CLIMATE INTERPOLATION
MATHEMATICAL NOTES ON METHODS
PART3

Benoit Parmentier

NCEAS, July 22, 2012

Notes assembled during the production of the climate interpolation review.

SPLINES AS SUM OF BASIS FUNCTIONS

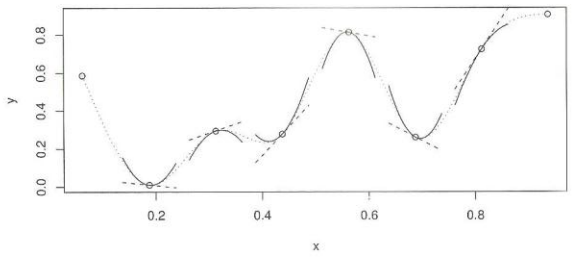


Figure 3.3 *A cubic spline is a curve constructed from sections of cubic polynomial joined together so that the curve is continuous up to second derivative. The spline shown (dotted curve) is made up of 7 sections of cubic. The points at which they are joined (◦) (and the two end points) are known as the* knots *of the spline. Each section of cubic has different coefficients, but at the knots it will match its neighbouring sections in value and first two derivatives. Straight dashed lines show the gradients of the spline at the knots and the continuous curves are quadratics matching the first and second derivatives at the knots: these illustrate the continuity of first and second derivatives across the knots. This spline has zero second derivatives at the end knots: a* 'natural *spline'. Note that there are many alternative ways of representing such a cubic spline, using basis functions: although all are equivalent, the link to the piecewise cubic characterization is not always transparent.*

Another example: A cubic spline basis

A univariate function can be represented using a *cubic spline*. A cubic spline is a curve made up of sections of cubic polynomial joined together so that they are continuous in value, as well as first and second derivatives (see figure 3.3). The points at which the sections join are known as the *knots* of the spline. For a conventional spline, the knots occur wherever there is a datum, but for the regression splines of interest here, the locations of the knots must be chosen. Typically the knots would either be evenly spaced through the range of observed x values, or placed at quantiles of the distribution of unique x values. Whatever method is used, let the knot locations be denoted by $\{x_i^* : i = 1, \cdots, q-2\}$.

Given knot locations, there are many alternative, but equivalent, ways of writing down a basis for cubic splines. A simple basis to use, results from the very general approach to splines that can be found in the books by Wahba (1990) and Gu (2002), although the basis functions are slightly intimidating when written down (and the link to the definition of a cubic spline given in figure 3.3 is rather opaque). For this

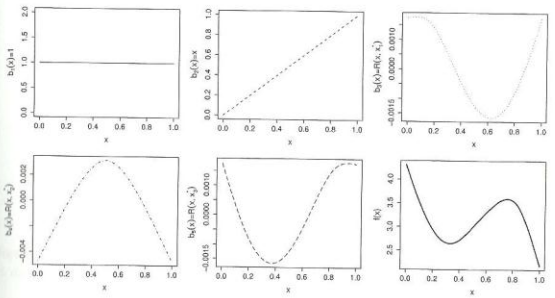


Figure 3.4 *Illustration of the representation of a smooth function using a rank 5 cubic regression spline basis (with knot locations $x_1^* = 1/6$, $x_2^* = 3/6$ and $x_3^* = 5/6$). The first 5 panels (starting from top left) illustrate the 5 basis functions, $b_j(x)$, for a rank 5 cubic spline basis. The basis functions are each multiplied by a real valued parameter, β_j, and are then summed to give the final curve $f(x)$, an example of which is shown in the bottom right panel. By varying the β_j we can vary the form of $f(x)$. See also figure 3.5*

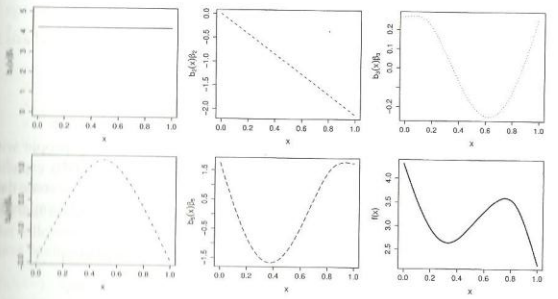


Figure 3.5 *An alternative illustration of how a function is represented in terms of cubic spline basis functions. This figure shows the same rank 5 cubic regression spline basis that is shown in figure 3.4, but in this case the basis functions, $b_j(x)$, are each shown multiplied by corresponding coefficients β_j (first five figures, starting at top left). Simply summing these 5 curves yields the function, $f(x)$, shown at bottom right.*

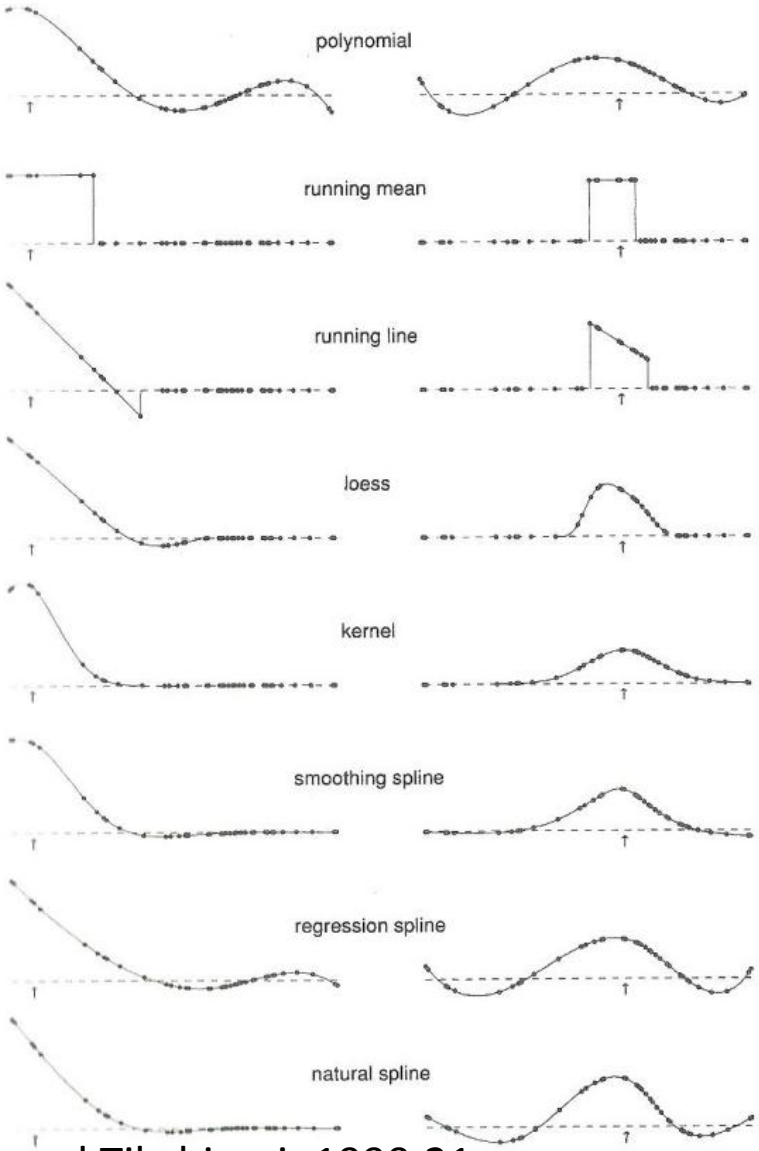
Function bases can be rescaled by their coefficients…

There are constraints at the knots so that the function is smooth.

Wood et al. 2006:124

Let a interval [0,1] in which we need to approximate a given set of points by a continuous analytical function: a sum of piecewise polynomials with smooth edges is a candidate….

SMOOTHING FUNCTION AND EQUIVALENT KERNELS



Hastie and Tibshirani: 1990:21

Fig. 2.5. *The equivalent kernels for the smoothers in Fig. 2.1. The bin smoother is omitted, since it resembles the running mean. The arrows indicate the target point in each case. The points are plotted at the pairs* (x_j, S_{0j}).

- Smoother scatterplots act on the set of points from the domain of prediction by removing details or noise and generalization locally…

- To smoothing functions we can associate **equivalent kernels** that represent the local effects of bases functions in a neighborhood around a particular observation.

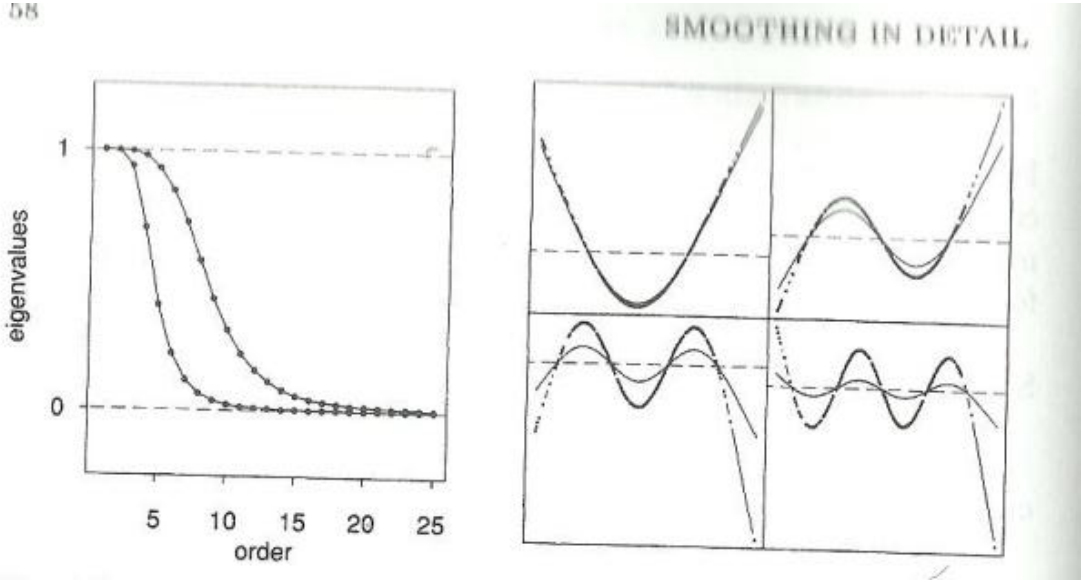
"The cubic smoothing spline, because of the implicit ways it is defined, doesn't appear to use local averaging. However, the equivalent-kernel weights in Fig.2.5 show that it does possess local behavior quite similar to kernels or local weighted lines. The equivalent kernel is nowhere nonzero, but it is close to zero far from the target point. "

"One might say, then that a cubic smoothing spline is approximately a kernel smoother. The smoothing parameter controls the shape of the kernel or weight function.

Hastie and Tibshirani, 1990:29.

SPECTRAL DECOMPOSITION OF SMOOTHING FUCTIONS

Hastie and Tibshirani 1990:58



- Smoothing functions can studied using eigen value decomposition.
- Figure on the right shows the third and sixth eigenvectors of a smoothing spline matrix.

The two different spectrum correspond to different values of the smoothing term. Larger smoothing term λ results in eigenvalue spectrum that decays more quickly.

The analysis of a linear scatterplot smoother through an eigenanalysis of the corresponding smoother matrix is closely related to the study of the transfer function of a linear filter for time series. This analogy can add insight, so we provide a brief summary here. Consider a time series yt: t=0,+/-1,…

$$\hat{y}_t = \sum_{j}^{b} c_j y_{t-j}.$$ → Linear smoother

$$\mathrm{Re}\Big[\sum_{j=a}^{b} c_j \exp(-i\omega j) r \exp\{i(\omega t + \phi)\}\Big],$$

See Hastie and Tibshirani 1990:59.

SPECTRAL DECOMPOSITION OF BASIS FUNCTION

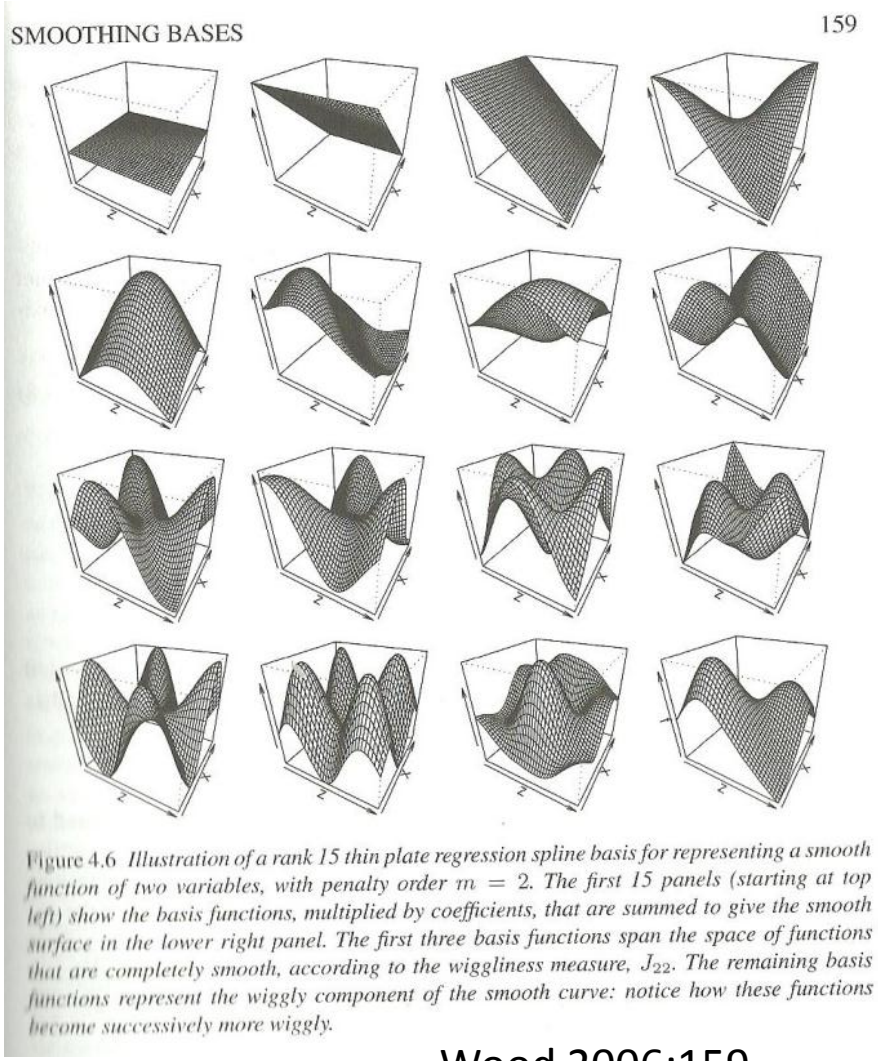


Figure 4.6 *Illustration of a rank 15 thin plate regression spline basis for representing a smooth function of two variables, with penalty order* $m = 2$. *The first 15 panels (starting at top left) show the basis functions, multiplied by coefficients, that are summed to give the smooth surface in the lower right panel. The first three basis functions span the space of functions that are completely smooth, according to the wiggliness measure,* J_{22}. *The remaining basis functions represent the wiggly component of the smooth curve: notice how these functions become successively more wiggly.*

Wood 2006:159

$$\hat{e}_k = \max_{\boldsymbol{\delta} \neq \mathbf{0}} \frac{\|(\mathbf{E} - \hat{\mathbf{E}}_k)\boldsymbol{\delta}\|}{\|\boldsymbol{\delta}\|}.$$

sary, since the upper norm otherwi
ıe spline is measured by the penal
ırst possible change in the shape of

$$\tilde{e}_k = \max_{\boldsymbol{\delta} \neq \mathbf{0}} \frac{\boldsymbol{\delta}^{\mathsf{T}}(\mathbf{E} - \tilde{\mathbf{E}}_k)\boldsymbol{\delta}}{\|\boldsymbol{\delta}\|^2}.$$

This is the spectral decomposition of a two dimensional thin plate regression spline (TPRS). Note that the larger eigenvalues show three polynomials of first degree (planes). The eigenvectors show higher degree of the polynomial terms for smaller eigenvalue terms…

TPS bases are expensive to calculate (of the $O(n^3)$ order) but its basis space can be approximated using eigenvectors from the spectral decomposition. This is cheaper in terms of operations using the Lanczos algorithm ($O(n^2k)$ operations).

SMOOTHING TERM: λ

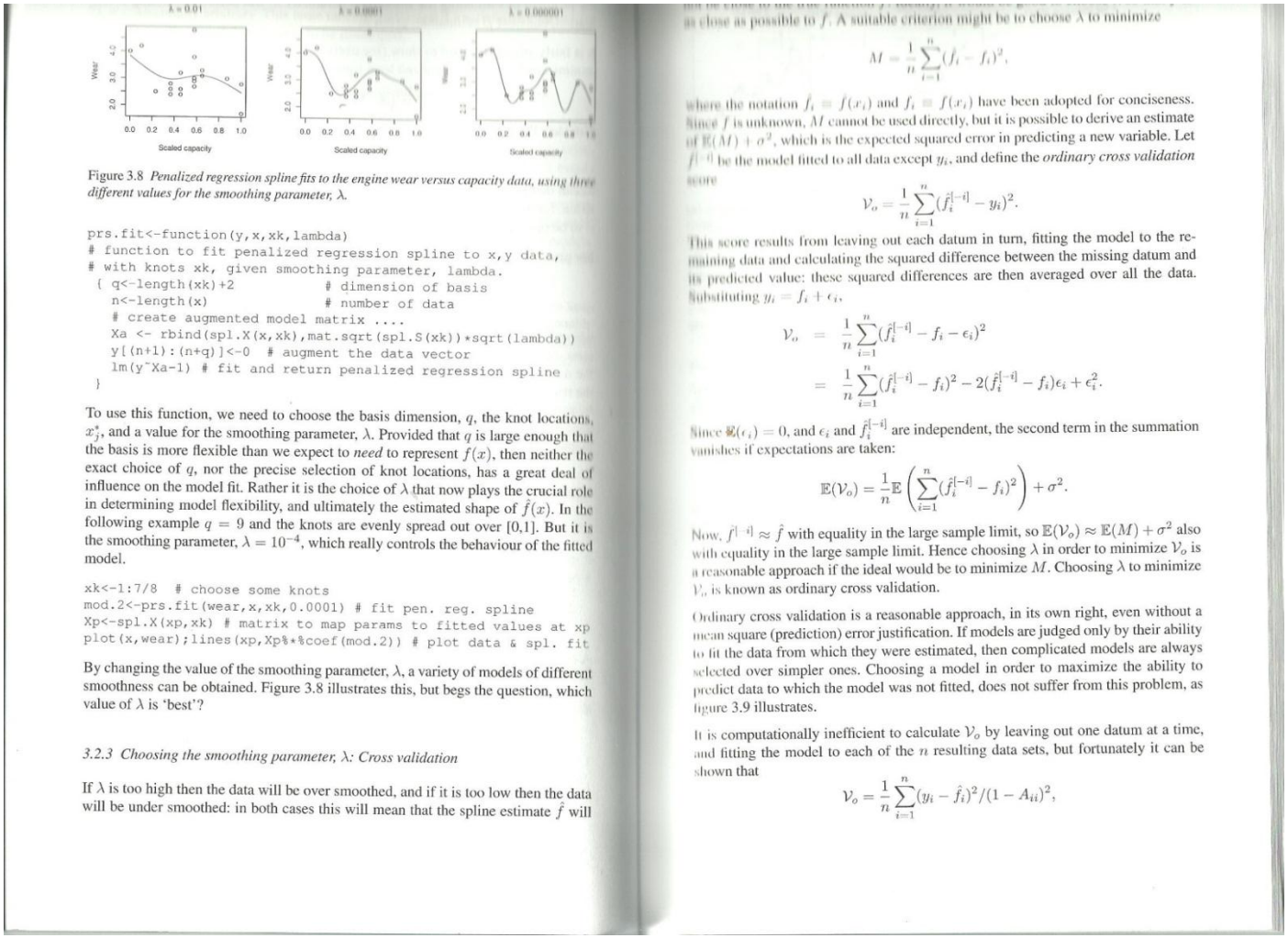


Figure 3.8 *Penalized regression spline fits to the engine wear versus capacity data, using three different values for the smoothing parameter, λ.*

```
prs.fit<-function(y,x,xk,lambda)
# function to fit penalized regression spline to x,y data,
# with knots xk, given smoothing parameter, lambda.
{ q<-length(xk)+2              # dimension of basis
  n<-length(x)                 # number of data
  # create augmented model matrix ....
  Xa <- rbind(spl.X(x,xk),mat.sqrt(spl.S(xk))*sqrt(lambda))
  y[(n+1):(n+q)]<-0  # augment the data vector
  lm(y~Xa-1) # fit and return penalized regression spline
}
```

To use this function, we need to choose the basis dimension, q, the knot locations, x_j^*, and a value for the smoothing parameter, λ. Provided that q is large enough that the basis is more flexible than we expect to *need* to represent $f(x)$, then neither the exact choice of q, nor the precise selection of knot locations, has a great deal of influence on the model fit. Rather it is the choice of λ that now plays the crucial role in determining model flexibility, and ultimately the estimated shape of $\hat{f}(x)$. In the following example $q = 9$ and the knots are evenly spread out over [0,1]. But it is the smoothing parameter, $\lambda = 10^{-4}$, which really controls the behaviour of the fitted model.

```
xk<-1:7/8  # choose some knots
mod.2<-prs.fit(wear,x,xk,0.0001) # fit pen. reg. spline
Xp<-spl.X(xp,xk) # matrix to map params to fitted values at xp
plot(x,wear);lines(xp,Xp%*%coef(mod.2)) # plot data & spl. fit
```

By changing the value of the smoothing parameter, λ, a variety of models of different smoothness can be obtained. Figure 3.8 illustrates this, but begs the question, which value of λ is 'best'?

3.2.3 *Choosing the smoothing parameter, λ: Cross validation*

If λ is too high then the data will be over smoothed, and if it is too low then the data will be under smoothed: in both cases this will mean that the spline estimate $\hat{f}$ will

as close as possible to f. A suitable criterion might be to choose λ to minimize

$$M = \frac{1}{n}\sum_{i=1}^{n}(\hat{f}_i - f_i)^2,$$

where the notation $\hat{f}_i = \hat{f}(x_i)$ and $f_i = f(x_i)$ have been adopted for conciseness. Since f is unknown, M cannot be used directly, but it is possible to derive an estimate of $\mathbb{E}(M) + \sigma^2$, which is the expected squared error in predicting a new variable. Let $\hat{f}^{[-i]}$ be the model fitted to all data except y_i, and define the *ordinary cross validation* score

$$\mathcal{V}_o = \frac{1}{n}\sum_{i=1}^{n}(\hat{f}_i^{[-i]} - y_i)^2.$$

This score results from leaving out each datum in turn, fitting the model to the remaining data and calculating the squared difference between the missing datum and its predicted value: these squared differences are then averaged over all the data. Substituting $y_i = f_i + \epsilon_i$,

$$\begin{aligned}\mathcal{V}_o &= \frac{1}{n}\sum_{i=1}^{n}(\hat{f}_i^{[-i]} - f_i - \epsilon_i)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(\hat{f}_i^{[-i]} - f_i)^2 - 2(\hat{f}_i^{[-i]} - f_i)\epsilon_i + \epsilon_i^2.\end{aligned}$$

Since $\mathbb{E}(\epsilon_i) = 0$, and ϵ_i and $\hat{f}_i^{[-i]}$ are independent, the second term in the summation vanishes if expectations are taken:

$$\mathbb{E}(\mathcal{V}_o) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n}(\hat{f}_i^{[-i]} - f_i)^2\right) + \sigma^2.$$

Now, $\hat{f}^{[-i]} \approx \hat{f}$ with equality in the large sample limit, so $\mathbb{E}(\mathcal{V}_o) \approx \mathbb{E}(M) + \sigma^2$ also with equality in the large sample limit. Hence choosing λ in order to minimize $\mathcal{V}_o$ is a reasonable approach if the ideal would be to minimize M. Choosing λ to minimize $\mathcal{V}_o$ is known as ordinary cross validation.

Ordinary cross validation is a reasonable approach, in its own right, even without a mean square (prediction) error justification. If models are judged only by their ability to fit the data from which they were estimated, then complicated models are always selected over simpler ones. Choosing a model in order to maximize the ability to predict data to which the model was not fitted, does not suffer from this problem, as figure 3.9 illustrates.

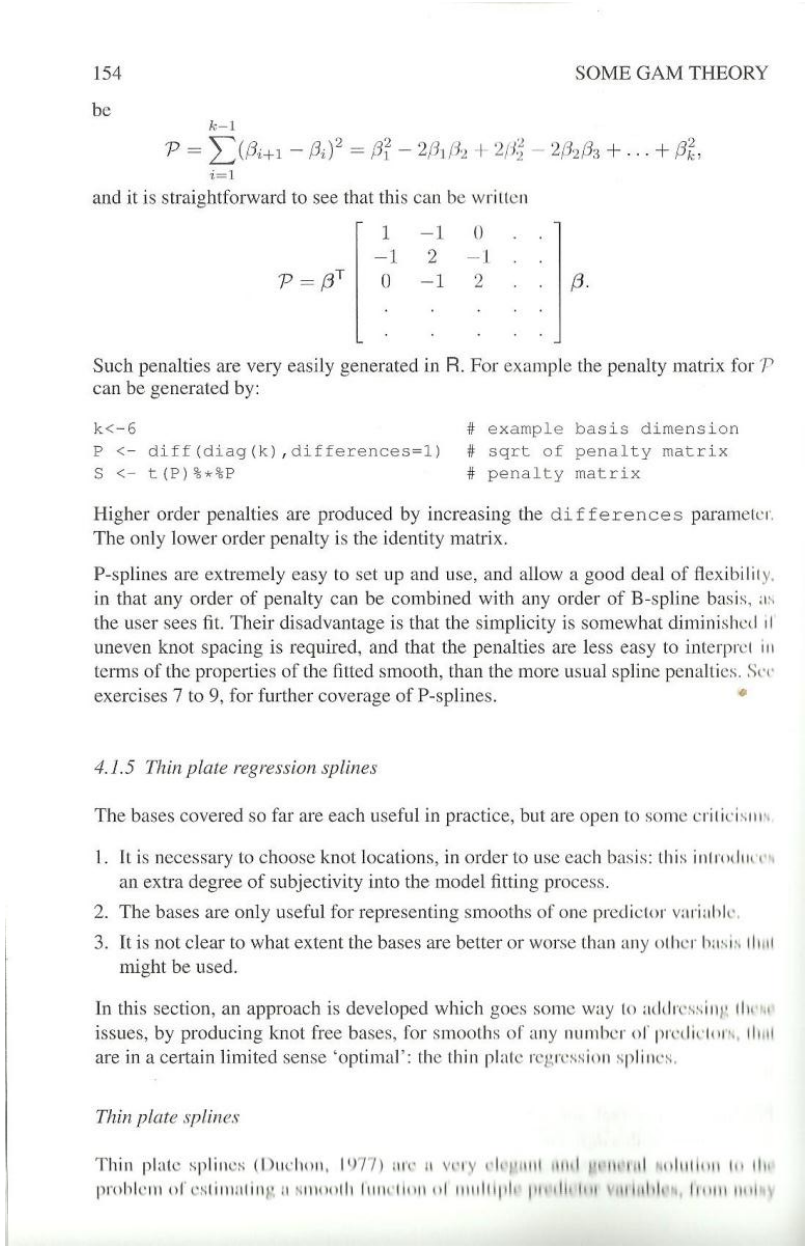
It is computationally inefficient to calculate $\mathcal{V}_o$ by leaving out one datum at a time, and fitting the model to each of the n resulting data sets, but fortunately it can be shown that

$$\mathcal{V}_o = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}_i)^2/(1 - A_{ii})^2,$$

- The smoothing term controls the size of the neighborhood.
- Large λ correspond to smoothing function that are flatter.

"The averaging is done in neighbourhoods around the target value. There are two main decisions to be made in scatterplot smoothing:
- How to average the response values in each neighbourhood, and
- How big to take the neighbourhoods."

B-SPLINE AND P-SPLINES



154 SOME GAM THEORY

be

$$\mathcal{P} = \sum_{i=1}^{k-1} (\beta_{i+1} - \beta_i)^2 = \beta_1^2 - 2\beta_1\beta_2 + 2\beta_2^2 - 2\beta_2\beta_3 + \ldots + \beta_k^2,$$

and it is straightforward to see that this can be written

$$\mathcal{P} = \boldsymbol{\beta}^{\mathsf{T}} \begin{bmatrix} 1 & -1 & 0 & . & . \\ -1 & 2 & -1 & . & . \\ 0 & -1 & 2 & . & . \\ . & . & . & . & . \\ . & . & . & . & . \end{bmatrix} \boldsymbol{\beta}.$$

Such penalties are very easily generated in R. For example the penalty matrix for $\mathcal{P}$ can be generated by:

```
k<-6                              # example basis dimension
P <- diff(diag(k),differences=1)  # sqrt of penalty matrix
S <- t(P)%*%P                     # penalty matrix
```

Higher order penalties are produced by increasing the `differences` parameter. The only lower order penalty is the identity matrix.

P-splines are extremely easy to set up and use, and allow a good deal of flexibility, in that any order of penalty can be combined with any order of B-spline basis, as the user sees fit. Their disadvantage is that the simplicity is somewhat diminished if uneven knot spacing is required, and that the penalties are less easy to interpret in terms of the properties of the fitted smooth, than the more usual spline penalties. See exercises 7 to 9, for further coverage of P-splines.

4.1.5 Thin plate regression splines

The bases covered so far are each useful in practice, but are open to some criticisms.

1. It is necessary to choose knot locations, in order to use each basis: this introduces an extra degree of subjectivity into the model fitting process.
2. The bases are only useful for representing smooths of one predictor variable.
3. It is not clear to what extent the bases are better or worse than any other basis that might be used.

In this section, an approach is developed which goes some way to addressing these issues, by producing knot free bases, for smooths of any number of predictors, that are in a certain limited sense 'optimal': the thin plate regression splines.

Thin plate splines

Thin plate splines (Duchon, 1977) are a very elegant and general solution to the problem of estimating a smooth function of multiple predictor variables, from noisy

Other common splines found in the literature are P-splines and B-splines.

B-SPLINES are another way to represent the cubic splines.

B-splines are interesting representation of cubic splines because they are truly local.

B-splines can be written recursively and were introduced by Boor in 1978. “They were developed as a very stable basis for large scale spline interpolation…” Wood et al. 2006:153

P-Splines can be developed from B-splines (Eilers and Marx 1996).

“P-splines are low rank smoothers using a B-spline basis, usually defined on evenly spaced knots, with a difference penalty applied directly to the parameters, βi. Values…”

P-splines are useful but need to have even spaced knots…

CLIMATOLOGY AIDED INTERPOLATION

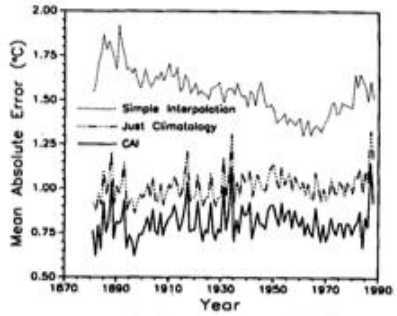


Figure 3. Time series of spatially integrated mean absolute interpolation error (MAE) from cross-validation analyses of (i) Willmott *et al.*'s (1985) interpolations from Jones *et al.*'s yearly station data (dotted line), (ii) climatologically aided interpolation (CAI) again interpolating with Willmott *et al.* (solid line), and (iii) climatology alone (mixed dashed and dotted line).

differences between the interpolated Legates and Willmott ($\bar{T}_j$) and Jones *et al.* (T_j) temperatures at station j, but they also contain information about station biases that influence differences between $\bar{T}_j$ and T_j. When the δT_i field is interpolated from the δT_j field, it (the δT_i field) then contains interpolated bias-correction information; in turn, when the δT_i field is added to the estimated $\bar{T}_i$ field, the grid-point estimates approximate the Jones *et al.* station values in quality, albeit at a much higher spatial resolution. It is this aspect of our somewhat unusual anomaly fields that allows CAI to effectively make use of the inhomogeneous—but spatially high-resolution—station records that reside in Legates and Willmott's climatology.

4.3. Interpolation errors for CAI

Cross-validation was performed (as in section 3.2) for CAI using Willmott *et al.* (1985) to interpolate both the $\bar{T}_j$ and the δT_j. Substantial improvements over the simple interpolation results (using only yearly air temperatures) are apparent (Figure 3). Interpolation errors are reduced by over 50 per cent, and approach interpolation errors associated with more standard anomaly data (Robeson, 1993). Scatter plots (Figure 4)

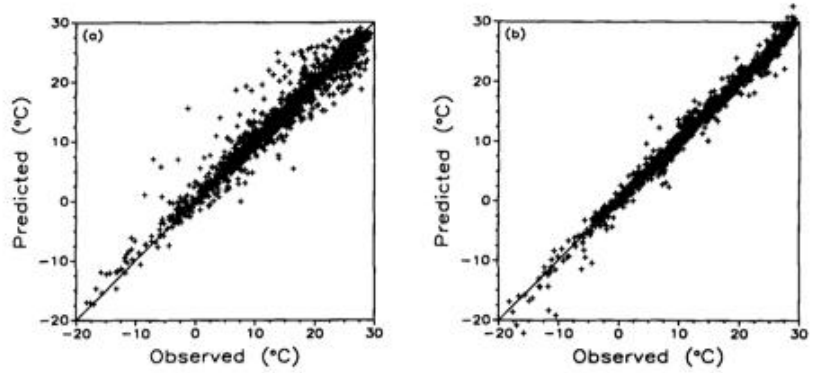


Figure 4. Scatter plots of observed air temperature versus cross-validation estimates for 1962 using (a) Willmott *et al.*'s algorithm with the yearly station data only and (b) climatologically aided interpolation (CAI).

Willmott and Robeson
1995: CAI has lower cross validation errors
In time with lower variance too.

T(interpolated)= T(climatology)+ T(deviation)

Wilmott 1995 makes the case for the use of higher quality climatology spatial surface for CAI. He presents an interpolated surface of mean annual temperature based on some 18,00 stations from the Willmott and Legate station database (1990). Such map can distinguish feature such as the Ethiopian highlands and the limit in eh Atacama desert and Andes mountain. When used as basis for interpolation in CAI it can improved prediction substantially.

"For many spatial applications, its high spatial resolution outweights the deleterious effects of variable averaging periods. Preliminary work work by the authors using precipitation data suggests that interpolation errors (when using traditional interpolation methods) are greater when station records from different averaging periods are left out rather than included."Wilmott and Robeson 1995"

CLIMATE INTERPOLATION: STATION DATABASE

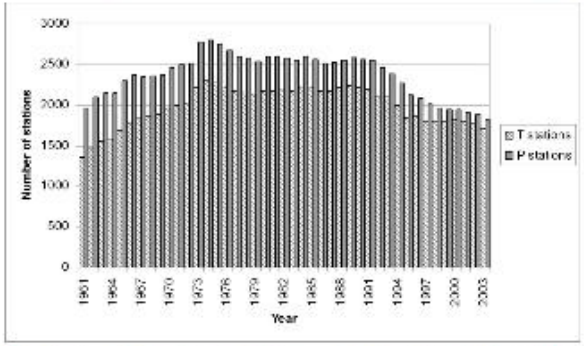


FIG. 1. Numbers of climate stations with daily temperature and precipitation data for 1961–2003.

thousands of precipitation values that would otherwise have been set to missing.

One issue that was not addressed was the difference in the climate day between ordinary climate stations and principal meteorological stations. Since July 1961 all principal stations have a climate day ending at 0600 UTC (around 0000 LST). On the other hand, ordinary climate stations, which make up the overwhelming

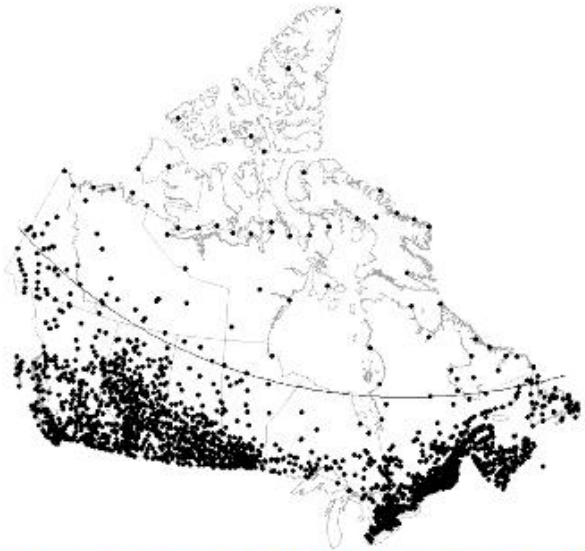


FIG. 2. Station locations used to generate daily precipitation models; example for yearday 250 in 1975. Line defined by latitude = −0.15 × longitude +42.0. Lambert conformal conic projection.

Hutchinson et al. 2009 Developing interpolation in Canada.

VARIATION OF FREQUENCY OF THE NETWORK IN TIME

There is some variation in the number of Station over the interpolation period (1961-2003).

"The period 1961-2003 was selected for the current work. The number of precipitation stations that were active in any give year over this period ranged from 2000-3000 while the number of temperature stations varied from about 1500 to 2200. as shown in Fig.1."

UNEQUAL DISTRIBUTION OF THE NETWORK IN SPACE

→ This figure shows the stations used for modeling on day of year 250 in 1975. Coverage is denser in the south…

BRIGS ET AL. 1996 TOPOGRAPHIC BIAS

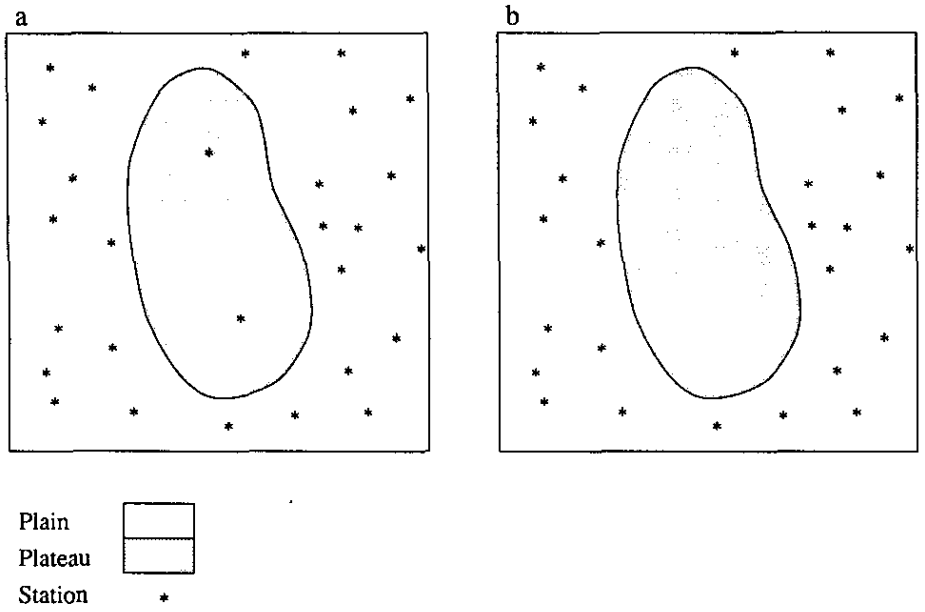


FIG. 1. Spatially biased sampling networks in which a major topographic feature is (a) resolved and (b) unresolved.

Relationships at higher elevations are often not represented properly because of bias in the distribution of the network.

Elevation bias is clearly present in the stations network we used, but it is unclear how much, and where, this affects the results. Briggs and Cogley (1996) showed that in the United States weather stations tend to be biased toward lower elevations, and our results confirm this, but show that while this pattern is common at high latitudes and in the subtropics, it tends to be reversed in the tropics. Hijmans et al. 2000:1977.

COVARIANCE AND SEMI-VARIOGRAMS

Note that the covariance is the inverse of the semivariance…

This is why it needs to be inverted to connect TPS and Kriging. See Hastie and Tibshirani 1990: 29 for details…

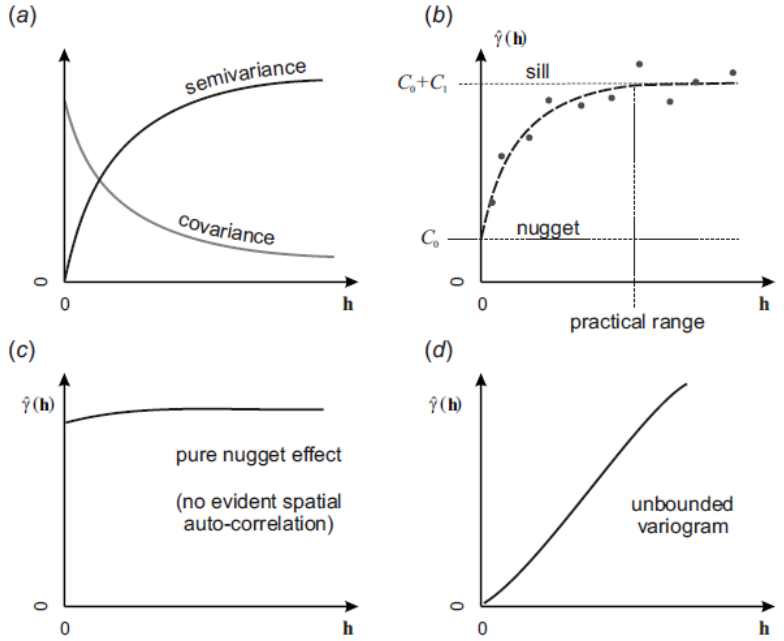
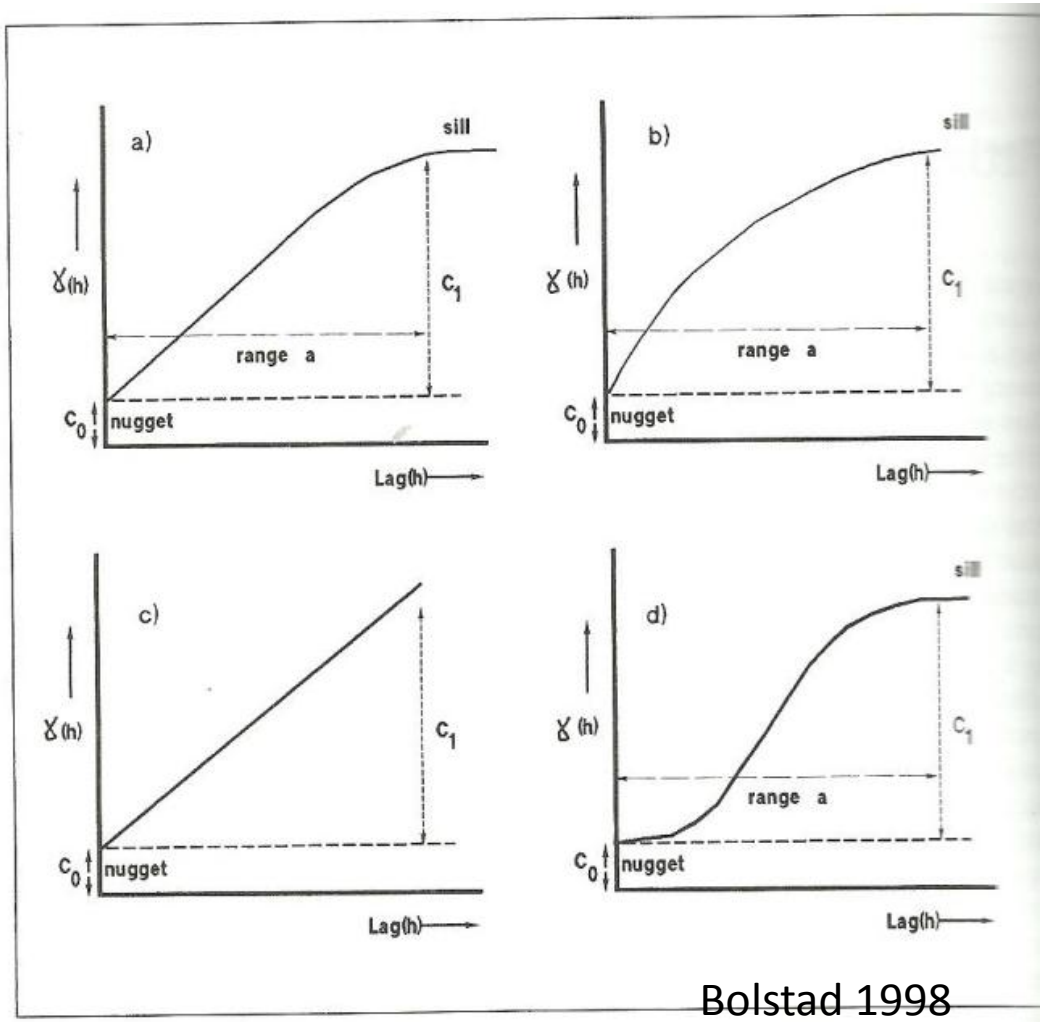


Fig. 1.10: Some basic concepts about variograms: (a) the difference between semivariance and covariance; (b) it is often important in geostatistics to distinguish between the sill variation ($C_0 + C_1$) and the sill parameter (C_1) and between the range parameter (R) and the practical range; (c) a variogram that shows no spatial correlation can be defined by a single parameter (C_0); (d) an unbounded variogram.

VARIOGRAM AND SEMI-VARIOGRAM MODELS



Bolstad 1998

Figure 6.3. Examples of the most commonly used variogram models: (*a*) spherical; (*b*) exponen
(*c*) linear; and (*d*) Gaussian

Interpolation of new values requires the fitting of a covariance function that will allow to predict new values at unknown places….

The function must be positive on its domain (positive definite matrix)….

"The shape of the semivariogram near the origin is of particular interest since it indicates the degree of smoothness of spatial continuity of the spatial variable under study. A parabolic shape near the origin arises with a very smooth spatial variable that is both continuous and differentiable. A linear shape near the origin reflects a variable that is continuous but not differentiable, and hence less regular. A discontinuity, or vertical jump, at the origin […] indicates that the spatial variable is not continuous and has highly irregular p.276 Waller and Gotway…

Variograms represent the relationship between pairwise observation as a function of distance (or lags)

NEW ET AL. 2001: COMPARISON IN ACCURACY ASSESSMENT

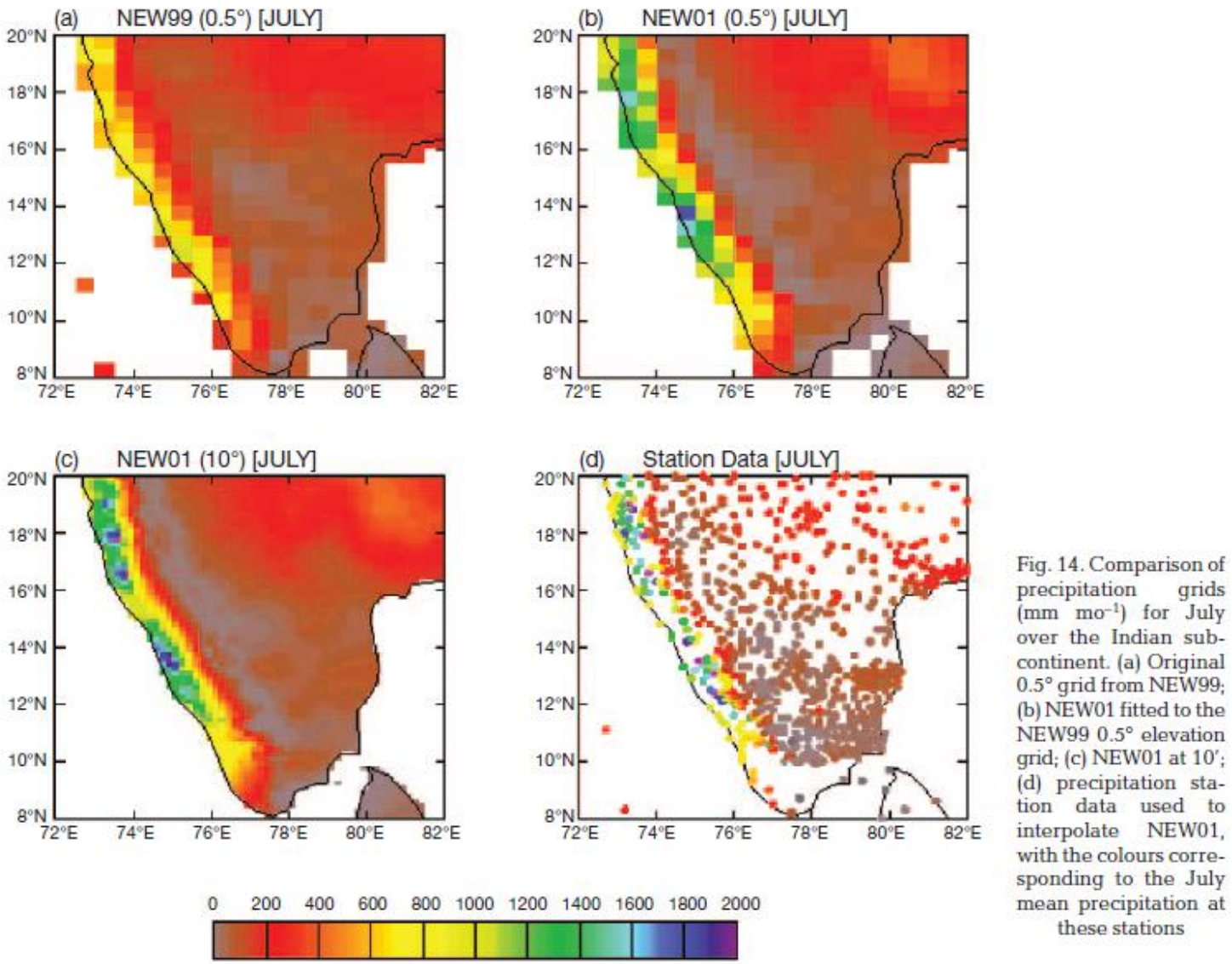


Fig. 14. Comparison of precipitation grids (mm mo^{-1}) for July over the Indian subcontinent. (a) Original 0.5° grid from NEW99; (b) NEW01 fitted to the NEW99 0.5° elevation grid; (c) NEW01 at 10′; (d) precipitation station data used to interpolate NEW01, with the colours corresponding to the July mean precipitation at these stations

New et al. 2002

INTER-PRODUCT PRODUCT COMPARISON

Clear improvements
In terms of details for
NEW01 compared to
Older product NEW99.

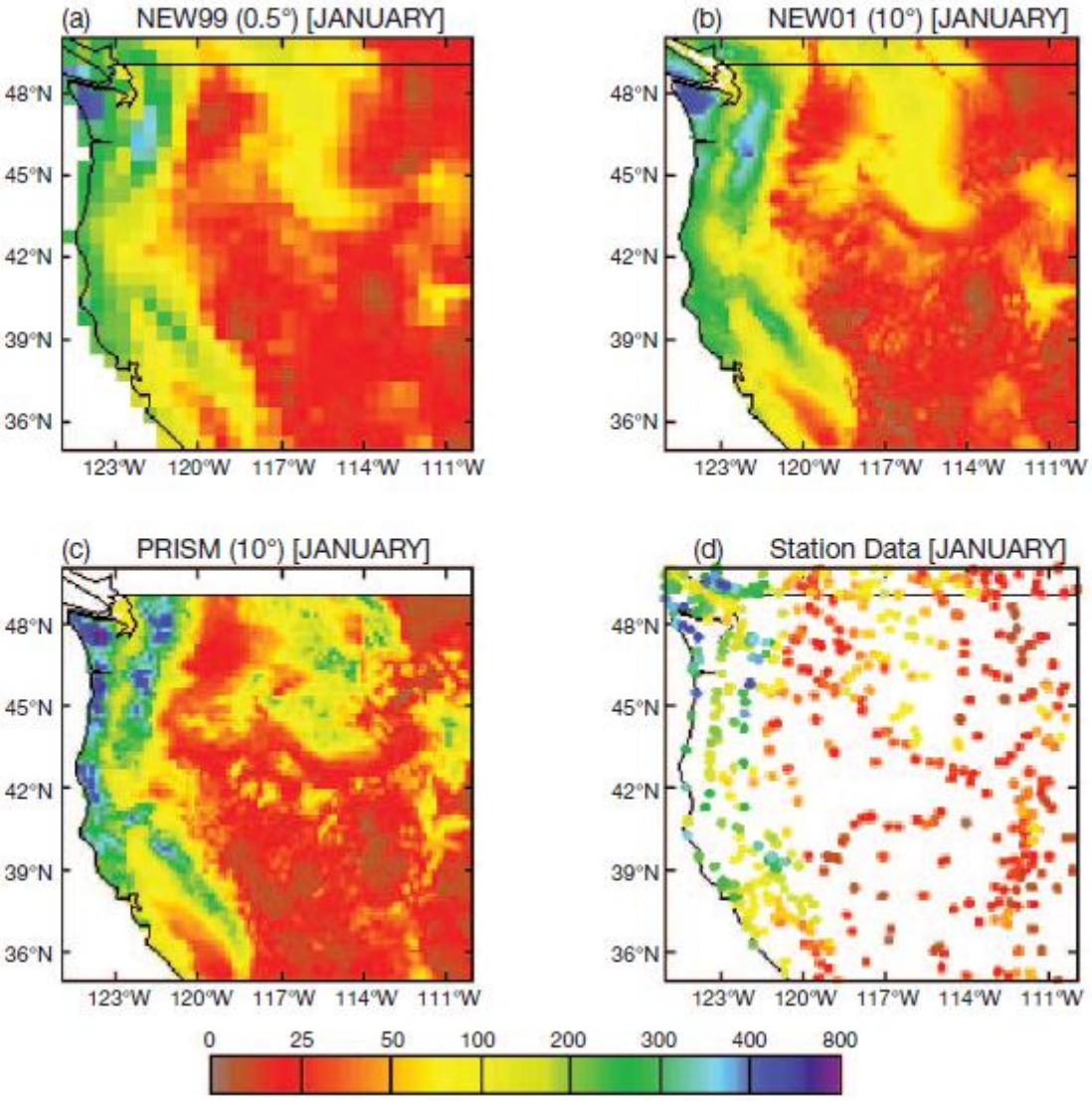


Fig. 15. Comparison of precipitation grids (mm mo^{-1}) for January over the western USA. (a) Original 0.5° grid from NEW99; (b) NEW01 at 10′; (c) the PRISM precipitation grid at 10′; (d) precipitation station data used to interpolate NEW01, with the colours corresponding to the January mean precipitation at these stations

New et al. 2002