WRF Developmental Testbed Center (DTC) – Visitor Program

Summary on visit achievements

by Barbara Casati

Period of the visit: 31st March - 25th April 2008.
Hosting laboratory: RAL division at NCAR, Boulder, CO.

Scale-decomposition verification tools for the WRF model: The MET intensity-scale verification tool

Scale-decomposition verification techniques belong to a new generation of diagnostic verification tools specifically designed for forecasts defined over spatial domains (Casati et al, 2008). Spatial verification approaches account for the presence of features and coherent spatial structure characterising the meteorological fields, and aim to to provide feedback on the physical nature of the forecast error (e.g. feature displacements or scale structure representation). Such targeted diagnostic verification can help the development and improvement of specific processes and physical parametrizations of NWP models.

Scale-decomposition verification techniques (e.g. Casati et al, 2004; Harris et al, 2001) assess the forecast quality on different spatial scales. Weather phenomena on different scales (e.g. fronts or convective showers) are often driven by different physical processes. Verification on different spatial scales can provide therefore a deeper insight into the model performance at simulating such processes. Moreover, scale-decomposition techniques enable to identify of the no-skill to skill transition scale, and can help diagnosing the highest resolution for which the forecast exhibits skill. Finally, scale-decomposition approaches can be used to assess the capability of forecasts in reproducing the observed scale structure: this can be used, as an example, to demonstrate the added value of high reslution models, which are expected to reproduce more relaistic featrues than their lower resolution counterparts.

The primary achievement of my project within the WRF Developmental Testbed Center (DTC) Visitor Program, has been the implementation of the Intensity-Scale verification technique introduced by Casati et al (2004) into the WRF model Meteorological Evaluation Toolkit (MET).

The intensity-scale verification technique is a scale-decomposition approach which measures the skill of spatial precipitation forecasts as a function of the intensity of the precipitation and scale of the error. The technique is capable of of identifying specific intensity-scale errors associated to individual cases, but the statistics can be also aggregated on multiple cases (e.g. monthly or seasonally) to more robustly assess the overall model performance and intensity-scale depensency of the forecast skill.

The intensity-scale tool within MET had been therefore coded to enable both the verification of a single case study (for the forecast-observation couple both intensity-scale statistics and forecast and observation wavelet components are produced), and for multiple model runs (aggregated statistics are evaluated, along with bootstrap 90% confidence intervals). Tools for the graphical display of the output ststistics have also been coded.

During my visit, the diagnostic capabilities of the intensity-scale technique have been extended to assess the bias on different scales and to enable the comparison of forecast and observed scale structures. For each intensity and scale, forecast and observation squared energy are calculated and compared to assess the bias. The precentage to which each scale contributes to the total energy have then been evaluated to enable the comparison of forecast and observation scale structures. More theoretical discussion on this new diagnostics of the intensity-scale technique, along with some examples and graphs, can be found in the Annex.

Some minor technical issues regading the implementation of the intensity-scale approach, which were not addressed in the original article (Casati et al, 2004), have been addressed in the MET intensity-scale verification tool. As an example, the technique in MET is implemented without prior performing a recalibration to remove the bias. Moeover, the requirement of a square spatial domain with dimension equal to $2^n x 2^n$ has been relaxed. Aggregation of the intensity-scale statistics for multiple cases, and evaluation of the statistics confidence intervals, has been included. Technical details for such improvements can be found also in the Annex.

Documentation for the MET intensity-scale tool have been carefully prepared and will be incorporated into the MET User's guide with the next MET version release (the Annex). A complete theoretical exposition on the intensity-scale technique and its new development, and further technical details on the MET intensity-scale verification tool, can be found in such documentation, along with few images and examples from a real case study.

The intensity-scale technique is based on a categorical approach, which is robust and resistent, and therefore suitable for non-nornally distributed variables characterised by the presence of many large values, such as precipitation. Moreover, the intensity-scale technique makes use of a 2D Haar wavelet filter to obtain the spatial scale components: wavelets, because of their local properties, are suitable for representing highly discontinuous fields characterised by the presence of features and few non-zero values (such as precipitation fields). The intensity-scale technique was specifically designed to cope with the difficult characteristics of precipitation fields, and for the verification of spatial precipitation forecasts. However the technique can be used also for the verification of different variables, such as cloud fraction, or different kinds of forecast, such as probabilistic forecasts (see Casati and Wilson, 2007).

Note that the intensity-scale technique bridges traditional categorical verification with the new scale-decomposition approaches. In such fashion, the traditional categorical verification gains from the new insights provided by of the spatial approach, so that the intensity-scale skill score can be though as an extension to the definition of the Heidke Skill Score for different scales. On the other hand, the scale-decomposition approach is stenghthen by the robustness and statistical meaning of the well established categorical scores.

Note also that the wavelet filetr routines included in MET for the implementation of the intensity-scale tool have great potential: in fact they provide the set-up for a frame work in which codes for other scale-decomposition techniques, such as Briggs and Levine (1997) or Harris et al. (2001), can be added, to create a MET scale-decoposition verification package for the WRF model.

Further achievements accomplished during the visit.
1. I participated to a meeting of the Verification Methods Inter-Comparison Project, led by Eric Gilleland. During the meeting some of the recently developed spatial verification methods are tested on the same case studies and their capabilities are compared. The Intensity-Scale technique revealed to be sensitive to displacment errors and, with the evaluation of the energy on different scales, could provide feedback on the bias on different scales and scale-structure errors. A paper illustrating these results is in preparation and will be submitted for a special issue on forecast verification in Weather and Forecasting.
2. I gave a presentation on a new wavelet-based verification approach to account for the variation in scale representativeness of observation networks across the domain. This technique is the natural extension of intensity-scale to account for sparsness of the observations. Discussions and feedback from NCAR scientists helped improving the approach, and hopefully will lead to future collaborations.
3. I participated to the WRF DTC verification workshop and gave a talk on verification of extreme events. My current project at Ouranos uses Extreme Value Theory (EVT) to analyse extreme events in climate. The talk outlied how the weather and climate communities use complementary approaches to tackle the analysis of extremes: both communities could greatly gain from a sinenergy in research efforts. Feedback and the many discussions with Rick Katz and Eric Gilleland helped greatly in the understanding of the EVT, and how it could be used both in climate and weather verification. I am currently collaborating with them on my climate extreme project.

Acknowleedgements: I wish to thank Barb Brown and the DTC Visitor Program for granting me this opportunity, which lead to a deeper undertanding and improvements in the intensity-scale verification approach, and hopefully to the spread of its use to many users. I wish also to thank John Halley-Gotway for dedicating many hours during my visit in discussing, understanding and coding the intensity-scale verification technique into MET: his great technical expertise and constructive feedback have been fundamental for the success of the project.

Annex – The Intensity-Scale Tool

1. Introduction

This annex provides a description of the MET Intensity-Scale Tool, which enables to apply the Intensity-Scale verification technique described by Casati et al. (2004).

The Intensity-Scale technique is one of the recently developed verification approaches which focus on verification of forecasts defined over spatial domains. Spatial verification approaches, as opposed to point-by-point verification approaches, aim to account for the presence of features and for the coherent spatial structure characterizing meteorological fields. Since these approaches account for the intrinsic spatial correlation existing between nearby grid-points, they do not suffer of point-by-point comparison related verification issues, such as double penalties. Spatial verification approaches aim to account for the observation and forecast time-space uncertainties, and aim to provide feedback on the forecast error in physical terms.

The Intensity-Scale verification technique, as most of the spatial verification approaches, compares a forecast field to an observation field. To apply the Intensity-Scale verification approach, observations need to be defined over the same spatial domain of the forecast to be verified.

Within the spatial verification approaches, the Intensity-Scale technique belongs to the scale-decomposition (or scale-separation) verification approaches. The scale-decomposition approaches enable to perform the verification on different spatial scales. Weather phenomena on different scales (e.g. frontal systems versus convective showers) are often driven by different physical processes. Verification on different spatial scales can therefore provide deeper insights into model performance at simulating these different processes.

The spatial scale components are obtained usually by applying a single band spatial filter to the forecast and observation fields (e.g. Fourier, Wavelets, ...). The scale-decomposition approaches measure error, bias and skill of the forecast on each different scale component. The scale-decomposition approaches provide therefore feedback on the scale dependency of the error and skill, on the no-skill to skill transition scale, and on the capability of the forecast of reproducing the observed scale structure.

The Intensity-Scale technique evaluates the forecast skill as a function of the precipitation intensity and of the spatial scale of the error. The scale components are obtained by applying a 2D Haar wavelet filter. Note that wavelets, because of their locality, are suitable for representing discontinuous fields characterized by few sparse non-zero features, such as precipitation. Moreover, the technique is based on a categorical approach, which is a robust and resistant approach, suitable for non-normally distributed variables, such as precipitation. The intensity-scale technique was specifically designed to cope with the difficult characteristics of precipitation fields, and for the verification of spatial precipitation forecasts. However, the intensity-scale technique can be applied to verify also other variables, such as cloud fraction.

2. Scientific and Statistical aspects

2.1 The method

Casati et al (2004) apply the Intensity-Scale verification to preprocessed and re-calibrated (unbiased) data. The preprocessing aimed mainly to normalize the data, and define so categorical threshold so that each categorical bin had a similar sample size. The recalibration was performed to eliminate the forecast bias. Preprocessing and recalibration are not strictly necessary for the application of the Intensity-Scale technique. The MET Intensity-Scale Tool do not perform either, and apply the Intensity-Scale approach to biased forecasts, for categorical thresholds defined by the user.

The Intensity Scale approach can be summarized in the following 5 steps:

1. For each threshold, the forecast and observation fields are transformed into binary fields: where the grid-point precipitation value exceeds the threshold it is assigned 1, where the threshold is not exceeded it is assigned 0. Figure 1 illustrates and example of a forecast and observation fields, and their corresponding binary fields for a threshold of 1mm/h. This case shows an intense storm of the scale of 160 km displaced almost its entire length. The displacement error is clearly visible from the binary field difference and the contingency table image obtained for the same threshold (Table 1).
2. The binary forecast and observation fields obtained from the thresholding are then decomposed into the sum of components on different scales, by using a 2D Haar wavelet filter (Fig 2). Note that the scale components are fields, and their sum adds up to the original binary field. For a forecast defined over square domain of 2^n x 2^n grid-points, the scale components are n+1: n mother wavelet components + the largest father wavelet (or scale-function) component. The n mother wavelet components have resolution equal to 1, 2, 4, ... 2^{n-1} grid-points. The largest father wavelet component is a constant field over the 2^n x 2^n grid-point domain with value equal to the field mean.

   Note that the wavelet transform is a linear operator: this implies that the difference of the spatial scale components of the binary forecast and observation fields (Fig 2) are equal to the spatial scale components of the difference of the binary forecast and observation fields (Fig. 3), and these scale components also add up to the original binary field difference (Fig. 1). The intensity-scale technique considers thus the spatial scale of the error. For the case illustrated (Fig 1 and Fig 3) note the large error associated at the scale of 160 km, due to the storm of 160 km displaced almost of its entire length.

Note also that the means of the binary forecast and observation fields (I.e. their largest father wavelet components) are equal to the proportion of forecast and observed events above the threshold, (a+b)/n and (a+c)/n, evaluated from the contingency table counts (Table 1) obtained from the original forecast and observation fields by thresholding with the same threshold used to obtained the binary forecast and observation fields. This relation is intuitive when observing forecast and observation binary fields and their corresponding contingency table image (Fig 1). The comparison of the largest father wavelet component of binary forecast and observation fields provides therefore feedback on the whole field bias.

3. For each threshold (t) and for each scale component (j) of the binary forecast and observation, the Mean Squared Error (MSE) is then evaluated (Figure 4). The error is usually large for small thresholds, and decreases as the threshold increases. This behavior is partially artificial, and occurs because the smaller the threshold the more events will exceed it, and therefore the larger would be the error, since the error tends to be proportional to the amount of events in the binary fields. The artificial effect can be diminished by normalization: because of the wavelet orthogonal properties, the sum of the MSE of the scale components is equal to the MSE of the original binary fields: $\mathrm{MSE}(t) = \Sigma_j\ \mathrm{MSE}(t,j)$. This enables to calculate in which percentage the MSE on each scale contributes to the total MSE, given a threshold: $\mathrm{MSE\%}(t,j) = \mathrm{MSE}(t,j)/\mathrm{MSE}(t)$. The MSE% does not exhibit the threshold dependency, and usually shows small errors on large scales and large errors on small scales, with the largest error associated to the smallest scale and highest threshold. For the NIMROD case illustrated note the large error at 160 km and between the thresholds of ½ and 4 mm/h, due to the 160 km storm displaced almost of its entire length.

   Note that the MSE of the original binary fields is equal to the proportion of the counts of misses (c/n) and false alarms (b/n) for the contingency table (Table 1) obtained from the original forecast and observation fields by thresholding with the same threshold used to obtained the binary forecast and observation fields: $\mathrm{MSE}(t) = (b+c)/n$. This relation is intuitive when comparing the forecast and observation binary field difference and their corresponding contingency table image (Fig 1).

4. The MSE for the random binary forecast and observation fields is estimated by $\mathrm{MSE}(t)_{random}=\mathrm{FBI}*\mathrm{Br}*(1-\mathrm{Br}) + \mathrm{Br}*(1-\mathrm{FBI}*\mathrm{Br})$, where $\mathrm{FBI}=(a+b)/(a+c)$ is the frequency bias index and $\mathrm{Br}=(a+c)/n$ is the sample climatology from the contingency table (Table 1) obtained from the original forecast and observation fields by thresholding with the same threshold used to obtained the binary forecast and observation fields. This formula follows by considering the

Murphy and Winkler (1987) framework, apply the Bayes' theorem to express the joint probabilities b/n and c/n as product of the marginal and conditional probability (e.g. Jolliffe and Stephenson, 2003; Wilks, 2006), and then noticing that for a random forecast the conditional probability is equal to the unconditional one, so that b/n and c/n are equal to the product of the corresponding marginal probabilities solely.

5. For each threshold (t) and scale component (j), the skill score based on the MSE of binary forecast and observation scale components is evaluated (Figure 5). The standard skill score definition as in Jolliffe and Stephenson (2003) or Wilks (2006) is used, and random chance is used as reference forecast. The MSE for the random binary forecast is equipartitioned on the n +1 scales to evaluate the skill score: $SS(t,j)=1-MSE(t,j)*(n+1)/MSE(t)_{random}$

   The Intensity-Scale (IS) skill score evaluates the forecast skill as a function of the precipitation intensity and of the spatial scale of the error. Positive values of the IS skill score are associated to a skillful forecast, whereas negative values are associated to no skill. Usually large scales exhibit positive skill (large scale events, such as fronts, are well predicted), whereas small scales exhibit negative skill (small scale events, such as convective showers, are less predictable), and the smallest scale and highest thresholds exhibit the worst skill. For the NIMROD case illustrated note the negative skill associated to the 160 km scale, for the thresholds ½ to 4 mm/h, due to the 160 km storm displaced almost its entire length.

In addition to the MSE and the SS, the energy squared is also evaluated, for each threshold and scale (Fig 6). The energy squared of a field X is the average of the squared values: $En2(X)= \Sigma_i x_i^2$. The energy squared provides feedback on the amount of events present in the forecast and observation fields for each scale, for a given threshold. Usually, small thresholds are associated to a large energy, since many events exceed the threshold. Large thresholds are associated to a small energy, since few events exceed the threshold. Comparison of the forecast and observed squared energy provide feedback on the bias on different scales, for each threshold.

The En2 bias for each threshold and scale is assessed by the En2 relative difference, equal to the difference between forecast and observed squared energies normalized by their sum: [En2(F)-En2(O)]/[En2(F)+End2(O)]. Since defined in such a fashion, the En2 relative difference accounts for the difference between forecast and observation squared energies relative to their magnitude, and it is sensitive therefore to the ratio of the forecast and observed squared energies. The En2 relative difference ranges between -1 and 1, positive values indicate over-forecast and negative values indicate under-forecast. For the NIMROD case illustrated the forecast exhibits over-forecast for small thresholds, quite pronounced on the large scales, and under-forecast for high thresholds.

As for the MSE, the sum of the energy of the scale components is equal to the energy of the original binary field: $En2(t) = \Sigma_j En2(t,j)$. This enables to calculate, given a threshold, in which percentage the energy on each scale contributes to the total energy: En2%(t,j)=En2(t,j)/En2(t). Usually, for precipitation fields, low thresholds exhibit most of the energy percentage on large scales (and less percentage on the small scales), since low thresholds are associated to large scale features, such as fronts. On the other hand, for higher thresholds the energy percentage is usually larger on small scales, since intense events are associated to small scales features, such as convective cells or showers. The comparison of the forecast and observation squared energy percentages provides feedback on how the events are distributed across the scales, and enable the comparison of forecast and observation scale structure.

For the NIMROD case illustrated, the scale structure is assessed again by the relative difference, but calculated of the squared energy percentages. For small thresholds the forecast over-estimates the number of large scale events and under-estimates the number of small scale events, in proportion to the total number of events. On the other hand, for larger thresholds the forecast under-estimates the number of large scale events and over-estimates the number of small scale events, again in proportion to the total number of events. Overall it appears that the forecast over-estimates the percentage of events associated to high occurrence, and under-estimate the percentage of events associated to low occurrence. The En2% for the 64 mm/h thresholds is homogeneously under-estimated for all the scales, since the forecast does not have any event exceeding this threshold.

Note that the energy squared of the observation binary field is identical to the sample climatology Br=(a+c)/n. Similarly, the energy squared of the forecast binary field is equal to (a+b)/n. The ratio of the squared energies of the forecast and observation binary fields is equal to the FBI=(a+b)/(a+c), for the contingency table (Table 1) obtained from the original forecast and observation fields by thresholding with the same threshold used to obtained the binary forecast and observation fields.

2.2 The spatial domain constraints

The Intensity-Scale technique is constrained by the fact that orthogonal wavelets (discrete wavelet transforms) are usually performed on square domains of 2^n x 2^n grid-points. The MET Intensity-Scale verification tool address this issue automatically, depending on the shape and dimension of the forecast domain, in the following 4 fashions:

1. Cropping – masking: if the domain size is slightly larger than a square domain of 2^n x 2^n grid-points, the used is suggested to provide a square mask of 2^n x 2^n grid-points, preferably over the domain region where would be more useful to perform the verification (e.g. the mask could eliminate part of the domain over the ocean, and focus on the continental domain, where usually observations are better estimated).
2. Padding: if the domain size is slightly smaller than a square domain of 2^n x 2^n grid-points, for certain variable (e.g. precipitation) is advisable to expand the domain to be a square domain of 2^n x 2^n grid-points by adding zeros at the boundaries.

3. Interpolating: if the domain dimension are of similar order, an interpolation could be applied to obtain forecast and observations over a square domain of 2^n x 2^n grid-points. Note that the interpolation should respect the field characteristics: nearest neighbor interpolation is advised for precipitation, because of its discontinuous nature and to preserve peak values; cubical interpolation is advised for smoother variables, such as temperature.
4. Tiling: this approach does not involve any field reduction, expansion or values alternation by interpolation. It consists in performing the Intensity-Scale verification on squared tiles, with dimension equal to the largest 2^n x 2^n grid-point tile contained in the domain given, and cover with these tiles the entire domain, allowing eventually tiles to overlap. The verification statistics are obtained by aggregating the statistics from all the tiles, as described in the following section. Note that with this approach the center of the domain will be sampled more than boundary grid-values, however the grid-points counted for a possible multiple case aggregation are not repeated when tiles overlap.

2.3 Aggregation of statistics on multiple cases

The Intensity-Scale analysis tool enables to aggregate the intensity scale technique results. Since the results are scale-dependent, it is sensible to aggregate results from multiple model runs (e.g. daily runs for a season) on the same spatial domain, so that the scale components for each singular case will be the same number and the domain, if not a square domain of 2^n x 2^n grid-points, will be treated in the same fashion. Similarly, the intensity thresholds for each run should be all the same.

The MSE and forecast and observation squared energy for each scale and thresholds are aggregated simply with a weighted average, where weights are proportional to the number of grid-points used in each single run to evaluate the statistics. If the same domain is always used (and it should) the weights result all the same, and the weighted averaging is a simple mean. For each threshold, the aggregated Br is equal to the aggregated squared energy of the binary observation field, and the aggregated FBI is obtained as the ratio of the aggregated squared energies of the forecast and observation binary fields. From aggregated Br and FBI, the MSE_{random} for the aggregated runs can be evaluated using the same formula as for the single run. Finally, the Intensity-Scale Skill Score is evaluated by using the aggregated statistics within the same formula used for the single case.

Bootstrapping by using re-samples obtained by a random selection with replacement, can provide confidence intervals for the Intensity-Scale technique aggregated results.

References

Briggs WM & Levine RA (1997) *Wavelets and field forecast verification*. Mon Wea Rev **125**: 1329-1341.

Casati B, Stephenson DB, Ross G (2004) *A new intensity-scale technique for the verification of spatial precipitation forecasts*. Met Apps **11**: 141-154.

Casati B & Wilson LJ (2007) *A New spatial scale decomposition of the Brier score for the verification of lightning probability forecasts*. Mon Wea Rev **135**: 3052-3069

Casati B, Wilson LJ, Stephenson DB, Nurmi P, Ghelli A,Pocernich M, Damrath U,Ebert EE, Brown BG, Mason S (2008) *Forecast verification: current status and future directions*. Met Apps **15**: 3-18.

Harris D, Foufoula-Georgiou E, Droegemeier KK, Levit JJ (2001) *Multiscale statistical properties of a high-resolution precipitation forecast*. J Hydromet **2**: 406-418.

Jolliffe, IT and Stephenson DB (2003): Forecast Verification, a practitioner's guide in atmospheric science. Wiley & Sons, 240 pp.

Murphy AH and Winkler RL (1987): A general framework for forecast verification. Monthly Weather Review **115**: 1330-1338.

Wilks DS (2006): Statistical methods in the atmospheric sciences. Academic Press, 627 pp.

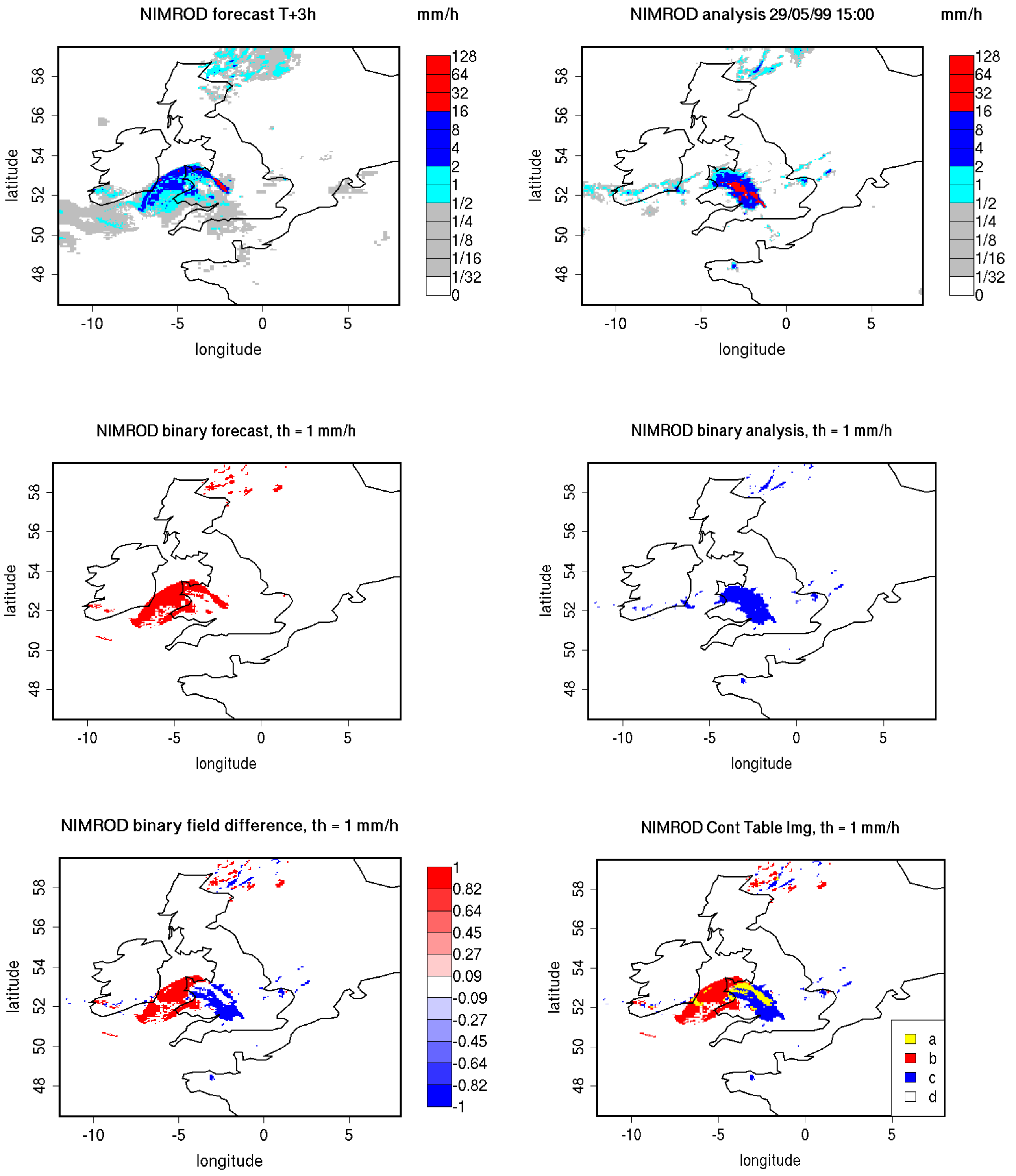
NIMROD forecast T+3h
mm/h
NIMROD analysis 29/05/99 15:00
mm/h
128
64
32
16
8
4
2
1
1/2
1/4
1/8
1/16
1/32
0
latitude
longitude
58
56
54
52
50
48
-10
-5
0
5
NIMROD binary forecast, th = 1 mm/h
NIMROD binary analysis, th = 1 mm/h
NIMROD binary field difference, th = 1 mm/h
1
0.82
0.64
0.45
0.27
0.09
-0.09
-0.27
-0.45
-0.64
-0.82
-1
NIMROD Cont Table Img, th = 1 mm/h
a
b
c
d

Figure 1: NIMROD 3h lead-time forecast and corresponding verifying analysis field (precipitation rate in mm/h, valid the 29/05/99 at 15:00 UTC); forecast and analysis binary fields obtained for a threshold of 1mm/h, the binary field difference their corresponding Contingency Table Image (see Table 1). The forecast shows a storm of 160 km displaced almost its entire length.

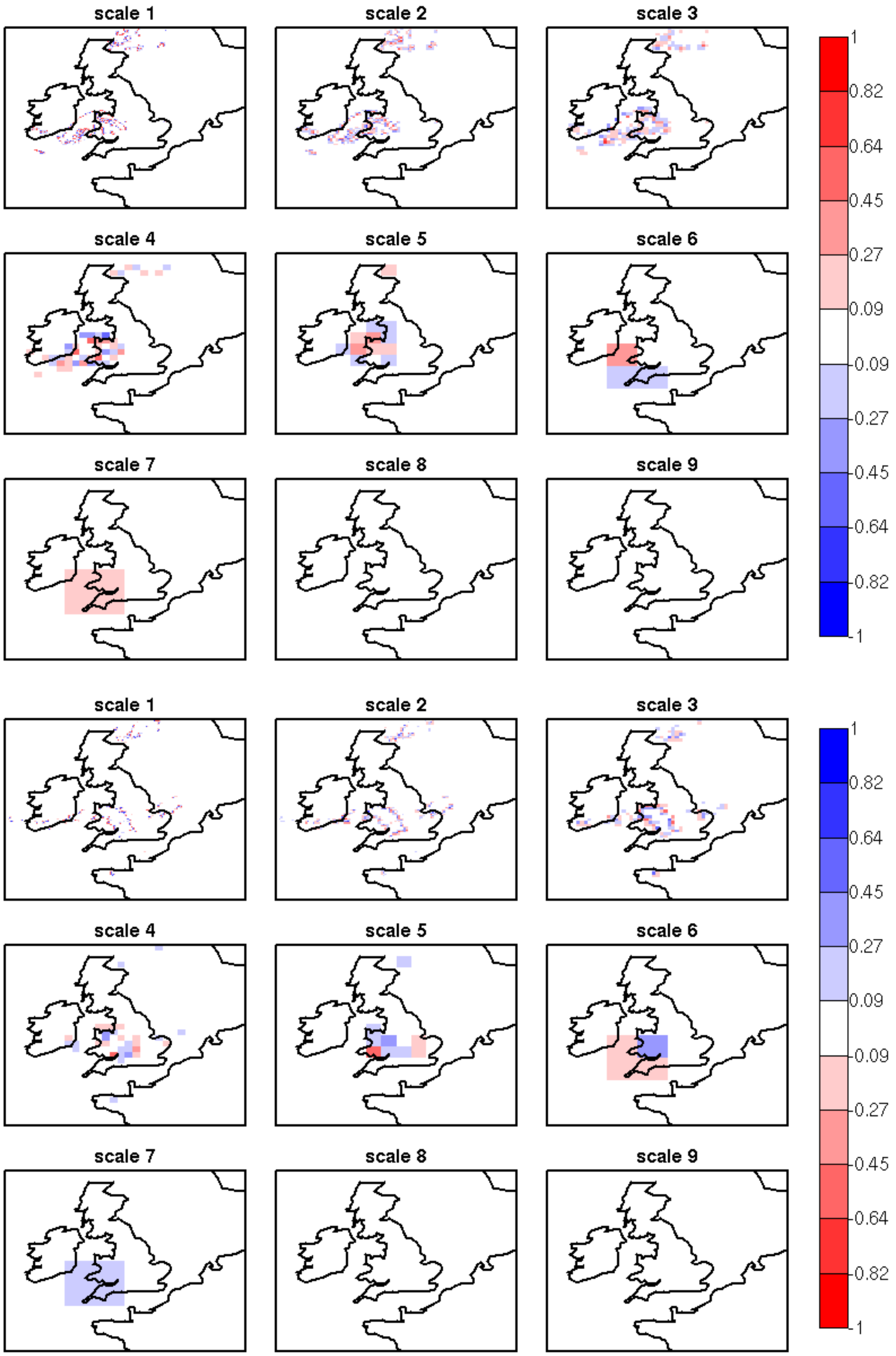

Fig 2. NIMROD binary forecast (top) and binary analysis (bottom) spatial scale components obtained by a 2D Haar wavelet transform (th=1 mm/h). Scale 1 to 8 refer to mother wavelet components (5, 10, 20, 40, 80, 160, 320, 640 km resolution); scale 9 refer to the largest father wavelet component (1280 km resolution)

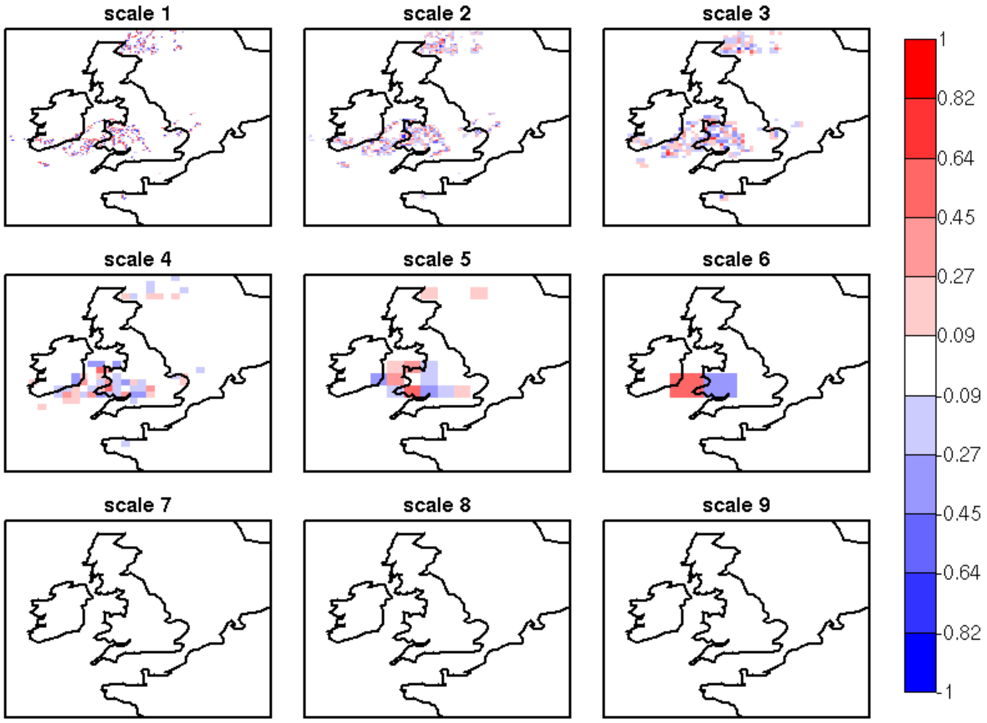

Fig 3. NIMROD binary field difference spatial scale components obtained by a 2D Haar wavelet transform (th=1 mm/h). Scale 1 to 8 refer to mother wavelet components (5, 10, 20, 40, 80, 160, 320, 640 km resolution); scale 9 refer to the largest father wavelet component (1280 km resolution). Note the large error associated at the scale 6 = 160 km, due to the storm of 160 km displaced almost of its entire length.

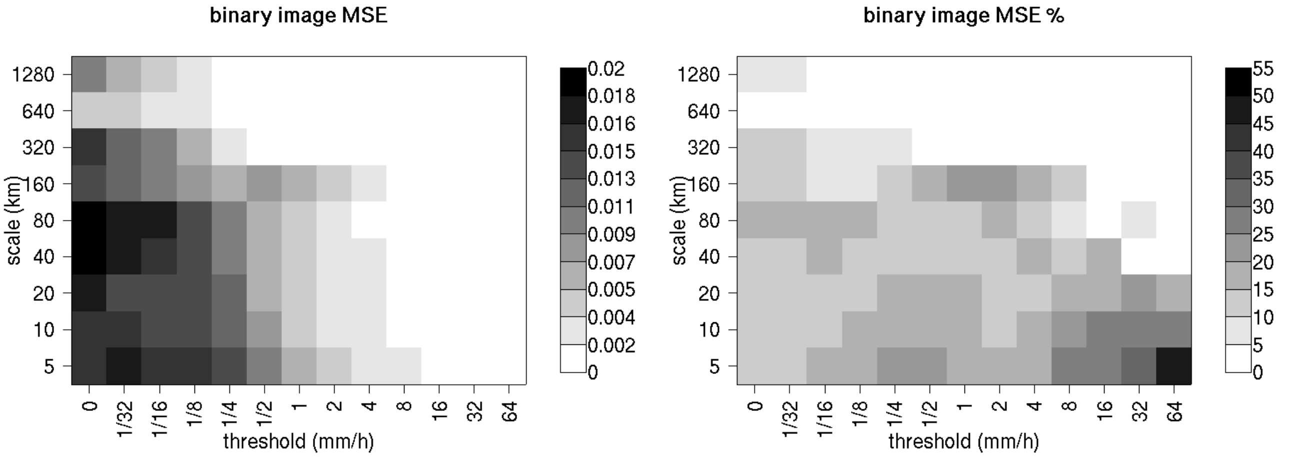

Fig 4. MSE and MSE % for the NIMROD binary forecast and analysis spatial scale components. In the MSE%, note the large error associated to the scale 6 = 160 km, for the thresholds ½ to 4 mm/h, associated to the displaced storm.

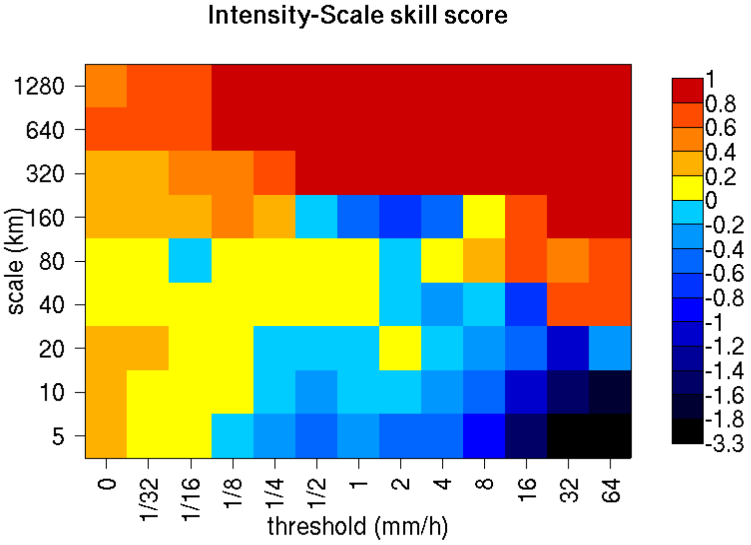

Fig 5. Intensity-Scale skill score for the NIMROD forecast and analysis shown in Fig 1. The skill score is a function of the intensity of the precipitation rate and spatial scale of the error. Note the negative skill associated to the scale 6 = 160 km, for the thresholds ½ to 4 mm/h, associated to the displaced storm.

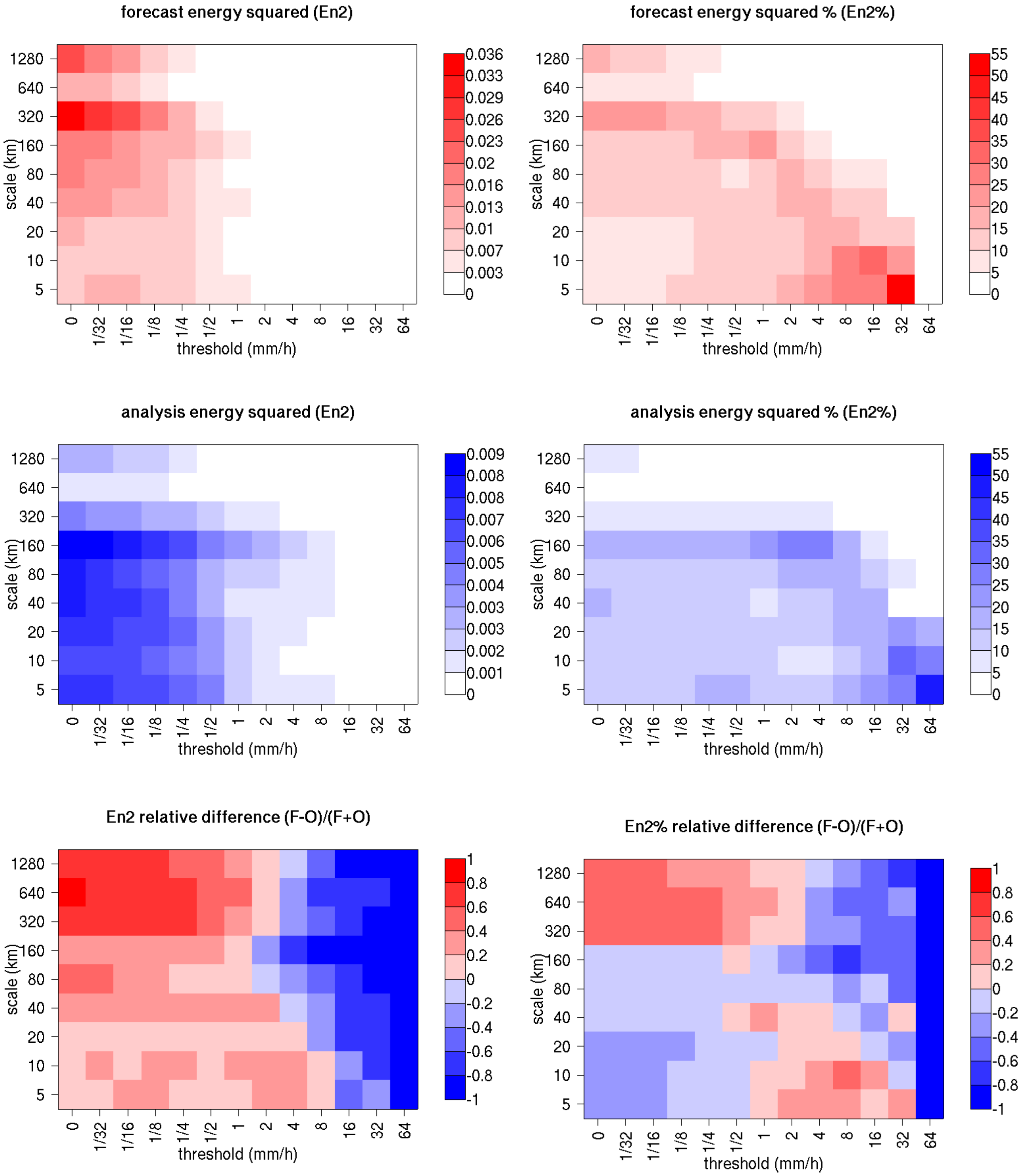

Fig 6. Energy squared and energy squared percentages, for each threshold and scale, for the NIMROD forecast and analysis, and forecast and analysis En2 and En2% relative differences.

	obs > th	obs < th	
forecast > th	a=hits	b=false alarms	a+b
forecast > th	c=misses	d=correct rejections	c+d
	a+c	b+d	n=a+b+c

Table 1: Contingency Table: the counts a,b,c,d correspond to the hits, false alarms, misses and correct rejections for the threshold th.