Section 10

Forecast verification: methods and studies.

Scale-separation diagnostics and the Symmetric Bounded Efficiency for the inter-comparison of gridded products with different spatial resolutions

B. Casati [1] (barbara.casati@ec.gc.ca), C. Lussana [2], A. Crespi [3]

1= Meteorological Research Division, Environment and Climate Change Canada, Dorval (QC), Canada
2 = Division for Climate Services, the Norwegian Meteorological Institute, Oslo, Norway
3 = Center for Climate Change and Transformation, Eurac Research, Bolzano, Italy

Scale-separation methods are a class of spatial verification methods (Gilleland et al., 2010) which enable i) the comparison of the scale structure of gridded products, such as reanalyses, forecasts from numerical models, and gridded observations; ii) the assessment of bias, error and skill on different scales; and iii) the analysis of the scale dependency of forecast predictability. In this study we apply novel scale-separation diagnostics and introduce the Symmetric and Bounded Efficiency for the comparison of precipitation reanalyses with different spatial resolutions. The COSMO-REA6 reanalysis (6km resolution) is compared against the ERA5 reanalysis control member (HRES, 31km resolution) and 10 ensemble members (EDA, 62km resolution). 24-hour accumulated precipitation fields are decomposed into the sum of components on different spatial scales by using a 2D Haar wavelet filter. The separate scale components are then compared by using the continuous verification statistics listed in Table 1.

$En_X^2 = \mu_{X^2} = \sigma_X^2 + \mu_X^2;\qquad En_Y^2 = \mu_{Y^2} = \sigma_Y^2 + \mu_Y^2$	$MSE = (\mu_Y - \mu_X)^2 + \sigma_Y^2 + \sigma_X^2 - 2\sigma_Y\sigma_X r_{Y,X}$
$NB_\mu = \frac{\mu_Y - \mu_X}{\lvert\mu_Y\rvert + \lvert\mu_X\rvert};\qquad NB_\sigma = \frac{\sigma_Y - \sigma_X}{\sigma_Y + \sigma_X}$	$NSE = 1 - \frac{MSE}{\sigma_X^2};\qquad SS_{rand} = 1 - \frac{MSE}{(\mu_Y - \mu_X)^2 + \sigma_Y^2 + \sigma_X^2}$
$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_Y}{\sigma_X} - 1\right)^2 + \left(\frac{\mu_Y}{\mu_X} - 1\right)^2}$	$SBE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_Y - \sigma_X}{\sigma_Y + \sigma_X}\right)^2 + \left(\frac{\mu_Y - \mu_X}{\lvert\mu_Y\rvert + \lvert\mu_X\rvert}\right)^2}$

Table 1: scale-separation diagnostics, where *X* and *Y* indicate the gridded observation and forecast compared, μ the mean, σ the standard deviation, *r* the correlation, following conventional statistical notation.

The scale-separation diagnostics are illustrated in Figure 1. The energy (En^2) and its square-root (En, Fig.1a) are proportional to magnitude and number of precipitation features on each scale. Comparison of the energies informs on biases on different scales and on the scale structure. All products show that the dominant precipitation features have scales ranging between 200-400km. As expected, the COSMO-REA6 (in grey) has larger energy, much more small-to-medium scale features, than the ERA5 products. The ERA5 control member (in red) has slightly more small-scale details than the lower resolution ensemble members (in blue). The energies on the different scales are compared by the Normalized Bias (NB), which is the difference between the two energies divided by their sum. For the wavelet components (which have null spatial average $\mu_X = 0$) the energy En_X^2 is the variance σ_X^2, whereas for the largest scale component (which is a constant field with $\sigma_X^2 = 0$) the En_X is the field spatial mean μ_X. Hence, the NB comparing square-root energies is the NB comparing the scale component standard deviations and the field spatial averages (Table 1). Normalized (and bounded) statistics facilitate the comparison and aggregation of verification results, e.g. for sites with different climatologies. The square-root energy NB is bounded (it ranges between -1 and 1) and symmetric (its absolute value does not change if we swap the two products compared): these two key properties enable the definition of the Symmetric and Bounded Efficiency.

While assessing the Mean Squared Error (MSE) and correlation on different scales (Fig.1b,c), the correlation separates the performance of the ERA5 products, whereas the MSE does not. In fact, the MSE highly depends on the variability of the products compared, both observation and forecast variabilities (Table 1), so that higher resolution forecasts (which have higher variability) tend to be penalized more heavily with respect to coarser/smoother forecasts. The reduction of variance, also known as Nash Sutcliffe Efficiency (NSE), is the MSE skill score against sample climatology ($MSE_{clim} = \sigma_X^2$), and it was introduced to reduce the effect of the variability while assessing the forecast performance. However, the NSE compensates solely for the observed variability (the MSE is normalized by the observed variance only). Then, the MSE shortcoming when comparing forecasts with different resolutions is still not addressed by the NSE: the higher resolution/noisier forecast is still penalized more heavily compared to the coarser/smoother one, and also the NSE cannot separate the performance of the two ERA5 products (Fig.1d). The MSE skill score against random chance (SS_{rand}, Table 1) normalizes the MSE by the variances of both observation and forecast products (the forecast variability is factored in), and is therefore capable of separating the performance of the ERA5 products (Fig.1e). The SS_{rand} allows a fairer comparison of products with different resolutions and enables the assessment of the added value of increasing resolution.

The Kling and Gupta Efficiency (KGE) combines in a single performance measure the comparison of forecast and observation variances and averages, and the two products correlation. The KGE also allows a fair comparison of forecasts with different resolutions, and is capable of separating the performance of the ERA5 products (Fig.8 of Casati et al., 2023). However, the KGE compares variances and averages by their ratio, which renders the KGE asymmetric and unbounded (the score can attain very large negative values e.g. when comparing a smooth forecast versus a noisy observation). The Symmetric and Bounded Efficiency (SBE) is then defined as the KGE, but the variances and averages are compared by the NB (Table 1). The SBE is then bounded and capable of separating the performance of the ERA5 products (Fig.1f). Both SS_{rand} and SBE are symmetric, hence they are invariant with-respect-to the definition of observation and forecast in the comparison of the two products, whereas NSE and KGE are not symmetric, and their values (as well as the no-skill to skill transition scale) change if the order of comparison of the two products is swapped.

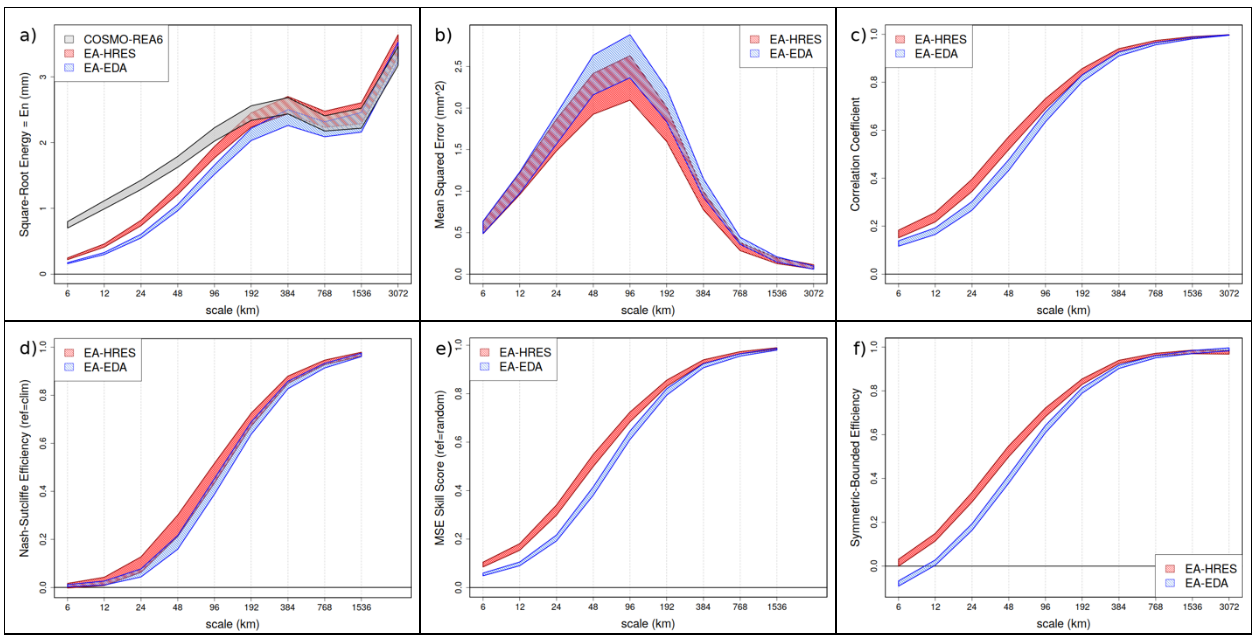

Figure 1: scale-separation diagnostics comparing the ERA5 products against COSMO-REA6. The shading shows the bootstrap 90% confidence intervals of the aggregated statistics on 50 intense precipitation case studies.

The scale-separation diagnostics in Figure 1 consistently show that the reanalyses are in strong agreement on large scales, and exhibit weaker agreement on (less predictable) small scales. The stronger agreement of HRES with COSMO-REA6 (compared to EDA) is due to a more similar representation of the variability at small-to-medium scales, as well as a better linear dependence (correlation). On the largest scale, on the other hand, HRES is slightly underperforming EDA due to over-forecast bias.

References

Casati, B., C. Lussana, and A. Crespi (2023) Scale-separation diagnostics and the Symmetric Bounded Efficiency for the inter-comparison of precipitation reanalyses. International Journal of Climatology, 43, 2287-2304.

Gilleland, E., Ahijevych, D. A., Brown, B. G. and Ebert, E. E. (2010) Verifying forecasts spatially. Bulletin of the American Meteorological Society, 91, 1365–1376.

Methods for Evaluating High Impact Hydrometeorological Features using METplus

Tracy Hertneky[1], Tara Jensen[1], and Michael Erickson[2]
[1]National Center for Atmospheric Research and Developmental Testbed Center
[2]Previously Cooperative Institute for Research in Environmental Sciences and NOAA Weather Prediction Center
Correspondence: hertneky@ucar.edu

Introduction

Evaluation of high impact hydrometeorological features in numerical weather prediction is challenging, but critical to understand due to the potential hazards they present. Features with strong gradients over a short distance, such as narrow snow bands, are more susceptible to the ‘double penalty’ when using traditional verification metrics due to displacement and size errors. This paper will demonstrate innovative tools for evaluating snowband features using object-based methods.

Software and Data

The enhanced Model Evaluation Tools (METplus) is a state-of the-art verification and diagnostics framework that hosts a suite of traditional statistics and diagnostic methods for applications over a wide range of temporal and spatial scales (Brown et al. 2021). Of particular use in assessing snowband events, is the Method for Object-based Diagnostic Evaluation (MODE) tool in METplus, which is used to identify and compare coherent spatial features (Davis et al. 2006). Seven cases of heavy-banded snowfall events over the Northeast, often associated with Nor’Easters, were selected and include initializations on 12/16/2020, 12/24/2020, 1/31/2021, 2/1/2021, 1/3/2022, 1/6/2022, and 1/16/2022. HRRR 1-hr forecast data was verified against the Multi Radar Multi Sensor (MRMS) product and the HRRR analysis. A masking region over the eastern quarter of the CONUS was applied in order to eliminate other snow events that may behave differently compared to the events of interest.

METplus Methods

Three novel METplus tools that enable object-based verification were used to evaluate snowband features including i) feature relative to diagnose systematic biases in the environment relative to the feature, ii) forecast consistency to provide a measure of forecast stability across cycles, and iii) multivariate MODE to identify and evaluate complex super objects from two or more input fields.

a. Feature Relative

The feature relative use-case utilized the seven cases initialized at 00 UTC to demonstrate the identification of systematic biases within the environment relative to the snowband features. The HRRR gridded forecast data was compared against the gridded MRMS accumulated precipitation and HRRR analysis for investigating various mechanisms that drive snowband behavior. This use-case was set up to use the following METplus tools, i) GenVxMask to mask the accumulated precipitation field using categorical snow and apply the east quarter mask, ii) MODE time domain (MTD) to identify and track the snowband events in time, iii) ExtractTiles to extract a tile relative to the snowband centroid, and iv) series analysis to accumulate statistics at each grid point separately within the extracted tile. Relevant configuration settings in MTD include a convolution threshold ≥0.05 inches precipitation, a convolution radius of 5 grid points (or 15 km), and a minimum volume of 1000 grid points. For ExtractTiles, a 30° x 30° tile with 0.25° grid spacing centered on the objects was extracted. Spatial mean error plots from series analysis are shown in figure 1. The aggregated biases are 2-3 mm in the 1-hour precipitation accumulations (fig. 1a), with a high

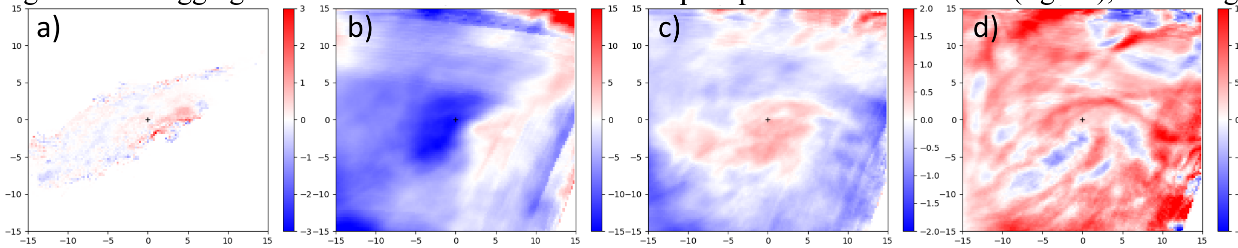

Figure 1. Spatial mean error plots from series analysis of a) 1h accumulated precipitation (mm), and b) geopotential height (m), temperature (K), and relative humidity (%) at 700 mb. The '+' indicates the feature centroid.

snowfall bias near the center and in the southeast quadrant. Geopotential heights in the mid-lower troposphere, such as at 700 mb (fig. 1b) and mean sea level pressure (not shown) show a low bias over the snowband centroid and downstream, which indicates the model low is too strong and too slow for these events. At 700 mb, the environment is too cold to the north and too warm to the south of the centroid (fig. 1c) and generally too moist (fig. 1d).

b. Forecast Consistency

For forecast consistency, the HRRR 1-hr forecasts were used for the 16 December 2020 snowband case. 1-hr cycles from 12/16/2020 12 UTC to 12/17/2020 06 UTC were used for measuring forecast stability as the event neared. This use-case first runs GenVxMask to mask the data using categorical snow and then applies the east quarter mask to focus on snowbands in the northeast. MTD is then run in reverse, starting with the longest lead time to the shortest, while keeping valid hour constant and thresholding accumulated precipitation ≥0.05” and a minimum volume of 1000 grid points. The difference from one time to the next, or the revisions, can then be computed for various object attributes, such as area, intensity, or displacement. The revisions of object area at valid time 12/17/2020 06 UTC are plotted in figure 2, showing small changes, generally less than 600 grid points, throughout the series. Considering the large size of the object (not shown), these revisions are <1% of the object area so the object size is not changing drastically from one time to another, showing consistency in the forecast.

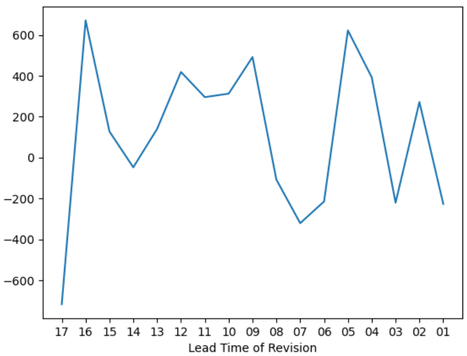

Figure 2. Time series at valid hour 12/17/2020 06 UTC of snowband object area revisions in grid points.

c. Multivariate MODE

The multivariate MODE use-case was applied to the 1 February 2021 snowband case, ingesting categorical snow and accumulated precipitation from the HRRR 1-h forecasts and MRMS. To identify and evaluate super objects in METplus, the use-case first runs multivariate MODE to identify the super objects where accumulated precipitation is identified as snow type. Then GenVxMask is used to apply the super object mask to the raw precipitation field, and finally MODE is run a second time on the masked field to provide attribute statistics. Figure 3 shows time series of object attributes, where it is observed that the forecast super objects are larger (fig. 3a), often more intense with respect to accumulated precipitation (fig. 3b), and have a north-east displacement (fig. 3c) compared to observations.

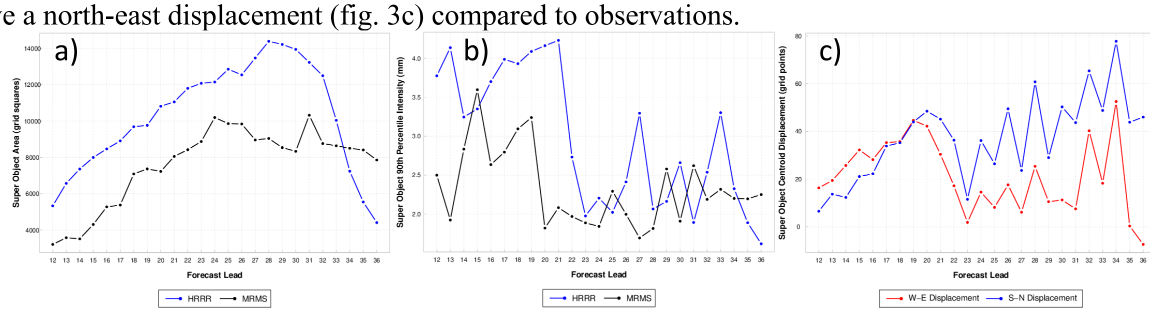

Figure 3. Output from running the multivariate MODE use-case showing time series plots of a) object area and b) 90 percentile accumulated precipitation intensity for the HRRR forecast (blue) and MRMS (black), and centroid displacement in the W-E (red) and S-N (blue) direction, where positive values indicate a E/N displacement and negative values indicate a W/S displacement.

References

Brown, B. G., T. G. Jensen, J. Halley Gotway, R. Bullock, E. Gilleland, T. Fowler, K. Newman, D. Adriaansen, L. Blank, T. Burek, M. Harrold, T. Hertneky, C. Kalb, P. Kucera, L. Nance, and J. Wolff, 2021. The Model Evaluation Tools (MET): More than a decade of community-supported forecast verification. Bull. Amer. Meteorol. Soc., 102 (4), E782 - E807, doi: 10.1175/BAMS-D-19-0093.1.

Davis, C.A., B.G. Brown, and R.G. Bullock, 2006a: Object-based verification of precipitation forecasts, Part I: Methodology and application to mesoscale rain areas. *Monthly Weather Review*, 134, 1772-1784.

Davis, C.A., B.G. Brown, and R.G. Bullock, 2006b: Object-based verification of precipitation forecasts, Part II: Application to convective rain systems. *Monthly Weather Review*, 134, 1785-1795.

METplus Verification and Diagnostics Framework for Model Evaluation Across Scales

Tara Jensen[1,2], John Opatz[1,2], Christina Kalb[1,2], Daniel Adriaansen[1,2], Kathryn Newman[1,2], Michelle Harrold[1,2], Mrinal Biswas[1,2], Tracy Hertneky[1,2], Will Mayfield[1,2], Weiwei Li[1,2], Brianne Nelson[1,2], Jonathan Vigh[1,2], Molly Smith[3,4], Jason English[3,4], Louisa Nance[1,2], Barbara Brown[1], Michael Ek[1,2]

[1]National Center for Atmospheric Research, [2]Developmental Testbed Center, [3]Cooperative Institute for Research in Environmental Sciences, [4]NOAA Global Systems Laboratory
Corresponding Author: jensen@ucar.edu

Introduction

The METplus system is a suite of verification and diagnostic tools in a consistent framework that are designed to facilitate consistent computation of statistics and metrics across applications and institutions. The highly configurable Python wrappers provide low-level workflow around the core Model Evaluation Tools (MET) package for computing verification statistics. Additional components of METplus include an Analysis Suite made up of a data input-output library (METdataio), an aggregation and synthesis tool (METcalcpy), and plotting (METplotpy). The tools are more fully discussed in Brown et. al (2021) and provide great flexibility to evaluate a range of numerical prediction across temporal scales, spatial scales, and model applications. To address the needs of a cadre of international research laboratories and operational centers, METplus is also being enhanced to provide systematic evaluation and diagnostics of many of the coupled components within an Earth system modeling framework.

Temporal and Spatial Scales

METplus was designed to work like most other Linux/Unix based tools, with each component being focused on a small subset of capability to allow for maximum flexibility in setting up the tools. Unless specifically designed to accept a time series of data, the tools focus on computing statistics for a give valid time and allow for aggregation of statistics, diagnostic attributes, or metrics, over the appropriate temporal scales for a model application. This allows the same tools to be used for evaluation of short-range (1-10 minute, hourly, daily), medium-range (3-14 days), sub-seasonal (weeks 2-4), seasonal (up to 9 months), and climate (yearly to multi-decadal) simulations. The capability includes the use of appropriate user-defined climatologies, thresholds, masking regions to define areas of interest, and interpolation methods.

Applications

The applications METplus has been applied to vary from evaluation forecasts for renewable energy, fire weather, severe weather, extremes, marine (standard variables as well as phenomena like chlorophyll), cryosphere, extra-tropical and tropical cyclones, monsoons, droughts, clouds, dust, aerosols, air quality fields, satellite brightness temperatures, land modeling diagnostics, and ionospheric fields for space weather. The tools are used in both research and operations. Current operational partners include the National Oceanic and Atmospheric Administration (NOAA) National Center for Environmental Prediction (NCEP) Centers, the United States Air Force, the Met Office for the United Kingdom, the Australian Bureau of Meteorology, the Indian National Center for Medium Range Weather Forecasting, and many of the other associated Unified Model partnership. Additionally, METplus is being adopted by US Naval Research Laboratory and US Army Research Laboratory, and the United Arab Emirates National Center for Meteorology.

Current and Future Capability

In METplus v5.1, released June 2023, there is support for over 150 traditional statistics and diagnostics methods. Table 1 provides a synopsis of what is included in the METplus framework. Future work is focused on optimizing the use of memory, adding support for parallelized computing and unstructured grids, developing database applications to store large quantities of forecast-observation matched pairs for analysis, and including more diagnostics for non-atmospheric components of an Earth system model.

Traditional	
Grid-Stat, Point-Stat, Series-Analysis Contingency table (CTS), Continuous, Probability forecast statistics, SEEPS	**Ensemble-Stat** CRPS, CRPSS, Rank probability, Prob. Integral Transform, and Relative position histograms, Spread/Skill, Ignorance
Spatial	
MODE Location and Geometric attribute differences, Intersection area, Intensity distributions, CTS measures for objects, ***Available for single and multi-variate fields***	**MODE-Time Domain** Time and location differences, Volume and Velocity differences, Intersection volume, Intensity distributions and differences
Wavelet-Stat Mean Square Error by scale, Energy by scale, Intensity-scale skill score	**Grid-Stat and Point-Stat** Fraction Skill Score, High Resolution Analysis, Distance Measures, Mean Error Distance, Baddeley, Hausdorff, Zhu, Fourier Decomposition
Diagnostics	
Grid-Diag Distribution of fields to assess multivariate relationships between fields	**Feature Relative** Used to assess presence of systematic errors associated with features or events
Physics Tendencies Computation of Physics tendencies (vertical cross section and plan-view)	
Tropical Cyclones Application	
TC-Pairs, TC-Stat, TC-Dland Track error (along, cross, total), Intensity (pressure, wind), Rapid Intensification/weakening errors, CTS measures for TC genesis	**TC-Gen** CTS measures for TC genesis, Spatial representation of TC genesis density function, TC density function
TC-RMW Errors and diagnostics in Radius of maximum wind projection	**TC-Diag** Errors and diagnostics in
S2S Application	
Realtime Multivariate Madden Julien Oscillation (MJO) Index (RMM) 1, RMM2, Outgoing Longwave Radiation (OLR) MJO Index, MJO-El Nino Southern Oscillation (ENSO) Index	
Identification of Weather Regimes and Blocking Regimes, Hovmoeller Diagrams, Zonal and Meridional Means, Empirical Orthogonal Functions (EOFs), Space-Time Coherence (or Cross-Spectra) Plots	
Statistical Synthesis Tools	
Scorecards, Contour Plots	Performance Diagrams, Taylor Diagrams

Table 1. High level summary of traditional statistics, spatial methods and model diagnostics, and Tropical Cyclone application support. MET Tool names are in blue with capability listed in black.

Summary

METplus is a state-of-the-science community verification and diagnostic package that is used by over 3,000 US and international institutions spanning the public, private, and academic sectors. It provides the Earth system modeling research community with the ability to share, and hence, standardize evaluation across entities. More information can be found at: dtcenter.org/community-code/metplus

References

Brown, B., T. G. Jensen, et al., **2021**: The Model Evaluation Tools (MET): More than a decade of community-supported forecast verification. *Bull. Amer. Meteor. Soc.*, E782-E807. DOI: https://doi-org.cuucar.idm.oclc.org/10.1175/BAMS-D-19-0093.1

Advances in METplus Verification for Subseasonal to Seasonal Model Evaluation

Christina Kalb[1], Douglas Miller[2], Maria Gehne[3,4], Zhou Wang[5], Minna Win-Gildenmeister[1], George McCabe[1], Hank Fisher[1], Tara Jensen[1], Weiwei Li[1]

[1]National Center for Atmospheric Research, [2]RWE Supply and Trading, LLC, [3]Cooperative Institute for Research in Environmental Sciences, [4]National Oceanic and Atmospheric Adminstration/Physical Sciences Laboratory, [5]University of Illinois at Urbana-Champaign

Email: kalb@ucar.edu

Introduction

The METplus system is a suite of verification and diagnostic tools in a consistent framework that are designed to facilitate quick setup. METplus contains several components, but at its core is the Model Evaluation Tools (MET) package for computing verification statistics. Additional components of METplus include METcalcpy, which contains python calculations of statistics and other indices and metrics, and METplotpy for creating graphics and displaying statistics. Recently, a set of process-oriented diagnostic and verification metrics have been added to the METplus system to examine the predictability of phenomena on subseasonal to seasonal time scales. The METplus GitHub repository (https://github.com/dtcenter/METplus) includes use cases, where users can change configuration settings to run the computations on different datasets. Here we will focus on four of these metrics, which fall into two categories, mid latitude weather and indices associated with the Madden Julian Oscillation (MJO).

Mid Latitude Metrics

There are two calculations added to METplus to evaluate different mid latitude weather events, atmospheric blocking and weather regimes. The weather regime classification begins by computing the optimal number of regimes using the sum of squared distances. From there, weather regime patterns and their frequency of occurrence are computed using K-means clustering. This clustering can be performed on the 500 hPa height data, or on reconstructed empirical orthogonal functions. Additionally, the frequency of occurrence of each classified weather regime can be computed over a user defined time period. Figure 1 shows the patterns of the first two weather regimes for the European Center for Medium Range Forecast Re-Analysis (ERA - top) and Global Forecast System (GFS - bottom). In total, there were six classified weather regimes. Output statistics can be performed on the classification and time frequency of regimes. Multi-category contingency table statistics computed on these classified weather regimes include a Heidke Skill Score (HSS) of 0.593, and a Hanssen-Kuipers Discriminant (HK) of 0.598.

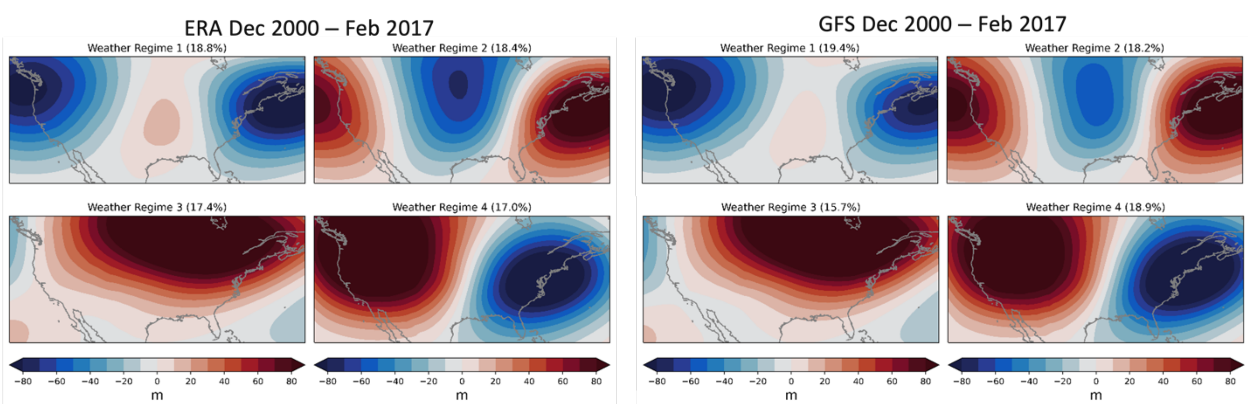

Figure 1. The first four weather regime patterns classified for the ERA (left) and GFS model (right). The frequency of each weather regime is given in parentheses.

The second mid latitude calculation, atmospheric blocking, uses approaches in both Pelly and Hoskins and Tibaldi and Molteni methods (Barnes et al, 2012). Specifically, atmospheric blocking events are identified by first locating reversals in the 500 hPa geopotential height gradient as blocked longitudes. This is followed by applying spatial and temporal thresholds to ensure the large-scale, quasi-stationary characteristics of blocking anticyclones are met. Some of the default characteristics are that a block must persist for at least 5 days and not travel more than 45 degrees downstream, although these options can be modified (Miller and Wang, 2019 and 2022).

MJO Indices

The two new MJO metrics that have been added to METplus include the Real-Time Multivariate MJO (RMM) Index and OLR-based MJO Index (OMI). Similar to the mid latitude calculations, these use ERA and GFS for the observations and model, however only graphics are output. The RMM is computed using latitudinal averages of outgoing longwave radiation (OLR), 850 hPa zonal winds, and 200 hPa zonal winds. The calculation follows Wheeler and Hendon 2004, and includes removal of the 120 day mean, regressing data onto the EOF patterns, and normalizing the principal components by the standard deviation. Figure 2 shows the phase diagram, RMM 1 and RMM 2 for 2022, and EOF1 for each variable from the RMM calculation.

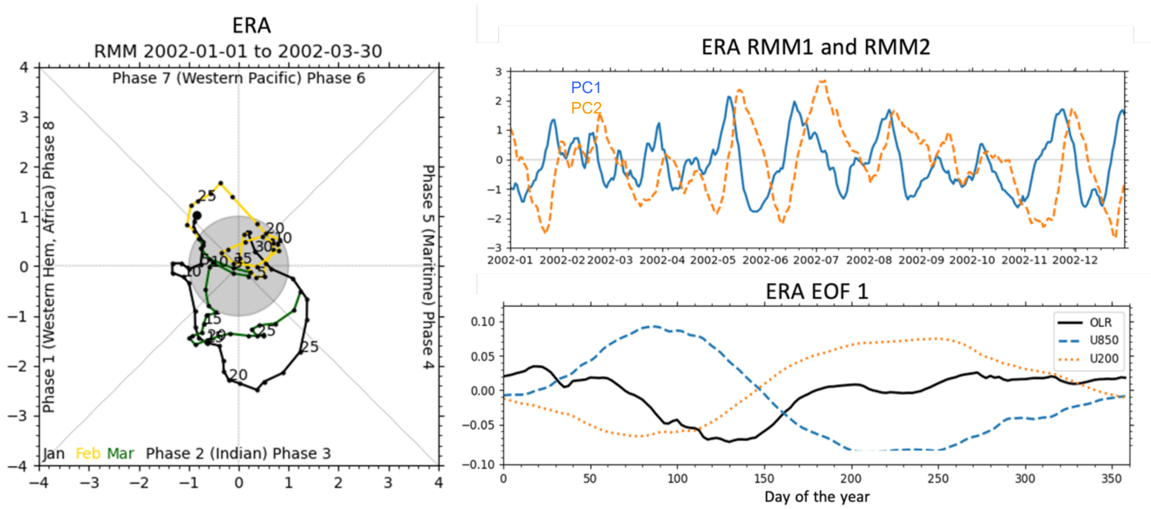

Figure 2. A phase diagram (left), RMM principal components 1 and 2 over 2022 (right, top), and the EOFs for all three variables (right, bottom).

The OMI computation is similar to RMM, but uses only outgoing longwave radiation. The first step is filtering to retain the frequencies associated with the MJO, which is set to be 29 – 90 days. Then, the OLR is projected on to the daily EOF patterns, and the principal components are normalized. Output from the OMI use case includes a phase diagram for both the model and observations which are like the one seen in Figure 2.

Summary

Many new process-oriented metrics have been added to the METplus system to examine the predictability of phenomena on subseasonal to seasonal time scales. These metrics include calculations of indices, output statistics, and graphics. The four metrics discussed here, along with all others, are designed to be flexible enough to use with different model and observation datasets, as well as user configurable options.

References

Barnes, E. A. ,J. Slingo, and T. Woollings, 2012: A methodology for the comparison of blocking climatologies across indices, models and climate scenarios. *Clim. Dyn.*, **38**: 2467–2481, 2012. doi: 10.1007/s00382-011-1243-6.

Miller, D. E., & Wang, Z., 2019: Skillful seasonal prediction of Eurasian winter blocking and extreme temperature frequency. Geophysical Research Letters, 46(20), 11530-11538.

Miller, D. E., & Wang, Z., 2022: Northern Hemisphere winter blocking: differing onset mechanisms across regions. Journal of the Atmospheric Sciences, 79(5), 1291-1309.

Wheeler, M. C., and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132**, 1917–1932, doi:10.1175/1520-0493(2004)132<1917:AARMMI >2.0.CO;2.

Using METplus for verification of COSMO-Ru/ICON modelling system

Alexander Kirsanov[1,2], Anastasia Bundel[1], Maria Tarasova[1,3,4], Elena Astakhova[1], Julia Shuvalova[1]

[1]Hydrometcentre of Russia, [2]*Obukhov Institute of Atmospheric Physics, Russian Academy of Sciences,* [3]Lomonosov Moscow State University, Department of Geography,[4]Research Computing Center, Lomonosov Moscow State University

METplus is a forecast verification system [McCabe, G. et al. 2022] with MET (Model Evaluation Tools) [Newman et al. 2022] as a core developed and supported to community via the Developmental Testbed Center (DTC) (https://dtcenter.org/community-code/metplus). It was decided to implement METplus at the Hydrometcentre of Russia because of its flexibility, availability of most necessary methods in one package, and a good support by developers via the forum. METplus with the tools for visualization of the forecast scores has been installed at the Hydrometcentre of Russia and applied to verify mesoscale high-resolution forecasts by the COSMO-Ru system [Rivin et al., 2015]. Figure 1 shows the operational COSMO-Ru model domains, for which the scores are calculated at present. The scores are also calculated in the experimental mode for ICON-Ru system. The plots are prepared in the METviewer package. It proved to be a convenient visualization tool. The synoptic station data are used as a reference.

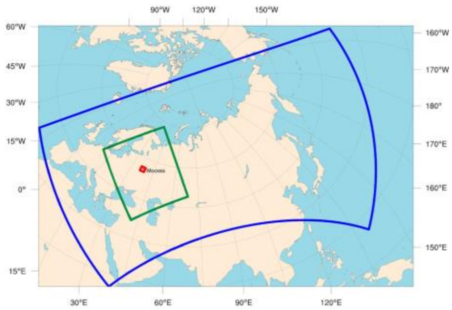

Figure 1. Configuration of the operational COSMO-Ru system [Rivin et al. 2015]. Blue is COSMO-Ru6ENA domain (grid step in horizontal 6.6 km), green is COSMO-Ru2By domain (grid step in horizontal 2.2 km), red is COSMO-Ru1Mr (grid step in horizontal 1 km)

Figure 2 shows how the diurnal cycle of the 2m temperature (T2m) mean error (ME) changes depending on the start time of the COSMO-Ru2By model with a 2.2 km grid step (green contour in Fig. 1). The score is aggregated over European Russia (the whole model domain) and over all lead times up to 24 h. Figure 2 demonstrates that the diurnal amplitude of the T2m changes is underestimated except for 18 UTC run, which overestimates T2m during almost the whole day. The 00 UTC runs have the smallest bias except for the evening hours.

Figure 3 shows the box plots of the wind gusts in COSMO-Ru2By and observations together with the ME and MAE of the wind gusts. It is a useful approach for comparing the distribution of the variable values and the errors. The model and observed values of wind gusts are in good agreement, and the ME is close to zero, although there are some outliers.

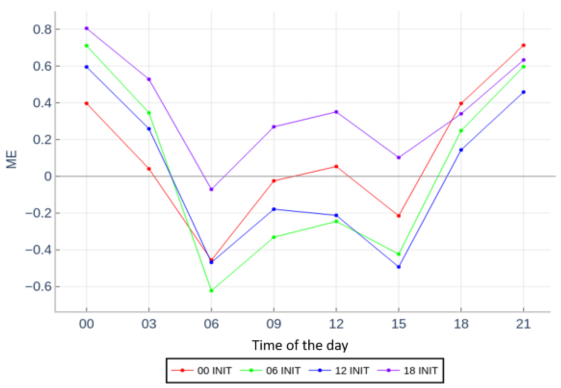

Figure 2. The mean error ME of the air temperature at 2 m (°C) for 00, 06, 12, and 18 UTC runs, COSMO-Ru2By, European Russia, spring 2023.

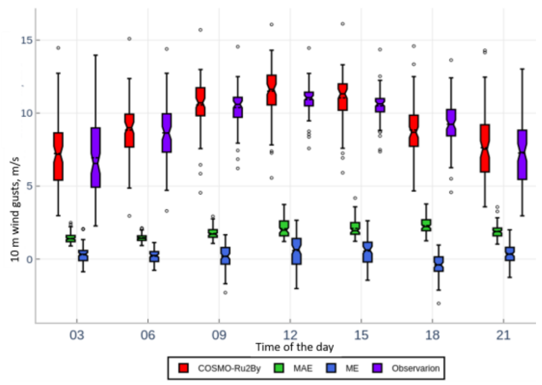

Figure 3. Forecast and observed wind gusts at 10 m (m/s), COSMO-Ru2By, European Russia, MAM 2023, 00 UTC run.

In Figure 4, precipitation performance diagrams are shown for summer 2022 (Fig. 4a) and spring 2023 (Fig. 4b). In summer 2022, COSMO-Ru6ENA overestimated precipitation exceeding low thresholds (0.1 and 1 mm/12h) and underestimated intense precipitation. Overall, the intense precipitation is forecasted worse than precipitation exceeding lower thresholds. There are similar conclusions for COSMO-Ru2By, spring 2023, but the precipitation scores are better compared to COSMO-Ru6ENA, summer 2022.

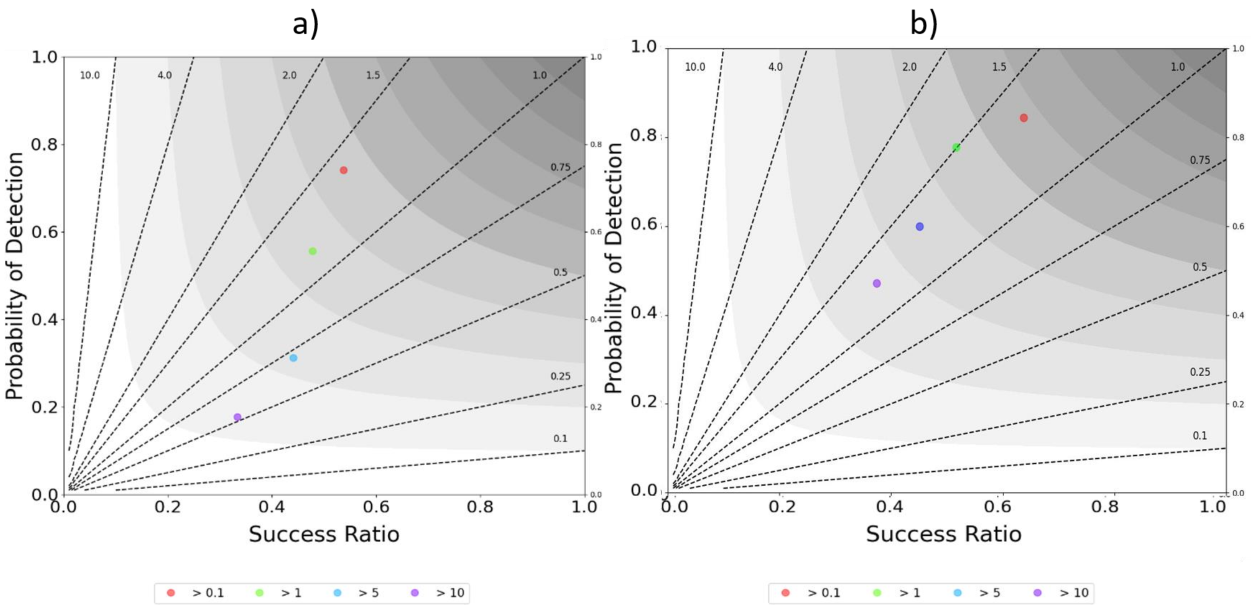

Figure 4. Performance diagrams of 12h precipitation accumulations exceeding different thresholds (0.1, 1, 5, and 10 mm/12h), Central Russia, 15 h lead time, 00 UTC run, (a) COSMO-Ru6ENA, JJA 2022 and (b) COSMO-Ru2By, MAM 2023

METplus is run to verify the test regional ensemble ICON-Ru2-EPS forecast based on the ICON-LAM model (2.2 km horizontal grid step) [Astakhova et al. 2021]. We also use METplus for neighborhood methods, object-based MODE method [Bundel et al. 2022], and for verifying the microphysical fields, such as the liquid water path using the MODIS satellite product as reference.

Conclusions: METplus was chosen as a basic verification tool for the COSMO-Ru system at the Hydrometcenter of Russia. It proved to be a convenient and versatile tool and helped to identify several model flaws. We plan to further expand the range of verification methods in operation, including the spatial ones applied to high-resolution ensembles. The research applications explore using the non-standard fields for both model and observations.

References

1. Astakhova Elena, Dmitry Alferov, Yuri Alferov and Anastasia Bundel, Ensemble approach to weather forecasting. 2021 J. Phys.: Conf. Ser. 1740 012070 doi:10.1088/1742-6596/1740/1/012070

2. Bundel Anastasia, Elena Astakhova, Elizaveta Olkhovaya, Alexander Kirsanov and Dmitry Alferov, Spatial verification of a regional ensemble precipitation forecasting system at the Hydrometeorological Research Center of the Russian Federation using a free verification package, MET // 2022 IOP Conf. Ser.: Earth Environ. Sci. 1023 012001 DOI 10.1088/1755-1315/1023/1/012001

3. McCabe, G., J. Prestopnik, J. Opatz, J. Halley Gotway, T. Jensen, J. Vigh, M. Row, C. Kalb, H. Fisher, L. Goodrich, D. Adriaansen, M. Win-Gildenmeister, J. Frimel, L. Blank, T. Arbetter, 2022: The METplus Version 4.1.4 User's Guide. Developmental Testbed Center. Available at: https://github.com/dtcenter/METplus/releases.

4. Newman, K., J. Opatz, T. Jensen, J. Prestopnik, H. Soh, L. Goodrich, B. Brown, R. Bullock, J. Halley Gotway, 2022: The MET Version 10.1.2 User's Guide. Developmental Testbed Center. Available at: https://github.com/dtcenter/MET/releases

5. Rivin, G.S., Rozinkina, I.A., Vil'fand, R.M. et al. The COSMO-Ru system of nonhydrostatic mesoscale short-range weather forecasting of the Hydrometcenter of Russia: The second stage of implementation and development. Russ. Meteorol. Hydrol. 40, 400–410 (2015). https://doi.org/10.3103/S1068373915060060