Automatic Construction and Natural-Language Description of Nonparametric Regression Models

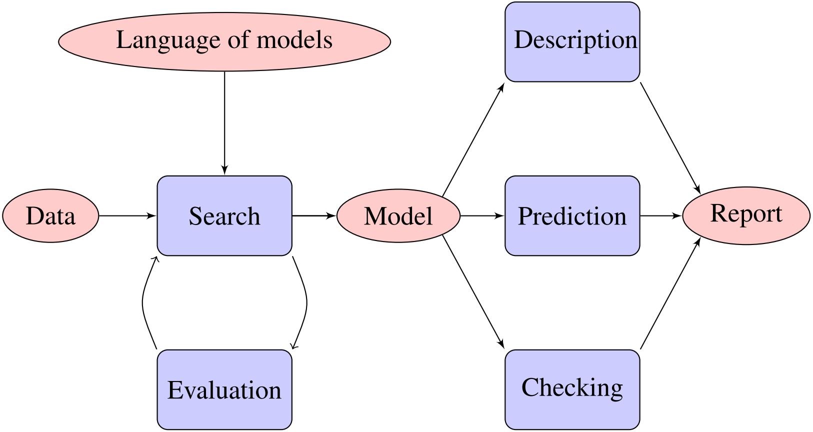


James Robert Lloyd[1], David Duvenaud[1], Roger Grosse[2],

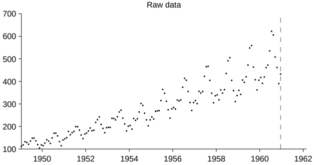

Joshua Tenenbaum[2], Zoubin Ghahramani[1]

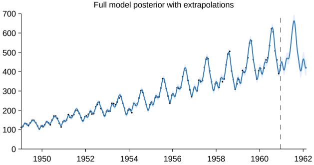
1: Department of Engineering, University of Cambridge, UK
2: Massachusetts Institute of Technology, USA

A SYSTEM FOR AUTOMATIC DATA ANALYSIS

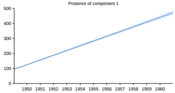


AN ENTIRELY AUTOMATIC ANALYSIS

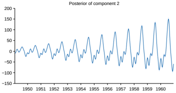




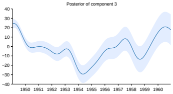
Four additive components have been identified in the data



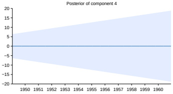
A linearly increasing function



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



A smooth function

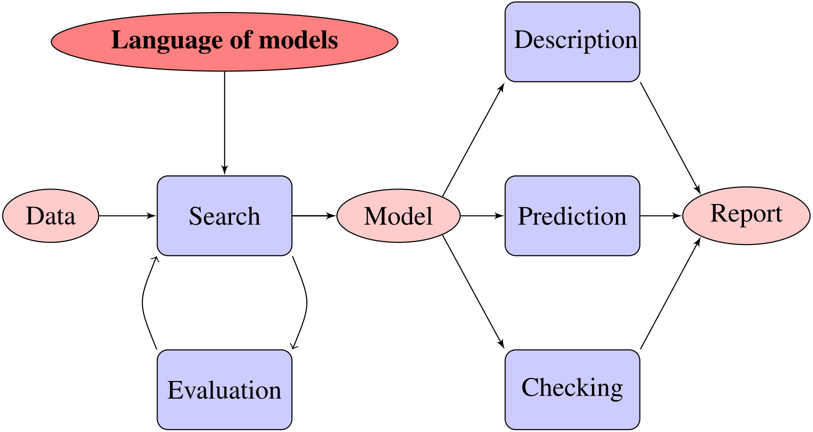


Uncorelated noise with linearly increasing standard deviation

Compositionally constructed statistical models

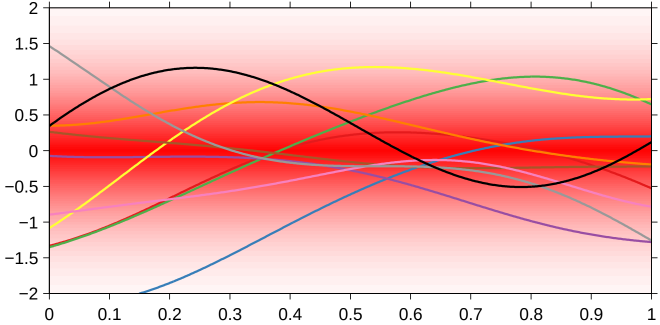
Compositionally constructed
natural-language descriptions

Defining a language of models

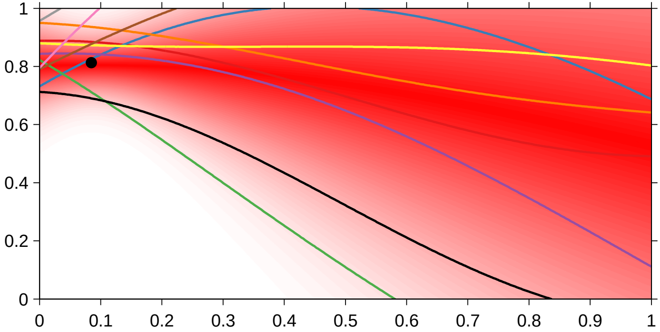


Gaussian process regression

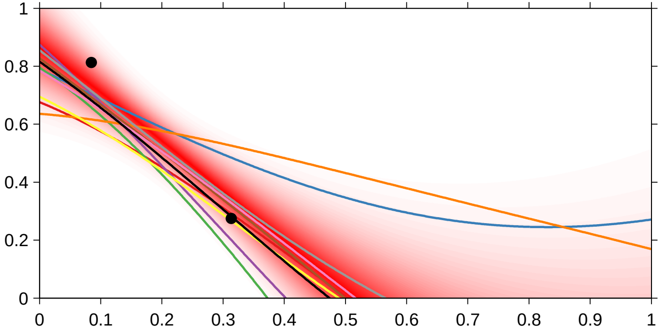
We can use Gaussian processes to place priors on functions and perform a Bayesian regression analysis



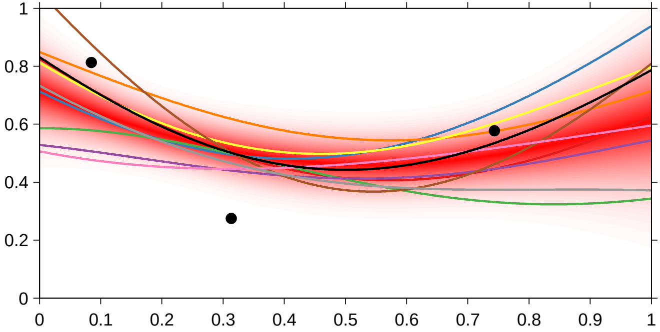
We can use Gaussian processes to place priors on functions and perform a Bayesian regression analysis



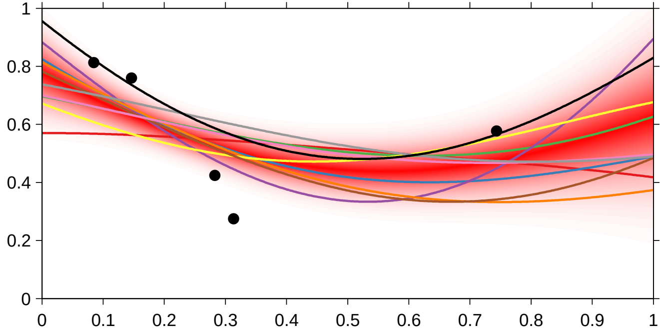
We can use Gaussian processes to place priors on functions and perform a Bayesian regression analysis



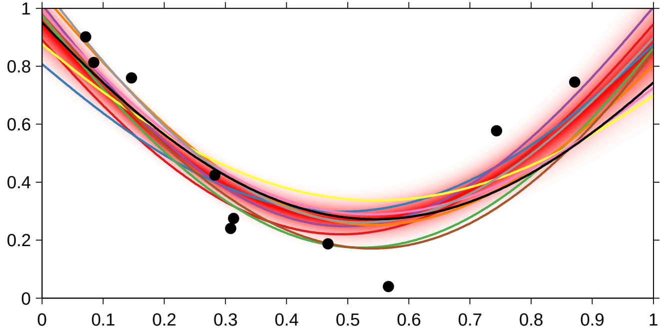
We can use Gaussian processes to place priors on functions and perform a Bayesian regression analysis



We can use Gaussian processes to place priors on functions and perform a Bayesian regression analysis

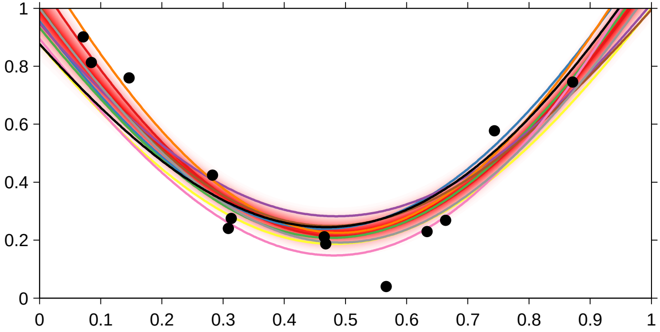


We can use Gaussian processes to place priors on functions and perform a Bayesian regression analysis

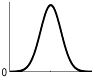


Gaussian process regression

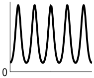
We can use Gaussian processes to place priors on functions and perform a Bayesian regression analysis



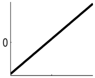
Five base kernels



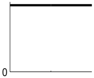
Squared exp. (SE)



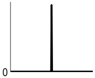
Periodic (PER)



Linear (LIN)

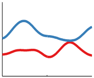


Constant (C)

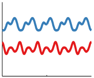


White noise (WN)

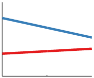
Encoding for the following types of functions



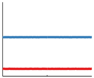
Smooth functions



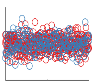
Periodic functions



Linear functions

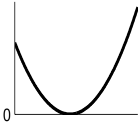


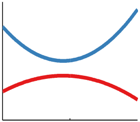
Constant functions



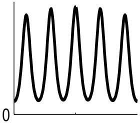
Gaussian noise

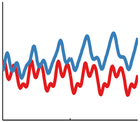
- Two main operations: addition, multiplication



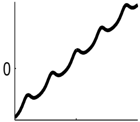


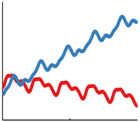
LIN × LIN — quadratic functions



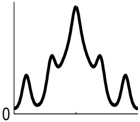


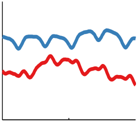
SE × PER — locally periodic





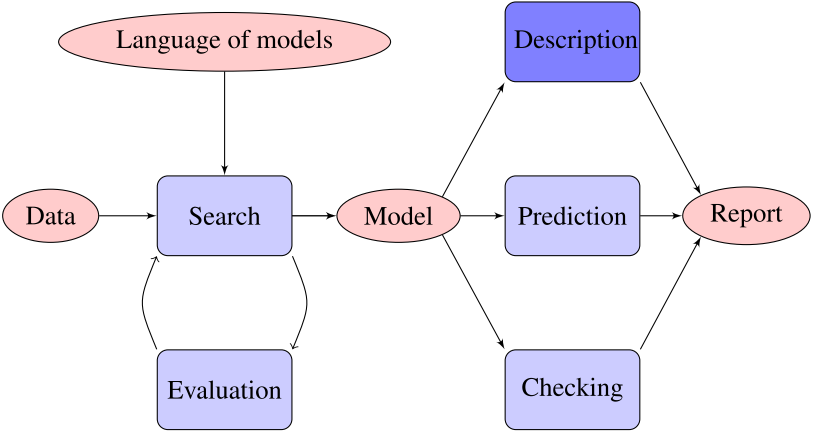
LIN + PER — periodic plus linear trend





SE + PER — periodic plus smooth trend

AUTOMATIC TRANSLATION OF MODELS

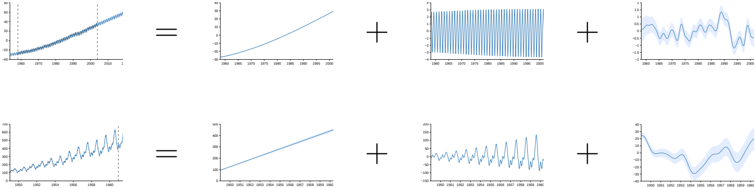


SUMS OF KERNELS ARE SUMS OF FUNCTIONS

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

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

e.g.

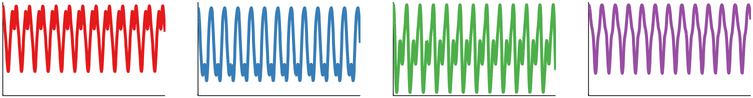


We can therefore describe each component separately

PRODUCTS OF KERNELS

$$\underbrace{\text{PER}}_{\text{periodic function}}$$

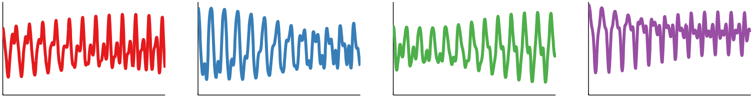
On their own, each kernel is described by a standard noun phrase



Products of kernels - SE

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

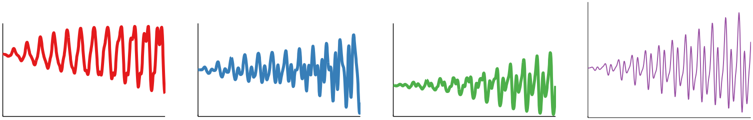
Multiplication by SE removes long range correlations from a model since $\text{SE}(x, x')$ decreases monotonically to 0 as $|x - x'|$ increases.



PRODUCTS OF KERNELS - LIN

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

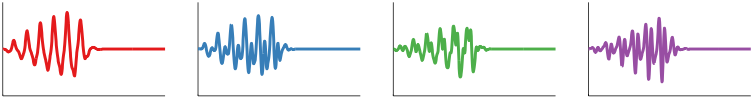
Multiplication by LIN is equivalent to multiplying the function being modeled by a linear function. If $f(x) \sim \mathrm{GP}(0, k)$, then $xf(x) \sim \mathrm{GP}(0, k \times \mathrm{LIN})$. This causes the standard deviation of the model to vary linearly without affecting the correlation.



PRODUCTS OF KERNELS - CHANGEPOINTS

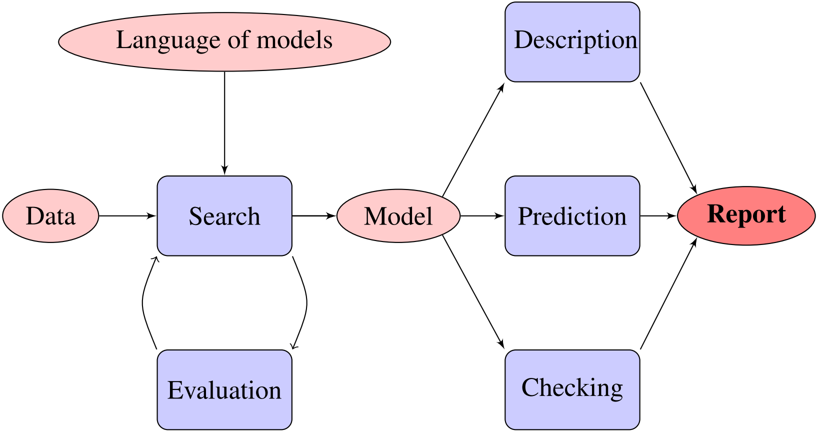
$$\underbrace{\text{SE}}_{\text{approximately}} \times \underbrace{\text{PER}}_{\text{periodic function}} \times \underbrace{\text{LIN}}_{\text{with linearly growing amplitude}} \times \underbrace{\boldsymbol{\sigma}}_{\text{until 1700}}$$

Multiplication by $\boldsymbol{\sigma}$ is equivalent to multiplying the function being modeled by a sigmoid.



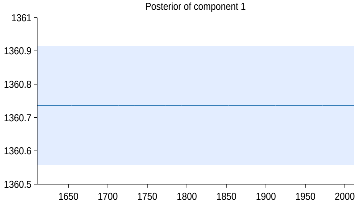
Kernel	Noun phrase	Postmodifier phrase
WN	uncorrelated noise	n/a
C	constant	n/a
SE	smooth function	whose shape changes smoothly
PER	periodic function	modulated by a periodic function
LIN	linear function	with linearly varying amplitude
$\prod_k \text{LIN}^{(k)}$	polynomial	with polynomially varying amplitude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

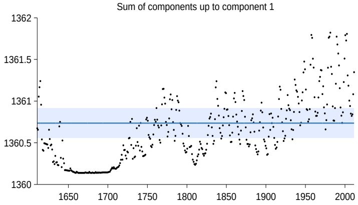
Automatically generated reports



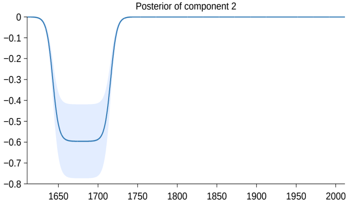
Example: Solar irradiance

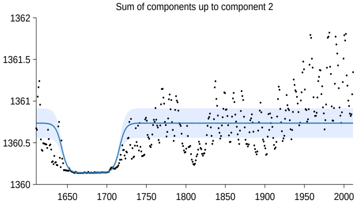
This component is constant.





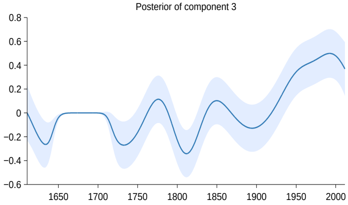
This component is constant. This component applies from 1643 until 1716.

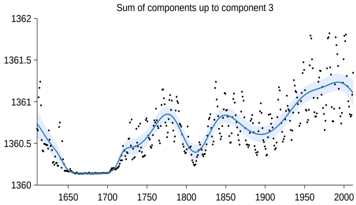




EXAMPLE: SOLAR IRRADIANCE

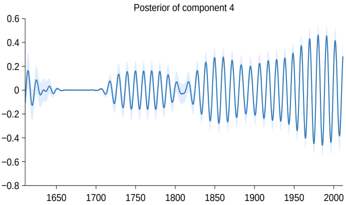
This component is a smooth function with a typical lengthscale of 23.1 years. This component applies until 1643 and from 1716 onwards.

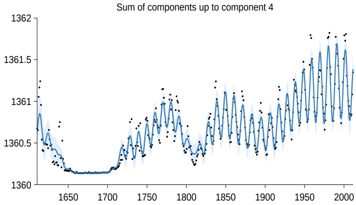




EXAMPLE: SOLAR IRRADIANCE

This component is approximately periodic with a period of 10.8 years. Across periods the shape of this function varies smoothly with a typical lengthscale of 36.9 years. The shape of this function within each period is very smooth and resembles a sinusoid. This component applies until 1643 and from 1716 onwards.





www.automaticstatistician.com