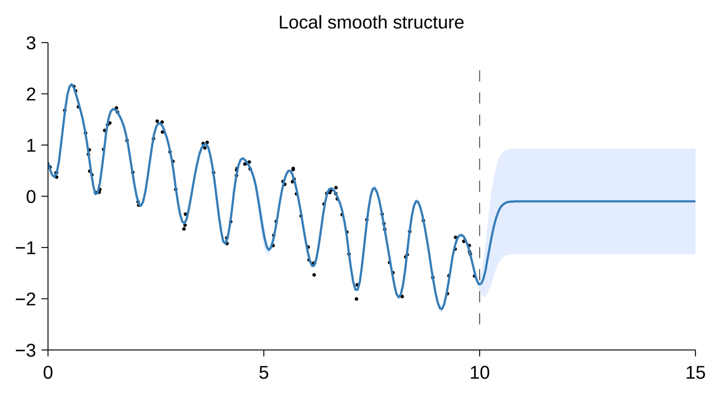
Structure Discovery in Nonparametric Regression through Compositional Kernel Search

Massachusetts Institute of Technology

David Duvenaud, James Robert Lloyd, Roger Grosse, Joshua B. Tenenbaum, Zoubin Ghahramani

UNIVERSITY OF CAMBRIDGE

Identifying structure is crucial for extrapolation

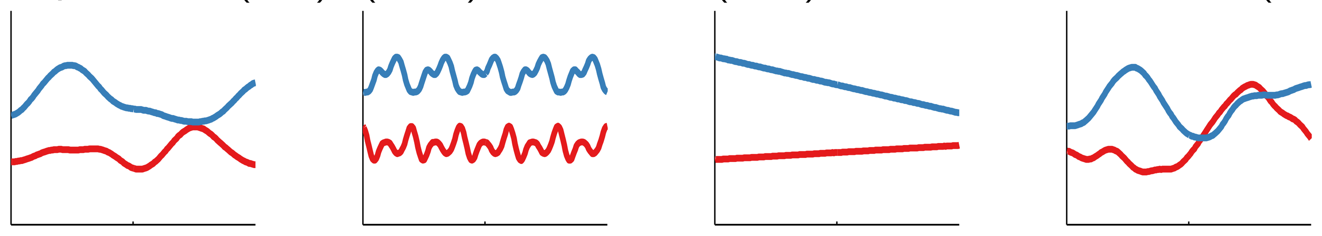


- Traditionally, a statistician proposes an appropriate model for the type of structures present
- Automatic model selection techniques already exist, typically choosing between a finite or restricted set of models
- Instead, we automate statistical model *construction*

Gaussian processes model structure through kernels

- The kernel specifies which structures are likely under the GP prior, which in turn determines the generalization properties of the model.
- Below we list standard base kernels, and draws from the corresponding GP priors:

Squared Exponential (SE) | Periodic (PER) | Linear (LIN) | Rational Quadratic (RQ)

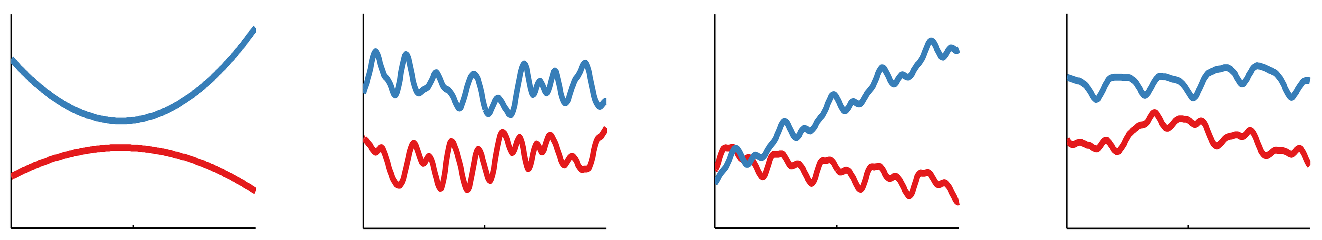


local variation | repeating structure | linear functions | multiscale variation

Kernels can be Composed

- Constructing appropriate composite kernels has previously been done by experts

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



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

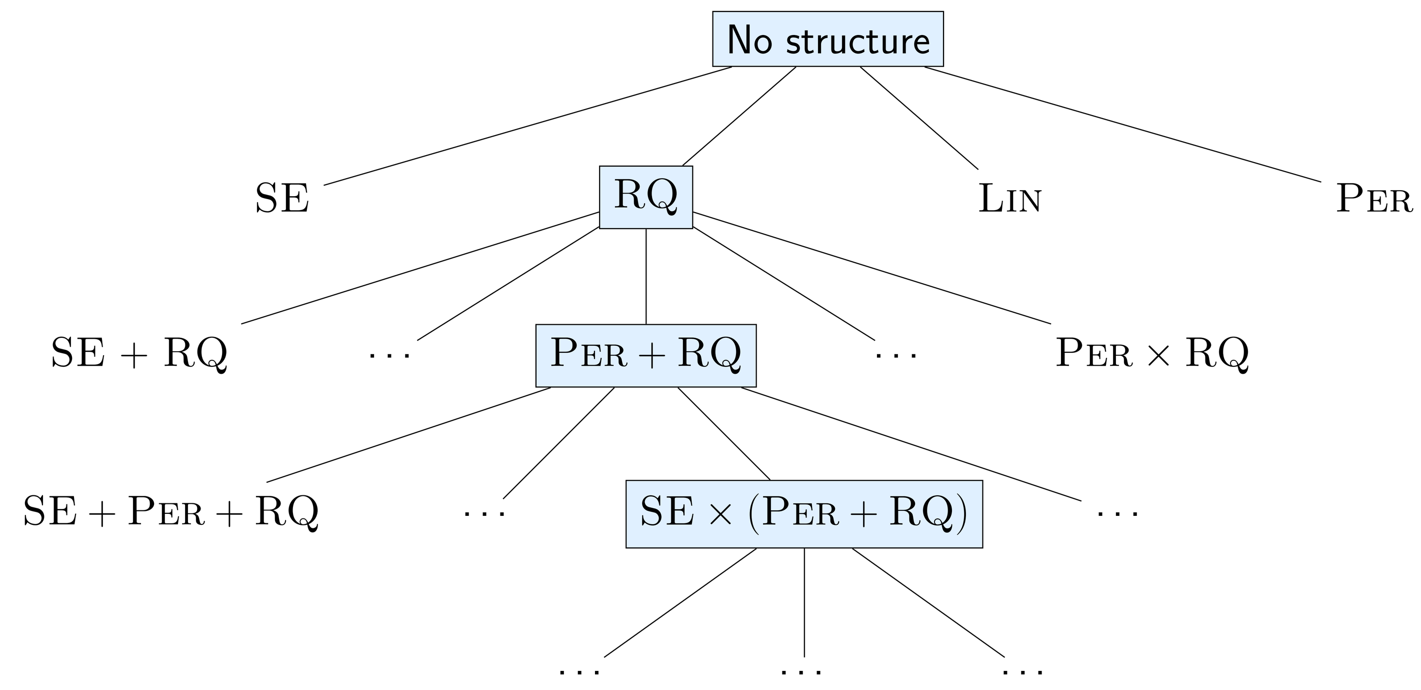
. . . defining an open-ended set of structures

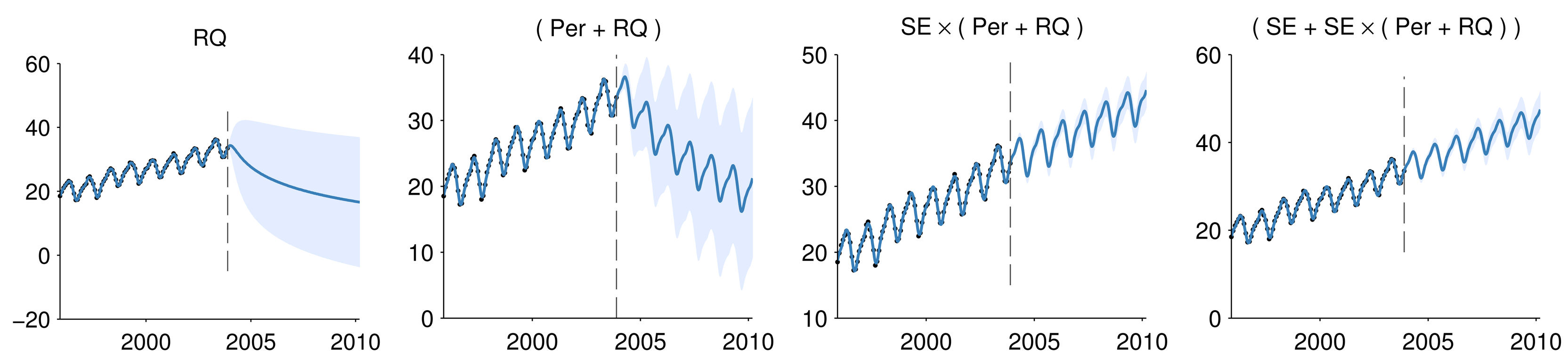
- We consider all algebraic expressions involving a small number of base kernels and the operations '+' and '×', including e.g.

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

. . . by a greedy search

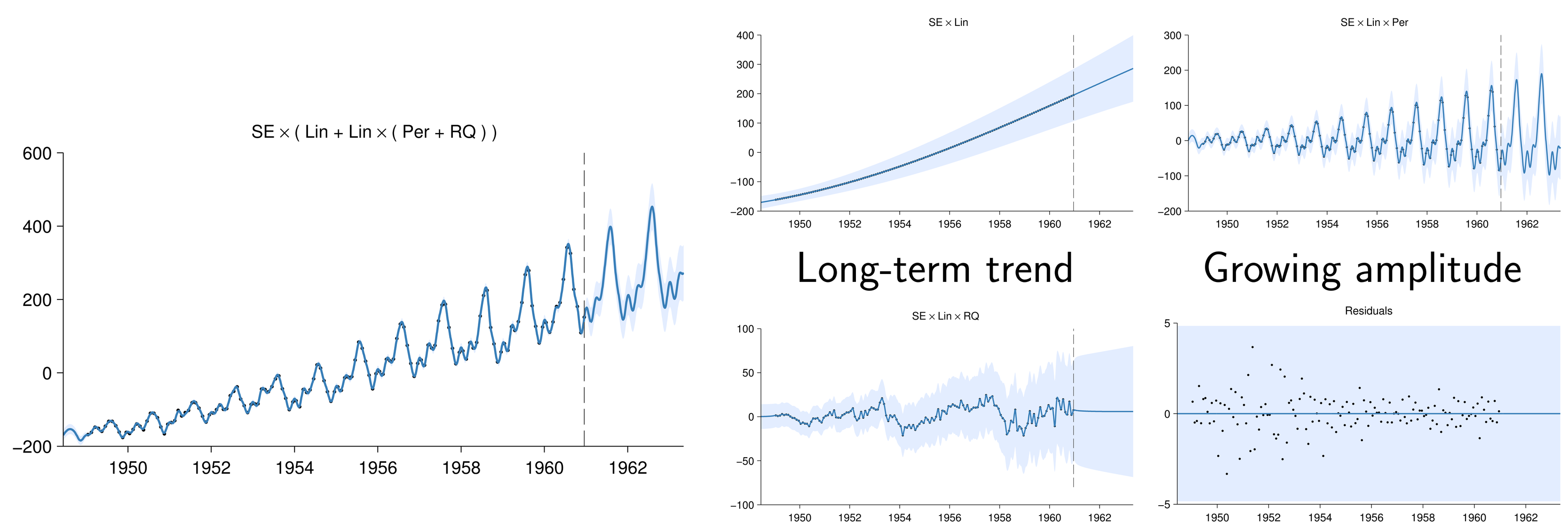
- We try all base kernels, selecting the one with the highest (approximate) marginal likelihood which balances data fit and model complexity
- The search continues by adding an extra term to the current best kernel, stopping when marginal likelihood no longer improves





Compound Kernels are Interpretable

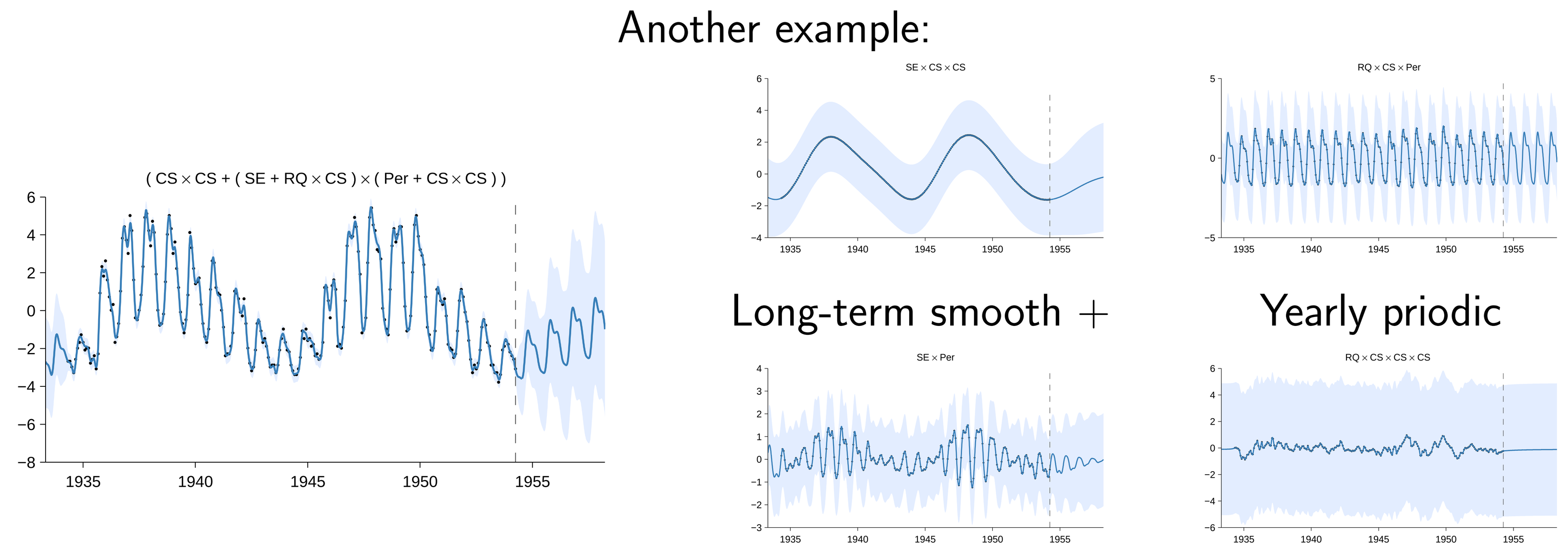
- Compound kernels decompose functions into additive components



Long-term trend | Growing amplitude

Short-term deviatinos | noise

Another example:



Long-term smooth + | Yearly priodic

Long-term amlpitude + | Short-term noise