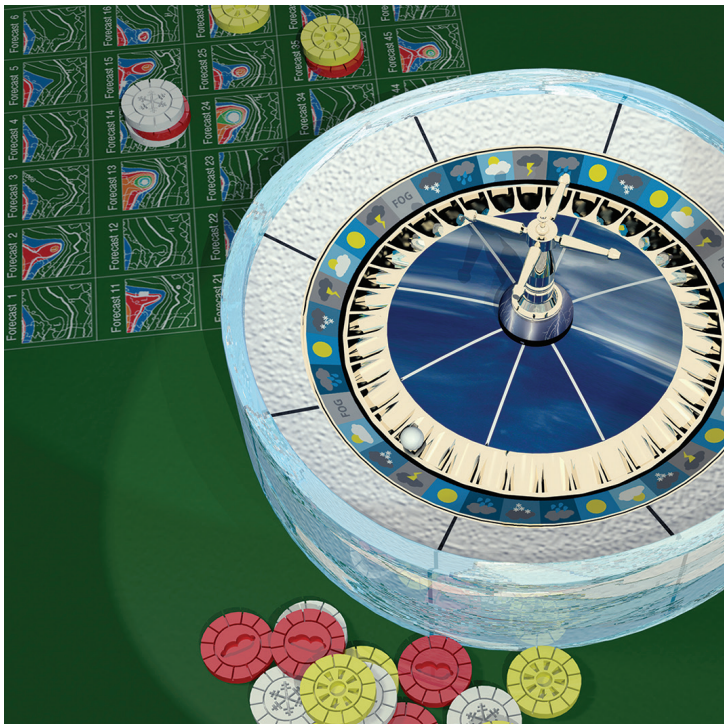
ECMWF Feature article

from Newsletter Number 106 – Winter 2005/06

METEOROLOGY

"Wavelet" J_b – A new way to model the statistics of background errors



www.ecmwf.int/en/about/news-centre/media-resources

doi:10.21957/ef5601ak43

This article appeared in the Meteorology section of ECMWF Newsletter No. 106 – Winter 2005/06, pp. 23–28.

"Wavelet" J_b – A new way to model the statistics of background errors

Mike Fisher

It has been estimated (*Cardinali et al.,* 2004) that only 15% of the information content of any recent ECMWF analysis is attributable to the observations assimilated during that analysis. The remaining 85% of the information is provided by the background: a short forecast from the preceding analysis. The rôle of the background is to incorporate into the current analysis information from observations that were assimilated during earlier cycles. Ultimately, any given analysis represents a synthesis of observations made over a five-to-ten day period.

The degree to which observational information can be brought forward in time to contribute to the current analysis depends both on the quality of the forecast model, and on the way in which the information is transferred from analysis cycle to analysis cycle via the background. Crucial to the success of this transfer is the statistical description of background errors.

The operational implementation of IFS cycle Cy29r1 in April 2005 saw the introduction of a new formulation for the background term of the analysis cost function, dubbed "Wavelet" J_b. This new formulation allows the modelled statistics of background error to exhibit spatial variation of vertical and horizontal correlations, while retaining important spectral characteristics. The formulation is described in rather mathematical terms by *Fisher* (2004b). The aim of this article is to provide a complementary, equation-free description of the concepts underlying "Wavelet" J_b, and the reasons for its adoption. But, let's start by putting things into context.

A brief description of the variational analysis method

Variational data assimilation defines the analysis in terms of a "cost" (or penalty) function, which is a sum of several components. Each component of the cost function measures how well the analysis meets some criterion. A perfect match for a given criterion is represented by a value of zero for the corresponding component of the cost function, whereas large values indicate that criteria have not been met. The final analysis represents the particular compromise between the different criteria that minimizes the overall "cost".

The terms of the cost function measure:

- Differences between the analysis and the background.
- Discrepancies between the analysis and observations.
- The amplitude of rapid, divergent oscillations.
- How far the evolution of the analysis deviates from a possible evolution of the model.

(This last term is zero in the current ECMWF analysis system, as the analysis is forced to evolve exactly as dictated by the model.)

In this article, we will consider only the first of these terms: the background cost function, conventionally denoted by J_b. Like the other terms of the cost function, J_b is defined statistically, and encodes our knowledge of the statistical properties of errors in the background. It heavily penalizes differences between the analysis and the background that are unlikely (in terms of magnitude, size, shape, etc.), while allowing more plausible departures from the background. The likelihood or otherwise of a given departure is measured using a covariance matrix. In principle, this matrix tabulates the covariances between all pairs of model grid-points. However, since even a low-resolution model can have well over a million grid-points, the number of covariances that must be specified is so huge that direct specification is not practical, even on a super-computer. Ways must be found to model the statistical properties of background error with fewer parameters, while retaining their chief characteristics.

Before discussing background covariance modelling in more detail, let us introduce the notion of a change-of-variable. In a variational analysis system, this is a transformation of the departures from the background that allows the background cost to be evaluated quickly and simply as the sum of the squares of the elements of the transformed vector (called the "control vector"). From the statistical point of view, the transformation makes the elements of the control vector statistically independent (i.e. their errors are uncorrelated) and with a variance of one.

 doi:10.21957/ef5601ak43

From the practical point of view, expressing the analysis problem in terms of the control vector has an important "preconditioning" effect. That is, the minimization algorithm is able to locate the minimum of the cost function much more rapidly when the function is expressed in terms of the control vector, than when it is expressed directly in terms of the "raw" model variables. During the course of the minimization, the control vector must be converted into equivalent values of the model variables so that, for example, they may be compared with observations. However, it is never necessary to perform the reverse transformation that converts model variables to a control vector.

In a variational analysis system, constructing a background covariance model boils down to specifying a transformation that converts a control vector of statistically independent, unit-variance elements, into model fields with the statistical structure of background error. This transformation implicitly defines the covariance matrix of background error, which is never explicitly represented. Before considering in more detail how such a transformation may be defined, let's consider which characteristics of background error we wish to retain in the model.

Some important characteristics of background error

Perhaps the most important characteristic of background error is that it tends to be balanced. That is, errors in temperature, surface pressure and wind are related to each other via geostrophic and hydrostatic balances. The *Derber & Bouttier* (1999) approach to accounting for balance is to express the control vector in terms of a single "balanced" variable, and a number of residual, "unbalanced" variables. They chose vorticity as the balanced variable. With this approach, the last step of the transformation from control vector to model fields is to calculate balanced components of other variables from the vorticity (using geostrophy, for example) and add them to the residual components. It is assumed that this process accounts for all the correlations between variables, so that the different variables of the control vector are assumed to be statistically independent. This approach to representing balance has been retained for "Wavelet"J_b.

A second important characteristic of background error is "non-separability". This simply means the tendency for broad horizontal error structures to be deep, and for narrow horizontal structures to be shallow. This property is illustrated in Figure 1, which shows the mean vertical correlation between temperature background errors at model level 49 (near 850 hPa) and temperature errors at other levels. The horizontal axis shows spherical wavenumber n: a measure of horizontal scale, with small scales corresponding to large values of *n*.

One important reason for wanting to retain non-separability in the covariance model is its interaction with balance. The strict, functional relationship between the balanced components of background error (e.g. between vorticity and the balanced part of the temperature error) means that specifying a covariance model for one variable (vorticity, say) implicity imposes a covariance model on other variables. It has been found that a separable model, in which all horizontal scales have the same vertical correlation, is unable simultaneously to represent the correlations of both wind and temperature (see *Bartello & Mitchell,* 1992).

A third feature of background error correlation is spatial variation. We expect background error correlations to vary geographically. Tropical error structures are different from those in mid-latitudes, and errors over data-dense regions are different from those over data-sparse regions.

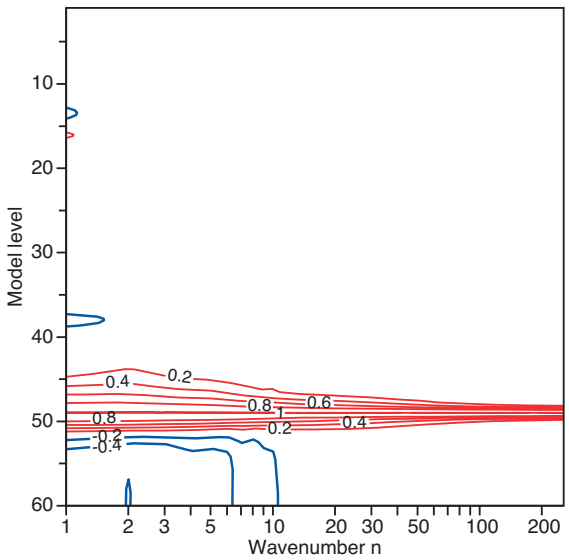


Figure 1 Mean correlation between background errors of temperature on model level 49 (near 850 hPa) and the corresponding errors on other model levels, as a function of spherical wave number.

doi:10.21957/ef5601ak43

The Derber-Bouttier J_b

The background covariance model devised by *Derber & Bouttier* (1999) (hereafter referred to as the "Derber-Bouttier J_b") was used operationally at ECMWF from May 1997 until April 2005, and had a very positive impact on forecast skill. It attempts to capture the first two properties described above: balance and non-separability. It also achieves a limited degree of spatial variability.

The treatment of balance has already been described. Non-separability is addressed by having different vertical correlation matrices for each spherical wavenumber. Because wavenumber is a global concept, this approach does not allow any horizontal variation of the correlations. However, this is only true of the variables that make up the control vector (vorticity, "unbalanced" temperature, etc.). Since the full temperature and surface pressure fields are calculated as the sum of an "unbalanced" residual and balanced fields derived from the vorticity, their covariance structure is determined in part by the statistics of the "unbalanced" component, and in part by the statistics of the vorticity error. In mid-latitudes, the balanced component is dominant, and the correlations of temperature error are effectively those implied through the balance equations by the vorticity correlations. In the tropics, by contrast, the residual components dominate, and the correlations are those prescribed for the "unbalanced" components. The result is that the implied temperature statistics vary with latitude.

Horizontal correlations are handled in the Derber-Bouttier J_b using "convolution". To create a horizontally-correlated field (e.g. vorticity on some model level), each horizontal grid-point of the field is calculated as a weighted average of the values of the control vector at nearby points. The weight given to each point is a function of distance from the central grid-point. A typical weighting function for vorticity is shown in Figure 2.

This type of averaging is used because it can be implemented very efficiently using spherical transforms. Specifically, convolution of a field with a function of distance f (such as that shown in Figure 2) can be achieved by multiplying the spectral coefficients of the field by coefficients $\hat{f}_n$ that depend only on the wavenumber n. There is a simple mathematical relationship between the coefficients $\hat{f}_n$ and the function f.

The disadvantages of this approach are first that the weighting function, f, cannot be varied from grid-point to grid-point, so that spatial variation of horizontal correlation is not allowed, and second that the averaging is isotropic (the same in all horizontal directions).

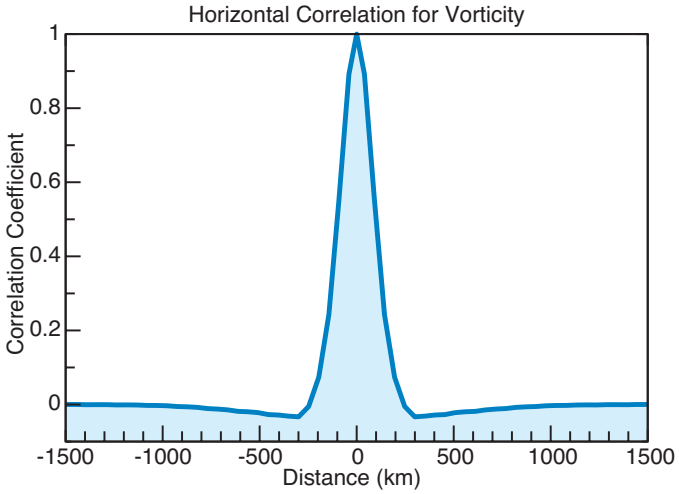


Figure 2 Mean horizontal correlation of vorticity at model level 49 (near 850 hPa) as a function of distance.

doi:10.21957/ef5601ak43

Musical interlude

By now, you must be wondering when I'm going to get round to talking about wavelets. I'll get there soon. But first, a little music.

Consider the snippet of a well-known tune shown in Figure 3. A purely spectral representation of the tune would identify the frequencies present, but would not identify when these frequencies appear. This is illustrated schematically by the graph to the left of the musical stave, showing amplitude as a function of frequency. By contrast, a purely temporal representation would identify the time at which each note is played and its loudness, but not the frequency, as illustrated by the graph below the stave. Clearly, neither the spectral nor the temporal representations capture the full nature of the music.

If we now replace time by spatial position, and frequency by spatial scale, we have rough analogues of two approaches to covariance modelling. The purely spectral approach, as exemplified by the Derber-Bouttier J_b, identifies vertical correlations as a function of scale, but does not identify where the correlations apply. It is like the spectral representation of the melody to the left of the stave. An alternative (separable) approach is to specify vertical correlations as a function of horizontal position, and apply them to columns of the model's grid. This provides spatial information, but applies the same correlations to all scales, rather in the way that the temporal description of the melody fails to identify the pitches of the notes. It is clear that, as with the musical example, neither the purely spectral nor the purely spatial approach captures all the characteristics of the correlations. What is needed is an equivalent of musical notation that identifies correlations as a function of both scale and location. This is the aim of "Wavelet" J_b.

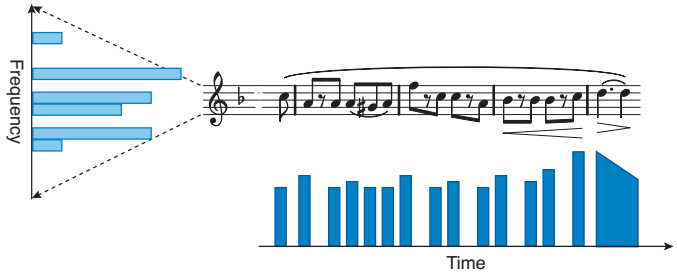


Figure 3 An illustration of the benefits of resolving both temporal and spectral information.

"Wavelet" J_b

Have another look at Figure 1. The variation of vertical correlation with wavenumber is rather smooth, yet the Derber- Bouttier J_b describes this spectral variation with individual matrices for each of the 256 wavenumbers of the T255 truncation. Horizontal correlations, too, are described by 256 spectral coefficients per model level, despite being smooth functions of wavenumber.

The first step towards "Wavelet" J_b is to realise that both the vertical and horizontal correlations may be described fairly accurately by specifying matrices and coefficients for a few selected wavenumbers, and simply interpolating between them. One way to do this would be to generate the 256 matrices and 256 coefficients required by the Derber-Bouttier J_b explicitly from the matrices and coefficients for the selected wavenumbers, and then proceed as before. This does not gain us much, except to provide a more compact description of the statistics. However, there is a different way to perform the interpolation, as described below.

The Derber-Bouttier J_b identifies horizontal scale using wavenumber, n, and there is a separate part of the control vector for each n. "Wavelet"J_b does things differently. It identifies horizontal scale with overlapping bands of wavenumbers centred on each of the selected wavenumbers to be interpolated. There is a separate part of the control vector for each band, and each part is assumed to be uncorrelated with other parts of the control vector. Because the wavenumber bands overlap, each wavenumber is represented in two or more different parts of the control vector, corresponding to two or more different bands. To construct the model background-departures for a particular wavenumber from the control vector, we multiply the corresponding wavenumber coefficients in each part of the control vector by the matrix and coefficient defined for the band. We then multiply each contribution to the wavenumber by a weight, and add the resulting values. The weights define the interpolation in wavenumber between the matrices and coefficients defined for the central wavenumbers of each band. The effect is identical to what we would have achieved by first interpolating the matrices and coefficients, and then applying the Derber-Bouttier J_b.

The weighting functions used in the current implementation of "Wavelet" J_b are shown in Figure 4. Four arbitrarily-chosen functions have been highlighted. The functions are, in fact, the square roots of triangular functions, and produce a linear interpolation of covariance between the wavenumber bands.

Now, the weighting functions are functions of wavenumber, n, and as described earlier, multiplication by a function of n is equivalent to convolution with a particular spatial function. Each point of the convolved field corresponds to a spatial average of nearby points.

Figure 5 shows the spatial averaging functions implied by the spectral functions highlighted in Figure 4. Note that the functions are quite localised, especially for the higher wavenumber bands.

Consider now what happens if we allow the matrices and coefficients that define the correlations to vary with latitude and longitude. For example, suppose we use different matrices and coefficients for points over North America than we do for points over Europe. For all but the lowest wavenumbers (corresponding to planetary scales) the spatial averaging functions for points over Europe give nearly zero weight to points over North America, and vice versa. So, the correlations in effect over Europe will essentially be those we would have got had we used the European correlations and coefficients everywhere. Likewise, the correlations over North America will effectively be those defined by the correlations and coefficients we specify for North America. In other words, we have succeeded in introducing spatial variation into the correlation statistics, while retaining the ability to describe their spectral variation. To return to the analogy of the previous section, we might say that we have improved on the constant drone of the Derber-Bouttier J_b and the drum-solo of the separable formulation, and produced a background covariance model that can represent the full "music" of the background error covariances! Some examples of the ability of "Wavelet" J_b to produce spatially-varying correlation structures are given in *Fisher* (2004a) and *Fisher* (2004b).

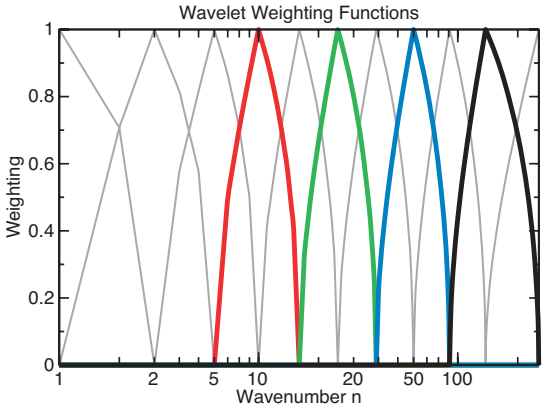


Figure 4 Weighting functions for the different wavenumber bands in "Wavelet" J_b. The coloured curves are referred to in Figure 5.

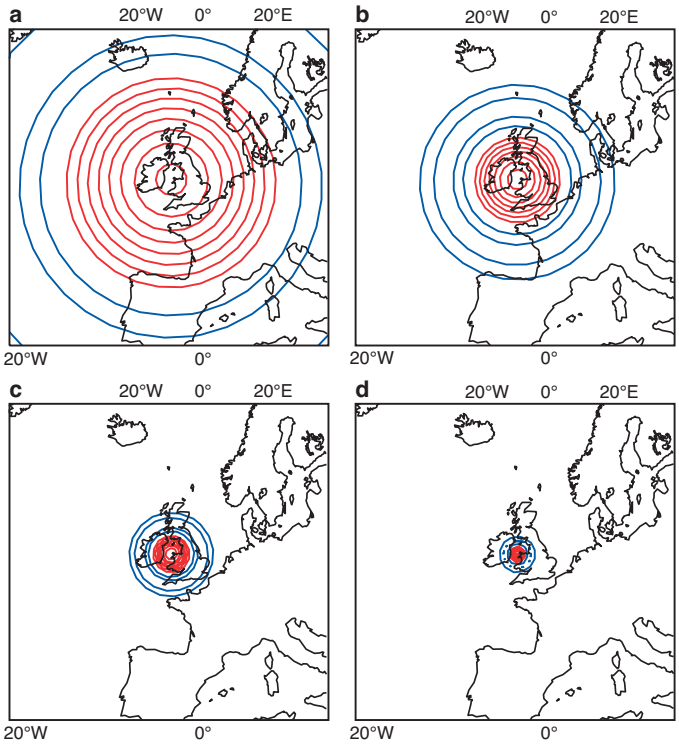


Figure 5 The spatial weighting functions corresponding to the functions of wavenumber highlighted in Figure 4, for a point over the Irish Sea. Red and blue contours represent positive and negative weights. The zero line is not plotted. Plots (a), (b), (c) and (d) refer to the red, green, blue and black curves of Figure 4.

doi:10.21957/ef5601ak43

Why call this "Wavelet" J_b?

The term "wavelet" describes a particular class of mathematical functions that are localised in both space and frequency. These functions have become quite popular in recent years for analysing problems for which a purely spectral or a purely spatial (or temporal) approach is insufficient. Applications include image compression, signal analysis and linear algebra.

Although the exact definition of what constitutes a "wavelet" varied a little after their introduction in the 1980's, it is now generally accepted that the term should be restricted to functions that have the mathematical property of orthogonality. The weighting functions used in "Wavelet" J_b do not have this property, for reasons explained in *Fisher* (2004b), and should not strictly be called "wavelets". Nevertheless, the term neatly sums up the most important property of the functions, which is their simultaneous localization in both wavenumber and space. Some other properties of orthogonal wavelets, such as the ability to define transforms, also apply (see *Fisher,* 2004b), making the distinction between true wavelets and the functions described here somewhat technical. Since the term "Wavelet" J_b is also much snappier than any more precise alternative, I have chosen to use it, and to indicate its inexactness with inverted commas.

Practical issues

Astute readers will have noted that the need to represent each wavenumber in more than one band results in some redundancy. (This can be regarded as a consequence of the lack of orthogonality between the weighting functions.) A practical consequence is that the control vector must be larger than the corresponding vector of model variables. Although the size of the control vector is not directly related to the computational cost of the minimization, it is nevertheless a good idea to reduce its size as much as possible. In the current implementation, we take advantage of the fact that each of the weighting functions is exactly zero outside its band, and store the corresponding part of the control vector on a grid appropriate to its spectral truncation. In the current implementation, the total dimension of the control vector is approximately three times the dimension of a grid-space representation of the model variables.

The storage required for the vertical correlation matrices is potentially huge. In principle, we could define a different matrix for each grid-point and for each band of wavenumbers. This is completely impractical, and would also require an enormous sample of background errors to generate stable statistics. To reduce the storage requirements, the matrices are stored on a lower resolution grid than the parts of the control vectors, with a maximum resolution (for higher wavenumber bands) of 5°x5°. This still results in statistics files that are uncomfortably large (a few gigabytes). Further ways to reduce their size will be investigated in the future.

Where next?

In conclusion, "Wavelet" J_b provides an elegant way of encapsulating in the covariance model an important property of the statistics of background error, that was not captured by its predecessor. There remain important properties that are not represented, such as the day-to-day variability of background error correlation, and the tendency for error structures to be strongly anisotropic (i.e. functions of direction as well as distance) and to tilt in the vertical. There is a range of possibilities that could be explored to address these issues, each of which captures some or other aspect of background error covariance. But, whatever approach is taken, it is impossible to escape the fact that the covariance model is a distillation of a vast matrix into a relatively small number of parameters. It is inevitable with any distillation that some of the "spirit" is lost, in this case to the detriment of the analysis.

Ultimately, the only way to produce a truly optimal analysis system is to eliminate the dependence of the analysis on a background error covariance matrix. I noted earlier that any given analysis can be regarded as a synthesis of observational information over a period of five-to-ten days, and that the function of the background error covariance model is to propagate this information between analysis cycles. But, what would happen if the analysis could simultaneously take into account all the observations over a five-to-ten day period? In this case, there would be no need to propagate information between cycles, and a background covariance model would be unnecessary. A 4D-Var analysis system of this sort, applied to a simple, low-dimension model of mid-latitude dynamics has been examined (*Fisher*, 2006). The analysis produced by this 4D-Var system is as good as that produced by a full extended Kalman filter. It is likely that attempts to make the background covariance model less important, by increasing the length of the analysis window, have even more potential to improve the analysis than attempts to improve the covariance model itself.

doi:10.21957/ef5601ak43

Further reading

Bartello, P. & **H.L Mitchell,** 1992: A continuous three-dimensional model of short-range forecast error covariances, *Tellus,* **44A,** 217–235.

Cardinali, C., S. Pezzulli & **E. Anderson,** 2004: Influence-matrix diagnostic of a data assimilation system. *Q.J.R. Meteorol. Soc.,* **130,** 2767–2786.

Derber, J.C. & **F. Bouttier,** 1999: Formulation of the background error covariances in the ECMWF global data assimilation system. *Tellus,* **51A,** 195-221.

Fisher, M., 2004a: Background error covariance modelling. *Proc. of ECMWF Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean,* Reading, 8–12 September 2003, 45–64.

Fisher, M., 2004b: Generalized frames on the sphere, with application to the background error covariance modelling. *Proc. of ECMWF Seminar on Recent Developments in Numerical Methods for Atmospheric and Ocean Modelling,* Reading, 6–10 September 2004, 87–102.

Fisher, M., 2006: On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation. *Q.J.R. Meteorol. Soc.* (in print).

 doi:10.21957/ef5601ak43