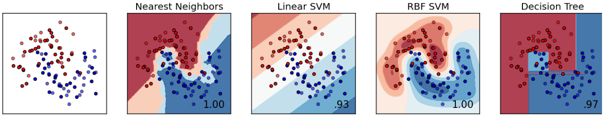
Automatic construction and description of nonparametric models

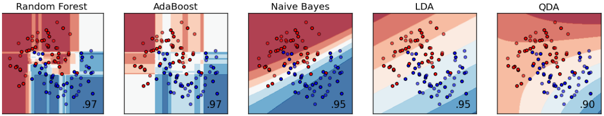


James Robert Lloyd, David Duvenaud, Roger Grosse,
Josh Tenenbaum, Zoubin Ghahramani

November 29, 2013

MOTIVATION

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

  - 

  - Just being nonparametric sometimes isn't good enough
  - Building by hand requires expertise, understanding of the dataset.

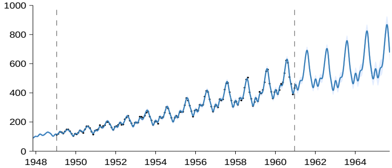
MOTIVATION

- Models today built by hand, or chosen from a fixed set.
    - Building by hand requires expertise, understanding of the dataset.
    - Follows cycle of: propose model, do inference, check model fit
        - Propose new model
        - Do inference
        - Check model fit
    - for high-dimensional data, this can silently fail
- Andrew Gelman asks: How would an AI do statistics?
- It would need a language for describing arbitrarily complicated models, a way to search over those models, nad a way of checking model fit.

MOTIVATION

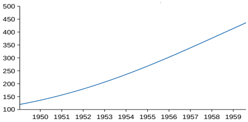
- Andrew Gelman asks: How would an AI do statistics?
- It would need a language for describing arbitrarily complicated models, a way to search over those models, nad a way of checking model fit.
- We built such a language over regression models, a procedure to search over them, and a method to describe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

Example



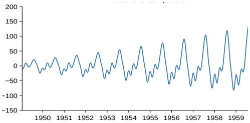
=

entire signal



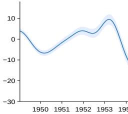
A very smooth monotonically increasing function

+



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

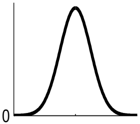
+



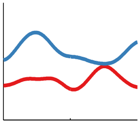
An exactly
tion with a
years but w
creasing an

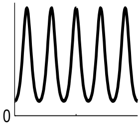
KERNEL CHOICE IS IMPORTANT

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

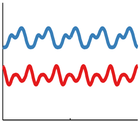


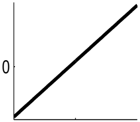
Squared-exp (SE)


local variation

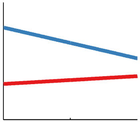


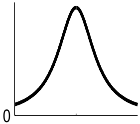
Periodic (PER)


repeating structure

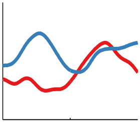


Linear (LIN)


linear functions

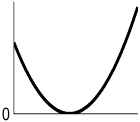


Rational-quadratic(RQ)

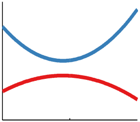

multi-scale variation

Kernels can be composed

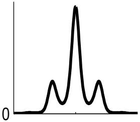
- Two main operations: adding, multiplying



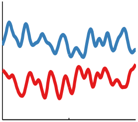
Lin $\times$ Lin



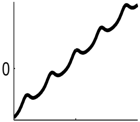
quadratic functions



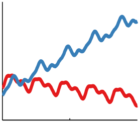
SE $\times$ Per



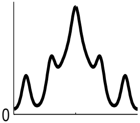
locally periodic



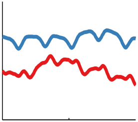
Lin $+$ Per



periodic with trend



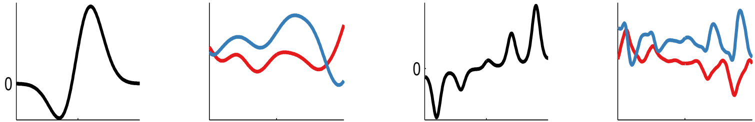
SE $+$ Per



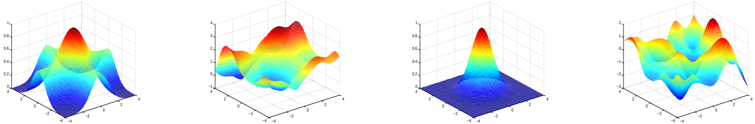
periodic with noise

Kernels can be composed

- Can be composed across multiple dimensions



Lin $\times$ SE | increasing variation | Lin $\times$ Per | growing amplitude



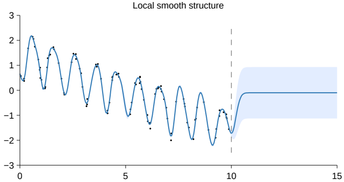
$\mathrm{SE}_1 + \mathrm{SE}_2$ | $f_1(x_1) + f_2(x_2)$ | $\mathrm{SE}_1 \times \mathrm{SE}_2$ | $f(x_1, x_2)$

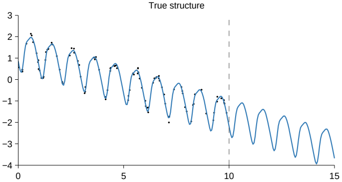
SPECIAL CASES

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

APPROPRIATE KERNELS ARE NECESSARY FOR EXTRAPOLATION

- SE kernel $\rightarrow$ basic smoothing.
- Richer kernels means richer structure can be captured.





KERNELS ARE HARD TO CHOOSE

- Given the diversity of priors available, how to choose one?
- Standard GP software packages include many base kernels and means to combine them, but *no default kernel*
- Software can't choose model for you, you're the expert (?)

Kernels are hard to construct

- Carl devotes 4 pages of his book to constructing a custom kernel for CO2 data
- requires specialized knowledge, trial and error, and a dataset small and low-dimensional enough that a human can interpret it.
- In practice, most users can't or won't make custom kernel, and SE kernel became *de facto* standard kernel through inertia.

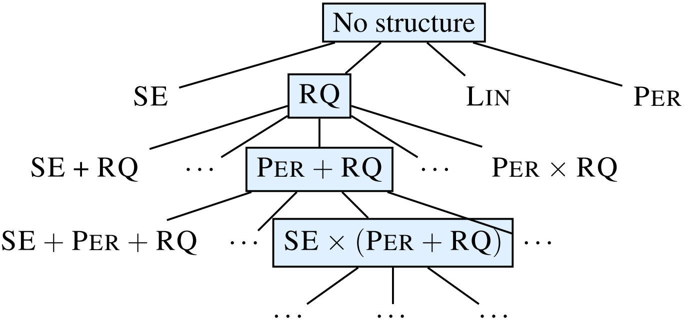
RECAP

- GP Regression is a powerful tool
- Kernel choice allows for rich structure to be captured - different kernels express very different model classes
- Composition generates a rich space of models
- Hard & slow to search by hand
- Can kernel specification be automated?

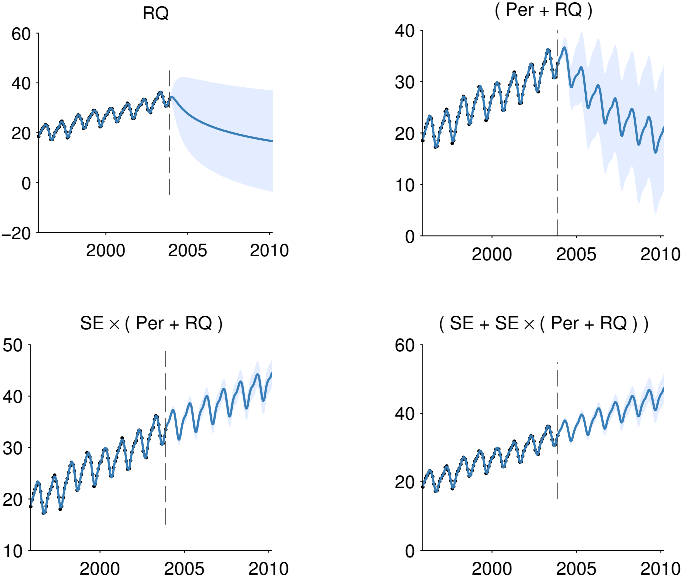
COMPOSITIONAL STRUCTURE SEARCH

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

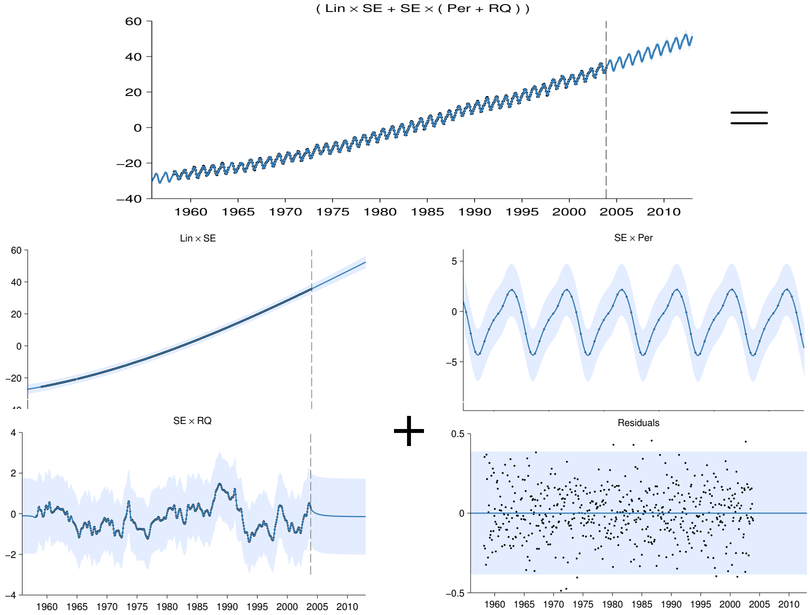
Compositional Structure Search



Example Search: Mauna Lua CO_2



Example Decomposition: Mauna Loa CO_2



Compound kernels are interpretable

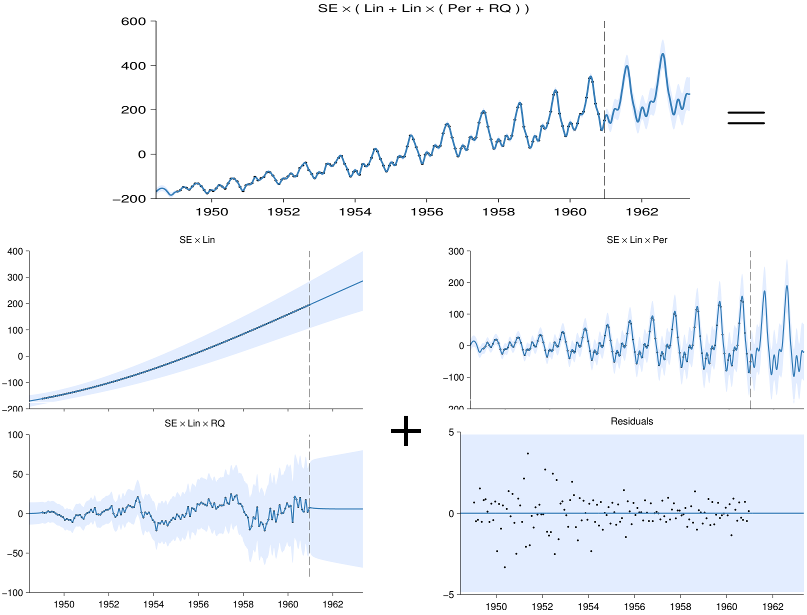
Suppose functions f_1, f_2 are draw from independent GP priors, $f_1 \sim \mathcal{GP}(\mu_1, k_1), f_2 \sim \mathcal{GP}(\mu_2, k_2)$. Then it follows that

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

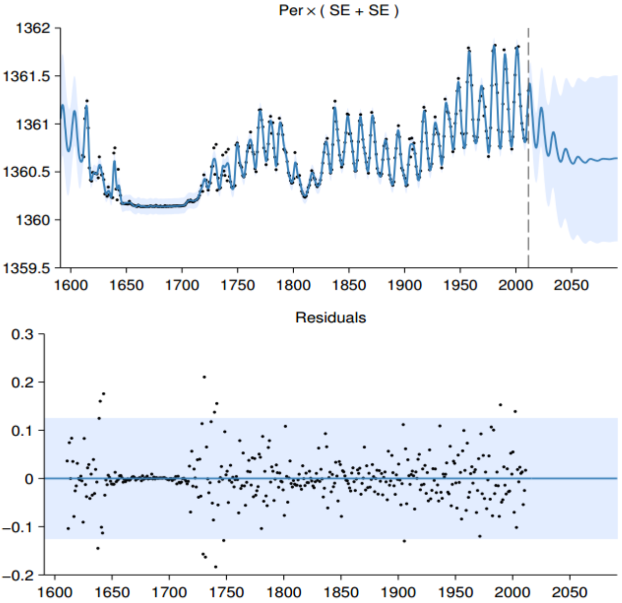
Sum of kernels is equivalent to sum of functions. Distributivity means we can write compound kernels as sums of products of base kernels:

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

EXAMPLE DECOMPOSITION: AIRLINE

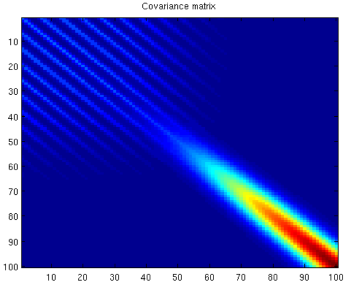


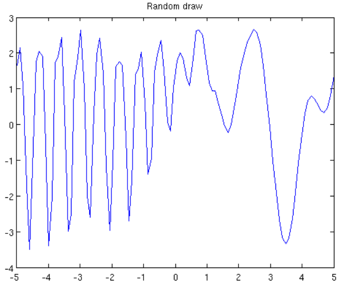
Example: Sunspots



CHANGEPOINT KERNEL

Can express change in covariance:

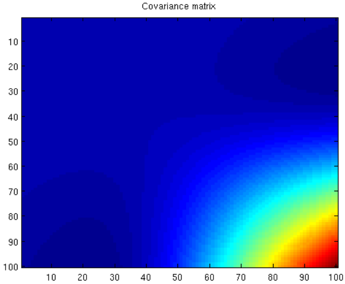


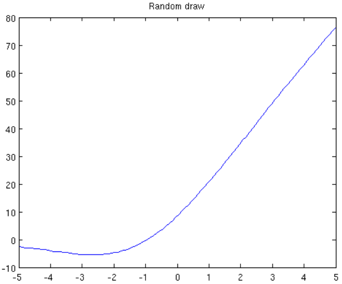


Periodic changing to SE

Changepoint Kernel

Can express change in covariance:





SE changing to linear

SUMMARY

- Choosing form of kernel is currently done by hand.
- Compositions of kernels lead to more interesting priors on functions than typically considered.
- A simple grammar specifies all such compositions, and can be searched over automatically.
- Composite kernels lead to interpretable decompositions.

Conclusions

- Model-building is currently done mostly by hand.
- Grammars over composite structures are a simple way to specify open-ended model classes.
- Composite structures often imply interpretable decompositions of the data.
- Searching over these model classes is a step towards automating statistical analysis.

CONCLUSIONS

- Model-building is currently done mostly by hand.
- Grammars over composite structures are a simple way to specify open-ended model classes.
- Composite structures often imply interpretable decompositions of the data.
- Searching over these model classes is a step towards automating statistical analysis.

Thanks!