Spatial Forecast Verification: Thin Plate Splines

Eric Gilleland* and Douglas W. Nychka†

18th June 2009

1. Spatial Forecast Verification

Traditional forecast verification is conducted for a single point location. When faced with a grid of points, the usual approach has been to perform forecast verification point by point. Such a strategy is inadequate because it ignores the spatial structure of the field, is highly susceptible to the double penalty problem (especially for higher resolution forecasts), and it does not provide useful diagnostic information about forecast quality. Numerous new techniques for analyzing spatial forecasts have been proposed, and an international project is underway to compare these methods; specifically, the Spatial Forecast Verification Inter-Comparison Project (ICP, http://www.ral.ucar.edu/projects/icp). A special collection of the *Weather and Forecasting* journal of the American Meteorological society is being prepared with papers pertaining to the ICP. The various methods involved are summarized in Gilleland

*Research Applications Laboratory, National Center for Atmospheric Research, Boulder, Colorado

†Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research, Boulder, Colorado

et al. (2009) and Ahijevych et al. (2009).

Gilleland et al. (2009) classify the various methods into four general categories: neighborhood, scale separation/decomposition, features-based, and field deformation. Although some techniques do not fit as well into these categories, they do represent well the majority of the approaches. Neighborhood approaches employ a type of low-pass filter and apply traditional-type verification approaches to increasingly coarser fields; these methods are nicely summarized in Ebert (2008) with applications to the ICP data cases in Ebert (2009). Scale separation techniques generally involve some type of spectral decomposition (e.g., Fourier, wavelets, etc.), and examples of this approach are numerous (e.g., Briggs and Levine 1997; Tustison et al. 2001; Casati et al. 2004; Casati 2009; Marzban and Sandgathe 2009b; Zepeda-Arce et al. 2000; Lack et al. 2009). While scale decomposition/separation methods can potentially provide forecast performance measures separated by different physical scales, features-based methods attempt to directly identify forecast features and provide detailed information about different types of errors such as spatial displacement, and intensity (e.g., Davis et al. 2006a,b, 2009; Micheas et al. 2006; Lack et al. 2009). Field deformation approaches are most similar to features-based approaches, but they attempt to manipulate the entire forecast field to better match the observed field. Information about the amount and type of manipulation required to reduce traditional scores is then given (e.g., Keil and Craig 2007; Marzban et al. 2009; Marzban and Sandgathe 2009a; Lindström et al. 2009)

The present paper introduces the thin-plate spline as a potentially useful tool for spatial forecast verification. The approach can be used as both a neighborhood type (i.e., low-pass filter) method as well as a scale decomposition approach. In particular, a thin-plate spline can be represented in a symbolically equivalent manner as a wavelet decomposition. Briggs

and Levine (1997) suggest the use of wavelets for forecast verification because of their ability to not only break a field down by physical scales, but also because the locations of the features at each scale are also preserved. For example, if one played a chord on a piano and applied Fourier analysis to decompose the chord, each individual note of the chord could be identified. However, if a second chord were played, again all of the notes can be identified, but no information is available as to which chords the notes belong. Wavelets, on the other hand, allow one to identify the individual notes, and to which chords they correspond (Ogden 1996; Meyer 1993).

2. Scale Decomposition

There are numerous types of scale decompositions that can represent a field. A very brief summary of the wavelet decomposition is given first in section a in order to demonstrate the similarities and differences between wavelets and thin-plate splines.[1] Background on thin-plate splines is then given in section b.

a. Wavelets

A function can be decomposed into wavelets in the following way.

$$f(x) = \sum_j \sum_k c_{j,k}\phi(2^j x - k) + \sum_j \sum_k d_{j,k}\psi(2^j x - k) \tag{1}$$

Here, the j progresses through the different scales, and the k progresses through translations (i.e., giving information about the locations). The ϕ are scaling functions that form

[1] In fact, thin-plate splines are a special case of a larger class of models commonly known as kriging.

a series of basis functions often referred to as father wavelets, and the ψ are basis functions with particular properties referred to as the mother wavelets. The primary distinguishing feature of the wavelet basis functions is that they have a very narrow support, meaning that they are zero outside of a small neighborhood. This allows for location information about features at different scales to be retained in the decomposition. It is important to note that this decomposition involves scales that are progressively halved; that is, one cannot pre-determine specific scales of interest. Finally, the values c and d are coefficients that are realizations of a random process. Note that Eq (1) can be re-written as $f(x) = \boldsymbol{B}\boldsymbol{\alpha}$, where $\boldsymbol{B}$ is a matrix containing the known basis functions, and $\boldsymbol{\alpha}$ is a vector of the coefficients. Subsequently,

$$\text{Var}(f(x)) = \text{Var}(\boldsymbol{B}\boldsymbol{\alpha}) = \boldsymbol{B}\text{Var}(\boldsymbol{\alpha})\boldsymbol{B}^T. \tag{2}$$

The basis functions in Eq (1) lend themselves easily to a decomposition of the estimated function by a sequence of curves representing different scales. That is, the representation provides a multi-resolution analysis.

b. Thin-Plate Splines

Thin-plate splines are commonly used to interpolate data spatially. But, they can also be used to perform a multiresolution analysis (cf. Nychka 2000, and references therein). The basic statistical model has the form

$$Y(\boldsymbol{s}_i) = f(\boldsymbol{s}_i) + \varepsilon_i \text{ for } 1 \leq i \leq n \tag{3}$$

where $Y(\boldsymbol{s}_i)$ is the observed response at the i^{th} combination of design variables, or i^{th} location, $\boldsymbol{s}_i \in \Re^d$, f is the function of interest, and ε_i are random components usually

associated with measurement error and assumed to be uncorrelated with zero-mean and variance σ^2/W_i.

Here, we consider the strategy of representing f as the sum of a low-order polynomial and a smooth function

$$f(\boldsymbol{s}) = P(\boldsymbol{s}) + h(\boldsymbol{s}). \tag{4}$$

Under the assumption that f is a realization of a spatial process, the polynomial term $P(\boldsymbol{s})$ is identified as the spatial trend or drift, and h is a mean-zero Gaussian process with

$$\mathrm{Cov}(h(\boldsymbol{s}), h(\boldsymbol{s}')) = \rho k(\boldsymbol{s}, \boldsymbol{s}'). \tag{5}$$

The thin-plate spline can be represented as an orthogonal series estimate. That is, a set of basis functions, $\{g_\nu\}$ are constructed for $1 \leq \nu \leq n$ such that if $u_\nu = \sum_{k=1,\ldots,n} g_\nu(\boldsymbol{s}_k) W_k Y_k$, then the function, $\sum_{\nu=1,\ldots,n} u_\nu g_\nu(\boldsymbol{s})$ interpolates the field, $Y(\boldsymbol{s}_k)$.

Furthermore, there is a sequence of increasing nonnegative weights, or eigenvalues, $\{D_\nu\}$ such that

$$\hat{f}(\boldsymbol{s}) = \sum_{\nu=1}^{n} \frac{1}{1 + \lambda D_\nu} u_\nu g_\nu(\boldsymbol{s}), \tag{6}$$

where λ is a spline smoothing parameter, or the signal to noise ratio, $\lambda = \sigma^2/\rho$.

For spatial process models, the usual kriging surface is a linear combination of low-order polynomial functions and n functions, $\psi_j(\boldsymbol{s}) = k(\boldsymbol{s}, \boldsymbol{s}_j)$. For the thin-plate spline, the $\psi_j(\boldsymbol{s})$ are given by radial basis functions.

In particular, the thin-plate spline estimator for f can be written as

$$f(\boldsymbol{s}) = \sum_{j=1}^{t} \beta_j \phi_j(\boldsymbol{s}) + \sum_{i=1}^{n} \delta_i \psi_i(\boldsymbol{s}) \tag{7}$$

where $\{\phi_j\}$ is a set of t polynomial functions that span the space of all d dimensional polynomials with degree less than or equal to $m-1$, and $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are parameters. The above

is very similar in form to Eq (1), but there are some important differences. In particular, Eq (7) does not impose any constraint on the scales that can be utilized in the reconstruction of f. Furthermore, the support for $\boldsymbol{s}$ in Eq (7) can be an irregular grid, or simply a smandering of points besprinkled about a region of space.

Although it is not as precise or compact as a wavelet representation, the orthogonal series representation also lends itself to a multi-resolution analysis. An advantage, however, is that it extends to irregularly spaced observations and provides a useful summary of how the estimated surface changes as a function of the smoothing parameter. A simple way to obtain such an analysis is to look at differences of thin-plate spline estimates for an increasing sequence of degrees of freedom (or, equivalently, a decreasing sequence of smoothing parameters). Specifically, let $\{\nu_j\}$, $0 < j \leq J$ be an increasing sequence of degrees of freedom with $\hat{d}_J(\boldsymbol{x}) \equiv \hat{f}_J(\boldsymbol{x})$, and for $j \geq 1$

$$\hat{d}_j(\boldsymbol{x}) = \hat{f}_{\nu_j}(\boldsymbol{x}) - \hat{f}_{\nu_{(j-1)}}(\boldsymbol{x}). \tag{8}$$

The differences, $\hat{d}_j$, are referred to as the detail surfaces because they capture the detail between the estimates at two different levels of smoothing. By the telescoping property of the differences, $Y_i = \sum_{j=1}^{J} \hat{d}_j(\boldsymbol{x}_i)$.

The thin-plate spline estimate itself is a low-pass filter applied to the data, and the differences of two such splines should behave as a a band-pass filter. That is, it passes components in a range of frequencies while omitting frequencies that are either very low or very high; and a band-pass filter is constructed by an appropriate combination of low- and high-pass filters. In terms of spatial forecast verification, therefore, the thin-plate spline can be used in a neighborhood approach or in a scale-separation approach. The former concept

is presented in section 4a and the latter in section 4b. First, simulated observed and forecast fields are described.

3. Baddeley's Δ metric

In order to make better use of the spatial information for both the neighborhood and scale-separation approaches, we use a very simple metric that is sensitive to size, shape and location errors. While it does not give information about these errors individually, it does give an excellent indication as to how similar two spatial patterns are. Specifically, we use Baddeley's Δ metric for comparing binary images (Baddeley 1992b,a). Gilleland et al. (2008) used this metric to merge and match objects obtained through the features-based MODE application, and found that it made very reasonable comparisons of spatial patterns.

The metric is a modification of the Hausdorff metric. Specifically, the Δ metric (for discrete pixel rasters) used here for comparing two binary images, A and B, is given by

$$\Delta(A,B) = \left[\frac{1}{N}\sum_{i=1}^{N}|\omega(d(\boldsymbol{x},A)) - \omega(d(\boldsymbol{x},B))|^2\right]^{1/2}, \tag{9}$$

where $\omega(z) = \min(z,c)$, with c a user-specified constant, the sum is over every pixel, $\boldsymbol{x}$, in the raster (in this case, every grid-point in the field), and $d(\boldsymbol{x},A)$ is the shortest (Euclidean) distance between $\boldsymbol{x}$ and the set A. The shortest distances, d, are calculated efficiently using the distance transform method.

4. Concept Example

The present treatment is intended as a proof of concept only. To apply these techniques to verify real meteorological forecasts, a few major issues would need to be resolved. These issues and proposed methods for addressing them are alluded to in section 5. Here, two relatively small fields are randomly generated, and a regular thin-plate spline is used to verify one field (field B) against the other (field A). The two fields are generated using `sim.rf` from the R (R Development Core Team 2008) package `fields` (Furrer et al. 2008). First, two 30×30 Gaussian random fields are generated using an exponential covariance function with range parameter 2.0 on the domain $[0, 1] \times [0, 1]$ independently of each other. Second, all negative values of the fields are set to zero and assigned the labels A and B, where A is considered to be the observed field and B its associated forecast. The resulting fields are shown in figure 1, and thin-plate spline smoothed versions of these fields are shown in figure 2 for different degrees of freedom.

A couple of properties stand out in figure 2. First, the fewer degrees of freedom, the smoother the field. For 5 degrees of freedom, the field is close to the average of the field, but there are higher values following a general pattern around the densest areas of activity so that for fewer degrees of freedom, the larger scale features are retained. One the other hand, for larger degrees of freedom, the fields pick out more of the smaller scale features of each field as expected. Second, the interpolation frequently yields negative values (white spaces in the figure). If one is investigating fields such as precipitation, where there are many zeros and negative values are not plausible, then one must address this point.

In the sections that follow, we apply the thin-plate spline decompositions of the randomly

generated fields to verify one field against the other in various different ways to glean different types of information. First, we apply traditional verification scores to the variedly smoothed fields and for increasing thresholds in a way analogous to the neighborhood methods (cf. Gilleland et al. 2009) in section a. In section b, we apply a scale separation technique that also utilizes the Baddeley binary image metric in order to take advantage of the spatial information allowed by the thin-plate spline approach for the different detail levels.

a. A thin-plate spline neighborhood approach

Some traditional verification scores for the raw fields for varying thresholds are shown in the top bands of the quilt plots of figure 3. The same scores are also shown in these quilts as applied to thin-plate spline smoothed fields for the same thresholds and varying degrees of freedom. For very low thresholds, the probability of detecting an event is reasonable, but this score drops to zero quickly as the thresholds increase. This can be attributed to the fact that the two fields are randomly generated from the same distribution so that much of the field is covered for a very low threshold, but as the threshold increases, the placement of higher values generally do not overlap. The percent correct incorporates getting the correct negatives along with the higher values so that the opposite is true. The frequency bias is good for most thresholds and smoothed fields, which makes sense given that both come from the same distribution. For higher thresholds, the "forecast" field B over predicts the spatial extent of very high values. Because the bias near one for most cases, the Gilbert Skill Score makes sense, and indicates that the forecast is generally poor. Because these fields were randomly generated, this is meaningful, but if the field B was a real forecast

that was slightly displaced, this result might not be much different. In terms of the scale of features, the results here do not show much difference in traditional scores except between the very coarsest scales (5 to 55 degrees of freedom) indicating that there is a fair amount of disagreement between the two fields for the large-scale structures. Checking the original fields in figure 1, this conclusion is substantiated given that the large-scale patterns are quite different.

In addition to the traditional verification scores, figure 3 (lower right panel) shows the Baddeley Δ metric from Eq (9) with $c = 1/2$ for each threshold and levels of smoothing. Using only 5 degrees of freedom (i.e., very smooth with fairly small values likely resulting from the numerous zero values in the field), we see that the skill is high (i.e., $\Delta \approx 0$) for most thresholds, except for thresholds of about 0.3 and 0.6 where there remain values above these thresholds, but the resulting spatial placement of excesses are wildly different. That is, the large scale structure of the two fields are spatially very different for higher values (especially for the threshold of about 0.6). Not at all surprising for randomly generated fields such as these, but this would be of major concern if this were a real forecast. For higher thresholds, and lower degrees of freedom, both fields are exactly zero, yielding $\Delta = 0$ exactly. As the degrees of freedom increases, the fields have more activity even at higher thresholds. The spatial placement of threshold excesses for the two randomly generated fields compare relatively well for moderate thresholds from zero to about 1.5 for all degrees of freedom above about 155. The string of oranges and reds shows threshold/degrees of freedom pairs where the spatial placement of threshold excesses is very poor. Again, such high values of Δ would not be likely for a real forecast, and would otherwise be of great concern; but are not at all unexpected for two randomly generated fields.

b. A thin-plate spline scale-separation approach

A few characteristic detail fields are shown in figure 4. Higher degrees of freedom used in predicting a field with a thin-plate spline lead to more detailed fields so that the difference between two predicted fields with high degrees of freedom (e.g., middle and bottom panels of figure 4) results in very fine-scale information (e.g., several small convective storms). Similarly, fewer degrees of freedom used in such spatial predictions result in highly smoothed fields, and difference between two fields with low degrees of freedom result in larger scale details (e.g., top panel of figure 4).

Figure 5 shows quilt plots for the Baddely Δ binary image metric from Eq (9) again with $c = 1/2$, the relative difference in energy between fields A and B, and the energy for fields A and B individually. The quilts show these results for several detail levels (ordinate axes) and thresholds (abscissas). Lower values of Baddeley's Δ indicate better matched fields. For higher thresholds, and especially for fewer degrees of freedom,[2] the two fields are often zero so that all of these statistics are perfectly zero beyond a certain point.

Note the stark contrast between the quilt plots for Baddeley's Δ metric of figure 3 (neighborhood approach) and figure 5 (scale-separation approach). The multi-resolution analysis yields metrics for different scaled features within the field, whereas the neighborhood approach yields metrics for amounts of smoothing. Therefore, the neighborhood approach reflects the fact that for these randomly generated fields, the highly smoothed fields (low degrees of freedom) are not very similar in terms of spatial pattern except for very low

[2]cf. figure 2 where the fewer degrees of freedom give coarser scale information, but also for these fields, the coarser scale fields have much smaller amplitudes.

thresholds, whereas the less smoothed fields (higher degrees of freedom) have more in common in terms of spatial pattern, at least for low to moderate thresholds. On the other hand, the scale-separation (or multi-resolution analysis) provides information about how well the spatial patterns of the detail fields match, with coarser scale details having good pattern matching only for very small thresholds, with poor matching at about 0.07; with both detail fields being exactly zero beyond this threshold. Smaller scale features have better spatial pattern matching for slightly higher thresholds, but also show that for more extreme small-scale details, the matching is quite poor. To a point, the two different methods yield very similar information about "forecast" performance, but in different ways.

5. Discussion

The concept example portrayed here is simplistic for a reason. The thin-plate spline in its most basic form is not computationally efficient. For a real forecast and analysis comparison, the grids are too big for this approach to work as it stands. This is perhaps the biggest hurdle over which to leap in order for this to be a manageable method. However, methods do exist **[citation]** for speeding up the computations so that it might be feasible to use the thin-plate spline approach with some modification.

Other major issues that need to be resolved include handling of zeros. This is an issue for fields such as quantitative precipitation, but for some fields, this is not an issue. As is evidenced by the concept examples herein, the thin-plate spline smoothes out the zeros so that the field values become very small for smoother fields. Finally, although prediction intervals fall out of the thin-plate spline approach rather naturally, obtaining confidence

intervals for the final summary scores needs to also be addressed.

Despite these above issues, the approach shows great promise, and both the computational inefficiency as well as for handling zeros better. With some care, it should also be possible to obtain parametric or even nonparametric confidence intervals for the final summary scores. Otherwise, this approach has some substantial tactical advantages over other neighborhood and scale-separation methods.

Finally, use of the Baddeley Δ metric for summarizing spatial pattern information is one very simple method for utilizing the spatial information and informing about forecast performance in this regard. Other more involved methods exist that can give more detailed information, such as the features-based methods or the field deformation approaches. Such methods could be applied to either the smoothed (neighborhood approach) fields, or to the detail fields (scale-separation approach). Nevertheless, the Baddeley Δ metric along with the energy information gives a pretty complete picture of forecast performance when used in conjunction with the neighborhood or scale-separation methods. That is, information is provided about intensity and spatial pattern errors for different scales of interest; either physical scales or resolution, though it is found here that the two are not necessarily practically different.

References

Ahijevych, D., E. Gilleland, B. Brown, and E. Ebert, 2009: Application of spatial verification methods to idealized and nwp gridded precipitation forecasts. *Submitted to Wea. Forecasting*, 1–34.

Baddeley, A., 1992a: An error metric for binary images. *Forstner and S. Ruwiedel (ed.) Robust Computer Vision Algorithms, Proceedings, International Workshop on Robust Computer Vision, Bonn. Karlsruhe: Wichmann, 559-78.*.

— 1992b: Errors in binary images and an l_p version of the hausdorff metric. *Nieuw Archief voor Wiskunde*, **10**, 157–183.

Briggs and Levine, 1997: Wavelets and field forecast verification. *Mon. Wea. Rev.*, **125**, 1329–1341.

Casati, B., 2009: New developments of the intensity-scale technique within the spatial verification methods inter-comparison project. *Submitted to Wea. Forecasting*.

Casati, B., G. Ross, and D. Stephenson, 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorol. Appl.*, **11**, 141–154.

Davis, C., B. Brown, and R. Bullock, 2006a: Object-based verification of precipitation forecasts, part i: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784.

— 2006b: Object-based verification of precipitation forecasts, part ii: Application to convective rain systems. *Mon. Wea. Rev.*, **134**, 1785–1795.

Davis, C., B. Brown, R. Bullock, and J. Halley Gotway, 2009: The method for object-based diagnostic evaluation (mode) applied to wrf forecasts from the 2005 spc spring program. *Submitted to Wea. Forecasting*.

Ebert, E., 2008: Fuzzy verification of high resolution gridded forecasts: A review and proposed framework. *Meteorol. Appl.*, **15**, 51–64, doi:DOI: 10.1002/met.25.

— 2009: Fuzzy (neighborhood) verification a strategy for rewarding close forecasts. *Submitted to Wea. Forecasting*.

Furrer, R., D. Nychka, and S. Sain, 2008: *fields: Tools for spatial data*. R package version 5.01.

URL `http://www.image.ucar.edu/Software/fields`

Gilleland, E., D. Ahijevych, and B. Ebert, 2009: Spatial forecast verification methods intercomparison. *Accepted to Wea. Forecasting*.

Gilleland, E., T. Lee, J. Halley-Gotway, R. Bullock, and B. Brown, 2008: Computationally efficient spatial forecast verification using baddeley's δ image metric. *Mon. Wea. Rev.*, **136**, 1747–1757.

Keil, C. and G. Craig, 2007: A displacement-based error measure applied in a regional ensemble forecasting system. *Mon. Wea. Rev.*, **135**, 3248–3259.

Lack, S., G. Limpert, and N. Fox, 2009: An object-oriented multiscale verification scheme. *Submitted to Wea. Forecasting*.

Lindström, J., E. Gilleland, and F. Lindgren, 2009: Image warping for forecast verification. *Manuscript in Preparation*.

Marzban, C. and S. Sandgathe, 2009a: Optical flow for verification. *Accepted to Wea. Forecasting*.

— 2009b: Verification with variograms. *Accepted to Wea. Forecasting*.

Marzban, C., S. Sandgathe, H. Lyons, and N. Lederer, 2009: Three spatial verification techniques: Cluster analysis, variogram, and optical flow. *Submitted to Wea. Forecasting*.

Meyer, Y., 1993: *Wavelets: Algorithms and Applications*. SIAM: Philadelphia, Pennsylvania, 133pp.

Micheas, A., N. Fox, S. Lack, and C. Wikle, 2006: Cell identification and verification of qpf ensembles using shape analysis techniques. *Submitted to J. of Hydrology*.

Nychka, D., 2000: *Smoothing and Regression: Approaches, computation and application*, Edited by M.G. Schimek, Wiley Series in Statistics, New York, chapter 13. Spatial process estimates as smoothers. 393–423.

Ogden, T., 1996: *Essential Wavelets for Statistical Applications and Data Analysis*. Birkhäuser, Boston, MA, 224pp.

R Development Core Team, 2008: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. URL `http://www.R-project.org`

Tustison, B., D. Harris, and E. Foufoula-Georgiou, 2001: Scale issues in verification of precipitation forecasts. *J. Geophys. Res.*, **106**, 11775–11784.

Zepeda-Arce, J., E. Foufoula-Georgiou, and K. Droegemeier, 2000: Space-time rainfall organization and its role in validating quantitative precipitation forecasts. *J. Geophys. Res.*, **105**, 10,129–10,146.

List of Figures

5 *Quilt plots for varying details as defined by the thin-plate spline predictions using varying degrees of freedom, $df_{j+1} - df_j$ (ordinate axes). Thresholds vary along the abscissa. Upper left graph shows the Baddeley's Δ binary image metric (Eq. (9) with $c = 1/2$), upper right the relative energy difference, lower left the energy for field A (i.e., observed), and lower right for field B (i.e., forecast).* 23

FIG. 1. *Simulated* 30×30 *Gaussian random fields using a stationary exponential covariance function with range 2.0 on the domain* $[0,1] \times [0,1]$*, with all values less than zero set equal to zero. Top is denoted field A and bottom field B.*

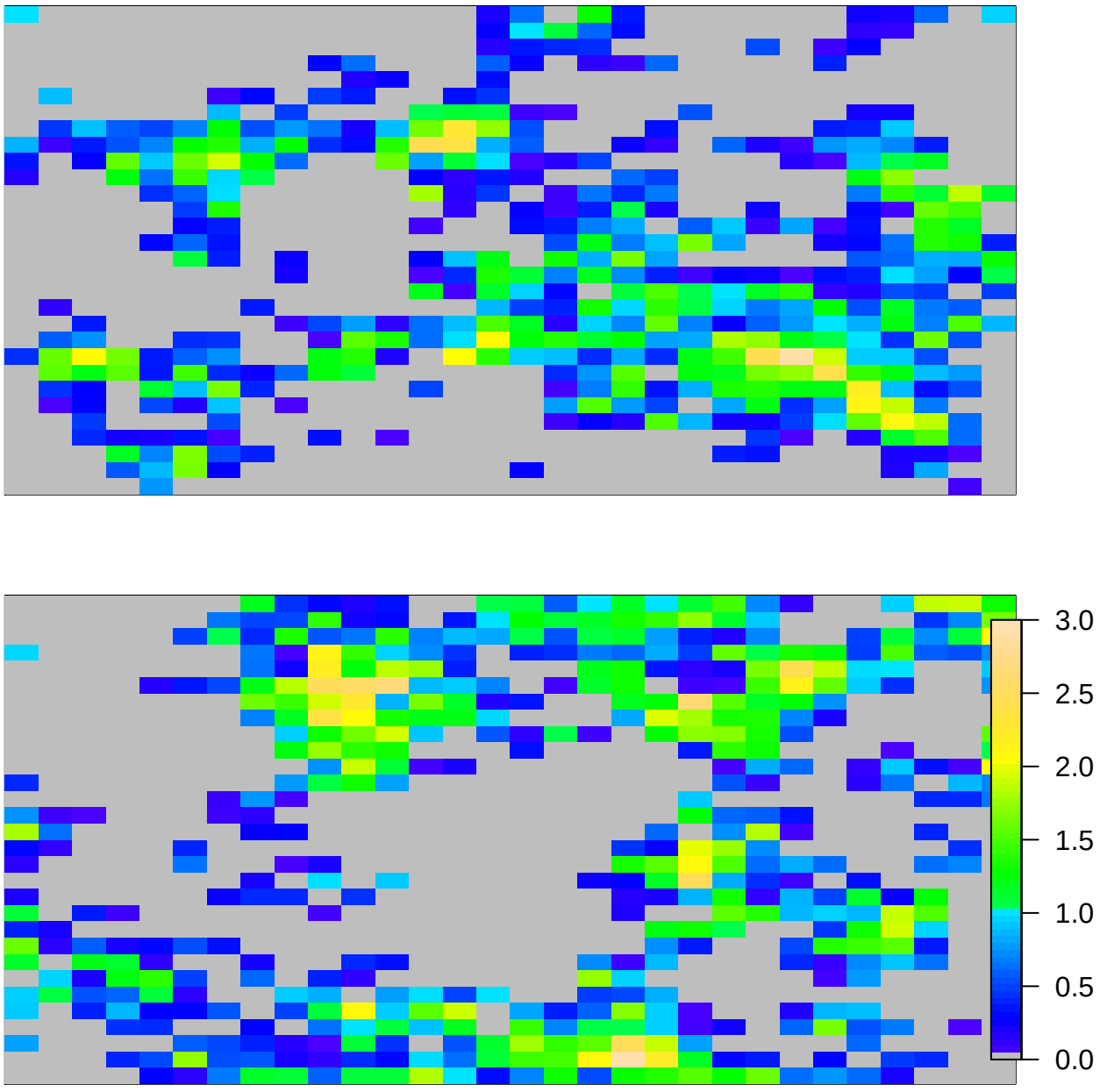

FIG. 2. *Thin-plate spline smoothed versions of the randomly generated fields from figure 1. Left column is for field A, and the right column for field B. From top to bottom, rows represent the thin-plate spline smoothes with 5, 155, 455 and 805 degrees of freedom, respectively. White spaces indicate that the smoothed field is less than zero at these grid points.*

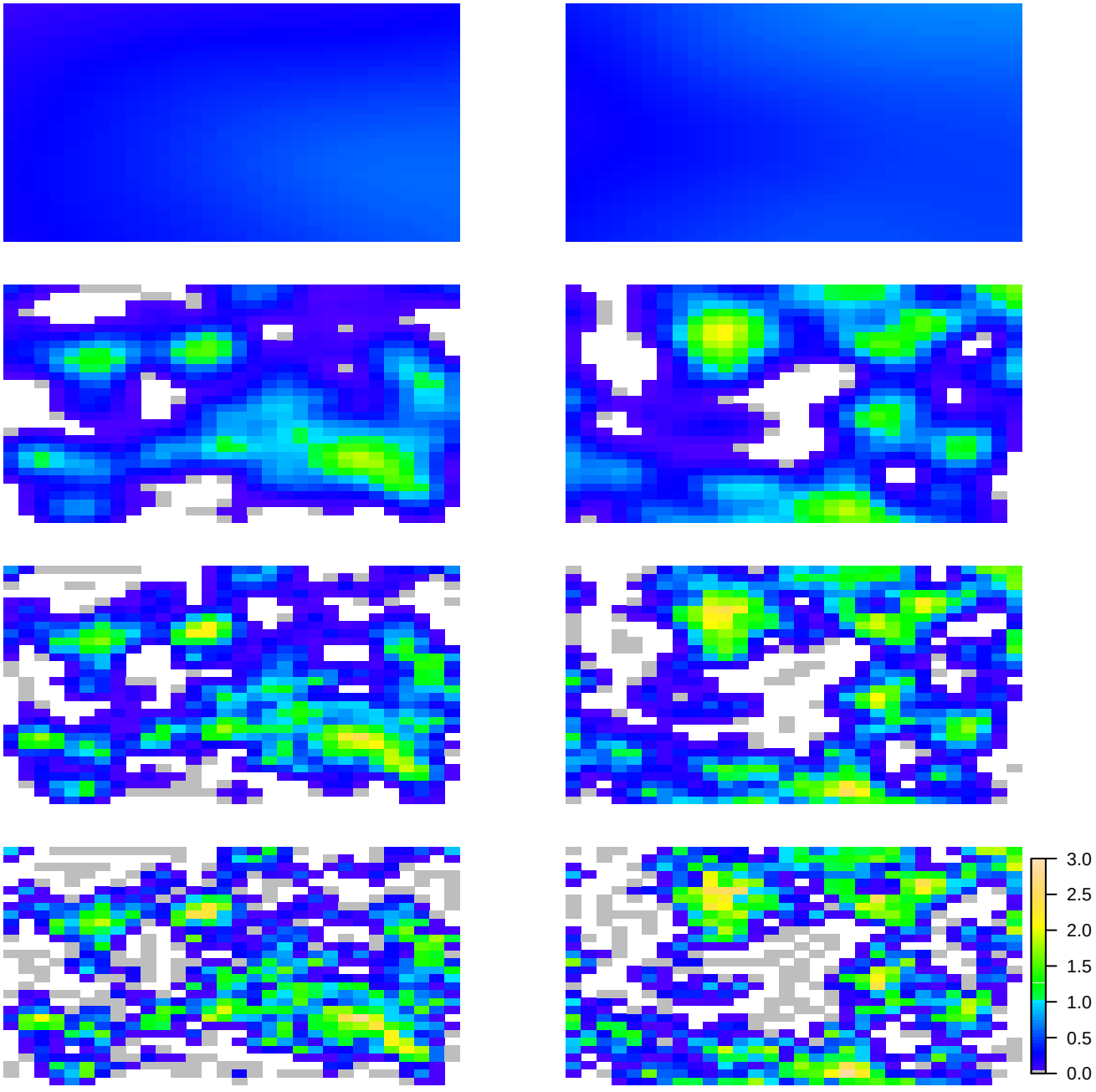

FIG. 3. *Quilt plots showing traditional verification scores (and the Baddeley Δ metric) for varying thresholds (abscissas) and thin-plate spline smoothes with varying degrees of freedom (ordinates). Top band of each quilt are the traditional scores applied to the raw fields (no smoothing). Top row of quilt plots are probability of detecting an event (PODy), Gilbert Skill Score (GSS), Heidke Skill Score (HSS). Bottom row of quilt plots are percent correct (PC), (frequency) bias, and Baddeley's Δ metric (Eq. (9) with $c = 1/2$).*

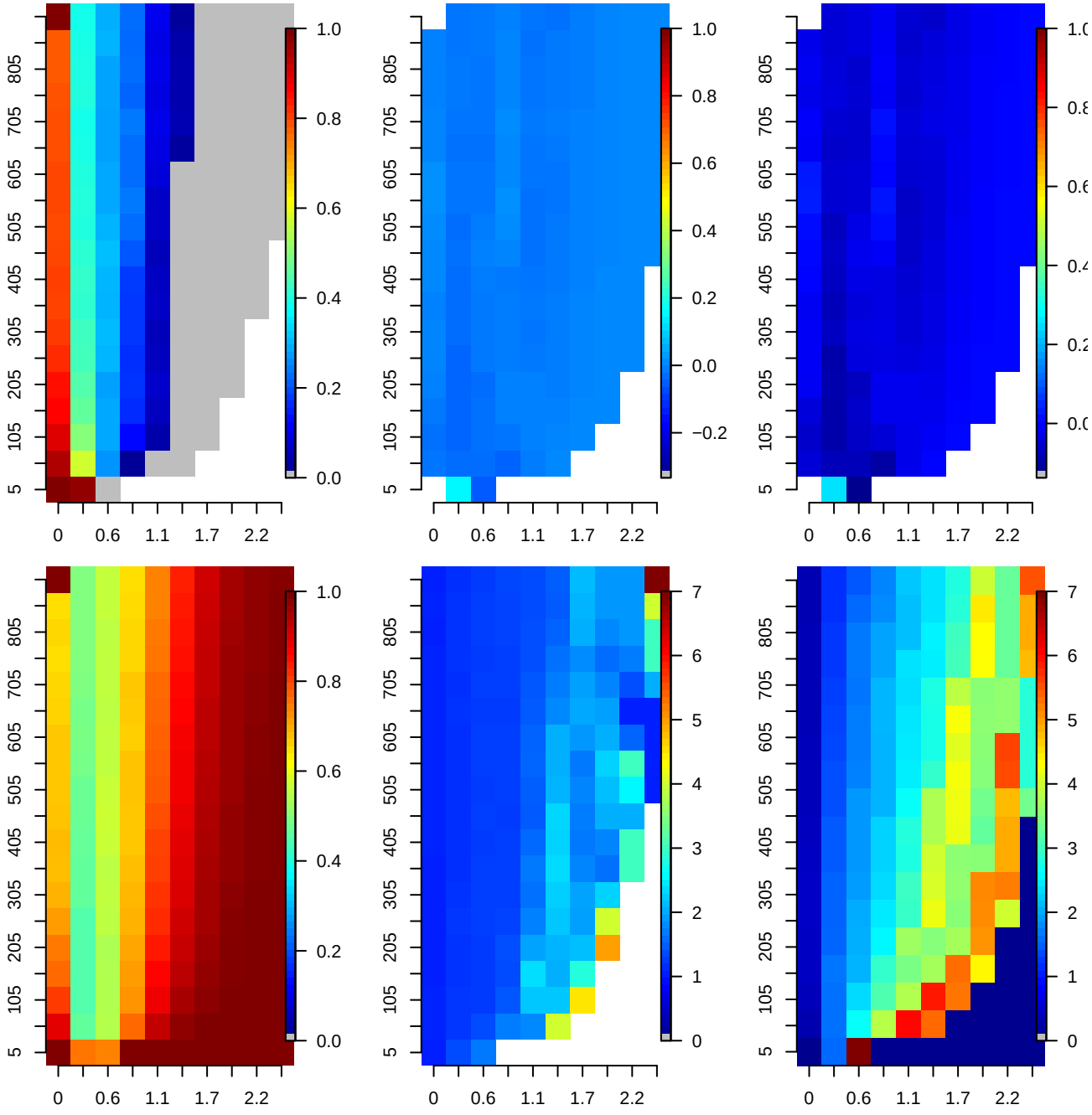

FIG. 4. *Detail fields based on thin-plate spline smoothers for degrees of freedom differences of:* $105 - 55$ *(top),* $405 - 355$ *(middle), and* $605 - 555$ *(bottom). First column is from simulated field A and the second from simulated field B.*

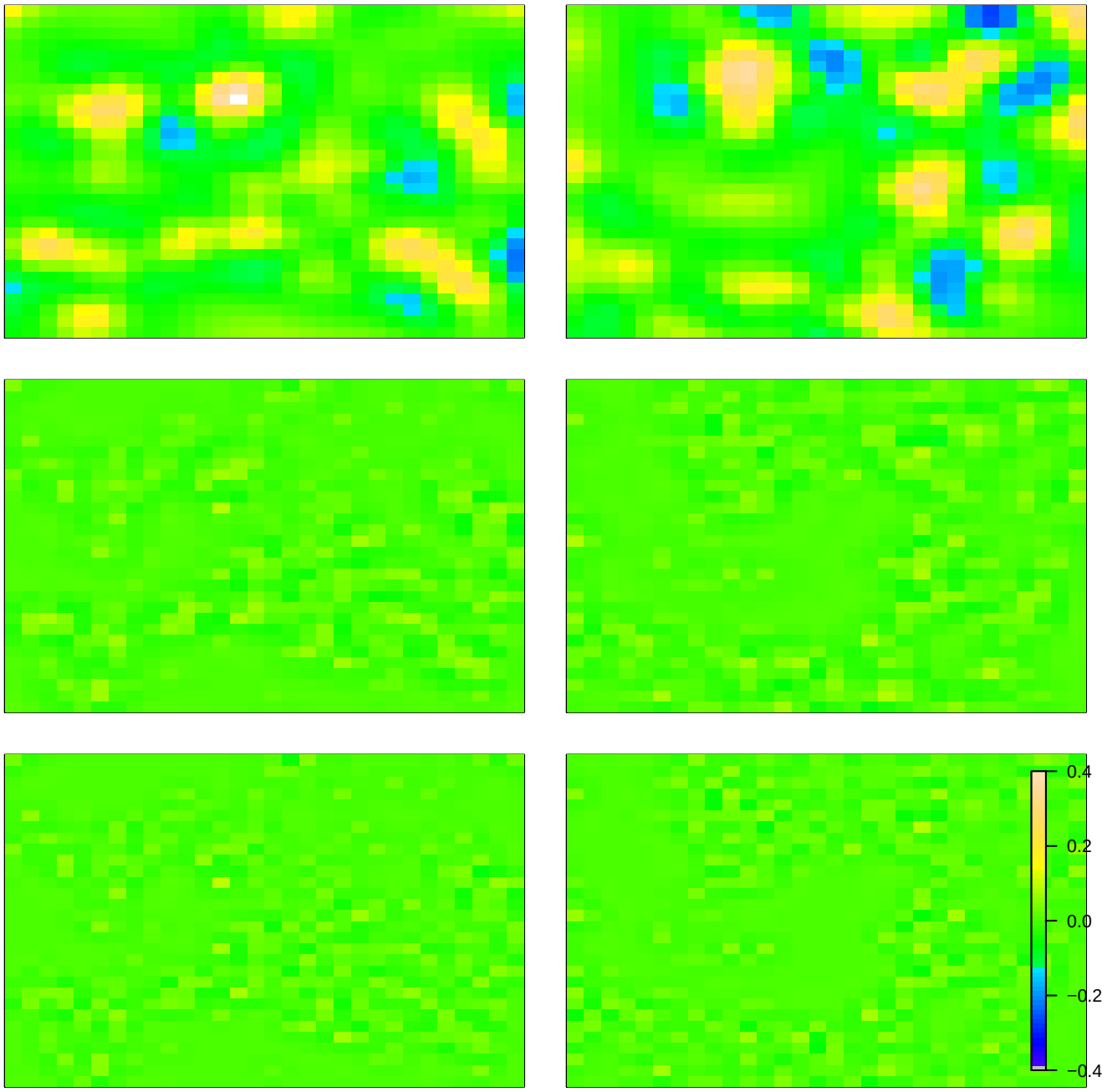

FIG. 5. *Quilt plots for varying details as defined by the thin-plate spline predictions using varying degrees of freedom, $df_{j+1} - df_j$ (ordinate axes). Thresholds vary along the abscissa. Upper left graph shows the Baddeley's Δ binary image metric (Eq. (9) with $c = 1/2$), upper right the relative energy difference, lower left the energy for field A (i.e., observed), and lower right for field B (i.e., forecast).*