

00:00UTC Session on Spatial Scores

Keynote talk discussion:

(Dominique Brunet) Can you please talk a bit more about the subjective score (Dominique Brunet)

Answer: Cases ranking are from 9 cases of the ICP, we put 20 or 30 people into a room with varied background. Basically, showed them slides of the cases, total of three models and an analysis. Asked which one was liked best? Gleaned some good information from that. There was a lot of disagreement. Not super scientific.

[7:20pm] Is $G(A, A\text{-complement})$ fixed, or dependent on beta? (Deryn Griffiths)

[7:21pm] Am I correct that this is designed for "events" and you could not change to treat the non-event area as the event and expect it to perform well? (Deryn Griffiths)

A: What trying to say you tune the beta to what kind of error you want to be "bad". It doesn't reward for the partially perfect. The lower the beta, the greater the penalty. Once you choose the beta, it's fairly consistent. Domain size can mean larger distances with larger domains.

[7:26pm] It is always dependent on beta. (Eric Gilleland)

[7:28pm] You can turn it around to think of the non-events as the interested part. If I understand correctly. (Eric Gilleland)

[7:37pm] Would beta depend on grid distance/model resolution? (Chakraborty, Paromita)

[7:37pm] Also a good idea. In my paper, I make a rose diagram that shows the bearing between points that "match" through sorting. The issue is in determining where to make those matches. You could do that inside a MODE-like approach. (Eric Gilleland)

[7:38pm] Patrick, we used event probabilities for fire spread simulations (Beth Ebert)

[7:38pm] Chakraborty, Paromita (National Centre for Medium Range Weather forecasting (NCMRWF), India) It depends more on the size of areas * distance that you are looking at. The size of the domain only comes in through the MED terms that mean that you could have longer distances for larger domains. What I didn't mention is that for the rare event, I chose the largest area times a translation error of 20 grid squares, which gave the vertical dashed line.

Patrick Skinner, Object-Based Verification Techniques for Short-Term Thunderstorm Forecasts

Discussion:

[7:35pm] After identifying the objects, and their location, why not evaluating also the distances (and orientation / trajectory), further than traditional contingency table scores? (Casati, Barbara)

A: In earlier papers, weighted by the area and the distance. The variation was such that different forecast configurations provided a different story. Can weight the scores by distance.

Q: Harvir Singh: Question about MODE, can this be used for heat and cold waves?

A: Yes, you can use this for different meteorological events, just need to define the object. Paper by Jamie Wolff (2014) is useful for this.

A: Tara Jensen comment, are you having trouble using MODE to identify cold wave and heat wave objects?

Harvir Singh: Cannot define the objects correctly, issues with reading GRIB data.

A: Tara Jensen, please write the MET help desk for assistance.

Q: Tara Jensen for Patrick: Regarding the points verification. What I am understanding is that you are looking at a spatial method, you have three different types of LSMs, are you expecting at least one member with each of the parameterizations?

A: That is correct, we have methods for PBL schemes and others. At least one of those members have to be matched to the observations for each of the PBL schemes

Comment: Interesting twist on introducing methods into ensemble stat, working with Marion Mittermaier on these issues and new requirements for METplus.

[7:48pm] We applied MODE to drought in this paper: doi: [10.1175/JHM-D-16-0125.1](https://doi.org/10.1175/JHM-D-16-0125.1))Eric Gilleland

[7:56pm] Barb - Sorry, I forgot to mention several relevant points! We did explore using orientation in object matching but found that the impacts were small relative to the spatial measures. However, plotting the centroid displacement at different lead times helped quantify storm motion biases. Also, we looked at the orientation of the composite reflectivity fields relative to observations to assess how well we were predicted storm size/structure. (Patrick Skinner)

[8:34pm] When you are matching forecast and observed objects how do you handle the situation when there is more than 20 min difference in when the objects initiate, but they match up later? Is that a hit or a miss/false alarm, or both? (Beth Ebert)

[8:35pm] Beth - we classify the objects at each available output time, so it would be a miss (assuming the observed object is there first) until the forecast object is able to clear the total interest threshold (<= 20 minutes for a perfect spatial forecast) (Patrick Skinner)

[8:36pm] Thanks, I thought maybe that was the case (Beth Ebert)

[8:37pm] Do you try to compute event statistics, i.e. was a given storm a hit, miss or false alarm? (Skinner)

[8:42pm] Beth - to be more specific, we haven't done that for full storm lifecycles (i.e. was the storm matched at any point) (Skinner)

Dominique Brunet: Insights from Image Quality Assessment

Discussion:

Q: Not familiar with a lot of scores in the presentation, can you talk about them? (Barb Casati)

A: SSIM, for example – referred to equation in presentation in slide 19.

[this part of the discussion referred to many of the equations in the presentation on slide 19 and slide 12, so difficult to capture in discussion text]

Q: Please add references to these metrics? (Barb Casati)

A: Yes, will add more information to presentation on references to the metrics.

Q: You prompted an idea not seen before, an aggregated score for combining all the scales together. Has there been much reaction to this score? (Beth Ebert)

A: This is research, still defining ideas.

Presentation for Bin Zhao was skipped due to presenter not logged into MSTeams.

Natali Giselle Aranda: Spatial verification of high-resolution precipitation of forecasts over southern South America

Discussion: No questions

Comment: Changing the values of R and T, the total interest value got bigger, like some other neighborhood methods. Is anyone using MODE in that way, to decide on which scale to go to to get a sufficiently good match? (Beth Ebert)

Comment: Brown It is similar to what you might do with Fraction Skill Score and other methods, you can get a pattern to see matching and how it changes based on scale. (Barb Brown)