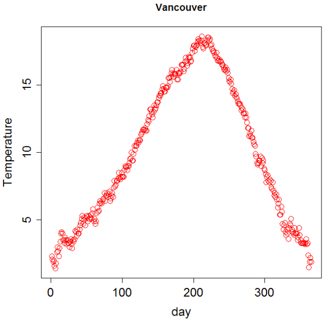
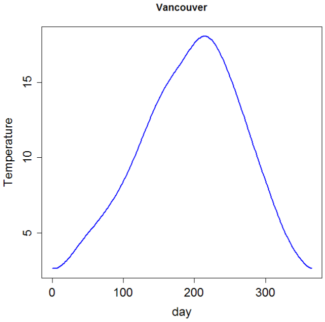
From Data To Functions

How do we go from

data to functions?





Basis Expansions

From multiple linear regression:

$$y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + \cdots + \epsilon_i$$

Or if there is curvature:

$$y_i = \beta_0 + x_i\beta_1 + x_i^2\beta_2 + x_i^3\beta_3 + \cdots + \epsilon_i$$

More generally

$$y_i = \sum_{j=1}^{K} c_j\phi_j(t_i) + \epsilon_i = f(t_i) + \epsilon_i$$

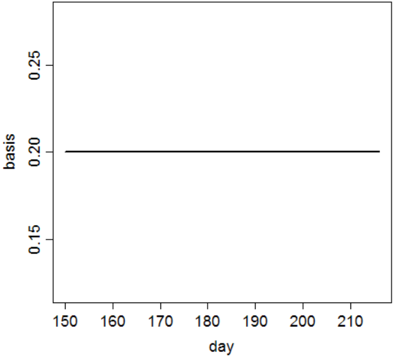
Which we write as being

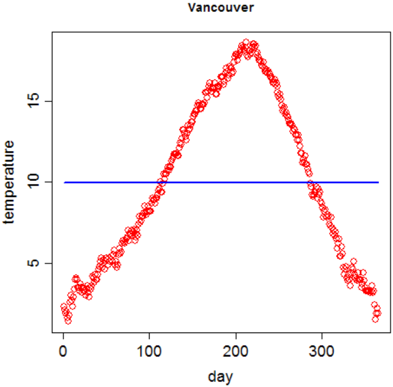
$$f(t) = \mathbf{c}^T\Phi(t)$$

And we say $\Phi(t)$ is a *basis system* for f.

The Monomial Basis

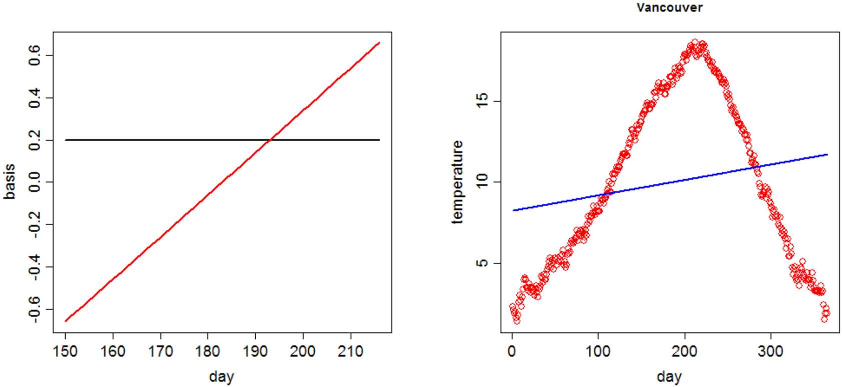
$$\Phi(t) = (1)$$





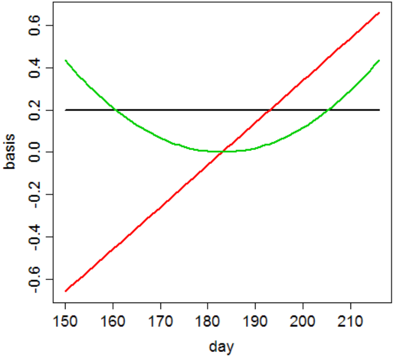
The Monomial Basis

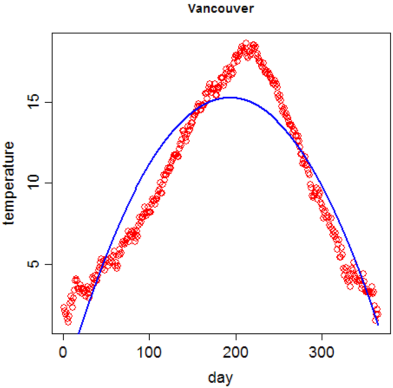
$$\Phi(t) = (1, t)$$



The Monomial Basis

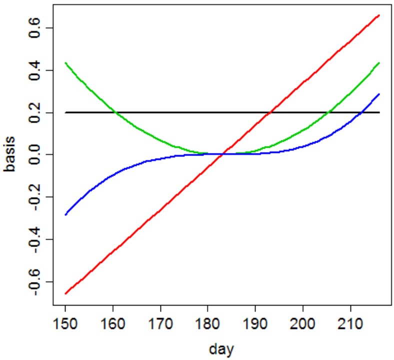
$$\Phi(t) = (1, t, t^2)$$

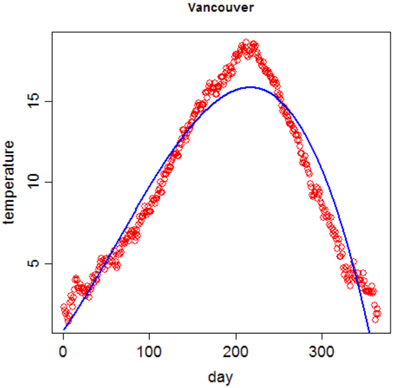




The Monomial Basis

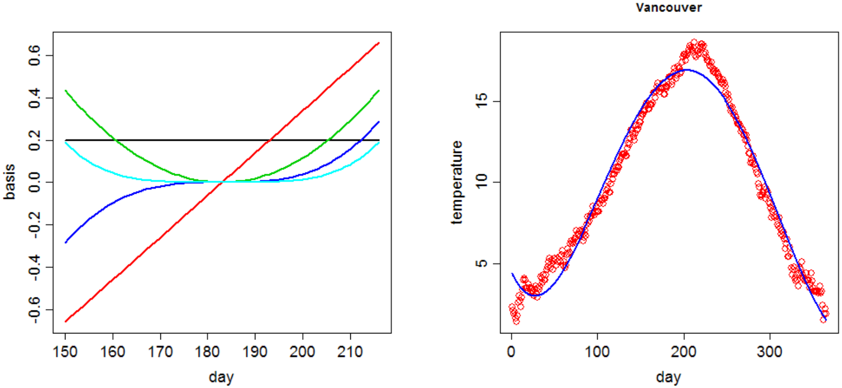
$$\Phi(t) = (1, t, t^2, t^3)$$





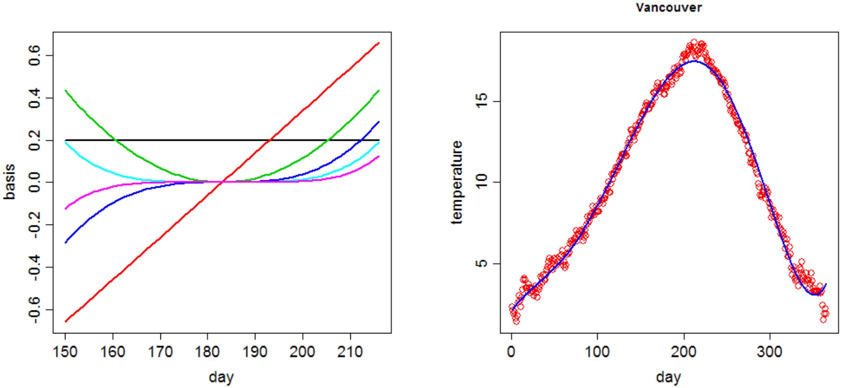
The Monomial Basis

$$\Phi(t) = (1, t, t^2, t^3, t^4)$$



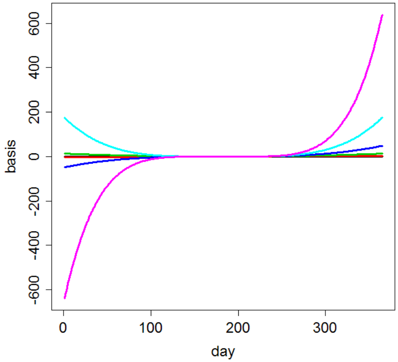
The Monomial Basis

$$\Phi(t) = (1, t, t^2, t^3, t^4, t^5)$$



Problems with the Monomial Basis

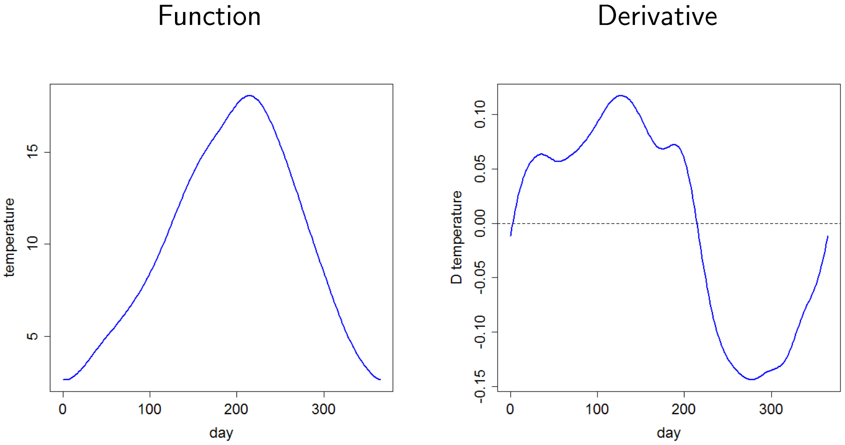
Numerically difficult for more than six terms



Larger terms over-run smaller ones; especially with unevenly-spaced observations.

Problems with the Monomial Basis

We are often interested in *rates of change*



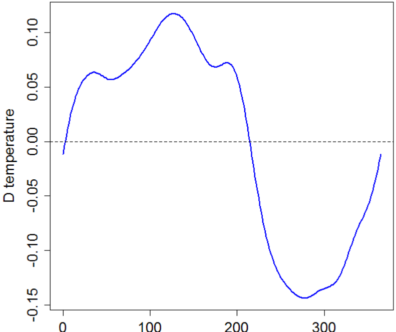
Problems with the Monomial Basis

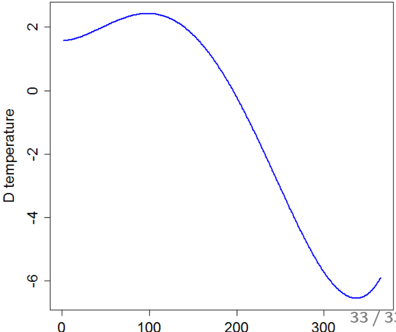
But monomial derivatives get simpler:

$$f(t) = \sum_{k=0}^{K} c_k t^k, \; Df(t) = \sum_{k=1}^{K-1} c_k k t^{k-1}$$

Derivative

Estimate

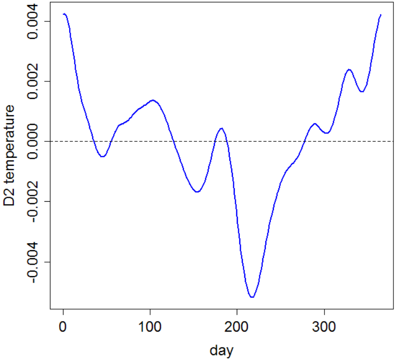




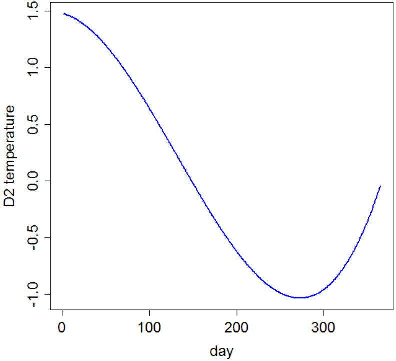
Problems with the Monomial Basis

Whereas the opposite happens in most real-world data:

Second Derivative



Estimate



The Fourier Basis

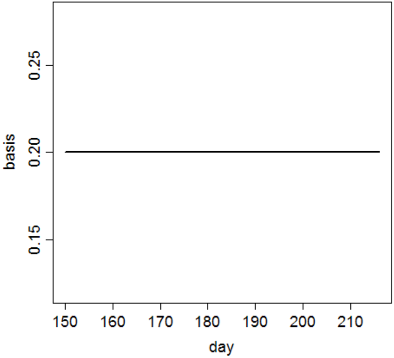
- basis functions are sine and cosine functions of increasing frequency:

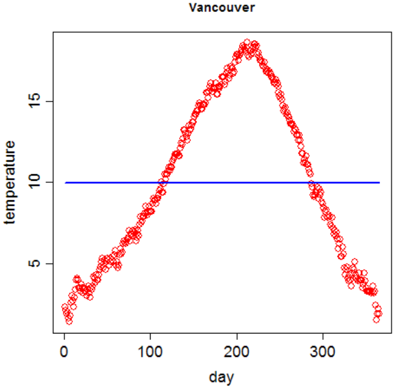
$$1, sin(\omega t), cos(\omega t), sin(2\omega t), cos(2\omega t), \ldots$$
$$sin(m\omega t), cos(m\omega t), \ldots$$

- constant ω defines the period of oscillation of the first sine/cosine pair. This is $\omega = 2\pi/P$ where P is the period.
- $K = 2M + 1$ where M is the largest number of oscillations required in a period of length P.

The Fourier Basis

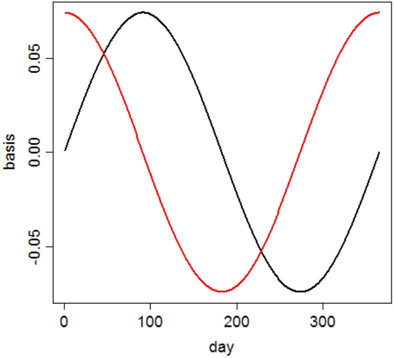
$$\Phi(t) = (1)$$

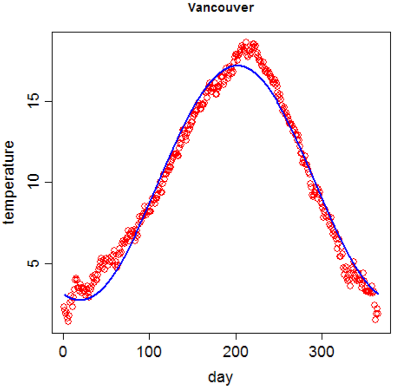




The Fourier Basis

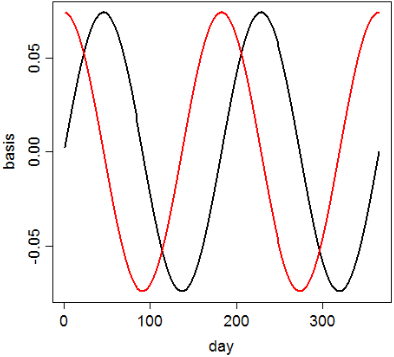
$$\Phi(t) = (1, sin(\omega t), cos(\omega t))$$

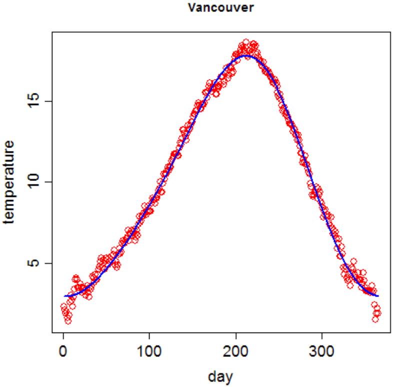




The Fourier Basis

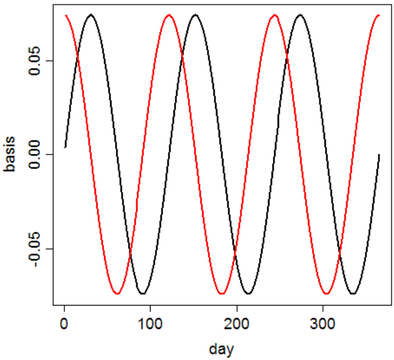
$$\Phi(t) = (1, sin(\omega t), cos(\omega t), sin(2\omega t), cos(2\omega t))$$

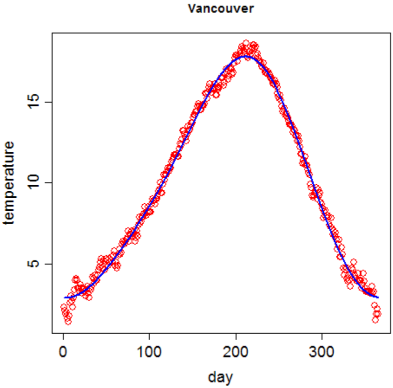




The Fourier Basis

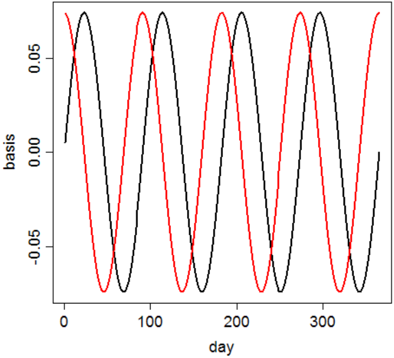
$$\Phi(t) = (1, sin(\omega t), cos(\omega t), sin(2\omega t), cos(2\omega t), sin(3\omega t), cos(3\omega t))$$

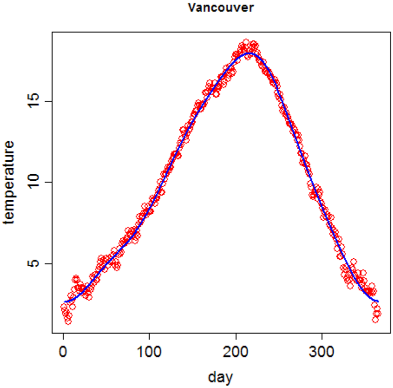




The Fourier Basis

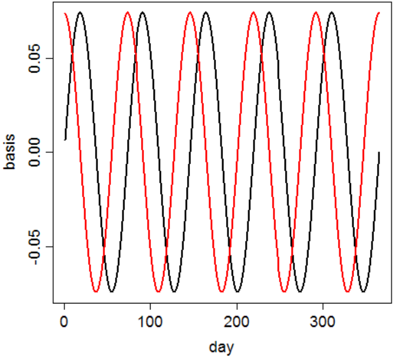
$$\Phi(t) = (1, sin(\omega t), cos(\omega t), \ldots, sin(4\omega t), cos(4\omega t))$$

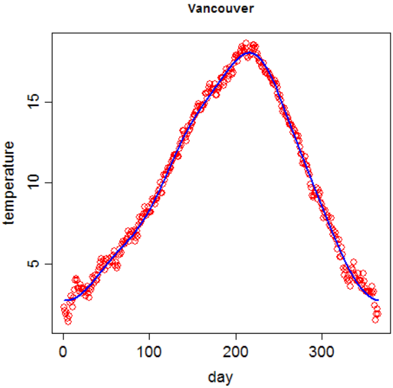




The Fourier Basis

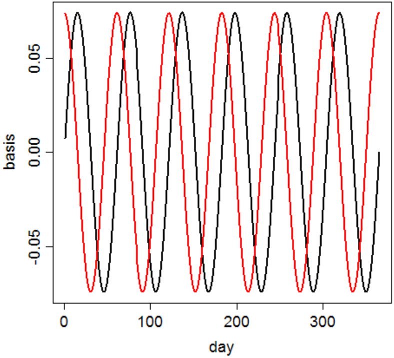
$$\Phi(t) = (1, sin(\omega t), cos(\omega t), \ldots, sin(5\omega t), cos(5\omega t))$$

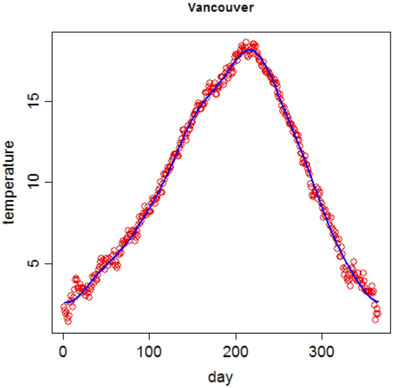




The Fourier Basis

$$\Phi(t) = (1, sin(\omega t), cos(\omega t), \ldots, sin(6\omega t), cos(6\omega t))$$





Advantages of Fourier Bases

- Only alternative to monomial bases until the middle of the 20th century
- Excellent computational properties, especially if the observations are equally spaced.
- Natural for describing periodic data, such as the annual weather cycle

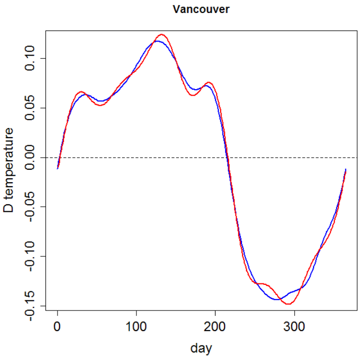
BUT functions are periodic; this can be a problem if the data are, for example, growth curves.
Fourier basis is still the first choice in many fields, such as signal analysis, even when the data are not periodic.

Fourier Derivatives

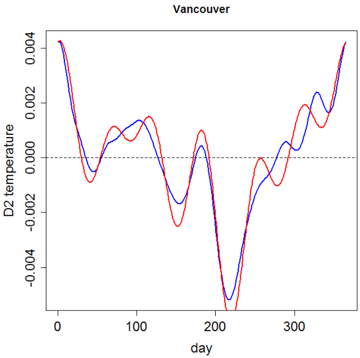
$$D\sin(\omega t) = -\omega\cos(\omega t), \;\; D\cos(\omega t) = -\omega\sin(\omega t)$$

So derivatives retain complexity, easy to compute

D



D^2

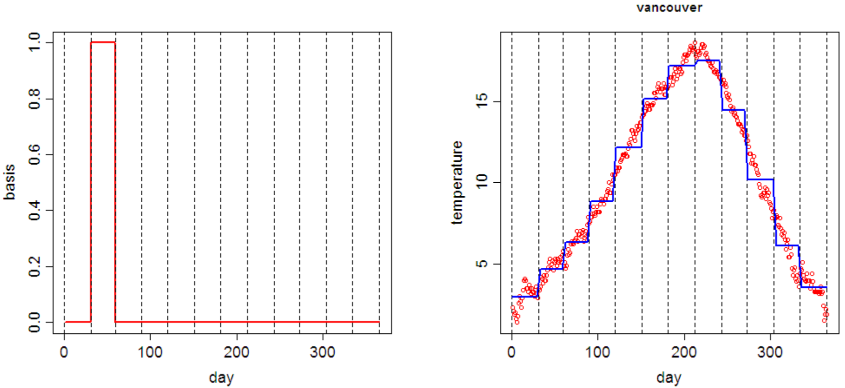


Splines

- Splines are polynomial segments joined end-to-end
- Segments are constrained to be smooth at the join
- The points at which the segments join are called *knots*
- The order m (order = degree+1) of the polynomial segments and
- the location of the knots define the system.
- **Bsplines** are a particularly useful means of incorporating the constraints.

Splines

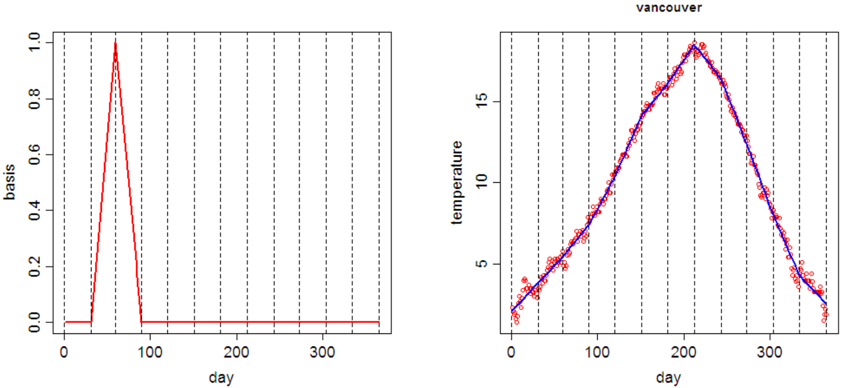
Vancouver temperature with knots at months.
Splines of order 1



Splines

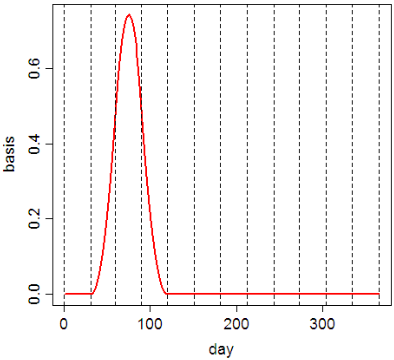
Vancouver temperature with knots at months.

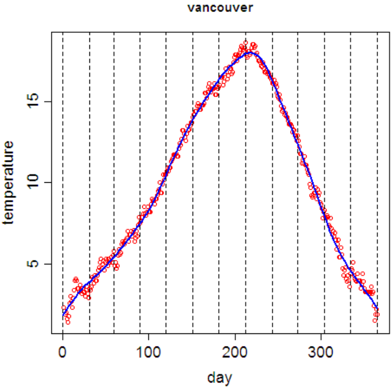
Splines of order 2



Splines

Vancouver temperature with knots at months.
Splines of order 3

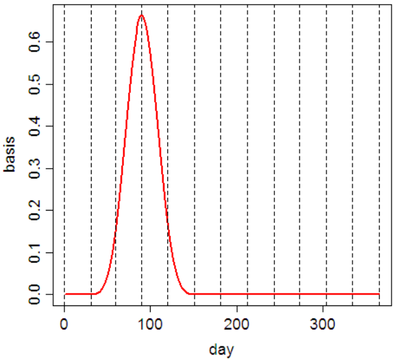


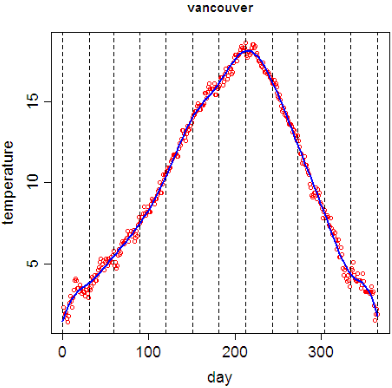


Splines

Vancouver temperature with knots at months.

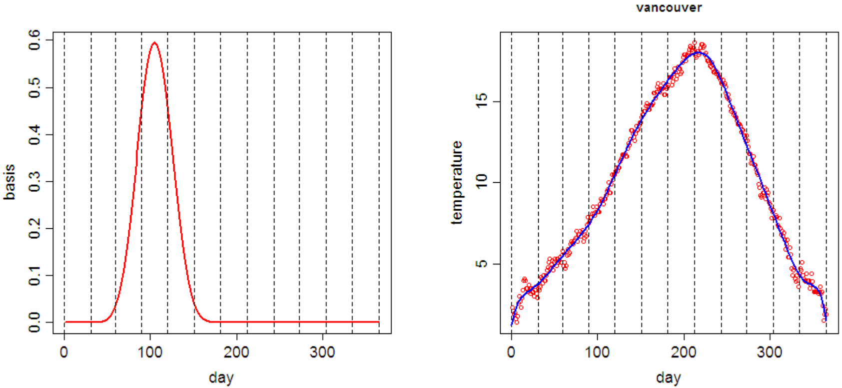
Splines of order 4





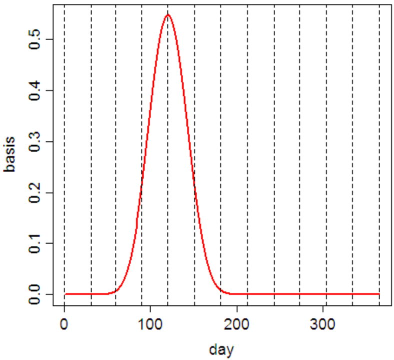
Splines

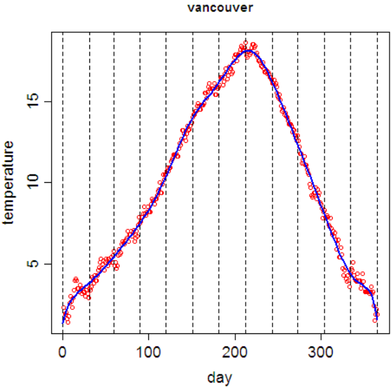
Vancouver temperature with knots at months.
Splines of order 5



Splines

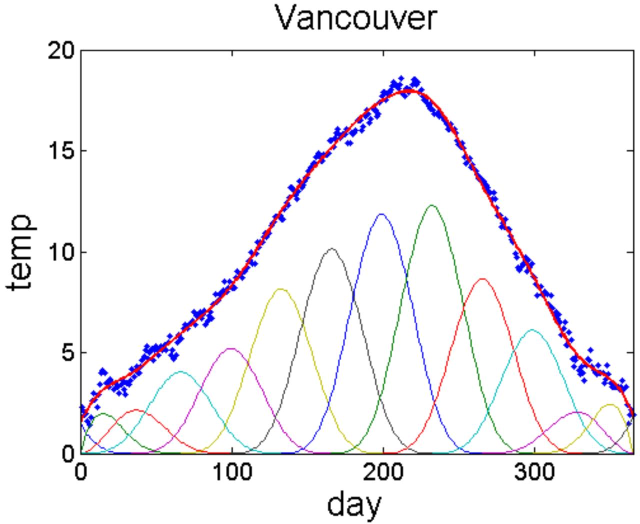
Vancouver temperature with knots at months.
Splines of order 6





Example

An illustration of basis expansions for *local* basis functions



Properties of B-splines

- Number of basis functions:

$$\textit{order} + \textit{number interior knots}$$

- Derivatives up to $m-2$ are continuous.
- B-spline basis functions are positive over at most m adjacent intervals $\rightarrow$ fast computation for even thousands of basis functions.
- Sum of all B-splines in a basis is always 1; can fit any polynomial of order m.
- Most popular choice is order 4, implying continuous second derivatives. Second derivatives have straight-line segments.

Bsplines: Choosing Knots and Order

- The order of the spline should be at least $k+2$ if you are interested in k derivatives.
- Knots are often equally spaced (a useful default)
- But there are two important rules:
  - Place more knots where you know there is strong curvature, and fewer where the function changes slowly.
  - Be sure there is at least one data point in every interval.
- Later, we'll discuss placing a knot at each point of observation.
- Co-incident knots reduce the number of continuous derivatives at each point. This can be useful (more later).

Other Bases

The `fda` library in `R` also allows the following bases:

Constant $\phi(t) = 1$, the simplest of all.

Power $t^{\lambda_1}, t^{\lambda_2}, t^{\lambda_3}, \ldots$, powers are distinct but not necessarily integers or positive.

Exponential $e^{\lambda_1 t}, e^{\lambda_2 t}, e^{\lambda_3 t}, \ldots$

Other possible bases include

Wavelets especially for sharp, local features

Empirical we will investigate functional Principal Components

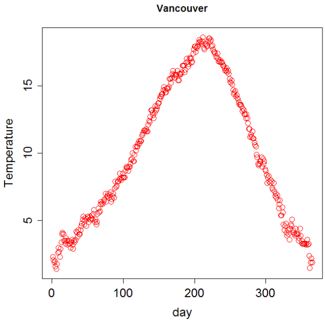
Designer for example on dynamic models: tailoring a basis to data (if you know something about the data) can be much more efficient.

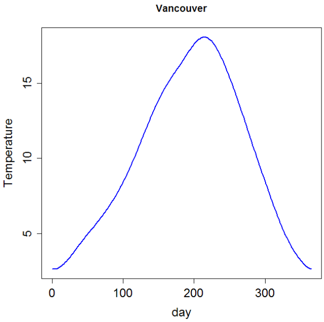
Summary

1. Basis expansions: just like adding different independent variables in linear regression
2. Monomial basis: direct extension of adding interaction and quadratic terms. Poor numerics, bad for derivatives.
3. Fourier basis: classical, common in signal processing etc. Great for periodic functions. Must be infinitely differentiable.
4. B-spline basis: locally polynomial. Allows control of smoothness and accuracy. Local definition $\Rightarrow$ good numerics.
5. Other basis systems also exist.
6. What is best depends on the data.

Fitting and smoothing

data to functions?





Least-Squares

Assume we have observations for a single curve

$$y_i = x(t_i) + \epsilon$$

and we want to estimate

$$x(t) \approx \sum_{j=1}^{j} c_j \phi_j(t)$$

Minimize the sum of squared errors:

$$SSE = \sum_{i=1}^{n} (y_i - x(t_i))^2 = \sum_{i=1}^{n} (y_i - \mathbf{c}^T \Phi(t_i))^2$$

This is just linear regression!

Linear Regression on Basis Functions

- If the N by K matrix $\mathbf{\Phi}$ contains the values $\phi_k(t_j)$, and $\mathbf{y}$ is the vector $(y_1, \ldots, y_N)$, we can write

$$SSE(\mathbf{c}) = (\mathbf{y} - \mathbf{\Phi c})^T (\mathbf{y} - \mathbf{\Phi c})$$

- The error sum of squares is minimized by the *ordinary least squares estimate*

$$\hat{\mathbf{c}} = \left(\mathbf{\Phi}^T \mathbf{\Phi}\right)^{-1} \mathbf{\Phi}^T \mathbf{y}$$

- Then we have the estimate

$$\hat{y}(t) = \Phi(t)\hat{\mathbf{c}} = \Phi(t) \left(\mathbf{\Phi}^T \mathbf{\Phi}\right)^{-1} \mathbf{\Phi}^T \mathbf{y}$$

The Standard Model for Residual Distribution

- least squares is optimal for residuals that are independently and identically normal with mean 0 and variance σ.
- That is

$$E\mathbf{y} = \mathbf{\Phi c} \text{ and } \mathrm{Var}\left[\mathbf{y}\right] = \sigma^2 \mathbf{I}$$

- Call this the *standard model* for the distribution of residuals.

Weighted Least Squares

The standard model is often overly simplistic

- $\text{Var}\,[\mathbf{y}]$ may vary with observation time
- The residuals may be correlated.

The first of these can be compensated for by weighting the observations

$$WMSE\,[x] = \sum w_i (y_i - x(t_i))^2$$

Set W to have w_i on the diagonal, we get

$$\hat{x}(t) = \Phi(t)\hat{\mathbf{c}} = \Phi(t)\left(\mathbf{\Phi}^T W \mathbf{\Phi}\right)^{-1} \mathbf{\Phi}^T W \mathbf{y}$$

When we look at the values of $\hat{x}$ at the observation points we have

$$\hat{\mathbf{y}} = \mathbf{\Phi}\left(\mathbf{\Phi}^T W \mathbf{\Phi}\right)^{-1} \mathbf{\Phi}^T W \mathbf{y} = S\mathbf{y}$$

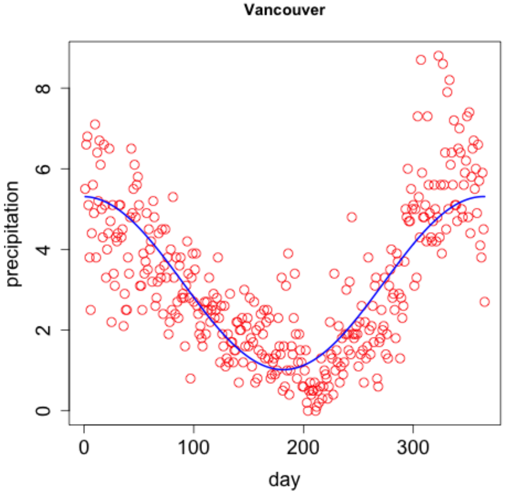
S is referred to as the *smoothing matrix*.

Choosing the Number of Basis Functions

- Small numbers of basis functions mean little flexibility
- Larger numbers of basis functions add flexibility, but may "overfit"
- For Monomial and Fourier bases, just add functions to the collection.
- Spline bases: adding knots or increasing the order changes the basis; but makes it more flexible.
- Spline bases: *changing* the knots may not help even if you add more of them; but this is unusual.

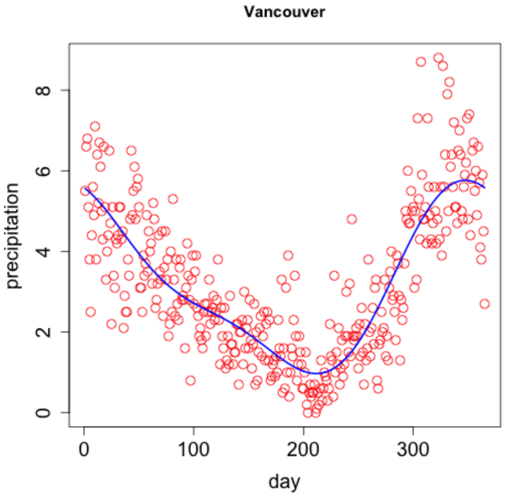
Choosing the Number of Basis Functions

Vancouver Precipitation: 3 Fourier Bases



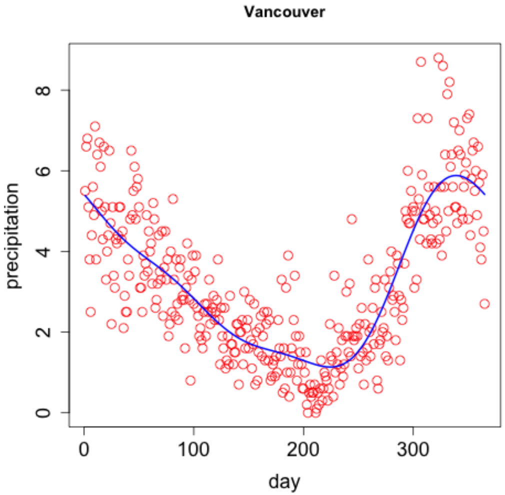
Choosing the Number of Basis Functions

Vancouver Precipitation: 5 Fourier Bases



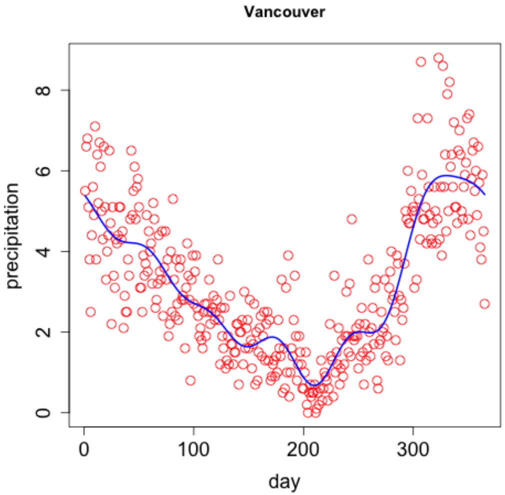
Choosing the Number of Basis Functions

Vancouver Precipitation: 7 Fourier Bases



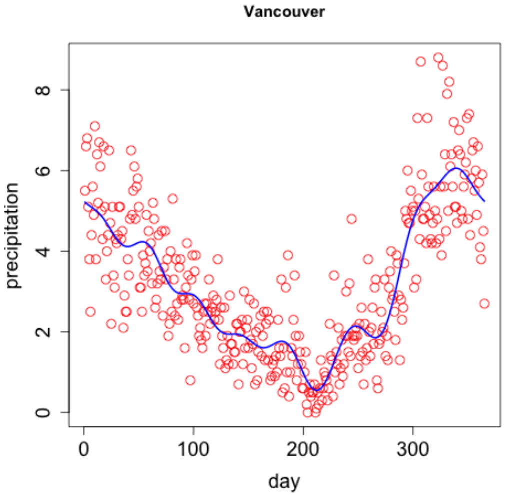
Choosing the Number of Basis Functions

Vancouver Precipitation: 13 Fourier Bases



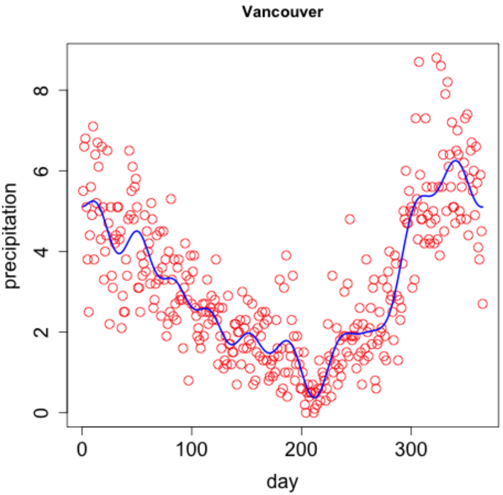
Choosing the Number of Basis Functions

Vancouver Precipitation: 19 Fourier Bases



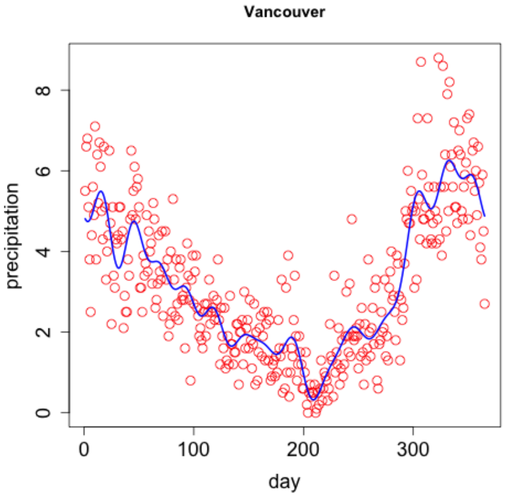
Choosing the Number of Basis Functions

Vancouver Precipitation: 25 Fourier Bases



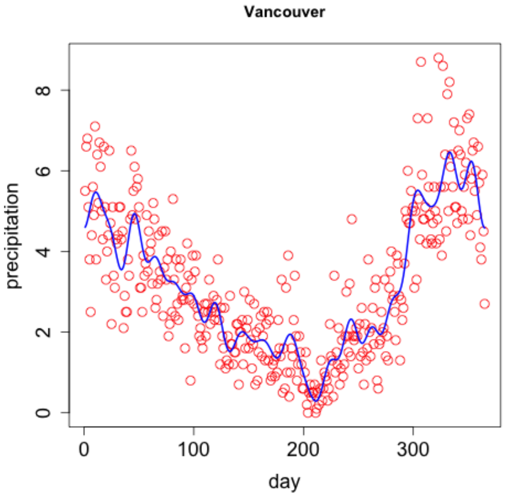
Choosing the Number of Basis Functions

Vancouver Precipitation: 31 Fourier Bases



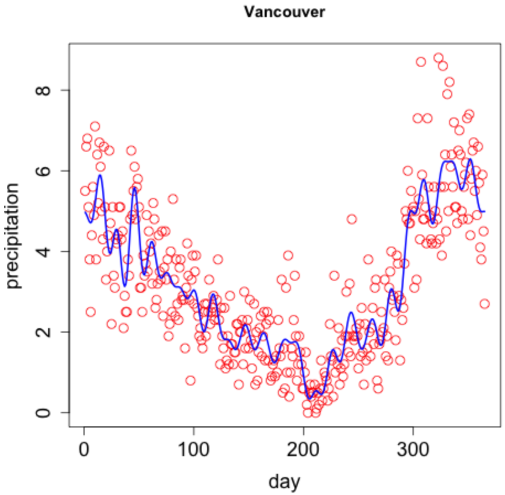
Choosing the Number of Basis Functions

Vancouver Precipitation: 41 Fourier Bases



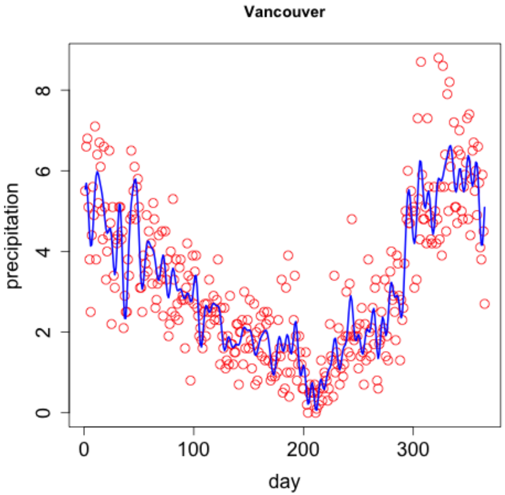
Choosing the Number of Basis Functions

Vancouver Precipitation: 53 Fourier Bases



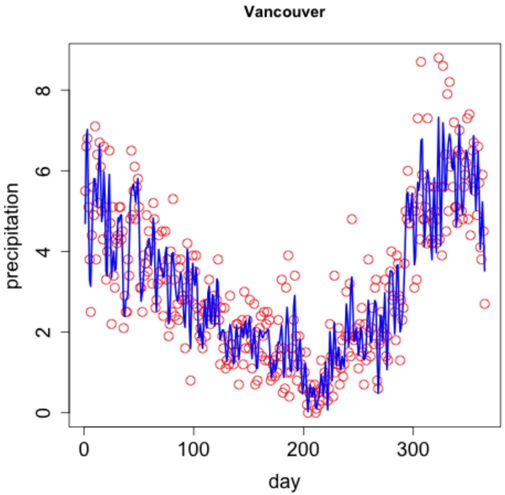
Choosing the Number of Basis Functions

Vancouver Precipitation: 105 Fourier Bases



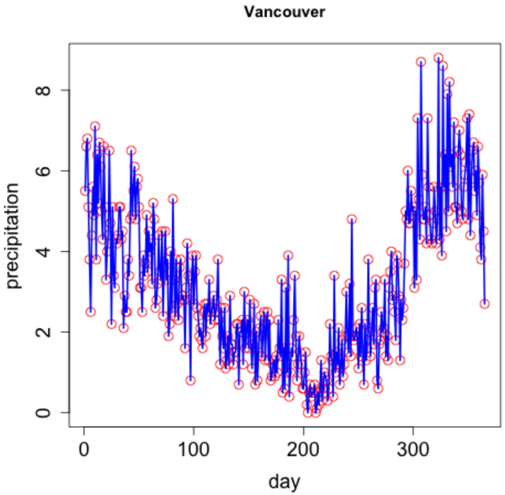
Choosing the Number of Basis Functions

Vancouver Precipitation: 207 Fourier Bases



Choosing the Number of Basis Functions

Vancouver Precipitation: 365 Fourier Bases



Choosing the Number of Basis Functions

Trade off:

- Too many basis functions over-fits the data and reflect errors of measurement
- Too few basis functions fails to capture interesting features of the curves.

Bias and Variance Tradeoff

- Express this trade-off in terms of
  - the *bias* of the estimate of $x(t)$:

$$\text{Bias}\left[\hat{x}(t)\right] = x(t) - E\hat{x}(t)$$

  - the *sampling variance* of the estimate

$$\text{Var}\left[\hat{x}(t)\right] = E\left[\{\hat{x}(t) - E\hat{x}(t)\}^2\right]$$

- Too many basis functions means small bias but large sampling variance.
- Too few basis functions means small sampling variance but large bias.

Mean Squared Error

- Usually, we would really like to minimize *mean squared error*

$$\text{MSE}\left[\hat{x}(t)\right] = E\left[\left\{\hat{x}(t) - x(t)\right\}^2\right]$$

- there is a simple relationship between MSE and bias/variance

$$\text{MSE}\left[\hat{x}(t)\right] = \text{Bias}^2\left[\hat{x}(t)\right] + \text{Var}\left[\hat{x}(t)\right]$$

- This is expressed for each t, in general, we would like to minimize the *integrated* mean squared error:

$$\text{IMSE}\left[\hat{x}(t)\right] = \int \text{MSE}\left[\hat{x}(t)\right] dt$$

A Simulation

- Fit Vancouver precipitation by B-splines, to get $x(t_i)$
- Pretend this is the "truth"
- Calculate "errors"

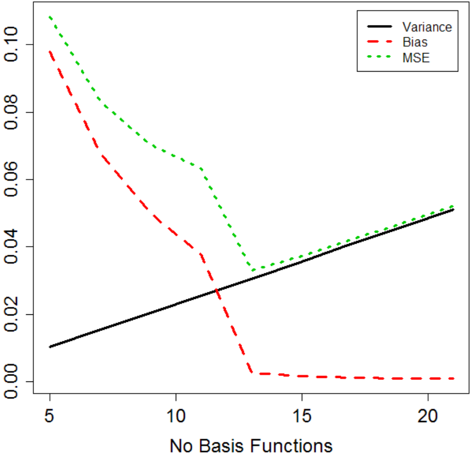
$$\epsilon_i = y_i - x(t_i)$$

- Create new "data" by randomly re-arranging the errors

$$y_i^* = x(t_i) + \epsilon_{i^*}$$

- Now fit the new data using a Fourier basis
- Repeat 1000 times; calculate bias and variance from sample.

Bias and Variance from Simulation



Cross-Validation

One method of choosing a model:

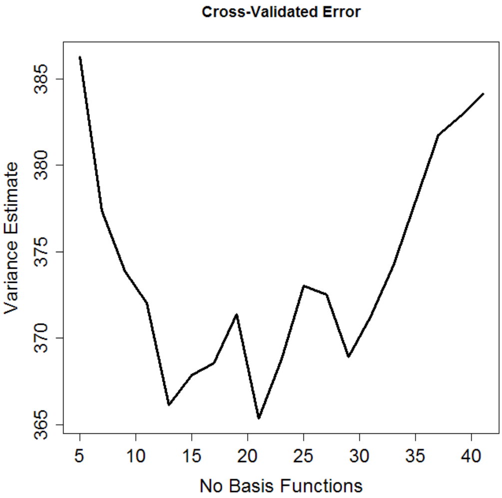
- leave out one observation (t_i, y_i)
- estimate $\hat{x}_{-i}(t)$ from remaining data
- measure $y_i - \hat{x}_{-i}(t)$
- Choose K to minimize the *ordinary cross-validation* score:

$$\mathrm{OCV}\left[\hat{x}\right] = \sum \left(y_i - \hat{x}_{-i}(t_i)\right)^2$$

- for a linear smooth $\hat{y} = Sy$,

$$\mathrm{OCV}\left[\hat{x}\right] = \sum \frac{\left(y_i - \hat{x}(t_i)\right)^2}{(1 - s_{ii})^2}$$

Cross Validation for Vancouver Precipitation



Estimating the Residual Covariance

- If we assume the standard model, then

$$\mathrm{Var}\left[\mathbf{y}\right] = \sigma^2 I$$

- An unbiased estimate is

$$\hat{\sigma}^2 = \frac{1}{N-K} MSSE$$

- Can be more sophisticated if residuals are correlated (will ignore here).

Sampling Variance of the Curve

- We know that $\hat{y} = c\boldsymbol{\Phi}$, $\mathbf{c} = C\mathbf{y}$ for $C = \left(\boldsymbol{\Phi}^T W \boldsymbol{\Phi}\right)^{-1} \boldsymbol{\Phi}^T W$
- Then under the standard model

$$\mathrm{Var}\left[\mathbf{c}\right] = \sigma^2 C I C^T$$

- More generally, if $\mathrm{Var}\left[\mathbf{y}\right] = \Sigma$, we have

$$\mathrm{Var}\left[\mathbf{c}\right] = C \Sigma C^T$$

- Then the sample variance of $\hat{y}(t)$ is

$$\mathrm{Var}\left[y(t)\right] = \Phi(t)^T C \Sigma C^T \Phi(t)$$

- And the variance-covariance matrix of the fitted values is

$$\mathrm{Var}\left[\hat{\mathbf{y}}\right] = \boldsymbol{\Phi} C \Sigma C^T \boldsymbol{\Phi}^T$$

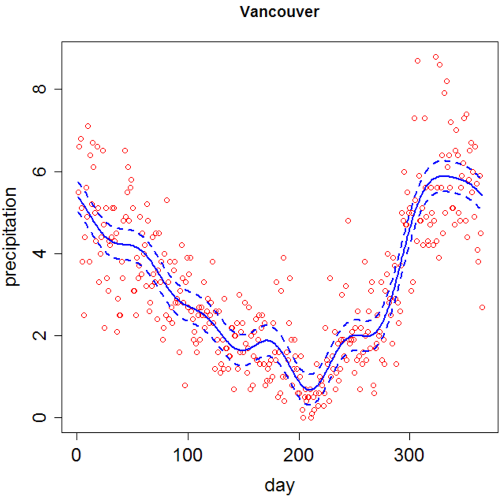
Pointwise Confidence Bands

- For each point we calculate lower and upper bands for $\hat{y}(t)$ by

$$\hat{y}(t) \pm 2\sqrt{\mathrm{Var}\left[\hat{y}(t)\right]}$$

- These bands are not confidence bands for the entire curve, but only for the value of the curve at a fixed point.
- Ignores bias in the estimated curve
- Provide an impression of how well the curve is estimated.

Fitted Vancouver Precipitation Data with 13 Fourier Bases



Summary

- Fitting smooth curves is just linear regression using basis functions as independent variables.
- Trade-off between bias and variance in choosing the number of basis functions
- Cross-validation is one way to quantitatively find the best number of basis functions
- Confidence intervals can be calculated using the standard model, but these should be treated with care
- We will see next time that there are better ways to control bias and variance.