A physically–motivated index of sub–gridscale pattern

Tim E. Jupp[1,2] and Sean D. Twiss[3]

Abstract. Earth Observation data on land surface properties such as albedo is typically collected at a pixel resolution of 1 km or less. Global climate models, on the other hand, are constrained by current limits on computing power to run at a gridbox resolution of 10 km or more. This mismatch in spatial scales means that large amounts of pixel–scale information are condensed into a small number of gridbox–scale summary statistics before use in global climate models. Subgridscale patterns in land surface properties may have a significant effect but the summary statistics currently in use – such as the gridbox mean and gridbox variance – are insensitive to the spatial arrangement of pixels. To address this gap in the information available to global climate models we define here a new gridbox–scale summary statistic – the Laplacian pattern index – that is sensitive to the spatial arrangement of pixels. This dimensionless index is based on the mean–squared value of the Laplacian filter ∇^2 within the gridbox, and is motivated by the physics of diffusive heat transport. We investigate the value of the index in some simple cases and show that it is a measure of the local correlation structure within a gridbox. This allows us to generate random gridboxes in which the index takes (on average) a prescribed value. The Laplacian pattern index is designed to be a useful measure of subgridscale pattern in numerical climate models, but can be used as a measure of pattern in any two–dimensional array of real–valued data.

1. Introduction

The aim of this paper is to enhance the flow of useful information from remotely–sensed observations of land–surface properties to numerical simulations of climate. Whilst the former are available at a relatively fine pixel scale $(\Delta x) \sim 1$ km or less, the latter must be undertaken at a relatively coarse gridbox scale $N(\Delta x) \sim 10-100$ km [e.g., *Essery et al.*, 2003]. It is therefore necessary for the observed pixel–scale land–surface properties to be summarized at the gridbox–scale before they can be used in climate models.

Subgridscale patterns in land–surface properties may have a significant effect on climate–atmosphere interactions [e.g., *Giorgi and Avissar*, 1997; *Koster and Suarez*, 1992; *Molod and Salmun*, 2002]. For example, it is reasonable to expect that a boundary between two land–surface types (e.g. forest and grassland) will have some localized effect on land–atmosphere fluxes. This edge effect will be confined to the region within some critical lengthscale of the boundary. The typical lengthscale of any pattern in land–surface properties is then of crucial importance [e.g., *Blyth*, 1995; *Raupach and Finnigan*, 1995; *Mahrt*, 2000]. If the lengthscale is sufficiently large then edge effects are relatively unimportant and each land–surface type interacts essentially independently with the atmosphere. In this situation a 'tile' (also known as 'mosaic') model for the land–surface is appropriate, with a separate surface energy balance being applied for each land–surface type. On the other hand, if the pattern lengthscale is sufficiently small then edge effects dominate and a 'mixture' land–surface model – in which the surface energy balance is solved for the mixed land–surface – is required [*Koster and Suarez*, 1992]. It should therefore be possible, using remote sensing, to determine whether a mosaic or mixture strategy is appropriate for each gridbox in a numerical climate model. In order to do so a numerical measure of subgridscale pattern is required.

There is no shortage of pattern metrics in existence for data in categorical form. For data of this type the pixels are assigned to qualitative classes and pattern can be analyzed with tools including fractal dimension, typical patch size and area:perimeter ratio [e.g., *McGarigal and Marks*, 1995; *Gustafson*, 1998; *Saura and Martínez-Millán*; 2001]. In this paper, however, the aim is to define a pattern metric for real–valued (quantitative) data. For data of this type geostatistical tools such as the variogram and correlogram provide measures of spatial correlation structure, while Fourier and wavelet analysis yield insight into the typical spatial scales of variation [*Cressie*, 1991]. The Laplacian pattern index defined here is designed to complement these geostatistical techniques. The difference is that it is motivated explicitly by consideration of the physics of diffusive processes.

As a particular example, consider the case in which the Earth Observation data consist of pixel–scale values of the co–albedo α within a gridbox, as in Figure 1. (The co–albedo α is defined by $\alpha = 1 - a$ where a is the albedo. It follows that the fAPAR – the fraction of absorbed photosynthetically active radiation – is the co–albedo for light of a particular color.) Let $\langle . \rangle$ denote the area average of pixel–scale values within the gridbox. Currently, the gridbox mean μ_0 (and occasionally the gridbox variance σ_0^2) of the pixel–scale values are used as gridbox–scale summaries of the land-surface properties:

$$\mu_0 = \langle \alpha \rangle, \quad \sigma_0^2 = \mathrm{var}(\alpha) = \langle \alpha^2 \rangle - \langle \alpha \rangle^2 \tag{1}$$

(Here and subsequently the subscript 0 refers to the original gridbox. An absence of subscripts will imply a generic

[1]Climate and Land-Surface Systems Interaction Centre, CEH Monks Wood, Huntingdon, Cambridgeshire, PE28 2LS, U.K.

[2]Climate and Land-Surface Systems Interaction Centre, CEH Wallingford, Crowmarsh Gifford, Oxfordshire, OX10 8BB, U.K.

[3]Climate and Land-Surface Systems Interaction Centre, Institute of Ecosystem Science, School of Biological and Biomedical Sciences, University of Durham, South Road, Durham DH1 3LE, U.K.

0148-0227/06/$9.00

expression with no particular gridbox being implied.) The statistics μ_0 and σ_0^2 are clearly useful, but both are insensitive to the spatial arrangement of pixels within a gridbox – that is, they are invariant under pixel shuffling. (We use the term shuffling to denote a random reordering of the pixels within a gridbox, with all pixels being regarded as distinguishable.) It follows that the two gridboxes in Figure 1 - which may interact very differently with the atmosphere - would be treated as equivalent by a climate model whose information was limited to the gridbox mean μ_0 and the gridbox variance σ_0^2. The purpose of this paper is to develop an additional, physically meaningful gridbox–scale statistic that is not invariant under pixel shuffling.

The structure of this paper is as follows. Section 2 contains a heuristic motivation for the Laplacian pattern index, using the particular example of energy balance at the earth's surface. This leads to the operational definition of the Laplacian pattern index in Section 3. Section 4 contains a discussion of the interpretations that can be placed on the index. To this end the index is calculated for some simple deterministic and random patterns as well as for some real remote sensing data. Conclusions are presented in Section 5.

2. Physical motivation

The motivation for the Laplacian pattern index comes from considering the surface energy balance of a pixel with co–albedo α. In particular it derives from an analysis of how the surface energy balance of a gridbox is changed when diffusive lateral fluxes between the pixels are suddenly switched on.

In the absence of lateral fluxes, the energy balance for each pixel is independent of the other pixels and can be written as:

$$\alpha R = A + BT$$

where R is the incoming solar radiation (in W m^{-2}) and T is the surface temperature [*North*, 1981]. A and B are constants in a linearization of the outgoing vertical energy flux.

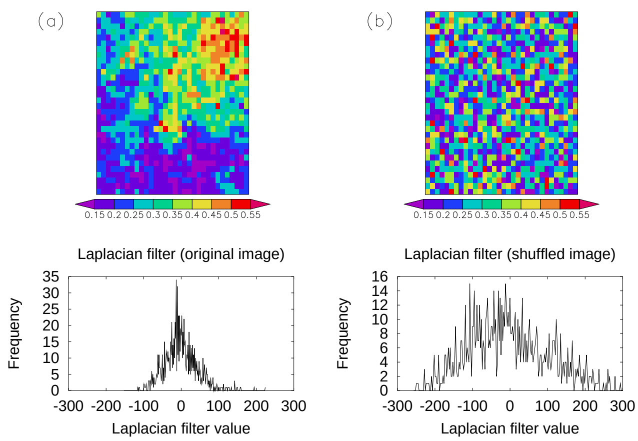

Figure 1. (a) Values of fAPAR α in an original (unshuffled) gridbox, and the corresponding distribution of Laplacian Filter values $\nabla^2\alpha$, which has variance s_0^2. (Data from the gridbox containing Cambridge, U.K., indicated by a cross in Figure 5b). (b) The same pixels, randomly shuffled and the subsequent effect on the distribution of $\nabla^2\alpha$. Pixels in the original gridbox are positively correlated with their neighbors, those in the shuffled gridbox are essentially uncorrelated. The unshuffled variance s_0^2 is therefore smaller than the shuffled variance s^2. The Laplacian pattern index λ_0 is based on the ratio of these two variances.

It follows that – in the absence of lateral fluxes – surface temperature is a linear function of co–albedo. Now imagine that the presence of the atmosphere allows lateral fluxes to be set up between neighboring pixels with different surface temperatures. For simplicity, suppose that these fluxes are of the form $-k\nabla T$ where k is a constant and ∇ denotes the horizontal spatial gradient. The energy balance equation for the rate of change of surface temperature becomes

$$C\dot{T} = \frac{R\alpha - A}{B} + k\nabla^2 T$$

where C is an area heat capacity in J m^{-2} K^{-1}. The term $k\nabla^2 T$ represents accumulation of energy due to lateral fluxes from neighboring pixels. Since the temperature is – when the lateral fluxes are first switched on – a linear function of the co–albedo, this motivates the use of the Laplacian filter $\nabla^2\alpha$ as a physically-meaningful parameter for each pixel. An interpretation of $\nabla^2\alpha$ is that it is (proportional to) *the pixel–scale energy accumulation due to lateral fluxes at the moment when they are first switched on.* To move from this pixel–scale variable to a gridbox–scale parameter it is helpful to consider the distribution of values of $\nabla^2\alpha$ within a gridbox.

At first sight one might expect that the mean value within a gridbox $\langle\nabla^2\alpha\rangle$ would prove useful since it is proportional to the mean energy accumulation per pixel. It follows from the divergence theorem, however, that $\langle\nabla^2\alpha\rangle \approx 0$. (The physical explanation is as follows: the energy accumulation due to lateral fluxes is positive at some pixels and negative at others. Every packet of energy exported by one pixel is imported by another and so the average energy accumulation per pixel is zero). Since the first moment of the distribution is unhelpful, attention can be turned to the second moment $\langle(\nabla^2\alpha)^2\rangle$ which quantifies the mean–squared energy accumulation per pixel. Specifically, consider the observed variance s_0^2 of the distribution of $\nabla^2\alpha$ in the original gridbox:

$$s_0^2 = \mathrm{var}(\nabla^2\alpha) = \langle(\nabla^2\alpha)^2\rangle - \langle\nabla^2\alpha\rangle^2 \qquad (2)$$

The standard deviation s_0 can be interpreted as being (proportional to) the root–mean–square energy accumulation per pixel due to lateral fluxes.

3. The Laplacian pattern index λ

A gridbox scale index of pattern can now be derived, inspired by the physical motivation in the previous section. Suppose that the pixel-level data α in a gridbox are known and that the pixels are square with size (Δx). (It will be shown later that one can set $(\Delta x) = 1$ without affecting the value of the index.) The Laplacian filter $\nabla^2\alpha$ at each pixel α_j is

$$\left(\nabla^2\alpha\right)_j = \frac{-4\alpha_j + \alpha_N + \alpha_S + \alpha_E + \alpha_W}{(\Delta x)^2} \qquad (3)$$

where the neighboring pixels $\alpha_N, \alpha_S, \alpha_E, \alpha_W$ occupy the north, south, east and west positions about the central pixel α_j in the 5–point Laplacian stencil:

$$\begin{array}{ccc} & \alpha_N & \\ \alpha_W & \alpha_j & \alpha_E \\ & \alpha_S & \end{array} \qquad (4)$$

(The 5–point stencil can be applied at edge pixels by imposing cyclic boundary conditions on the gridbox.) This yields a distribution of values of $\nabla^2\alpha$ within the gridbox as shown in Figure 1a. The variance of this distribution is then given by equation 2. In general, small values of s_0^2 correspond to structured gridboxes in which pixels are positively correlated with their neighbors and hence lateral fluxes are relatively small. Unstructured gridboxes in which pixels are uncorrelated with their neighbors produce larger lateral fluxes and hence larger values of s_0^2. The special case in which pixels are negatively correlated with their neighbors produces the largest values of s_0^2. (This is discussed in more detail in Section 4.1).

The next step is to define a dimensionless measure of whether an observed value of s_0^2 is large or small. One way to do this is to compare s_0^2 – calculated for the original gridbox – with the value that one would expect to obtain in a randomly shuffled version of the gridbox (e.g. Figure 1b). Accordingly, define the random variable

$$S^2 = \langle(\nabla^2\alpha)^2\rangle - \langle\nabla^2\alpha\rangle^2 \text{ in a shuffled gridbox}$$

It is important to stress that S^2 is a random variable whose value depends on which of the many possible shufflings of the original gridbox happens to have been picked. To remove this ambiguity consider its expected value $E(S^2)$ where $E(.)$ denotes the expectation taken over the ensemble of all possible shufflings of the original gridbox.

The following ratio might then be chosen as a dimensionless gridbox–scale index of pattern:

$$\lambda_F = \frac{s_0^2}{E(S^2)} \qquad (5)$$

The subscript F emphasizes that this is a formal definition rather than a practical one. For a real gridbox, exact calculation of $E\left(S^2\right)$ is prohibitively expensive since the number of possible shufflings of a gridbox is so large. For this reason, a more practical index is defined in the next section.

3.1. A practical Laplacian pattern index: λ_0

In this section a practical Laplacian pattern index λ_0 is defined. It is helpful to consider the limit of an infinitely large randomly shuffled gridbox. In this limit all non–identical pixels are uncorrelated, so that $\langle\alpha_j\alpha_N\rangle = 0$, $\langle\alpha_S\alpha_N\rangle = 0$ and so on, while the pixel variance is $\sigma_0^2 = \langle\alpha_i^2\rangle - \langle\alpha_i\rangle^2$ (equation 1). It therefore follows (from considering the expected value of the square of equation 3) that

$$\begin{aligned}\text{var}\left(\nabla^2\alpha\right) &= \frac{(-4)^2+1^2+1^2+1^2+1^2}{(\Delta x)^4}\sigma_0^2 \\ &= \frac{20\sigma_0^2}{(\Delta x)^4}\end{aligned}$$

in the limit of an infinitely large shuffled gridbox. The pixels in a randomly shuffled gridbox are not, strictly speaking, uncorrelated, but they will be approximately so when the number of pixels in the gridbox is large. It follows that the expected value of S^2 is given approximately by

$$E(S^2) \approx \frac{20\sigma_0^2}{(\Delta x)^4} \qquad (6)$$

For this reason, it is reasonable to replace the strict definition of the Laplacian pattern index (equation 5) with a practical definition:

$$\lambda_0 = \frac{(\Delta x)^4 s_0^2}{20\sigma_0^2} \quad \text{for} \quad \sigma_0^2 \neq 0 \qquad (7)$$

(The index is left undefined for a uniform gridbox $\sigma_0^2 = 0$.) This constitutes the operational definition of the Laplacian pattern index (along with equations 1 and 2). Note that the index is dimensionless and that the pixel size (Δx) cancels between equations 2, 3 and 7. For the purpose of calculating the index, therefore, units may chosen such that $(\Delta x) = 1$ in these equations.

4. Examples and interpretation

It is helpful to consider the values the Laplacian pattern index takes in a few simple cases. In this section upper and lower bounds for the index are considered, and gridboxes that are equivalent to the original gridbox (in the sense of having similar values of μ, σ^2 and λ) are constructed. Finally the use of the index is illustrated with some remote sensing data.

4.1. Example: Binary data

Consider first of all the simplest case in which the pixel–level data take one of two values, which may be taken as 0 (black) and 1 (white) [*Cressie*, 1991]. Let p denote the proportion of white pixels, so that $\mu_0 = p$ and $\sigma_0^2 = p(1-p)$. For each pixel define its 4 immediate neighbor pixels to be those in the north, south, east and west positions in the 5–point Laplacian stencil of equation 4, with cyclic boundary conditions being imposed for edge pixels. Each pixel can now be placed into one of 5 classes according to the number $i = 0, 1, 2, 3, 4$ of its immediate neighbors that are of the complementary color. Thus, a black pixel surrounded by 4 white pixels is in class $i = 4$, while a white pixel surrounded by 4 white pixels is in class $i = 0$. Setting $(\Delta x) = 1$ for simplicity it follows from equation 3 that a pixel in class i has a Laplacian filter value $\pm i$. We now let p_i denote the proportion of all pixels within the gridbox that lie in class i. For example, in the special case in which all pixels are the same color, every pixel is in class 0 and so $p_0 = 1$ while $p_1 = p_2 = p_3 = p_4 = 0$. Returning to the case of an arbitrary binary gridbox, it follows from equation 7 that

$$\begin{aligned}\lambda_0 &= \frac{(\pm1)^2p_1+(\pm2)^2p_2+(\pm3)^2p_3+(\pm4)^2p_4}{20p(1-p)} \\ &= \frac{p_1+4p_2+9p_3+16p_4}{20p(1-p)}\end{aligned} \qquad (8)$$

This formula allows the Laplacian pattern index λ_0 to be calculated for any binary gridbox. In particular it can be used to illustrate the range of values of λ_0.

Consider the case of a unit chessboard in which every pixel is of a different type to its immediate neighbors (Figure 2a). Half the pixels are white ($p = 0.5$), while all pixels

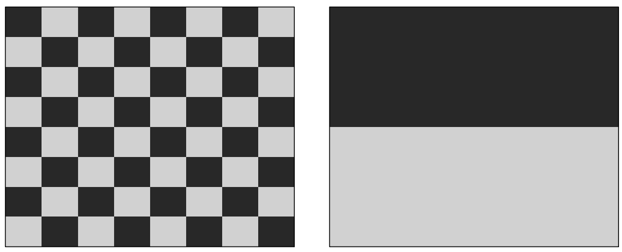

Figure 2. Examples of binary (black–and–white) gridboxes, each containing $M \times N$ pixels. (a) A unit chessboard gridbox in which each pixel is of a different color to its immediate neighbors. This gridbox has the maximum possible Laplacian pattern index $\lambda_0 = 3.2$. (b) A highly structured block gridbox containing equal blocks of color, each of size $M/2 \times N$ pixels. The Laplacian pattern index for this block pattern is $\lambda_0 = 0.8/M$ which tends to zero as the number of pixel rows $M \to \infty$.

have 4 neighbors of complementary color ($p_4 = 1$). This means that $p_0 = p_1 = p_2 = p_3 = 0$ and hence $\lambda_0 = 3.2$ (equation 8). It is intuitively clear that this gives the maximum possible value of λ_0 amongst all binary gridboxes since each pixel has the maximum possible value of $(\nabla^2\alpha)^2$. In fact, the result applies more generally. It can be shown that the upper bound $\lambda_0 \leq 3.2$ applies for all types of gridboxes and not just those with binary pixel values (Appendix A).

The Laplacian pattern index λ_0 is by definition non–negative, and it is intuitively clear that minimal values of λ_0 are associated with structured gridboxes in which neighboring pixels are positively correlated. As an example of a very low value of λ_0, consider the case of two equal blocks of color, each of size $M/2 \times N$ pixels (Figure 2b). Half the pixels are white ($p = 0.5$) and since cyclic boundary conditions have been imposed, a total of $4N$ out of the MN pixels have one neighbor of complementary color ($p_1 = 4/M$). The remaining pixels have no neighbors of complementary color ($p_0 = 1 - 4/M$, $p_2 = p_3 = p_4 = 0$) and hence $\lambda_0 = 0.8/M$ (equation 8). This represents a very low value of the Laplacian pattern index.

4.2. An equivalent deterministic gridbox

Suppose that observed values of μ_0 σ_0^2 and λ_0 have been obtained from an original gridbox (e.g. Figure 1a). In order to gain understanding of what a given value of λ_0 means it is helpful to ask what other gridboxes would yield similar values for these three statistics. Such gridboxes will be labelled 'equivalent'.

An example of a regular pattern for which the statistics μ_0, σ_0^2 and λ_0 can be calculated occurs when the pixel values are given by a sine wave with wavenumber k:

$$\begin{aligned} \alpha &= A + B \sin kx \sin ky \\ \nabla^2\alpha &= -2Bk^2 \sin kx \sin ky \end{aligned}$$

For simplicity suppose that the gridbox has side–lengths $M(\Delta x)$ and $N(\Delta x)$ that are integer multiples of the wavelength $2\pi/k$. Assume also that the wavelength is much larger than the pixel size $k(\Delta x) \ll 1$ so that the sinusoidal pattern is resolvable at the pixel scale. The variances of the pixel–values and the Laplacian Filter (equations 1 and 2) are then:

$$\sigma_0^2 = \frac{1}{2}B^2, \quad s_0^2 = 2B^2k^4$$

It then follows that

$$\lambda_0 = \frac{k^4(\Delta x)^4}{5}$$

from equations 6 and 7. Thus, given observed values of μ_0, σ_0^2 and λ_0 from an original gridbox, there is an equivalent sine–wave gridbox with wavenumber $k = (5\lambda)^{1/4}/(\Delta x)$ in which the pixel values are:

$$\mu_0 + \sqrt{2\sigma_0^2} \sin\left[(5\lambda_0)^{1/4}\frac{x}{(\Delta x)}\right] \sin\left[(5\lambda_0)^{1/4}\frac{y}{(\Delta x)}\right]$$

In other words, a gridbox with Laplacian pattern index λ_0 is equivalent to one with a sine–wave pattern of wavelength

$$l = 2\pi(5\lambda_0)^{-1/4} \cdot (\Delta x) \tag{9}$$

4.3. Equivalent stochastic gridboxes

An alternative way of generating gridboxes equivalent to the original gridbox is to define a statistical distribution from which random gridboxes can be generated. These random gridboxes should have values for the gridbox mean μ, gridbox variance σ^2 and Laplacian pattern index λ that are similar to those in the original gridbox (denoted by μ_0, σ_0^2 and λ_0). In this section a procedure is outlined for producing random gridboxes with these properties. It is straightforward to rescale the pixel values in random gridboxes to ensure that $\mu = \mu_0$ and $\sigma^2 = \sigma_0^2$. The Laplacian pattern index λ, on the other hand, is unaffected by the rescaling and remains a random variable whose value varies from random gridbox to random gridbox. It is, however, possible to set the expected value of λ equal to λ_0 by imposing an appropriate spatial correlation structure on the random gridboxes.

It is helpful to represent a random gridbox – in reality an $M \times N$ array of pixel values – with a vector $\underline{\alpha}$ of length MN. The aim is to define a suitable probability density function $f(\underline{\alpha})$ for the random vector $\underline{\alpha}$. Define the vector $\underline{U}$ and the matrix Q as follows:

$$\begin{aligned} \underline{U} &= \tfrac{1}{MN}(1, 1, 1, \ldots, 1)^T \\ Q &= \tfrac{1}{MN}I - \underline{U} \cdot \underline{U}^T \end{aligned}$$

where the superscript T denotes a transpose and I denotes the $MN \times MN$ identity matrix. The gridbox mean μ and gridbox variance σ^2 can then be expressed as an inner product and a quadratic form:

$$\begin{aligned} \mu(\underline{\alpha}) &= \underline{U}^T\underline{\alpha} \\ \sigma^2(\underline{\alpha}) &= \underline{\alpha}^T Q \underline{\alpha} \end{aligned}$$

Similarly, the values of the Laplacian filter within a gridbox can be represented (to within a multiplicative factor) by the vector $L\underline{\alpha}$, where L is the matrix representing the linear operator

$$\frac{1}{\sqrt{20MN}}\nabla^2$$

with cyclic boundary conditions. The factor $\sqrt{20MN}$ is included in the definition for convenience and means that the Laplacian pattern index can be expressed as a ratio of quadratic forms:

$$\lambda(\underline{\alpha}) = \frac{\underline{\alpha}^T L^T L \underline{\alpha}}{\underline{\alpha}^T Q \underline{\alpha}}$$

A gridbox $\underline{\alpha}$ satisfying $\underline{\alpha}^T Q \underline{\alpha} \neq 0$, can be rescaled by the transformation:

$$\underline{\alpha} \to \sigma_0 \frac{\underline{\alpha} - MN \cdot (\underline{U}^T\underline{\alpha})\underline{U}}{\sqrt{\underline{\alpha}^T Q \underline{\alpha}}} + \mu_0 \cdot MN \cdot \underline{U} \tag{10}$$

This transformation rescales the gridbox mean and variance to the desired values μ_0 and σ_0^2, but since it is linear, it leaves the Laplacian pattern index λ unchanged. To ensure that the index is typically close to λ_0, it is reasonable to impose the constraint

$$E_f(\lambda(\underline{\alpha})) = \lambda_0 \tag{11}$$

where $E_f(.)$ denotes an expectation taken over the distribution of random gridboxes. (The subscript f emphasizes that this is not the same as an expectation over all shufflings as in Section 3.) For simplicity, some natural symmetries can be imposed on the distribution f. Accordingly, the gridbox $\underline{\alpha}$ is assumed to be statistically isotropic and invariant under translation so that the pixels are statistically indistinguishable. With these symmetries, the correlations between a pixel and its neighbors can be expressed by the $M \times N$

array:

$$\begin{array}{ccccccc} . & . & . & . & . & . & . \\ . & . & . & r_{02} & . & . & . \\ . & . & r_{11} & r_{01} & r_{11} & . & . \\ . & r_{02} & r_{01} & 1 & r_{01} & r_{02} & . \\ . & . & r_{11} & r_{01} & r_{11} & . & . \\ . & . & . & r_{02} & . & . & . \\ . & . & . & . & . & . & . \end{array} \qquad (12)$$

where $r_{01}, r_{02}, r_{11}, \ldots$ denote correlation co–efficients, so that, for example, $\mathrm{cor}(\alpha_i, \alpha_i) = 1$ and $\mathrm{cor}(\alpha_i, \alpha_E) = r_{01}$. For clarity only the correlation coefficients closest to the central pixel have been shown. With $(\Delta x) = 1$ the Laplacian pattern index for a random gridbox $\underline{\alpha}$ can be written

$$\lambda(\underline{\alpha}) = \frac{\mathrm{var}(-4\alpha_i + \alpha_N + \alpha_S + \alpha_E + \alpha_W)}{20\mathrm{var}(\alpha_i)}.$$

The variances in this expression can be replaced by covariances according to the identity $\mathrm{var}(X) = \mathrm{cov}(X, X)$, and then by correlations according to the identity $\mathrm{cor}(X, Y) = \mathrm{cov(X,Y)}/\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}$. It then follows that the constraint of equation 11 can be rewritten as a constraint on the local correlation structure:

$$E_f(\lambda(\underline{\alpha})) = \lambda_0 = \frac{5 - 8r_{01} + r_{02} + 2r_{11}}{5} \qquad (13)$$

This demonstrates that the Laplacian pattern index λ_0 can be interpreted as a measure of local correlation structure. (A helpful analogy can be drawn with the structure functions of homogeneous isotropic turbulence in which correlations are expressed as functions of spatial separation [e.g., *Frisch*, 1995]).

A gridbox distribution having the required correlation structure (equation 13) can now be defined by rescaling (equation 10) gridboxes drawn from a zero–mean multivariate normal distribution with covariance matrix Σ:

$$f(\underline{\alpha}) = \frac{1}{(2\pi)^{MN/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\underline{\alpha}^T\Sigma^{-1}\underline{\alpha}\right) \qquad (14)$$

Here, the covariance matrix has determinant $|\Sigma|$ and inverse Σ^{-1} [*Kendall and Stuart*, 1963]. Each row of the covariance matrix Σ has length MN and can be interpreted as a rescaled, reshaped version of the $M \times N$ correlation array in equation 12. The difference is that the covariance matrix contains rows of covariances $\Sigma_{ij} = \mathrm{cov}(\alpha_i, \alpha_j)$ while the stencil is an array of correlations $\mathrm{cor}(\alpha_i, \alpha_j)$. It is helpful to normalize the covariance matrix to ensure that it is also a matrix of correlations. This requires that all the diagonal elements of Σ be unity. Since all pixels are assumed to be statistically indistinguishable this is equivalent to imposing the constraint $\mathrm{tr}(\Sigma) = MN$ where $\mathrm{tr}(.)$ denotes the trace of a matrix.

It can be shown that when $\underline{\alpha}$ is drawn from a zero–mean multivariate normal distribution with covariance matrix Σ an arbitrary quadratic form $\underline{\alpha}^T P \underline{\alpha}$ is a random variable with mean $\mathrm{tr}(P\Sigma)$ and variance $2\mathrm{tr}(P\Sigma P\Sigma)$ [*Kendall and Stuart*, 1963]. Since the gridbox variance σ^2 can be rescaled to unity without changing the pattern index (equation 10), it follows that the Laplacian pattern index $\lambda(\underline{\alpha})$ is a random variable with mean $\mathrm{tr}(L^T L\Sigma) = \lambda_0$ and variance $2\mathrm{tr}(L^T L\Sigma L^T L\Sigma)$.

Thus, for a zero–mean multivariate normal distribution with the assumed symmetries (equation 14), two constraints are to be imposed on the covariance matrix Σ:

$$\begin{array}{rcl} \mathrm{tr}(\Sigma) & = & MN \\ \mathrm{tr}(L^T L\Sigma) & = & \lambda_0 \end{array} \qquad (15)$$

The first constraint normalizes the covariance matrix to ensure that its elements are correlations in the range $[-1, 1]$. The second constraint ensures that the expected value of the Laplacian pattern index $\lambda(\underline{\alpha})$ is equal to the imposed value λ_0 taken from the original gridbox.

4.3.1. The city–block correlation structure

One way to satisfy the constraints of equation 15 is to impose the correlation structure:

$$\Sigma_{ij} = r^{d(i,j)} \qquad (16)$$

where r is an appropriate constant (determined below) and $d(i, j)$ is the distance between pixels α_i and α_j in the gridbox, calculated according to the city–block metric with cyclic boundary conditions. (The city–block distance between two pixels is the least number of north–south and/or east–west moves required to travel between them.) The city-block correlation structure is represented by the array

$$\begin{array}{ccccccc} . & . & . & . & . & . & . \\ . & . & . & r^2 & . & . & . \\ . & . & r^2 & r & r^2 & . & . \\ . & r^2 & r & 1 & r & r^2 & . \\ . & . & r^2 & r & r^2 & . & . \\ . & . & . & r^2 & . & . & . \\ . & . & . & . & . & . & . \end{array} \qquad (17)$$

in the notation of equation 12. The constant r can be interpreted as an immediate neighbor correlation since it is the correlation coefficient between a pixel and any of its four immediate neighbors. Comparison with the generic correlation structure (equation 12) shows that the required constraints (equation 13) can be satisfied provided that the immediate neighbor correlation r satisfies:

$$r = \frac{4 - \sqrt{1 + 15\lambda_0}}{3} \qquad (18)$$

This relationship is illustrated in Figure 3. Note that $\lambda_0 = 0$ corresponds to $r = 1$ (immediate neighbors perfectly positively correlated), $\lambda_0 = 1$ corresponds to $r = 0$ (immediate neighbors uncorrelated) and $\lambda_0 = 3.2$ corresponds to $r = -1$ (immediate neighbors perfectly negatively correlated). Random gridboxes generated using the city–block correlation structure are shown in Figure 4.

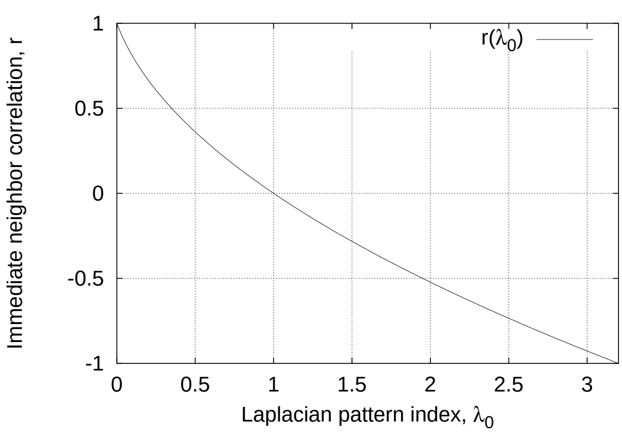

Figure 3. The immediate neighbor correlation structure for the city–block covariance matrix (equation 16). The immediate neighbor correlation r is shown as a function of the expected Laplacian pattern index λ_0 (equation 18).

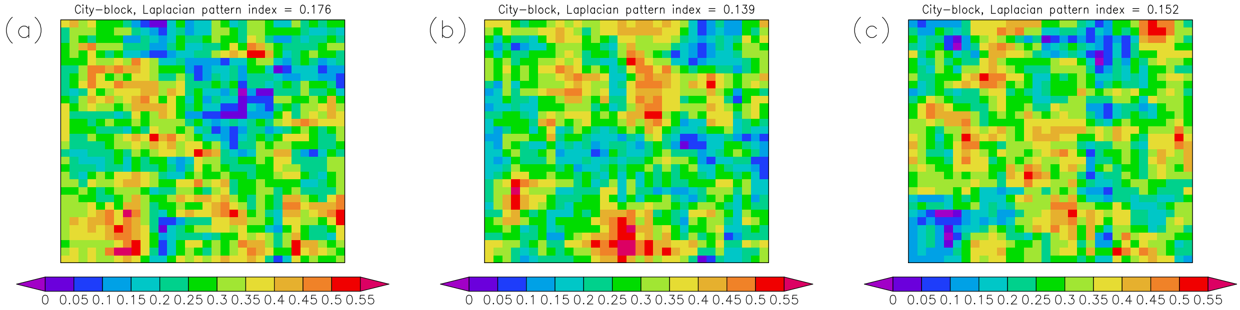

Figure 4. Random gridboxes equivalent (in the sense defined in the text) to the original gridbox in Figure 1a. Each gridbox has the imposed gridbox mean μ_0 and the imposed gridbox variance σ_0^2.

4.3.2. Stochastic gridboxes: summary

A probability distribution for generating gridboxes equivalent to an original gridbox has been outlined. The gridbox distribution is a zero–mean multivariate normal distribution with a suitable correlation structure defined by its covariance matrix Σ (equation 14). The procedure for generating an equivalent random gridbox is as follows: (1) Generate a random gridbox $\underline{\alpha}$ from the probability density function (equation 14). (2) Rescale the gridbox to have the required gridbox mean μ_0 and gridbox variance σ_0^2 (equation 10). The Laplacian pattern index λ of the rescaled gridbox is then a random variable with mean $\mathrm{tr}(L^T L\Sigma) = \lambda_0$ and variance $2\mathrm{tr}(L^T L\Sigma L^T L\Sigma)$.

4.4. Example: fAPAR data

It is helpful at this point to illustrate the calculation of the Laplacian pattern index for some real remote sensing data. In the illustrative example given in Figure 5 the data consist of fAPAR values for Western Europe in September 2000 (downloaded from `http://fapar.jrc.it`). The original data consist of a grid of 1531×1376 pixels whose resolution is approximately 2 km (Figure 5a). After masking of non–land pixels, these data were aggregated up to an array of 43×47 gridboxes, each of size 32×32 pixels ($M = N = 32$) so that the gridboxes are nominally of size 64 km $\times$ 64 km. For each gridbox, values of the gridbox mean μ_0 and gridbox variance σ_0^2 can easily be calculated (Figure 5b,c). A traditional climate model would have access to these two statistics as gridbox level summaries of the fAPAR field. It would therefore be aware – for example – of the low fAPAR in North Africa and Iberia, and the similarity in fAPAR between Ireland and the Low Countries. This is clearly useful information, but the next step is to examine the impact of additional information about subgridscale pattern provided by the Laplacian pattern index.

The Laplacian pattern index λ_0 was calculated for each gridbox and the results are shown in Figure 5d. It is clear that there are coherent regional patterns in the Laplacian pattern index λ_0 with noticeably high values in Ireland, Scotland and Eastern Europe (Figure 5d). These high values suggest that the lengthscale of land–surface fAPAR patterns in these areas is relatively small. From the point of view of climate modelling this highlights the potential for regional differences in subgridscale land-surface processes. For example, Ireland and Northern France have very similar values of the gridbox mean and gridbox variance (Figure 5b,c), but strikingly dissimilar values of the Laplacian pattern index (Figure 5d). It follows that different parameterizations of subgridscale processes might be appropriate for these two regions.

Interpretation of the Laplacian pattern index can be helped by considering the two parameters defined in this paper that are simple functions of it. These two parameters are effectively rescaled versions of the Laplacian pattern index, designed in such a way as to yield physically meaningful interpretations. Firstly, the wavelength of an equivalent sinusoid, l, is given by equation 9 and is shown in Figure 5e. This represents the wavelength of a sinusoidal pattern that would produce the observed value of the Laplacian pattern index. In the example shown this varies from < 5 pixels in Ireland and Eastern Europe to around 7 pixels in Southern Europe. Secondly, the immediate neighbor correlation of an equivalent gridbox, r, is given by equation 18 and shown in Figure 5f. This represents the correlation coefficient between neighboring pixels in a stochastic pattern that would produce the observed value of the Laplacian pattern index. With this interpretation the variations in land-surface pattern are revealed by the correlation between neighboring pixels being lower than 0.4 in Ireland but greater than 0.6 in Southern Europe.

In summary, the example data in Figure 5 illustrate how the Laplacian pattern index can be used to quantify subgrid patterns in land–surface properties, and to show how the lengthscale of subgrid variability changes regionally. This information could be used in deriving subgrid parameterizations within numerical models.

5. Conclusions

In this paper a new index of pattern – the Laplacian pattern index λ_0 – has been defined and its properties investigated. The index can be calculated for any 2–d field of real-valued data on a lattice, such as fAPAR, temperature or topographic height. It can be interpreted as a dimensionless measure of gradient–driven flux accumulation, and in particular as a measure of the similarity of this accumulation under pixel shuffling. The following properties of λ_0 have been derived. (1) $\sqrt{\lambda_0}$ can be interpreted as the root–mean–square flux accumulation in the original gridbox divided by the root–mean–square flux accumulation in a randomly shuffled version of the gridbox. (2) $\lambda_0 = 1$ means that pixels are uncorrelated with their immediate neighbors and that there is complete invariance under pixel shuffling. (3) $\lambda_0 < 1$ means that pixels tend to be positively correlated with their immediate neighbors. (4) $\lambda_0 > 1$ means that pixels tend to be negatively correlated with their immediate neighbors. (5) The maximum possible value of λ_0 is 3.2 and occurs for a unit chessboard pattern of alternate black and white pixels (Appendix A). (6) $l = 2\pi(5\lambda_0)^{-1/4}$ can be interpreted as the wavelength (in pixel units) of an equivalent sine wave (equation 9). (7) The index can be interpreted as a measure of local correlation structure. For random gridboxes with the city–block correlation structure (Section 4.3.1), the immediate neighbor correlation is $r = (4 - \sqrt{1 + 15\lambda_0})/3$ (Figure 3).

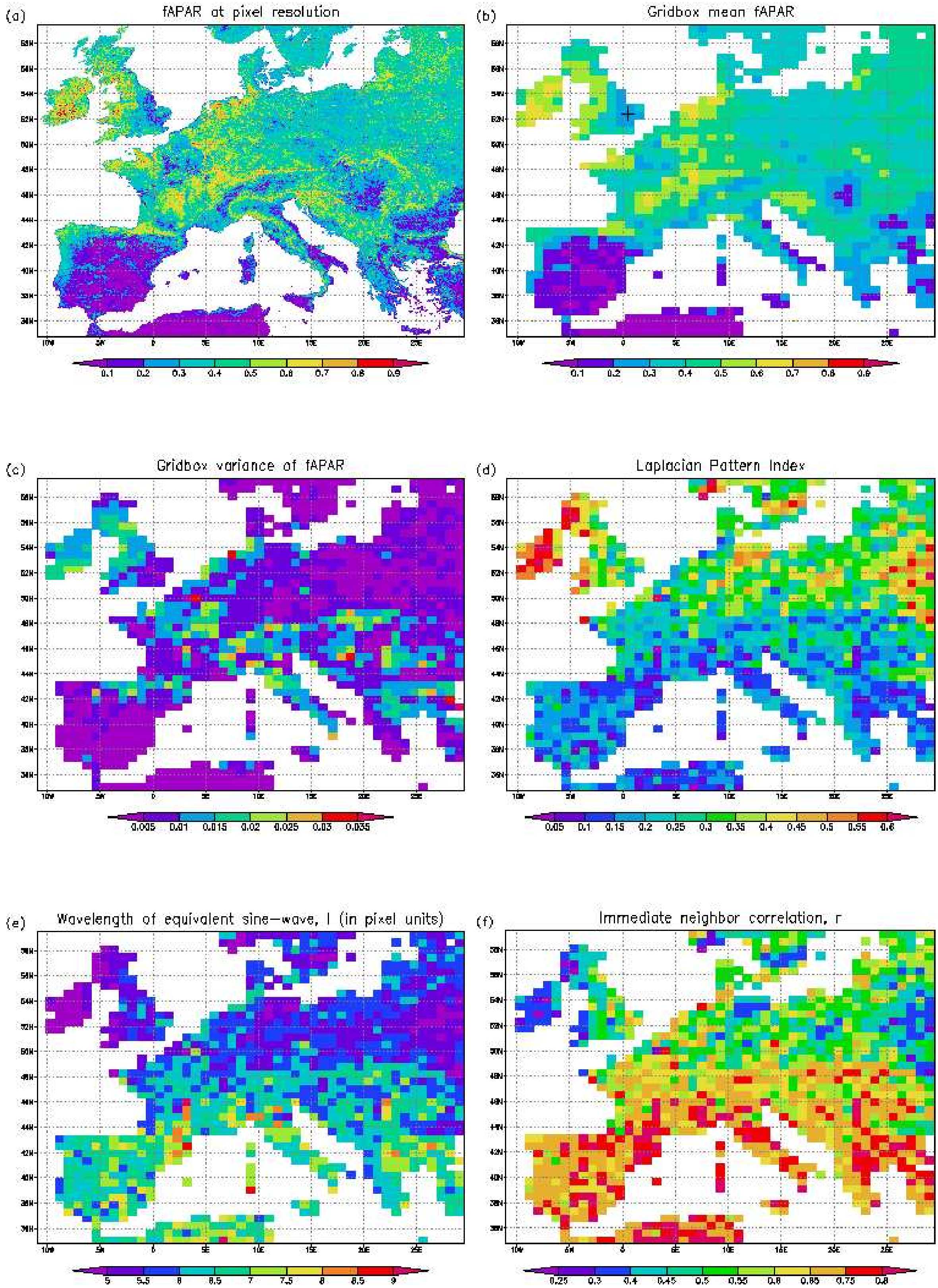

Figure 5. Illustration of the Laplacian pattern index using Western European fAPAR data for September 2000. (a) Original data at pixel resolution (2 km). (b) Data aggregated to 64 km to give the gridbox mean μ_0. (c) The gridbox variance σ_0^2. (d) The Laplacian pattern index λ_0. (e) The wavelength l of an equivalent sine–wave pattern (equation 9). (f) The immediate neighbor correlation r for an equivalent city–block pattern (equation 18).

the spatial arrangement of pixels – can now be added the Laplacian pattern index λ as a measure of pattern. The additional information provided by the Laplacian pattern index reveals the typical lengthscales of subgrid variability and how they change regionally. These lengthscales could be used in the derivation of subgrid parameterizations in models.

The Laplacian pattern index index has a number of advantages over other ways of quantifying pattern. It produces a single summary statistic (unlike Fourier or wavelet methods), it applies to real–valued continuous data (unlike indices such as the area:perimeter ratio which are restricted to classified data), and it has a natural physical interpretation. It is hoped that the Laplacian pattern index will prove useful in numerical climate models as a means of parameterizing subgrid scale effects (e.g. determining whether the mosaic or mixture surface energy balance is appropriate). More generally, the index could be used as a physically–motivated measure of pattern for any two–dimensional array of real–valued data.

Acknowledgments. We thank Will Grey, Heiko Balzter, Chris Taylor, Graham Weedon, Charles George and France Gerard for helpful comments. Reviews from three anonymous referees were greatly appreciated. This work was supported by the NERC Climate and Land-Surface Systems Interaction Centre (CLASSIC).

Appendix A: Proof that $\lambda \leq 3.2$

This is a sketch proof that the maximum value of the Laplacian pattern index arises for a pixel–scale chessboard pattern (Section 4.1, Figure 2a). Suppose for simplicity that $\mu_0 = 0$ and label the pixels surrounding α_i in a natural way:

$$\begin{array}{ccccc} & & \alpha_{NN} & & \\ & \alpha_{NW} & \alpha_N & \alpha_{NE} & \\ \alpha_{WW} & \alpha_W & \alpha_i & \alpha_E & \alpha_{EE} \\ & \alpha_{SW} & \alpha_S & \alpha_{SE} & \\ & & \alpha_{SS} & & \end{array}$$

Define

$$F = \sum_i \left(-4\alpha_i + \alpha_N + \alpha_S + \alpha_E + \alpha_W\right)^2 , \quad G = \sum_i \alpha_i^2$$

One can extremize F subject to the constraint $G = \sigma_0^2$ by extremising $H = F - aG$ where a is a Lagrange multiplier. It follows (after some algebra), that

$$\begin{aligned} \frac{\partial H}{\partial \alpha_j} &= (40 - a)\alpha_j \\ &- 16(\alpha_N + \alpha_S + \alpha_E + \alpha_W) \\ &+ 2(\alpha_{NN} + \alpha_{SS} + \alpha_{EE} + \alpha_{WW}) \\ &+ 4(\alpha_{NE} + \alpha_{NW} + \alpha_{SE} + \alpha_{SW}) \end{aligned}$$

Noting that $\partial^2 H/\partial\alpha_j^2 = 40 - a$ it is then clear by setting $a = 128$ that a pixel–scale chessboard pattern $\alpha_N = \alpha_S = \alpha_E = \alpha_W = -\alpha_j$, $\alpha_{NN} = \alpha_{SS} = \alpha_{EE} = \alpha_{WW} = \alpha_j$ and $\alpha_{NE} = \alpha_{SE} = \alpha_{NW} = \alpha_{SW} = \alpha_j$ is a solution for which $\partial H/\partial\alpha_j = 0$ and $\partial^2 H/\partial\alpha_j^2 < 0$. It therefore maximizes F while satisfying $G = \sigma_0^2$.

References

Blyth, E.M. (1995), Using a simple SVAT scheme to describe the effect of scale on aggregation, *Boundary-Layer Met.*, *72*, 267 – 285.

Cressie, N.A.C. (1991), *Statistics for spatial data*, Wiley.

Essery, R.L.H., Best, M.J., Betts, R.A. and P.M. Cox (2003), Explicit representation of subgrid heterogeneity in a GCM land surface scheme, *J. Hydromet.*, *4*, 530 – 543.

Frisch, U. (1995), *Turbulence – the legacy of A.N. Kolmogorov*, Cambridge University Press.

Giorgi, F. and R. Avissar (1997), Representation of heterogeneity effects in earth system modeling: experience from land surface modeling, *Rev. Geophys., 35(4)*, 413 – 438.

Gustafson, E.J. (1998), Quantifying landscape spatial pattern: What is the state of the art?, *Ecosystems*, *1*, 143 – 156.

Kendall, M.G. and A. Stuart (1963), *The Advanced Theory of Statistics Volume 1: Distribution Theory*, Griffin.

Koster, R.D. and M.J. Suarez (1992), A comparative analysis of two land surface heterogeneity representations, *J. Climate*, *5*, 1379 – 1390.

McGarigal, K. and B.J. Marks (1995), FRAGSTATS: spatial pattern analysis program for quantifying landscape structure, *USDA For. Serv. Gen. Tech. Rep.*, *PNW-351.*

Mahrt, L. (2000), Surface heterogeneity and vertical structure of the boundary layer, *Boundary-Layer Met.*, *96*, 33 – 62.

Molod, A. and H. Salmun (2002), A global assessment of the mosaic approach to modeling land surface heterogeneity, *J. Geophys. Res., 107*, doi:10.1029/2001JD000588.

North, G.R., Cahalan, R.F. and J.A. Coakley (1981), Energy Balance Climate Models, *Rev. Geophys., 19*, 1, 91 – 121.

Raupach, M.R. and J.J. Finnigan (1995), Scale issues in boundary-layer meteorology: surface energy balances , *Hydrol. Proc., 9*, 6, 589 – 612.

Saura, S. and J. Martínez–Millán (2001), Sensitivity of landscape pattern metrics to map spatial extent, *Photogrammetric Engineering and Remote Sensing*, *67*, 9, 1027 –1036.

T. E. Jupp, CEH Monks Wood, U.K. (tju@ceh.ac.uk)