

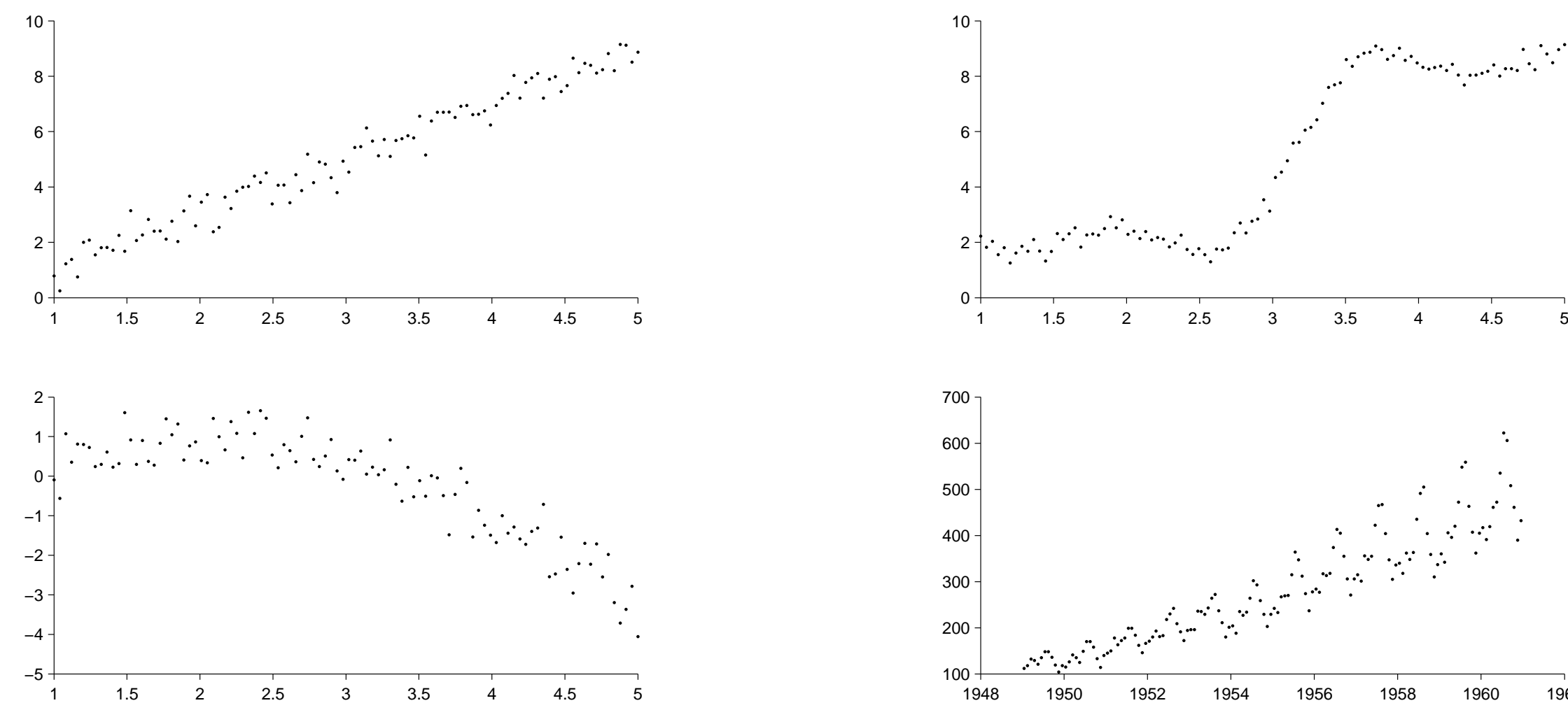


# Automating pattern discovery and the statistical process for regression

David Duvenaud<sup>1</sup>, James Robert Lloyd<sup>1</sup>, Roger Grosse<sup>2</sup>,  
Joshua B. Tenenbaum<sup>2</sup>, Zoubin Ghahramani<sup>1</sup>

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

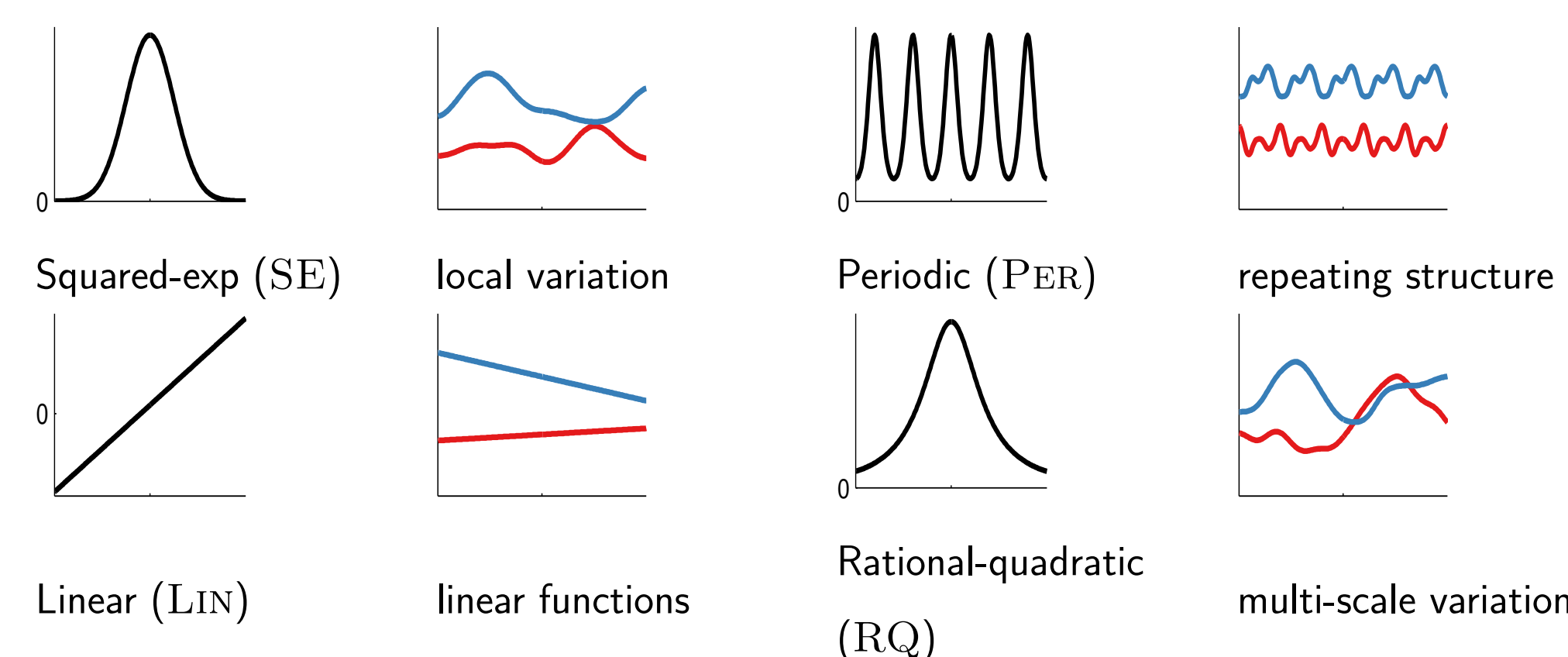
## Data often exhibits high level structure



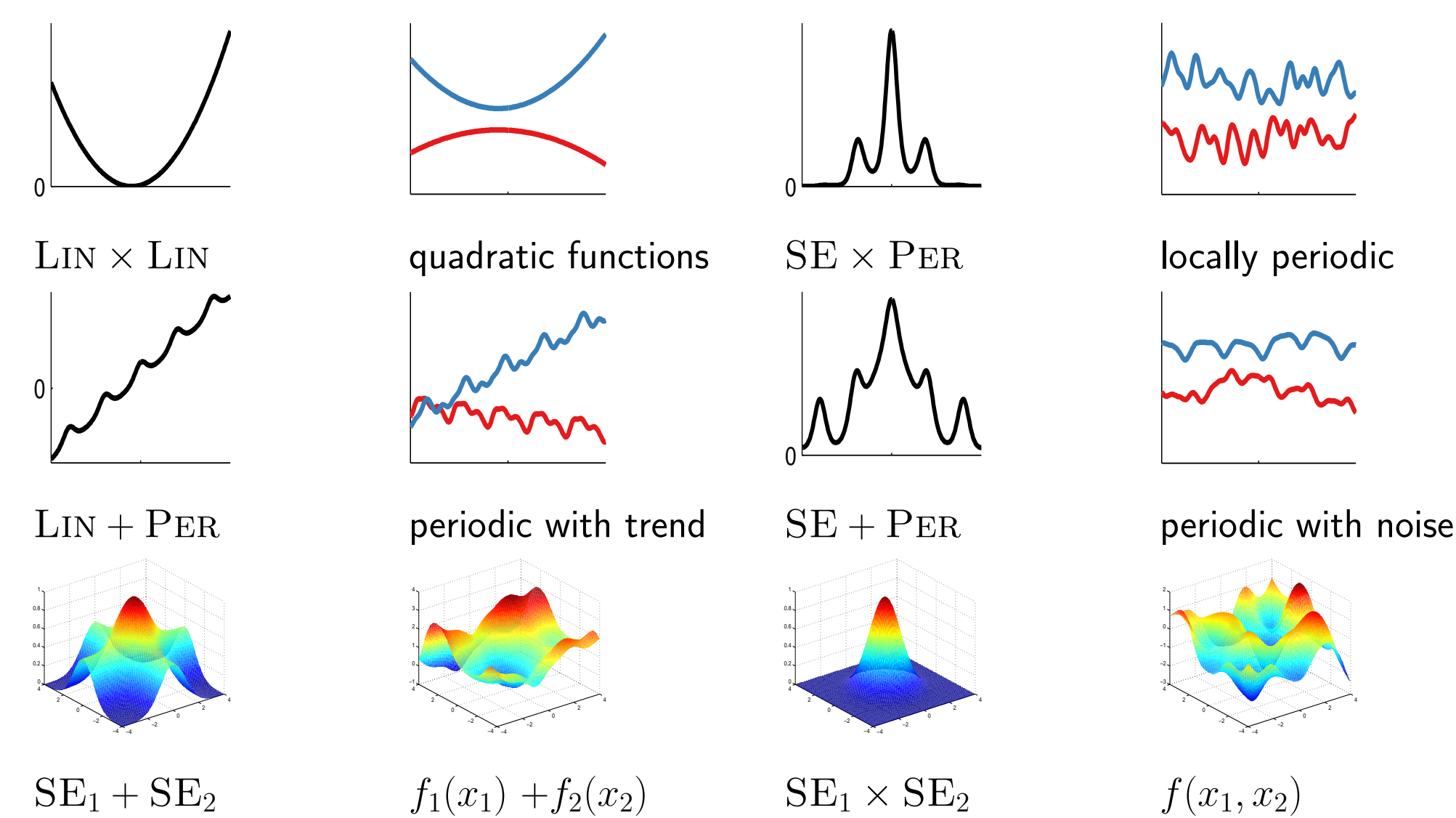
- Traditionally, a researcher / scientist / statistician would select an appropriate model for the type of structures present
- Automatic model selection techniques already exist, typically choosing between a finite or restricted set of models
- Instead, we automate statistical model *construction*, which allows for a very large set of models to be considered

## Gaussian process regression can model many structures with an appropriately chosen kernel

- The kernel encodes the inductive bias of the model i.e. the types of functions the model 'believes in'
- Below we depict standard base kernels, and examples of functions the model believes in (samples from the prior)



- Base kernels can be combined to create more complicated structural assumptions



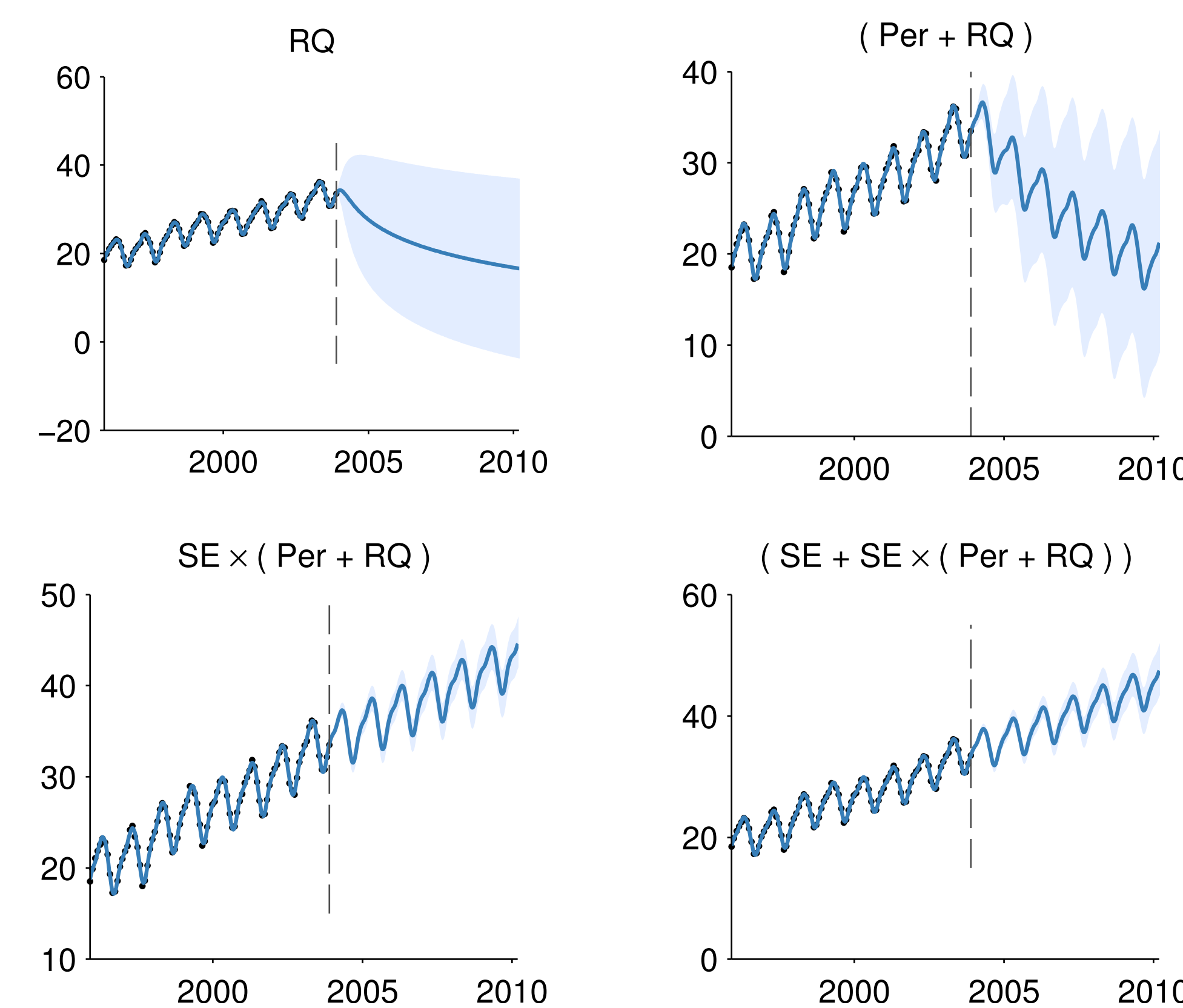
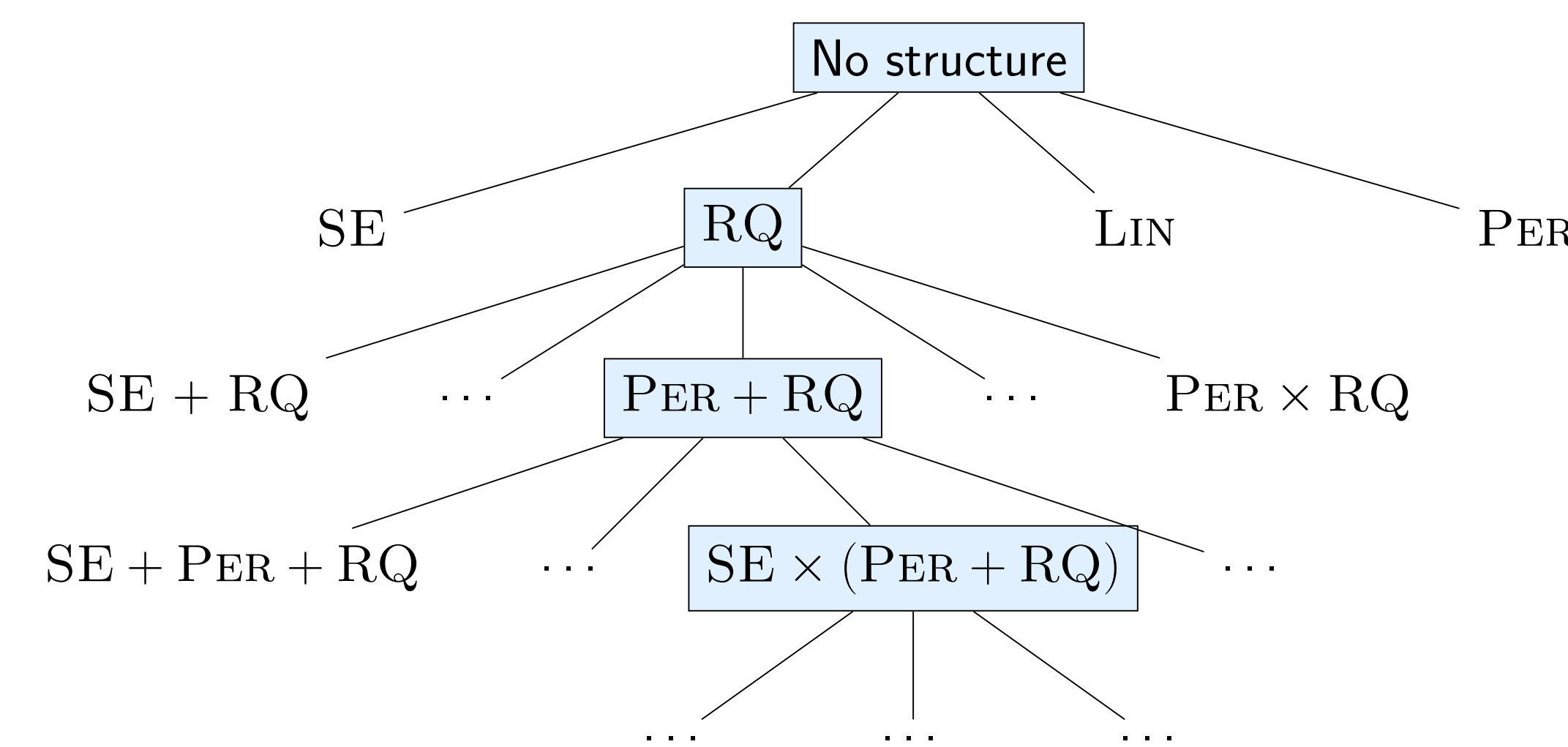
## We consider all kernel expressions derived from a generative grammar...

- Constructing appropriate composite kernels has previously been the domain of Gaussian process experts
- We consider all algebraic expressions involving a small number of base kernels and the operations '+' and 'x', which includes

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

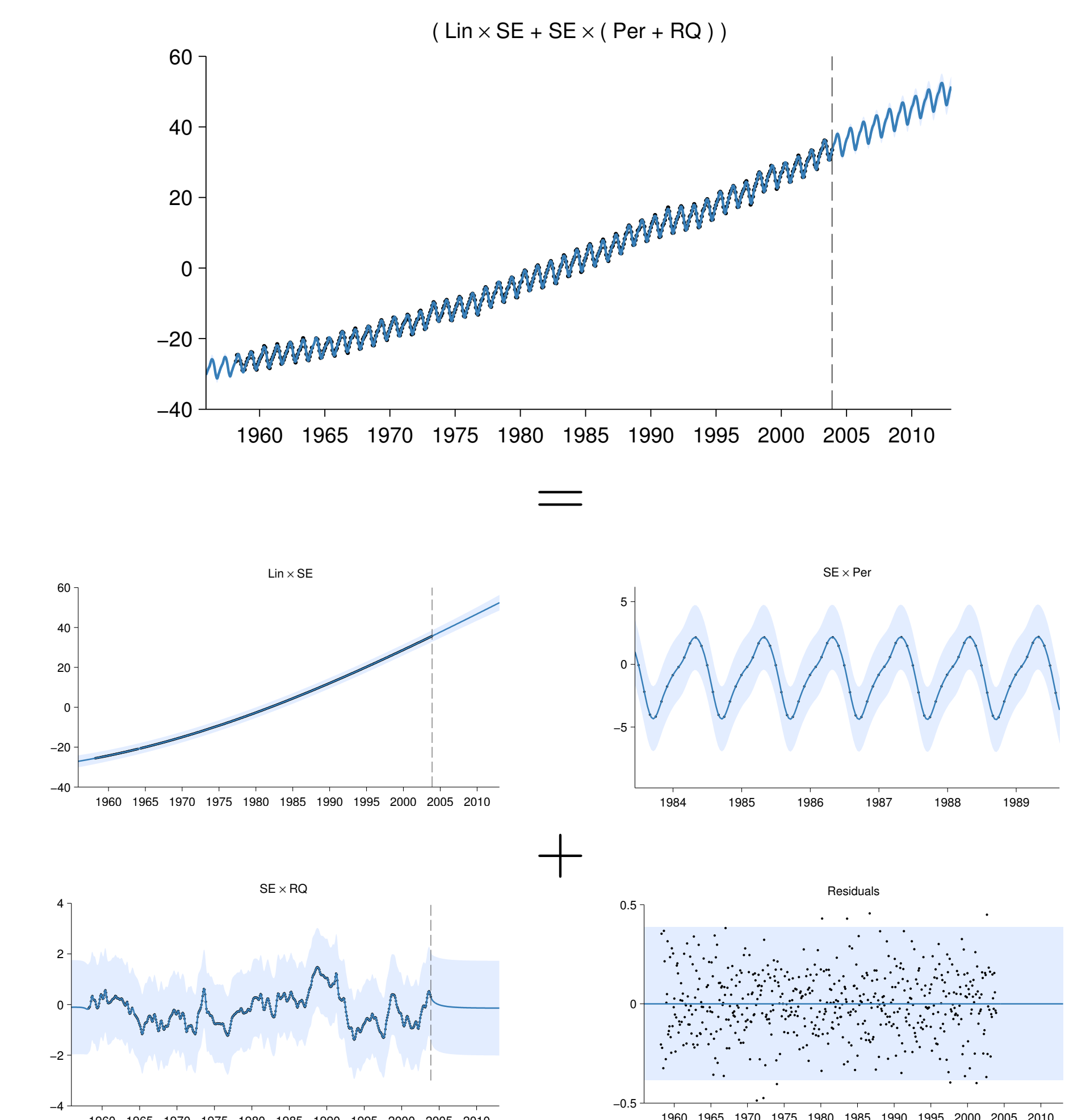
## ... which we search greedily, producing progressively better statistical models

- 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



## Example: Mauna Loa CO<sub>2</sub> concentration

- By automatically inferring an appropriate kernel, we can also automatically decompose functions into additive components



## Example: International airline passengers

