Chapter 6
Mapping Data to a Grid and Data Assimilation

6.1 Placing data on a regular grid

In dynamical meteorology, oceanography, and numerical prediction one is often presented with the following problem. Data are available at a number of observation points (usually located near cities or at field stations, along ship cruise tracks, at moorings, or perhaps located by the observation points of an orbiting satellite) that are unevenly distributed over the domain of interest (the globe, for example). In order to compute derivatives of the field variables, as would be required in diagnostic studies or in the initialization of a numerical model, or simply to perform a sensible averaging process, one often requires values of the variables at points on a regular grid. Assigning the best values at the grid points, given data at arbitrarily located stations and perhaps a first guess at regular grid points, is what has traditionally been called objective analysis when done on a computer rather than graphically by hand.

We will use the example of making weather maps from rawinsonde data as the particular example of the mapping problem here. In fact the methods described are applicable to any problem where the data you are given do not fill the domain of interest fully, and/or where the data must be interpolated to a regular grid. The regridding can be in space, in time, or both. You may also find yourself in the position of wanting to plot a continuous function of an observation in two parameter dimensions, and have samples at only a few points. We will proceed through some of the methods in the order that they arose in the history of numerical weather forecasting. In this way we show the weaknesses of some of the most obvious methods such as function fitting, to the correction method, and ultimately to statistically optimized correction methods such as optimum interpolation. Current assimilation schemes in numerical forecast models us a combination of optimum interpolation and use of the governing equations of the model, which we can call *Kalman filtering*, which is discussed in elementary terms in Chapter **??**.

6.1.1 Interpolation with polynomial fits

Let's say we want to estimate the temperature at a point. However, we don't have any observations at that exact location. How might we use our observations to still get an estimate of the temperature at our point? The answer could be to perform some sort of interpolation. Probably one of the most intuitive methods for interpolating is to fit some polynomial to all of our station values, and then use that curve to get the temperature at a location between the observations. For example,

$$\Phi(x, y) = a_0 + a_1 x + a_2 x^2 + b_2 y^2 + 2c_2 xy + \dots \tag{6.1}$$

It turns out this isn't a very good method when you have sparse data due to the unstable nature of the polynomial fit. Removing just one point can wildly change the curve/interpolation in the vicinity of this point and will impact the values at many other points too. The problem gets worse as the order of the polynomial is increased. An example of this is depicted in **Fig. 6.1**. Note how wildly the two curves depart from each other in the vicinity of the missing point. Such problems can be avoided by stepping away from

polynomial fits and rather, utilizing a reasonable "first guess", and then only modifying it when and where data are available. Also, if the new data departs too wildly from the first guess, one suspects that the data are faulty.

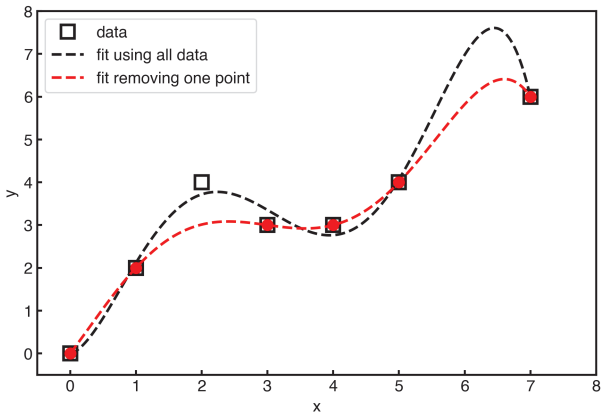

Figure 6.1 Illustration of the unstable nature of polynomial fits when one data point is removed using a 5th order polynomial.

A polynomial fit that actually got adopted by the US National Meteorological Center for its routine operational products was proposed by Flattery (1971). In this scheme, *Hough functions* were used as the interpolating polynomials. These functions are an orthogonal set that are the solutions of the linearized equations for a resting atmosphere (the tidal equations). The idea was that if you expressed the data in terms of actual solutions of the dynamical equations, then your fit between the data points would have some dynamical consistency. The Hough functions are global functions and so all of the observations were used simultaneously to define the global Hough function coefficients and produce a global map. Only the Hough functions describing slowly varying rotational modes were used. The gravity wave modes were zeroed out to produce a well-initialized field. This method replaced Cressman's correction method (Cressman, 1959) for global analyses in about 1972 and was replaced by Optimum Interpolation (see Chapter 6.1.2) in 1978.

This method has some dynamical and mathematical appeal, but is in truth just a glorified polynomial fit and has all of the problems of polynomial fits. First of all, the atmosphere is highly nonlinear and strongly forced by heating, especially in the tropics. The Hough modes chosen were primarily the free, non-divergent Rossby modes, which constitute a large, but not dominant, fraction of the variance. Therefore this aspect of the Flattery method did not buy much. In the tropics, where highly divergent motions forced by heating are important, the analyses constructed with the Flattery method are very much in error, especially in their estimates of divergence, which they set to essentially zero. In addition the Hough function fits are wildly unstable in regions of sparse data, like any polynomial fit. The NMC tropical analyses produced before 1978 are almost totally useless because they were made with the Flattery analysis system. Normal mode fits are still used in numerical initialization schemes to remove fast gravity waves, but this does not really affect the slowly changing meteorological flow. Modern reanalysis data products are based on data assimilation methods that take into account both the data and the model forecast and the uncertainty in both.

6.1.2 Optimum Interpolation

> *"The interpolation which is linear relative to the initial data and whose root-mean- square error is minimum is called the optimum interpolation." - Wiener, 1949*

The difference between optimum interpolation and linear regression is that the coefficients are not determined anew each time. Suppose we consider deviations from some "normal" state. This could be climatology or a

first guess, depending upon the application.

$$\phi^{'} = \phi - \phi_{norm} \quad \phi_{norm} = \overline{\phi} \text{ or a first guess} \tag{6.2}$$

Then we try to approximate the value of ϕ at a grid point, ϕ_g, in terms of a linear combination of the values of ϕ at neighboring station points, ϕ_s.

$$\phi_g^{'} = \sum_{i=1}^{N} p_i \phi_i^{'} \tag{6.3}$$

The coefficients p_i are to be determined by minimizing the mean squared error

$$E = \overline{\left(\phi_g^{'} - \sum_{i=1}^{N} p_i \phi_i^{'}\right)^2} \tag{6.4}$$

We can write the normalized error as

$$\epsilon \equiv \frac{E}{\overline{\phi_g^{'2}}} = 1 - 2\sum_{i=1}^{N} p_i r_{gi} + \sum_{i=1}^{N}\sum_{j=1}^{N} p_i p_j r_{ij} \tag{6.5}$$

$$\text{where } r_{gi} = \frac{\overline{\phi_g^{'}\phi_i^{'}}}{\overline{\phi_g^{'2}}} \quad r_{ij} = \frac{\overline{\phi_i^{'}\phi_j^{'}}}{\overline{\phi_g^{'2}}} \tag{6.6}$$

Differentation with respect to the coefficients leads to the condition of minimization used to determine them.

$$\frac{\partial \epsilon}{\partial p_i} = -2r_{gi} + 2\sum_{j=1}^{N} p_j r_{ij} = 0 \quad i = 1, 2, \ldots, N \tag{6.7}$$

(6.7) constitutes a system of N linear equations for the N p's. By substituting the conditions **(6.7)** into the expression for the error **(6.5)**, it can be shown that the error obtained after fitting the coefficients is

$$\epsilon = 1 - \sum_{i=1}^{2} r_{gi} p_i \tag{6.8}$$

Note that in this simple example, if one of the observation points, k, coincides with a grid point, then $r_{gk} = 1$, and we expect the regression procedure to return $p_k = 1$ and all the other weights zero. In this case the error is zero, $\epsilon = 0$, since we have assumed the data are perfect. If the station points are uncorrelated with the grid point in question, then $p_i = 0$ and $\epsilon = 1$, the climatic norm. That is, the error will equal the standard deviation, but no worse.

6.1.2.1 Adding measurement error

In what we have done so far the observations have been assumed to be perfect. Let us now consider what happens if we explicitly take account of the fact that our observations will always contain some error, δ_i.

$$\phi_i^{'} = \phi_{ia}^{'} + \delta_i \tag{6.9}$$

Let's assume, as is usually reasonable, that the error is unbiased (zero mean) and ucorrelated with the true value, that is,

$$\overline{\phi_{ia}^{'}\delta_i} = 0 \tag{6.10}$$

and that the errors at the various stations where we have data are also uncorrelated

$$\overline{\delta_i \delta_j} = \begin{cases} \overline{\delta^2} \text{ for } i = j \\ 0 \text{ for } i \neq j \end{cases} \tag{6.11}$$

In this case, rather than **(6.5)**, we obtain

$$\epsilon = 1 - 2\sum_{i=1}^{N} p_i r_{gi} + \sum_{i=1}^{N}\sum_{j=1}^{N} p_i p_j r_{ij} + \eta \sum_{i=1}^{N} p_i^2 \tag{6.12}$$

where r_{ij} is the correlation between the two points and where η is the ratio of the error variance to the measurement variance - in other words, the *signal-to-noise ratio.*

$$\eta = \frac{\overline{\delta^2}}{\overline{\phi_g'^2}} \tag{6.13}$$

Minimization of the error leads to the condition

$$\sum_{j=1}^{N} r_{ij} p_j + \eta p_i = r_{gi} \quad \text{for } i = 1, 2, 3, \ldots, N \tag{6.14}$$

In this case the normalized error is

$$\epsilon = 1 \sum_{i=1}^{N}\sum_{j=1}^{N} p_i p_j r_{ij} + \eta \sum_{i=1}^{N} p_i^2 \tag{6.15}$$

6.1.2.2 What is the effect of including noise in the measurements?

In order to see how optimum interpolation treats the *a priori* information that the measurements include some error, it is instructive to compare the results **(6.14)** and **(6.15)** with the results **(6.5)** and **(6.7)** obtained assuming perfect data. In the case of perfect data, **(6.7)** gives

$$r_{ij} p_j = r_{gi} \quad \text{or } p_j = r_{ij}^{-1} r_{gi} \tag{6.16}$$

When noise is included we get, rather, the result **(6.14)**, which can be written

$$\{r_i j + \eta \mathbf{I}_{ij}\} p_j = r_{gi} \quad \text{or } p_j = \{r_i j + \eta \mathbf{I}_{ij}\}^{-1} r_{gi} \tag{6.17}$$

where $\mathbf{I}_{ij}$ is the unit matrix. Looking at the right-hand member of the pair of equations in **(6.17)**, it is easy to see that the coefficients p_j will be smaller when the error is large. This is most obvious if we assume that $r_i j$ is diagonal. Thus we see that the inclusion of error makes the coefficients in **(6.3)** smaller and that therefore, by **(6.2)**, the estimate we make will be closer to climatology. If we include error, then Optimum Interpolation will draw more closely to climatology or the first guess and tend to weight new observations less heavily. This is desirable. By putting different values of η_j along the diagonal, one can put information on the confidence one has in individual stations into the analysis scheme and weight more heavily those stations in which one has more confidence.

6.1.2.3 What do we need to make Optimum Interpolation work?

In order to make the above schemes work, we need the correlation matrices r_{ij} and r_{gi}. The first of these is easily calculable from observations, but the second is not since it involves correlations between the station points and the grid points. We do not have data at the grid points, or we would not need an analysis scheme. In practice, not even the r_{ij} are calculated in full generality. It is possible to assume that correlations between

points depend only on the distance between them and not on location or direction (although it would be possible to include directionally dependent (anisotropic) correlations). In this case the single isotropic correlation function can be estimated from station data. This is a crude approximation since correlations between stations depend on the location of the stations and whether longitude or latitude separates them. An example illustrating the anisotropy of correlation functions in 500 hPa geopotential heights is shown in **Fig. 6.2**.

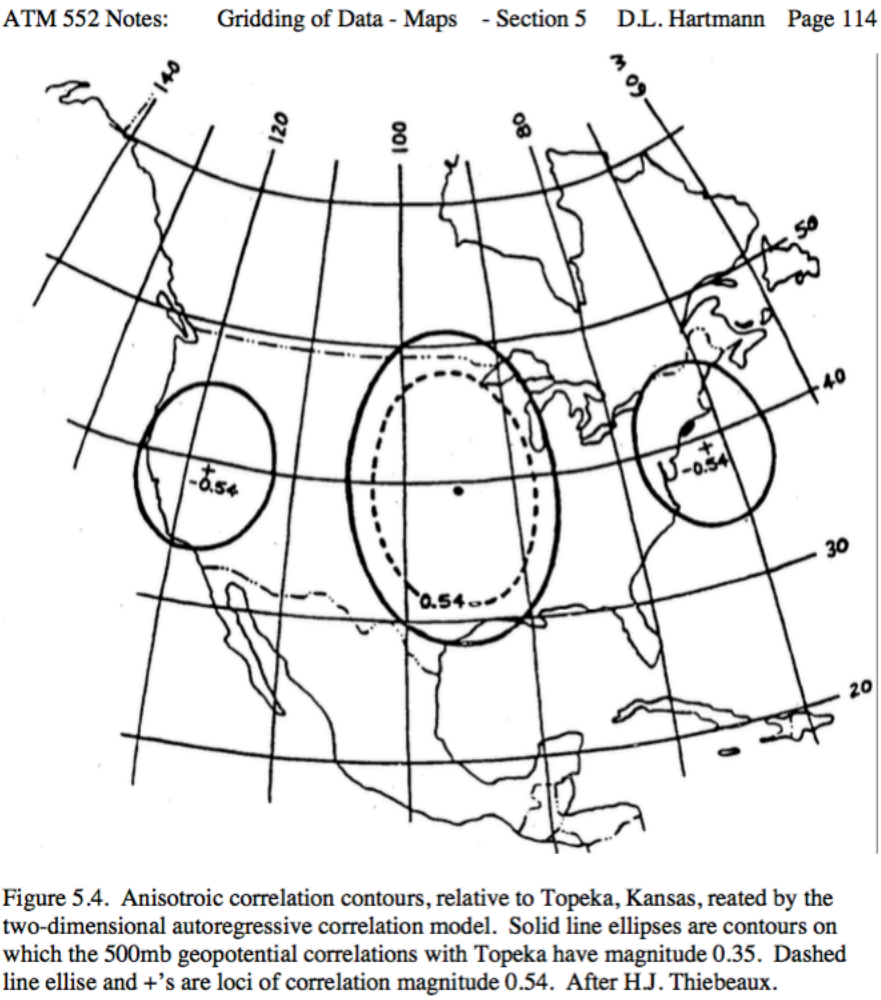

Figure 5.4. Anisotroic correlation contours, relative to Topeka, Kansas, reated by the two-dimensional autoregressive correlation model. Solid line ellipses are contours on which the 500mb geopotential correlations with Topeka have magnitude 0.35. Dashed line ellise and +'s are loci of correlation magnitude 0.54. After HJ. Thiebeaux.

Figure 6.2 Anisotropic correlation contours, relative to Topeka, Kansas, created by a two-dimensional autoregressive correlation model. Solid line ellipses are contours on which the 500 hPa geopotential correlations with Topeka have magnitude 0.35. Dashed lines denote the correlation of 0.54 and +'s are loci of correlation magnitude 0.54. After H.J. Thiebeaux.

Libby: *stopped at Dennis' Chapter 5.4*