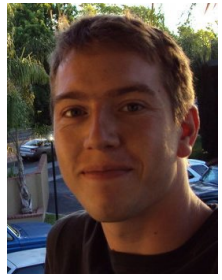
Automatic Construction and Natural-Language Description of Nonparametric Regression Models

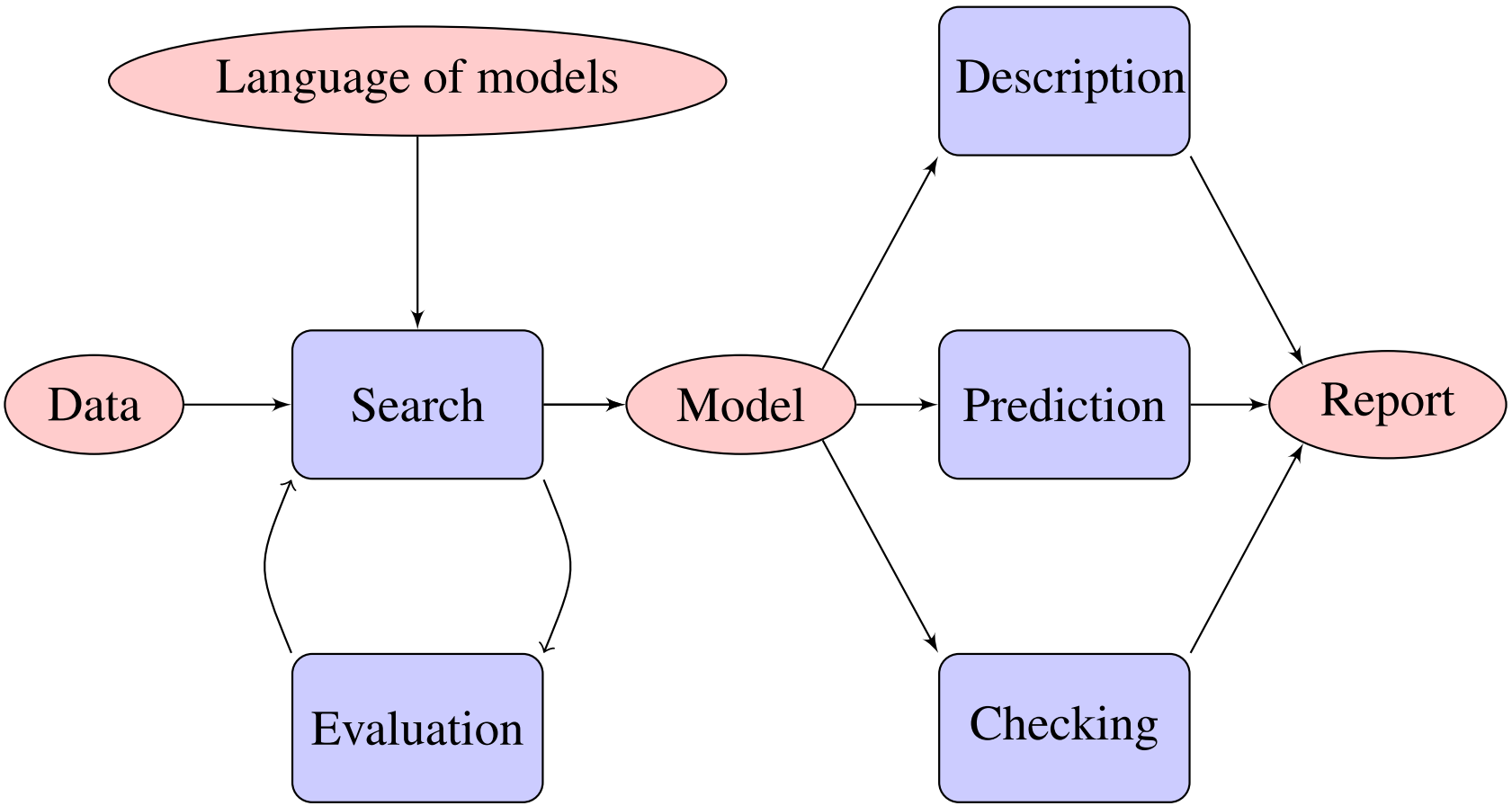




James Robert Lloyd[1], David Duvenaud[1], Roger Grosse[2],

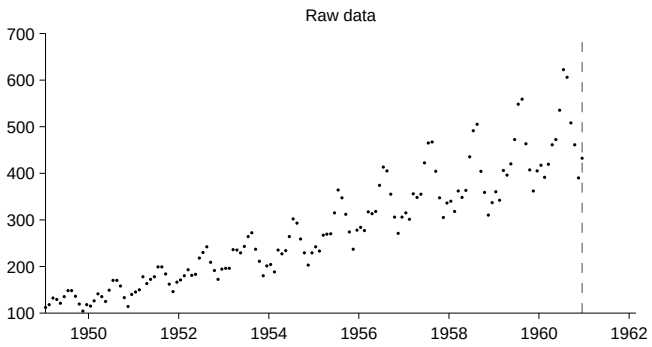

Joshua Tenenbaum[2], Zoubin Ghahramani[1]

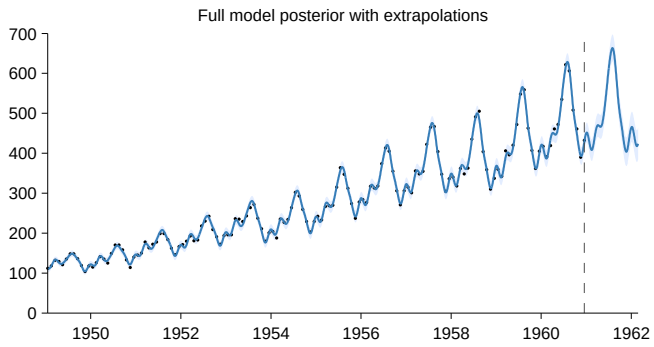
1: Department of Engineering, University of Cambridge, UK
2: Massachusetts Institute of Technology, USA

A SYSTEM FOR AUTOMATIC DATA ANALYSIS

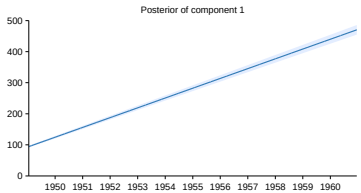


An entirely automatic analysis

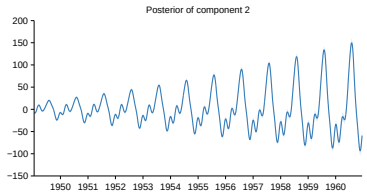




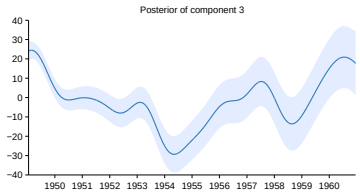
Four additive components have been identified in the data



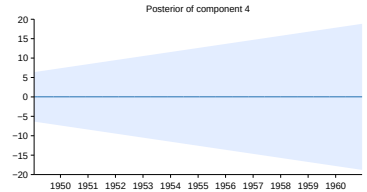
A linearly increasing function



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



A smooth function



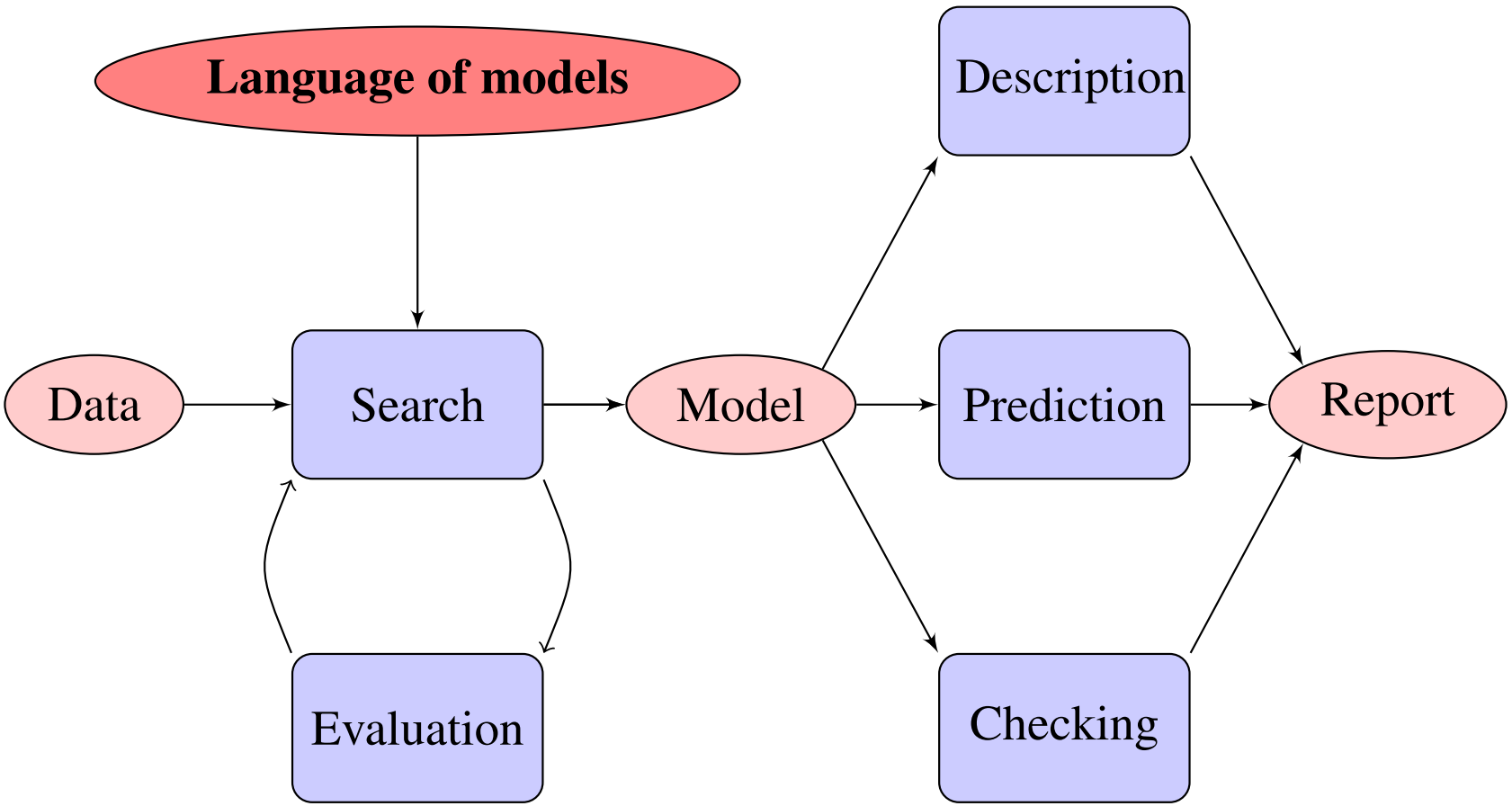
Uncorelated noise with linearly increasing standard deviation

NATURAL LANGUAGE DESCRIPTIONS OF MODELS

Compositionally constructed statistical models

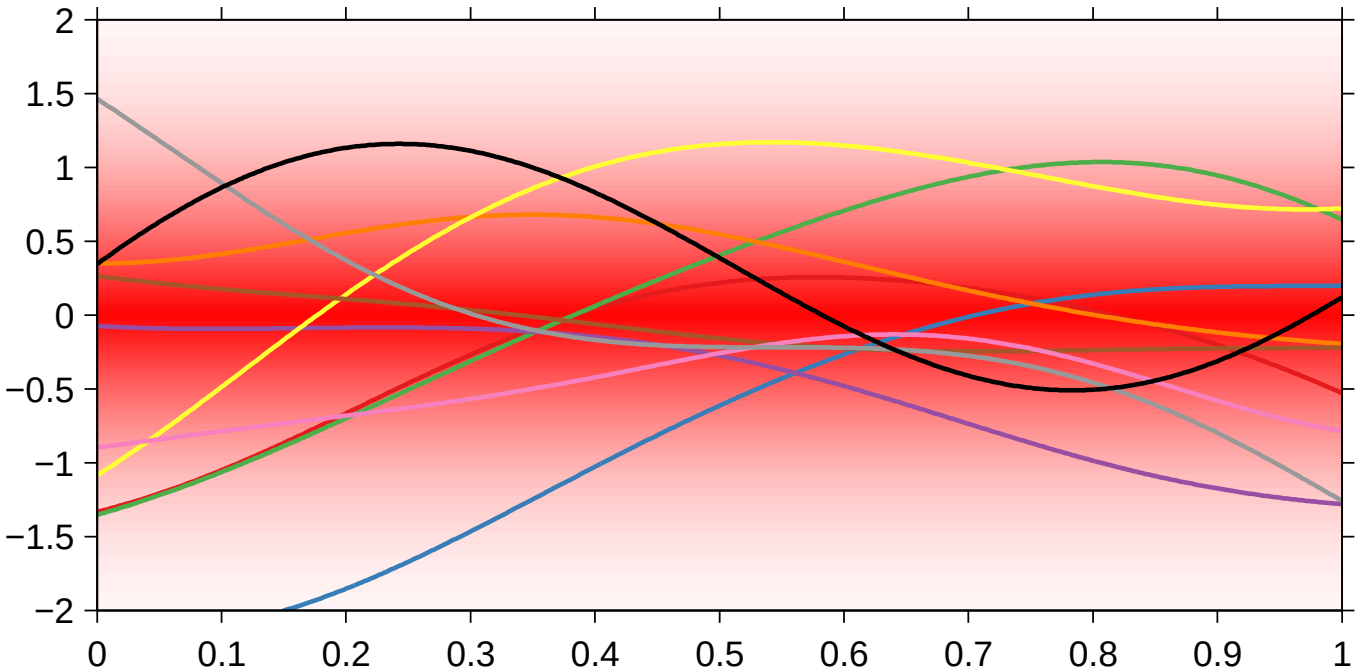
Compositionally constructed
natural-language descriptions

Defining a language of models



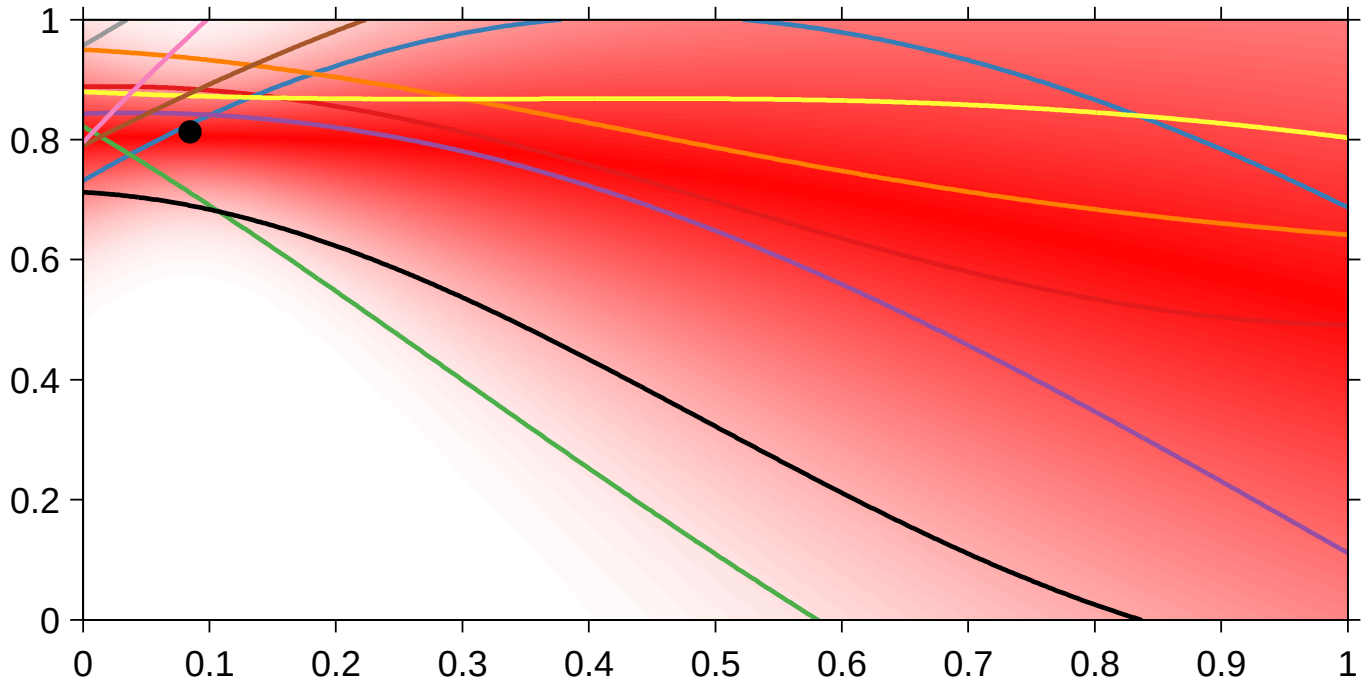
A language of Gaussian processes

- Define probability distributions on functions
- Used to perform Bayesian (nonlinear) regression



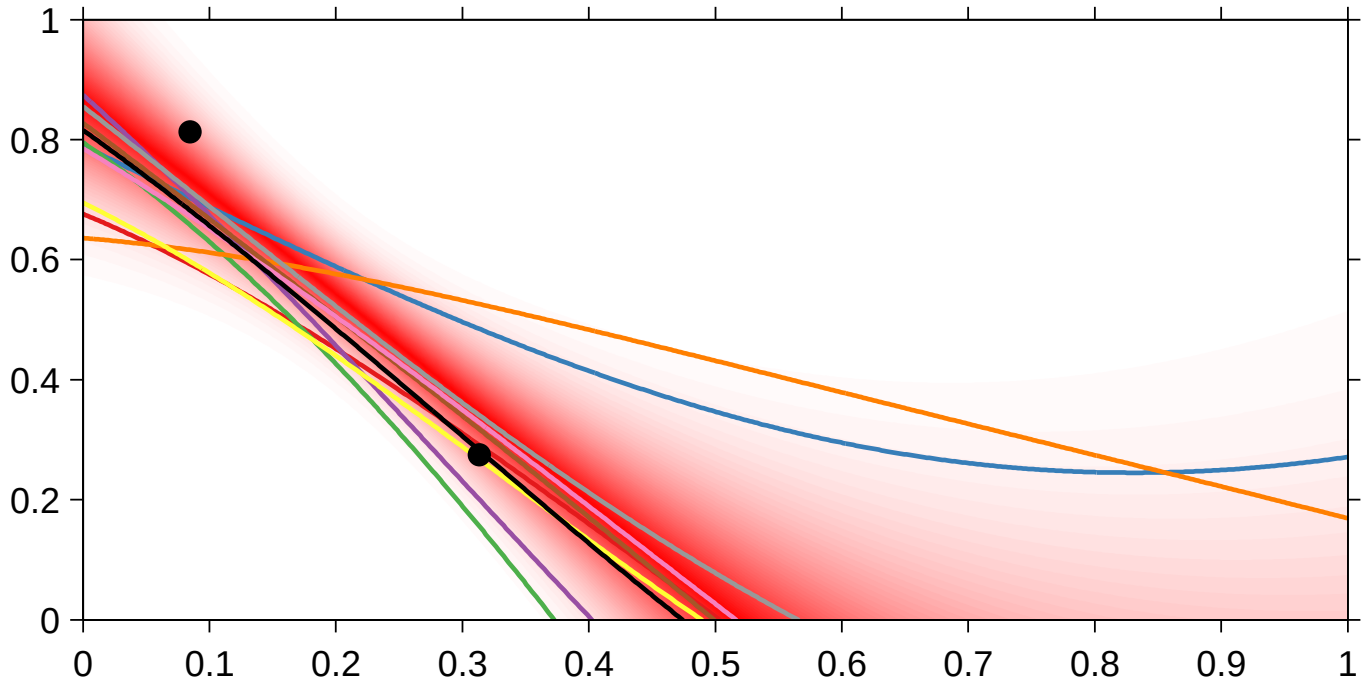
A language of Gaussian processes

- Define probability distributions on functions
- Used to perform Bayesian (nonlinear) regression



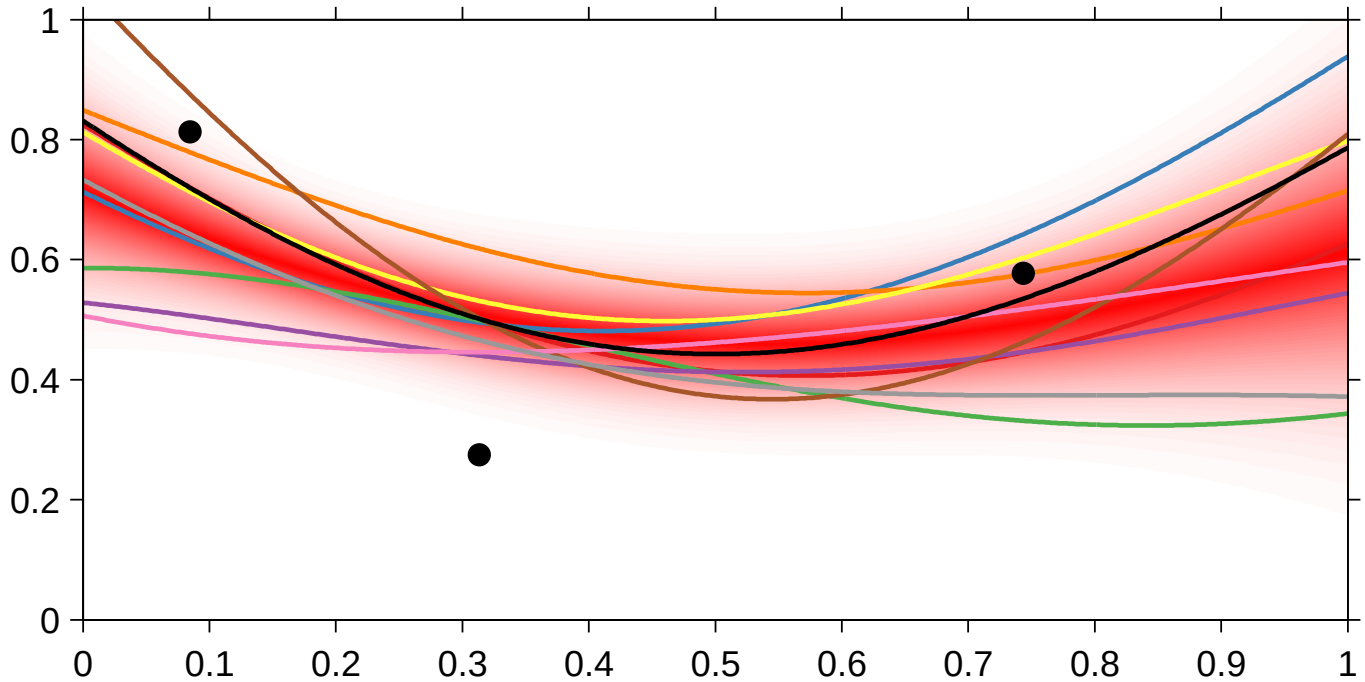
A language of Gaussian processes

- Define probability distributions on functions
- Used to perform Bayesian (nonlinear) regression



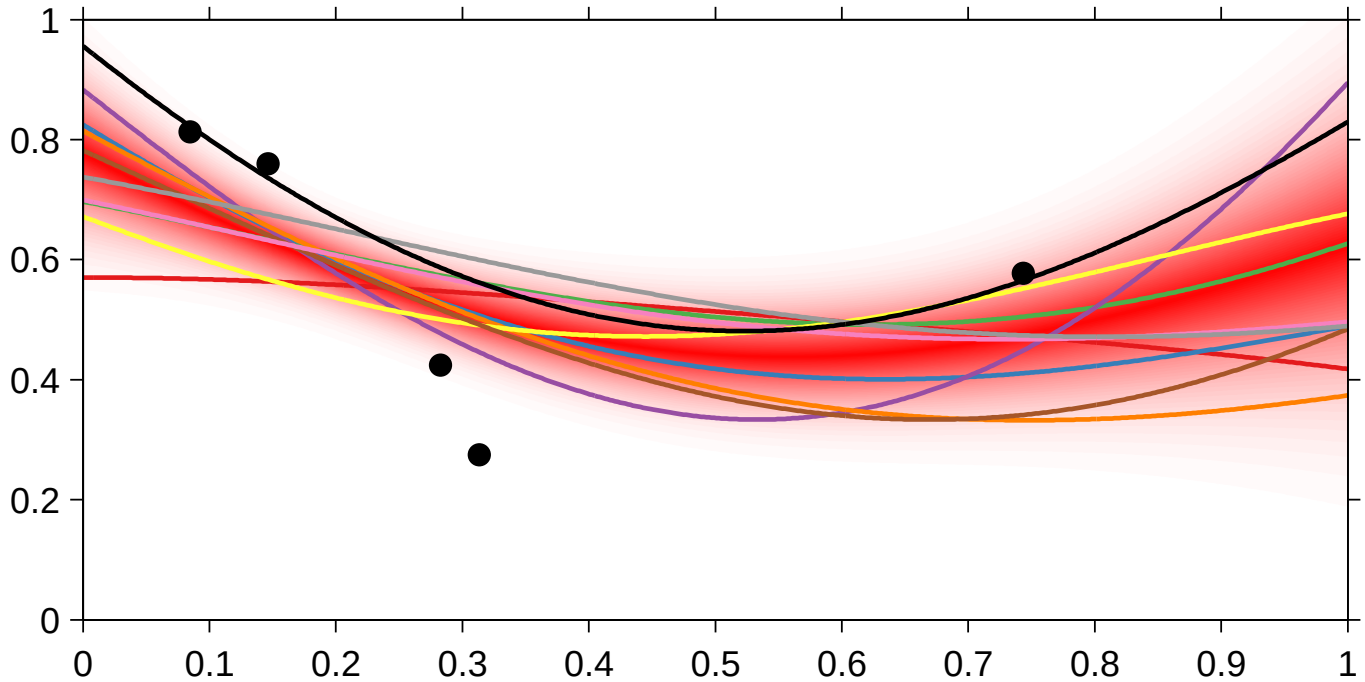
A language of Gaussian processes

- Define probability distributions on functions
- Used to perform Bayesian (nonlinear) regression



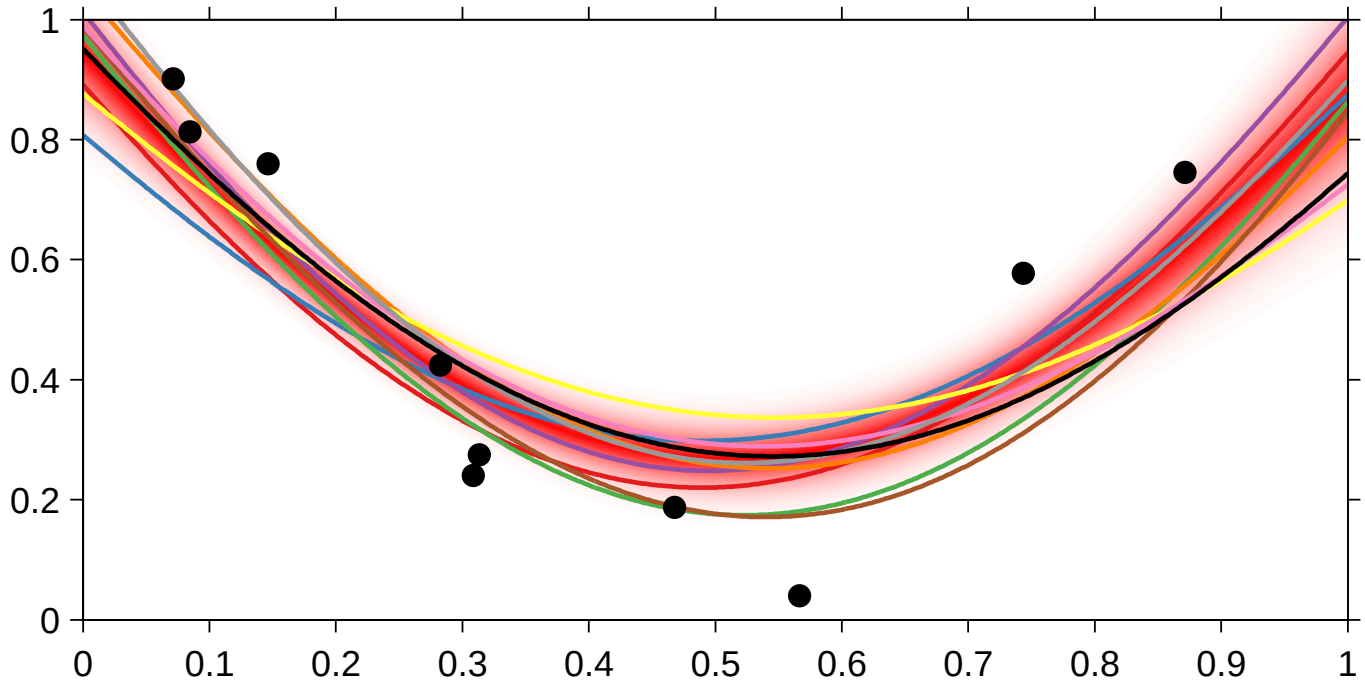
A language of Gaussian processes

- Define probability distributions on functions
- Used to perform Bayesian (nonlinear) regression



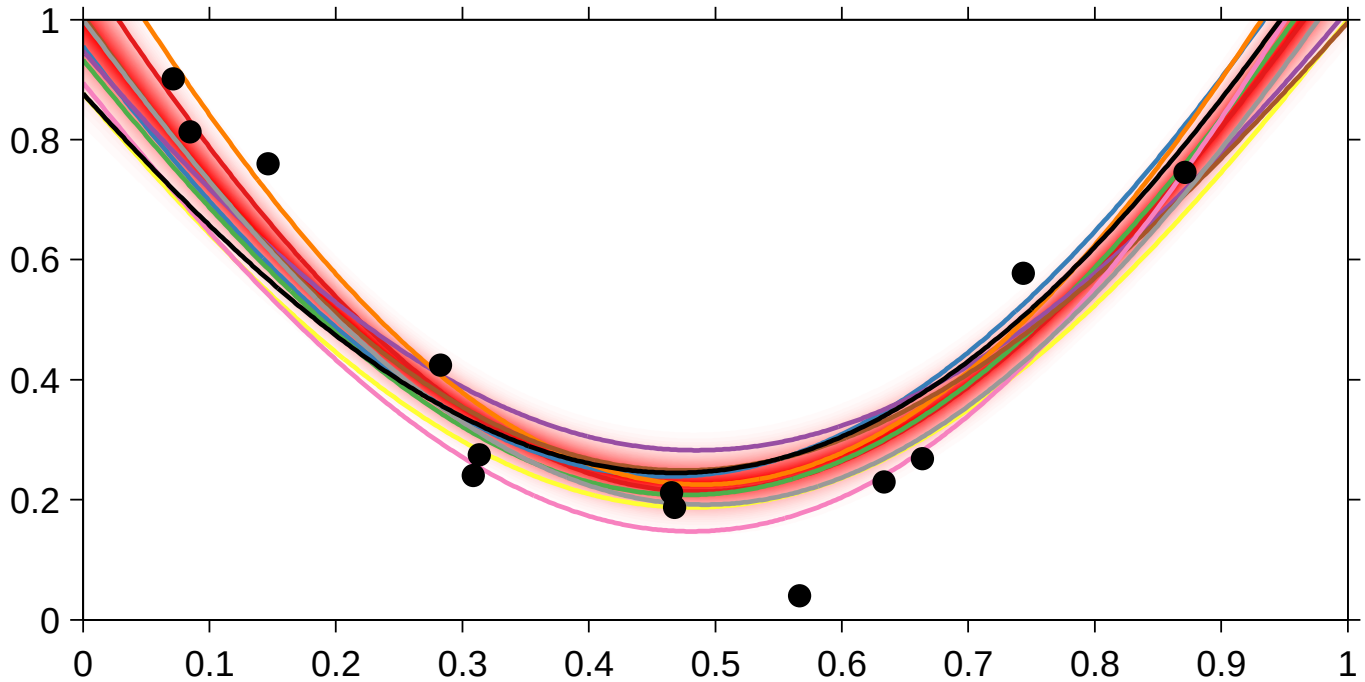
A language of Gaussian processes

- Define probability distributions on functions
- Used to perform Bayesian (nonlinear) regression



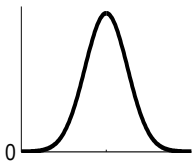
A language of Gaussian processes

- Define probability distributions on functions
- Used to perform Bayesian (nonlinear) regression

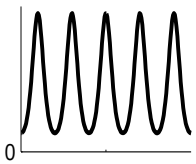


THE ATOMS OF OUR LANGUAGE

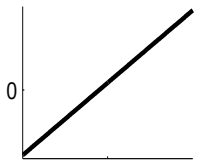
Five base kernels...



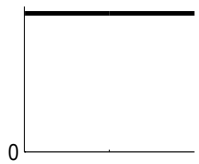
Squared exp. (SE)



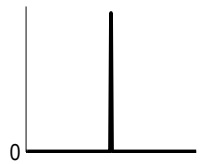
Periodic (PER)



Linear (LIN)

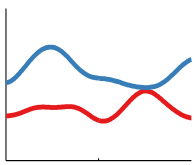


Constant (C)

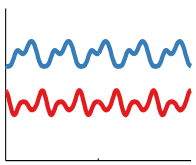


White noise (WN)

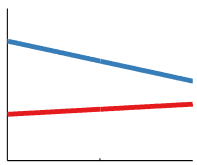
...encoding for the following types of functions



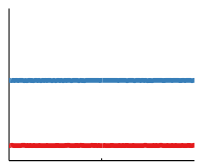
Smooth functions



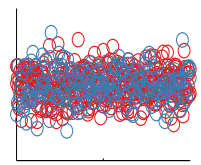
Periodic functions



Linear functions



Constant functions

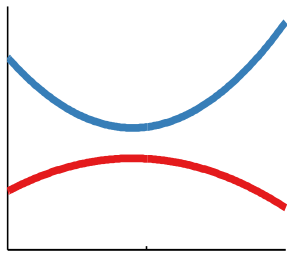


Gaussian noise

THE COMPOSITION RULES OF OUR LANGUAGE

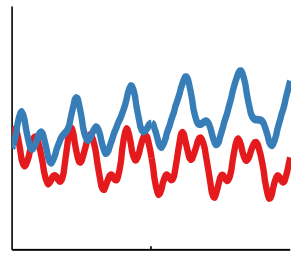
- Two main operations: addition, multiplication

LIN × LIN



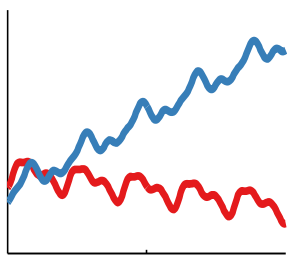
quadratic functions

SE × PER



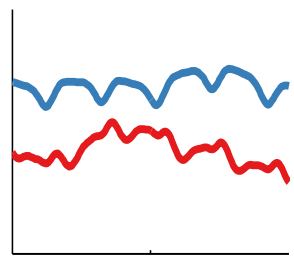
locally periodic

LIN + PER



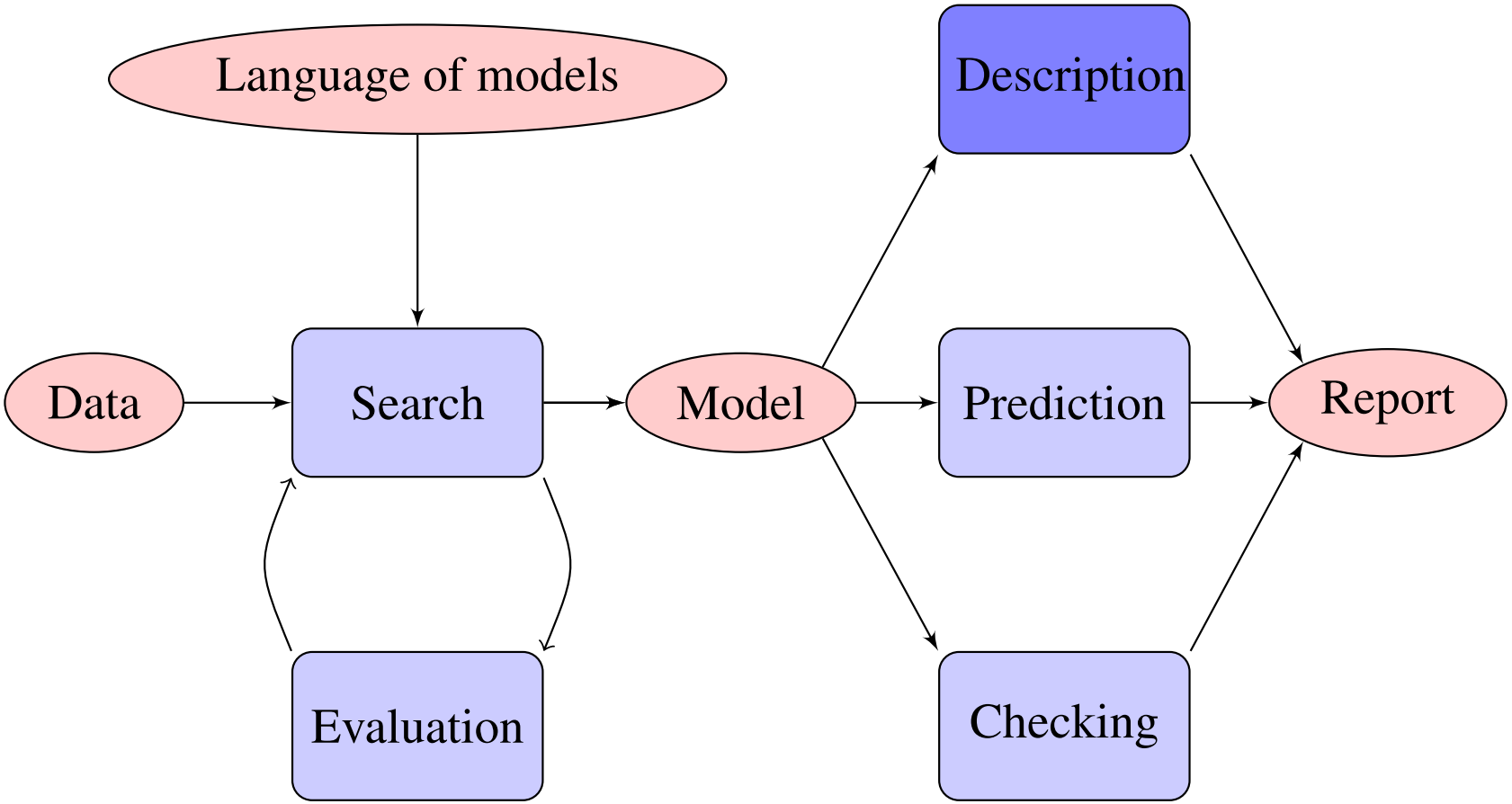
periodic plus linear trend

SE + PER



periodic plus smooth trend

Automatic translation of models

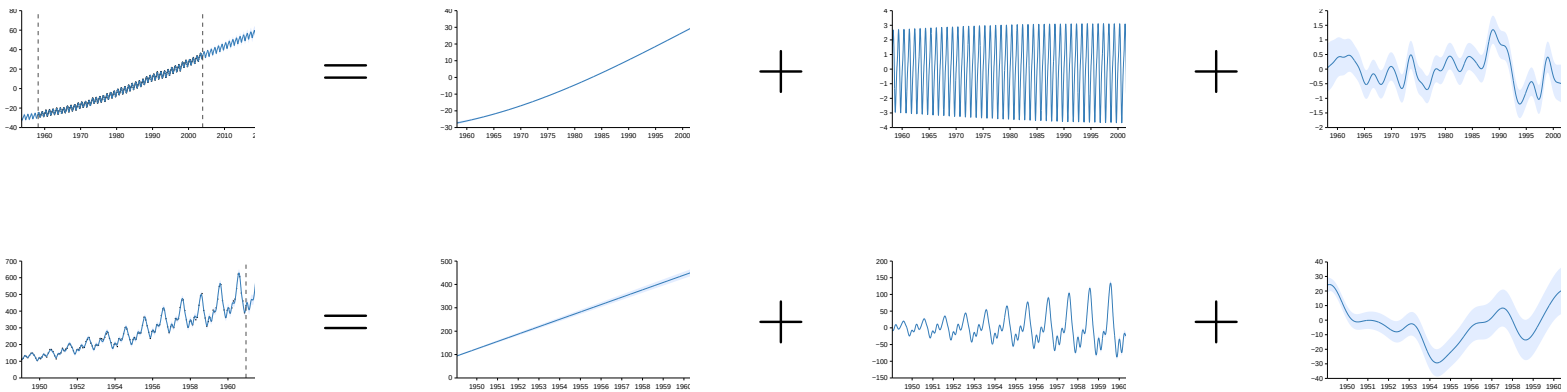


SUMS OF KERNELS ARE SUMS OF FUNCTIONS

If $f_1 \sim \mathrm{GP}(0, k_1)$ and independently $f_2 \sim \mathrm{GP}(0, k_2)$ then

$$f_1 + f_2 \sim \mathrm{GP}(0, k_1 + k_2)$$

e.g.



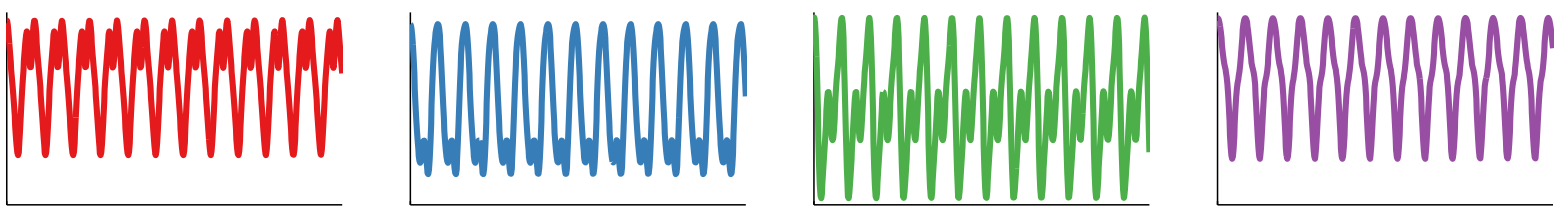
We can therefore describe each component separately

PRODUCTS OF KERNELS

PER

periodic function

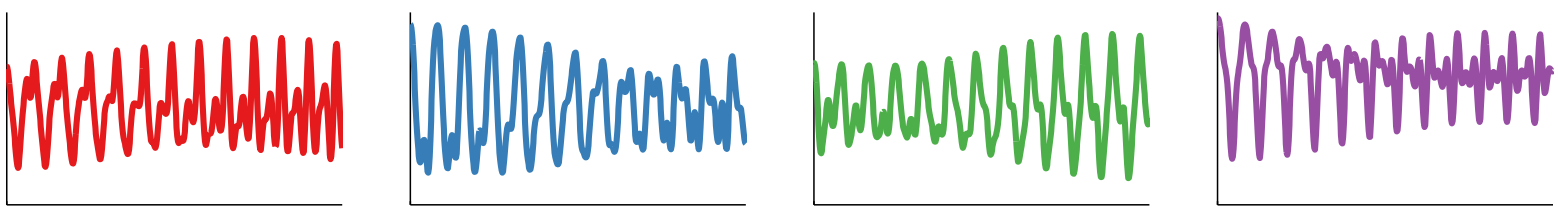
- Properties of individual kernels well understood
- Can be described with standard noun phrase



PRODUCTS OF KERNELS

$$\underbrace{\text{SE}}_{\text{approximately}} \times \underbrace{\text{PER}}_{\text{periodic function}}$$

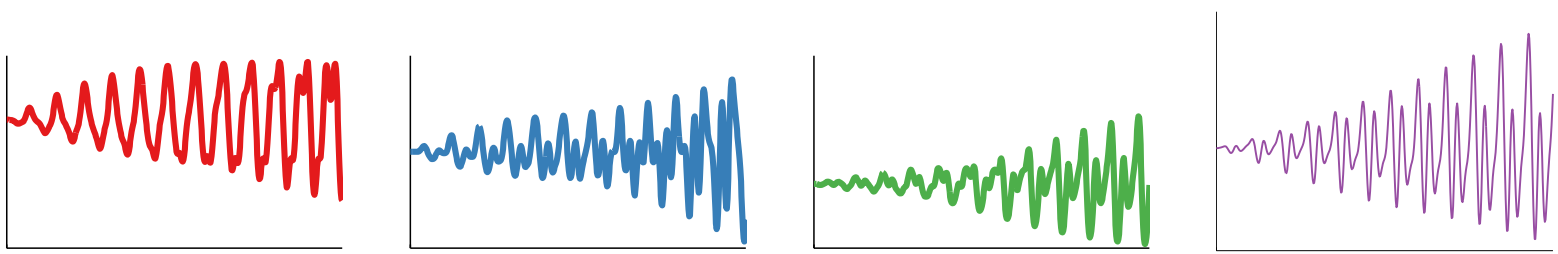
- Multiplying by each kernel has a consistent effect
- Can be described with consistent adjectives / modifiers



PRODUCTS OF KERNELS

$$\underbrace{\text{SE}}_{\text{approximately}} \times \underbrace{\text{PER}}_{\text{periodic function}} \times \underbrace{\text{LIN}}_{\text{with linearly growing amplitude}}$$

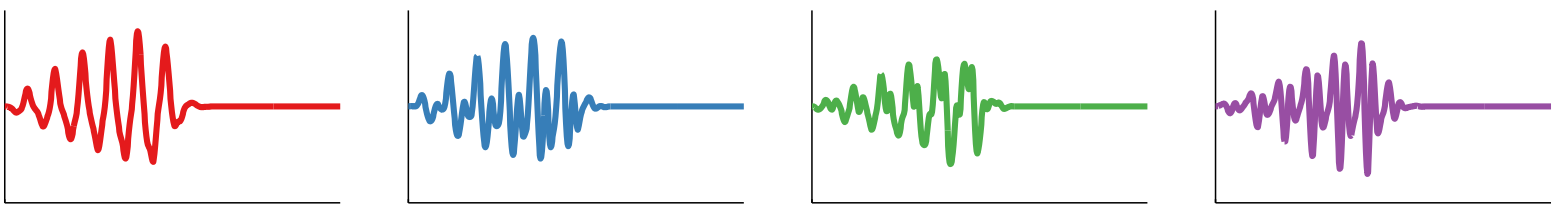
- Multiplying by each kernel has a consistent effect
- Can be described with consistent adjectives / modifiers



PRODUCTS OF KERNELS

$$\underbrace{\text{SE}}_{\text{approximately}} \times \underbrace{\text{PER}}_{\text{periodic function}} \times \underbrace{\text{LIN}}_{\text{with linearly growing amplitude}} \times \underbrace{\sigma}_{\text{until 1700}}$$

- Multiplying by each kernel has a consistent effect
- Can be described with consistent adjectives / modifiers



VISIT OUR WEBSITE - TRY THE (SIMPLER) DEMO

www.automaticstatistician.com