Spatial Verification Discussion, 17 November 2020 08 UTC

Bent Sass – Forecasting spatial structure of local precipitation extremes

Deryn Griffiths – I like that you are tackling minima as well as maxima in this score.

> Bent Sass - Thank you

Chiara Marsigli – How does the score behave when there are multiple maxima in the area?

> Bent Sass – It depends on whether there is small or large distance between the features. If there are several showers, then the score will be an average of how well you forecast each shower.

Seonaid Anderson – The NFLICS project: Nowcasting FLood Impacts of Convective Storms in the Sahel

Marion Mittermaier – So, am I right in thinking that you are using neighbourhoods to create the forecast and neighbourhoods to verify?

> Seonaid Anderson – Correct. We use neighbourhoods to generate probabilities and also apply the same spatial filter to the observation images to verify the forecast.

Manfred Dorninger – Do you take into account strengthening and weakening of the systems or building up of new systems?

> Seonaid Anderson – Really interesting point, highlights differences between this approach and cell tracking. If climatologically the cells tend to grow and propagate toward the coast, then this method will pick it up. Similarly, in the early morning hours cells tend to drop off. It's an indirect effect.

Carlo Cafaro – Interesting talk Seonaid. So, did you apply the "maximum" method to create probabilities? Do you apply this to observations as well? Can this method be applied to deterministic model to extend the lead time?

> Seonaid Anderson – For historical storms calculate convective structures, convert to 1's and 0's, probability is fraction of neighbourhood with 1's. Same neighbourhood size is applied to the observed (analysis) rainfall. The performance depends on the size of the convective structures.

Deryn Griffiths – You seem to use different spatial scales for different lead times and times of day. How do you communicate or use your forecast if the definition doesn't have stability?

> Seonaid Anderson – Thanks for the question! The spatial scale varies with lead time but not time of day. We deliberately avoided a time-of-day variation to avoid confusion. Also hopefully by showing the scale on the nowcast diagrams as a function of lead-time helps clarify things

Chiara Marsigli - How much is the neighbourhood affected by anisotropy? Due to local climatology/topography, is it possible to restrict the spatial area of a certain size to a "corridor", depending on the location?

Seonaid Anderson – Thanks for the question. This is something that we have considered and would be possible. We decided not to look further for now to keep the system as simple as possible. Maintaining some anisotropy/local information is definitely a motivator for using as small a neighborhood as possible!

Raghu Ashrit – What happens when you have a forecast storm outside the locations of climatological probabilities??

Seonaid Anderson – Raghavendra - Thanks for the question. This is a good point, we only consider storms falling within our selected domain where we have climatological probabilities. The domain was selected to cover the historical storms which reach Senegal, but other locations/countries would need a different domain. This would be possible as long as the satellite data were available.

Joel Stein – Neighborhood-based Continuous Ranked Probability Score for Ensemble Predictions

Marion Mittermaier – So for CRPSno the observed pdf is no longer a step/Heaviside function?

Joel Stein – Yes, exactly. Using the neighborhood observation "forgets" which grid point the observation comes from.

Marion – Do you still see that perturbing the ensemble members with obs uncertainty makes sense when CRPSno is used?

Joel Stein – Yes, it still makes sense.

Manfred Dorninger – You have used 1km gridded obs. Can you use this method if you have point observations?

Joel Stein – can apply a mask that corresponds to a list of points, then perform the computation in the same way.

Manfred – wouldn't it be better to instead choose a number of points rather than a size a of neighborhood?

Joel Stein – The size of the neighborhood is what's relevant. Computing CRPSno is computationally expensive when applied to high resolution NWP.

Zied Ben Bouallegue – Is CRPSso still a proper score?

Joel Stein -

Fabien Stoop – Application of neighborhood-based contingency scores to AROME verification

Gunter Mahringer - Experience with PSS – sensitive to rare events. PSS doesn't punish false alarms much for rare events.

Fabien Stoop – Use PSS for all thresholds. For heavy rains HSS may be better adapted.

Gunter Mahringer – Used to combine PSS and HSS to give complementary information.

Fabien Stoop – Need to communicate POD and EDI

Joel Stein – Given a contingency table can do what you want, including SEDI

Barbara Casati – I understand that the neighbourhoods overlap: how do you evaluate the Confidence Intervals in the score, how do you account for the spatial dependence due to the overlapping neighbourhoods? Bootstrap gives Cis that are too small.

Marion Mittermaier - This is no different than for the FSS for example Barbara. We do have an issue with interdependence and overestimation of statistical significance if this spatial correlation is not accounted for. I think thus far this effect has been ignored? Or has someone looked at this?

Fabien Stoop – we have not addressed this aspect for spatial fields

Joel Stein – at Meteo France we only apply bootstrap temporally distinct (e.g. daily) results

Barbara Casati – Can do block bootstrap in space, use the neighborhood size as the block bootstrap size. Larger neighborhoods give fewer samples.

Roger Harbord - It's very difficult to account for spatial dependence in forming CIs, it's easier to base CIs on temporal variation only, i.e. average over space for each timepoint first then use time-series for inference

Marion Mittermaier - I think there are ways of investigating through sampling, i.e. only using non-overlapping neighbourhoods to compute the score.

Raghu Ashrit – In any of the neighborhood Vx methods, as the neighborhood size is increased how to ensure that the obs-fcst pairs stay related and we are not picking unrelated pairs??

Chiara Marsigli - @Raghu: very good question ... I guess there is not a unique answer. One possible criterion could be not to exceed the neighbourhood size at which the forecast stops to be relevant: for the forecasters, the size of the region over which they should issue a warning; for the model developers, the scale of the phenomenon they intend to describe ...

Carlo Cafaro – Do short-range convection-permitting ensembles lead to more skilful probabilistic rainfall forecasts over tropical East Africa?

(no specific questions for this presenter)

General discussion

Barbara Casati – Neighborhood scores are mainly used as headline scores, or are they also used by modellers for diagnostics?

Joel Stein – For modellers neighborhood scores show the benefit of higher resolution model compared to lower resolution.

Marion Mittermaier – It's difficult to assess model upgrades in the presence of the double penalty problem. Are perceived differences in comparisons actually random effects? Spatial methods are part fo the model development process. When we start looking at probabilistic metrics and want to estimate statistical significance, if we don't have obs uncertainty in our measures we might change the direction of the results. Need to be careful (refer Zied's talk).

Chiara Marsigli – Regarding model development, one model did better with model orography. Spatial verification methods helped detect this impact.

Manfred Dorninger – as we go to higher and higher resolutions, is it still useful to use spatial methods, or should we just compare to observations?

Marion – really interesting question. It depends on the observation. Sub-km modelling runs on very small domains with not very many point observations. But gridded observations on sub-km skills aren't available! We ill need to tackle this soon to verify urban modelling. In future will we have to upscale everything to the scale that we can actually observe at?

Manfred – agree, another 10 years of verification science may be needed!

Raghu – Probably going to high resolution is more about resolving the processes rather than capture location-specific information. So upscaling for verification is definitely relevant.

Deryn – Neighborhooding is a form of post-processing. If you have a good post-processed version of two models, maybe these should be compared. Is this a reasonable way to think of the verification process for various scales of models?

Marion – That's kind of what we're doing in the Met Office, wanting to see what value is added by post-processing. Now debating whether one should use neighborhood verification methods to evaluate post-processed forecasts. Probably not, as you're applying the post-processing twice. Better then to verify locally.

Deryn – post-processing can give very well-defined forecast, i.e. optimised for a point. Makes the verification clearer.

Marion – agree.