

GEOVISUAL ANALYTICS ON THE VERIFICATION OF THE PAGASA OPERATIONAL NUMERICAL WEATHER PREDICTION MODEL RAINFALL FORECAST

R. A. A. Flores¹

¹ RDTD, Philippine Atmospheric, Geophysical, Astronomical Services Administration, BIR Road, Quezon City, Philippines -
renflores@pagasa.dost.gov.ph

Commission IV

KEY WORDS: Geovisual Analytics, Numerical Weather Prediction, Forecast Verification, WRF-ARW, Tropical Meteorology

ABSTRACT:

Assessment of NWP model performance is an integral part of operational forecasting as well as in research and development. Understanding the bias propagation of an NWP model and how it propagates across space can provide more insight in determining underlying causes and weaknesses not easily determined in traditional methods. The study aims to introduce the integration of the spatial distribution of error in interpreting model verification results by assessing how well the operational numerical weather prediction system of PAGASA captures the country's weather pattern in each of its climate type. It also discusses improvements in model performance throughout the time-frame of analysis. Error propagation patterns were identified using Geovisual Analytics to allow comparison of verification scores among individual stations. The study concluded that a major update in the physics parameterization of the model in 2016 and continued minor updates in the following years, surface precipitation forecasts greatly improved from an average RMSE of 9.3, MAE of 3.2 and Bias of 1.36 in 2015 to an RMSE of 7.9, MAE of 2.5 and bias of -0.63 in 2018.

1. INTRODUCTION

The use of numerical weather prediction (NWP) model is one of the most relevant components of routine and severe weather forecasting and warnings at PAGASA. The WRF-ARW, the operational NWP model of the agency has been running since 2012. It provides forecast up to seven days, covering two domains, the one centered on Philippine Island having a 3x3 horizontal resolution that is mainly used for research and the other which covers the Philippine Area of Responsibility (PAR) with a 12 x 12 km resolution extensively used for operational forecasting. With recent upgrades in physics parameterization, its performance must be evaluated.

Operational usage of NWP entails understanding the model's behaviour and errors. Such errors in forecast parameters like precipitation and temperature are characterized by their biases that may be present for certain periods of the day or season. These characteristics, if known by the forecasters, will allow them to adjust accordingly and create more insightful analysis of the NWP outputs.

Currently, verification analysis on WRF-ARW is centered on determining errors arising from the diurnal cycle, seasonal changes, and parameterization scheme (Sun, Bi, 2019). The proposed study will provide statistical analytics, using forecast verification, on the performance and behaviour of the operational WRF-ARW to provide more contexts in interpreting the model output. Specifically it will look into the geographical distribution of errors using point analysis and spatial visual analytics. Through these methods, it can be determined if local topography and climate type have substantial impacts on the forecast quality. Moreover, spatially analysing verification statistics will help determine any geographical variability of the model error and its error distribution pattern. It will also aid in determining local processes that are overlooked or unresolved in the model.

2. DATA AND METHODOLOGY

2.1 Study Area

The study area is based on the operational WRF domain which encompasses the Philippine Area of Responsibility (PAR); the basis of the Operational WRF domain. PAR is the area defined by the World Meteorological Organization (WMO) as a part wherein the National Weather Service shall undertake operational activities pertaining to the observation, moderation, modification, reporting of the weather and issue forecasts and warnings of weather and flood conditions affecting national safety, welfare and economy. (P.D No. 78). The PAR is bounded by imaginary lines on the surface of the earth that makes equal oblique angles with all meridians joining the following points: 5°N 120°E, 25°N 135°E, 5°N 135°E, 5°N 115°E, 15°N 115°E, 21°N 120°E and back to the beginning. Since WRF only accepts rectangular grids, the area of the operational domain is bounded by 25°N 114°E, 25°N 135°E, 2°N 114°E, 2°N 135°E. The area covers almost all of the land territory of the Philippines, except for the southernmost portions of the province of Tawi-Tawi, and some of the country's claimed islands in the Spratlys. The area also includes the main island of Palau, most of Taiwan, as well as portions of the Malaysian state of Sabah.

2.2 Observed Rainfall Data

The observed surface precipitation was retrieved from the archives of the Climate Data Section, PAGASA that were collected from 52 synoptic stations. The data consists of accumulated rainfall for an interval of 3 hours, 6 hours and 24 hours. The study utilized the 6-hourly accumulation for a more complete data set; the 3-hourly and 24-hourly were also used to interpolate missing data. The retrieved data has undergone quality check from the Climate Data Section and was measured by standardized and calibrated tipping buckets issued by PAGASA.

2.3 Forecasted Rainfall Data

The operational WRF model is managed by the Numerical Modelling Section of PAGASA. The model is run at the Integrated High Performance Computing System (IHPC) facility of the agency. It runs every three hours beginning at 0000 UTC (08:00 AM PST). The model inputs include topography and land use classification data from the United States Geological Survey (USGS). For the initial boundary conditions, inputs are Global Forecasting System (GFS) forecast data from the latest initialization; GFS model runs every six hours beginning at 0000 UTC (08:00 AM PST). Another model input is the Sea Surface temperature data sourced from National Centers for Environmental Prediction (NCEP). The model has two domains. Domain one has a spatial resolution of 12 KM and can forecast up to 144 hours lead time. This domain is used for operations such as forecasting and product derivation mainly because of the long range of lead time. Domain two has a spatial resolution of 3 KM and can forecast up to 48 hours lead time. Domain two is mostly used for research purposes such as event analysis, test case analysis, and model performance comparison.

The study makes use of WRF post processed data from all initializations from the years 2015 to September 2018; all are available in standard Grib2 format. Data for the months of January to March of 2017 were not available due to model malfunction. The surface precipitation forecast for each synoptic station was then extracted and stored as CSV files. The corresponding grids of the 52 stations were obtained from the WRF output using the Grid Analysis and Display Software (GrADS) nearest-neighbour algorithm. After extraction, the surface precipitation forecast was accumulated to 6-hourly to match the time interval of the observed data.

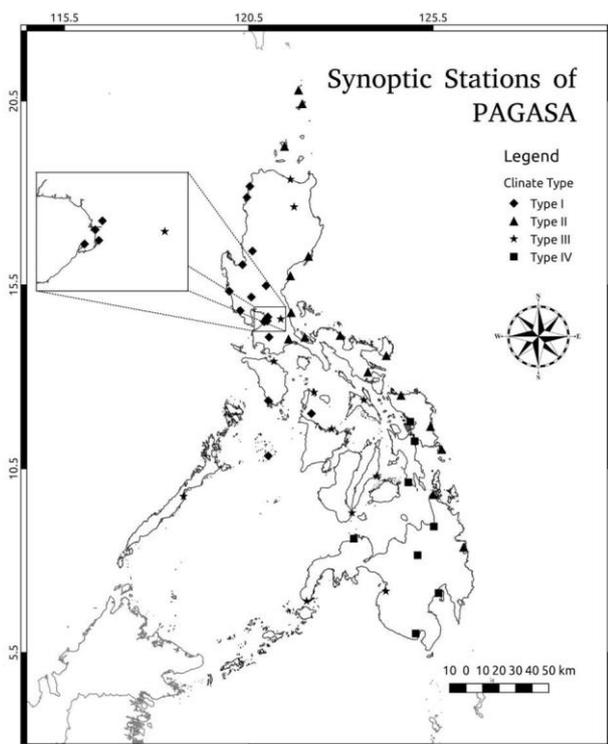


Figure 1. Locations of the synoptic stations used as validation points; stations are classified based on climate type of location.

2.4 Verification Metrics

To quantitatively evaluate the predicted rainfall, a set of statistical metrics was utilized. Among the standard methods, we used Bias, Root Mean Square Error (RMSE), and the Mean Absolute Error (MAE). (Brooks, et.al, 2017) These metrics measure the magnitude of forecast errors and how well the magnitudes of forecasts correspond with the observed. MAE is a measure of average error magnitude with scores ranging from zero to infinity, with zero as the perfect score. The RMSE is also a measure of “average” error, weighted according to the square of the error. It puts greater influence on large errors but is not indicative of the deviation direction. This score ranges from zero to infinity with zero as a perfect score. Bias, on the other hand, is the summation of the difference between the forecast and the observed value. It is used as an indicator of deviation direction; the score ranges from 0 to infinity with 0 as a perfect score, a negative bias is considered as under forecasted and a positive score is likewise considered as over forecasting.

$$Bias = \frac{1}{N} \sum_{i=1}^N (F_i - O_i) \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (F_i - O_i)^2} \quad (2)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |F_i - O_i| \quad (3)$$

where N = total number of samples
 F = forecasted rainfall intensity
 O = observed rainfall intensity

2.5 Visualization Development

The main consideration for the visualization was that it can show the error growth for each station per month. This was accomplished by plotting the locations of the synoptic stations; the color of each point is based on the climate type of the area of the station. The point size was dependent on the magnitude of the error. A map was generated for each month of 2015 to September 2018, for all verification tools used, and was converted into a GIF animation to monitor the changes in error magnitude.

The final visualization product was characterized by points that represent the stations used in the study, color coded according to climate type, type I uses red, type II is blue, type III is yellow and type IV is green. The magnitude of the verification score is depicted by the size of the point, hence larger points connote larger errors.

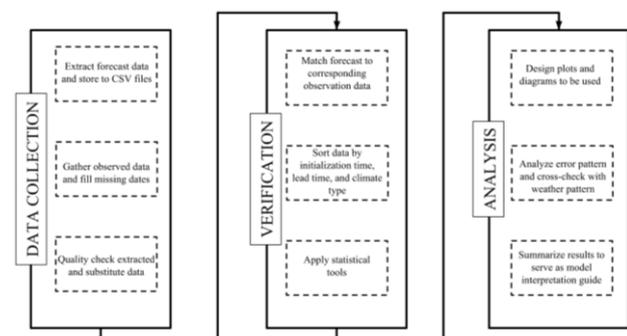


Figure 3. Flowchart showing the series of steps in implementing geo-visual analytics on the verification results

3. RESULTS AND DISCUSSIONS

PAGASA has a total of 52 stations scattered across the country. Among these 52, 16 are in the climate type one area, another 16 are in the climate type two area, 12 are in the climate type three and the last 8 are in the climate type four area. Ten stations in climate type one is located in coastal areas and the rest are situated in inland areas. Climate type two stations are all located at coastal areas. Stations classified as climate type three comprise of ten stations in coastal areas and two in inland areas. As for climate type four, the stations area equally divided into inland and coastal areas.

3.1 Yearly Forecast Quality Evaluation

The data were initially evaluated regardless of station location to assess the model's performance through the years since it undergoes yearly updates which often include different parameterization schemes and/or changes in the numerical methods and codes used to represent certain physical processes. Based on the resulting scores from the verification tools, forecast quality greatly improved beginning in 2016. Based on the history of updates there were a lot of changes employed during March 2016. The WRF version used jumped from version 3.3.1 to 3.7.1. The changes are summarized in the following table:

Category	Old Setting	New Setting	Remarks
Land Surface	Land Categories: 24 (USGS)	Land Categories: 20 (MODIS)	The MODIS land surface data has a higher resolution
Cloud Physics	Kain-Fritsch Scheme	Kain-Fritsch Scheme	Scheme codes and numerical methods were updated and improved
Micro-physics	Thompson Scheme	WRF 6 Scheme	WRF 6 Scheme is faster compared to Thompson
Vertical Levels	32	36	Increase in vertical levels can improve model's ability to simulate convection

Table 1. Summary of changes made in WRF configuration.

The version update improved forecast quality. 2015 forecast were over-forecasted, especially during the onset of the Southwest Monsoon. Error propagation with respect to the

increasing lead time is also prominent for 2015. Diagrams of the resulting scores per month and forecast lead-time are provided in the appendix.

The largest MAE and RMSE occurrence was during July 2015. This month is during the onset of the Southwest Monsoon, the large errors may have been related to the El Nino Southern Oscillation (ENSO) onset enhanced by the presence of the Madden-Julien Oscillation (MJO). The compounded effects of these events caused enhanced convections in the Maritime Continent. (Islam, et. Al, 2018). These convections may not have been well modelled by the operational model due to the coarse resolution. Although in comparison with the other months against those of the following years, there is still a great difference even without the ENSO and MJO which was brought about by the WRF version update.

The bias shows directional deviation. As seen in the plots above, 2015 exhibits a pattern wherein surface precipitation during the dry months (March, April, May) has a high dry bias and has a large wet bias during the peak rainfall season (July, August, September). While the following years, appear to have a consistent dry bias regardless of the season.

The following table summarizes the overall analysis for the model's error propagation.

Year	Error Pattern
2015	RMSE shows that there is a slow error growth as lead time increases Bias and MAE shows error is most prominent during the rainy season (Jul-Sep) Bias shows that there is substantial over-forecasting
2016	RMSE shows improvement from 2015, less prominent error during rainy season Bias shows that there is still over-forecasting during August. Dry season also appears to be more consistent regardless of lead time
2017	RMSE shows that errors are slightly higher compared to last year but is more evenly distributed throughout the year. Bias and MAE show under-forecasting during the wet season
2018	RMSE shows errors prominent during July but still greater improvement compared to past years. MAE and Bias also appear more consistent throughout lead times and show under-forecasting

Table 2. Comparison of the error distribution per year

3.2 Error Patterns based on Climate Type

The climate types used in the study was based on the climate map provided by PAGASA. The map is based on the Modified Coronas' Climate Classification. The process used yearly type of rainfall distribution modals from 1951 to 2010 encompassing 45 synoptic stations and 66 climate stations. (PAGASA, n.d.).

3.2.1 Climate Type One (CT1): Climate type one has two pronounced seasons: dry from November to April and wet during the rest of the year, experiencing maximum rain during June to September. Overall, CT1 exhibits large errors during the months of June to September and has reliable forecasts during October and April. The error pattern coincides with the area's peak wet season (high error) and dry season (low error). Additionally, June to September is the onset of the Southwest Monsoon which enhances the formation of localized thunderstorms and small scale convective systems; some weak spots of WRF. It is also possible that the notable errors over Metro Manila are mostly due to localized convection due to urban heating.

Year	Error Pattern
2015	Consistently large error in Manila area Large errors in South Luzon in January Peak errors in July to September Minimal error in November to February
2016	Peak errors in May to August Large errors in Manila from May to September Minimal error in January to April, and November and December
2017	Notable errors July to August Consistent error decrease from September to December
2018	Minimal errors in January to April Steady increase from June to September

Table 3. Summary of the observed CT1 error patterns per year.

3.2.2 Climate Type Two (CT2): Climate type two is characterized by wet season, which peaks from December to February. Minimum rain period is from December to February. Throughout the timeframe of analysis, this area exhibited relatively higher errors during the months of May to September and also has notable errors during December. CT2 covers the typhoon prone provinces which may explain the high errors during typhoon season. Typhoons and other large scale systems are actually the strong points of WRF and the low scores may be due to the limitations of the point matching verification statistics that do not fully assess the systems displacement error. The area is also affected by the easterlies wind system that aids in the formation of local convective systems.

Year	Error Pattern
2015	Error dominant in January Minimal error in February to May Steady error increase in June to August Large error in November and December
2016	Minimal error from January to April Error increases in May then becomes steady towards September Larger errors from November to December
2017	Small error during April Slow error increase from June to September

	Relatively large error during December
2018	Notable errors in January Consistent, low errors from February to April Slow and steady error increase from May to September

Table 4. Summary of the observed CT2 error patterns per year.

3.2.3 Climate Type Three (CT3): Climate type three has no pronounced maximum rain period with a dry season either from December to February or March to May. CT3 has the most erratic errors among the other types; there was no specific month that consistently had high errors however, November to December has notably reliable forecasts. In CT3, most stations are situated in islands and coastlines. Such locations are prone to local wind systems, also a cause of localized convection

Year	Error Pattern
2015	Small, steady increase in June to July Decreases during November to December Minimal during February to May
2016	Overall minimal error Least error during January to May Small Increase during September to October Decreases again During November to December
2017	Fluctuating errors throughout the year December has notable increase in error
2018	Peak error at January Consistently minimal errors from February to September

Table 5. Summary of the observed CT3 error patterns per year.

3.2.4 Climate Type Four (CT4): This climate type has more or less evenly distributed rainfall throughout the year. This CT has the most reliable forecasts which consistently have minimal errors. This consistency may be due to the area being shielded from most weather drivers of the country, therefore minimizing localized convection.

Year	Error Pattern
2015	Minimal changes throughout the year Largest error during August to September
2016	Minimal changes throughout the year Largest error during September
2017	Minimal changes throughout the year
2018	Minimal changes throughout the year

Table 6. Summary of the observed CT4 error patterns per year.

4. CONCLUSION AND RECOMMENDATION

4.1 Conclusion

Relating the verification scores to their location provides insight as to what are the probable causes of errors which can help in the model use and continued development.

4.1.1 Operations: In terms of operation, the study shows that the model is yielding better results over the years. It can be noted that 2018 forecasts have greatly improved compared to the forecasts of 2015. Based on previous studies, surface precipitation forecast during 2014 to 2015 were highly over-forecasted. The said years exhibited large diurnal wet biases and notably over-forecasts every 12 hours (PAGASA-NMS, 2016). As discussed in the previous chapter, after a major change in the physics parameterization scheme, forecasts for 2016 onwards greatly improved. The PBL scheme was able to address the wet bias every 12 hours and lowered the wet bias. However, this created a dry bias or under forecasting but the error was minimal compared to that of the two previous years. It is also notable that error growth with respect to the increasing lead time was minimal during 2016 to 2018 unlike in 2015 where the longer lead times had large error growths.

4.1.2 Model Interpretation Guidance: For model interpretation guidance, the study established that forecast biases area affected by climate type. Each climate type exhibits a unique pattern of errors which coincide with the peak rain season of each type. This leads to the conclusion that model interpretation should not solely focus on the general seasonal bias: surface precipitation is under-forecasted during the onset of the Southwest Monsoon. It also revealed that forecasts are most reliable in the climate type four areas; this type showed the lowest errors with minimal variance. In contrast with this, the study determined that the most erratic errors were exhibited by stations located in climate type three areas. This type did not exhibit a particular pattern of error changes throughout the time frame of analysis and also displayed a large variance. In addition, the study shows that there is an uneven distribution of stations based on climate type classification; majority of the stations are in climate types one and two. Since these types comprise almost half of the validation points, the said stations greatly influence the overall evaluation. Verification done as a whole can lead to the assumption that the model has poor and unreliable forecasts due to the underlying forecast errors based on climate type. Also, using WRF as a basis of rainfall should be done with caution and an understanding of the model's errors.

4.2 Recommendations

The study can be furthered by test case analysis per climate type; this can include focusing on different phenomena i.e. localized thunderstorms to provide a concrete evaluation of WRF's reliability in forecasting events due to local convective systems. Since the results imply that the operational domain does not model convection-caused rainfall adequately, PAGASA forecasters should consider the use of the 2nd domain, which has a higher spatial resolution at three kilometers per grid. Ideally, WRF forecasts at this resolution can resolve or model convective systems better. Test case comparison between the two domains can support this hypothesis.

It can also be improved by using alternative sources of observed data to create a denser network of validation points or using gridded observation data to allow evaluation of the whole WRF domain. This can allow comparison between model performance over land mass and water bodies.

Using gridded data can also improve the integration of geo-visual analytics such as creating a composition chart of which climate type has the highest difference between the forecast and observed value aggregated by province without bias caused by the uneven distribution of stations. This can also highlight specific areas that have the highest errors; the analysis used in the study was limited by the presence of a synoptic station which is not available in all provinces. Furthermore, gridded analysis can also highlight the seasonal variation of error based on climate type.

REFERENCES

- Brooks, H., Brown, B., Ebert, B., Ferro, C., Jenkner, J., Jolliffe, I., ... Stephenson, D. (2015, January 26). WWRP/WGNE Joint Working Group on Forecast Verification Research. Retrieved from <https://www.cawcr.gov.au/projects/verification/>
- Islam, M., Chan, A., Ashfold, M., Ooi, C., & Azari, M. (2018). Effects of El-Niño, Indian Ocean Dipole, and Madden-Julian Oscillation on Surface Air Temperature and Rainfall Anomalies over Southeast Asia in 2015. *Atmosphere*, 9(9), 352. doi: 10.3390/atmos9090352
- Li, Y., Lu, G., Wu, Z., He, H., Shi, J., Ma, Y., & Weng, S. (2016). Evaluation of Optimized WRF Precipitation Forecast over a Complex Topography Region during Flood Season. *Atmosphere*, 7(11), 145. doi:10.3390/atmos7110145
- Maisha, T. R. (2014). The influence of topography and model grid resolution on extreme weather forecasts over South Africa (Doctoral dissertation, The Influence of Topography and Model Grid Resolution on Extreme Weather Forecasts over South Africa).
- Moya-Álvarez, A., Gálvez, J., Holguín, A., Estevan, R., Kumar, S., Villalobos, E., Silva, Y. (2018). Extreme Rainfall Forecast with the WRF-ARW Model in the Central Andes of Peru. *Atmosphere*, 9(9), 362. DOI:10.3390/atmos9090362
- NCAR. (2018). WRF-Weather Research & Forecasting Model. Retrieved from National Center for Atmospheric Research Mesoscale and Microscale Meteorology Laboratory: <https://www.mmm.ucar.edu/weather-research-and-forecasting-model>
- NOAA. (2017). NOAA-National Weather Service. Retrieved October 25, 2018, from National Oceanic Atmospheric Administration: <https://www.weather.gov/media/ajk/brochures/NumericalWeatherPrediction.pdf>
- Philippine Atmospheric, Geophysical, Astronomical, Services Administration (n.d.). Climate of the Philippines. Retrieved from <http://bagong.pagasa.dost.gov.ph/information/climate-philippines>
- Rontu, L. (2007). Studies on Orographic Effects in a Numerical Weather Prediction Model (Doctoral dissertation, University of Helsinki)[Abstract].

Sun, B.-Y., & Bi, X.-Q. (2019). Validation for a tropical belt version of WRF: sensitivity tests on radiation and cumulus convection parameterizations. *Atmospheric and Oceanic Science Letters*, 12(3), 192–200. DOI: 10.1080/16742834.2019.1590118

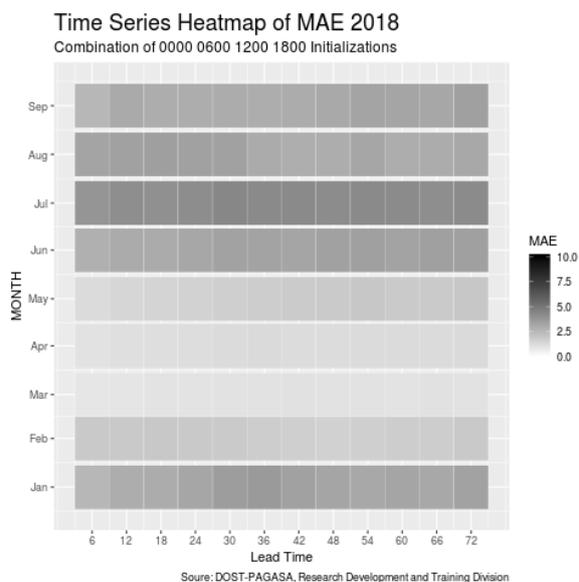
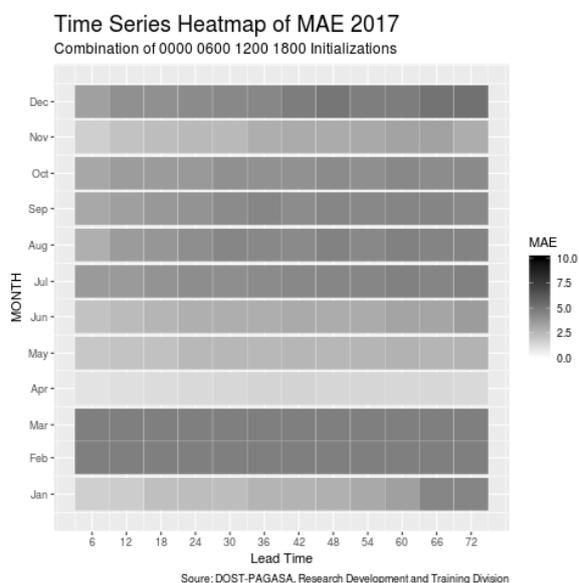
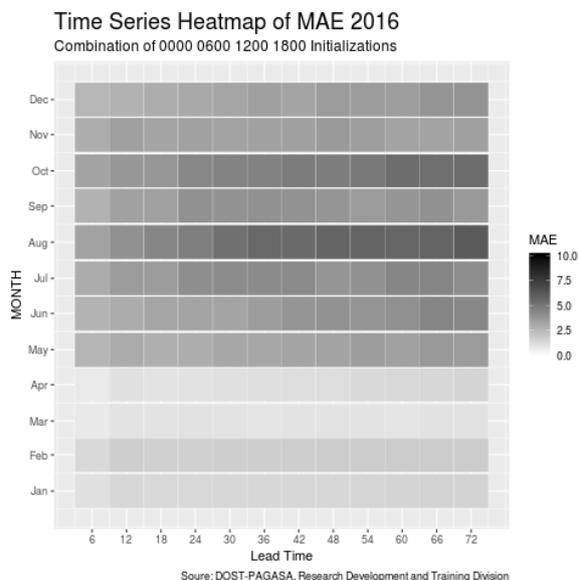
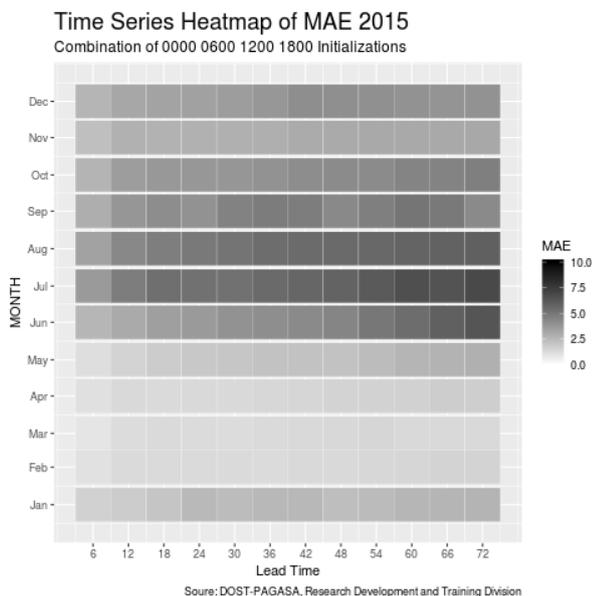
Wee, T.-K., Kuo, Y.-H., Lee, D.-K., Liu, Z., Wang, W., & Chen, S.-Y. (2012). Two Overlooked Biases of the Advanced Research WRF (ARW) Model. *MONTHLY WEATHER REVIEW*, 3907-3918

Yáñez-Morróni, G., Gironás, J., Caneo, M., Delgado, R., & Garreaud, R. (2018). Using the Weather Research and Forecasting (WRF) Model for Precipitation Forecasting in an Andean Region with Complex Topography. *Atmosphere*, 9(8), 304. DOI:10.3390/atmos9080304

APPENDIX

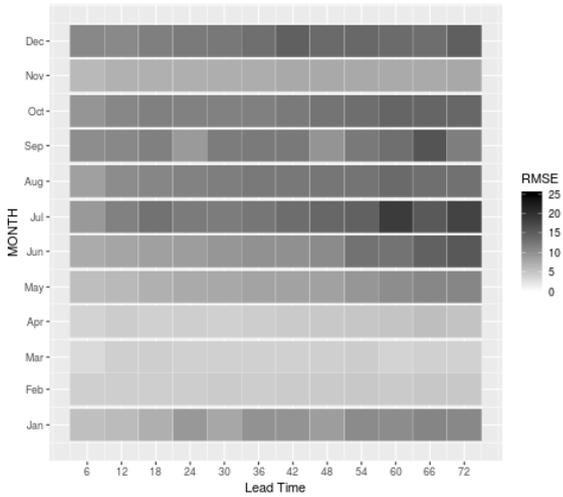
Animated plots used for the geovisual analysis can be found here: shorturl.at/bdorH.

The following plots show the resulting verification scores for the yearly performance evaluation of the model.



Time Series Heatmap of RMSE 2015

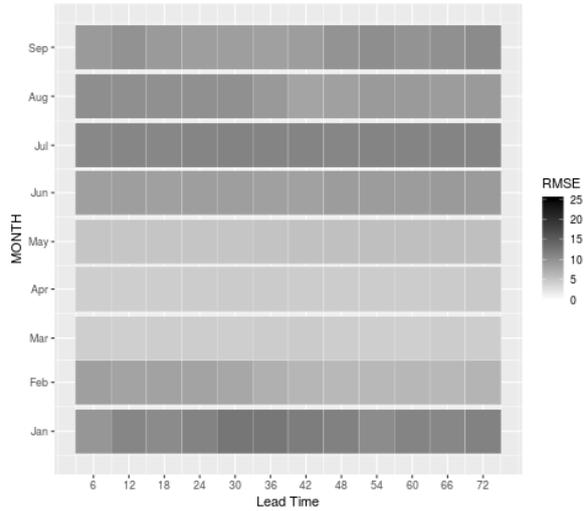
Combination of 0000 0600 1200 1800 Initializations



Source: DOST-PAGASA, Research Development and Training Division

Time Series Heatmap of RMSE 2018

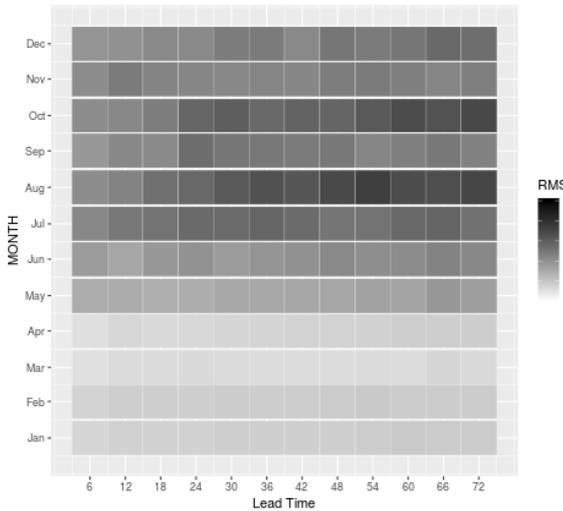
Combination of 0000 0600 1200 1800 Initializations



Source: DOST-PAGASA, Research Development and Training Division

Time Series Heatmap of RMSE 2016

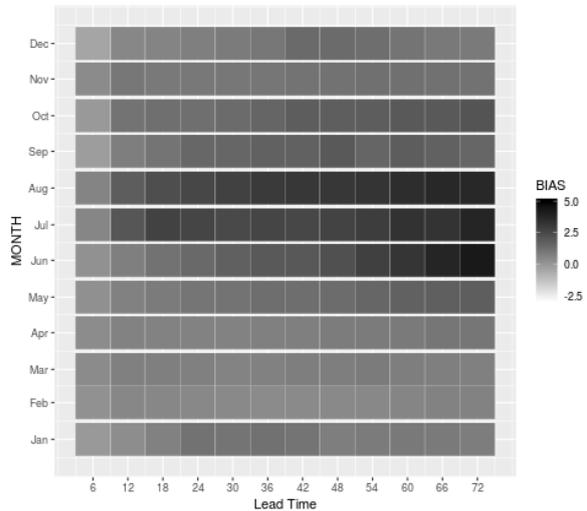
Combination of 0000 0600 1200 1800 Initializations



Source: DOST-PAGASA, Research Development and Training Division

Time Series Heatmap of BIAS 2015

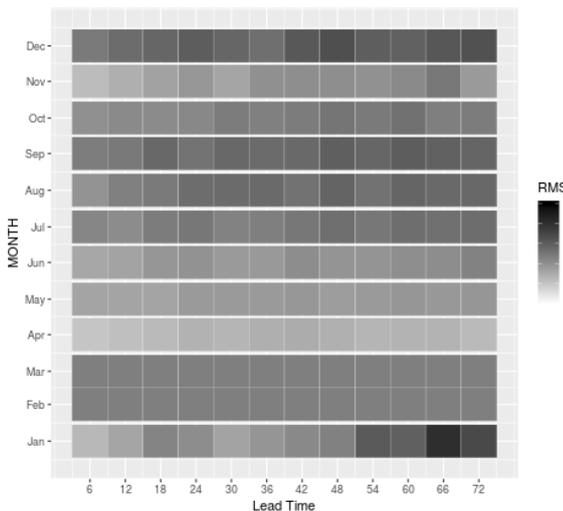
Combination of 0000 0600 1200 1800 Initializations



Source: DOST-PAGASA, Research Development and Training Division

Time Series Heatmap of RMSE 2017

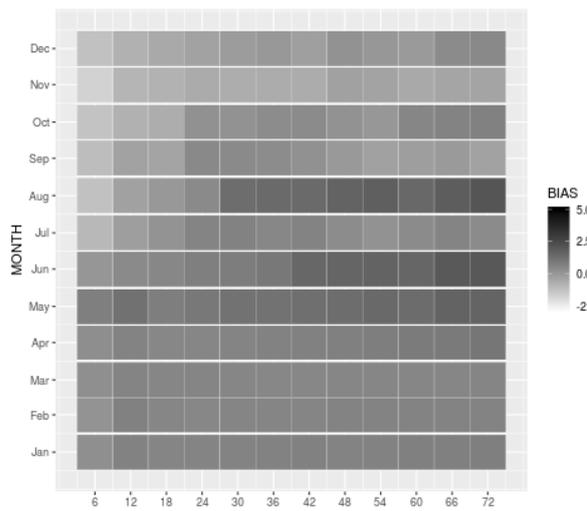
Combination of 0000 0600 1200 1800 Initializations



Source: DOST-PAGASA, Research Development and Training Division

Time Series Heatmap of BIAS 2016

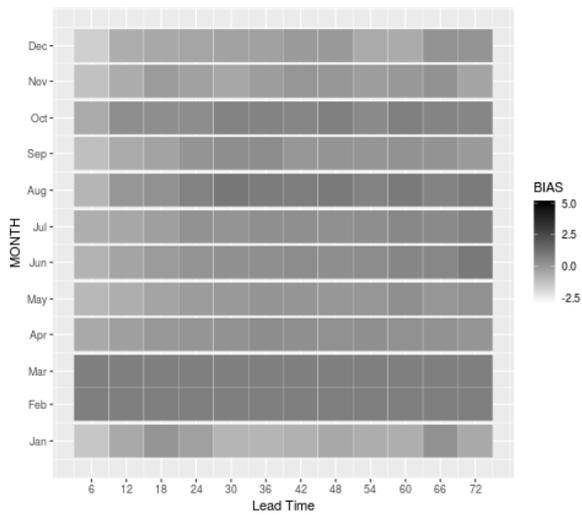
Combination of 0000 0600 1200 1800 Initializations



Source: DOST-PAGASA, Research Development and Training Division

Time Series Heatmap of BIAS 2017

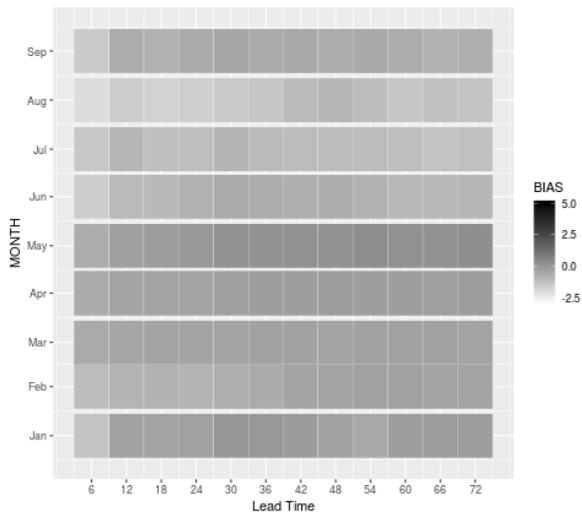
Combination of 0000 0600 1200 1800 Initializations



Source: DOST-PAGASA, Research Development and Training Division

Time Series Heatmap of BIAS 2018

Combination of 0000 0600 1200 1800 Initializations



Source: DOST-PAGASA, Research Development and Training Division