

2021 DTC visitor program final report

Developing a GridDiag use case for analysing 14-day precipitation accumulations for weeks 3 and 4 from the GloSea monthly forecast system

Marion Mittermaier¹

Global Model Evaluation and Development, Foundation Science, Met Office, United Kingdom

Aim

Develop a METplus GridDiag use case to read in Met Office GloSea5 ensemble forecasts pre-processed into 14-day accumulations spanning weeks 3 and 4, alongside pre-prepared 14-day GPM accumulations regridded to the GloSea grid. Then use the GridDiag tool to investigate the degree of correspondence between the model and observed accumulations and how these evolve over time.

1. Region of interest and scientific objectives

This study forms part of a larger body of work utilising several METplus tools and datasets with the scientific objective of exploring the predictability of dynamical precursors responsible for flooding in the Kerala region of India, which has seen severe flooding impacts in successive monsoon seasons between 2018 and 2020. The floods of 2019 were some of the most impactful seen. In the first instance the 2019 monsoon season is used for setting up and testing the GridDiag use case.

The Western Ghats (Fig. 1) is a mountain range along the W coast of India, between 900 and 2600 m in height with an average of around 1600 m. Strong, predominantly westerly flow with high humidity content is common during the monsoon season. The presence of the complex terrain leads to the forced uplift of this strong and moist low-level jet. These combine to locally enhance precipitation. The hypothesis is that the combination of these factors can potentially provide enhanced predictability at longer lead times than what may otherwise be thought likely.

2. Datasets

GloSea5-GC2 (GloSea for short) provided operational sub-seasonal and seasonal forecasts between July 2013 and January 2021 (MacLachlan et al., 2015, Scaife et al., 2014). The model uses the GC2.0 Unified Model (UM) configuration and is a coupled atmosphere-land-ocean-sea-ice ensemble forecast system with a horizontal resolution of 0.8° degrees in latitude and 0.5° degrees in longitude, which equates to ~60 km in the mid-latitudes and ~95 km at the Equator. GC2.0 uses the Global Atmosphere (GA6) component (Walters et al., 2017) which is coupled to the Joint UK Land Environment Simulator (JULES; Best et al., 2011) and is forced with data from the Japanese re-analysis (JRA-55), NEMO (Nucleus for European Modelling of the Ocean; Madec, 2008) and the Los Alamos Sea Ice Model (CICE) (Hunke and Lipscomb, 2010). A stochastic kinetic energy backscatter scheme (SKEB2; Bowler et al., 2009) is used to introduce small grid-level perturbations throughout the integrations to enhance ensemble spread. The forecasting system is run in three parts: i) sub-seasonal ii) seasonal and iii) hindcast. Four members are initialized at 00 UTC every day. Two members are run out to 64 days (sub-seasonal forecast system). Another two members are run out

¹ Corresponding address: marion.mittermaier@metoffice.gov.uk

to 216 days (seasonal forecast system). These four members are used to generate a 40-member lagged ensemble forecast with 10 days of lag time. Here all 40 lagged members are used. At NWP time scales (short- to medium-range), forecast skill is primarily derived from the initial conditions. Beyond this (e.g., sub-seasonal-to-seasonal scales) forecast skill is derived from external forcing (such as SSTs). The cross-over point between the different primary drivers is anecdotally somewhere between days 10 and 15, but very much dependent on the geographical region and climate.

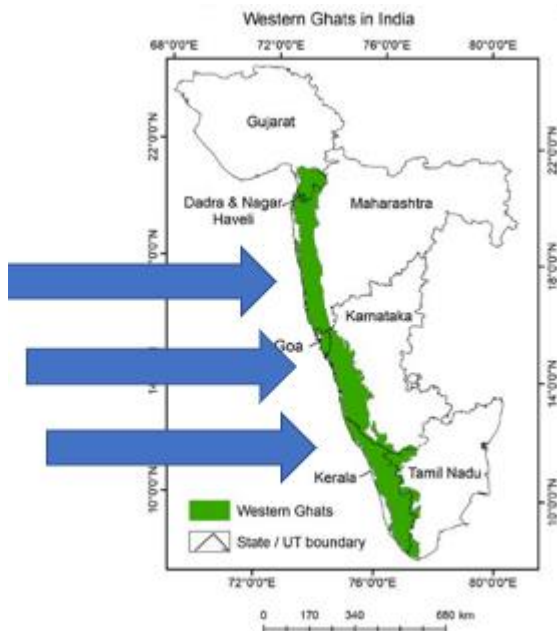


Figure 1 Area defined as the Western Ghats.

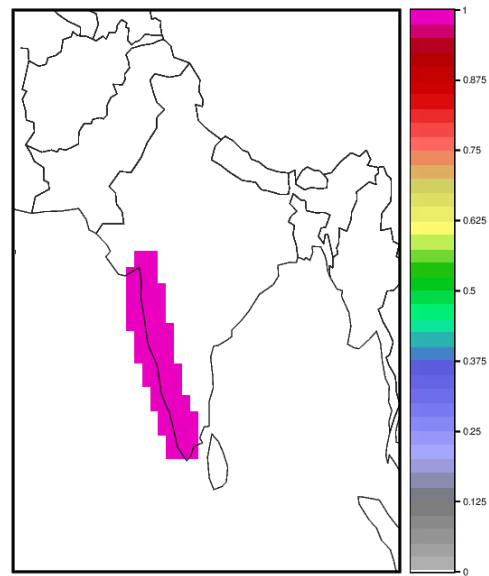


Figure 2 Mask created using the model topography as input to GenVxMask to create a West Coast + Western Ghats area mask.

Here we are particularly interested in the utility of two-week accumulations for weeks 3 and 4 and whether they have any skill at pointing to potentially large (and potentially impactful) accumulations some weeks ahead. The premise here is that the timing of rainfall events is no longer reliable at the longer lead times, but large multi-day accumulations may be able to pick out time windows which suggest large accumulations, which are in turn worthy of watching from an early-warning perspective.

Global Precipitation Measurement (Huffman et al., 2020) Integrated Multi-satellite Retrievals for GPM (GPM IMERG) (Hou et al., 2014; Skofronick-Jackson et al., 2017) final precipitation products are used in this study. The data are available at a horizontal resolution of $0.1^\circ \times 0.1^\circ$ ($\sim 10 \text{ km} \times 10 \text{ km}$) with 30 minutes temporal resolution. To evaluate monthly forecasts, the GPM data were interpolated to the GloSea grid ($\sim 93 \text{ km}$ at the Equator) and aggregated to produce 14-day accumulations.

3. Methodology

First a brief word on the data processing, which was not specifically designed for use with METplus, but is more generic. Given this is a lagged ensemble, the 40 members are assembled as a first step. From here the ensemble forecasts valid for a specific time horizon are extracted and the 14-day accumulations calculated. These were then written to a netCDF file which contains *all* the members for a validity day (reference date at the end of the accumulation window).

METplus, as a rule, reads ensemble members that are in individual files. Having all the ensemble members and validity days in the same file was something different for METplus wrappers to deal with. The GridDiag use case was constructed on the back of the predefined (non-standard METplus) data model. Appendices A and B provide the wrapper and the netCDF file headers for reference. To get the use case to work required utilising the custom loop and the dynamic level selection capability in METplus. It demonstrated the flexibility that these capabilities provide in coping with datasets which are organised differently to the norm.

NetCDF output from GridDiag was then read in and manipulated for visualisation. GridDiag was run with very fine bins to provide the ability to provide downstream flexibility in the analysis. The number of bins was rationalised for plotting to provide a clear distribution. The distributions were normalised and plotted using a log scale given that the values can vary over several orders of magnitude.

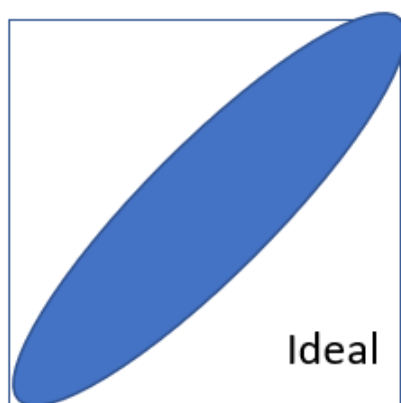


Figure 3 The ideal outcome for a joint distribution comparing the same variable is that all the values lie on the diagonal or are clustered close to it.

Joint distributions measure the degree of association between two variables or quantities. The ideal is that there is perfect correspondence on a paired grid-point-by-grid-point basis then only the bins along the diagonal would be populated. This represents the perfectly unbiased scenario. Realistically speaking there is always a bias and there are other factors (e.g., location errors) which affect how far GloSea-GPM pairs are away from the diagonal. Good correspondence would find the pairs clustered around the diagonal as shown in Fig. 3. Flat distributions in either direction suggest a lack of discrimination in one of the data sets, i.e., there are many similar values in one dataset associated with a large range of values in the other.

4. Results

The strength of association between 14-day accumulations from a monthly forecast ensemble and GPM was analysed for the JJAS 2019 monsoon season using GridDiag. To get as rich, but also as clear, a distribution as possible all the ensemble members are analysed for the Western Ghats region as shown in Fig. 2 as model-GPM pairs. Output was calculated in a rolling 14-day manner (using the First In, First Out or FIFO principle), such that one can see whether there is a temporal evolution in the association as the monsoon progresses.

The full time series of joint histograms is provided in Appendix C. Until mid-July there seems little agreement as the monsoon gets going in the model way too early. From July 20th onwards a picture of reasonable agreement emerges, though there is a surprising over-forecast bias in GloSea. By early August the relationship begins to collapse (the model is unable to discriminate), noting that the biggest impacts from the flooding were reported between 6-11 August 2019. Maps of the *daily* GPM rainfall show that the wettest period in the region of interest ahead of the floods was between July 19th and 24th (Fig 4); the heavier rain in the Western Ghats region migrating north after this (not shown). There is therefore some hint that the period with the highest degree of association was collocated in time with the week of heaviest rainfall. Appendix C shows that all the 14-day

accumulations which include this period, especially those ending between July 20th and 31st show very good association between the GloSea and GPM.

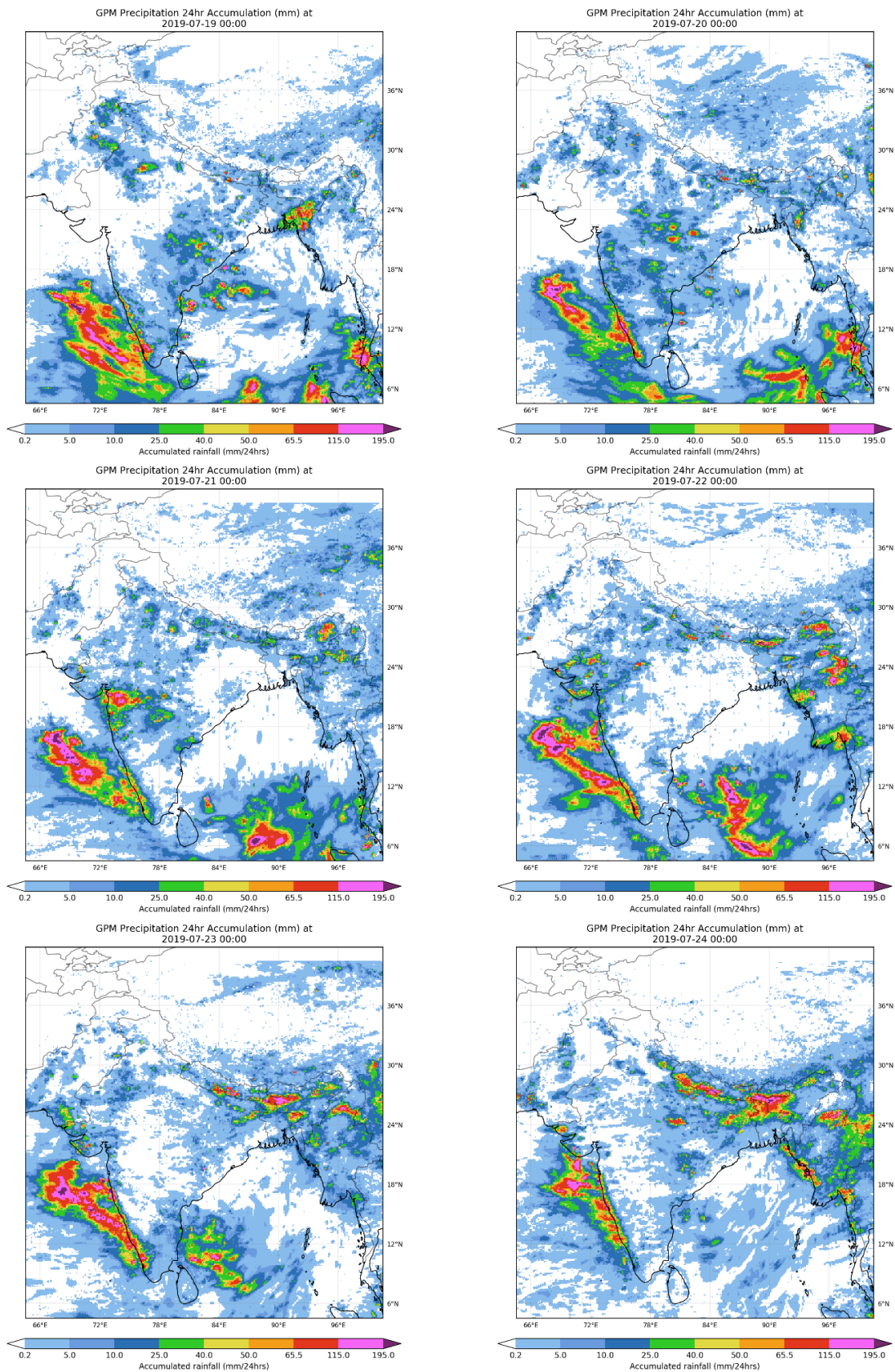


Figure 4 Sequence of daily GPM rainfall accumulations showing the week of intense rains over the Kerala region which encompasses the southern half of the Western Ghats. (Images courtesy, Rachel North)

There appear to be distinct phases in the monsoon season from the full time series presented in Appendix C. A representative of each of these evolutionary phases was selected from the rolling set of joint histograms to summarise the behaviour. This is shown in Fig 5. At the start of the 2019 monsoon season in (a), the model is too fast for reality. What follows is a messy transition whereby in (b) the model continues to be keen on producing large amounts of rainfall but not necessarily where it is observed in the region. At the peak of the season (c) shows a strong association signal emerges, though the model continues to be keen on over-estimating the higher totals compared to GPM. At the lower end of accumulation totals, the model appears to potentially be under-estimating slightly. Beyond the monsoon peak (d) shows there follows a messy degeneration in the association whereby the model is far less keen on larger totals, unlike earlier in the season. The final diminution in the season shown in (e) is roughly symmetrical, with a slight tendency towards too little rain in a 14-window in weeks 3-4.

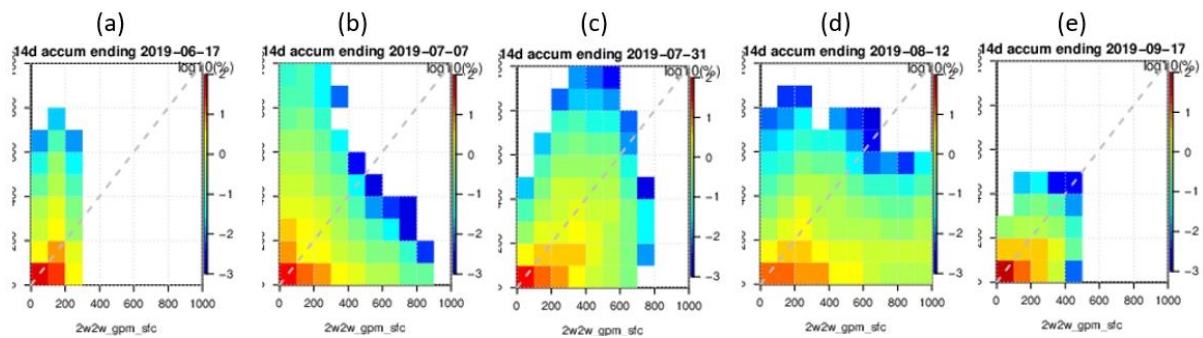


Figure 5 Selection of joint distributions for the 2019 monsoon season, charting the evolution of the distributions as the season progresses.

5. Conceptualising the findings and further work

It has been insightful to document the temporal evolution of the association in 14-day accumulations for weeks 3 and 4. What has been surprising is the level of agreement, even without any form of hindcast adjustment, that is evident near the peak of the 2019 season. The behaviour in Fig 5 has been conceptualised as a schematic, and depicted in Fig. 6. The plan is to repeat this analysis for other monsoon seasons to establish whether a pattern emerges which gives this behaviour further credibility in terms of repeatability, and more specifically, if this clear peak in association can be linked with the other Kerala flooding events.

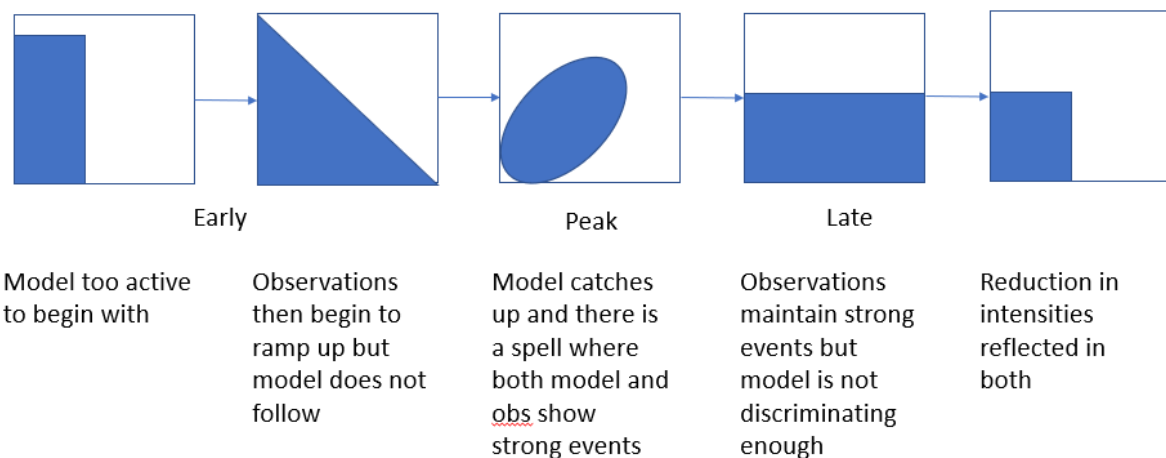


Figure 6 Conceptual model based on the evolution of the 2019 season.

6. Conclusions

A GridDiag use case has been developed which will ingest a time series of pre-processed lagged ensemble members all in the same file. This is a rather unusual data model for METplus to have to deal with and makes use of some of the flexible features which are now available to accommodate such data sets.

The tool has shown it can provide some very useful and swift analysis of associations between variables.

Funding statement

This work is being funded under the Met Office Weather and Climate Science for Service Partnership (WCSSP) India as part of the Newton Fund.

Acknowledgements and thanks

At the Met Office:

- Dr. Caroline Jones for providing technical support for the GloSea ensemble extraction.
- Dr Seshagiri Rao Kolusu for developing the GPM and GloSea forecast data processing chain.
- Dr Rachel North for graciously allowing me to reproduce her GPM daily “quick looks” for the 2019 season for this report.

At the NCAR node of the DTC:

- George McCabe and John Halley Gotway for their never-ending patience, help and advice in getting the use case off the ground.
- Eric Gilleland, Louisa Nance, and Tara Jensen for giving me the opportunity and being amazing hosts.

References

- Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J. (2011) The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes. *Geosci. Model Dev.*, 4, 677–699, <https://doi.org/10.5194/gmd-4-677-2011>, 2011.
- Bowler, N. E., Arribas, A., Beare, S. E., Mylne, K. R., & Shutts, G. J. (2009). The local ETKF and SKEB: Upgrades to the MOGREPS short-range ensemble prediction system. *Q. J. R. Met. Soc.*, 135(640), 767-776.
- Hou, A. Y., and Coauthors, (2014) The Global Precipitation Measurement Mission. *Bull. Amer. Meteor. Soc.*, 95 (5), 701-722, doi:10.1175/BAMS-D-13-00164.1.
- Hunke, E. C. and Lipscomb, W. H. (2010) CICE: the Los Alamos sea ice model documentation and software user's manual, Version 4.1, LA-CC-06-012, Los Alamos National Laboratory, New Mexico, 76 pp., 2010.
- Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K.L., Joyce, R.J., Kidd, C., Nelkin, E.J., Sorooshian, S., Stocker, E.F., Tan, J. and Wolff, D.B., (2020) Integrated multi-satellite retrievals for the Global Precipitation Measurement (GPM) Mission (IMERG). In *Satellite precipitation measurement* (pp. 343-353). Springer, Cham.
- MacLachlan C., A. Arribas, K.A. Peterson, A. Maidens, D. Fereday, A.A. Scaife, M. Gordon, M. Vellinga, A. Williams, R. E. Comer, J. Camp and P. Xavier. (2015) Description of GloSea5: the Met Office high resolution seasonal forecast system. *Q. J. R. Met. Soc.*, DOI: 10.1002/qj.2396.
- Madec, G.: NEMO ocean engine, Technical Report 27, Note du Pole demodélisation, Institut Pierre-Simon Laplace (IPSL), Paris, France, 300 pp., 2008.
- Scaife A.A., A. Arribas, E. Blockley, A. Brookshaw, R. T. Clark, N. Dunstone, R. Eade, D. Fereday, C. K. Folland, M. Gordon, L. Hermanson, J. R. Knight, D. J. Lea, C. MacLachlan, A. Maidens, M. Martin, A. K. Peterson, D. Smith, M. Vellinga, E. Wallace, J. Waters and A. Williams, 2014. Skilful Long Range Prediction of European and North American Winters. *Geophys. Res. Lett.*, 41, 2514-2519, DOI:10.1002/2014GL059637.
- Skofronick-Jackson, G., and Coauthors, 2017: The Global Precipitation Measurement (GPM) Mission for Science and Society. *Bull. Amer. Meteor. Soc.*, 98 (8), 1679--1695, doi:10.1175/BAMS-D-15-00306.1.
- Walters, D., Boutle, I., Brooks, M., Melvin, T., Stratton, R., Vosper, S., ... & Xavier, P. (2017). The Met Office unified model global atmosphere 6.0/6.1 and JULES global land 6.0/6.1 configurations. *Geoscientific Model Development*, 10(4), 1487-1520. <https://doi.org/10.5194/gmd-10-1487-2017>.

Appendix A: GridDiag config file for the use case

```
#
# CONFIGURATION
#
[config]

# Looping by times: steps through each 'task' in the PROCESS_LIST for each
# defined time, and repeats until all times have been evaluated.
LOOP_ORDER = times

# 'Tasks' to be run
PROCESS_LIST = GridDiag
GRID_DIAG_CUSTOM_LOOP_LIST = begin_end_incr(0, 39, 1)
LOOP_BY = VALID

# The init time begin and end times, increment, and last init hour.
VALID_TIME_FMT = %Y%m%d%H
VALID_BEG = 2019060100
VALID_END = 2019073100

# This is the step-size. Increment in seconds from the begin time to the end time
# set to 6 hours = 21600 seconds
VALID_INCREMENT = 86400

#LEAD_SEQ = 24, 48, 72, 96, 120

# frequency to run the tool
# valid options include:
# RUN_ONCE, RUN_ONCE_PER_VALID_OR_VALID, RUN_ONCE_PER_LEAD, RUN_ONCE_FOR_EACH
GRID_DIAG_RUNTIME_FREQ = RUN_ONCE_FOR_EACH

#LOG_GRID_DIAG_VERBOSITY = 2

GRID_DIAG_DESC = GloSea5

BOTH_VAR1_NAME = precipitation_amount
BOTH_VAR1_LEVELS = ((valid?fmt=%Y%m%d %H?shift=43200), {custom?fmt=%d}, *, *)
BOTH_VAR1_OPTIONS = set_attr_name="2w2w_glosea"; set_attr_level="sfc"; n_bins = 10; range = [0, 1000];

BOTH_VAR2_NAME = precipitation_flux
BOTH_VAR2_LEVELS = ((valid?fmt=%Y%m%d %H), *, *)
BOTH_VAR2_OPTIONS = set_attr_name="2w2w_gpm"; set_attr_level="sfc"; n_bins = 10; range = [0, 1000];

# The following variables set values in the MET
# configuration file used by this example
# Leaving these values commented will use the value
# found in the default MET configuration file
#GRID_DIAG_REGRID_TO_GRID =
#GRID_DIAG_REGRID_METHOD =
#GRID_DIAG_REGRID_WIDTH =
#GRID_DIAG_REGRID_VLD THRESH =
#GRID_DIAG_REGRID_SHAPE =

GRID_DIAG_MASK_POLY = /home/h02/frmm/WCSP/India/glosea/R/western_ghats_mask.nc

#
# DIRECTORIES
#
[dir]

# location of configuration files used by MET applications; NOT NEEDED
#CONFIG_DIR =
# Configuration file
#GRID_DIAG_CONFIG_FILE =
#GRID_DIAG_INPUT_DIR =
GRID_DIAG_OUTPUT_DIR = {OUTPUT_BASE}/GridDiag

[filename_templates]

GRID_DIAG_INPUT_TEMPLATE = /path/to/forecast/data/.nc, /path/to/obs/data/.nc
GRID_DIAG_OUTPUT_TEMPLATE =
grid_diag_glosea_on_glosea_grid_2w2w_precip_wghats_{valid?fmt=%Y%m%d%H}_{custom?fmt=%d}_2w2w.nc
```


Appendix B: NetCDF headers of input files

The input files are structured such that all (40) 2-week accumulations for each of the ensemble members and each of the dates spanning weeks 3 and 4 (or days 14 to 28) are stored in the same file together, i.e. the lagged ensemble was assembled upstream and the whole monsoon season effectively grouped and written to one file. This was feasible given the size of the domain.

```
netcdf GLOSEA5_5226_w2w2_2019_JJAS_IND_marion_JHG {
dimensions:
    time = UNLIMITED ; // (122 currently)
    bnds = 2 ;
    longitude = 48 ;
    latitude = 70 ;
    realization = 40 ;
variables:
    double time(time) ;
        time:standard_name = "time" ;
        time:bounds = "time_bnds" ;
        time:units = "hours since 1970-01-01 00:00:00" ;
        time:calendar = "gregorian" ;
        time:axis = "T" ;
    double time_bnds(time, bnds) ;
    float longitude(longitude) ;
        longitude:standard_name = "projection_x_coordinate" ;
        longitude:long_name = "longitude" ;
        longitude:axis = "X" ;
        longitude:units = "degrees_east" ;
    float latitude(latitude) ;
        latitude:standard_name = "projection_y_coordinate" ;
        latitude:long_name = "latitude" ;
        latitude:axis = "Y" ;
        latitude:units = "degrees_north" ;
    int latitude_longitude ;
        latitude_longitude:grid_mapping_name = "latitude_longitude" ;
        latitude_longitude:longitude_of_prime_meridian = 0. ;
        latitude_longitude:earth_radius = 6371229. ;
    double realization(realization) ;
        realization:standard_name = "realization" ;
        realization:units = "1" ;
        realization:axis = "Z" ;
    float precipitation_amount(time, realization, latitude, longitude) ;
        precipitation_amount:standard_name = "precipitation_amount" ;
        precipitation_amount:units = "kg m-2" ;
        precipitation_amount:grid_mapping = "latitude_longitude" ;
        precipitation_amount:um_stash_source = "m01s05i226" ;
        precipitation_amount:cell_methods = "time: mean (interval: 1 hour) time: sum" ;

// global attributes:
    :CDI = "Climate Data Interface version 2.0.3 (https://mpimet.mpg.de/cdi)" ;
    :Conventions = "CF-1.7" ;
    :source = "Data from Met Office Unified Model" ;
    :um_version = "10.1" ;
    :CDO = "Climate Data Operators version 2.0.3 (https://mpimet.mpg.de/cdo)" ;
    :history = "Wed Aug 24 19:30:04 2022: ncatted -a units,latitude,o,c,degrees_north -a
units,longitude,o,c,degrees_east GLOSEA5_5226_w2w2_2019_JJAS_IND_marion.nc -o
GLOSEA5_5226_w2w2_2019_JJAS_IND_marion_JHG.nc\nMon May 30 13:15:27 2022: cdo -selmon,6,7,8,9
GLOSEA5_5226_w2w2_2019_MJJASO_IND_marion.nc GLOSEA5_5226_w2w2_2019_JJAS_IND_marion.nc" ;
    :NCO = "netCDF Operators version 4.7.5 (Homepage = http://nco.sf.net, Code =
http://github.com/nco/nco)" ;
}
```

```

netcdf GPM_w2w2_2019_JJAS_IND_marion {
dimensions:
    time = 122 ;
    latitude = 70 ;
    longitude = 48 ;
    bnds = 2 ;
variables:
    float precipitation_flux(time, latitude, longitude) ;
        precipitation_flux:units = "mm" ;
        precipitation_flux:cell_methods = "time: sum" ;
    double time(time) ;
        time:axis = "T" ;
        time:bounds = "time_bnds" ;
        time:units = "days since 2019-01-01 00:00:00" ;
        time:standard_name = "time" ;
        time:calendar = "proleptic_gregorian" ;
    double time_bnds(time, bnds) ;
    float latitude(latitude) ;
        latitude:axis = "Y" ;
        latitude:units = "degrees_north" ;
        latitude:standard_name = "latitude" ;
    float longitude(longitude) ;
        longitude:axis = "X" ;
        longitude:units = "degrees_east" ;
        longitude:standard_name = "longitude" ;

// global attributes:
    :CDI = "Climate Data Interface version 2.0.3 (https://mpimet.mpg.de/cdi)" ;
    :CDO = "Climate Data Operators version 2.0.3 (https://mpimet.mpg.de/cdo)" ;
    :history = "Mon May 30 12:17:28 2022: cdo -selmon,6,7,8,9
GPM_2019_mmday_daysum_w2w2_marion.nc GPM_2019_mmday_daysum_w2w2_JJAS_marion.nc" ;
    :Conventions = "CF-1.7" ;
}

```

Appendix C: Sequence of daily rolling joint histograms for the 2019 season

