KEMENTERIAN SUMBER ASLI, ALAM SEKITAR
DAN PERUBAHAN IKLIM
Ministry of Natural Resources, Environment and Climate Change

MALAYSIAN METEOROLOGICAL DEPARTMENT
MINISTRY OF NATURAL RESOURCES, ENVIRONMENT AND CLIMATE CHANGE

Technical Note No. 4/2022

Improving WRF-MMD's Precipitation Forecasts Consistency using Time-Lagged Ensemble Method

Muhamad Sofian Bin Muhamad Yusof and
Muhammad Firdaus Ammar Bin Abdullah

TECHNICAL NOTE NO. 4/2022

Improving WRF-MMD's Precipitation Forecasts Consistency using Time-Lagged Ensemble Method

By
Muhamad Sofian Bin Muhamad Yusof and
Muhammad Firdaus Ammar Bin Abdullah

Perpustakaan Negara Malaysia Data Pengkatalogan-dalam-Penerbitan

Published and printed by:
Jabatan Meteorologi Malaysia
Jalan Sultan
46667 Petaling Jaya
Selangor Darul Ehsan
Malaysi

Contents

Improving WRF-MMD's Precipitation Forecasts Consistency using Time-Lagged Ensemble Method

Muhamad Sofian Bin Muhamad Yusof and
Muhammad Firdaus Ammar Bin Abdullah

ABSTRACT

Time-Lagged Ensemble (TLE) technique has been applied to the output of numerical weather prediction (NWP) in MET Malaysia, also known as WRF-MMD (Weather Research Forecast – Malaysian Meteorological Department), to identify the consistency of heavy rainfall prediction in spatial aspect. Existing rainfall prediction from WRF-MMD are presented in a deterministic way and does not have any information about uncertainty in the prediction or forecast. Furthermore, the latest deterministic forecast often differs from the previous forecasts for the same valid time, in other words, the forecasts are inconsistent. Moreover, the latest one not guaranteed to has better accuracy than the previous one, which can cause wrong interpretation of WRF-MMD prediction by users. Therefore, the usage of the TLE method hopefully can improve the consistency of WRF-MMD rainfall forecast and finally can improve the interpretation of its products.

Through this study, it is found that the TLE method can give better consistency for WRF-MMD precipitation forecasts. This can be achieved because all the deterministic forecasts for the same valid time are considered in the TLE calculation to produce probability values for selected 24-hour accumulated rainfall thresholds. After TLE is applied to the WRF-MMD output, the predictions for heavy rainfall caused by strong monsoon surges become more consistent and trustworthy compared to the deterministic forecasts. Besides building trust and confidence in WRF-MMD products, TLE also helps improve weather forecasting, especially over the Malaysia, through better interpretation of WRF-MMD products.

1. INTRODUCTION

Forecasts inconsistency and uncertainty are the main problem in the NWP which can undermine confidence in its predictions. It is caused by sensitiveness of the NWP model's output to the input data or the initial condition. In other words, small errors in the initial condition can result in significant differences in the NWP model's predictions. Therefore, the development of weather forecasting, which was previously deterministic, has changed to a more probabilistic one, which can evaluate the uncertainty of NWP predictions and thus help improve its consistency.

One of the methods to make NWP predictions that have better consistency and, at the same time, can measure the NWP uncertainty is to use the Ensemble Prediction System (EPS). The way this system is run is to assume that the initial condition, or can also be called model analysis, is not 100% accurate. In fact, this assumption makes sense because the analysis was generated from observation data whose accuracy depends on the error of the apparatus or the sensors. In addition, there are also areas where very few observations can be made, such as in some parts of the ocean, which requires data interpolation and further contributes to the errors in model analysis.

To take into account the uncertainty in this analysis, additional sets of model analyses will be generated by perturbing the meteorological parameters from the original analysis, also known as control analysis[1]. This control analysis is input for the deterministic forecasts which has longer forecast period and higher spatial resolution. Then, all these analyses, including the control analysis, will be used as initial conditions for the NWP model to produce the same number of forecast sets, called ensemble members. These sets usually have fewer periods of forecasts and coarser spatial resolution than deterministic forecasts because of limited computing resources. There are no certain fixed amount of these members, and some can reach up to more than 50 sets[2]. Due to the sensitivity of the NWP output to the initial condition, each forecast's outcomes will be different from the other and require a statistical approach to interpret the results, as depicted in Figure 1.

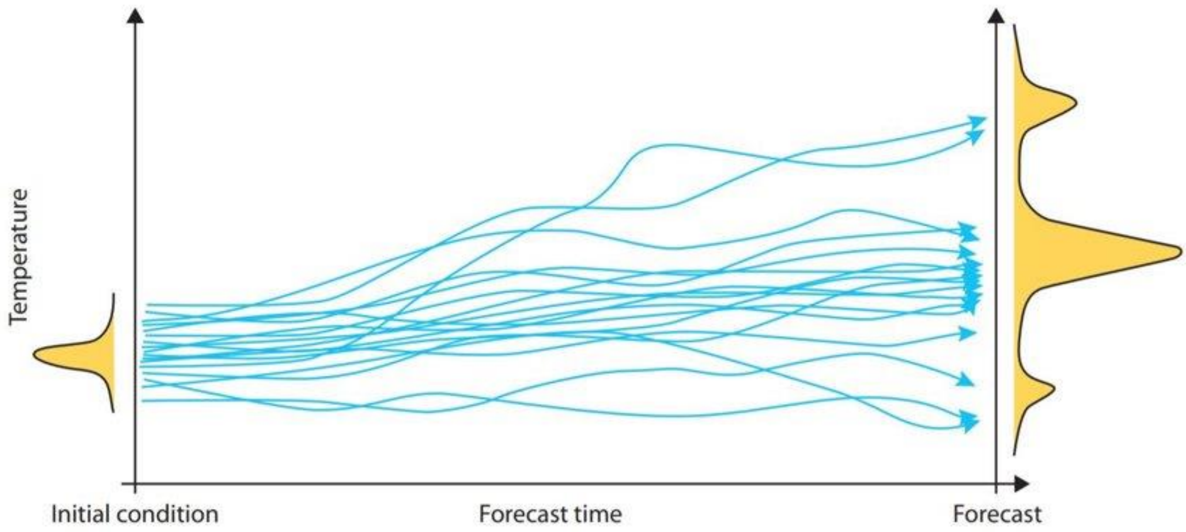
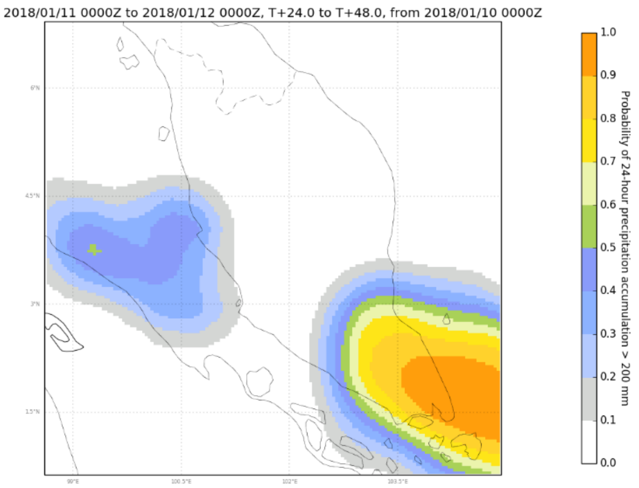


Figure 1: Sets of forecasts derived from many analyses (or initial conditions) in Ensemble Prediction System (EPS). (Source: ECMWF – European Centre for Medium-Range Weather Forecasts)

Besides the usual statistical values such as mean and variance, a value of probability also can be calculated from these members to represent the uncertainty of the forecast. For example, a high probability value for heavy rainfall forecast over a certain area means most of the ensemble members predicted that there will be heavy rainfall over the area, and only a few members predicted otherwise. This can be understood more from Figure 2 were Figure 2(a) shows EPS probabilities calculated from individual EPS members forecasts in Figure 2(b). From the figure, EPS forecasted a higher probability for accumulated rainfall over 200mm in 24 hours over the southern part of Peninsular Malaysia (PM) compared to Malacca Strait (MS). For users or forecasters, the interpretation of this EPS forecast was heavy rainfall event over the southern part of PM has higher confidence, thus can be trusted more, compared to heavy rainfall event over MS.

EPS has many advantages over Time-lagged Ensemble (TLE) method because of larger samples and longer lead time. However, the EPS technique requires very high computational resources with expensive costs. Therefore, there is a need to find a cheaper and simpler alternative method to calculate the uncertainty in the forecast such as the TLE method.

a) EPS forecast for probabilities of accumulated rainfall over 200mm in 24 hours on 10 Jan 2018.



b) Individual members forecast for accumulated rainfall in 24 hours (mm) on 10 Jan 2018.

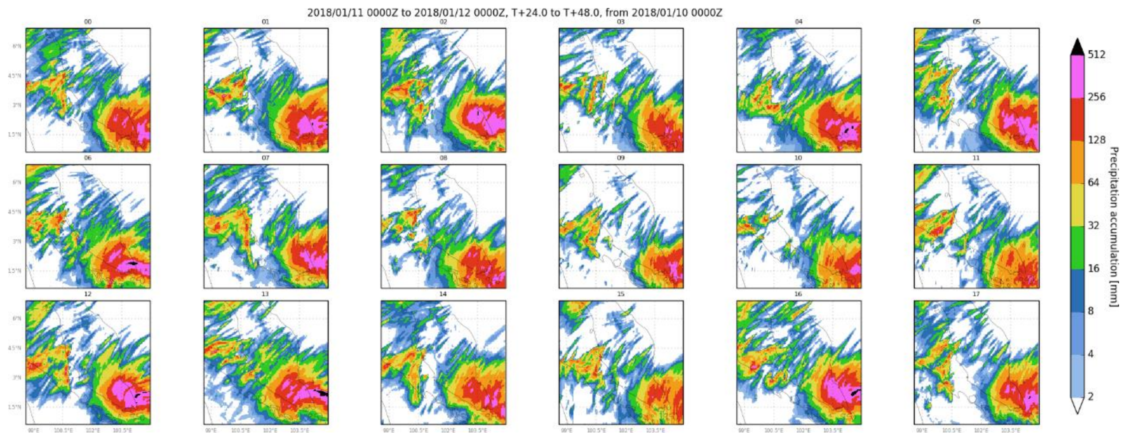


Figure 2: (a) EPS probability forecast and, (b) individual members forecast for 24-hour accumulated rainfall in next 24-48 hours. (Source: Unified Model Global EPS, UK MET Office)

TLE first is introduced by Hoffman and Kalnay in 1983 in the publication entitled "Lagged average forecasting, an alternative to Monte Carlo forecasting"[3]. Through this technique, multiple forecasts for the same valid time are obtained from the forecasts generated using initial conditions at different times, as shown in Figure 3. In this figure, NWP for 48 hours forecast is run in two cycles per day which are on 00:00 and 12:00 UTC. Thus, for the forecast from the present and the next 12 hours, there are four sets of deterministic forecasts that can become members of the TLE.

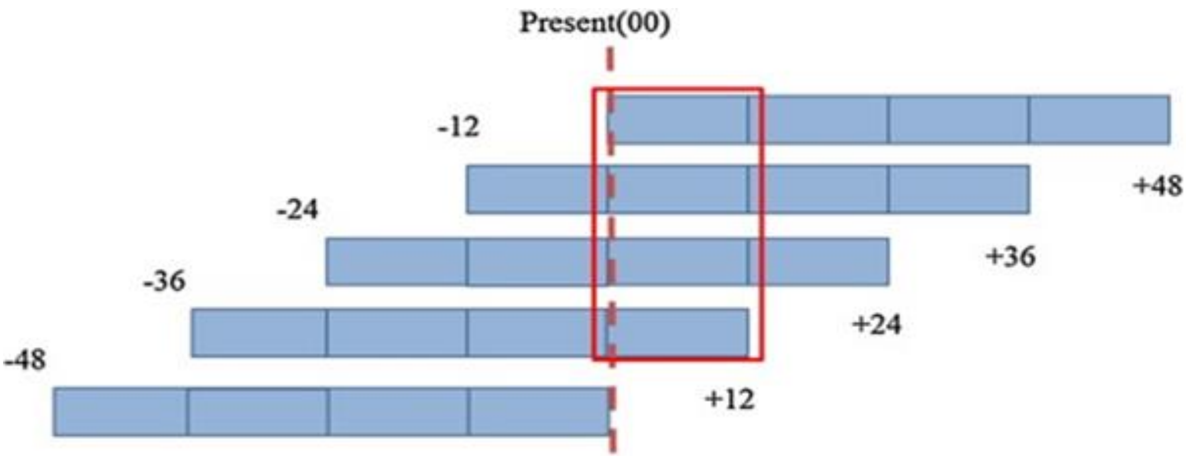


Figure 3: Time-lagged Ensemble (TLE) members come from forecasts derived from different initialization.

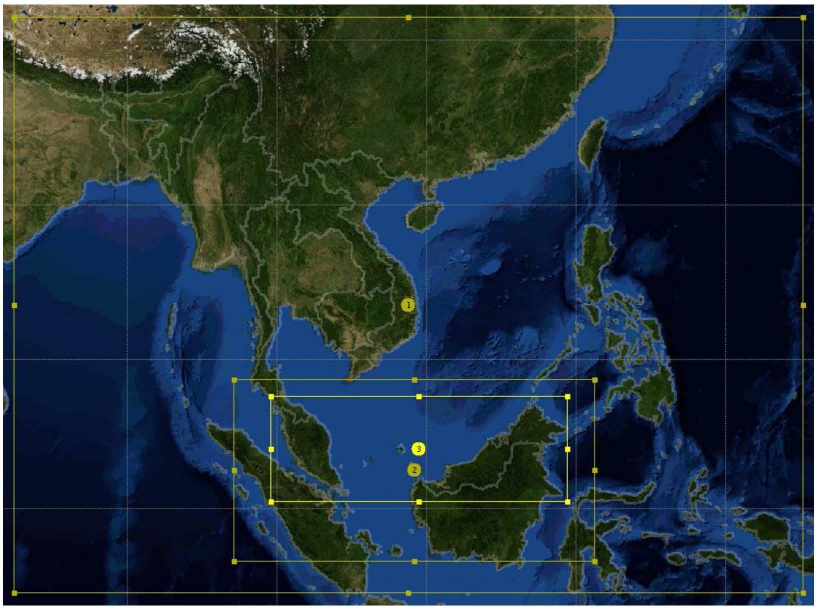
Among the meteorological centers that started using the TLE method is the European Centre for Medium-Range Weather Forecast (ECMWF), which published a technical memorandum about the implementation and results of TLE on June 1989[4]. This memorandum explains how the TLE technique is applied to a spectral NWP model with T63 resolution before the results are compared to the deterministic outputs. From the study, it is found that the NWP forecast from TLE has higher skill than the average of all deterministic forecasts and the latest deterministic forecast. However, TLE is not able to properly predict the formation of prominent weather systems such as the North Pacific Ridging. This does not mean that the NWP did not capture this event, but there were only a few of the members predicted this event, thus making TLE hard to capture it. Therefore, this finding shows that worst-case scenario judgement is still relevant when forecasting the prominent weather pattern like this.

ECMWF once again published a technical memorandum about the comparison between EPS and TLE performance in March 2008[5]. By comparing EPS generated by 51 members of low-resolution deterministic forecast and TLE generated by 6 members of high-resolution deterministic forecast, it is found that EPS forecast outperforms the TLE on each day of forecast. This means that the amount of members is important to increase the performance of the ensemble systems.

2. METHODOLOGY

For the time being, Met Malaysia is running WRF-MMD to generate weather forecasts covering Malaysia over four domains with different resolutions as shown in Figure 4. The coarsest resolution is 9km, covering the ASEAN region and up to the Himalayas Mountain ranges. Inside this largest domain is the 3km resolution domain covering both Peninsular Malaysia (PM) and Sabah/Sarawak (SS). Nested inside this is a 1km resolution domain which also covers the whole of Malaysia. The smallest domain has a 333m resolution, covering just the Klang Valley area. TLE in this study involved just forecast from domain 1km resolution.

a) Domains



b) Domains area

Domain	Lower left lat-lon	Upper right lat-lon	Resolution
1	-5.606888 , 82.27888	31.34004 , 135.1853	9km
2	-3.453232 , 97.15967	8.397476 , 121.0337	3km
3	0.5326843 , 99.55428	7.443436 , 119.3143	1km
4	2.662277 , 101.0997	3.567162 , 101.9609	333m

Figure 4: Domains selection for WRF-MMD. (a) map and (b) latitude-longitude boundaries for each domain.

WRF-MMD generates forecast four times per day, which is at 00:00, 06:00, 12:00 and 18:00 UTC cycles. For 00:00 and 12:00 UTC cycles, the forecasts are up to 7 days, while only three days for other cycles to save computing costs. For the selection of the TLE members, only forecasts above 12 hours from each cycle will be considered because it is assumed that the model was still in spin-up time to achieve statistical equilibrium[6] for forecasts below this hour. For the purpose of this study, TLE is applied to 24-hour accumulated rainfall parameters. The TLE members' selection can be understood more by referring to Figure 5. From this figure, it clearly can be seen that as forecast days increase, the TLE members will decrease.

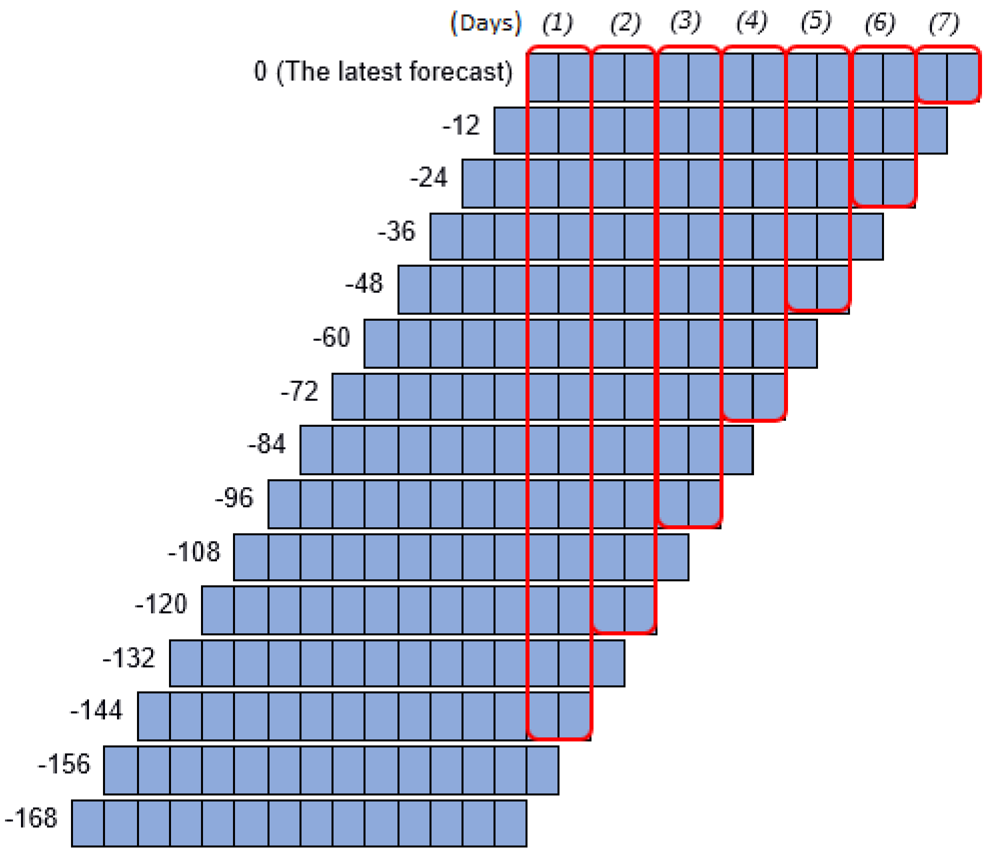


Figure 5: TLE forecasts from day 1 to 7 calculated from each deterministic member (red border rectangles).

Only one parameter is chosen for this TLE study which is 24-hour accumulated rainfall. This is because the accumulated precipitation is the most sought forecast and has vast inconsistency and high uncertainty. Before calculating the probability for the TLE, the thresholds for the accumulated rainfall must be chosen first. The thresholds are 50mm, 100mm, 150mm, 200mm and 250mm of accumulated rainfall in 24 hours. The rainfall forecast data will then be converted into binary numbers by using these thresholds. For example, using the 50mm threshold, all grids which have accumulated rainfall that is less than this threshold will be assigned as “0”, while “1” if it is equal to or more than this value. Further details about this can be referred in Figure 6.

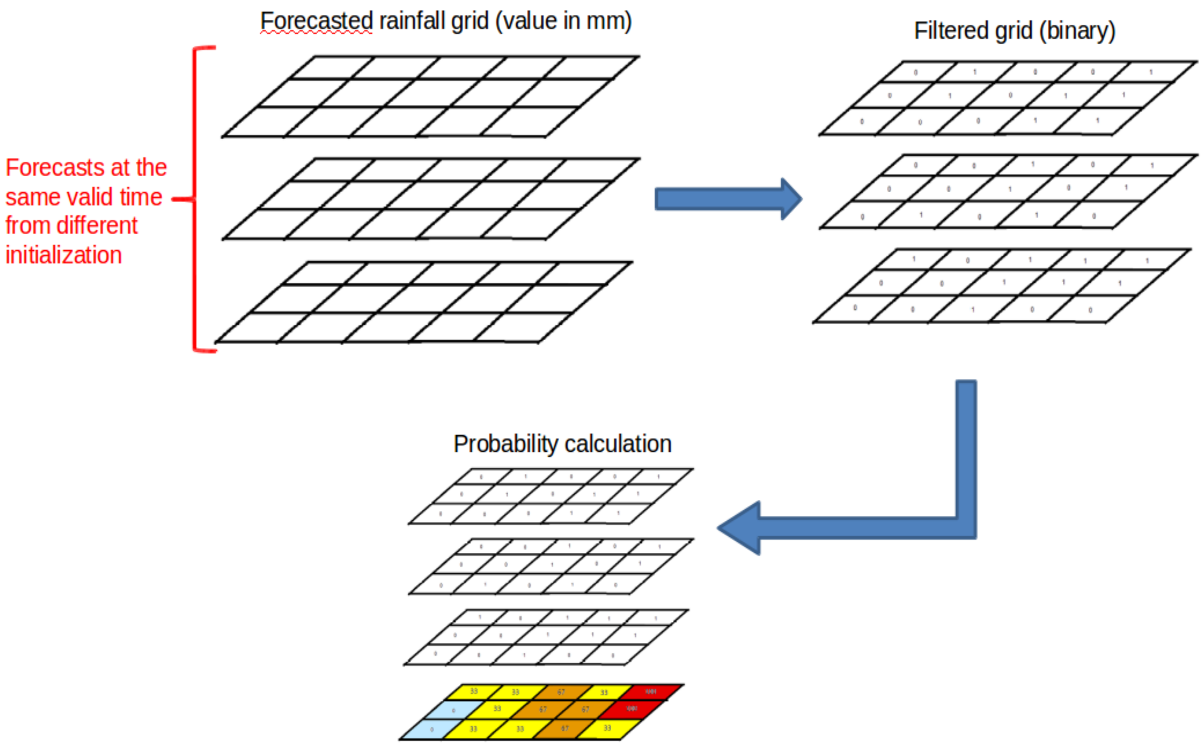


Figure 6: Conversion of precipitation data into binary values in each grid before can be calculated as probabilities value.

Before the probability value is calculated, all the forecast data with the same valid time generated from different initial data must be identified. This can be done by matching the forecast data file's name with the valid date. The standard format for the forecast data file's name for WRF-MMD is

wrfout_d*AA_YYYY-MM-DD-HH_MM_SS*

where *AA* is for domain number, *YYYY* is for a year, *MM* is for a month, *DD* is for a day, *HH* is for an hour, *MM* is for a minute and *SS* is for a second of the valid date and time. This means all the forecast data with the same valid time will have the same name and cannot been placed in the same location or directory.

The WRF-MMD outputs are in NetCDF format, which requires programming language with NetCDF library to read and convert it into other formats. One such language is FORTRAN which is especially suited for numeric computation and scientific computing[7]. APPENDIX A shows FORTRAN program to reads each WRF-MMD NetCDF files and convert the precipitation values into the binary figures before being used to calculate probability values for TLE. In this program, the NetCDF library is called at line 2 of the program while NetCDF calls that has been used is described in Table 1 below.

NetCDF calls	Descriptions	Lines
nf90_open()	Open existing NetCDF file or dataset.	42,63,83
nf90_noerr	Error status	44,66,85
nf90_inq_varid()	Get variable IDs	45,46,68,69,70,86,87,88
nf90_get_var()	Get values of variables	47,48,72,73,74,89,90,91
nf90_close()	Close NetCDF file or dataset	49,77,93,210
nf90_create()	Create NetCDF file or dataset	180
nf90_def_dim()	Define dimensions	182,183
nf90_def_var()	Define variables	191,192,193
nf90_put_att()	Define attributes	197,198,199
nf90_enddef()	Check definitions, leave define mode	201
nf90_put_var()	Provide new variable values	205,206,207

Table 1: NetCDF calls in FORTRAN to read/write NetCDF data.

To avoid filename conflict between the forecast files, the filename needs to be renamed first into other forms such as defined in line 6 in the program. From this line to 39 of the program is where the variables such as WRF-MMD input and outputs filename, rainfall forecast data, and others are defined. After that, the main program is written where the actual calculation for TLE is done.

In this program, all the filtered forecasts will be computed together to produce probability percentages. If the grids achieved 100% probability, that means all the members of TLE shown the accumulated rainfall over that grid is equal to or exceeded the threshold value. If the value is less than 100%, then there are members of TLE that predict the accumulated rainfall over the grids is less than the threshold value. The percentage calculation is used by this formula

$$p = \frac{100}{N} \sum_{i}^{N} r_i$$

where *p* is a probability in percentage, *N* is a number of TLE members, *r* is a binary value and *i* is an index for each TLE member. In this calculation, it is assumed that all forecasts can be trusted equally, thus equal weightage should be given.

There are some methods that give each member a different weightage based on the initial time of the forecasts. Mostly the methods will set larger weightage on the latest forecasts compared to the earlier forecasts[8]. However, on the contrary, equal weightages are chosen in this study. The reason that different weightages are not used in this study is simply that there are no clear indicators of which leading time of forecasts is the most accurate to the actual event, at least in this period of study. It is found that the recent forecasts are not necessarily better than previous forecasts, thus making the relationship between weightage and leading time obscure. This study is done for the North East Monsoon (NEM)

2021-2022 season because this is when heavy monsoon rainfall occurs and WRF-MMD always gives an inconsistent forecast as discussed in this study later. In a nutshell, the calculation of TLE can be summarized as follow:

1. Classification of forecasts data based on valid time to create TLE members.
2. Convert the rainfall forecasts data into binary mode based on the threshold given.
3. Calculation of probability values in percentage.

For verification purposes, two primary sources of precipitation observations data are used, which are from meteorological stations and Global Precipitation Measurement (GPM) Satellite[9]. Stations or in situ observations are regarded to have better accuracy than any other remote sensing observations, such as from the satellite. However, in situ observations lack spatial continuity, making it difficult to get the full picture of the heavy precipitation events. Therefore, the satellite data, which has better spatial coverages, need to be referred together with the in situ data for a better representation of the true events.

Three cases were selected for this verification purposes where the heavy downpours occur in November, December 2021, and February 2022. All the cases show heavy precipitation over the east coast of PM, especially Kelantan and Terengganu, where monsoon rainfall and floods are yearly events.

3. RESULTS

The chaotic nature of weather forecasting makes long-term weather predictions are unreliable and useless. Therefore, it makes sense to trust the most recent forecast more than the previous forecast. However, in reality, the NWP model is far from perfect. Besides inaccurate initial atmospheric state, imperfect representation of atmospheric processes by NWP models can also cause forecast errors that defy the general understanding. Such processes are called physical processes, which are too small to be explicitly represented dynamically by the grid model and need to be parameterized in the model by a statistical approach. Examples of these processes are shortwave and longwave radiation, cloud cover, soil-vegetation-water-atmosphere transfer, urban areas, planetary boundary layer, convection, microphysics and orographic drag[10].

In the NWP models, there are complex interactions between model dynamics and parameterization processes, as well as between parameterization processes themselves. These interactions can sometimes over-enhance the storm development in one location and suppress storm development in a nearby location. For example, storm development over the sea and nearby coastal areas where both developing storms compete for the source of moisture. Too much storm development over the sea by NWP simulation will inhibit storm development over the land, which results in lesser rainfall over the land, as depicted in Figure 7. This may be the case for the inaccurate recent forecast compared to the previous forecast, where observations usually show storms over the sea began to increase at that time.

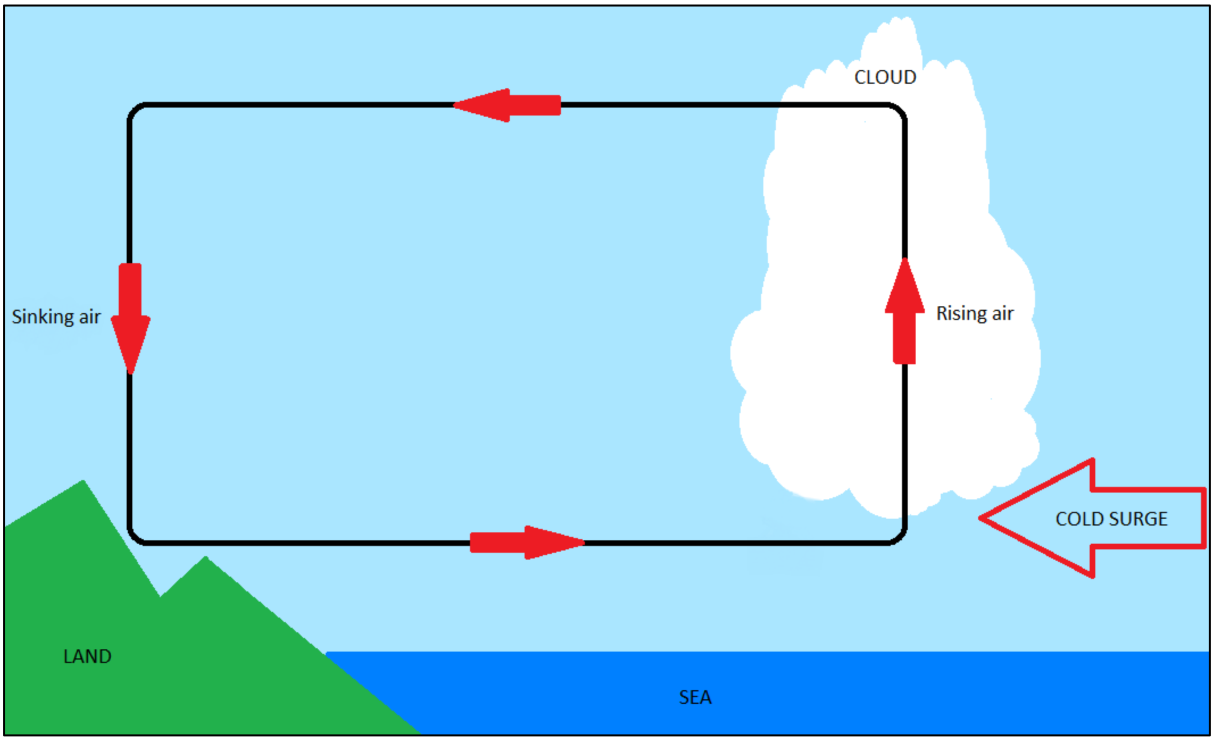


Figure 7: Enhanced circulation caused by storm development over the sea will inhibit storm development over the land in NWP simulation.

Before looking straight into the TLE results, the individual members of TLE or the deterministic forecasts must be examined first. This is to see how far the inconsistency and uncertainty are in the model forecast before applying the TLE method. Figure 8 shows 24 hours of accumulated rainfall forecasts at a valid time from 10 Nov 2021 00:00 UTC to 11 Nov 2021 00:00 UTC. Generally, most forecasts show that heavy rainfall will hit the PM's northern part of the east coast. Nevertheless, there are also forecasts that show heavy rainfall occurs only over the sea and not hitting the land. This is the case for the most recent forecast, which is from initials 24-hour before the end of the valid time. As said before, this may happen because too many storms are present over the sea, which caused the NWP model to enhance the circulations to support the systems and simultaneously inhibit storm development over the land. Due to this, users can misinterpret the model's forecasts since they usually will be more accurate when closer to the true event.

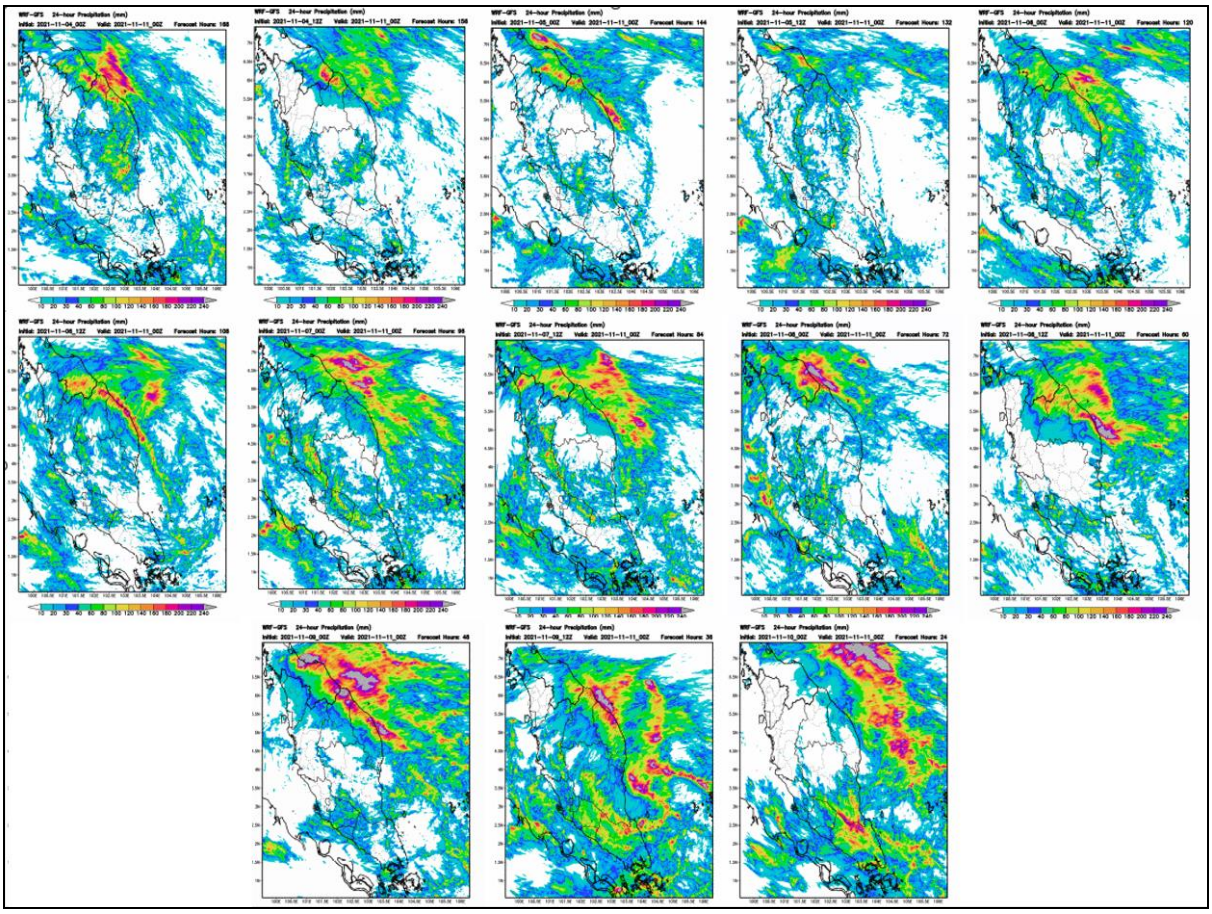


Figure 8: TLE individual member or deterministic forecasts for 24-hour accumulated rainfall valid from 10 Nov 2021 to 11 Nov 2021 (0 to 0 UTC). Initial was in order from top-left to bottom-right begin with 168-hour (top-left) to 24-hour (bottom-right) before the end of valid time.

To make the NWP model's forecasts take the uncertainty into account and have better consistency, TLE is applied to the deterministic forecasts, as shown in APPENDIX B. The forecast's valid time in this appendix is between 10 Nov 2021 00:00 UTC to 11 Nov 2021 00:00 UTC. The table also shows changes in TLE values as new members are added into the calculation for each threshold. At first, the TLE forecast showed single color presenting over 90% probability or likelihood because there was only one member available at that time which is from the initial, or forecast-hour, 144-168 hours before the valid time. The colors and likelihood ranges will increase as more members are added to the TLE calculation. These plots have ten colors representing the equal divided likelihood ranges from 0 to 100%. For thresholds above 50mm, all these probability ranges began to appear on the TLE plot at forecast-hour of 72-96 hours.

As thresholds increase, only a lower probability value can be achieved, indicating that more uncertainty in predicting the heavy precipitation compared to the light rain. For example, at the forecast-hour 72-96 hours, the highest probability on the plot for the threshold over 50mm is over 90%, but only 20% for the plot for the threshold over 250mm. However, a lower probability value for a higher threshold does not mean that the event unlikely to occurs, which will be explained later.

Next, to evaluate the TLE output, a comparison with actual observations should be made. This can be seen from the 24-hour accumulated precipitation observations plots between 10 Nov 2021 00:00 UTC to 11 Nov 2021 00:00 UTC in APPENDIX C. There are two observation plots from different sources, which are in situ and GPM satellite. In situ observation shows that the highest rainfall amount is between 200 to 225mm in 24-hour over Kota Bharu Airport station, which is located close to the coastal area of Kelantan (Figure 9). Two other stations on the east coast show 24-hour accumulated rainfall between 75 to 100mm (TUDM Gong Kedak) and 125 to 150 mm (K. Terengganu Airport). Other stations that recorded rainfall during this time are in the southern part of PM which is not significant compared to the stations on the east coast.

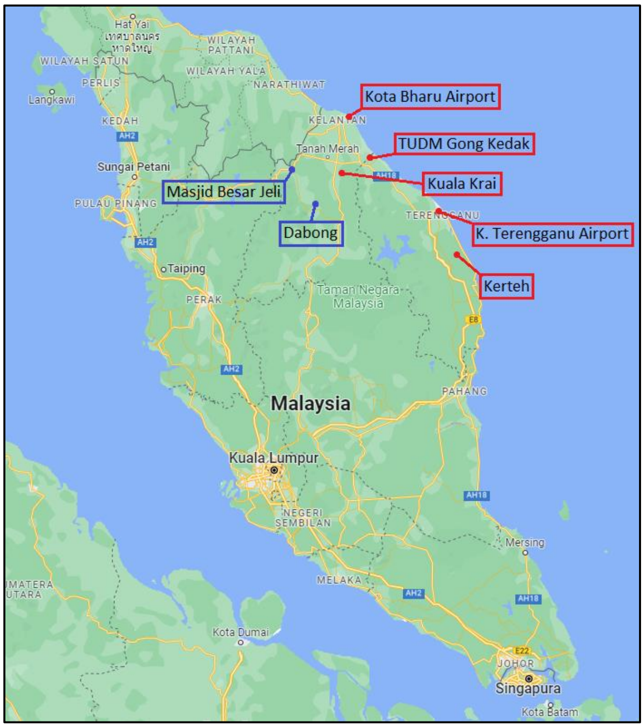


Figure 9: Location of meteorological main stations (red box) and auxiliary stations (blue box) in east coastal of Peninsular Malaysia.

In situ observations can only show points value but not spatial coverages. Because of that, remote sensing observations, such as from the GPM satellite, need to be referred together for better representations of the actual event. From the plot in APPENDIX C, the GPM plot shows heavy rainfall centered over the sea and expanded to the nearby coastal area. The amount of rainfall in 24 hours estimated by this remote sensing data over the stations on the east coast of PM is between 75 to 150 mm. This is an underestimated value, especially when compared to the Kota Bharu Airport station. Nevertheless, the conclusion from both observations is the heavy rainfall only concentrated over the coastal area of Kelantan and Terengganu and did not penetrate far inland.

After knowing the real picture, now it is time to verify the TLE plot mentioned earlier. Almost all TLE plots at an initial time between 72-96 hours before the true event show a higher probability of rainfall exceeding respective thresholds value over the coastal area of Kelantan and Terengganu, similar to the recorded observations. For instance, the TLE plot for a threshold more than 50mm shows a 90% probability covering almost the exact area as shown by observation plots. The same goes for others TLE plots with more significant thresholds. However, the probability values are much lesser than the 50mm threshold plot because intense precipitation cannot be forecast as consistent as slight precipitation.

Nevertheless, a small probability value for larger thresholds does not mean that the precipitation will not exceed the threshold, but it still can happen, as shown by observation plots. This information given by the TLE is far better if compared to the corresponding deterministic forecast, which only shows 24-hour accumulated precipitation of less than 120mm over the coastal area of Kelantan and Terengganu. By referring to the TLE at this initial, users or forecasters will be aware that heavy downpours can still happen over the coastal area even if the deterministic forecast does not predict the event.

The second heavy rainfall event selected for TLE forecast verification is between 30 Dec 2021 00:00 UTC to 31 Dec 2021 00:00 UTC (APPENDIX D and E). This time observations show downpours more concentrated inland and less rainfall over the coastal areas of Kelantan and Terengganu. Some stations in this area have recorded accumulated rainfall between 125 to 150mm during that time (Kuala Krai and Dabong). GPM plot also shows precipitation inland, thus supporting the station’s data, but estimated less rainfall. Based on these observations, it was found that WRF-MMD deterministic forecasts can predict this heavy rainfall episode. However, the inconsistency of the forecasts is still present like in the previous case. This can be seen when comparing the earlier and later forecasts, where the former is closer to the observations compared to the latter. After TLE is plotted, it is obvious that heavy precipitation has a higher chance occurs over inland rather than coastal, especially when looking at the output from forecast-hour 48-72 hours onwards.

The final date for the TLE verification is chosen on the 25th of February 2022 (UTC). By referring to APPENDIX G, the highest amount of rainfall on this day is recorded by the station in Terengganu, which is Kerteh station. The precipitation recorded in 24 hours was between 200 to 225mm. Two stations in Kelantan also recorded quite heavy rain, which is more than 150mm but less than 200mm (Kuala Krai and Masjid Besar Jeli). The GPM rainfall spatial coverage also seems to agree with the station's record. It shows heavy precipitation concentrated over inland rather than coastal areas.

WRF-MMD deterministic forecasts during this valid time show the worst consistency if compared with two previous cases, which can be referred to in APPENDIX F. Earlier forecasts show that the rainfall will hit the central and southern part of the east coast of PM. On the other hand, forecast-hour 36-60 hours onwards show that heavy precipitation will hit the northern part of the east coast of PM (Kelantan and Terengganu). These later forecasts are more accurate when compared to the observations data (APPENDIX G). This inconsistency influences the TLE's performance, where it shows a higher probability of downpours over the southern part of the east coast of PM. Nevertheless, TLE began to show more accurate predictions when forecast-hour approached 36-60 hours ahead. At this valid time, there is an obvious signal in TLE for thresholds above 250mm, where can be seen a rather large coverage of probability above 10% just over the meteorological station mentioned before (Kerteh). This corresponds to the large area of heavy rainfall over the location predicted in the deterministic forecast. This signal is very important and cannot be neglected because extreme precipitation like this is very hard to predict by WRF-MMD but can occur and has a serious impact, such as widespread flooding.

4. CONCLUSION

WRF-MMD deterministic forecasts can predict heavy rainfall as early as seven days ahead of the actual events. Still, there are inconsistencies in the output, which sometimes can mislead the users. So, the TLE method is used to make the output more consistent and trustworthy. Besides can improve the consistency of the forecast, TLE also requires only small computing resources. Nevertheless, the real challenge of TLE is to improve the probability for larger rainfall thresholds, in this case, more than 100mm of precipitation in 24 hours. This only can be achieved if the sample size is big enough and only can be obtained by using the intensive computing EPS method.

The TLE can give better consistency in WRF-MMD rainfall prediction, unlike deterministic forecasts, which always give different results in each initialization. However, the output is in probability values and needs to be divided into several thresholds. In this study, the chosen thresholds are 50, 100, 150, 200 and 250 mm of 24-hour accumulated rainfall. The probability for rainfall over 50mm in 24 hours can easily achieve over 90% but become lesser for more significant thresholds. Nevertheless, the lower probability value for higher thresholds does not mean the rainfall cannot exceed the threshold. For example, in this study, the probability value of 20% for a threshold above 200 mm can be considered a high chance of occurrence.

APPENDIX A

Fortran program for calculating the probability values in TLE method

```
program lag_ensemble
  use netcdf
  implicit none

  ! This is the name of the data file we will read.
  character (len = *), parameter :: WRFOUTC ="wrfoutc", WRFOUTB = "wrfoutb"
  character (len = *), parameter :: EXT_NAME =".nc"
  character (len = *), parameter :: FMT1 ="(A7,I0.2,A3)"

  character (len = 12) :: FILENAMEC, FILENAMEB

  integer, parameter :: FILE_NUM = 28

  integer :: i, j
  integer :: statc, statb, statll

  ! This will be the netCDF ID for the file and data variable.
  integer :: ncid2, varidrainnc2, varidrainc2, varidrainsh2
  integer :: ncid1, varidrainnc1, varidrainc1, varidrainsh1
  integer :: ncidll, varidlat, varidlon

  ! We are reading 2D data, a 6 x 12 grid.
  integer, parameter :: NX = 2196, NY = 771
!  integer, parameter :: NDIMS = 2
  real :: data_inrainnc2(NX, NY), data_inrainc2(NX, NY), data_inrainsh2(NX, NY)
  real :: data_inrainnc1(NX, NY), data_inrainc1(NX, NY), data_inrainsh1(NX, NY)
  real :: data_m(NX, NY), data_lat(NX, NY), data_lon(NX, NY)
  real :: data_rain2(NX, NY), data_rain1(NX, NY), data_diff(NX, NY)
  real :: lat1d(NY), lon1d(NX)

  integer :: data_thresh(NX, NY), data_thresh_acc(NX, NY)
  real :: thresh
  character (len=32) :: arg

  CALL get_command_argument(1, arg)
  READ (arg, '(F5.0)') thresh
  print *,thresh

  data_thresh_acc = 0

  ! read latitude & longitude first
```

```
 statll = nf90_open("wrfoutc01.nc", NF90_NOWRITE, ncidll)
 call check(statll)
 if(statll == nf90_noerr) then
        call check( nf90_inq_varid(ncidll, "XLAT", varidlat) )
        call check( nf90_inq_varid(ncidll, "XLONG", varidlon) )
        call check( nf90_get_var(ncidll, varidlat, data_lat) )
        call check( nf90_get_var(ncidll, varidlon, data_lon) )
        call check( nf90_close(ncidll) )
        call lat1d_conv(data_lat,lat1d)
        call lon1d_conv(data_lon,lon1d)
!         call lat1d_conv(data_lat,lat1d)
 end if

 j=0

 ! read the data
 do i = 1, FILE_NUM
      write(FILENAMEC, FMT1), WRFOUTC, i , EXT_NAME
      print *,FILENAMEC
 ! Open the file. NF90_NOWRITE tells netCDF we want read-only access to
 ! the file.
      statc = nf90_open(FILENAMEC, NF90_NOWRITE, ncid2)
 !     print *,stat
      call check(statc)
      if(statc == nf90_noerr) then
 ! Get the varid of the data variable, based on its name.
        call check( nf90_inq_varid(ncid2, "RAINNC", varidrainnc2) )
        call check( nf90_inq_varid(ncid2, "RAINC", varidrainc2) )
        call check( nf90_inq_varid(ncid2, "RAINSH", varidrainsh2) )
 ! Read the data.
        call check( nf90_get_var(ncid2, varidrainnc2, data_inrainnc2) )
        call check( nf90_get_var(ncid2, varidrainc2, data_inrainc2) )
        call check( nf90_get_var(ncid2, varidrainsh2, data_inrainsh2) )
 ! Close the file, freeing all resources.
        print *, ncid2, "Closing nc file"
        call check( nf90_close(ncid2) )
        data_rain2 = data_inrainnc2 + data_inrainc2 + data_inrainsh2
      end if
 ! Previous file
      write(FILENAMEB, FMT1), WRFOUTB, i , EXT_NAME
      print *,FILENAMEB
      statb = nf90_open(FILENAMEB, NF90_NOWRITE, ncid1)
      call check(statb)
      if(statc == nf90_noerr) then
        call check( nf90_inq_varid(ncid1, "RAINNC", varidrainnc1) )
        call check( nf90_inq_varid(ncid1, "RAINC", varidrainc1) )
```

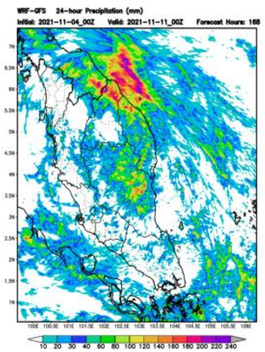
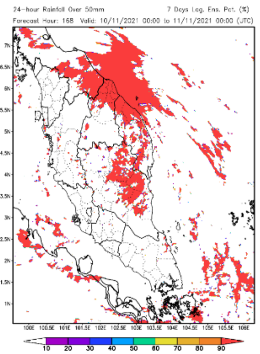
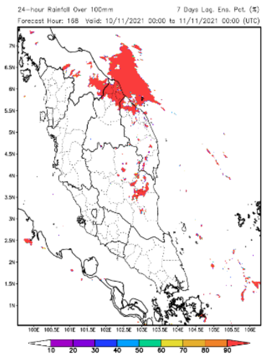
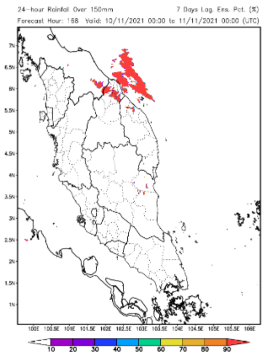
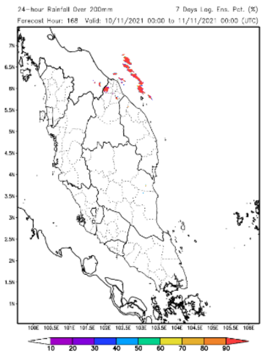
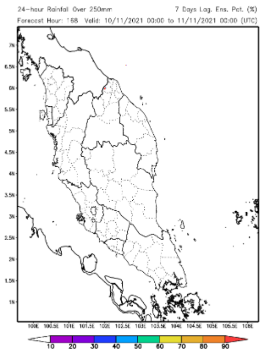
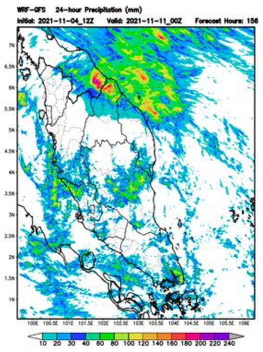
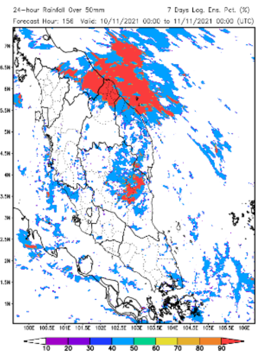
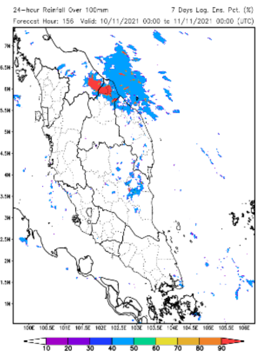
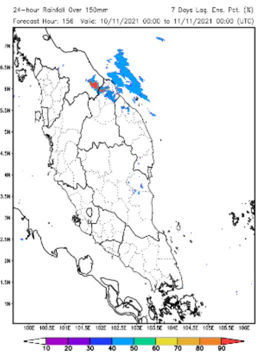
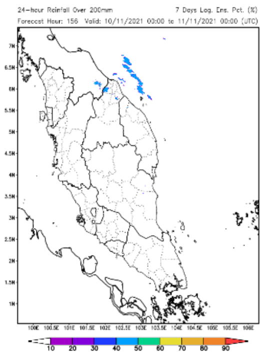
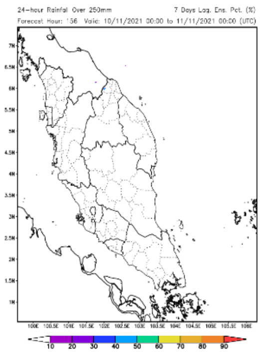
```
        call check( nf90_inq_varid(ncid1, "RAINSH", varidrainsh1) )
        call check( nf90_get_var(ncid1, varidrainnc1, data_inrainnc1) )
        call check( nf90_get_var(ncid1, varidrainc1, data_inrainc1) )
        call check( nf90_get_var(ncid1, varidrainsh1, data_inrainsh1) )
        print *, ncid1, "Closing nc file"
        call check( nf90_close(ncid1) )
 ! Calculate accumulated rainfall
        data_rain1 = data_inrainnc1 + data_inrainc1 + data_inrainsh1
        data_diff = data_rain2 - data_rain1
        call setthresh(data_diff,data_thresh,thresh)
        call percentage(data_thresh_acc, data_thresh)
        j=j+1
!        call writeout(lat1d,lon1d,data_diff,i)
!        call writeout(lat1d,lon1d,data_thresh_acc*1.0,i)
      end if
 end do
!  call writeout(lat1d,lon1d,data_tot*1.0,i)
 call writeout(lat1d,lon1d,data_thresh_acc/(j*1.0)*100.0,int(thresh))

contains
 subroutine check(status)
   integer, intent (in) :: status

   if(status /= nf90_noerr) then
     print *, trim(nf90_strerror(status))
 !    stop "Stopped"
   end if
 end subroutine check
 subroutine lat1d_conv(lat2d,lat1d)
      real, intent (in) :: lat2d(NX,NY)
      integer :: k
      real, intent (out) :: lat1d(NY)
      do k=1, NY
        lat1d(k) = lat2d(1,k)
      end do
      return
 end subroutine lat1d_conv
 subroutine lon1d_conv(lon2d,lon1d)
      real, intent (in) :: lon2d(NX,NY)
      integer :: k
      real, intent (out) :: lon1d(NX)
      do k=1, NX
        lon1d(k) = lon2d(k,1)
      end do
      return
 end subroutine lon1d_conv
```

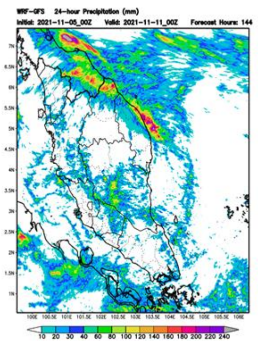
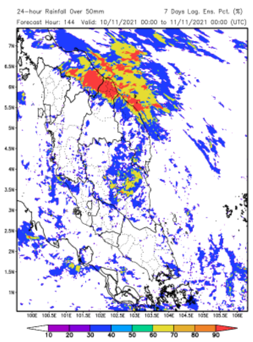
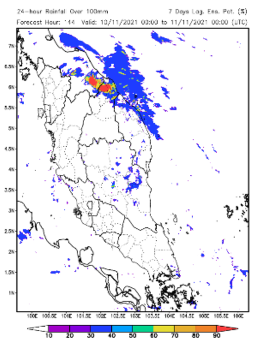
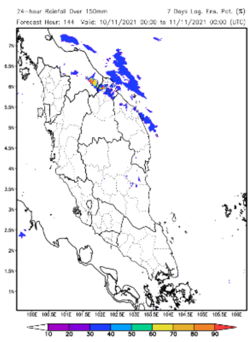
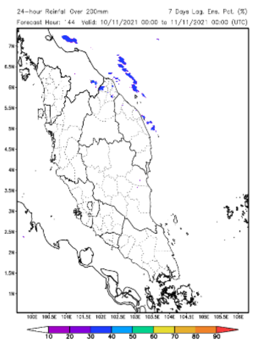
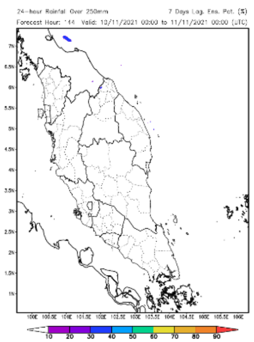
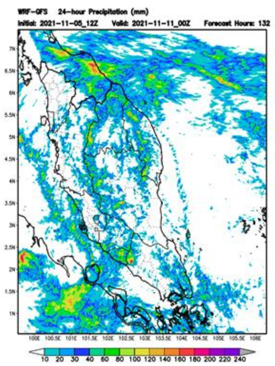
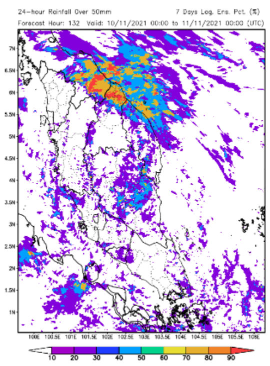
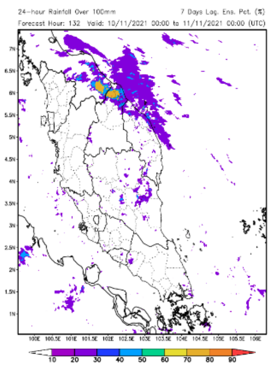
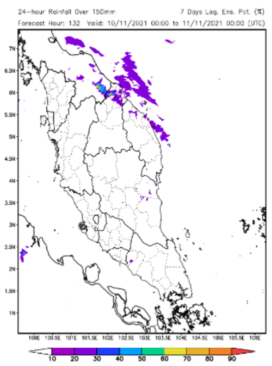
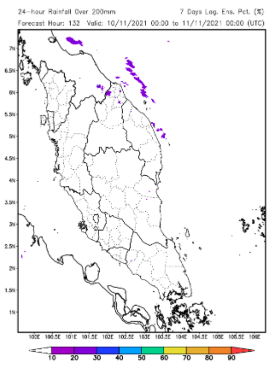
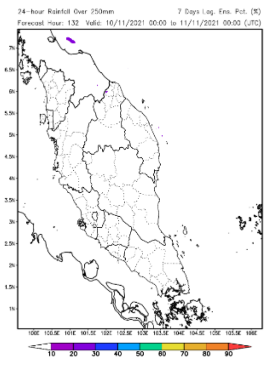
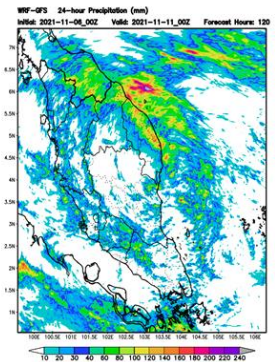
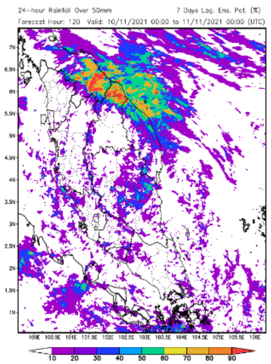
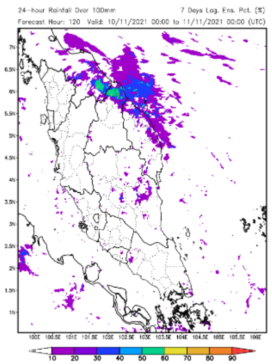
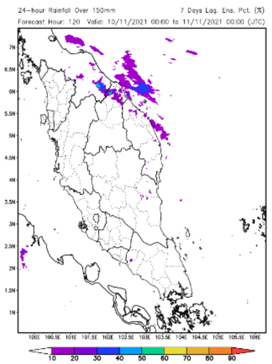
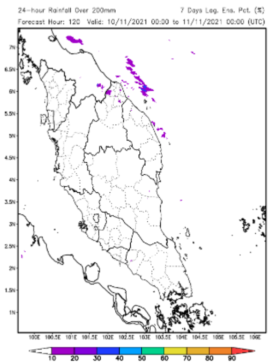
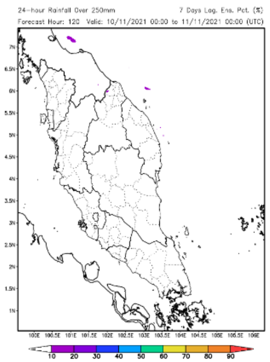
```
subroutine setthresh(datain,dataout,thresh)
    real, intent (in) :: thresh
    real, intent (in) :: datain(NX,NY)
    integer, intent (out) :: dataout(NX,NY)
    integer :: j,k
    do j=1, NX
      do k=1, NY
        if (datain(j,k) .gt. thresh) then
          dataout(j,k) = 1
        else
          dataout(j,k) = 0
        end if
      end do
    end do
    return
end subroutine setthresh
subroutine percentage(data1, data2)
    integer, intent (in) :: data2(NX, NY)
    integer, intent (inout) :: data1(NX, NY)

    data1 = data1 + data2
    Return
end subroutine
subroutine writeout(lat,lon,dataout,i)
! subroutine writeout(lat,lon,dataout)
  character (len = *), parameter :: OUTFILE ="lagens"
  character (len = *), parameter :: FMT1 ="(A6,I0.3,A3)"
  integer, parameter :: NDIMS = 2
  real, intent (in) :: lat(NY), lon(NX) , dataout(NX,NY)
!   integer, intent (in) :: dataout(NX,NY)
  integer, intent (in) :: i
  integer :: ncido
  character (len = 12) :: OUTFILENAME
  integer :: x_dimid, y_dimid, dimids(NDIMS)
  integer :: varido, varidlato, varidlono
 !Attribute
  character (len = *), parameter :: UNITS = "units"
  character (len = *), parameter :: LAT_UNITS = "degrees_north"
  character (len = *), parameter :: LON_UNITS = "degrees_east"

 ! print *, i

  write(OUTFILENAME, FMT1), OUTFILE, i , ".nc"
!   write(OUTFILENAME, FMT1), OUTFILE, ".nc"
 ! Create the netCDF file. The nf90_clobber parameter tells netCDF to
 ! overwrite this file, if it already exists.
```

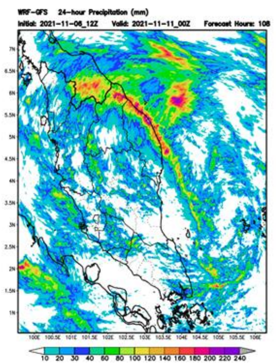
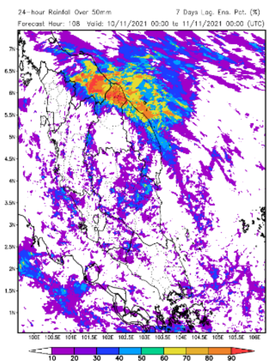
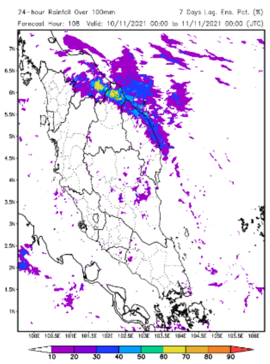
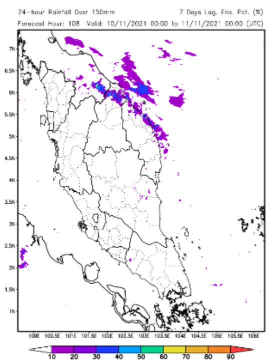
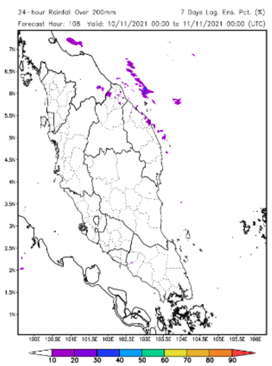
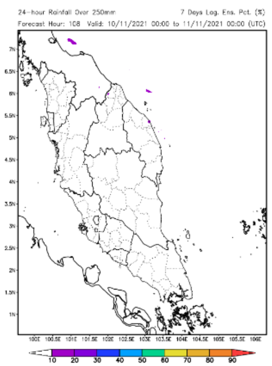
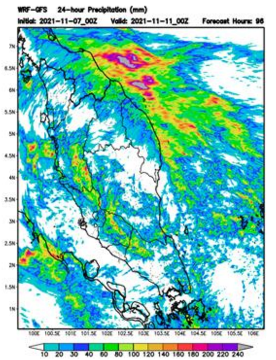
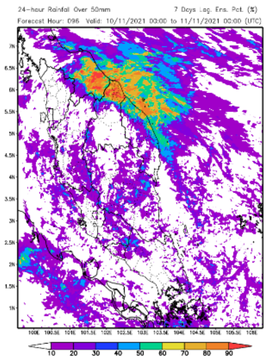
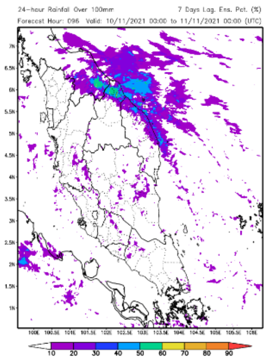
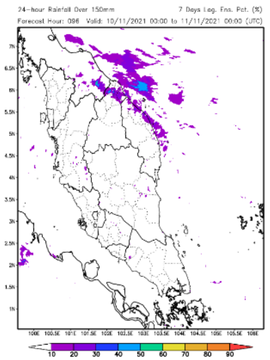
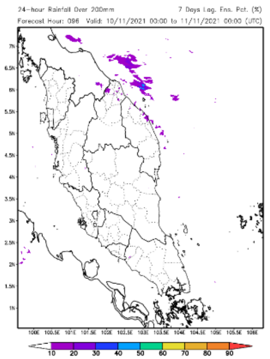
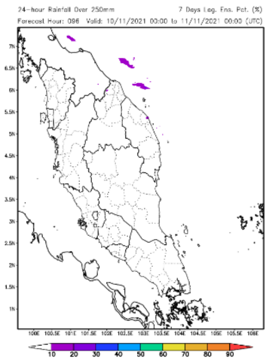
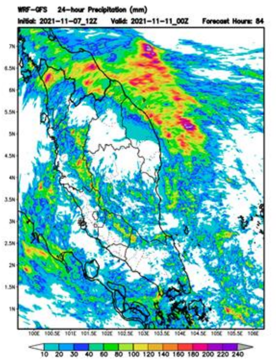
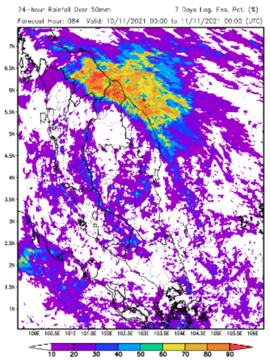
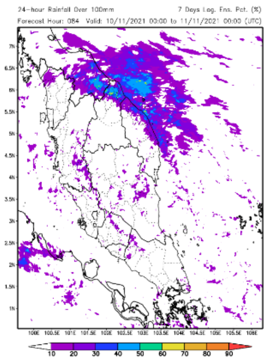
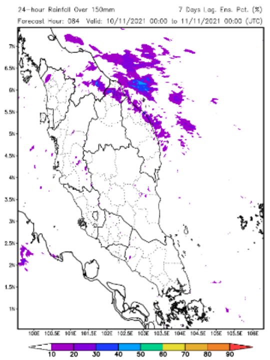
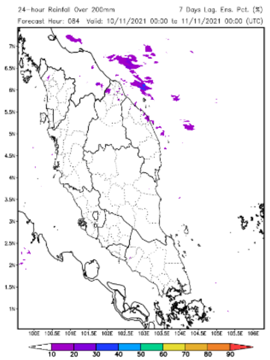
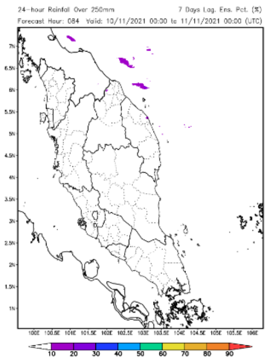
```
    call check( nf90_create(OUTFILENAME, NF90_CLOBBER, ncido) )
  ! Define the dimensions. NetCDF will hand back an ID for each.
    call check( nf90_def_dim(ncido, "longitude", NX, x_dimid) )
    call check( nf90_def_dim(ncido, "latitude", NY, y_dimid) )
  ! The dimids array is used to pass the IDs of the dimensions of
  ! the variables. Note that in fortran arrays are stored in
  ! column-major format.
    dimids =  (/ x_dimid, y_dimid /)
  ! Define the variable. The type of the variable in this case is
  ! NF90_INT (4-byte integer).
!    call check( nf90_def_var(ncido, "ensemble_percentage", NF90_INT, dimids, varido) )
    call check( nf90_def_var(ncido, "lag_ensemble", NF90_FLOAT, dimids, varido) )
    call check( nf90_def_var(ncido, "latitude", NF90_FLOAT, y_dimid, varidlato) )
    call check( nf90_def_var(ncido, "longitude", NF90_FLOAT, x_dimid, varidlono) )
  ! Assign units attributes to coordinate var data. This attaches a
  ! text attribute to each of the coordinate variables, containing the
  ! units.
    call check( nf90_put_att(ncido, varidlato, UNITS, LAT_UNITS) )
    call check( nf90_put_att(ncido, varidlono, UNITS, LON_UNITS) )
    call check( nf90_put_att(ncido, varido, UNITS, "percentage") )
  ! End define mode. This tells netCDF we are done defining metadata.
    call check( nf90_enddef(ncido) )
  ! Write the pretend data to the file. Although netCDF supports
  ! reading and writing subsets of data, in this case we write all the
  ! data in one operation.
    call check( nf90_put_var(ncido, varido, dataout) )
    call check( nf90_put_var(ncido, varidlato, lat) )
    call check( nf90_put_var(ncido, varidlono, lon) )
  ! Close the file. This frees up any internal netCDF resources
  ! associated with the file, and flushes any buffers.
    call check( nf90_close(ncido) )
  end subroutine writeout
end program lag_ensemble
```

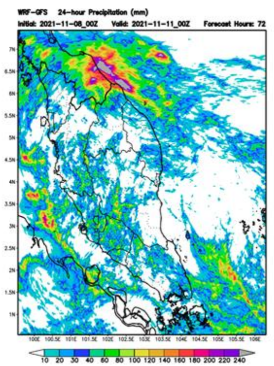
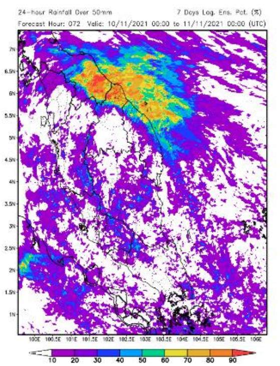
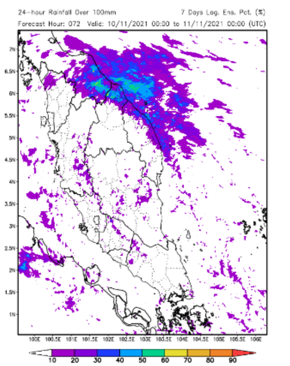
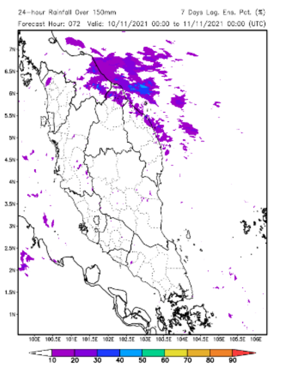
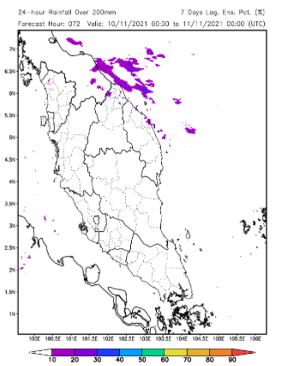
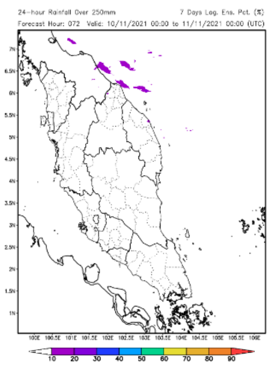
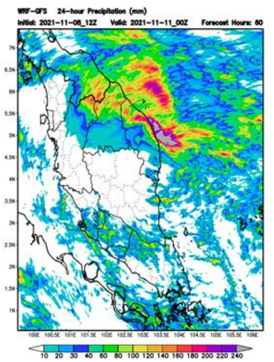
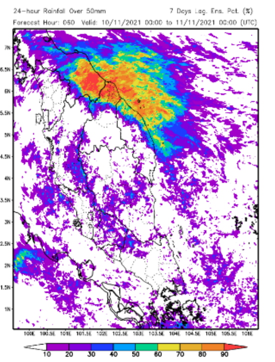
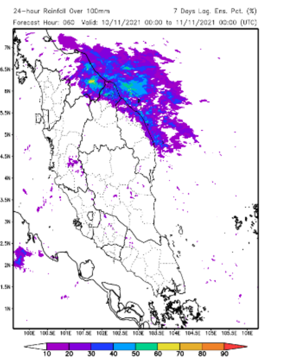
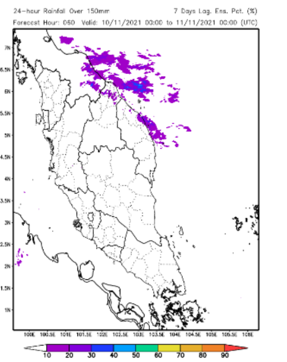
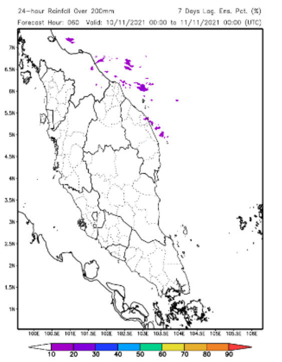
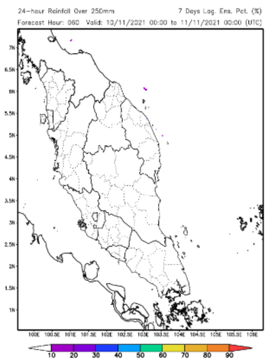
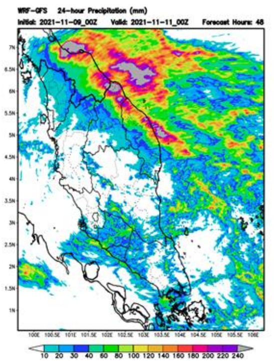
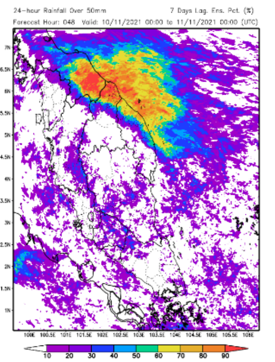
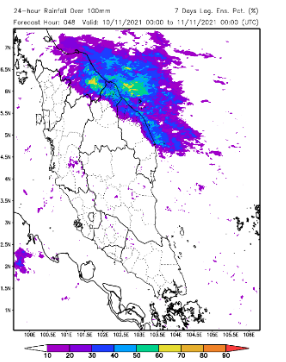
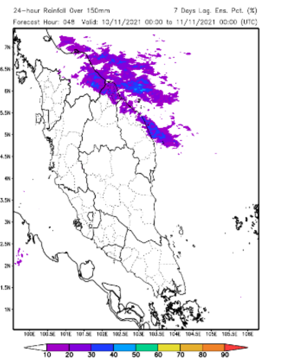
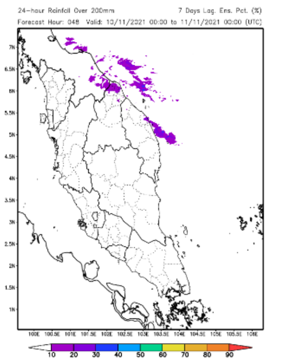
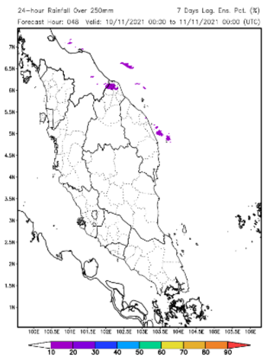
APPENDIX B

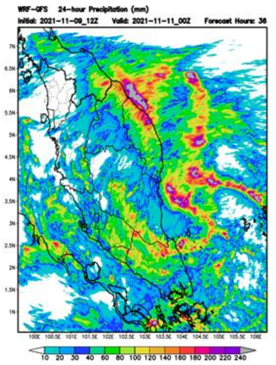
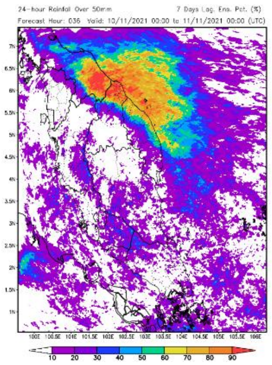
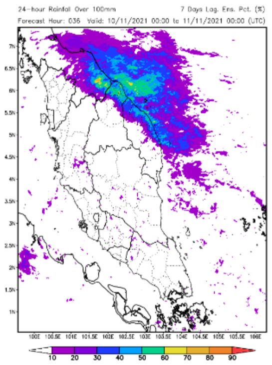
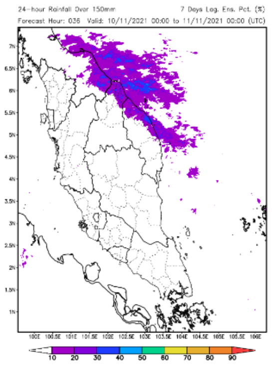
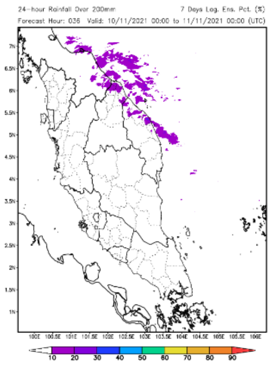
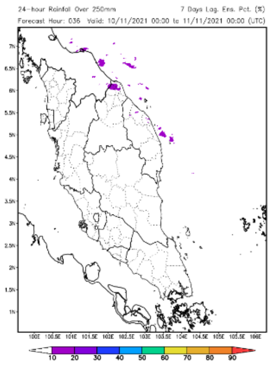
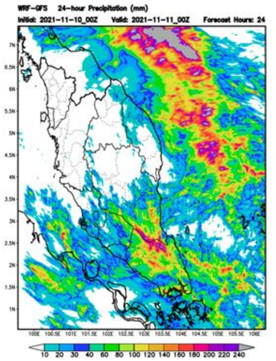
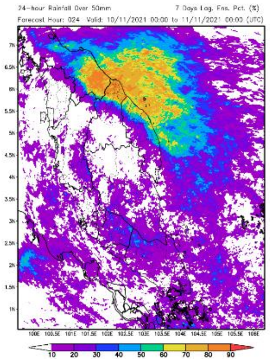
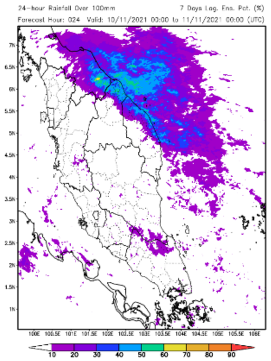
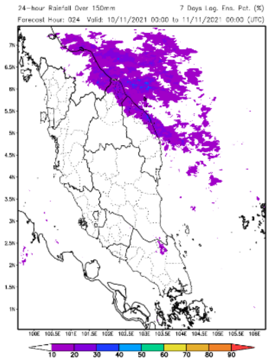
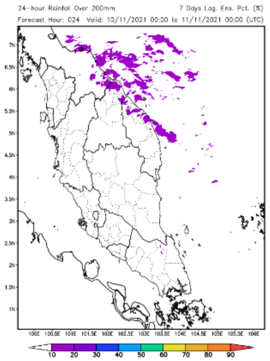
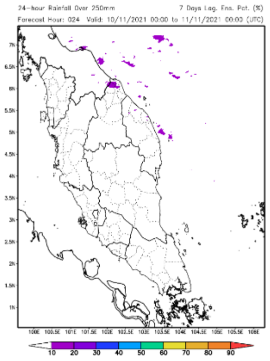
Deterministic (precipitation amount in mm) and Time-Lagged Ensemble (probability value in %) forecasts for 24-hour accumulated precipitation on 10th Nov 2021 (UTC)

Forecast-hour (hours)	Deterministic NWP forecast	TLE NWP forecast: Threshold 50mm and above	TLE NWP forecast: Threshold 100mm and above	TLE NWP forecast: Threshold 150mm and above	TLE NWP forecast: Threshold 200mm and above	TLE NWP forecast: Threshold 250mm and above
144-168						
132-156						

120-144						
108-132						
96-120						

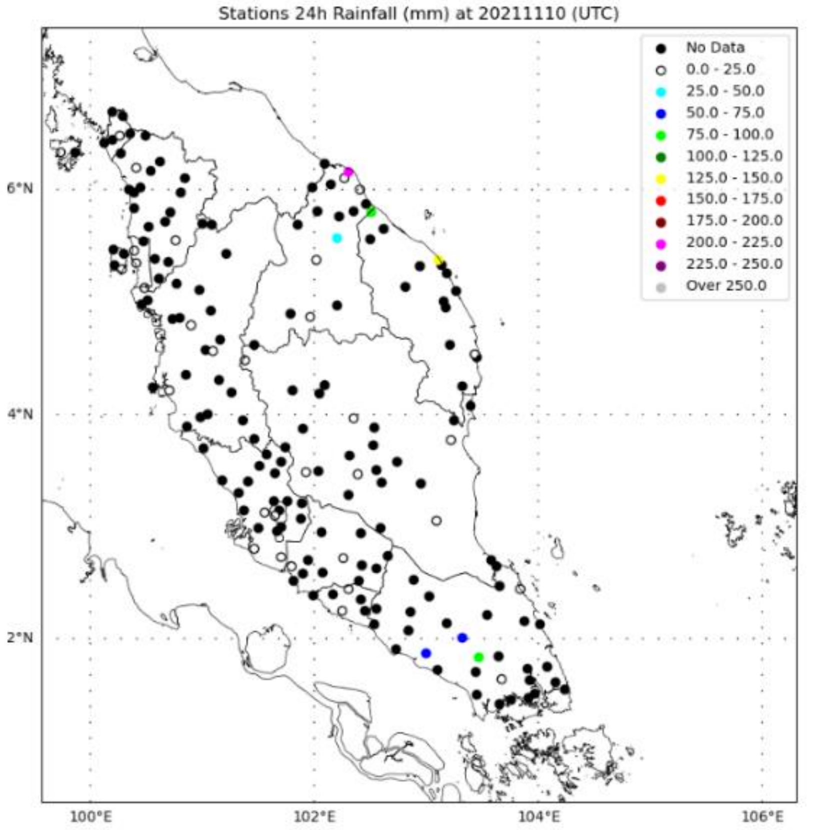
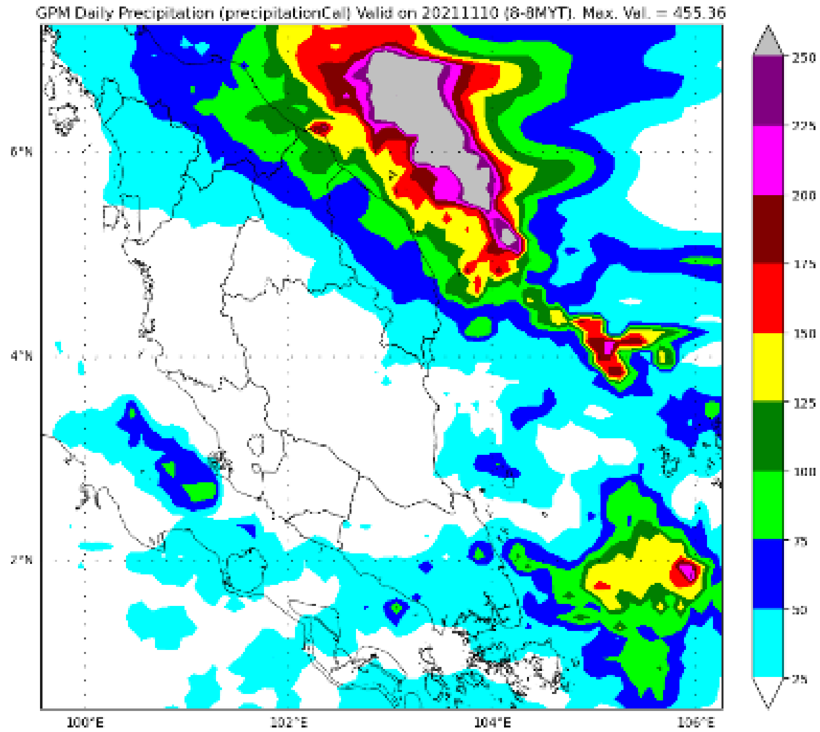
84-108						
72-96						
60-84						

48-72						
36-60						
24-48						

12-36						
0-24						

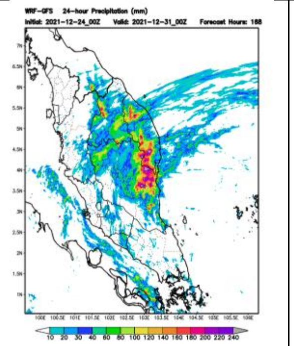
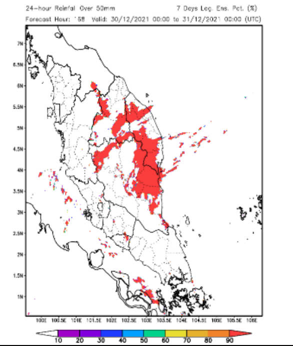
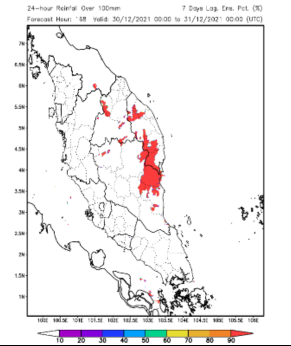
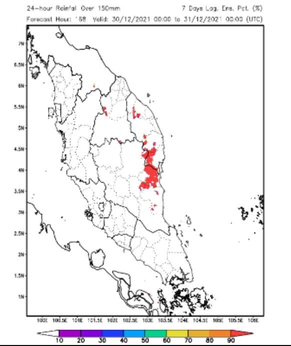
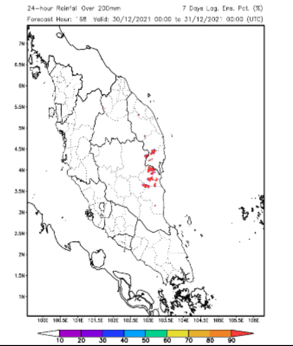
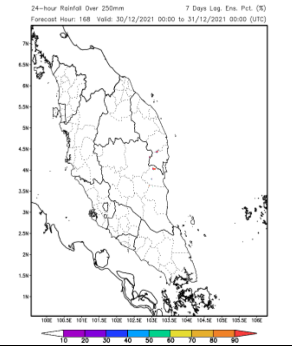
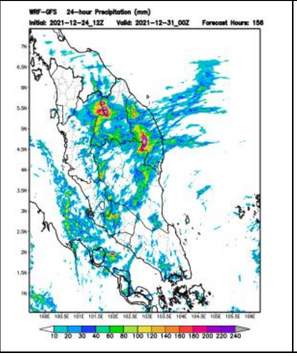
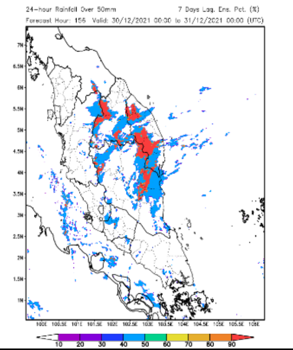
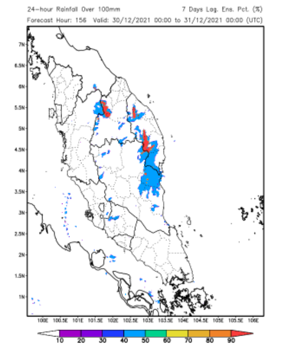
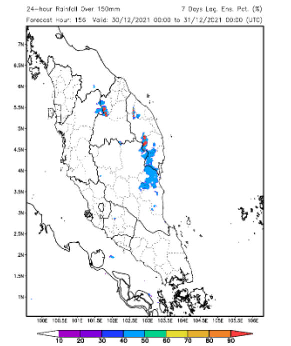
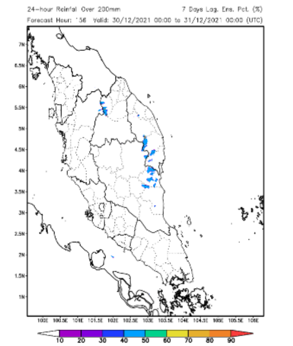
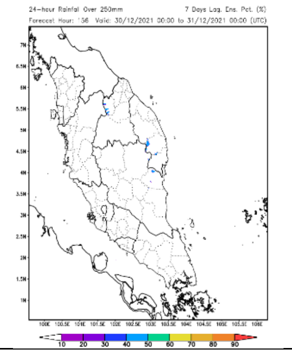
APPENDIX C

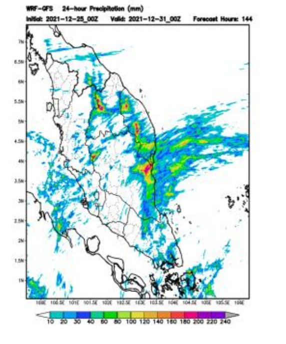
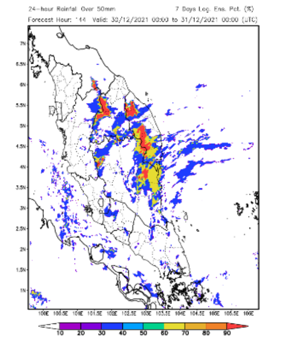
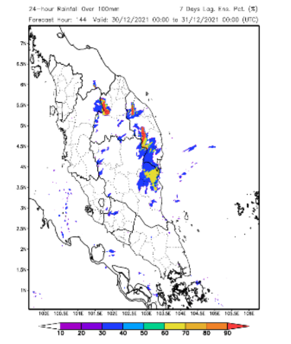
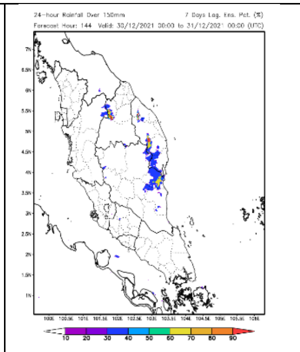
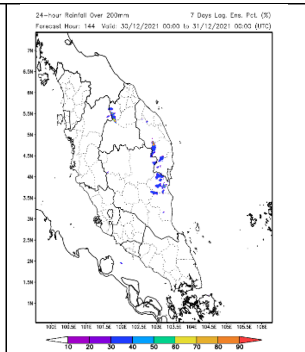
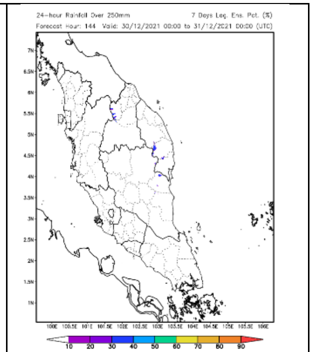
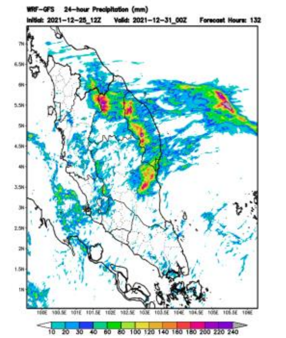
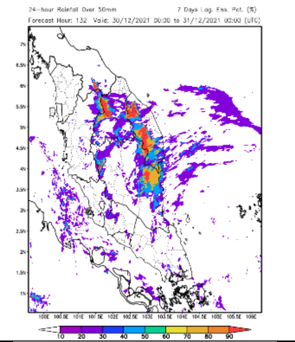
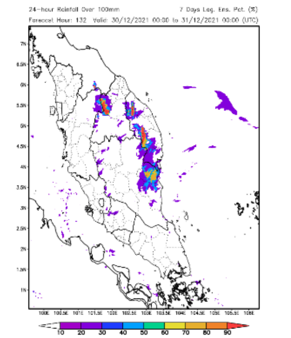
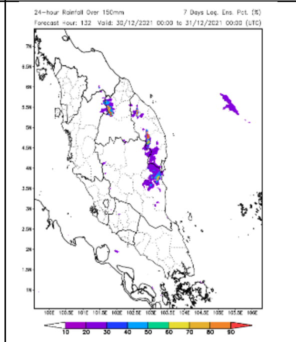
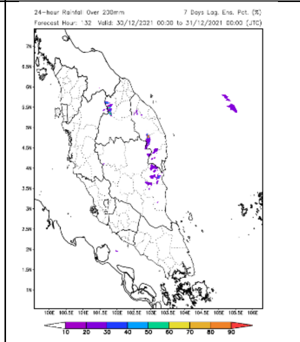
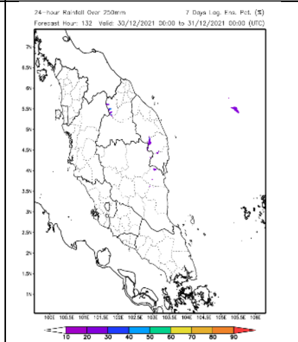
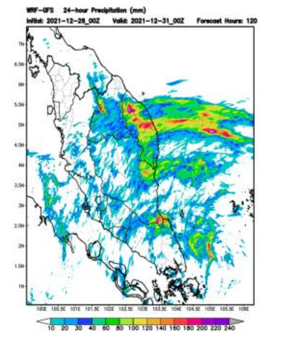
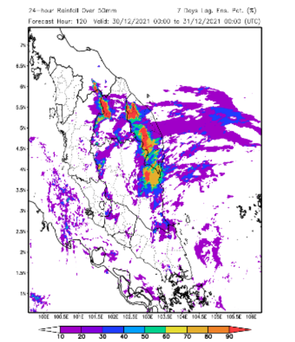
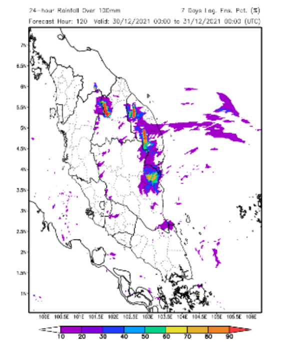
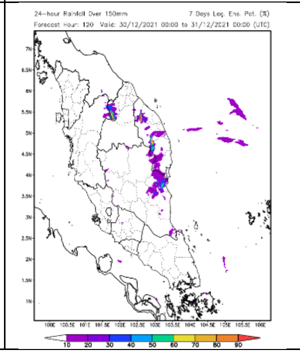
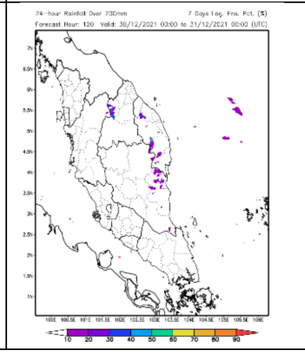
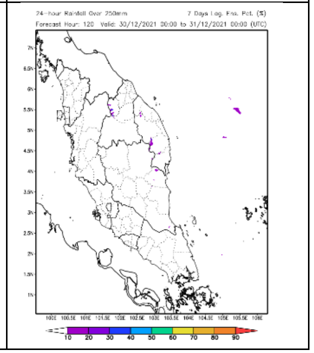
Observation of 24-hour accumulated precipitation plotted from meteorological stations and Global Precipitation Measurement (GPM) satellite (precipitation in mm) on 10th Nov 2021 (UTC)

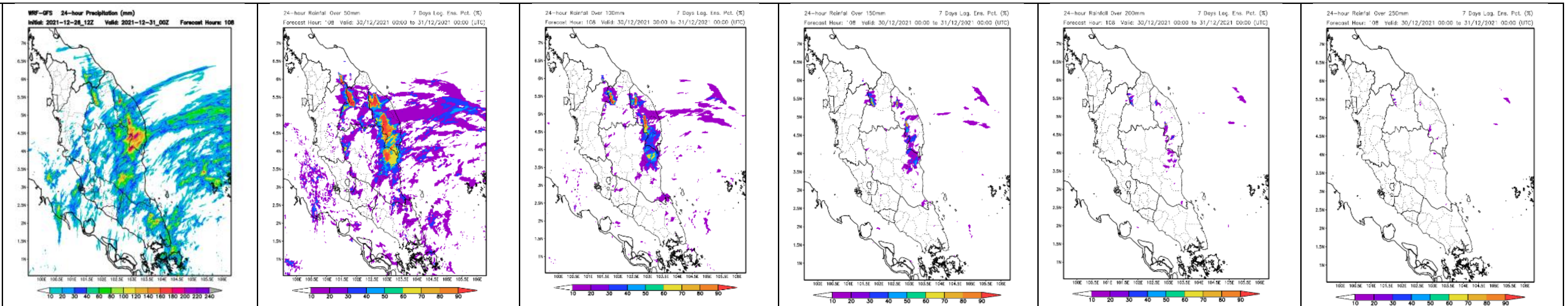
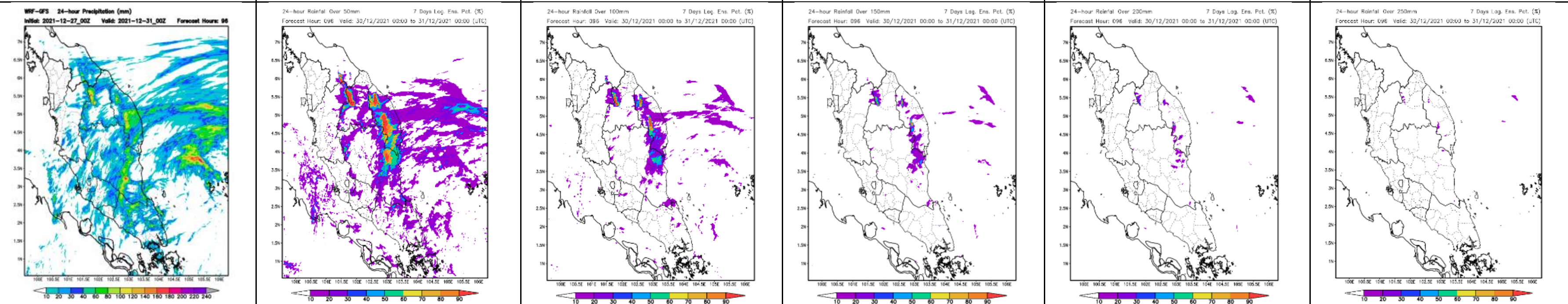
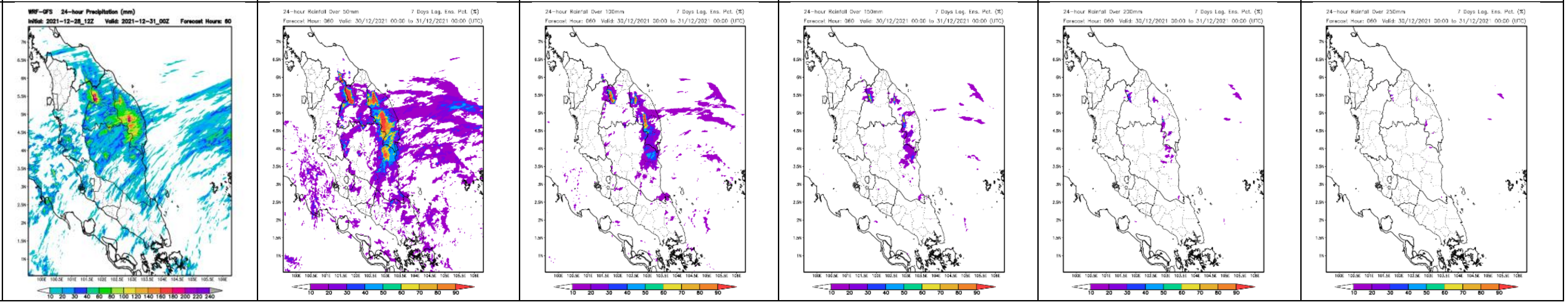
Date (UTC)	STATIONS	GPM
10/11/2021		

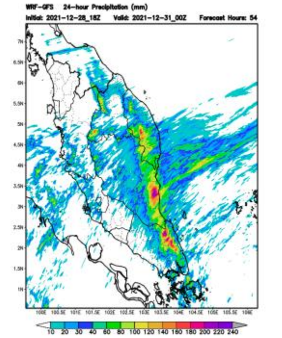
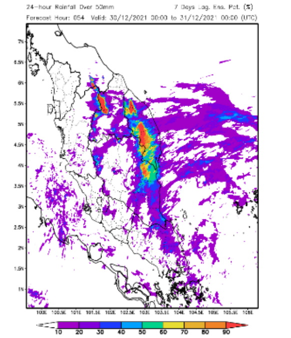
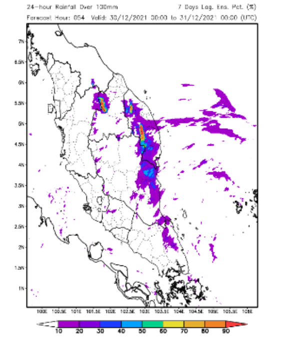
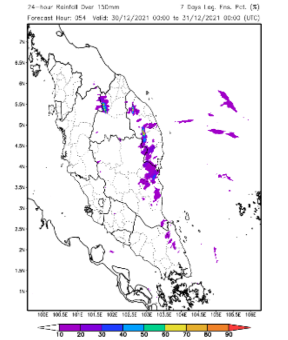
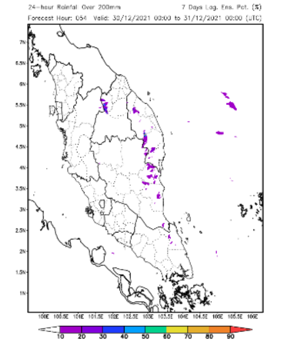
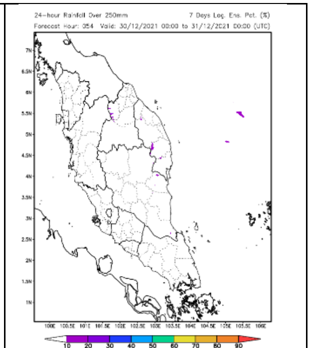
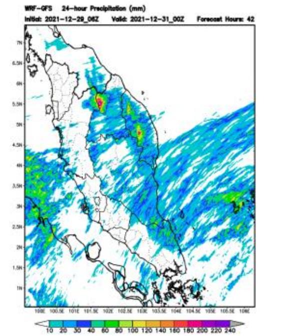
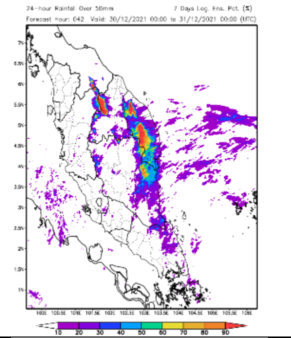
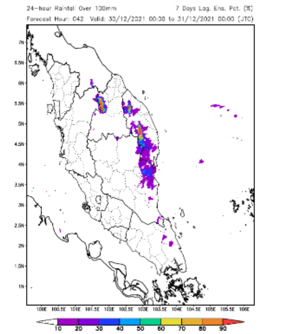
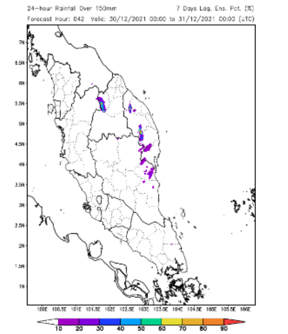
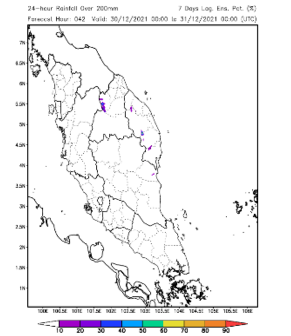
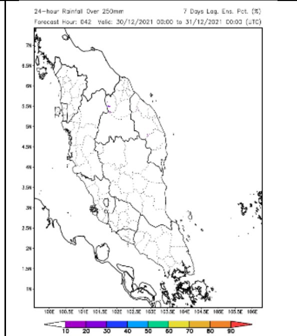
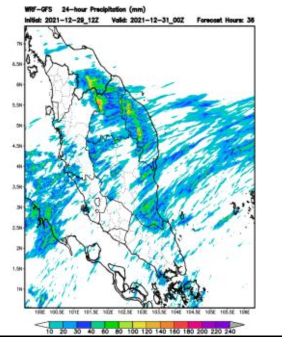
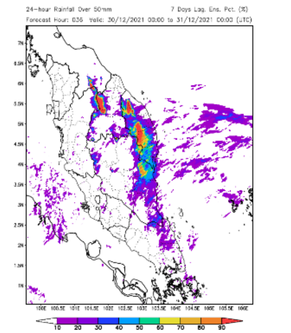
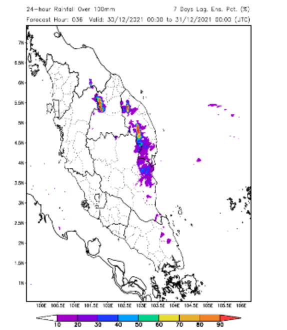
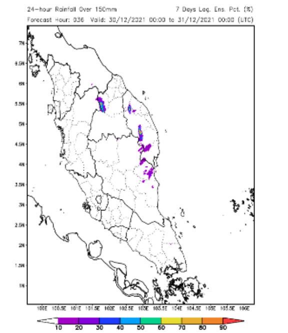
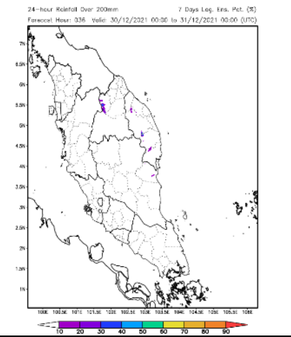
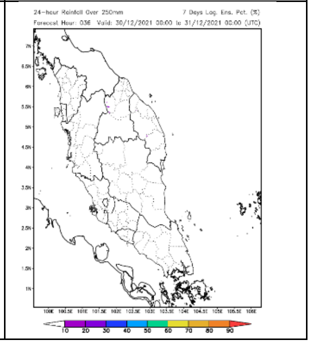
APPENDIX D

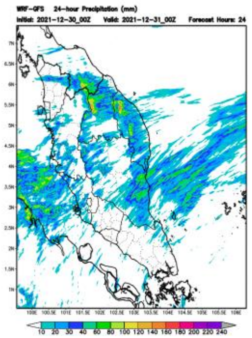
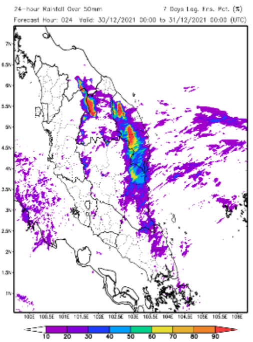
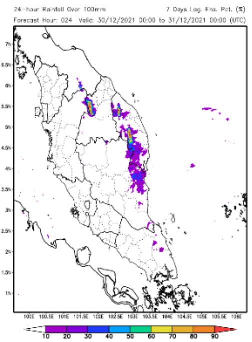
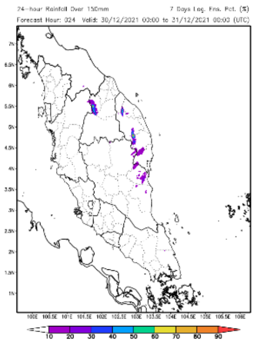
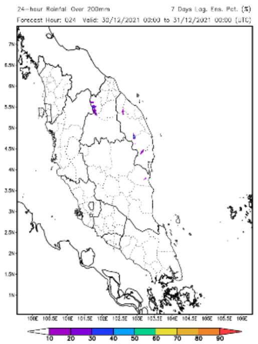
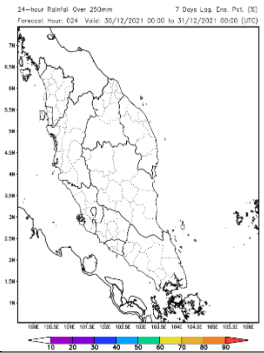
Deterministic (precipitation amount in mm) and Time-Lagged Ensemble (probability value in %) forecasts for 24-hour accumulated precipitation on 30th Dec 2021 (UTC)

Forecast-hour (hours)	Deterministic NWP forecast	TLE NWP forecast: Threshold 50mm and above	TLE NWP forecast: Threshold 100mm and above	TLE NWP forecast: Threshold 150mm and above	TLE NWP forecast: Threshold 200mm and above	TLE NWP forecast: Threshold 250mm and above
144-168						
132-156						

120-144						
108-132						
96-120						

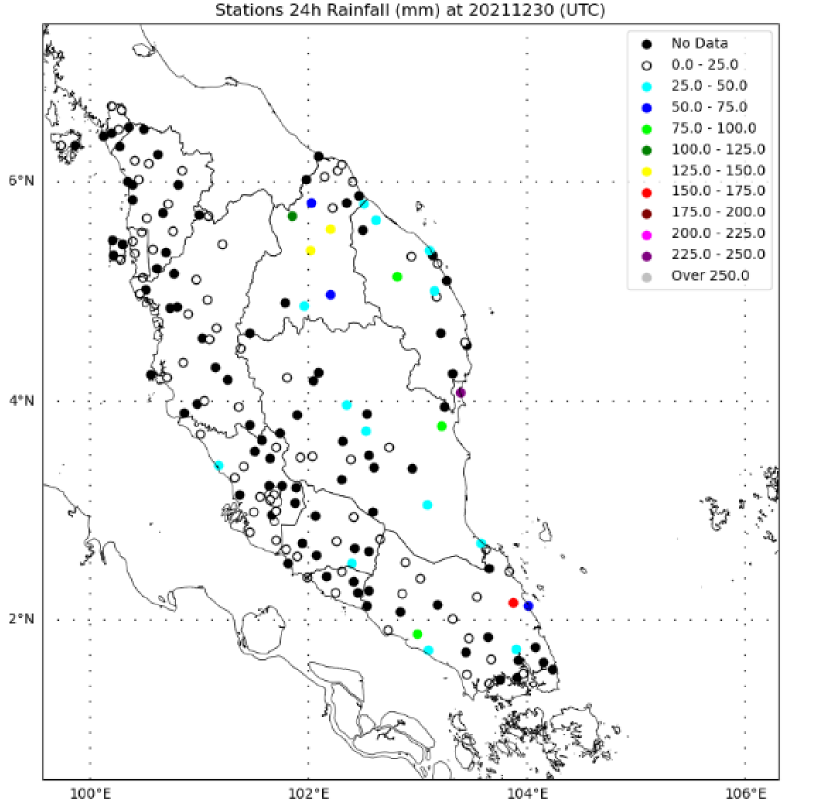
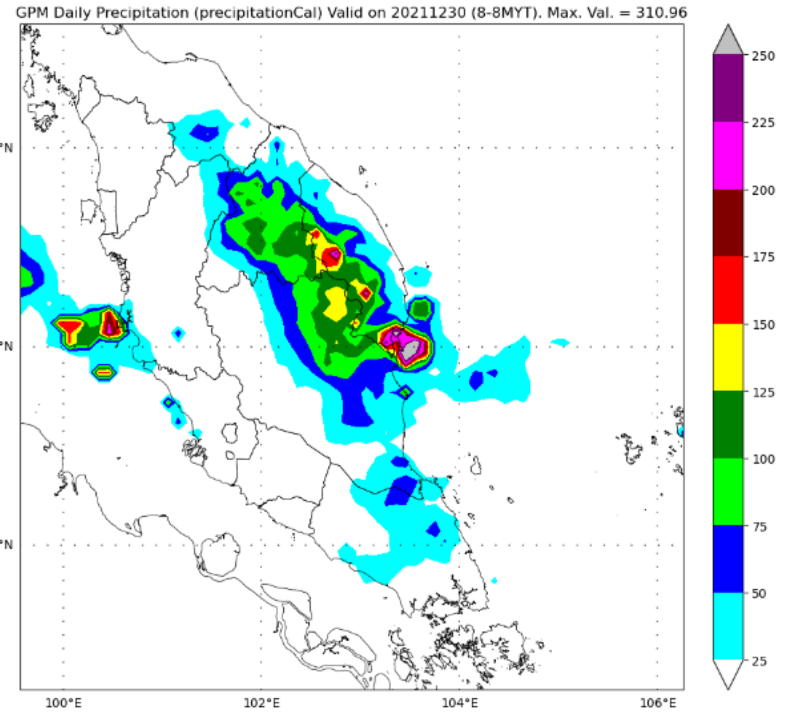
84-108	
72-96	
48-72	

36-60						
24-48						
12-36						

<table>
<tr>
<td>0-24</td>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td>
</tr>
</table>

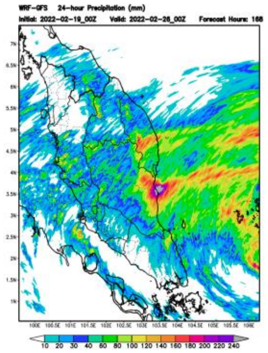
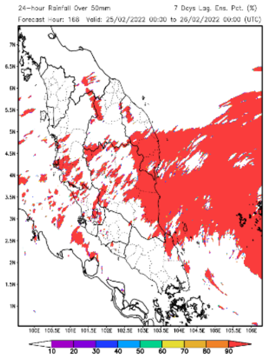
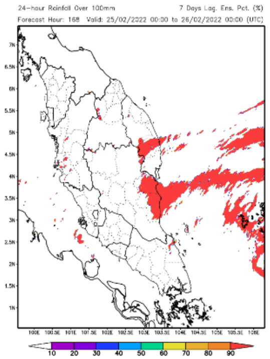
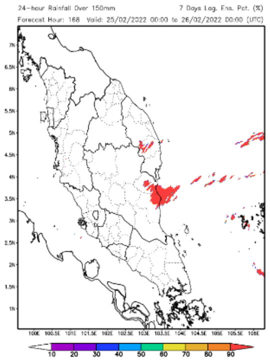
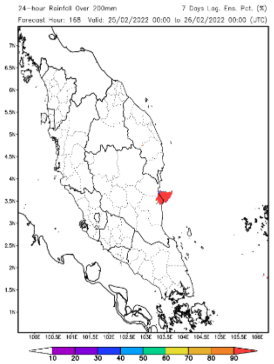
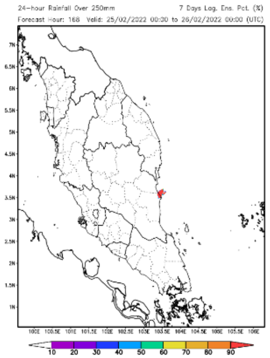
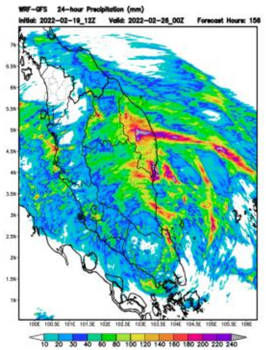
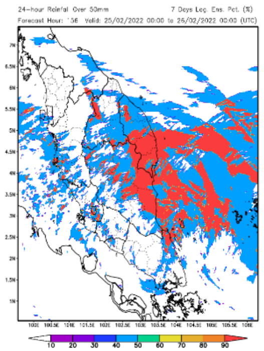
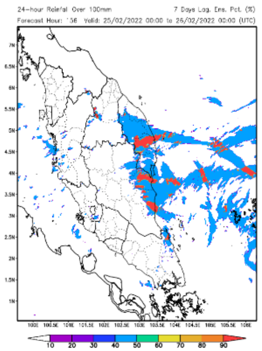
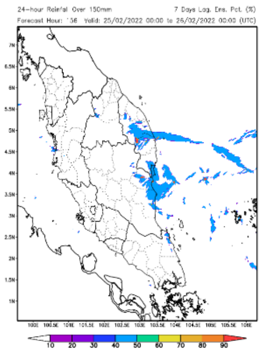
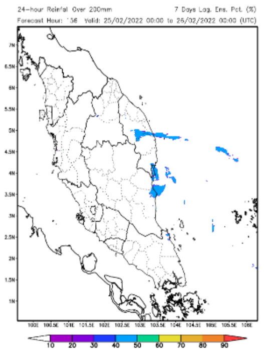
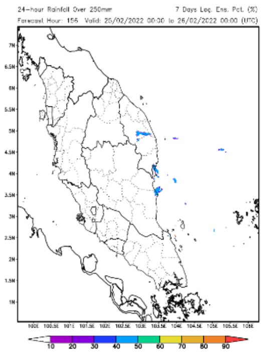
APPENDIX E

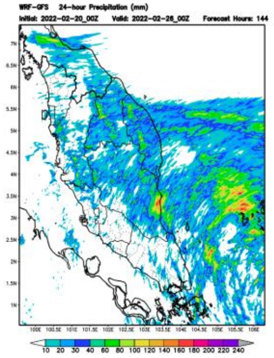
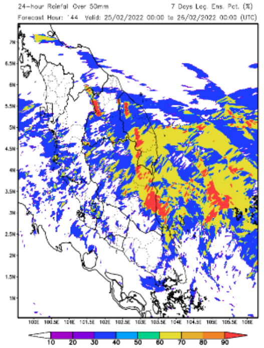
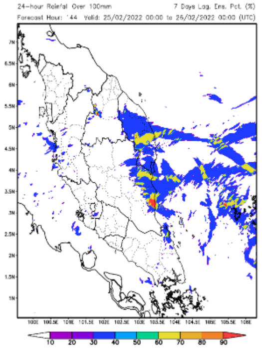
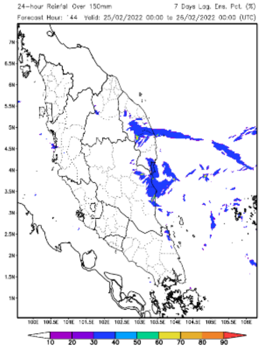
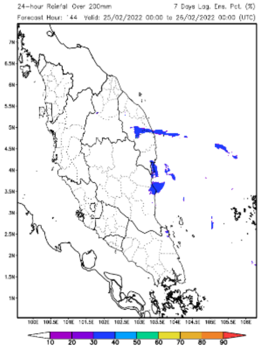
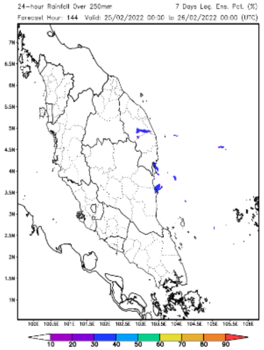
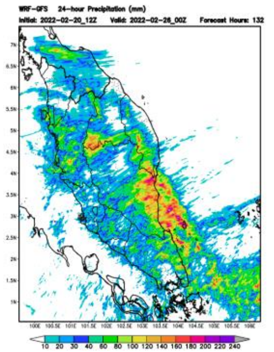
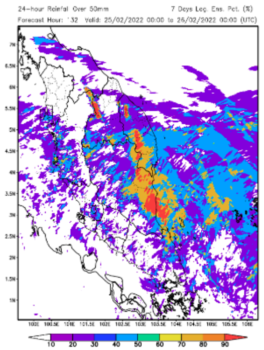
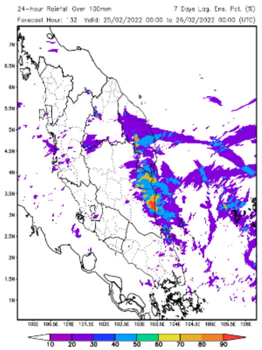
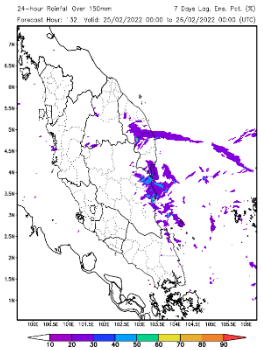
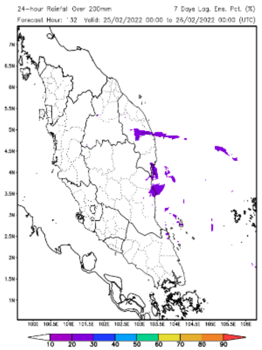
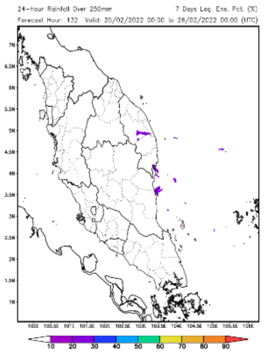
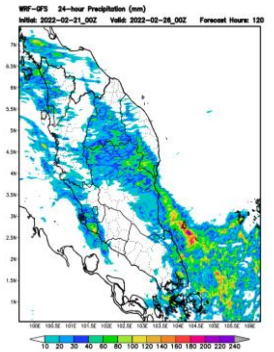
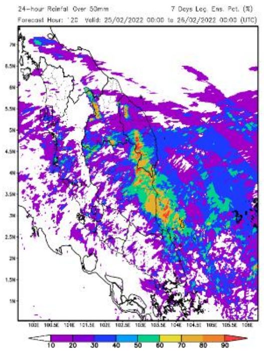
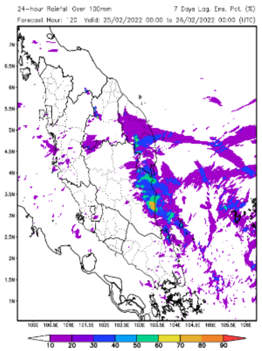
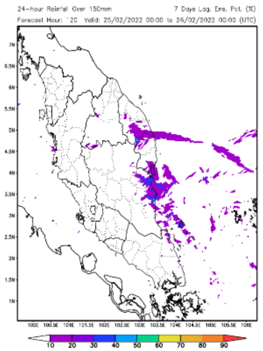
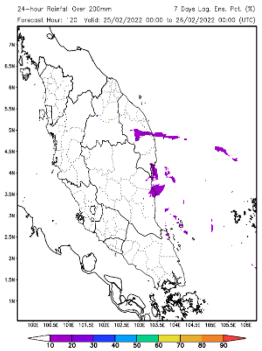
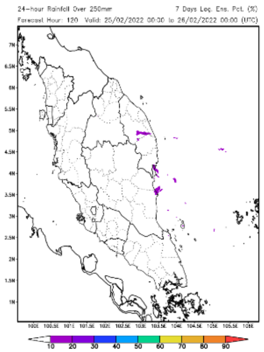
Observation of 24-hour accumulated precipitation plotted from meteorological stations and Global Precipitation Measurement (GPM) satellite (precipitation in mm) on 30th Dec 2021 (UTC)

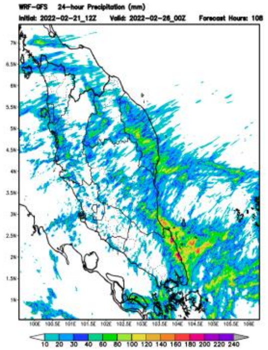
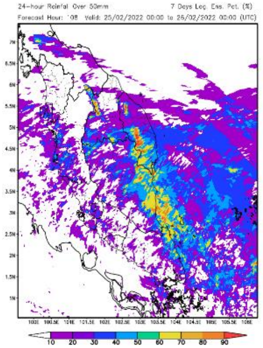
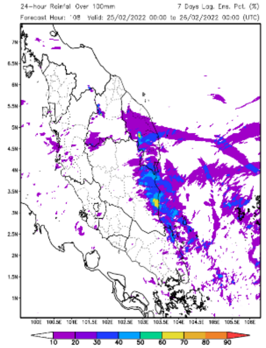
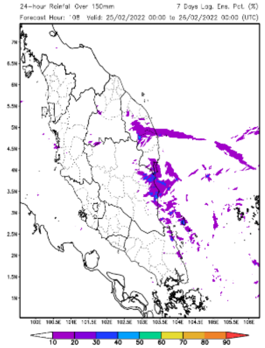
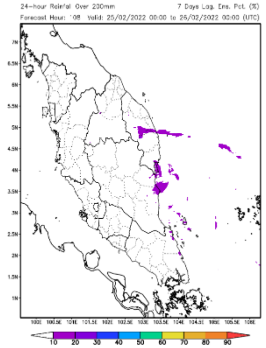
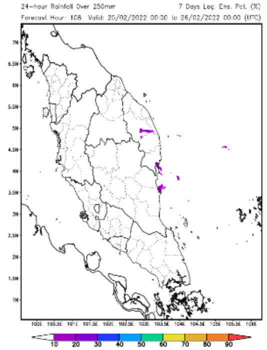
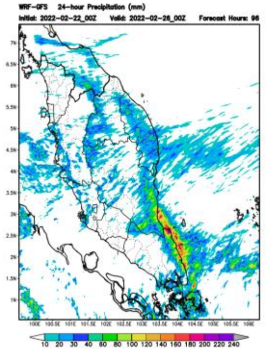
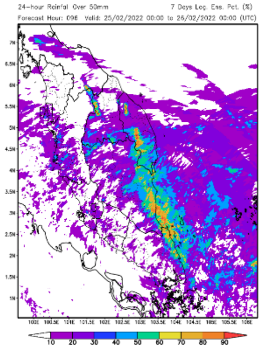
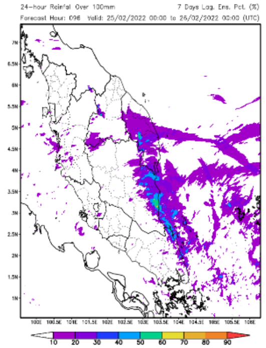
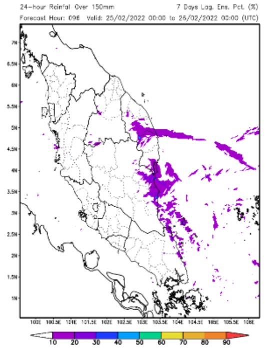
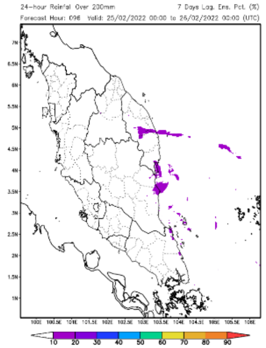
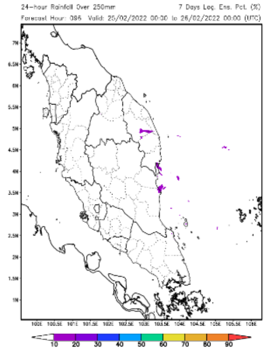
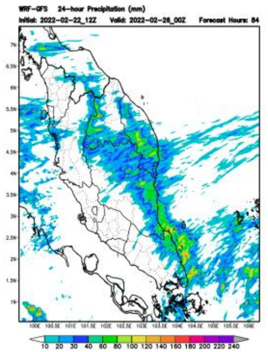
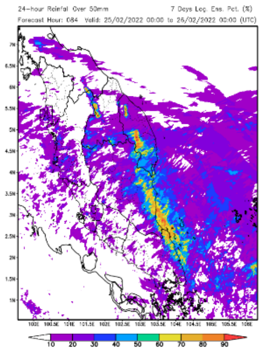
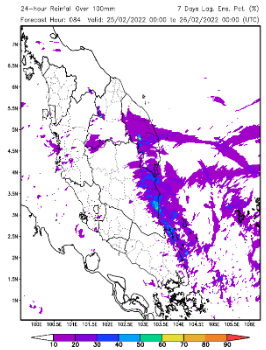
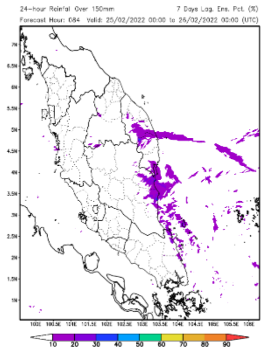
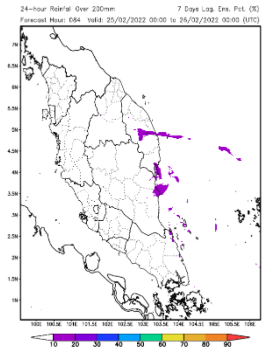
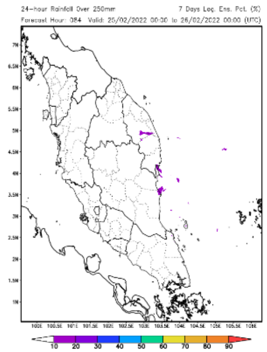
Date (UTC)	STATIONS	GPM
30/12/2021		

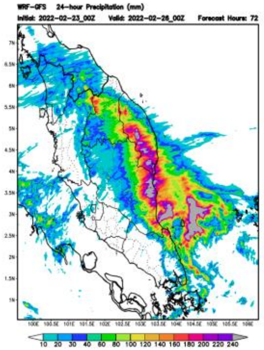
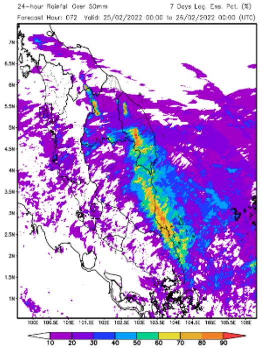
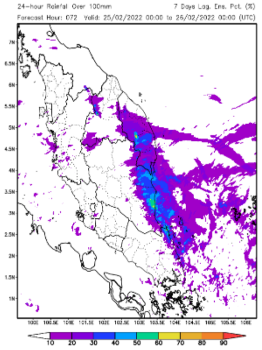
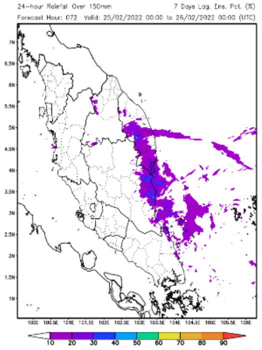
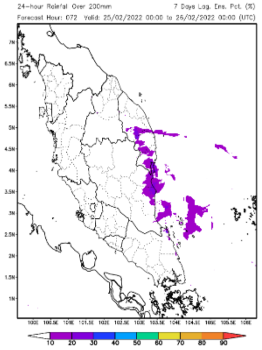
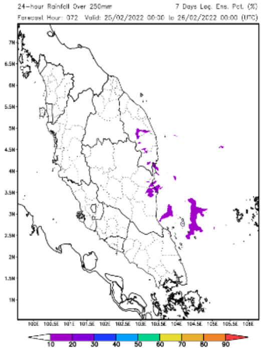
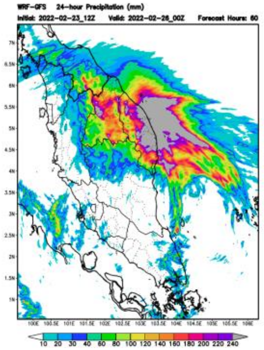
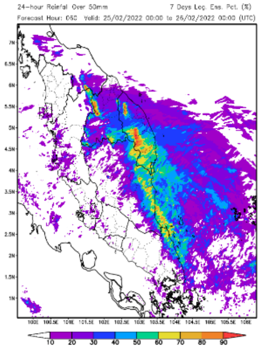
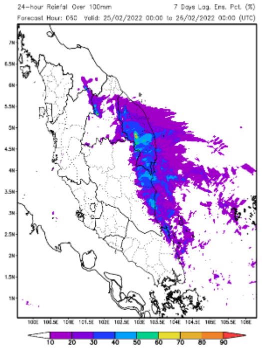
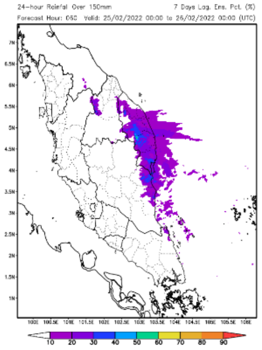
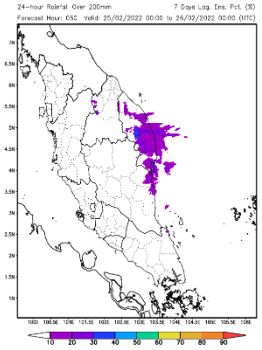
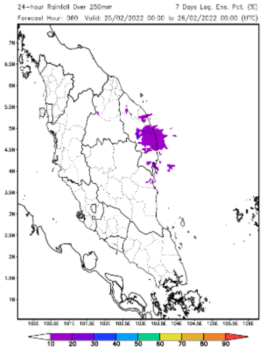
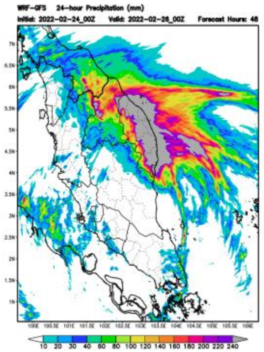
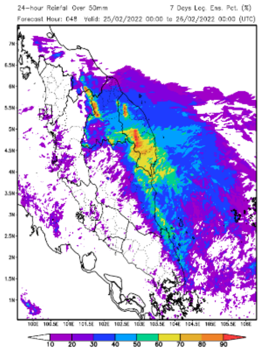
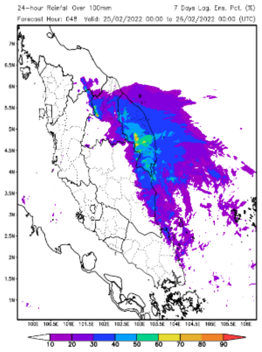
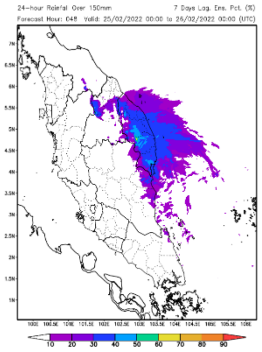
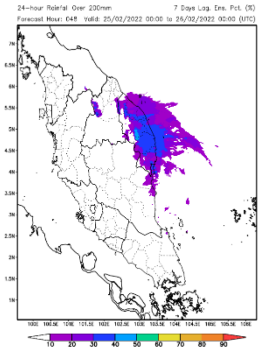
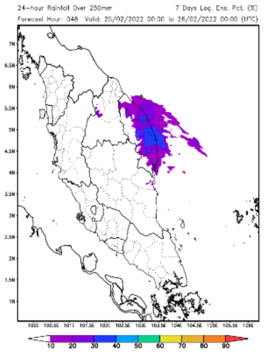
APPENDIX F

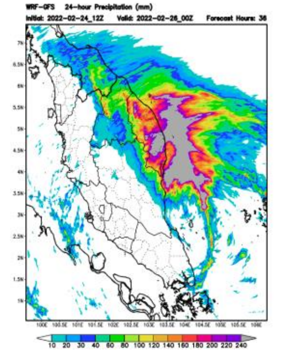
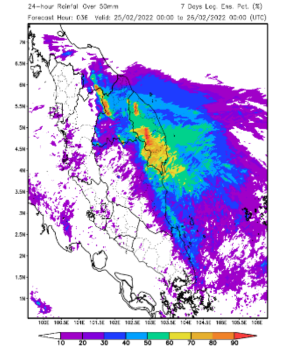
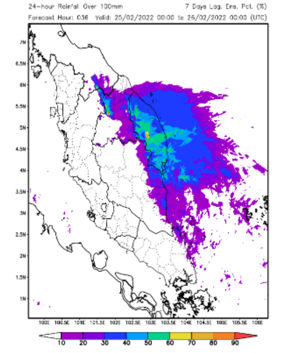
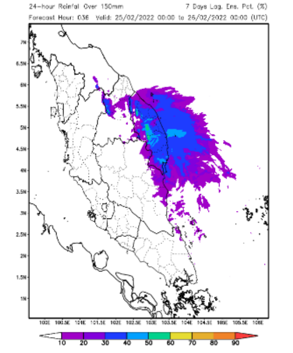
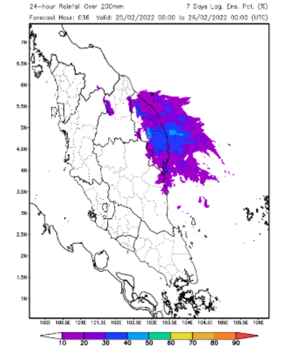
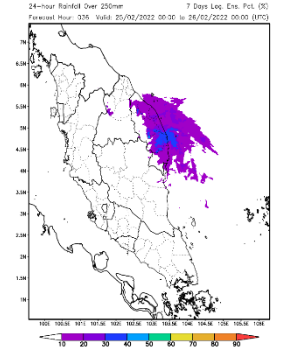
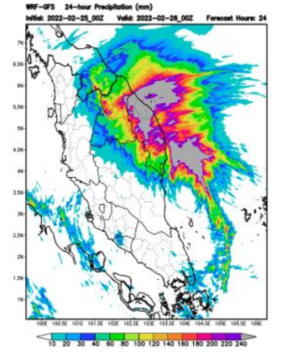
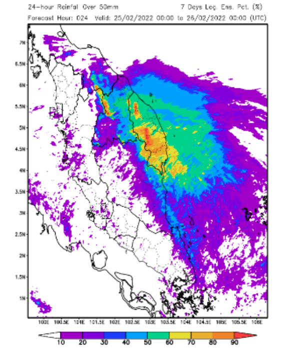
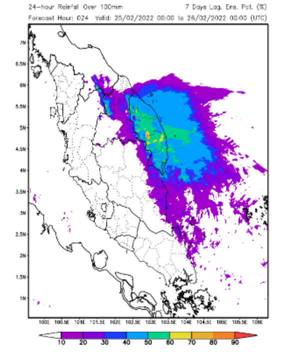
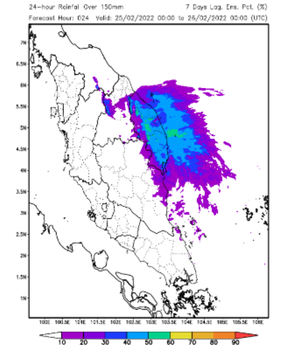
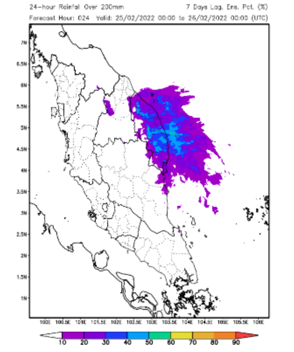
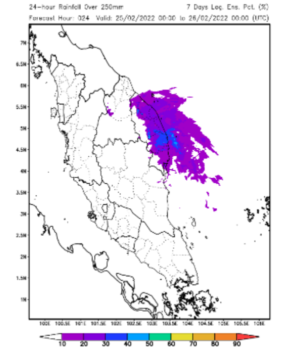
Deterministic (precipitation amount in mm) and Time-Lagged Ensemble (probability value in %) forecasts for 24-hour accumulated precipitation on 25th Feb 2022 (UTC)

Forecast-hour (hours)	Deterministic NWP forecast	TLE NWP forecast: Threshold 50mm and above	TLE NWP forecast: Threshold 100mm and above	TLE NWP forecast: Threshold 150mm and above	TLE NWP forecast: Threshold 200mm and above	TLE NWP forecast: Threshold 250mm and above
144-168						
132-156						

120-144						
108-132						
96-120						

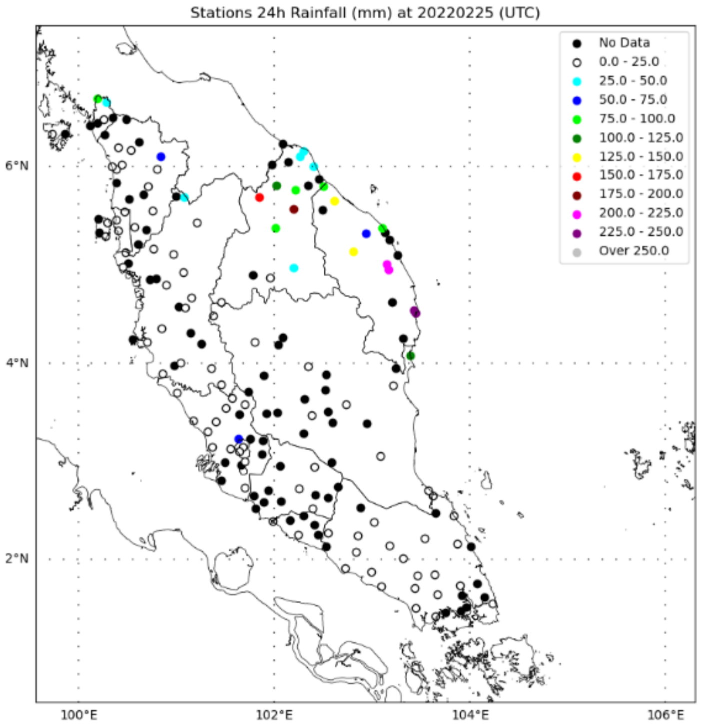
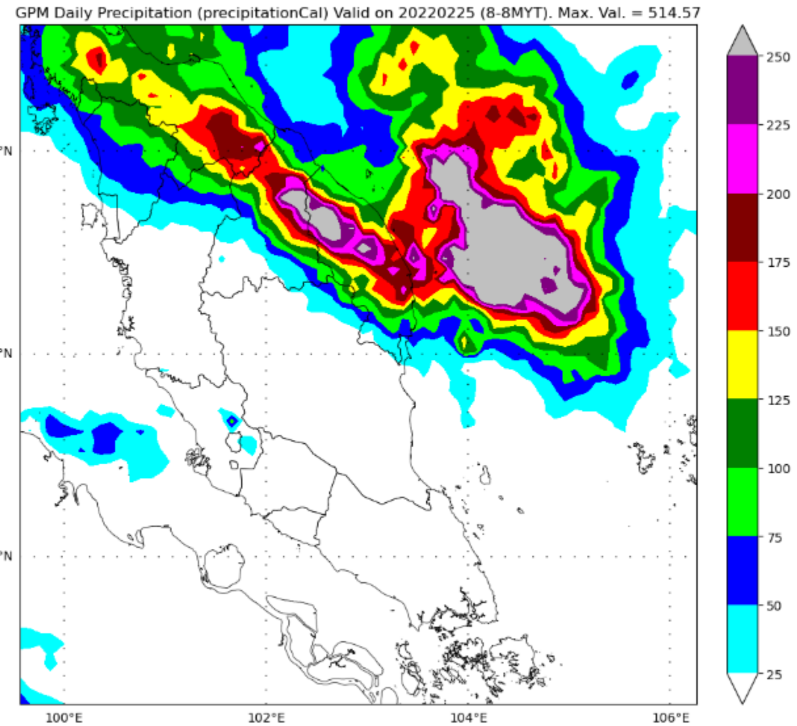
84-108						
72-96						
60-84						

48-72						
36-60						
24-48						

12-36						
0-24						

APPENDIX G

Observation of 24-hour accumulated precipitation plotted from meteorological stations and Global Precipitation Measurement (GPM) satellite (precipitation in mm) on 25th Feb 2022 (UTC)

<table>
<tr><th>Date (UTC)</th><th>STATIONS</th><th>GPM</th></tr>
<tr><td>25/2/2022</td><td>
</td><td>
</td></tr>
</table>

REFERENCES

1. Andrew W. Robertson, Frederic Vitart (2018). *Sub-seasonal to Seasonal Prediction, The Gap between Weather and Climate Forecasting, 1st Edition*. *Chapter 13 – Ensemble Generation: The TIGGE and S2S Ensembles*
2. ECMWF. *The ECMWF Ensemble Prediction System*. Obtained on March 2022 from https://www.ecmwf.int/sites/default/files/elibrary/2012/14557-ecmwf-ensemble-prediction-system.pdf
3. Ross N. Hoffman and Eugenia Kalnay (1983). *Lagged average forecasting, an alternative to Monte Carlo forecasting.*
4. C. Brankovic, T.N Palmer, F. Molteni & U. Cubash (1989). *Extended range predictions with ECMWF models. III. Time lagged ensemble forecasting*. ECMWF Technical Memorandum
5. Roberto Buizza. (2008). *Comparison of a 51-member low-resolution (T_L399L62) ensemble with a 6-member high-resolution (T_L799L91) lagged-forecast ensemble*. ECMWF Technical Memorandum
6. Chris J. Short, Jon Petch. (2022). *Reducing the spin-up of a regional NWP system without data assimilation*. Royal Meteorological Society
7. Fortran community. *Fortran Wiki HomePage*. Obtained on March 2022 from https://fortranwiki.org/fortran/show/HomePage
8. Marion P. Mittermaier (2007). *Improving short-range high-resolution model precipitation forecast skill using time-lagged ensembles*.
9. NASA. *Global Precipitation Measurement (GPM)*. Obtained on March 2022 from https://gpm.nasa.gov/missions/GPM
10. D.J.Stensrud, M.C.Coniglio, K.H.Knopfmeier, A.J.Clark (2015). *Numerical Models | Model Physics Parameterization*. Encyclopedia of Atmospheric Sciences (2nd Edition)

MALAYSIA METEOROLOGICAL DEPARTMENT
JALAN SULTAN
46667 PETALING JAYA
SELANGOR DARUL EHSAN
Tel : 603-79678000
Fax : 603-79550964
www.met.gov.my

ISBN 978-967-2327-11-0