OBJECTIVE MAPPING AND DESIGNING MAPPING ARRAYS

One useful application of linear estimation is "objective mapping". This procedure combines a finite number of of discrete observations (taken at particular times and places) into a map, or a continuous time series of maps, of the field under examination. This application is discussed here for two reasons: (a) it is a useful technique in its own right; (b) it provides a specific context in which to explore linear estimation.

1. Mapping scalar fields

First, consider the problem of mapping some scalar property such as surface temperature θ. Our objective is to use linear estimation and a number of data $\theta(n) = \theta(\mathbf{x}_n, t_n)$ to draw a continuous map of $\theta(\mathbf{x}, t)$. The first, and most crucial, step in any such mapping is to be very specific about what is to be mapped. Obviously it will not be possible to map every little detail of temperature. More likely, there are particular scales (in time and space) which are of interest and may span enough data to be accurately mapped. It would be stupid to employ a scheme which does its best at the impossible while being sub-optimal at the real purpose. Thus we must first *define* what is to be estimated. One typical objective is to map a smoothed temperature Θ defined as a filtered version of actual temperature θ such as

$$\Theta(\mathbf{x}, t) = \int d\xi \int d\tau \;\; \theta(\xi, \tau)\, W(\xi - \mathbf{x}, \tau - t) \tag{1}$$

where W is a smoothing filter centered on the point $\xi = \mathbf{x}$ and $\tau = t$. A simple filter is the multidimensional analog of the running mean filter where all values near $\mathbf{x},\ t$ are added with equal weight. In general, as W includes a larger $\xi,\ \tau$ volume, resolution is lost (small scales are filtered out) but the mapping accuracy is improved because more observations are included within W's range.

With the estimand (here Θ) defined, the mapping procedure must be selected to optimize specific properties of the estimate. The procedure should produce a small error $\hat{\Theta} - \Theta$ but there is no way of knowing this error without knowing the answer. All that can be asked is that the procedure, when applied to many sets of data, usually produce a small error, say have minimum mean (over many mapping exercises) square error. The minimum MSE estimate is

$$\hat{\Theta}(\mathbf{x}, t) = \boldsymbol{\theta} \cdot \mathbf{D}^{-1} \cdot \langle \theta \Theta(\mathbf{x}, t) \rangle \qquad \text{or} \qquad \hat{\Theta}(\mathbf{x}, t) = \sum_n \theta(n) \sum_m \mathbf{D}^{-1}(n, m) \langle \theta(m) \Theta(\mathbf{x}, t) \rangle \tag{2}$$

where the required statistics can be computed from the 2-point, 2-time covariance $C(\mathbf{x}, \mathbf{y}, t, s) = \langle \theta(\mathbf{x}, t)\, \theta(\mathbf{y}, s) \rangle$ according to

$$D(n, m) = C(\mathbf{x}_n, \mathbf{x}_m, t_n, t_m) \qquad b(m) = \langle \theta(m) \Theta(\mathbf{x}, t) \rangle = \int d\xi \int d\tau\, W(\xi - \mathbf{x}, \tau - t)\, C(\mathbf{x}_n, \xi, t, \tau) \tag{3}$$

Note how $\langle \theta \Theta \rangle$ is a covariance of linear operations on θ and is, therefore, computable as an operation on C. The components of the estimate (2) have physical meaning: the matrix $\mathbf{D}$ describes how the data are related to each other (Are they redundant?) while $\mathbf{b}$ describes how the data are related to the estimand Θ.

To employ (2) one needs to know the mean product of θ at two places and times. Recall that the average $\langle\ \rangle$ is meant to be over many repetitions of the map drawing exercise. This is meaningless unless some class of position-time pairs are statistically similar to each other so that the average can be over more than one example. Most commonly time averages are employed, saying that the procedure is to be optimized for data taken with the same relative time separations but at

arbitrary absolute times. Then $C(\mathbf{x},\mathbf{y},t,s) = C(\mathbf{x},\mathbf{y},t-s)$ can be computed by averaging over time holding $t-s$ fixed. This is, in principle, sufficient to define C if there are many observations at all possible positions. It might also be that, over the scales of interest, statistics (now meaning time averages) depend weakly on position (are homogeneous) so that $C(\mathbf{x},\mathbf{y},t-s) = C(\mathbf{x}-\mathbf{y},t-s)$ can be computed by averaging over positions while holding $\mathbf{x}-\mathbf{y}$ fixed. This makes it possible to implement the mapping procedure with many fewer observations. It might further be that the field is approximately isotropic so that the covariance of two θs depends only on their time separation and the distance between them. Then $C(\mathbf{x}-\mathbf{y},t-s) = C(|\mathbf{x}-\mathbf{y}|,t-s)$ and yet fewer observations are needed to specify the requisite statistics. The statistical assumptions about C define the kind of field for which the mapping procedure will be optimal. As the assumption are less valid the procedure becomes less optimal but hopefully it will still be good. In any case, to improve it one would need to know how the assumptions are violated; if this is not known then the procedure is optimal.

In applying the minimum MSE technique it is essential that the estimated mean product, $C(\mathbf{x},\mathbf{y},t-s)$ be a possible mean product. In particular, all $D(n,m) = C(\mathbf{x}_n,\mathbf{x}_m,t_n-t_m)$ must be positive semi-definite matrices. By this we mean that the mean square of any linear combination of θs must be nonnegative, that is

$$q = \sum_m \alpha(m)\theta(m), \quad \langle q^2\rangle = \sum_m\sum_n \alpha(n)\langle\theta(n)\theta(m)\rangle\alpha(m) = \boldsymbol{\alpha}\cdot\mathbf{D}\cdot\boldsymbol{\alpha} \ge 0 \tag{4}$$

for all $\boldsymbol{\alpha}$ α. If $\boldsymbol{\alpha}\cdot\mathbf{D}\cdot\boldsymbol{\alpha} \ge 0$ for any $\boldsymbol{\alpha}$ then $\mathbf{D}$ is said to be a positive semi-definite matrix. From a singular value decomposition of the square matrix $\mathbf{D}$ it is seen that this means $\mathbf{D}$ must have no negative eigenvalues. Not all functions $C(\mathbf{x},\mathbf{y},s)$ have this property and if objective mapping is attempted with one that does not the result will likely be garbage. The mapping estimator is based on extremizing $\langle(\hat{y}-y)^2\rangle$; if D_{nm} has negative eigenvalues the extremum is not the minimum and an α which extremizes the MSE at a large negative value may be chosen.

In the case where θ is homogeneous and stationary the easiest way to insure that C is a possible mean product is to insure that it corresponds to a positive semi-definite spectrum (*i.e.* no negative spectral values). Recall that for stationary data the power spectrum of a time series is proportional to the eigenvalue spectrum of the time-lagged covariance. The same things works in many dimensions. Let

$$\theta(\mathbf{x},t) = \sum_\omega\sum_k a(\omega,\mathbf{k})\ \exp(i\omega t + i\mathbf{k}\cdot\mathbf{x}) \tag{5}$$

$$C(\mathbf{x},\mathbf{y},t,s) = \sum_\omega\sum_k\sum_\sigma\sum_l a(\omega,\mathbf{k})a^*(\sigma,\mathbf{l})\ \exp[i\omega(t-s)+i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})]\ \exp[i(\omega-\sigma)s+i(\mathbf{k}-\mathbf{l})\cdot\mathbf{y}] \tag{6}$$

If this is to depend only on $t-s$ and $\mathbf{x}-\mathbf{y}$ then the mean product $\langle a(\omega,\mathbf{k})a^*(\sigma,\mathbf{l})\rangle$ must vanish except when $\sigma=\omega$ and $\mathbf{k}=\mathbf{l}$ so

$$C(\mathbf{x},t) = \sum_\omega\sum_{\mathbf{k}}\langle|a(\omega,\mathbf{k})|^2\rangle\exp(i\omega t + i\mathbf{k}\cdot\mathbf{x}) \tag{7}$$

Thus if the wavenumber-frequency spectrum made from $|a(\omega,\mathbf{k})|^2$ is positive, $C(\mathbf{x},t)$ is a possible covariance and will yield positive semi-definite matrices $D(n,m)$ regardless of the sampling positions and times.

One practical procedure is to express $C(\mathbf{x},t)$ in terms of a set of prescribed functions of $\mathbf{x}$ and t, constrain this combination to adjust the representation to fit observed sample mean products and

have a nonnegative Fourier transform (which is the spectrum). Of course, one should make sure that the sampling error of these mean products is small enough to give reliable estimates and not to insist that the function reproduce every sample mean product exactly. When the process is not stationary and homogeneous one procedure is to generate the covariance as a sum of the form

$$C(\mathbf{x},\mathbf{y},t,s)=\sum_n \lambda_n^2 V_n(\mathbf{x},t)V_n(\mathbf{y},s) \tag{8}$$

where the functions V and the λs are adjusted so C approximates the observed mean products. This EOF-like structure of C insures that it is a possible covariance and thus leads to appropriate $\mathbf{D}$s.

Objective analysis is a great way to draw maps, but its real virtue is that quantitative estimates of mapping error can be made. The MSE takes a familiar form:

$$\langle(\hat{\Theta}-\Theta)^2\rangle=\langle\Theta^2\rangle\left[1-\langle\Theta^2\rangle^{-1}\langle\Theta\boldsymbol{\theta}\rangle\cdot\mathbf{D}^{-1}\cdot\langle\boldsymbol{\theta}\Theta\rangle\right] \tag{9}$$

The last term is called the **squared multiple correlation** between Θ and the set of data θ. If the θ are all uncorrelated it is simply the sum of the squared correlations between Θ and each of the $\theta(n)$.

Note that the MSE (9) depends only on statistics, not actual observations. Thus if the mean product C is known then the things that determine MSE are (a) the location (and time) at which Θ is to be estimated, (b) the filter W used to define Θ, and (c) the location of the observations. For a given array (the times and locations of observations) it is then easy to determine mapping accuracy. It will depend on the filter and the location where Θ is to be estimated. If no filtering is applied, mapping error will vanish at measurement locations and will grow away from these points. If a smoothing filter is applied, the estimate will never equal the observation (since Θ and θ are different quantities), there will be some error even at measurement locations, but the error will grow more slowly as a function of distance from the observations. Again, if the objective is clearly specified then performance can be determined quantitatively.

Because mapping error depends only on the statistics of the observations and the field to be mapped, it is practical to use objective mapping to design sampling arrays and determine their performance before observations are made (assuming the statistics can be determined from existing data). To do this the objective and statistics are used to optimize the performance of candidate arrays. The mapping error is examined and the array geometry changed until a desired balance of accuracy and mapping domain size is achieved. Examples of this are provided by Bretherton, Davis and Fandry (1976) in Deep-Sea Res. *23*, 559-582.

2. Unbiased estimates

Objective mapping assumes an estimate of the form

$$\hat{\Theta}=\boldsymbol{\alpha}\cdot\boldsymbol{\theta} \tag{10}$$

and requires knowing the mean product of the field θ. If the mean value of θ does not vanish, this estimate is sub-optimal. To account for mean values an additional constant should be appended to the model and adjusted just like the αs; a simple way of doing this is to add a constant as one component of $\boldsymbol{\theta}$. Regardless of the method, both the mean value and the variance of θ must be known . Unfortunately, in the real world energetic low frequency variability is the rule rather than the exception, and the mean is not well defined nor easily measured. For example, in mapping temperature would the mean be over several ice-age cycles or over recent time be more appropriate?

Because precise definition of mean and variance do not seem central to mapping over a finite time span it, is intuitively appealing to see that good maps can be drawn without this information. Formally, it is assumed that θ is statistically stationary and homogeneous so that the mean and variance are constants. Then the bias of the estimate can be made to vanish by constraining the sum of the weights, $\alpha(n)$, to be unity so that

$$\langle\hat{\Theta}\rangle = \sum_n \alpha(n)\langle\theta(n)\rangle = \langle\theta\rangle \sum_n \alpha(n) = \langle\theta\rangle. \tag{11}$$

Minimizing the MSE subject to the constraint $\sum\alpha(n) = 1$ is accomplished by modifying the extremization operation using a Lagrange multiplier, λ. Thus we seek the minimum, with respect to variations of α, of

$$\langle[\hat{\Theta}-\Theta]^2\rangle = \sum_n\sum_m \alpha(n)\langle\theta(n)\theta(m)\rangle\alpha(m) - 2\sum_m \alpha(m)\langle\theta(m)\Theta\rangle + \langle\Theta^2\rangle + 2\lambda[\sum_m \alpha(m) - 1] \tag{12}$$

Extremizing this gives $\alpha(n)$ as a function of λ from

$$\sum_m \langle\theta(n)\theta(m)\rangle\alpha(m) = \langle\Theta\theta(n)\rangle - \lambda\, u(n) \tag{13}$$

where $\mathbf{u}$ is a vector with unit elements, *i.e.* $u(n) = 1$. Adjustment of λ to meet the constraint $\sum\alpha = 1$ gives

$$\lambda = -\frac{1 - \sum_{n,m} u(n)D^{-1}(n,m)\langle\theta(m)\Theta\rangle}{\sum_{n,m}\ u(n)\ D^{-1}(n,m)\ u(m)} \tag{14}$$

where $\mathbf{D}^{-1}$ is the inverse of the matrix $D(n,m) = \langle\theta(n)\theta(m)\rangle$. This leads to

$$\boldsymbol{\alpha} = \mathbf{D}^{-1}\cdot\left[\mathbf{b} + \mathbf{u}\,\frac{1 - \mathbf{u}\cdot\mathbf{D}^{-1}\cdot\mathbf{b}}{\mathbf{u}\cdot\mathbf{D}^{-1}\cdot\mathbf{u}}\right]. \tag{15}$$

where $\mathbf{b} = \langle\boldsymbol{\theta}\Theta\rangle$ as below (2). The first term will be recognized as the weight for the unconstrained estimate (2). The extra terms with $\mathbf{u}$ look messy but are not; for example $\mathbf{u}\cdot\mathbf{D}^{-1}\cdot\mathbf{u}$ is simply the sum of all the elements in the inverse of $\mathbf{D}$.

Since no constant was added to the data set θ, this procedure clearly does not involve knowing $\langle\theta\rangle$. Less obviously, the unbiased estimator also does not require knowledge of the variance. To see this note that mean products, like $\langle\theta(n)\theta(m)\rangle$, can be expressed as

$$\langle\theta(n)\theta(m)\rangle = -\frac{1}{2}\langle[\theta(n) - \theta(m)]^2\rangle + \frac{1}{2}[\langle\theta^2(n)\rangle + \langle\theta^2(m)\rangle] \tag{16}$$

Statistics of the form $[\theta(n) - \theta(m)]^2$ are called **structure functions** and involve only changes of θ; thus they do not depend on knowing the mean (which subtracts out) or the variance. Substitution of (16) (and the analogous expression for $\langle\theta(m)\Theta\rangle$) show that when $\sum\alpha = 1$ the MSE is

$$\langle[\hat{\Theta} - \Theta]^2\rangle = -\frac{1}{2}\sum_n\sum_m \alpha(n)\langle[\theta(n) - \theta(m)]^2\rangle\alpha(m) + \sum_m \alpha(m)\langle[\theta(m) - \Theta]^2\rangle + 2\lambda[\sum_n \alpha(n) - 1]. \tag{17}$$

The variances of θ and Θ subtract out of this so that the optimal weights and the MSE (for this constrained procedure) do not depend on the mean or the variance. The constrained map will have higher error (since the αs which exactly minimize MSE don't add to unity) but this is usually a small price to pay for having a good procedure for any mean and variance, especially when these can't be well determined.

3. An example

Some insight into the behavior of different mapping procedures can be obtained from the following simple example. Let $\theta(t)$ be a stationary process with covariance $\exp(-|t|)$ and suppose we have two observations $x(1) = \theta(-1)$ and $x(2) = \theta(1)$. The data-data mean product

$$D(1,1) = D(2,2) = 1, \qquad D(1,2) = D(2,1) = e^{-2} \tag{18}$$

has inverse

$$D^{-1}(1,1) = D^{-1}(2,2) = (1-e^{-4})^{-1}, \qquad D^{-1}(1,2) = D^{-1}(2,1) = -e^{-2}/(1-e^{-4}) \tag{19}$$

and

$$\langle x(1)\theta(t)\rangle = e^{-|t+1|}, \quad \langle x(2)\theta(t)\rangle = e^{-|t-1|} \tag{20}$$

It is now relatively simple to examine the performance of three mapping methods. First is the straight minimum MSE estimate of θ using weights

$$\alpha(n) = \sum_m D^{-1}(n,m)\langle\theta(t)x(m)\rangle \tag{21}$$

whose MSE is plotted as E_1 in Figure 1. The second procedure is the unbiased estimate based on minimizing MSE while holding $\sum\alpha = 1$. The weights $\alpha(n)$ are given by (15) and the MSE is plotted as the curve E_2. The final case is an unconstrained estimate of a smoothed field

$$\Theta(t) = \int_{t-0.5}^{t+0.5} \theta(t')dt' \tag{22}$$

A little integration of $\langle\theta(0)\theta(t')\rangle$ provides $\langle\theta(0)\Theta(t)\rangle$ from which the weights and MSE can be found. The MSE is the plot E_3.

There are three general observations which this example typifies:
(a) The unconstrained estimate has lower error than the constrained unbiased estimate **if** the mean values (here all zero) are known. Of course the comparison can not be made if a mean is not assumed and if the assumption is wrong the unconstrained model may have the poorer performance.
(b) The smoothed field estimate is in error even at the data points ($t = \pm 1$) because the measurements are of θ not Θ.
(c) Away from the data points, the error in estimating Θ is smaller than the error in estimating θ; this simply shows that when the rapid fluctuations are filtered out, the smooth remainder becomes easier to estimate. Remember that $\hat{\Theta}$ could be obtained by filtering $\hat{\theta}$ and this filtering would remove some of the error too. If Θ were measured directly, its estimate would everywhere be more accurate than the estimate of θ.

4. Vector fields

A special case of objective mapping worth examination is mapping a vector field, say the two-dimensional $\mathbf{U}(\mathbf{x})$ where, for simplicity we treat different times as separate realizations. The fact that two components components are to be predicted adds nothing new but notation and converts (2) to

$$\vec{\hat{U}}(\vec{x}) = \mathbf{u}\cdot\mathbf{D}^{-1}\cdot\langle\mathbf{u}\,\vec{U}(\vec{x})\rangle \tag{23}$$

where $\mathbf{u}$ is a collection of observations of $\vec{U}$ taken at particular positions (which make up the observational array). The notation here is that an observation of $\vec{U}(\vec{x}_1)$ contributes two elements

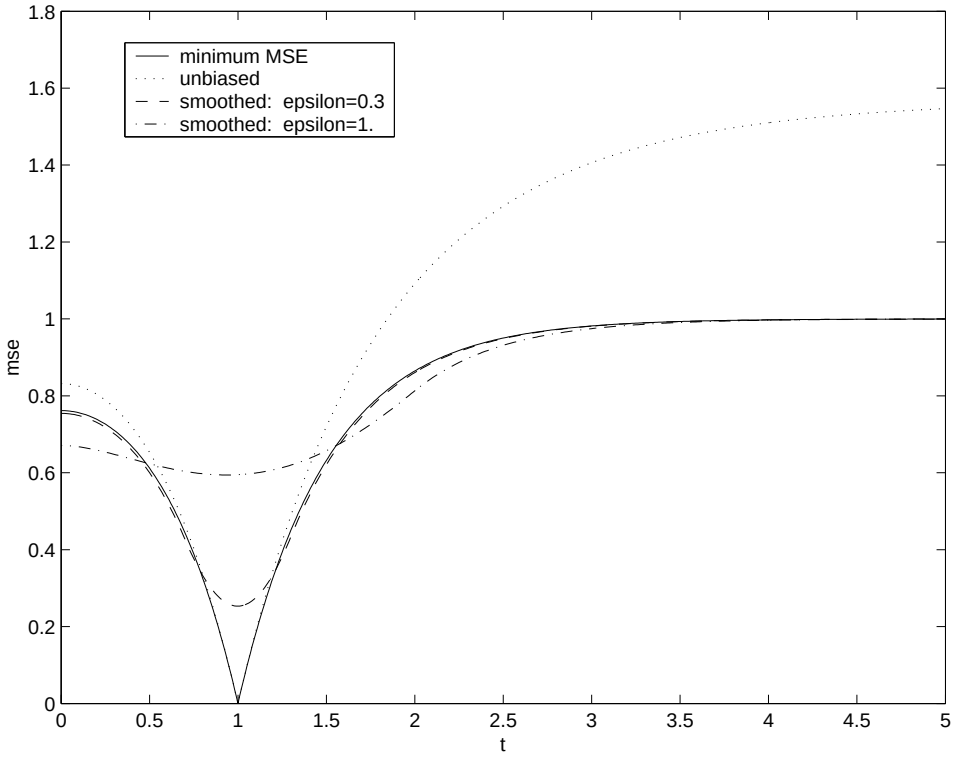

Figure 1: The effect of smoothing and zero-bias constraint on optimal interpolation error. The data are two values at $t = \pm 1$. $\langle\theta\rangle$ and $\theta(t)\theta(s)\rangle = \exp(-|t-s|)$. Solid curve is Mean Square Error (MSE) for the minimum MSE (possibly biased) estimator. Upper (dotted) curve is MSE for the optimal unbiased estimator. The other curves show MSE for estimating a smoothed field $\Theta(t) = \frac{1}{2\epsilon}\int_{t-\epsilon}^{t+\epsilon}\theta(s)\,ds$ for $\epsilon = 0.3$ and $\epsilon = 1.0$.

of $\mathbf{u}$, namely $u(1) = U_x(\vec{x}_1)$ and $u(2) = U_y(\vec{x}_1)$. The new thing here is dealing with vector data and the way they appear in $\mathbf{D}$.

The elements of $\mathbf{D}$ and $\langle\mathbf{u}\,\vec{U}(\vec{x})\rangle$ involve mean products or, if the means are removed covariances, of the form $\langle U_i(\vec{x}_0)U_j(\vec{x}_1)\rangle$ where the subscripts i and j indicate directions. You will find other special forms of tensors of various orders in Chapter 3 of G.K. Batchelor's *The theory of homogeneous turbulence* or any good text covering tensors.

It would be rare to have enough information to map out this function over all pairs of positions put the problem is closer to feasible if the field is homogeneous, in which case $\langle U_i(\vec{x}_0)U_j(\vec{x}_0+\vec{r})\rangle = \langle U_i(0)U_j(\vec{r})\rangle$ and observations from different locations but the same separation are equivalent. There is still a lot of information here because the covariance is a function of a vector argument.

The simplest case is when the vector field is **isotropic**, that is all directions are equivalent just as all positions are equivalent in a homogeneous field. There is still directional information in the covariance since both the velocity $\mathbf{U}$ and the separation vector $\vec{r}$ are vectors and their relative orientation can affect the velocity covariance. Tensor analysis shows that all isotropic second order tensors, such as the dyad $\langle\vec{U}\vec{V}\rangle$, take a simple form which for the case of interest here is

$$\langle U_i(0)U_j(\vec{r})\rangle = \gamma_i\gamma_j[R(r) - S(r)] + \delta_{ij}S(r) \tag{24}$$

where γ_i is the cosine of angle between the ith component of $\vec{U}$ and $\vec{r}$, δ_{ij} is the Kronecker Delta, and R and S are general functions of r, the magnitude of $\vec{r}$. Here $R(r)$ is the covariance of the

velocity components parallel to the separation vector (called the longitudinal covariance) and $S(r)$ is the covariance of the components perpendicular to $\vec{r}$ (called the transverse covariance). In the special case of the covariance of a velocity field there are restrictions on R and S. Because these are essentially covariances of scalars they must result in a positive semi-definite covariance matrices between observations taken from all possible sampling arrays. This is equivalent to saying they must have nonnegative Fourier (wavenumber) spectra, just as all time-lagged covariances must lead to nonnegative frequency spectra. Necessary conditions for these functions are $R(0) = S(0)$ and that both functions be smaller than $R(0)$ for $r > 0$.

It is more helpful to write (24) as

$$\langle U_i(0)U_j(\vec{r})\rangle = \frac{r_i r_j}{r^2}[R(r) \; - \; S(r)] \; + \; \delta_{ij}S(r) \tag{25}$$

As you might suppose, if the vector field is nondivergent there is a relation between R and S which can be determined by computing $\sum_j \partial/\partial r_j \langle U_i(0)U_j(\vec{r})\rangle$ using (25) and requiring it to vanish. In doing this note that

$$\frac{\partial}{\partial r_n}R(r) = \frac{r_n}{r}\frac{dR(r)}{dr} \qquad \frac{\partial}{\partial r_m}\frac{r_n r_m}{r^2}R(r) = (N-1)\frac{r_n}{r^2}R(r) + \frac{1}{r}\frac{dR(r)}{dr} \tag{26}$$

where N is the dimensionality of the space spanned by $\mathbf{r}$. This gives $S(r) = R(r) + \frac{r}{N-1}\frac{dR(r}{dr}$. Thus the statistics of a homogeneous, isotropic and nondivergent vector field are determined by a single function.

5. Testing models with data

One use of data is to test the differential equations used in analytical or numerical models. For example, one might attempt to verify geostrophy by comparing maps of pressure and velocity. Or the assumption of approximate horizontal velocity nondivergence might be tested directly from velocity maps. This raises a subtle but very crucial point about objective mapping in particular and model testing in general.

A great deal of the structure of objective maps is built in by the statistics used in the mapping procedure. This follows because the minimum MSE estimator of Y based on the data $\mathbf{q}$ can be written as either

$$\hat{y} = \boldsymbol{\alpha} \cdot \mathbf{q} \qquad \text{or} \qquad \hat{y} = \boldsymbol{\beta} \cdot \mathbf{q}y \tag{27}$$

where, if the data-data product matrix is $D(n,m) = \langle q(n)q(m)\rangle$,

$$\boldsymbol{\alpha} = \mathbf{D}^{-1} \cdot \langle \mathbf{q}y\rangle, \qquad \boldsymbol{\beta} = \mathbf{q} \cdot D^{-1}. \tag{28}$$

This shows why linear operations on y (such as filtering or differentiation) are optimally estimated by applying that operation to $\hat{y}$. It also shows the subtle way that $\langle \mathbf{q}y\rangle$ builds properties into the estimate. Say, for example, that estimates $\hat{u}$ and $\hat{v}$ of the two velocity components were made from the same data $\mathbf{q}$. These estimates would be

$$\hat{u}(\mathbf{x}) = \boldsymbol{\beta} \cdot \langle \mathbf{q}\, u(\mathbf{x})\rangle, \qquad \hat{v}(\mathbf{x}) = \boldsymbol{\beta} \cdot \langle \mathbf{q}\, v(\mathbf{x})\rangle, \tag{29}$$

where $\boldsymbol{\beta}$ is the same for both. Note that $\hat{\mathbf{u}}$ will be nondivergent if the statistics used to draw the map obey $\langle \mathbf{q}\boldsymbol{\partial_x u}\rangle + \langle \mathbf{q}\boldsymbol{\partial_y v}\rangle = \mathbf{0}$, regardless of what the observations $\mathbf{q}$ are. There are two lessons:
(a) The statistics assumed in order to map data can affect the conclusions as much as the data do. In fact, one could map a given data set using different statistical descriptions, each obeying

different “dynamics”, and verify any number of mutually contradictory hypotheses. Thus one must be extremely careful that the test outcome is not determined by the test methodology.
(b) The test of hypotheses is often carried out much more clearly and accurately on the statistics of all available data than on a particular subset of the data suitable for a deterministic test. For example, rather than mapping $\mathbf{u}$ one might do better to determine $\langle \mathbf{u}(\mathbf{y})\mathbf{u}(\mathbf{x})\rangle$ directly from the data and see if $\boldsymbol{\nabla}_x \cdot \langle \mathbf{u}(\mathbf{y})\mathbf{u}(\mathbf{x})\rangle = \mathbf{0}$.