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.

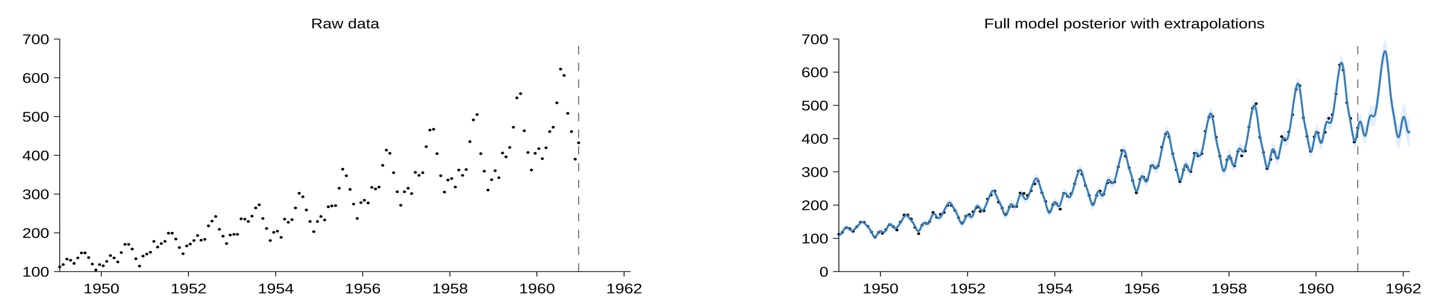


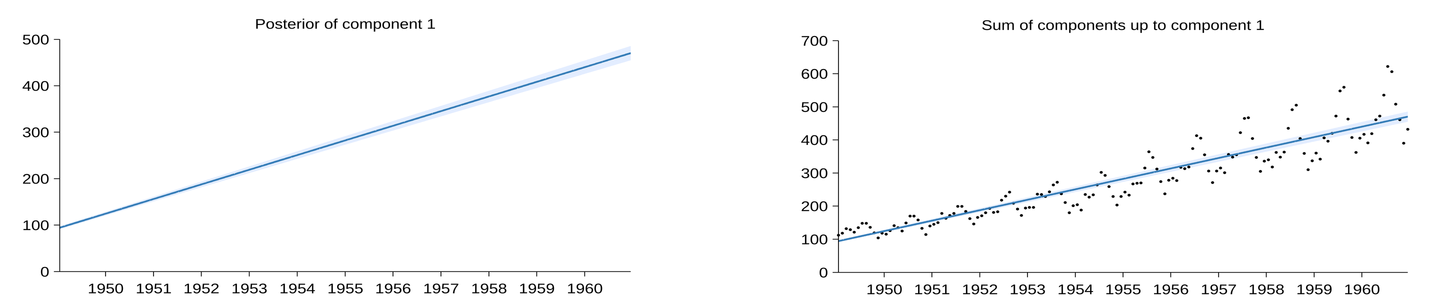
Figure 1: Raw data (left) and model posterior with extrapolation (right)

The Automatic Bayesian Covariance Discovery system has discovered four additive components in the data:

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

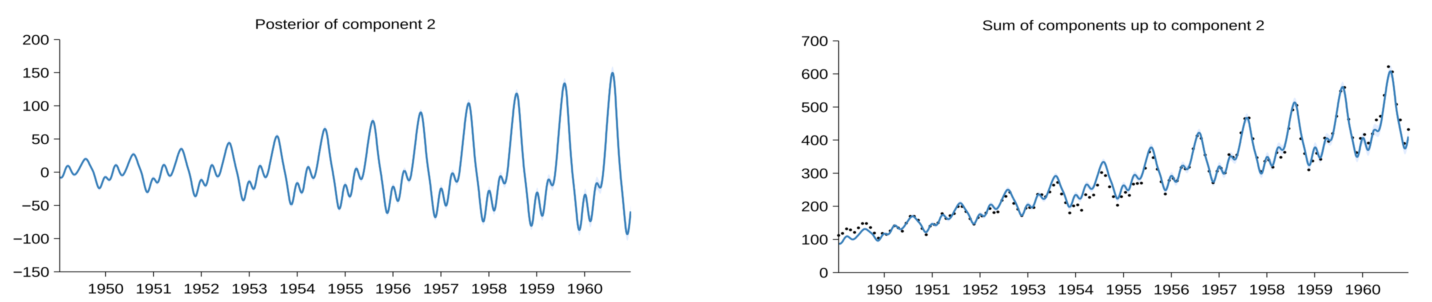
2.1 Component 1 : A linearly increasing function

This component is linearly increasing.



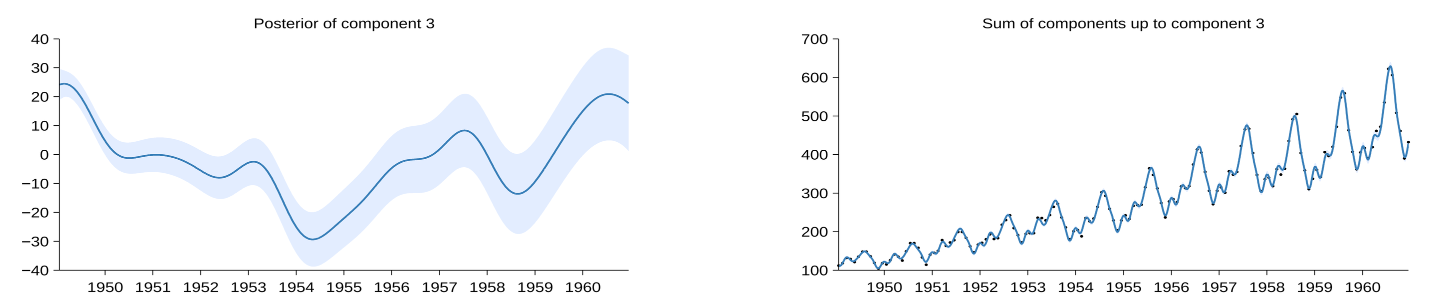
2.2 Component 2 : An approximately periodic function with a period of 1.0 years and with linearly increasing amplitude

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 linearly. The shape of this function within each period has a typical lengthscale of 6.0 weeks.



2.3 Component 3 : A smooth function

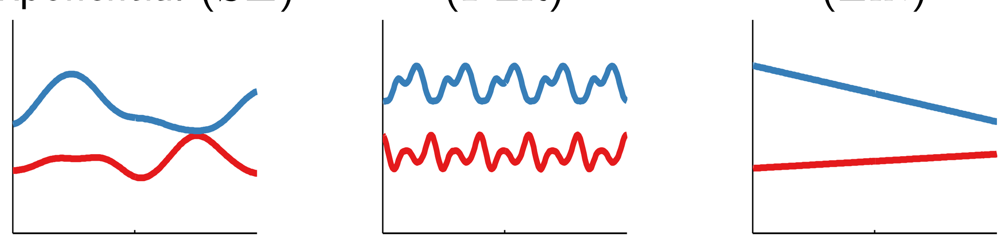
This component is a smooth function with a typical lengthscale of 8.1 months.



Modelling structure through Gaussian process kernels

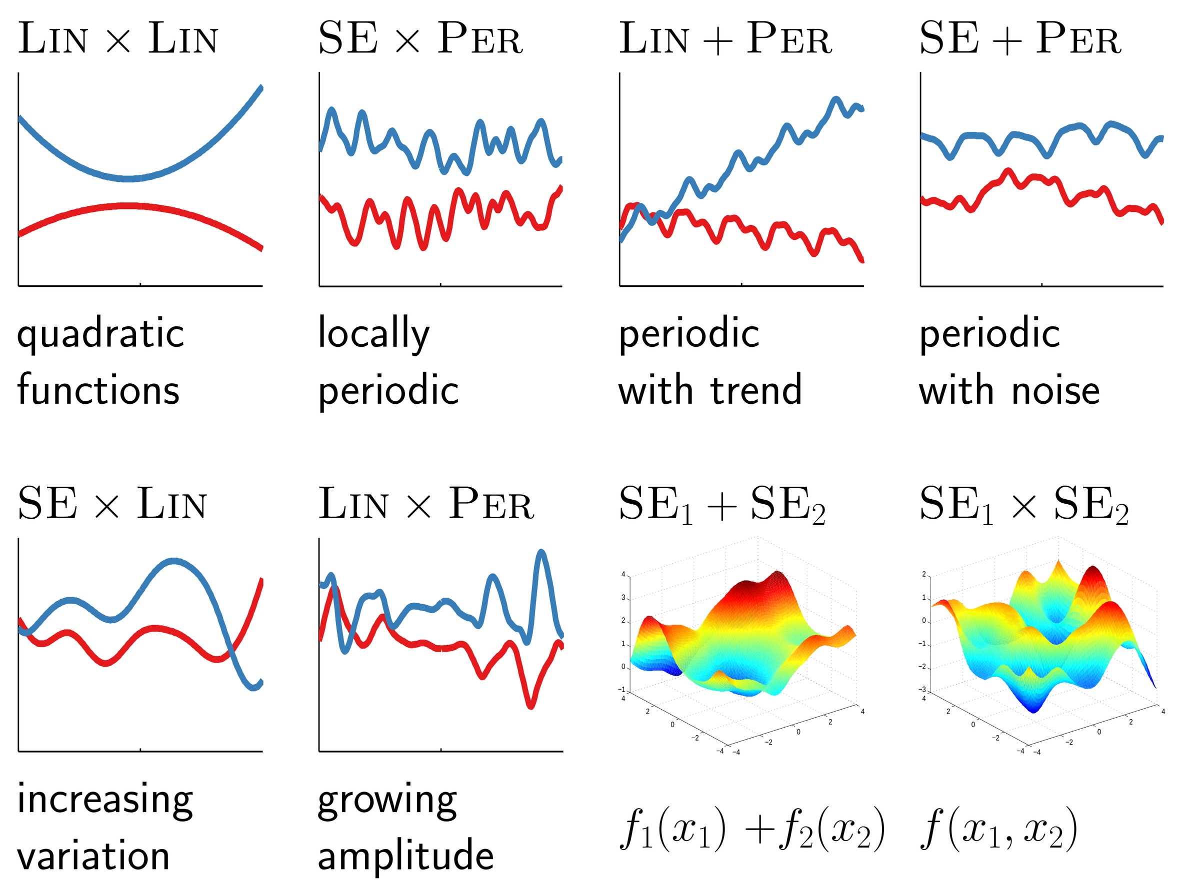
- The kernel specifies which structures are likely under the GP prior - which determines the generalisation properties of the model.

Squared Exponential (SE)	Periodic (Per)	Linear (Lin)

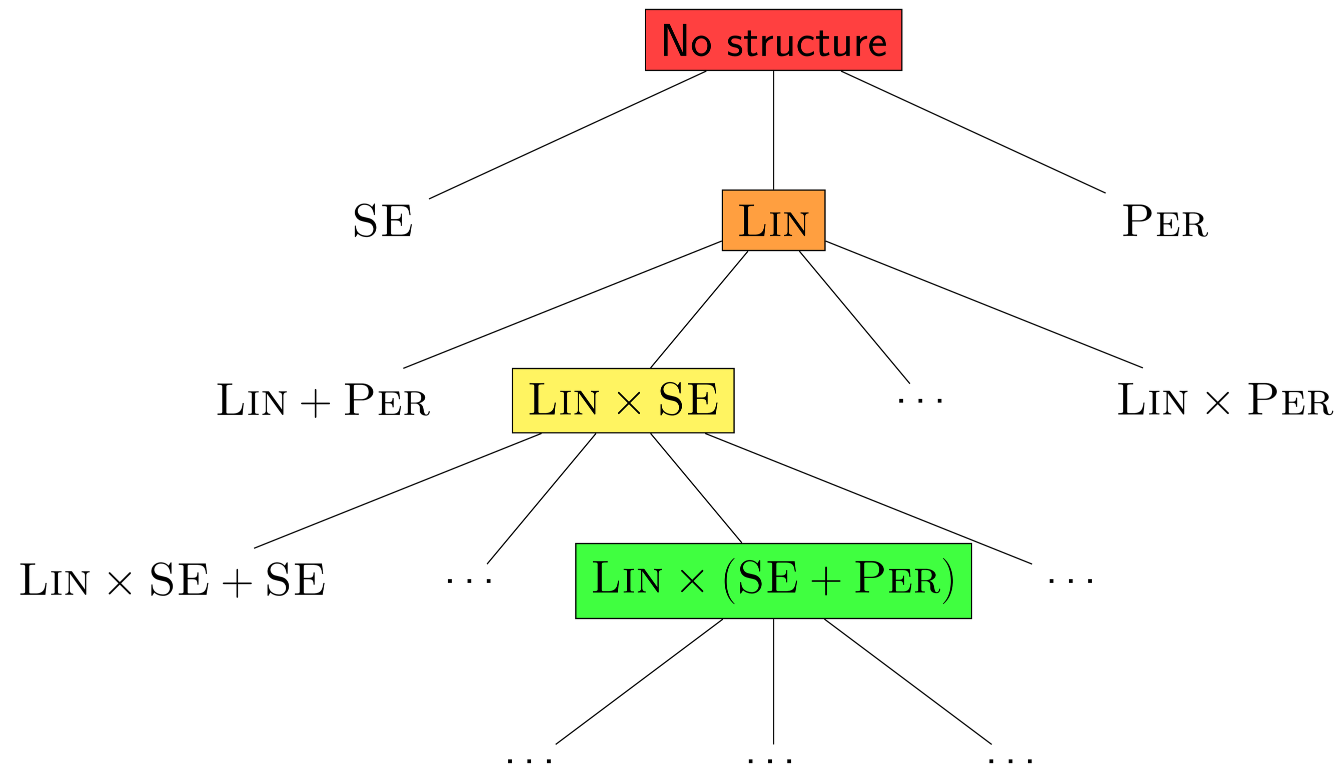


local variation	repeating structure	linear functions

- Composite kernels can express many types of structure



We build models by a greedy search



Automatically describing model properties

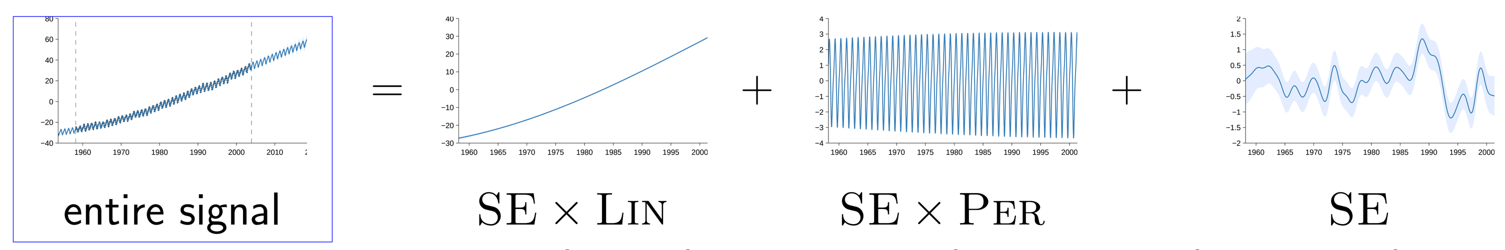
Kernels can be distributed into a sum of products

$$\text{SE} \times (\text{Lin} + \text{Per} + \text{SE})$$

becomes (after simplification)

$$(\text{SE} \times \text{Lin}) + (\text{SE} \times \text{Per}) + (\text{SE}).$$

Sums of kernels correspond to sums of functions



entire signal	=	SE × Lin smooth trend	+	SE × Per periodicity	+	SE short-term deviation

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 functions.

The compositional structure of products of kernels maps onto compositionally constructed sentences

Kernel	Noun phrase	Postmodifier phrase
WN	uncorrelated noise	n/a
C	constant	n/a
SE	smooth function	whose shape changes smoothly
Per	periodic function	modulated by a periodic function
Lin	linear function	with linearly varying amplitude
$\prod_k \text{Lin}^{(k)}$	polynomial	with polynomially varying amplitude
$\prod_k \boldsymbol{\sigma}^{(k)}$	n/a	which applies until / from [changepoint]

Example description

$$\underbrace{\text{Per}}_{\text{periodic function}} \times \underbrace{\text{Lin}}_{\text{with linearly growing amplitude}} \times \underbrace{\boldsymbol{\sigma}}_{\text{which applies until [date]}}$$

Per has been chosen to act as the noun while Lin and $\boldsymbol{\sigma}$ modify the description

Visit the website - try the (simpler) demo

This is part of a larger research project called the Automatic Statistician project. See the website for more details and try the linear modelling demo.

www.automaticstatistician.com