15:00UTC Session on Spatial Scores

Keynote talk discussion:
[10:13 AM] Ashrit, Raghavendra (National Centre for Medium Range Weather Forecasting, National Informatics Centre, India)

What makes a skill score 'proper'?

Barbara Casati (BC): difficult to hedge, cannot change forecast to improve score (e.g. by overforecasting)

Brooks, Harold (NOAA/National Severe Storms Laboratory)

Here's how I teach the definition of scoring rules. Strictly proper is the highest-a forecaster gets their best score if and only if they forecast their true beliefs about the outcome. A proper score rule is one for which a forecaster gets their best score by forecasting their true belief, but could get that same score by forecasting something other than their true belief. A consistent score is one for which you get a better score by forecasting your true belief. For many kinds of forecasts, there are no proper or strictly proper scoring rules. For probability forecasts of yes/no events, there are two strictly proper rules: Brier score and the ignorance score.

[10:30 AM] Marsigli,Chiara (Deutscher Wetterdienst, Germany)

What were the key factors to permit the usage of the MET for so many different applications?

Barbara Brown (BB): Major key to it is that there are a lot of options for reformatting forecasts and observations, which make it possible to apply the same type of technique for different applications (e.g. algae)

Tara Jensen (TJ): There is a lot of reasons… Many config files allow to reuse code, automated regridding tools, etc.

BB: Some statistics computed EMC are included, these things add to the MET capabilities.

[10:32 AM] Jonathan Day (Guest)

Is MET still focussed on applications to WRF, or is the package usable with any FC system?

TJ: Can handle Grib1, Grib2 and NetCDF, it does not matter what the model format the model is in.

[10:33 AM] Mohan S Thota (Guest)
How does MET or METplus account for the quality of the observations used for verification?

TJ: we don't do QC assessment, the user can configure MET to only use QC flags already done beforehand, but it is beyond the scope to do all QC

[10:33 AM] Casati,Barbara (ECCC)
MET is usable by any FC community, several operational centres are adopting it for operational verification.

[10:34 AM] Brooks, Harold (NOAA/National Severe Storms Laboratory) (Guest)
As a historical note, Kim Hoogewind and I are starting a project to use environmental conditions, similar to what we use now to forecast tornadoes, derived from the 20th Century Reanalysis to redo Finley's forecast experiment. We (Kim Hoogewind and I) are going to take 20th Century Reanalysis fields that would be what we use to forecast storms now. We'll use those fields to make forecasts like Finley did. Then, we can use an "environmental skill score" that Alex Anderson-Frey and I have in press to evaluate those forecasts. Essentially, we're going to try to re-do the Finley experiment using our modern understanding of the ingredients for severe thunderstorms.

[10:35 AM] Ashrit, Raghavendra (National Centre for Medium Range Weather Forecasting, National Informatics Centre, India) (Guest)

At NCMRWF in India MET is used for verification of GFS and NCMRWF Unified Model forecasts.

[11:05 AM] Mittermaier, Marion (MM)

MET is a little bit US-centric but becoming less so all the time!

Gregor Stok talk:

MM: For gridded precip fields, do you use gridded analysis from ECWMF?

- Yes.

MM: If you would use radar analysis, what it would influence?

- Not so much.

MM: Second question: The proposed score is for continuous field, does it still work with zeros in precip in contrast to binary fields?

- Yes

Raghavendra Ashrit: Idealized type of evolution. Do you have precip in mind or applicable to any field?

- Easier with precip. You can define events more natural, for other field I would to need think about it.

SK Sagar: DNSS score with kind of science we can expect from this score how the score will indicate how to improve the model?

- I don’t know, you will get average displacement, but what exactly to do reduce displacement is not very clear.

SK Sagar: If we use CRNA(?), we will get three fields.

- DNSS only get displacement

Carlo Cafaro (Guest):
Thanks Greg for the great talk. So it is not clear to me whether the NSS is sensitive to the bias or not.

- The fields need to be unbiased. By multiplying by some constant before.

[11:52 AM] Skok, Gregor (Guest): Since the bias is removed prior to the calculation of the score the dNSS is not sensitive to the bias. This also means that dNSS cannot be hedged by multiplying one or both fields with a constant value..

Rachel North talk:

[11:41 AM] Casati,Barbara (ECCC)
The differences between gauges and satellite climatologies can be used to identify sites affected by representativeness errors (or other gauge Quality Control issues)

[11:43 AM] Casati,Barbara (ECCC)
I would be curious to know if these gauges, showing large discrepancies, are assimilated in the NWP analysis. Moreover, are they blacklisted for the WMO score exchange?

- That’s something that we need to look at, restrict ourselves to a list of specific stations that we know they have good quality… I would be interested to see what ECWMF is doing.

BC: Can we use error statistics from data assimilation?

- Yes, good idea, thank you.

[11:49 AM] Casati,Barbara (ECCC)

- how can the observation uncertainty (e.g. from the DA error statistics on the satellite products) be brought into the score?

MM: The Gauge climatology comes from ECMWF…. we need to weed out the locations which have strange climatologies, mostly because there are some spurious values that make it into the climatology calculation.

[11:52 AM] Marsigli,Chiara (Deutscher Wetterdienst, Germany) (Guest)

It seems to be a very good tool to identify systematic model error and eventually to perform process verification

BC: gridded verification exchange with MET to be used to evaluate the scores

MM: seeps as example, is gauge-based score, to turn it in spatial score by element of upscaling because global model is finer resolution than TRMM gird ~25km. Seeps not picked because cannot get consistent climatology across centers. TRMM climatology offers a way a) common grid, b) common climatology once seeps is implemented in MET. Agree. We are converging towards something feasible and doable.

Ashrit: TRMM how it can represent extremes TRMM has serious bias according to some of my studies, any bias correction done for TRMM, not bias adjustment done on it yet.

- We are using local data source that is more reliable than TRMM or GPM, it would make sense to use local data sources when possible.

BC: Gridded observation, two elephants obs uncertainty and representativeness, can benefit from DA community, in one sense we want a gridded product to combine the two, but also want to independent data sets if they agree we are set,

- I agree to look at both direction to get uncertainty

MM: agree struggle how to make it to agree if they measure something fundamentally different. In the paper we look how representative the sampling of the land area is. Gauge are where they are people, we have little ground based measurement in remote locations. Colocated grid squares, where we are sampling is not representative of the model as a whole. Performance tuned at certain locations but does not mean it is good everywhere.

TJ: One thing MET has capabilities to include ensemble uncertainty for spread skill type diagram, got around and around to include observation uncertainty. We would like to encourage to provide us some methods to integrate in METplus framework if someone stumble about method that is scientifically robust, please reach out to us.

Chiara: missing evaluation a lot is needed forecast spatially, how to evaluate spread skill spatially? I would help as much as I can to put inside MET tool.

Paramita: Is SEEPS part of MET package.

TJ: It is scheduled to put it within next 6 months, collaborate with MET office, we do have resources now to implement it

Jessica Baker: A comment : Found results very interesting gauges sparse in tropics south America Africa hardly any station data so need satellite data one step towards evaluating TRMM

[12:06 PM] Jess Baker (Guest)

Very interesting result Marion and Rachel! I look forward to reading the paper. In the tropics gauges especially scarce, particularly Africa. Makes it hard to assess the quality of remote-sensing products.

BC: I have three major questions, but I will keep it for MET session.

Raghavendra: what is the forecast perspective how they use the spatial verification scores? I have the impression in India it is more expert friendly but not usasble to day to day forecaster, how to communicate to make it more useful to forecasters?

Chiara: experience I had in DWD part of spatial verification is neighborhood, the forecasters like to see precip prob upscale in the neighborhood, same thing COSMO in … the forecaster would like to see the skill of the forecast over an area (e.g. warning area) represent a skillful scale for forecast. Appreciated by forecasters.

BC: At EC we are exploring we realize traditional scores don't tell the full picture. We find different answers with spatial verif. Quantile plots discover something we did not know before

A lot of eye ball verification. We need objective verification that we see by eye.

MM: At Met office neighborhood popular of model development score cards ingrained in scientist in developer they get exposed by forecaster not feedback, but they look at systematic bias, subjective verification very important part.

TJ: In the states, use MODE at diagnostic tool but also on website, use it a pp product generation to id obje ens of MODE object for intence rain fall. Storm pred center use obj based evalutation,. Sptial method ave nto infiltrated forecast workflow that much

Paromita: underdispersion of ens, what do we need to look on a model perspective? If bias, we do bias correction.

BC: for PP or for tweaking the model? Modellers should know.

Chiara: very good question. Recently what has been done is to different component IC model or Bond C, what part is lacking unc perturb, then combine all together, b/c sometimes then sometimes they can cancel each others.

Paromita: can we really segregate unc arising between different components? Seen a paper saying not. Error cancelling it. IC and model unc overlapping

BC: one recommendation, verify at each update, you can disatanggle. QQ plots cn show us simulation correcting better bias, but because bias cancelling out for low and high values, it was actually worse

Tim Bullock: about forecaster, thoughts and considerations. Answer some questions about model developpers. Question related to decision they have to make. As we can engage more user differerent set of decision different needs differnet metrics to be used.

[12:07 PM] Ashrit, Raghavendra (National Centre for Medium Range Weather Forecasting, National Informatics Centre, India) (Guest)

Rachel, Marion pl share the paper details

[12:12 PM] Mittermaier, Marion (Guest)

At Met Office neighbourhood methods and FSS are also the most common method that is used. SOme of this info is communicated with the forecasters but it would be fair to say that we have very little feedback.

[12:13 PM] Ashrit, Raghavendra (National Centre for Medium Range Weather Forecasting, National Informatics Centre, India) (Guest)

Same here

[12:13 PM] Mittermaier, Marion (Guest)

Within the model development community it is very different, where it is used all the time.

[12:14 PM] Mittermaier, Marion (Guest)

Subjective verification is still a key part of model upgrade assessment.

[12:18 PM] Mittermaier, Marion (Guest)

MODE is definitely in that direction.

Dominique Brunet: I glad that we are discussing about subjective (eye-ball) verification as it is something that I was not sure if it is considered in verification. Basically, verification scores for users should be made to match their subjective judgement. (Plugging own talk in next session.)

[12:21 PM] Casati,Barbara (ECCC)

For Dominique: most of the spatial methods originated to mimic what an eye-ball verification would diagnose