

Forecast verification using image warping

Johan Lindström* Eric Gilleland† Finn Lindgren‡

14th October 2009

Abstract

TODO

1 Introduction

Higher resolution weather models have greatly improved their ability to forecast weather phenomena, and subsequently their usefulness. However, traditional forecast verification scores (e.g., root mean squared error (RMSE), probability of detection (POD), etc.; Jolliffe and Stephenson (2003); Wilks (2006)) generally suggest that the models do not perform as well because they are calculated by comparing single grid-points from the forecast field to their corresponding single grid-points in the observation field allowing slight displacement errors to heavily penalize the overall score. This is a well known problem, and numerous new verification methods have recently been proposed to give more useful information about forecast performance. All of the methods proposed loosely fall into four categories: (i) features-based, (ii) scale-decomposition, (iii) neighborhood, and (iv) field verification. Many of the methods are summarized in ?.

The features-based methods attempt to define features of interest from both the forecast and observation fields (usually by applying a threshold to the raw fields), and then compute information on different sources of error such as spatial displacement, feature intensities, spatial coverage, among others. These methods primarily differ from one another in the way they identify features within a field and match features across fields, and in how they characterize the various sources of error. Ebert and McBride (2000) merge features within a field if an

*Mathematical Statistics, Centre for Mathematical Sciences, Lund University, Sweden

†Research Applications Laboratory, National Center for Atmospheric Research, U.S.A.

‡Mathematical Statistics, Centre for Mathematical Sciences, Lund University, Sweden

enlarged feature area overlaps, and matches across fields are determined through an optimization of a pattern matching criterion of translations of the features. Several modifications to this procedure have been proposed (e.g., Ebert et al., 2004; Grams et al., 2006). Davis et al. (2006a,b) use a convolution threshold procedure to identify features, and a fuzzy logic algorithm applied to various feature attributes (such as size, centroid distance, etc.) to determine merges, matches and overall forecast performance. Micheas et al. (2006) use Procrustes shape analysis and a user-defined penalty function to determine overall forecast performance. Nachamkin (2004) compares only the distributions of features in each field through a composite approach. Marzban and Sandgathe (2006) employ a very different approach involving statistical cluster analysis.

Scale-decomposition methods were first introduced for the purpose of forecast verification by the seminal paper of Briggs and Levine (1997). Several variations on the idea have subsequently been proposed (e.g., Casati et al., 2004; Harris et al., 2001). Generally, these methods compute various error metrics at different scales of a single-band spatial filter (e.g., Fourier transforms, wavelets, etc.).

Numerous neighborhood methods have been introduced, and they are nicely summarized in ? (referred to there as *fuzzy* methods). These methods are a slight modification of the traditional verification procedures. Instead of calculating the various scores on a grid-point to grid-point basis, functions of the values at neighboring grid points are computed and compared. This has the effect of blurring (or smoothing) the two fields and then applying the usual traditional techniques. The techniques are performed using neighbors within a given radius, and usually this is done for multiple radii so that information about the useful resolution of a forecast is gleaned.

Of interest to the present work are the field verification methods, which attempt to deform one field (or subregion of the field) to match the other field (subregion) in some optimal way. Keil and Craig (2007) combine the magnitude of a displacement vector calculated with a pyramid algorithm and the local squared difference of observed and morphed forecast brightness fields and Marzban (personal communication) is investigating the use of optical flow techniques for verification purposes. The method of Nehrkorn et al. (2003) distorts the forecast field using a technique referred to as feature calibration and alignment (FCA), which is perhaps more closely related to the scale-decomposition techniques, but does involve a distortion of the forecast field to better match the observed field. Dickinson and Brown (1996) and Alexander et al. (1999) applied an image warping technique using polynomial warp functions. Dickinson and Brown (1996) chose their

landmarks automatically using covariate information, whereas Alexander et al. (1999) manually selected landmarks (referred to as “tie points” therein) to match important features and ensure physically meaningful warps. The field verification approaches are related to the features-based approaches in that both attempt to manipulate the respective fields to match up as best as possible. The difference being that the features-based approaches are usually concerned with individual features of each field whereas the field verification approaches are concerned with the field (or subregion of the field) as a whole entity. Features-based approaches typically require an algorithm for finding optimal merges and matches of features, and techniques used for the field verification methods should prove useful for this purpose as well.

In Åberg et al. (2005) an image warping approach to short term precipitation forecasts was introduced. Our goal is to use a similar method for forecast verification. Image warping has previously been used in a number of different applications; Good introductions and overviews are given in Glasbey and Mardia (1998) and Dryden and Mardia (1998). Examples of the use of image warping includes such diverse applications as construction of non-stationary fields in spatial statistics (Sampson and Guttorp, 1992; Guttorp et al., 1994; Meiring et al., 1998; Damian et al., 2001), image registration problems (Zitova and Flusser, 2003) and matching of biological images (Sorzano et al., 2005). The principal motivation for those series of papers is to transform co-ordinates of measurement stations to co-ordinates where the dependence between the stations can be modeled using a stationary covariance structure. When the field is transformed back to the original co-ordinates this results in a non-stationary covariance structure. It has also been pointed out (Åberg et al., 2005) that the warping approach can be interpreted as a special case of the spatio-temporal modeling using integro-difference equations as presented and used in Wikle and Cressie (1999); Wikle (2002) and Xu et al. (2005).

Here the warping will be used to deform a set of forecast fields to match an observation field. This gives an estimate of the spatial errors, represented by the warping deformations of the forecasts. Using this approach we hope to quantify the location error in the forecasts, gathering valuable information about the spatial errors and uncertainties. Due to the statistical formulation of the image warping problem we are also able to give approximate confidence intervals for several features of the deformation; Such as global translation, global rotation and the reduction in the amplitude error that the warping deformation gives rise to, i.e. how much better than the original forecast the deformed forecast is. The approach presented here differs markedly from previous studies (e.g., Alexander et al., 1999; Dickinson and Brown,

1996) because we employ thin plate spline-based warp functions. As pointed out by Glasbey and Mardia (1998) thin plate splines allows for more local features than polynomial warping functions. The thin plate splines also allows for smooth warps and the parametric approach using pseudo-landmarks provides a way of building a statistical hierarchical model for the entire warp (Glasbey and Mardia, 2001). **We should probably change this, since we're just putting up a grid of landmarks** Further, we employ a procedure that enables for computationally efficient and automatic selection of landmark points without relying on outside information or other pitfalls associated with trying to match features across fields.

When deforming a forecast field to match an observation field a perfect match might be attainable but require unrealistically rough (non-linear) deformations. Thus some sort of restriction on the warping function that penalizes very rough deformations is needed. This results in a penalized likelihood problem where the likelihood measures the similarity between the observation and (deformed) forecast field. While the penalty, which is expressed as a prior on the warping function, measures the distortion of the deformations (Glasbey and Mardia, 2001).

2 Method

The method consists of three main parts: A stochastic model that relates the observation field to the forecast field, a way of estimating the stochastic model, and an ability to construct approximate confidence intervals from the model. However to construct the model a fourth part, the image warping, is needed. This section describes, in turn, the image warping, the stochastic model, the estimation and the construction of approximate confidence intervals.

2.1 Image warping

Image warping is essentially a smooth mapping between two image planes, that maps each point in one plane to a point in another plane. A pixel at point \mathbf{s} in the new image, \mathcal{Y} , is assigned the pixel value of the undeformed image, Y , taken at $\Psi(\mathbf{s})$, i.e.

$$\mathcal{Y}(\mathbf{s}) = Y(\Psi(\mathbf{s})), \forall \mathbf{s} \in \mathcal{D},$$

where \mathcal{D} is the support of the image. Note that the warping function

$$\Psi : \mathbb{R}^2 \mapsto \mathbb{R}^2,$$

maps coordinates from the deformed image to the undeformed image, i.e. the pixel values are moved from Y to \mathcal{Y} , while the warping function Ψ describes a co-ordinate transform from \mathcal{Y} to Y .

Two problems now arise: First the images are represented by pixels defined on a discrete grid, if the mapping, $\Psi(\mathbf{s})$, of a pixel does not coincide with a discrete pixel in Y some interpolation is needed to obtain the pixel value that should be assigned to $\mathcal{Y}(\mathbf{s})$. Here we use bilinear interpolation of the neighbouring pixels since this gives a reasonable trade off between speed and accuracy. Second if a point is mapped to a point outside of the domain of the image, $\Psi(\mathbf{s}) \notin \mathcal{D}$, some way of selecting this pixel value is needed. Glasbey and Mardia (2001) set the values to zero while Åberg et al. (2005) used kriging to extrapolate the image values. For the forecast verification a choice between these, and other, approaches is going to be determined by the kind of data that is being verified. In the result section we analyse cases where adding zeros as well as kriging is appropriate.

In the following we are going to assume that the warping function is a continuous function that interpolates a set of control points from the deformed image plane to the undeformed image plane, i.e.

$$\Psi(\boldsymbol{\rho}_k) = \mathbf{p}_k, \quad k \in [1, P], \quad (1)$$

where \mathbf{p}_k and $\boldsymbol{\rho}_k$ denotes the k^{th} (of P) control points in the undeformed and deformed images respectively. If the control points are known they uniquely define the deformation and we write, $\Psi(\mathbf{s}; \boldsymbol{\rho}, \mathbf{p})$. Moving the control points will alter Ψ , and thus change the deformed image, see Figure 1. Since the deformation is a function of the control points we're going to assume that $\Psi(\mathbf{s}; \boldsymbol{\rho}, \mathbf{p})$ is twice differentiable w.r.t. \mathbf{p} . Here we will use the common choice of Thin plate splines for Ψ (e.g. Sampson and Guttorp, 1992; Glasbey and Mardia, 2001; Åberg et al., 2005), although several other families of transformations exist (Goshtasby, 1987; Lee et al., 1997; Glasbey and Mardia, 1998, for some alternatives, see).

2.1.1 Thin plate splines

Selecting a thin plate spline function implies that the warping function, Ψ , can be seen as a pair of thin plate splines

$$\Psi(\mathbf{s}) = \left(\Psi_x(\mathbf{s}), \Psi_y(\mathbf{s}) \right),$$

where $\Psi_x(\mathbf{s})$ and $\Psi_y(\mathbf{s})$ are functions from \mathbb{R}^2 to \mathbb{R}^1 that minimise the bending energy,

$$\mathcal{J}(\Psi_x) = \iint_{\mathbb{R}^2} \left(\frac{\partial^2 \Psi_x(\mathbf{s})}{\partial s_x^2} \right)^2 + 2 \left(\frac{\partial^2 \Psi_x(\mathbf{s})}{\partial s_x \partial s_y} \right)^2 + \left(\frac{\partial^2 \Psi_x(\mathbf{s})}{\partial s_y^2} \right)^2 \mathrm{d}\mathbf{s}, \quad (2)$$



Figure 1: Examples of an image warping

while interpolating the control points in accordance with (1) (see Wahba, 1990; Bookstein, 1998, for details). According to Duchon (1976) the general function that minimises (2) is

$$\Psi_x(\mathbf{s}) = a_{x,0} + a_{x,x}\mathbf{s}_x + a_{x,y}\mathbf{s}_y + \sum_{k=1}^P w_{x,k}U(\|\mathbf{s} - \boldsymbol{\rho}_k\|),$$

where \mathbf{s}_x and \mathbf{s}_y are the x - and y -coordinates of \mathbf{s} , and $U(r) = r^2 \log r^2$. The $P + 3$ unknown coefficients $\mathbf{a}_x = [a_{x,0} \ a_{x,x} \ a_{x,y}]^\top$ and \mathbf{w}_x are determined by requiring Ψ_x to interpolate the control points according to (1), and by requiring that

$$\sum_{i=1}^P w_{x,k} = \sum_{i=1}^P w_{x,k} \boldsymbol{\rho}_{x,k} = \sum_{i=1}^P w_{x,k} \boldsymbol{\rho}_{y,k} = 0.$$

The three last requirements ensure that $\Psi_x(\mathbf{s})$ has square integrable second derivatives (see Wahba, 1990).

Applying similar conditions for the Ψ_y function the coefficients, \mathbf{a} and \mathbf{w} , in the two thin plate spline function can be found by solving the following system of linear equations

$$\begin{bmatrix} \mathbf{K} & \mathcal{P} \\ \mathcal{P}^\top & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{w}_x & \mathbf{w}_y \\ \mathbf{a}_x & \mathbf{a}_y \end{bmatrix} = \mathbf{L} \begin{bmatrix} \mathbf{w}_x & \mathbf{w}_y \\ \mathbf{a}_x & \mathbf{a}_y \end{bmatrix} = \begin{bmatrix} \mathbf{p}_x & \mathbf{p}_y \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

where $\mathbf{K}_{ij} = U(\|\boldsymbol{\rho}_i - \boldsymbol{\rho}_j\|)$ and the k^{th} row of \mathcal{P} is $\mathcal{P}_k = [1 \ \boldsymbol{\rho}_{x,k} \ \boldsymbol{\rho}_{y,k}]$. Decomposing the inverse of the \mathbf{L} -matrix as

$$\mathbf{L}^{-1} = \begin{bmatrix} \mathbf{L}^{11} & \mathbf{L}^{12} \\ \mathbf{L}^{21} & \mathbf{L}^{22} \end{bmatrix},$$

where \mathbf{L}^{11} is a $P \times P$ -matrix, the minimum bending energy (2) becomes

$$\mathcal{J}(\boldsymbol{\Psi}) = \mathcal{J}(\Psi_x) + \mathcal{J}(\Psi_y) = \mathbf{p}_x^\top \mathbf{L}^{11} \mathbf{p}_x + \mathbf{p}_y^\top \mathbf{L}^{11} \mathbf{p}_y.$$

It is also worth noticing that if the coordinates of the N pixels, $\{\mathbf{s}_i\}_i^N$, and the positions of the deformed landmarks, $\boldsymbol{\rho}$, are known, then most of the transformation can be precalculated. Let \mathbf{W} be a $N \times P$ -matrix with elements $\mathbf{W}_{ij} = U(\|\mathbf{s}_i - \boldsymbol{\rho}_j\|)$ and let

$$\mathbf{B} = [\mathbf{W} \ \mathbf{1} \ \mathbf{s}_x \ \mathbf{s}_y] \begin{bmatrix} \mathbf{L}^{11} \\ \mathbf{L}^{21} \end{bmatrix}, \quad (3)$$

now the thin plate spline warping function becomes $\Psi(\mathbf{s}; \boldsymbol{\rho}, \mathbf{p}) = \mathbf{B}(\mathbf{s}, \boldsymbol{\rho})\mathbf{p}$, i.e. the warping function is linear in the undeformed landmarks, \mathbf{p} .

2.2 Stochastic model

We are now ready to construct a stochastic model. The model consists of two parts: An error function f that models the deviation in intensity between the observed field and the deformed forecast and a smoothness prior on the deformations $p(\Psi)$ that penalizes unrealistic deformations.

Let Y^F be a forecast of the observation field, Y^O , and introduce control points \mathbf{p}^F and $\mathbf{p}^O (= \boldsymbol{\rho}^F)$ in the respective images. Since we want to deform the forecast field to match the observation field a deformed version of the forecast field,

$$\mathcal{Y}^F(\mathbf{s}) = Y^F(\Psi(\mathbf{s})), \forall \mathbf{s} \in \mathcal{D},$$

is introduced. To measure the dissimilarity between the deformed forecast and the observation field an error log-likelihood is introduced as:

$$\log p(Y^O | Y^F, \mathbf{p}^F, \boldsymbol{\rho}^F, \boldsymbol{\theta}) = f(Y^O, \mathcal{Y}^F; \boldsymbol{\theta}). \quad (4)$$

Here $\boldsymbol{\theta}$ are parameters, e.g. variance, of the of the likelihood that we may want to estimate.

Selection of the error log-likelihood, $f(Y^O, \mathcal{Y}^F; \boldsymbol{\theta})$, will depend on the type of forecast that is being verified; Åberg et al. (2005) used i.i.d. Gaussian errors, $Y^O(\mathbf{s}) - \mathcal{Y}^F(\mathbf{s}) \in \mathbf{N}(0, \sigma_\epsilon^2)$. Cases with different error log-likelihoods are given in the result section.

A smoothness prior for the warping function that penalizes unrealistic deformations is also needed. Since the behavior of the warping function is determined by the position of the control points, \mathbf{p}^F and \mathbf{p}^O , this translates into a distribution on the control points $p(\mathbf{p}^F, \mathbf{p}^O)$. Additionally, the control points in the observation field are going to be assumed fixed and apriori known which reduces the prior on the warping function to $p(\mathbf{p}^F | \mathbf{p}^O)$. The posterior likelihood for the control points and unknown parameters now becomes

$$p(\mathbf{p}^F, \boldsymbol{\theta} | Y^O, Y^F, \mathbf{p}^O) \propto p(Y^O | Y^F, \mathbf{p}^F, \mathbf{p}^O, \boldsymbol{\theta}) p(\mathbf{p}^F | \mathbf{p}^O) p(\boldsymbol{\theta}), \quad (5)$$

where we have assumed that the control points \mathbf{p}^F are conditionally independent of the forecast observations Y^F and parameters $\boldsymbol{\theta}$ given \mathbf{p}^O , i.e. $p(\mathbf{p}^F | Y^F, \mathbf{p}^O) = p(\mathbf{p}^F | \mathbf{p}^O)$.

2.3 Estimation

Finding the optimal deformation is now a question of maximising the log of the likelihood in (5). In Åberg et al. (2005) this is done numerically using a robust Newton method (Branch et al., 1999; Steihaug, 1983; Byrd et al., 1988), which is possible since the derivatives and Hessian of the log-likelihood can be calculated, see Appendix A for details.

2.3.1 Only rotation, scale and translation

2.4 Approximate confidence intervalls

TODO

3 Optimisation strategies, scaled optimisation

TODO

4 Data

In order to demonstrate the usefulness and adaptability of the method, we utilize different data fields having very different characteristics. We demonstrate the warping approach using both simple geometric fields and several types of meteorological forecasts: (i) quantitative precipitation forecasts (QPF), (ii) wind vector fields, and (iii) an aerosol optical thickness (AOT) field. It is useful to examine the behavior of any spatial verification method in the face of very simple geometric cases as an illustration of how the approach works and how to interpret the results, in order to demonstrate how the information provided is meaningful. Using the simple geometric fields is also useful for showing how the technique might be extended to binary object fields in conjunction with features-based verification techniques (e.g., Davis et al., 2006a).

The geometric and precipitation examples make use of some of the test cases from the Spatial Forecast Verification Methods Inter-Comparison Project (ICP), which is an international meta-verification project for verifying gridded forecast and observed field pairs. Information on this project is found on the web at <http://www.ral.ucar.edu/projects/icp>. Test cases from the ICP used here are described in sections 4.1 and 4.2, where the geometric cases used here are described in the former and the QPF fields and their respective Stage II Analysis fields used for the “observations” are detailed in the latter. Section 4.3 describes the wind vector field, and section 4.4 the AOT field.

4.1 Geometric cases

A total of five geometric cases are available with the ICP, and although all five have been tested on the image warping method, we select two representative cases for the current paper. They are shown in figures 2

and 3. In each case, the mean squared error (MSE) is reduced to practically zero after warping because the intensities are essentially the same. In the first case, the forecast is identical to the observation except for being spatially displaced too far to the east (by 50 grid squares).

We also demonstrate the approach on actual meteorological fields of interest to the weather forecasting community. In particular, we use three wildly different fields: (i) a quantitative precipitation forecast (QPF) field, (ii) a wind vector field, and (iii) an air quality example (specifically, aerosol optical thickness (AOT)). Each pair of fields represent very different situations. The QPF example involves large areas with no precipitation (i.e., a lot of zeros), but is reasonably continuous (and positive) where precipitation exists. Wind vector fields generally have (positive and negative) values at each location, but involve two separate components (u and v), and there is interest in identifying performance for both of these components separately, as well as for two derivatives (speed (w) and direction (θ)). For the AOT field, there are typically positive and reasonably continuous values at each location, and for our example, the model does not have any noticeable spatial displacement errors, only amplitude errors. Such a case allows for investigating whether the method is able to recognize a forecast that performs well in terms of spatial coverage and placement, and not, for example, over compensate the bending to account for unnecessary affine transformations.

4.2 Precipitation

For the QPF example, we take advantage of data provided by Mike Baldwin through the Spatial Forecast Verification Methods Inter-Comparison Project (ICP) (<http://www.ral.ucar.edu/projects/icp>). Nine cases of three versions of the Weather Research and Forecast (WRF) model from the 2005 Storm Prediction Center-National Severe Storms Laboratory Spring Program (Kain et al., 2006) were chosen for the ICP. Specifically, we use the Advanced Research WRF model run of the National Center for Atmospheric Research (henceforth, WRF) from 1 June 2005. We chose this case because it represents a particularly interesting snapshot where there are spatial placement errors in different directions for different regions.

The field represents a one hour accumulation of precipitation ****Johan, it occurred to me that the field is presently in hundredths-of-an-inch. We should probably convert it to mm. I think you just multiply by 0.254, or some such****, and is interpolated onto a 501×601 grid at approximately 4-km spatial resolution over the

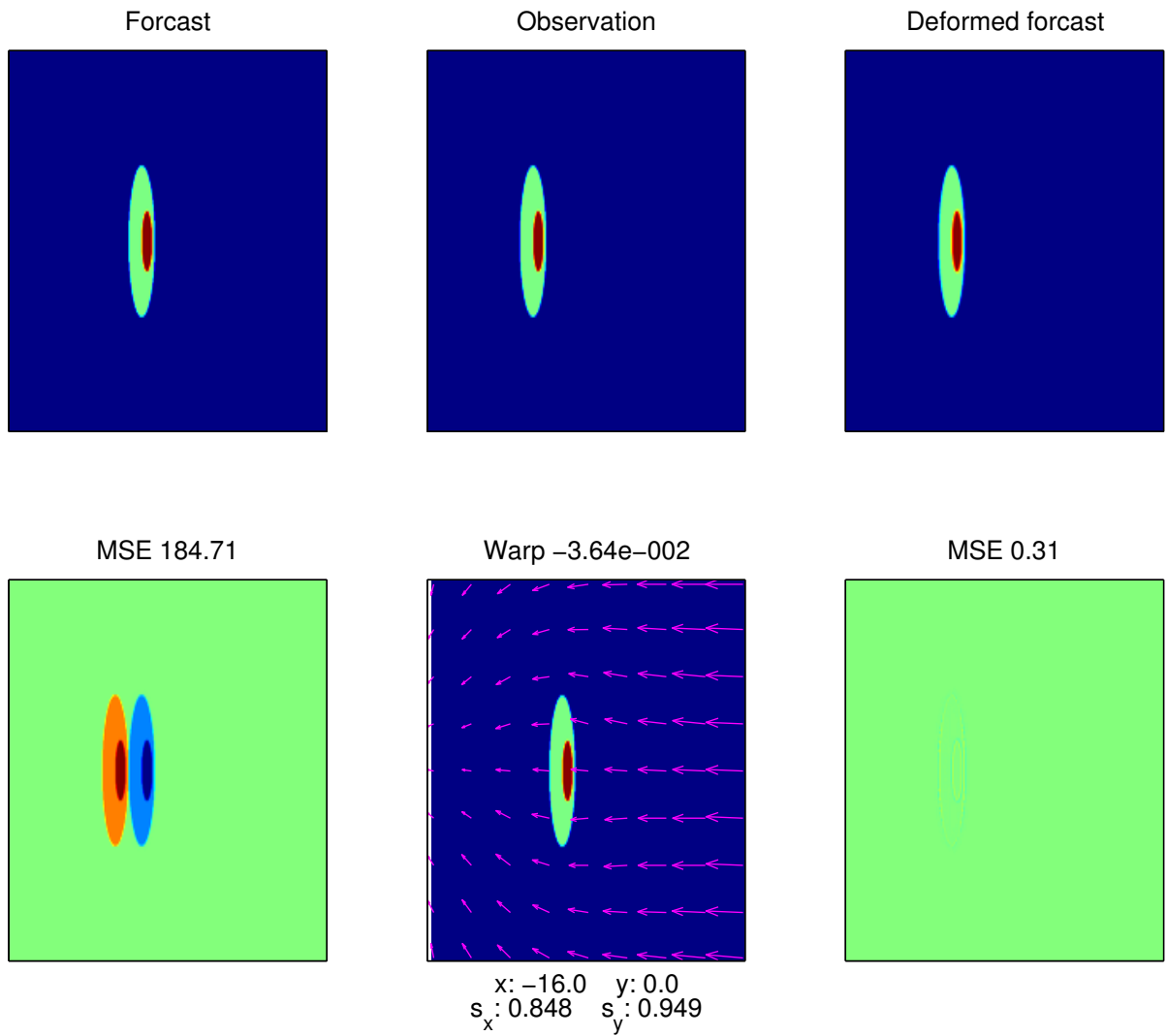


Figure 2: Geometric case 1 from the Spatial Forecast Verification Methods Inter-Comparison Project (ICP). Forecast has same pattern and intensities, but has a spatial displacement of 50 grid points too far to the east.

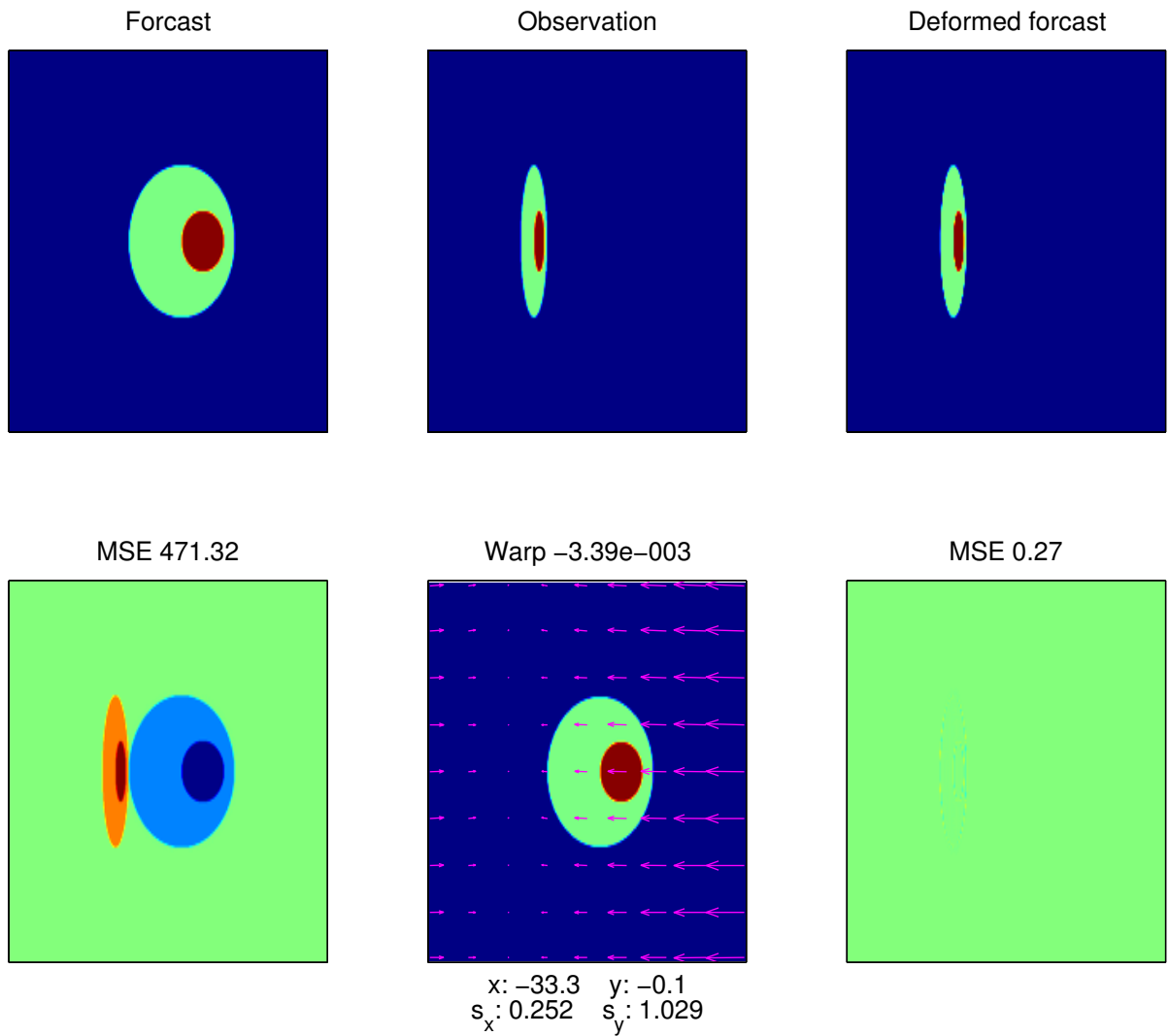


Figure 3: Geometric case 3 from the Spatial Forecast Verification Methods Inter-Comparison Project (ICP). Forecast has larger spatial coverage (over-forecast the area), and has a spatial displacement of 125 grid points too far to the east.

central United States. Verification data are taken from the National Center for Environmental Prediction’s Stage II Analyses, which are based on radar reflectivity fields (e.g., Lin and Mitchell, 2005). Both fields are on the same grids, and are shown in Fig. ?.

4.3 Wind Vector Field

4.4 Aerosol Optical Thickness (AOT)

Hodzic et al. (2006) investigated the evolution of AOT over Europe during the August 2003 heat wave using simulations from the chemistry-transport model, CHIMERE. The output is on a 256×256 grid of approximately 30-km resolution. They compare this model output with AOTs measured at 865 nm by the Polarization and Directionality of the Earth’s Reflectances (POLDER) sensor, averaged to the same grid resolution as CHIMERE. Here, we apply the image warping verification technique to one case identified by Hodzic et al. (2006) on 5 August 2005 (Fig. ?), which corresponds to their Fig. 2 (a).

5 Results

5.1 Geometric Cases

For all of the geometric cases from the ICP, the image warping approach behaves as expected. Each case is a simple transformation of the observed case; and given the geometric shapes of the objects, the image warp easily manipulates the forecast image to match the observed image nearly perfectly so that the MSE is reduced to near zero in each case. The transformation types and amounts are also quite well recovered by the warp in every case.

In the case of figure ?, a human observer might want to rotate the forecast object to match the observed object instead of the scale transformations invoked by the image warp. Indeed, it is because of the location of the high intensity inner oval that the object is re-scaled instead of rotated. Figure ? shows a similar case as figure ?, but where the error is a true rotational error; and the image warp picks up the rotation for this case. This example shows that different transformations can reproduce essentially the same deformed image. Therefore, some care should be taken in defining the penalty function in order to elicit physically meaningful information about forecast performance.

5.2 Precipitation

The image warp procedure suggests that the forecast MSE can be reduced by nearly 50% by deforming the image. Specifically, it shows that the forecast is displaced, on average, 7-km too far to the west. Additional local warping is also required, suggesting that there may be some region-specific phenomena that the forecast is missing. The specific regions can be seen in the figure. For example, the forecast has roughly the same spatial coverage of precipitation in the southeast, but the shape of the system is a bit different. Similarly, the northwest corner shows similarities between the forecast and observed fields, but the general shape is different and small cells are spatially displaced in various ways.

5.3 Wind vector

Initial results for the wind vector field are promising. **Still need to add results here.**

5.4 Aerosol Optical Thickness

6 Discussion

TODO. Johan/Finn, I am starting a brainstorm list for this section. After we finish the rest of the paper, we can begin this section with a summary and conclusions. Then, we can talk about possible future work, which we can add to the brainstorm list below, and put it into wonderful wording later on. Feel free to add as you think of things.

- Potential to incorporate other dimensions (e.g., time, vertical space, etc.) into the analysis

Acknowledgement

The authors would like to acknowledge STINT (The Swedish Foundation for International Cooperation in Research and Higher Education) Grant IG2005-2047.

A Derivatives of the log-likelihood

From (5) we have that the log-likelihood is

$$\begin{aligned} l(\mathbf{p}^F, \boldsymbol{\theta} | Y^O, Y^F, \mathbf{p}^O) &= \log p(Y^O | Y^F, \mathbf{p}^F, \mathbf{p}^O, \boldsymbol{\theta}) + \log p(\mathbf{p}^F | \mathbf{p}^O) + \log p(\boldsymbol{\theta}) \\ &= f(\mathcal{Y}^F, Y^O; \boldsymbol{\theta}) + \log p(\mathbf{p}^F | \mathbf{p}^O) + \log p(\boldsymbol{\theta}). \end{aligned}$$

Here \mathbf{p}^F and \mathbf{p}^O are the control points represented as $P \times 2$ -matrices, with the individual coordinates of each point denoted $\mathbf{p}_{x,i}^F$ or $\mathbf{p}_{y,i}^F$. The forecast and observation fields Y^F and Y^O are $m \times n \times d$ images with $N = m \cdot n \cdot d$ individual pixel values (i.e. a $m \times n$ grid of d -variate observations). Additional parameters are denoted $\boldsymbol{\theta} = \{\theta_k\}_{k=1}^K$.

Generalising the results in Åberg et al. (2005) to a general error and warp function, the gradient of the loglikelihood is

$$\begin{aligned} \nabla l(\mathbf{p}^F, \boldsymbol{\theta}) &= \\ & \begin{bmatrix} \mathbf{J}^\top \nabla_{\mathcal{Y}} f(\mathcal{Y}^F, Y^O; \boldsymbol{\theta}) + \nabla_p \log p(\mathbf{p}^F | \mathbf{p}^O) \\ \nabla_{\boldsymbol{\theta}} f(\mathcal{Y}^F, Y^O; \boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{\theta}) \end{bmatrix}, \end{aligned} \quad (6)$$

where $\nabla_{\mathcal{Y}} f(\mathcal{Y}^F, Y^O)$ is the gradient of the error log-likelihood w.r.t. the pixel values

$$\nabla_{\mathcal{Y}} f(\mathcal{Y}^F, Y^O; \boldsymbol{\theta}) = \left[\frac{\partial f}{\partial \mathcal{Y}_1^F} \quad \dots \quad \frac{\partial f}{\partial \mathcal{Y}_N^F} \right]^\top,$$

\mathbf{J} is a Jacobian, whose rows are the transposed gradients $\nabla_p \mathcal{Y}_i^F(\mathbf{p}^F)$,

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \mathcal{Y}_1^F}{\partial \mathbf{p}_{x,1}^F} & \dots & \frac{\partial \mathcal{Y}_1^F}{\partial \mathbf{p}_{x,P}^F} & \frac{\partial \mathcal{Y}_1^F}{\partial \mathbf{p}_{y,1}^F} & \dots & \frac{\partial \mathcal{Y}_1^F}{\partial \mathbf{p}_{y,P}^F} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial \mathcal{Y}_N^F}{\partial \mathbf{p}_{x,1}^F} & \dots & \frac{\partial \mathcal{Y}_N^F}{\partial \mathbf{p}_{x,P}^F} & \frac{\partial \mathcal{Y}_N^F}{\partial \mathbf{p}_{y,1}^F} & \dots & \frac{\partial \mathcal{Y}_N^F}{\partial \mathbf{p}_{y,P}^F} \end{bmatrix}, \quad (7)$$

and $\nabla_p \log p(\mathbf{p}^F | \mathbf{p}^O, \boldsymbol{\theta})$ is the gradient of the smoothness prior w.r.t. the control points,

$$\nabla_p \log p(\mathbf{p}^F | \mathbf{p}^O, \boldsymbol{\theta}) = \left[\frac{\partial \log p}{\partial \mathbf{p}_{x,1}^F} \quad \dots \quad \frac{\partial \log p}{\partial \mathbf{p}_{x,P}^F} \quad \frac{\partial \log p}{\partial \mathbf{p}_{y,1}^F} \quad \dots \quad \frac{\partial \log p}{\partial \mathbf{p}_{y,P}^F} \right]^\top.$$

The elements in the Jacobian (7) are

$$\begin{aligned} \frac{\partial \mathcal{Y}_i^F}{\partial \mathbf{p}_{x,j}^F} &= \frac{\partial Y^F(\Psi_x(\mathbf{s}_i; \mathbf{p}_x^F), \Psi_y(\mathbf{s}_i; \mathbf{p}_y^F))}{\partial \mathbf{p}_{x,j}^F} \\ &= \frac{\partial Y^F(\Psi_x(\mathbf{s}_i; \mathbf{p}_x^F), \Psi_y(\mathbf{s}_i; \mathbf{p}_y^F))}{\partial \Psi_x(\mathbf{s}_i; \mathbf{p}_x^F)} \frac{\partial \Psi_x(\mathbf{s}_i; \mathbf{p}_x^F)}{\partial \mathbf{p}_{x,j}^F} \\ &= \widetilde{\partial_x Y_i^F} \frac{\partial \Psi_x(\mathbf{s}_i; \mathbf{p}_x^F)}{\partial \mathbf{p}_{x,j}^F} = \widetilde{\partial_x Y_i^F} \mathbf{B}_{ij} \\ \frac{\partial \mathcal{Y}_i^F}{\partial \mathbf{p}_{x,j}^F} &= \text{as above} = \widetilde{\partial_y Y_i^F} \mathbf{B}_{ij} \end{aligned} \quad (8)$$

where $\widetilde{\partial_x Y_i^F}$ denotes warped partial derivative along the x -axis of the forecast Y^F (analogously for $\widetilde{\partial_y Y_i^F}$, see A.0.1 for details). The appearance of \mathbf{B}_{ij} comes from the definition of a TPS as $\Psi(\mathbf{s}; \boldsymbol{\rho}, \mathbf{p}) = \mathbf{B}(\mathbf{s}, \boldsymbol{\rho})\mathbf{p}$, with \mathbf{B} given by (3).

A.0.1 Warped partial derivatives

The warped partial derivatives $\widetilde{\partial_x Y_i^F}$ in (8) above will depend on the type of interpolation used in the warping. To find the derivatives we must first study the interpolation used to determine the value of a pixel in the deformed field, $\mathcal{Y}^F(\mathbf{s})$. The pixel value is given by interpolation of the pixels in the original field Y^F closest to $\Psi(\mathbf{s})$.

First we need some notation, let Y_{00} denote the pixel given by $\lfloor \Psi(\mathbf{s}) \rfloor$ (where $\lfloor t \rfloor$ gives the largest integer such that $\lfloor t \rfloor \leq t$), and let x and y be the fractional part of $\Psi_x(\mathbf{s})$ and $\Psi_y(\mathbf{s})$ respectively (see Figure for details), i.e.

$$x = \Psi_x(\mathbf{s}) - \lfloor \Psi_x(\mathbf{s}) \rfloor.$$

Now the value of a pixel in the deformed field is given by

$$\mathcal{Y}^F(\mathbf{s}) = \sum_l \sum_k b_l(x) b_k(y) Y_{lk},$$

where the $b_l(x)$ and $b_k(y)$ as well as the range of the two sums depend on the type of interpolation used. For bilinear interpolation the sums over l and k run from 0 to 1 with $b_0(x) = 1 - x$ and $b_1(x) = x$. For bicubic interpolation the sums run from -1 to 2 and the interpolation polynomials are

$$\begin{aligned} b_{-1}(x) &= (2x^2 - x^3 - x)/2 & b_0(x) &= (3x^3 - 5x^2 + 2)/2 \\ b_1(x) &= (4x^2 - 3x^3 + x)/2 & b_2(x) &= (x - 1)x^2/2. \end{aligned}$$

The corresponding derivatives become

$$\widetilde{\partial_x Y^F}(\mathbf{s}) = \frac{\partial \sum_l \sum_k b_l(x) b_k(y) Y_{lk}}{\partial x} = \sum_l \sum_k \frac{\partial b_l(x)}{\partial x} b_k(y) Y_{lk},$$

and similar for $\widetilde{\partial_y Y^F}(\mathbf{s})$. The higher derivatives needed when calculating the Hessian in Section A.1 can be calculated in a similar way.

A.1 Hessian

The Hessian of the loglikelihood can be divided in to three submatrices,

$$\Delta l(\mathbf{p}^F, \boldsymbol{\theta}) = \begin{bmatrix} \mathbf{H}_{pp} & \mathbf{H}_{p\theta} \\ \mathbf{H}_{p\theta}^\top & \mathbf{H}_{\theta\theta} \end{bmatrix}. \quad (9)$$

Here $\mathbf{H}_{\theta\theta}$ is the Hessian of the likelihood with respect to the parameters, calculation of which is trivial. The of diagonal matrix, $\mathbf{H}_{p\theta}$, is the cross derivative between the points and parameters. Differentiating the first part of the gradient (6), we obtain

$$\frac{\partial}{\partial \theta_k} \mathbf{J}^\top \nabla_{\mathcal{Y}} f(\mathcal{Y}^F, Y^O; \boldsymbol{\theta}) + \nabla_p \log p(\mathbf{p}^F | \mathbf{p}^O) = \mathbf{J}^\top \frac{\partial}{\partial \theta_k} \nabla_{\mathcal{Y}} f(\mathcal{Y}^F, Y^O; \boldsymbol{\theta}).$$

Finally the last part of the Hessian, H_{pp} is

$$\mathbf{H}_{pp} = \mathbf{J}^\top \Delta_{\mathcal{Y}} f(\mathcal{Y}^F, Y^O) \mathbf{J} + \sum_{l=1}^N \frac{\partial f(\mathcal{Y}^F, Y^O)}{\partial \mathcal{Y}_l^F} \Delta \mathcal{Y}_l^F + \Delta_p \log p(\mathbf{p}^F | \mathbf{p}^O). \quad (10)$$

Here $\Delta_{\mathcal{Y}} f(\mathcal{Y}^F, Y^O)$ is the Hessian of the error log-likelihood w.r.t. the pixel values, $\Delta \log p(\mathbf{p}^F | \mathbf{p}^O)$ is the Hessian of the smoothness prior, and $\Delta \mathcal{Y}_l^F$ are the Hessian of $\mathcal{Y}_l^F(\mathbf{p}^F)$ w.r.t. the control points,

$$\Delta \mathcal{Y}_l^F = \begin{bmatrix} \frac{\partial^2 \mathcal{Y}_l^F}{\partial (\mathbf{p}_{1,x}^F)^2} & \cdots & \frac{\partial^2 \mathcal{Y}_l^F}{\partial \mathbf{p}_{1,x}^F \partial \mathbf{p}_{p,y}^F} \\ \vdots & & \vdots \\ \frac{\partial^2 \mathcal{Y}_l^F}{\partial \mathbf{p}_{p,y}^F \partial \mathbf{p}_{1,x}^F} & \cdots & \frac{\partial^2 \mathcal{Y}_l^F}{\partial (\mathbf{p}_{p,y}^F)^2} \end{bmatrix}.$$

The elements of $\Delta \mathcal{Y}_l^F$ are

$$\begin{aligned} \frac{\partial^2 \mathcal{Y}_l^F}{\partial \mathbf{p}_{i,x}^F \partial \mathbf{p}_{j,x}^F} &= \frac{\partial^2 Y^F(\Psi_x(\mathbf{s}_l; \mathbf{p}_x^F), \Psi_x(\mathbf{s}_l; \mathbf{p}_x^F))}{\partial \mathbf{p}_{i,x}^F \partial \mathbf{p}_{j,x}^F} \\ &= \frac{\partial^2 Y^F(\Psi_x(\mathbf{s}_l; \mathbf{p}_x^F), \Psi_x(\mathbf{s}_l; \mathbf{p}_x^F))}{\partial \Psi_x(\mathbf{s}_l; \mathbf{p}_x^F) \partial \Psi_x(\mathbf{s}_l; \mathbf{p}_x^F)} \frac{\partial \Psi_x(\mathbf{s}_l; \mathbf{p}_x^F)}{\partial \mathbf{p}_{i,x}^F} \frac{\partial \Psi_x(\mathbf{s}_l; \mathbf{p}_x^F)}{\partial \mathbf{p}_{j,x}^F} \\ &= \widetilde{\partial_{xx} Y_l^F} \frac{\partial \Psi_x(\mathbf{s}_l; \mathbf{p}_x^F)}{\partial \mathbf{p}_{i,x}^F} \frac{\partial \Psi_x(\mathbf{s}_l; \mathbf{p}_x^F)}{\partial \mathbf{p}_{j,x}^F} \\ &= \widetilde{\partial_{xx} Y_l^F} \mathbf{B}_{li} \mathbf{B}_{lj} \\ \frac{\partial^2 \mathcal{Y}_l^F}{\partial \mathbf{p}_{i,x}^F \partial \mathbf{p}_{j,y}^F} &= \cdots = \widetilde{\partial_{xy} Y_l^F} \mathbf{B}_{li} \mathbf{B}_{lj} \\ \frac{\partial^2 \mathcal{Y}_l^F}{\partial \mathbf{p}_{i,y}^F \partial \mathbf{p}_{j,y}^F} &= \cdots = \widetilde{\partial_{yy} Y_l^F} \mathbf{B}_{li} \mathbf{B}_{lj} \end{aligned}$$

or where $\widetilde{\partial_{xx}Y_i^F}$ denotes warped, finite second differences of the forecast Y^F . Note that since the TPS warp functions is $\Psi(\mathbf{s}; \boldsymbol{\rho}, \mathbf{p}) = \mathbf{B}(\mathbf{s}, \boldsymbol{\rho})\mathbf{p}$ all second derivatives of Ψ with respect to \mathbf{p} are zero.

A.2 Example

Assume $Y_i^O - \mathcal{Y}_i^F \in \mathbf{N}(0, \sigma_\epsilon^2)$,

$$\begin{bmatrix} \mathbf{p}_x^F \\ \mathbf{p}_y^F \end{bmatrix} \in \mathbf{N} \left(\begin{bmatrix} \mathbf{p}_x^O \\ \mathbf{p}_y^O \end{bmatrix}, \frac{1}{\kappa} \begin{bmatrix} \mathbf{Q}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^{-1} \end{bmatrix} \right),$$

and $\boldsymbol{\theta} = \sigma_\epsilon^2$.

Now we have that

$$f(\mathcal{Y}^F, Y^O; \boldsymbol{\theta}) = -\frac{N}{2} \log(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^N (\mathcal{Y}_i^F - Y_i^O)^2,$$

$$\nabla_{\mathcal{Y}} f(\mathcal{Y}^F, Y^O; \boldsymbol{\theta}) = \left[-\frac{1}{\sigma_\epsilon^2} (\mathcal{Y}_1^F - Y_1^O) \quad \dots \quad -\frac{1}{\sigma_\epsilon^2} (\mathcal{Y}_N^F - Y_N^O) \right]^\top,$$

and

$$\nabla_{\boldsymbol{\theta}} f(\mathcal{Y}^F, Y^O; \boldsymbol{\theta}) = -\frac{N}{2\sigma_\epsilon^2} + \frac{1}{2\sigma_\epsilon^4} \sum_{i=1}^N (\mathcal{Y}_i^F - Y_i^O)^2.$$

For the Hessian we obtain

$$\Delta_{\mathcal{Y}} f(\mathcal{Y}^F, Y^O; \boldsymbol{\theta}) = \begin{bmatrix} -\frac{1}{\sigma_\epsilon^2} & 0 \\ 0 & \ddots \end{bmatrix}^\top,$$

$$\Delta_{\boldsymbol{\theta}} f(\mathcal{Y}^F, Y^O; \boldsymbol{\theta}) = \frac{N}{2\sigma_\epsilon^4} - \frac{1}{\sigma_\epsilon^6} \sum_{i=1}^N (\mathcal{Y}_i^F - Y_i^O)^2,$$

and

$$\frac{\partial}{\partial \boldsymbol{\theta}} \Delta_{\mathcal{Y}} f(\mathcal{Y}^F, Y^O; \boldsymbol{\theta}) = \left[\frac{1}{\sigma_\epsilon^4} (\mathcal{Y}_1^F - Y_1^O) \quad \dots \quad \frac{1}{\sigma_\epsilon^4} (\mathcal{Y}_N^F - Y_N^O) \right]^\top.$$

References

- Åberg, S., Lindgren, F., Malmberg, A., Holst, J., Holst, U., 2005. An image warping approach to spatio-temporal modelling. *Environmetrics* 16 (8), 833–848.
- Alexander, G., Weinman, J., Karyampudi, V., Olson, W., Lee, A., 1999. The effect of assimilating rain rates derived from satellites and lightning on forecasts on the 1993 superstorm. *Mon. Wea. Rev.* 127, 1433–1457.

- Bookstein, F., 1998. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11 (6), 567–585.
- Branch, M., Coleman, T., Li, Y., 1999. A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems. *SIAM Journal on Scientific Computing* 21 (1), 1–23.
- Briggs, W., Levine, R., 1997. Wavelets and field forecast verification. *Mon. Wea. Rev.* 125, 1329–1341.
- Byrd, R., Schnabel, R., G.A., S., 1988. Approximate solution of the trust region problem by minimization over two-dimensional subspaces. *Mathematical Programming* 40, 247–263.
- Casati, B., Ross, G., Stephenson, D., 2004. A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorol. Appl.* 11, 141–154.
- Damian, D., Sampson, P. D., Guttorp, P., 2001. Bayesian estimation of semi-parametric non-stationary spatial covariance structures. *Environmetrics* 12 (2), 161–178.
- Davis, C., Brown, B., Bullock, R., 2006a. Object-based verification of precipitation forecasts, part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.* 134, 1772–1784.
- Davis, C., Brown, B., Bullock, R., 2006b. Object-based verification of precipitation forecasts, part ii: Application to convective rain systems. *Mon. Wea. Rev.* 134, 1785–1795.
- Dickinson, S., Brown, R., 1996. A study of near-surface winds in marine cyclones using multiple satellite sensors. *J. Appl. Meteorol.* 35, 769–781.
- Dryden, I., Mardia, K., 1998. *Statistical Shape Analysis*. John Wiley.
- Duchon, J., 1976. Splines minimizing rotation invariant seminorms in sobolev spaces. In: Schempp, W., Zeller, K. (Eds.), *Constructive Theory of Functions of Several Variables*. Springer-Verlag, pp. 85–100.
- Ebert, E., McBride, J., 2000. Verification of precipitation in weather systems: determination of systematic errors. *J. Hydrology* 239, 179–202.

- Ebert, E., Wilson, L., Brown, B., Nurmi, P., Brooks, H., Bally, J., Jaeneke, M., 2004. Verification of nowcasts from the wrwp sydney 2000 forecast demonstration project. *Wea. Forecasting* 19, 73–96.
- Glasbey, C., Mardia, K., 1998. A review of image warping methods. *Journal of Applied Statistics* 25 (2), 155–171.
- Glasbey, C., Mardia, K., 2001. A penalized likelihood approach to image warping. *Journal of the Royal Statistical Society. Series B (Methodology)* 63 (3), 465–514.
- Goshtasby, A., 1987. Piecewise cubic mapping functions for image registration. *Pattern Recognition* 20, 523–533.
- Grams, J., Gallus Jr., W., Koch, S., Wharton, L., Loughe, A., Ebert, E., 2006. The use of a modified ebert-mcbride technique to evaluate mesoscale model qpf as a function of convective system morphology during ihop 2002. *Wea. Forecasting* 21, 288–306.
- Guttorp, P., Meiring, W., Sampson, P. D., 1994. A space-time analysis of ground-level ozone data. *Environmetrics* 5 (3), 241–254.
- Harris, D., Foufoula-Georgiou, E., Droegemeier, K., Levit, J., 2001. Multiscale statistical properties of a high-resolution precipitation forecast. *J. Hydrometeorology* 2, 406–418.
- Hodzic, A., Vautard, R., Chepfer, H., Goloub, P., Menut, L., Chazette, P., Deuzé, J., Apituley, A., Couvert, P., 2006. Evolution of aerosol optical thickness over europe during the august 2003 heat wave as seen from chimere model simulations and polder data. *Atmos. Chem. Phys.* 6, 1853–1864.
- Jolliffe, I., Stephenson, D., 2003. *Forecast Verification. A Practitioner’s Guide in Atmospheric Science.* Wiley and Sons Ltd, 254 pp.
- Kain, J., Weiss, S., Levit, J., Baldwin, M., Bright, D., 2006. Examination of convection-allowing configurations of the wrf model for the prediction of severe convective weather: The spc/nssl spring program 2004. *Wea. Forecasting* 21, 167–181.
- Keil, C., Craig, G., 2007. A displacement-based error measure applied in a regional ensemble forecasting system. *Mon. Wea. Rev.* 135, 3248–3259.
- Lee, S., Wolberg, G., Shin, S., 1997. Scattered data interpolation with multilevel b-splines. *IEEE Transactions on Visualization and Computer Graphics* 3 (3), 228–244.

- Lin, Y., Mitchell, K., 2005. The ncep stage ii/iv hourly precipitation analyses: Development and applications. available at <http://www.emc.ncep.noaa.gov/mmb/ylin/pcpanl/refs/stage2-4.19hydro.pdf>. 19th Conf on Hydrology. Amer. Meteorol. Soc. 1.2.
- Marzban, C., Sandgathe, S., 2006. Cluster analysis for verification of precipitation fields. *Wea. Forecasting* 21 (5), 824–838.
- Meiring, W., Sampson, P. D., Guttorp, P., 1998. Space-time estimation of grid-cell hourly ozone levels for assessment of a deterministic model. *Environmental and Ecological Statistics* 5 (3), 197–222.
- Micheas, A., Fox, N., Lack, S., Wikle, C., 2006. Cell identification and verification of QPF ensembles using shape analysis techniques. Submitted to *J. of Hydrology*.
- Nachamkin, J., 2004. Mesoscale verification using meteorological composites. *Mon. Wea. Rev.* 132, 941–955.
- Nehrkorn, T., Hoffman, R., Grassotti, C., Louis, J.-F., 2003. Feature calibration and alignment to represent model forecast errors: Empirical regularization. *Q.J.R. Meteorol. Soc.* 129, 195–218.
- Sampson, P. D., Guttorp, P., 1992. Nonparametric estimation of non-stationary spatial covariance structure. *Journal of the American Statistical Association* 87 (417), 108–119.
- Sorzano, C., Thevenaz, P., Unser, M., 2005. Elastic registration of biological images using vector-spline regularization. *IEEE Transactions on Biomedical Engineering* 52 (4), 625–663.
- Steihaug, T., 1983. The conjugate gradient method and trust regions in large scale optimization. *SIAM Journal on Numerical Analysis* 20, 626–637.
- Wahba, G., 1990. *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics.
- Wikle, C. K., 2002. A kernel-based spectral model for non-gaussian spatio-temporal processes. *Statistical Modelling* 2, 299–314.
- Wikle, C. K., Cressie, N., 1999. A dimension-reduced approach to space-time kalman filtering. *Biometrika* 86 (4), 815–829.
- Wilks, D., 2006. *Statistical Methods in the Atmospheric Sciences*. An Introduction, 2nd Edition. Academic Press, San Diego, 627 pp.

- Xu, K., Wike, C., Fox, N., 2005. A kernel-based spatio-temporal dynamical model for nowcasting weather radar reflectivities. *Journal of the American Statistical Association* 100 (472), 1133–1144.
- Zitova, B., Flusser, J., 2003. Image registration methods: a survey. *Image and Vision Computing* 21 (11), 977–1000.