Gaussian process regression (2)

ISTA 410 / INFO 510: Bayesian Modeling and Inference

U. of Arizona School of Information

November 9, 2020

Outline

Last time:

- Simple intro to Gaussian process regression

Today:

- Example: fitting a sum of GPs to time series data

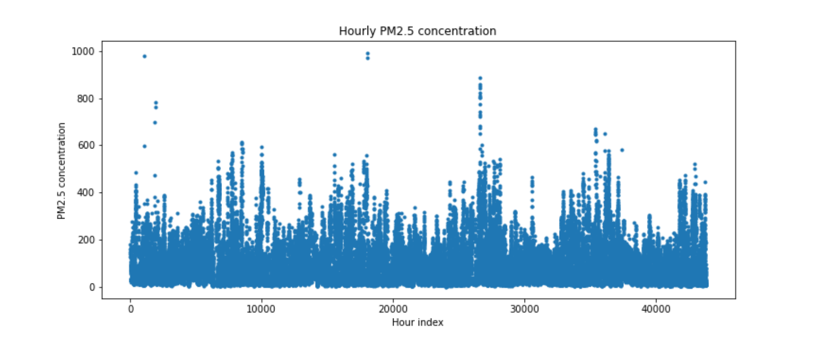
Example

PM2.5 pollution

Data:

- PM2.5 air pollution levels in
- Particulate matter, size ¡ 2.5 micron; produced by combustion processes, hazardous to health
- Data from 01/2010 - 12/2014, measured hourly
- Liang, X., Zou, T., Guo, B., Li, S., Zhang, H., Zhang, S., Huang, H. and Chen, S. X. (2015). "Assessing Beijing's PM2.5 pollution: severity, weather impact, APEC and winter heating." *Proceedings of the Royal Society A*, 471, 20150257.

What does the data look like?



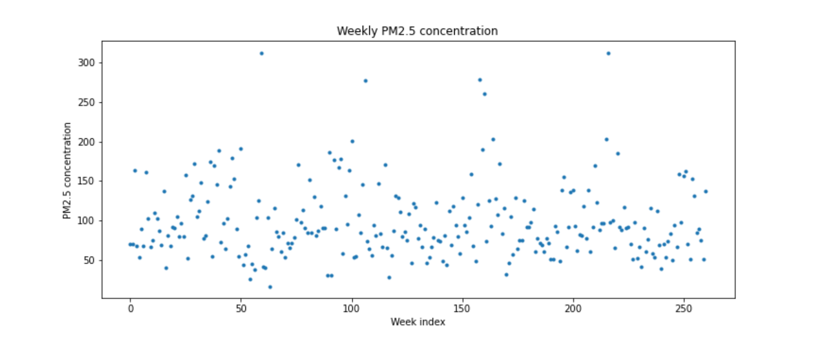
Observations

Observations:

- Very noisy
- Apparent yearly quasi-periodic trend
- Zooming in, daily fluctuations

To reduce computation, we'll average over weeks and focus on longer-term variation.

What does the data look like?



Putting together a model

Additive GP model

We want to decompose the PM2.5 readings into several components:

$$\begin{aligned} y_t = &\ \text{seasonal pattern} \\ &+ \text{short-term fluctuation} \\ &+ \text{noise} \end{aligned}$$

Additive GP model

In other words, our model is

$$
\begin{aligned}
y_i &\sim \mathrm{Normal}(f(t), \sigma) \\
\sigma &\sim \mathrm{HalfCauchy}(1) \\
f(t) &= f_s(t) + f_{fluc}(t) \\
f_s &\sim \mathcal{GP}(0, k_s) \\
f_{fluc} &\sim \mathcal{GP}(0, k_{fluc})
\end{aligned}
$$

Covariance choices

Covariance for the seasonal variation:

$$k(x, x') = \eta_s^2 \exp\left(-\frac{\sin^2(\pi|x - x'|/T)}{2\ell_p^2}\right) \exp\left(-\frac{(x - x')^2}{2\ell_s^2}\right)$$

overall scale $\times$ periodic correlations $\times$ trend

Allows periodic behavior to vary over time.

Covariance choices

Covariance for the short-term fluctuations:

$$k(x, x') = \text{Matèrn}_{5/2}(x, x')$$

The Matèrn family adds a *smoothness* parameter, here $\nu = 5/2$.

- $\nu = 1/2$ recovers the absolute exponential covariance (continuity but no smoothness)
- $\nu \to \infty$ recovers the squared exponential (infinite smoothness)

What the Matèrn covariance looks like

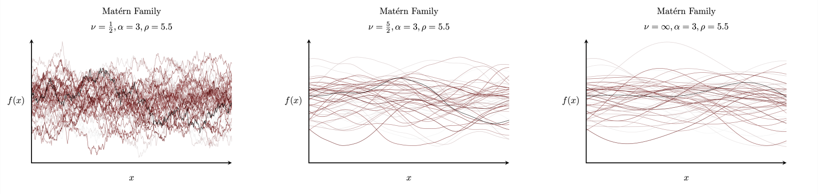
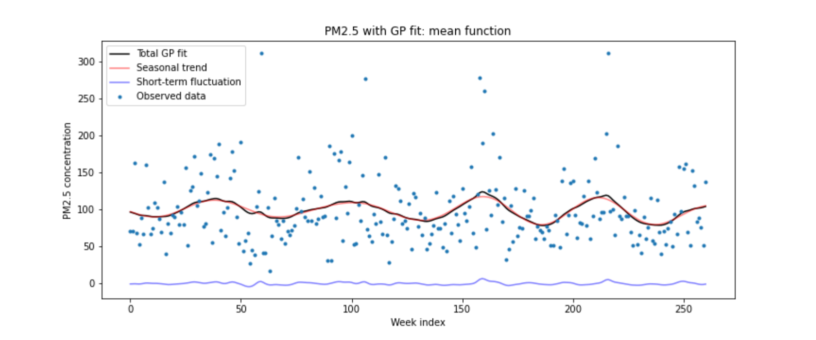
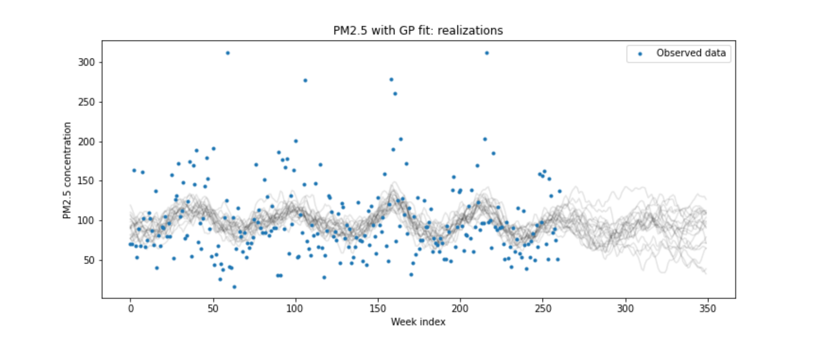


Image credit: Michael Betancourt
`https://betanalpha.github.io/assets/case_studies/gaussian_processes.html`

Estimate of underlying mean function



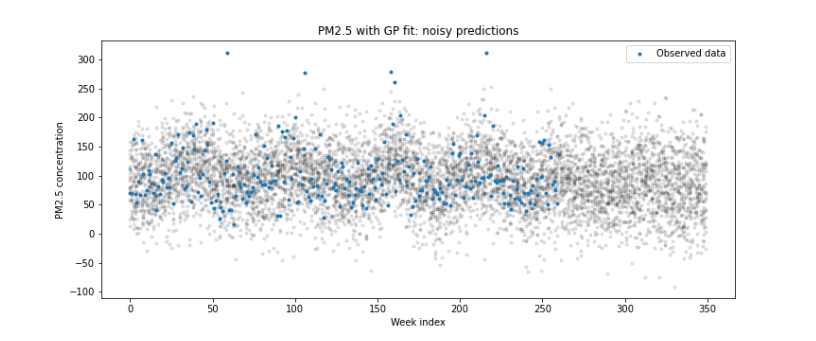
Sample realizations of GP



Realizations can be projected out to the future. Notice these are still realizations of the underlying function $f(t)$, not including the normal noise.

What happens if we re-introduce the noise?

Sample realizations of GP (with noise)



What's wrong with this picture?

Trying to improve the model

GP with non-Gaussian noise

Our model is, recall:

$$
\begin{aligned}
y_i &\sim \mathrm{Normal}(f(t), \sigma) \\
f(t) &= f_s(t) + f_{fluc}(t) \\
f_s &\sim \mathcal{GP}(0, k_s) \\
f_{fluc} &\sim \mathcal{GP}(0, k_{fluc})
\end{aligned}
$$

Problem: noise is symmetric – can't reach the extreme high points without dipping below 0 on the other end

GP with non-Gaussian noise

Substitute model:

$$\begin{aligned} y_i &\sim \mathrm{SkewNormal}(f(t), \sigma, \alpha) \\ f(t) &= f_s(t) + f_{fluc}(t) \\ f_s &\sim \mathcal{GP}(0, k_s) \\ f_{fluc} &\sim \mathcal{GP}(0, k_{fluc}) \\ \sigma &\sim \mathrm{HalfCauchy}(1) \\ \alpha &\sim \mathrm{Normal}(0, 2) \end{aligned}$$

Computation a bit more complicated – no longer conjugate

PyMC3 `Latent` implementation

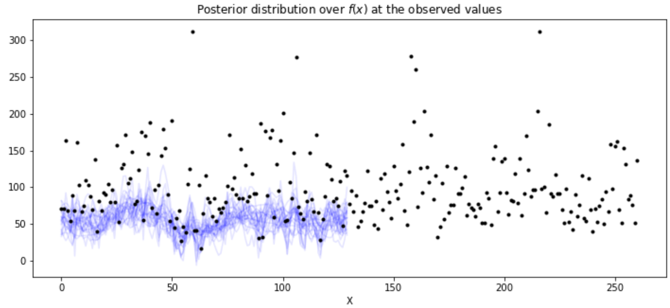
The `Latent` class in PyMC3 can be used to do this: it implements the function $f(t)$ as a multivariate normal at the given grid points.

Then we can use this value as a location parameter for our SkewNormal likelihood.

Let's go over to the code...

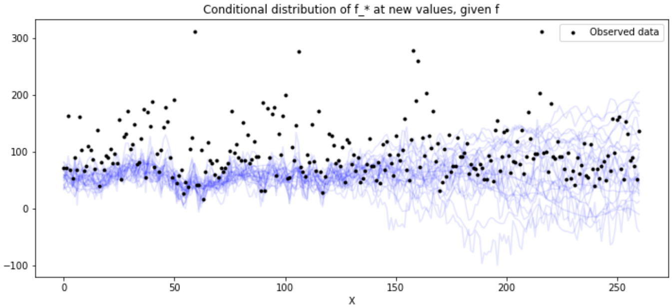
Result

Several realizations from the trace *at training points*:



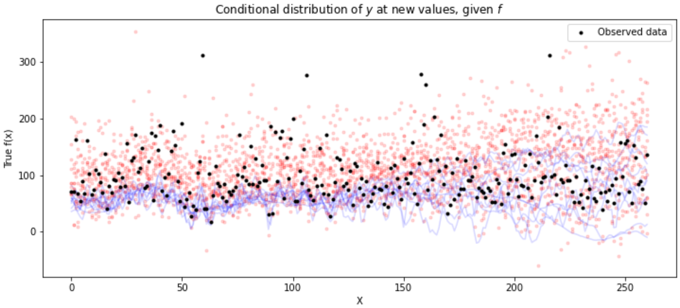
Result

Several realizations from the trace *at training points*:



Result with noise

Adding in skew-normal noise:



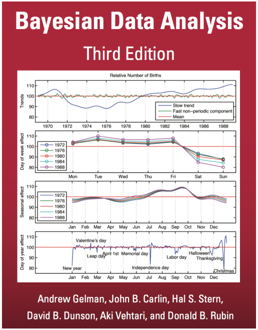
What could we do better?

Our model with skew-normal errors is a little better at avoiding absurd predictions, but not perfect:

- Noise model still probably doing too much
- In addition to occasional negative predictions, also too much noise in the other direction
- Possible improvement: add another additive component to handle whatever is happening around the New Year

Further examples

From BDA:



Mauna Loa CO_2:

`https://docs.pymc.io/notebooks/GP-MaunaLoa.html`

Summary

Today:

- Additive Gaussian process models can be used to decompose processes into several components for modeling time series data

Next week:

- General probabilistic graphical models

Further reading (not specifically GP related):

- "Bayesian workflow": `https://arxiv.org/abs/2011.01808`