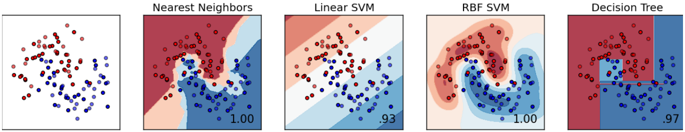
Automatic construction and description of nonparametric models

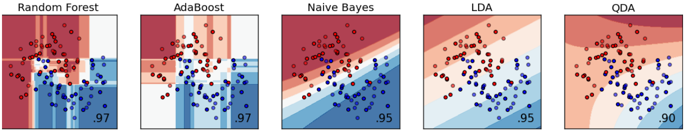
  

James Robert Lloyd, David Duvenaud, Roger Grosse,
Josh Tenenbaum, Zoubin Ghahramani

December 2, 2013

Motivation

- Models today built by hand, or chosen from a fixed set.
  - Example: Scikit-learn
  - 

  - 

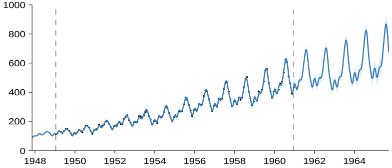
Problems with this approach

- Building by hand requires expertise, understanding of the dataset
- just being nonparametric isn't good enough
- can silently fail.

Motivation

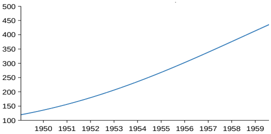
- Andrew Gelman asks: How would an AI do statistics?
- would need:
  - a language for describing arbitrarily complicated models
  - a way to search over those models
  - a way of checking model fit
- We built such a language over regression models, a procedure to search over them, and a method to describe in english language the properties of the resulting models.
  - Working on automatic model-checking.

Example



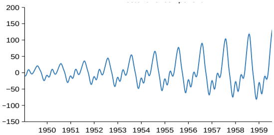
=

entire signal



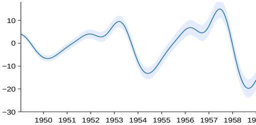
A very smooth, monotonically increasing function

+



An approximately periodic function with a period of 1.0 years and with approximately linearly increasing amplitude

+



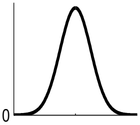
An exactly periodic functio with a period of 4.3 years bu with linearly increasing ampli tude

How to build a language of models?

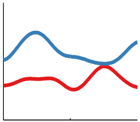
- We'll do this by defining a language on GP kernels
- Simple rules to combine them give diverse structures

Kernel Determines Structure

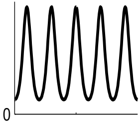
- Kernel determines almost all the properties of the prior.
- Many different kinds, with very different properties:



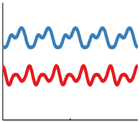
Squared-exp (SE)



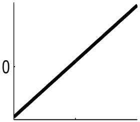
local variation



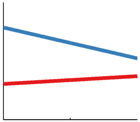
Periodic (Per)



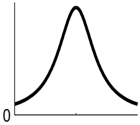
repeating structure



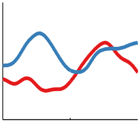
Linear (Lin)



linear functions



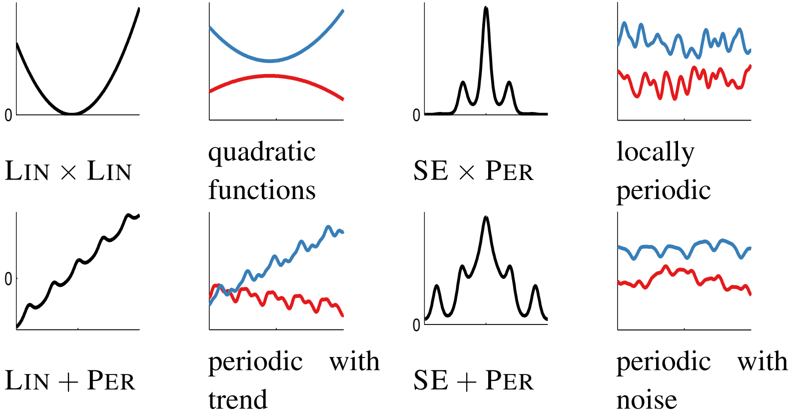
Rational-quadratic(RQ)



multi-scale variation

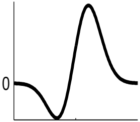
Kernels can be composed

- Two main operations: adding, multiplying

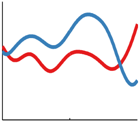


Kernels can be composed

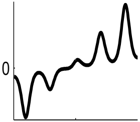
- Can be composed across multiple dimensions



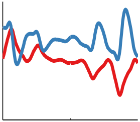
$\textsc{Lin} \times \textsc{SE}$



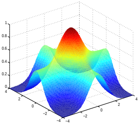
increasing variation



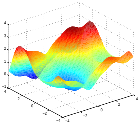
$\textsc{Lin} \times \textsc{Per}$



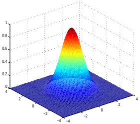
growing amplitude



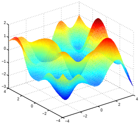
$\text{SE}_1 + \text{SE}_2$



$f_1(x_1) + f_2(x_2)$



$\text{SE}_1 \times \text{SE}_2$



$f(x_1, x_2)$

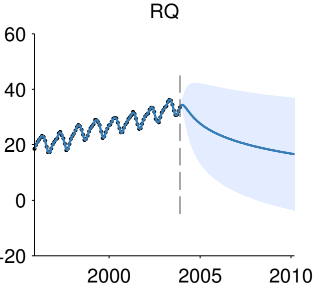
SPECIAL CASES

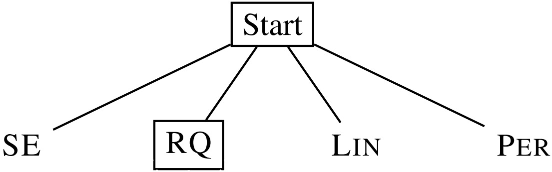
Bayesian linear regression	LIN
Bayesian polynomial regression	LIN $\times$ LIN $\times \ldots$
Generalized Fourier decomposition	PER $+$ PER $+ \ldots$
Generalized additive models	$\sum_{d=1}^{D} \mathrm{SE}_d$
Automatic relevance determination	$\prod_{d=1}^{D} \mathrm{SE}_d$
Linear trend with deviations	LIN $+$ SE
Linearly growing amplitude	LIN $\times$ SE

COMPOSITIONAL STRUCTURE SEARCH

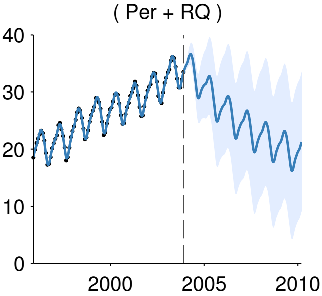
- Define grammar over kernels:
  - $K \to K + K$
  - $K \to K \times K$
  - $K \to \{\text{SE}, \text{LIN}, \text{PER}\}$
- Search the space of kernels greedily by applying production rules, checking model fit (approximate marginal likelihood).

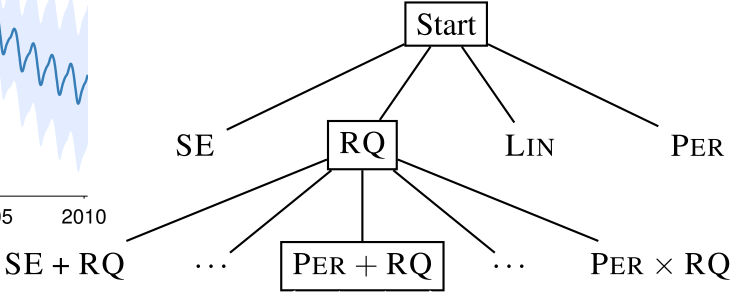
Compositional Structure Search



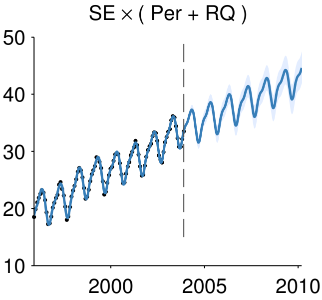


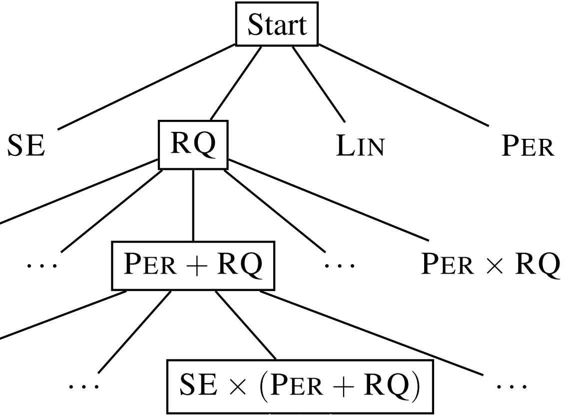
COMPOSITIONAL STRUCTURE SEARCH



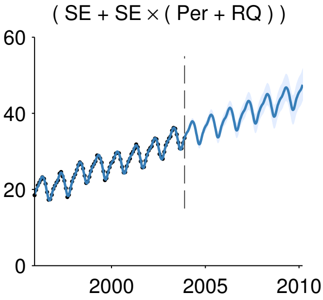


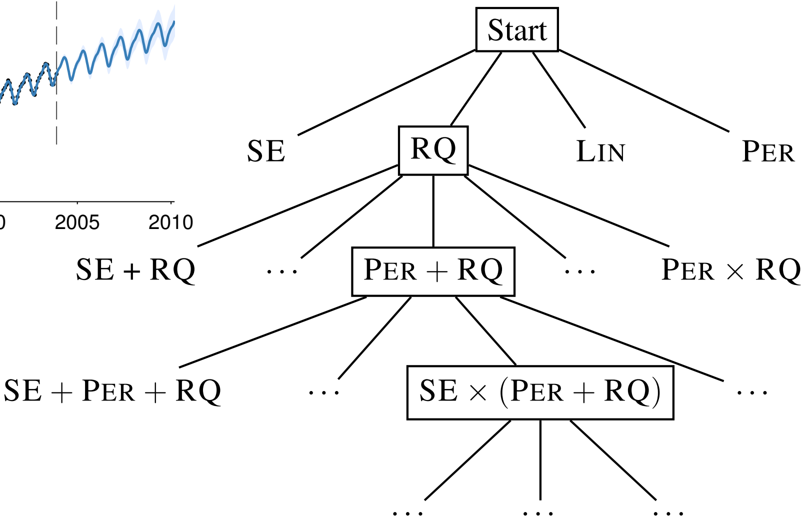
Compositional Structure Search





Compositional Structure Search





DISTRIBUTIVITY HELPS INTERPRETABILITY

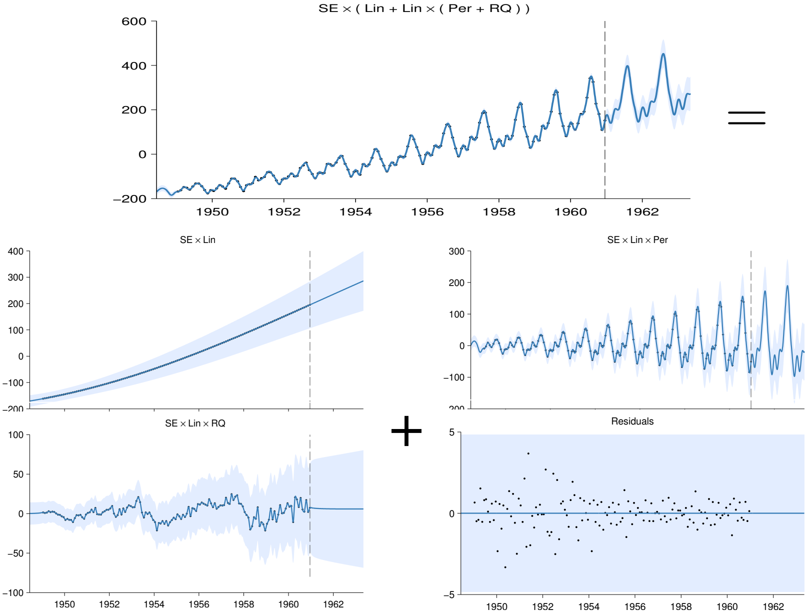
We can write all kernels as sums of products of base kernels:

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

Sums of kernels are equivalent to sums of functions.
If f_1, f_2 are independent, and $f_1 \sim \mathcal{GP}(\mu_1, k_1), f_2 \sim \mathcal{GP}(\mu_2, k_2)$.
Then it follows that

$$(f_1 + f_2) \sim \mathcal{GP}(\mu_1 + \mu_2, k_1 + k_2)$$

EXAMPLE DECOMPOSITION: AIRLINE



EXAMPLE KERNEL DESCRIPTIONS

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

This analysis was automatically generated

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

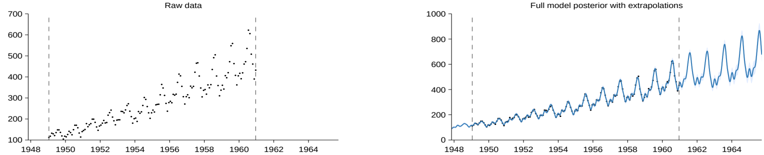


Figure 1: Raw data (left) and model posterior with extrapolation (right)

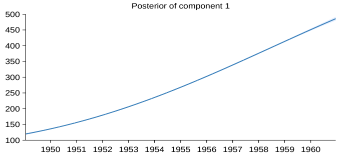
The structure search algorithm has identified four additive components in the data:

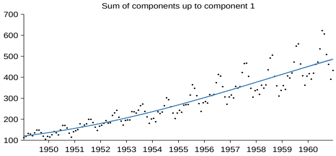
- A very smooth monotonically increasing function.
- An approximately periodic function with a period of 1.0 years and with approximately linearly increasing amplitude.
- An exactly periodic function with a period of 4.3 years but with linearly increasing amplitude.
- Uncorrelated noise with linearly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

This analysis was automatically generated

2.1 Component 1

This component is a very smooth and monotonically increasing function.

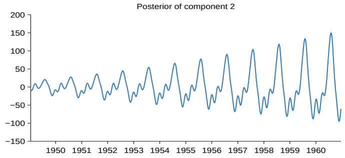


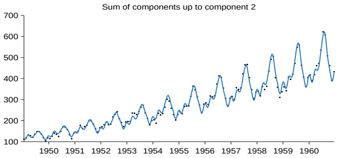


THIS ANALYSIS WAS AUTOMATICALLY GENERATED

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.

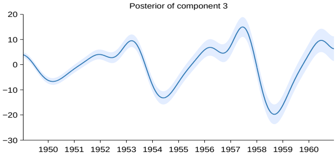


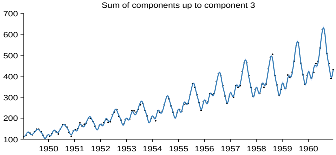


THIS ANALYSIS WAS AUTOMATICALLY GENERATED

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.





Summary

- Compositions of kernels give a language of models.
- Can search over models automatically.
- Kernels modify prior in predictable ways, allowing automatic english description of models.

SUMMARY

- Compositions of kernels give a language of models.
- Can search over models automatically.
- Kernels modify prior in predictable ways, allowing automatic english description of models.

Thanks!