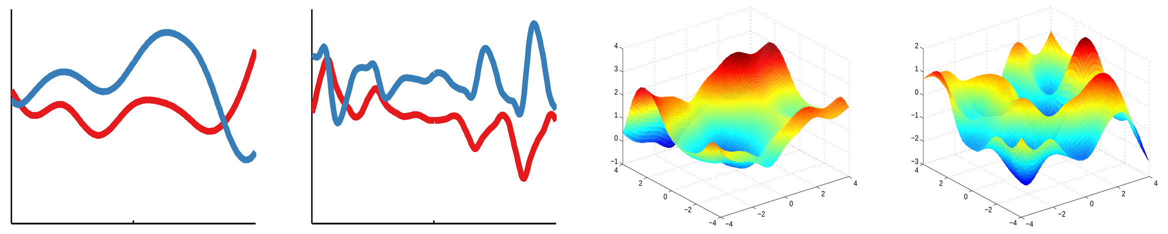
- Composite kernels can express many types of structure

LIN × LIN | SE × PER | LIN + PER | SE + PER



quadratic functions | locally periodic | periodic with trend | periodic with noise

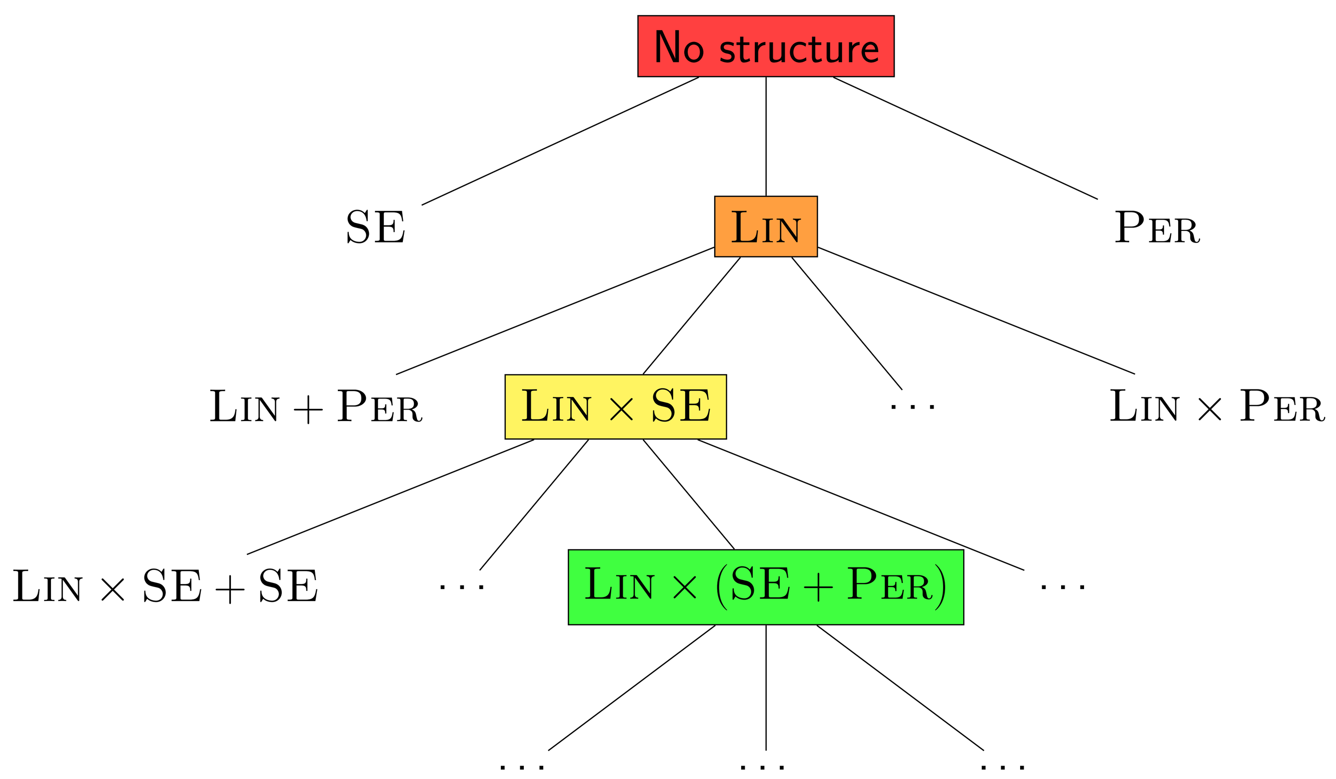
SE × LIN | LIN × PER | $\text{SE}_1 + \text{SE}_2$ | $\text{SE}_1 \times \text{SE}_2$



increasing variation | growing amplitude | $f_1(x_1) + f_2(x_2)$ | $f(x_1, x_2)$

- Building composite kernels has often required much human ingenuity

We build models via a greedy search



Automatically describing model properties

- How to automatically describe complex kernels?

1. Break each kernel into a sum of products.
2. For each product, look up the properties of each kernel in that product.
3. Combine the properties into one sentence.
4. Plot contribution of this compoenent to the model.

Kernels can be distributed into a sum of products

For example:

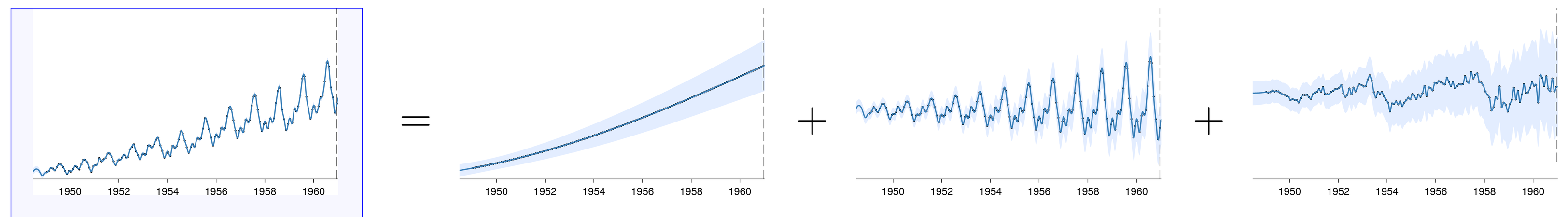
$$\text{SE} \times (\text{LIN} + \text{LIN} \times (\text{PER} + \text{SE}))$$

becomes

$$(\text{SE} \times \text{LIN}) + (\text{SE} \times \text{LIN} \times \text{PER}) + (\text{SE} \times \text{LIN}).$$

Sums of kernels correspond to sums of functions

If $f_1(x) \sim \text{GP}(0, k_1)$ and $f_2(x) \sim \text{GP}(0, k_2)$ then $f_1(x) + f_2(x) \sim \text{GP}(0, k_1 + k_2)$. Therefore, a sum of kernels can be described as a sum of independent functions.



entire signal = SE × LIN smooth trend + SE × LIN × PER growing periodic + SE × LIN short-term deviation

Each kernel in a product corresponds to an adjective

Kernel	How it modifies the prior
SE	functions change smoothly
PER	functions repeat
LIN	amplitude increases linearly

On its own, each kernel simply modifiyes the constant function $f(x) = c$.

Example short descriptions

Product of Kernels	Description
PER	An exactly periodic function
PER × SE	An approximately periodic function
PER × SE × LIN	An approximately periodic function with linearly varying amplitude
LIN	A linear function
LIN × LIN	A quadratic function
PER × LIN × LIN	An exactly periodic function with quadratically varying amplitude

Code available at `github.com/jamesrobertlloyd/gpss-research`