

Verificaton and Evaluation of Environmental Prediction Systems at the NOAA Environmental Modeling Center – Jason Levit

Discussion:

Michael Foley (MF): I have a question about the forecaster involvement in the evaluation. When you get the subjective view on the model performance, how well does that match with the verification metrics? How to get more forecaster involvement in interpreting metrics (I think)? Are the forecasters involved with decisions about what the right quantities to look at are?

Jason Levit (JL): We're always looking for ways to get forecasters more involved. They can join teams such as the MEG and have enormous capabilities to look at verification products/assess forecasters, and provide feedback.

MF: Do you tend to get consistent views from forecaster evaluations?

JL: Yes we do, we're looking for systematic errors in the systems, though there is case-to-case verification. There are instances where forecasters identify specific forecast errors (e.g. slight biases in TC tracks) that are missed by bulk metrics.

(from chat):

Barbara Casati (BC): How much is subjective interpretation of model output weighted relative to objective verification metrics, and has a model ever been stopped from being implemented?

JL: The relative weights are case dependent. There have been cases where model changes have not been implemented owing to poor performance, but not in a while. Different centers have different priorities, which may inform the relative weighting of subjective/objective verification measures. Each center is given an opportunity to present their evaluation of model changes and provide a thumbs up/thumbs down vote. We usually know enough going into an evaluation period to be confident that forecast performance will improve, but surprises do happen.

BC: Do you evaluate multiple changes at once or one at a time?

JL: One at a time.

Deryn Griffiths (from chat): Are you using a common truth dataset for model verification or the model analysis?

JL: We are using the model analysis. Not sure what specific time is used for GFS, thinks it's slightly past the 0 hour.

JL: Chat about common obs base/analysis across different international centers – that would be a great thing to look into together.

Methods and Tools Used to Verify Convection-Allowing Model Guidance at the NCEP/Environmental Modeling Center – Logan Dawson et al.

Patrick Skinner (PS): Is interested in using SPC outlook areas as masks for thunderstorm verification. Does using a smaller threshold like the general thunder outlook change the results? How do you interpret the results since, as you mentioned, you would expect better predictability with organized convection in outlook areas. Do lower skill forecasts outside of outlooks provide useful information?

Logan Dawson (LD): Convective outlook areas are generally defined as Marginal or greater, could use the general thunder line and that may be a good way to look at this a bit farther and tease out what we can in the verification.

Exploring Spatial Distributions of Systematic Errors in the NCEP's Global Ensemble Precipitation Forecast Products - Yan Luo

questions deferred in interest of time

A New METplus-based Verification System for the Global Forecast System (GFS) – Mallory Row and Jason Levit

Tom Robinson (TR): Who is the contact for the content on the verification statistic webpage? (for the GFS verification page)

Mallory Row (MR): I am.

A New Webpage for Visualizing Verification Statistics from the National Centers for Environmental Prediction Production Suite – Alicia Bentley et al.

Beth Ebert (BE): Are the plots on the webpage static plots or is interaction via mouseover possible?

Alicia Bentley (AB): Most are static but some (under analyses) are interactive.

BE: Are you thinking of moving towards interactive images?

AB: For now we are planning on staying with static images, but long term, interactive plots are a good idea.

Tara Jensen (TJ): There are some plots available through MET Express and METViewer that support interactivity. She is not sure how they would work within a web interface.

AB: Some of the interactive options on the webpage are direct links to interactive MET Viewer plots

General discussion:

TJ: It resonates with me that verification is getting as complex as the model. Especially as the manager of a tool that is growing exponentially ... to be able to provide verification for coupled model systems. I hope that out of these new and exciting things that we as a community can learn and draw from what other communities have learned so that things that might be useful in hydrology may be useful for things like space weather. There are more common threads than people think there are and I encourage this community to think of the common threads to keep the problem from being too complex.

The Model Evaluation Group (MEG) at the Environmental Modeling Center – Geoff Manikin et al.

TR: The operational meteorologists are working shifts and do some of the things you're (MEG) doing, and maybe some other things you aren't depending on the clients for the forecasts (like flood response). The culmination of the objective and subjective scores is something that we haven't done to the extent you have – not really a question but a comment that I like what you're doing.

GM: The stats are absolutely critical but there are specific forecast errors that are really tough to tease out in stats (missed the specific examples), and by looking at large numbers of cases that's something we feel comfortable noting as a concern in the model. And as we go live we can pass those findings on to the forecasting community.

BE: Spotting the convection along the coast and linking that to the climb in SSTs is something that can't be teased out in stats, but a lot of organizations don't have the resources of the MEG. Is there any way to automate detection of these features?

GM: That's a tough one, something widespread will show up in the stats, but something small like a SST bias along a coast, unless you have a specific regional sector for verification, it's tough to see. We were starting to see it in the stats a bit, but caught it much sooner by looking at the maps. Something that dramatic will become manifest in the stats eventually, but daily interrogation of the maps allows you to spot these issues before you get into trouble.

Deryn Griffiths (DG): Are you looking at verification of the National Blend of Models?

JL: We are just verifying EMC forecasts, but Jeff Craven does his own verification for the NBM.

BE: NOAA has a performance and evaluation branch that does verification of all the official forecasts (and maybe the NBM). Do you have an affiliation with the PEB?

JL: The performance and evaluation branch does work with us and is responsible for verification of all official forecasts (e.g. tornado warnings). They do request some model verification stats, particularly the 120 hr 500 mb height anomaly correlation coefficient for the GFS.

Harold Brooks (HB): "Making a good forecast is hard, evaluating it is just as hard" - Alan Murphy

Eric Gilleland (EG): On complexity of verification – that's a good reason to use the distance measures.

LD: Comment on verifying against analysis – precip verification uses analysis of QPE, radar verified against somewhat independent analysis, can also verify operational CAMS using RTMA analyses. RTMA is being converted into a 3-d system to provide analyses of variables such as CAPE/reflectivity. Should become available for verification of fields that we don't have apples-to-apples comparisons for right now. When we switch will need to be careful to consider accuracy of model background and availability of obs.

TJ: RTMA is more CONUS scale, what is its applicability to the world?

LD: Good question – including more satellite-based observations might make it more available across the world. We're focused on high-resolution modeling in North America and will need to find a solution to be able to verify outside of CONUS. Satellite-based analyses will help there.