Research Article

Environmental Science and Climate Research

Evaluation of Heat Wave Predictability Skills of Numerical Weather Models

Ademola Akinbobola[1] and Raji Ibraheem[2*]

[1]*Department of Meteorology and Climate Science, Federal University of Technology, Akure, Ondo State, Nigeria.*

[2]*Department of Meteorology and Climate Science, Federal University of Technology, Akure, Ondo Statec, Nigeria.*

[*]Corresponding Author
Raji Ibraheem, Department of Meteorology and Climate Science, Federal University of Technology, Akure, Ondo Statec, Nigeria.

Submitted: 2023, Sep 08; **Accepted**: 2023, Oct 27: **Published**: 2023, Oct 31

Citation: Akinbobola, A., Ibraheem, R. (2023). Evaluation of Heat Wave Predictability Skills of Numerical Weather Models. *Env Sci Climate Res, 1*(1), 13-32.

Abstract

Significant changes is being experienced in the climate system due to the unprecedented rate of global warming. This has resulted in the increased frequency of weather extreme events which such as heatwave occurrence in the northern Nigeria. In order to mitigate the effects of heatwaves, early warning systems are needed to be implemented. Insufficient knowledge about the performance of the models is partly a factor that hinders the development of such systems. This study thus, addresses the gap by assessing the predictability skills of sub-seasonal to seasonal numerical weather model over different time lead and as well improves the predictability skills through the incorporation of deep learning to bias correct the model output at a 30-day lead period. The Excess heat index (EHI) was used to detect heatwave occurrence over the study area, using both observational and forecast data from selected S2S models at 5 -, 7 -, 15 -, and 30 – days lead time. Metrics employed to evaluate the skills of the models are; the Anomaly corelation coefficient (ACC), Symmetric External Dependency Index (SEDI) with each evaluating different strength of the models. The result of the analysis shows that the three models considered in this research overestimates the heat wave frequency in the region. This results in reduced reliability of the models in the region. Further analyses shows that the use of deep learning to bias the model output increases the forecast reliability in the region significantly.

Keywords: Heatwave, Numerical Weather Models, Excess heat Index, Sub-Seasonal to Seasonal.

1. Introduction

Heat wave refers to periods of extended heat and or weather in periods of high humidity in reference to normal climate patterns of an area. It is responsible for immediate changes in lifestyle and depending on the intensity may have severe impact on health of the population. Heatwave has been seen experiencing an increasing trend in major part of the world, both at regional and global scale according too several studies among other climate extremes due to climate change. According to the Intergovernmental Panel on Climate Change (IPCC), it was agreed upon that the continuous increase in earth's global average temperature is unequivocal as it is clearly observed from the increase in earth's average air and sea temperatures, rising sea level and as well melting of the glaciers and sea ice (IPCC 2007d). Within the past century, earth's average temperatures have increased by approximately 0.75°C and the level of sea rise is over 4 centimeters (IPCC 2007c). Future projection warnings shows increase in severity of thermal discomfort for humans due to global warming.

Heatwave impact which can be felt on several sector such as health sector, Agricultural sector, energy sector, economic sector and several other socio-economical sectors as shown in respectively. As different locations (or regions) have different weather characteristics, geographical features, human activities, and population the effect and severity of heatwave occurrences differs. The variabilities in the climate trend experienced in recent times thus arouse interest of researchers in investigating the variability and extremes associated with temperature in the tropical region. Analysis by shows significant increase in the frequencies of extreme heat waves and reduced frequency of cold extreme events. It is also found in a study to investigate the changes in daily temperature indices and precipitation extremes in Northwestern Nigeria that significant increase in temperature indices that are related to temperature increase. Changes in extreme temperature and precipitation indices in Kaduna (Northern Nigeria) by shows similar trend.

Several indices have also been used to define and calculate warm or cold spells (heat or cold wave). This index employs the use of a number of time steps in daily maximum or minimum temperature above a particular threshold value which has been widely used.

The combination of threshold exceeding certain values of both daytime maximum or nighttime minimum temperatures has also been used. The latter defines heat wave event as at least three days with a maximum temperature above the 90th percentile (based on climatic reference period) for that day followed by a minimum temperature above the 80th percentile for that day. This was done to include periods with warm daily maximum temperatures with much of the heat still trapped at night causing a warmer night. This study assess the performance of sub-seasonal to seasonal numerical models in predicting heat wave. It also aim to improve heat wave predictability using Deep learning approach to post process the model output.

2. Data and Methods

The National Centers for Environmental Prediction (NCEP) Climate Forecast System version 2 (CFSv2), the ECMWF seasonal forecast system (SEAS5) and the UK Met Office seasonal forecast system (GloSea5-GC2-LI) were selected as medium-range forecast models to evaluate the prediction skills of Heatwaves in the Northern Nigeria at the synoptic to subseasonal lead-times. These three models provides hindcasts for daily maximum temperature (Tmax). The main reason for the preference of this models is that the ECMWF has been proven to be highly skilful. Also, the National meteorological services can easily access the models high resolution forecast data (including the 2 metre maximum temperature).

2.1 Forecast Models

The Forecast model used in this study is summarized below;

MODEL	CFsv2	SEAS5	GloSea5-GC2-LI
Department	National Centers for Environmental Prediction (NCEP)	ECMWF	UK Met Office
Group/ Project	Climate Forecast System version 2		
Hindcast Period	1982-2010	1981-2016	
Frequency	Daily	Monthly	weekly
Time range	9 months	30 days	60 days
Atmospheric Model	CFSR	IFS Cy43R1	Met Office Unified Model (UM) - Global Atmosphere 6.0
Ocean Model	GFDL MOM4	NEMO3.4	NEMO v3.4 - Global Ocean 5.0
Sea Ice Model	Sea-Ice Model	LIM2	CICE v4.1 - Global Sea-Ice 6.0
Land Surface Model	NOAH land	HTESSEL	JULES - Global Land 6.0
Initial Condition	The Climate Forecast System Reanalysis (CFSR)	ERA-Interim Reanalysis, ORAS5 reanalysis	Met Office operational numerical weather prediction (NWP) 4D-Var data assimilation system

Table 2.1: Seasonal to Sub-seasonal Numerical Model and their properties

2.2 Heat Wave Metrics

Excess heat index (EHI) is a well-known heat wave metric used for heat waves identification and intensity measure. It is high heat arising from a high daytime temperature that is not sufficiently discharged overnight due to unusually high overnight temperature. A three-day running mean maximum temperature is compared to a climate reference value to indicate the unusually high heat in an excess heat index. This is referred to as long-term (climatescale) temperature anomaly.

EHI is a perfect metrics for nationally consistent heat wave service since it includes analysis and forecasts of low-intensity heat waves and for this reason, it will be used for identifying heat wave events in this study.

Excess heat index is calculated as follows:

$$EHI_{sig} = \frac{Ti+Ti+1+Ti+2}{3} - T95 \quad (2.1)$$

Where;
$T95$ is the 95th percentile of daily temperature,
Ti is the maximum temperature on day i.
The EHI_{sig} is in effect an anomaly of three-day daily maximum temperature with respect to climatological 95th percentile of the daily maximum temperature. If the EHI_{sig} is positive, then the Three-Day Period (TDP) is unusually warm with respect to local annual climate. Conversely, if $EHIsig$ is negative or zero, then the TDP cannot be considered as unusually hot, and so for a heatwave to be present, we require EHI_{sig} to be positive.

The calculated EHI_{sig} will be compared between ERA 5 reanalysis products and the forecast models, so the daily maximum, temperature from ERA 5 reanalysis products are aggregated to match the model resolution. As a result, the daily EHI_{sig} from the ERA 5 reanalysis products is calculated from Tmax within each of the forecast models individually. Using the method described above, heatwaves in the ERA5 dataset and the model datasets are defined for the EHI_{sig} at each grid cells. The model dataset at each grid point consists of the combination of the hindcast dataset and the forecast dataset. The combination of the hindcast and the forecast datasets thus span from 2000-2020 for each model at different initialization periods.

2.3 Forecast Evaluation Metrics

To evaluate the skills of the models in predicting heatwave, evaluation metrics as presented in are used;

2.3.1 Anomaly Correlation Coefficients

Anomaly Correlation Coefficient (ACC) is one of the most widely used measures in the verification of spatial fields. It is the correlation between anomalies of model data and observational value. ACC is mathematically defined as;

$$ACC = \frac{\sum_{i=1}^{n} w_i(f_i-\overline{f})(a_i-\overline{a})}{\sqrt{\sum_{i=1}^{n} w_i(f_i-\overline{f})^2 \sum_{i=1}^{n} w_i(a_i-\overline{a})^2}} (-1 \le ACC \le 1) \quad (2.2)$$

Where;
N= number of samples
and *fi, f* , *ai* and *a* are given by the following equations:

$$fi = Fi - Ci; ai = Ai - Ci;$$

$$\overline{f} = \left(\sum_{i=1}^{n} w_i fi\right) \Big/ \sum_{i=1}^{n} w_i;$$

$$\overline{a} = \left(\sum_{i=1}^{n} w_i ai\right) \Big/ \sum_{i=1}^{n} w_i;$$

Where; *Fi*, *Ai*, and *Ci* represent forecast, verifying value, and reference value such as climatological value, respectively. Also, *f* is the mean of *fi*, a is the mean of *ai*, and *wi* represents the weighting coefficient. It indicates the strength of the association between the observed and the predicted values of the heat wave index. If the variation pattern of the anomalies of forecast is perfectly coincident with that of the anomalies of verifying value, ACC will take the maximum value of 1. In turn, if the variation pattern is completely reversed, ACC takes the minimum value of -1.

2.3.2 Symmetric External Dependency Index

According to Ferro and Stephenson 2011, common measures of forecast quality of rare events such as heatwaves struggles to show real indications of model skill for extreme events as they degenerate to trivial values with increasingly rare events. This triggers the need to design non-degenerative metrics to assess the skills associated with rare events. suggested the use of SEDI and which is agreed to be the best as shown in similar heatwave studies. SEDI is computed using the hit rate score and the false rate score which is derived from a two by two contingency table shown in Table 2 below;

$$Hitrate(H) = \frac{hits}{hits + misses}$$

$$Falserate(F) = \frac{Falsealarms}{Falsealarms + CorrectNegatives}$$

Hits - represents the instances heatwaves were detected in the observational data as well as the model data
Misses - represents the instances heatwaves were detected in the observational data but not detected from the model data
False alarms - represents the instances heatwaves were not detected in the observational data but detected from the model data
Correct Negatives - represents the instances heatwaves were not detected in the observational data as well as from the model data
The SEDI is thus obtained using;

$$SEDI = \frac{lnF - lnH - ln(1-F) + ln(1-H)}{lnF + lnH + ln(1-F) + ln(1-H)}$$

The possible values of SEDI score ranges from -1 to 1. Positive values indicate the model forecast is better than random with 1 as perfect score, while values below zero implies a forecast system which is worse.

		Observed	
		Yes	**No**
Forecast	Yes	Hits	False alarm
	No	Misses	Correct negative

Table 2.2: Contingency Table of heatwave occurrence between a deterministic forecast and the observation

2.3.3 Equitable Threat Score

To evaluate the ensemble mean heat wave forecast, the ETS (a deterministic prediction skill score) is utilized. The calculation of ETS is as follows:

$$ETS = \frac{a_n - a_r}{a_n + b_n + c_n}$$

$$a_r = \frac{(a_n)}{lnF + lnH + ln(1-F) + ln(1-H)}$$

2.3.4 Reliability Score

The reliability score metric for evaluating the heatwave prediction skills of numerical models involves assessing the agreement between the observed frequency of heatwave occurrence and the hindcast probabilities of dichotomous outcome (e.g., occurrence/non-occurrence) of heatwaves.

Reliability score measures the accuracy of a forecast model in predicting extreme events. It is typically calculated as the ratio of the number of observed events that were correctly forecast to the total number of events predicted by the model.

The formula for the reliability score is:

$$\text{Reliability score} = \frac{Numberofobservedeventscorrectlyforecast}{Totalnumberofeventsforecasted}$$

The reliability score can take values between 0 and 1, with 0 indicating that the model's forecasts are completely unreliable and 1 indicating that the model's forecasts are perfectly reliable. This range of values is widely accepted in the literature, as discussed in sources such as.

2.4 Deep Learning for bias correction

Deep Learning provides the potential avenue to develop and improve S2S forecasts systems with significantly lower computational costs. Here, a grid specific data-driven deep learning heat wave bias correction model is developed. The model takes the daily output predictions of the three seasonal to sub-seasonal numerical model used in this study as features and takes a binary data of heat wave occurrence computed from the Era-5 reanalysis datasets as labels (0 means no heat wave occurrence while 1 means heat wave occurrence). The model is designed for each grid point using data from years 1990 – 2005 as training dataset and the model is validated on datasets from 2005 -2015. The data from 2015-2020 are used as test sets. The model is then used to predict a dichotomous output of heat wave occurrence within the study period.

The model architecture is designed as shown in the table below;

Model Layers	Activation Funtction
64	Relu
40	Relu
32	Relu
24	Relu
12	Relu

6	Relu
1	Sigmoid

Table 2.3: Deep learning model architecture

Other functions included in the model at compilation stage:
Loss = 'binary_crossentropy'
Optimizer = 'Adam'
Metrics = 'accuracy'
Best model fit functions such as epochs and batch sizes were selected per grid point using the Grid Search option to get the best loss magnitude size.

3. Results and Discussion

The temperature hindcast of the models are evaluated against the re-analysis (ERA5) dataset using Pearson correlation. Fig 3.1 reveals that the correlation shows high relationship between the NCEP's CSFV2 model at all-time leads in major part of the study area except the 7-days' time lead. The model shows maximum correlation with values of 0.6 in the northern part to 0.8 in the southern part of the study area at 15-days' time lead. The average correlation at is observed to be 0.4 at both 5-days and 30-days' time leads in major part while the northern part shows correlation of 0.2. The 7-days' time lead generally shows a weak negative correlation with value of -0.2 in major part of the study area.

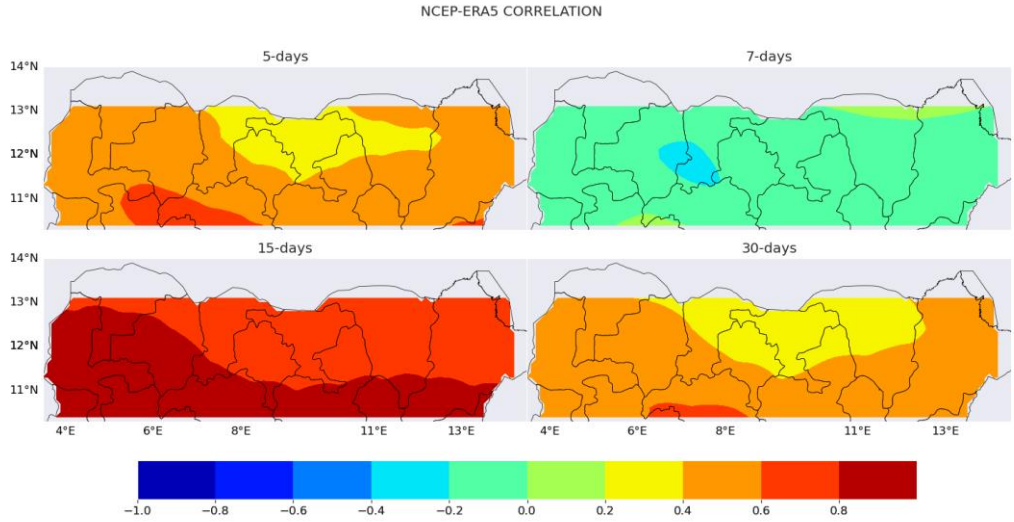

Figure 3.1: Correlation co-efficient between the NCEP CSFV2 and ERA5 reanalysis data of maximum temperature at 5-days, 7-days, 15-days and 30-days time lead.

The UKMO hindcast correlation shows that the 5-days lead period have the weakest correlation especially in the northern region with value of 0.5 and increase southward with observed value of 0.6 in the in the central region. Both the 7-days and 15-days time lead shows average spatial correlation of 0.7 while the 30-days lead period shows 0.6. Generally, a local maximum in the southern part of the region is observed at all time lead.

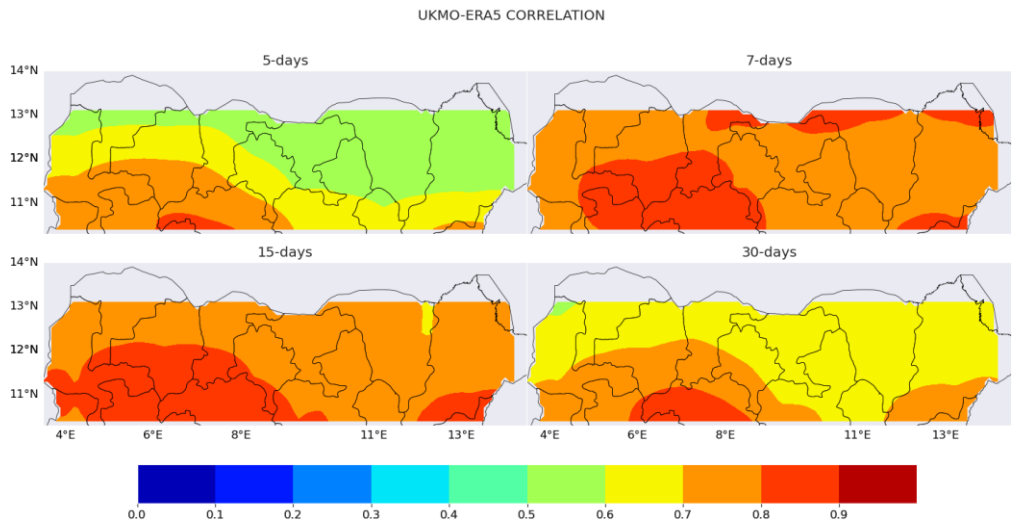

Figure 3.2: Correlation co-efficient between the UKMO GloSea5-GC2-LI and ERA5 reanalysis data of maximum temperature at 5-days, 7-days, 15-days and 30-days time lead.

The ECMWF model only available at 30-days time lead shows a southward increase in the correlation coefficient. The plot shows 0.5, 0.6 and 0.7 magnitude for the correlation coefficient in the norther, central and southern part of the study area.

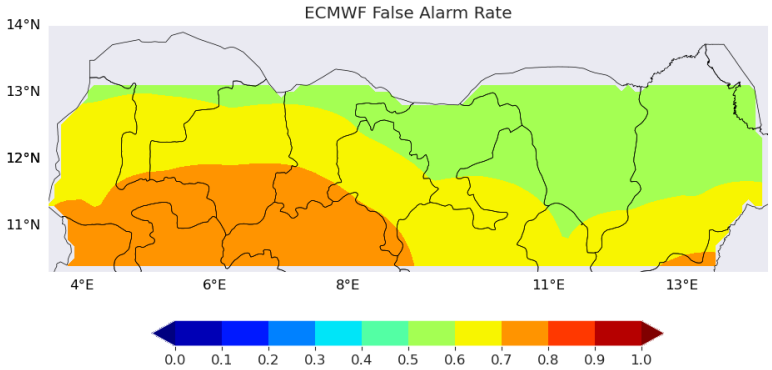

Figure 3.3: Correlation co-efficient between the ECMWF SEAS5 and ERA5 reanalysis data of maximum temperature at 30-days time lead.

The evaluation of the predictability skills of the models are done using complimentary metrics: Reliability; Equitable Threat Score (ETS), Symetric External Dependency Index (SEDI). These metrics are selected as they evaluate different facet of the forecast skills of the model. AUC measures the discrimination, this implies that the higher the AUC score the more the forecast have prediction hits and the lower the false positives. Both the reliability and ETS score metrics rewards the consistency in the models while penalizing false positives and false negatives. However, Reliability assess the ensemble member while the ETS assesses the ensemble mean. SEDI score offers the possibility of assessing the general performance of a forecast system and its ability to predict rare extreme events. An important property of this metric is its resistance to 'hedging' (i.e when a forecast that is issued differs from the forecast which would have been given). The metrics, when all evaluated on a model provides a detailed evaluation of each model heat wave predictability skill. The forecast lead time in this research are 5, 7, 15, 30- day lead period for the NCEP's CFsv2 and UK Met Office GloSea5-GC2-LI while the ECMWF SEAS5 model has a 30-day lead time forecast only.

3.1. NCEP's CFsv2

Fig 3.1.1 shows reliability maps for dichotomous heat wave hindcasts with lead times of 5, 7, 15, and 30 days. The model's reliability decreases as values approach 0, and increases as they approach 1. The model's estimation of heatwave occurrence is either an underestimation or overestimation when compared to the observational data (Perfect Forecast). Based on the reliability plots for 5-day, 15-day, and 30-day lead times, it appears that the model tends be less reliable in predicting heatwave occurrence in most of the region as the reliability score is averagely between 0.2 and 0.3. However, the reliability of the model further decrease at the 5-day lead time.

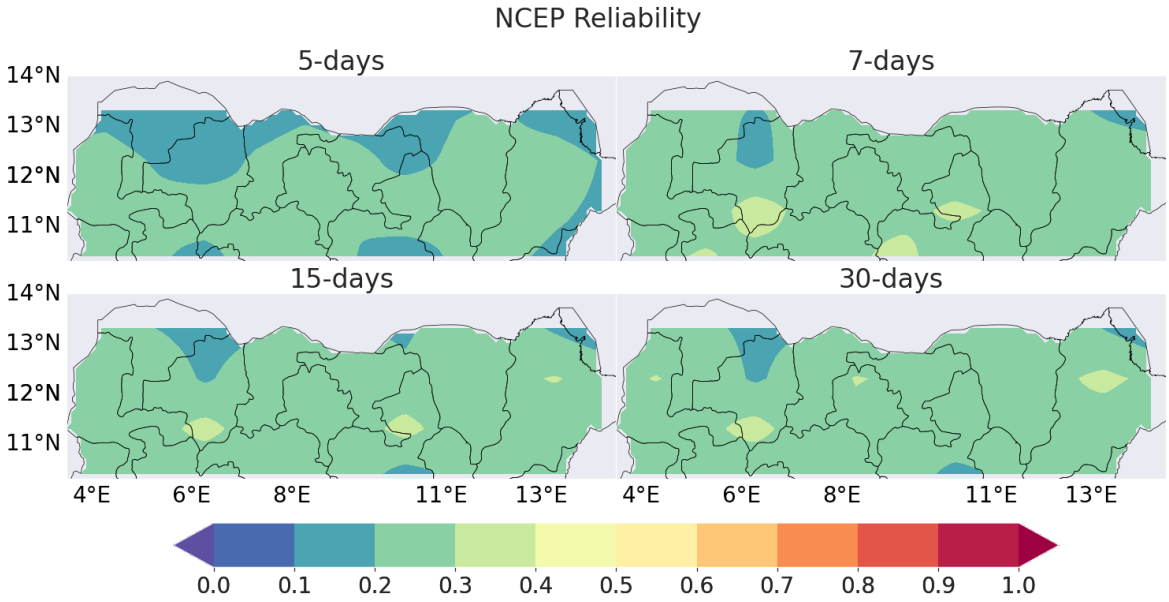

Figure 3.1.1: NCEP CFsv2 Reliability Map at 5 -, 7 -, 15 -, and 30 – days lead period

Temporal Autocorrelation bias maps of dichotomous heat wave hindcasts are plotted for 5-day, 7-day, 15-day, 30-day lead time (fig 2 (a-d)). The plots at 5-day, 15-day and 30-day lead time indicates that the CFsv2 model overestimates the autocorrelation (i.e persistence of maximum temperature). This thus implies that the model overestimates the heat wave frequency in the whole region causing the model to be less reliable in the whole region at these lead periods. Although, at 30-day lead time, the over-estimation is relatively low when compared to the other lead time. It is also observed that with decreasing lead time, the model over-estimation of heat wave increases. Also, the 7-day lead time model output shows the maximum over-estimation of heat wave frequency.

Finally, the model does not under-estimate the heat wave frequency has no negative value across all the lead times.

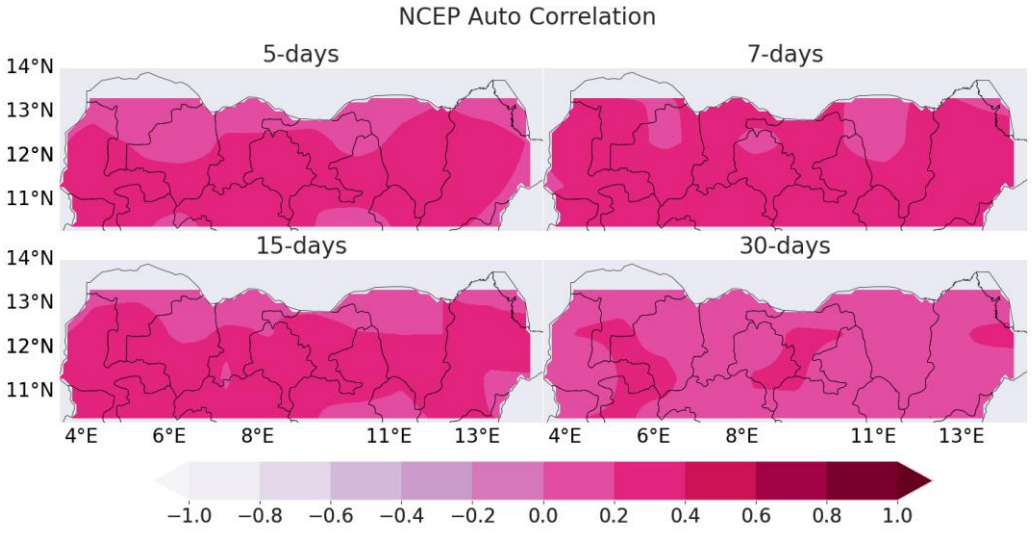

Figure 3.1.2: NCEP CFsv2 Auto-correlation bias at 5 -, 7 -, 15 -, and 30 – days lead period

The CFsv2 model's equitable threat score (ETS) maps for heat wave prediction indicate that the model's heat wave prediction abilities are generally weak in the region, as demonstrated by the low ETS scores generated at all lead times. However, the ETS scores improve at 15-day and 30-day lead times, with values reaching up to 0.15 and 0.13, respectively. As the lead time decreases, the ETS score also decreases, with the maximum score being less than 0.1 at 7-day and 5-day lead times. Additionally, a longitudinal peak is observed at 15-day and 30-day lead times (6.27oE, 10.27oE, and 14.27oE), which corresponds to the area of maximum reliability. Regions where the forecast model performs more along those longitudes are observed to have certain features. The max-

ima at the North-East part of the region can be attributed to the presence of lake in the region while the maxima at the North-West and North-Central can be attributed to presence of vegetation in both region which is absent in the surrounding region.

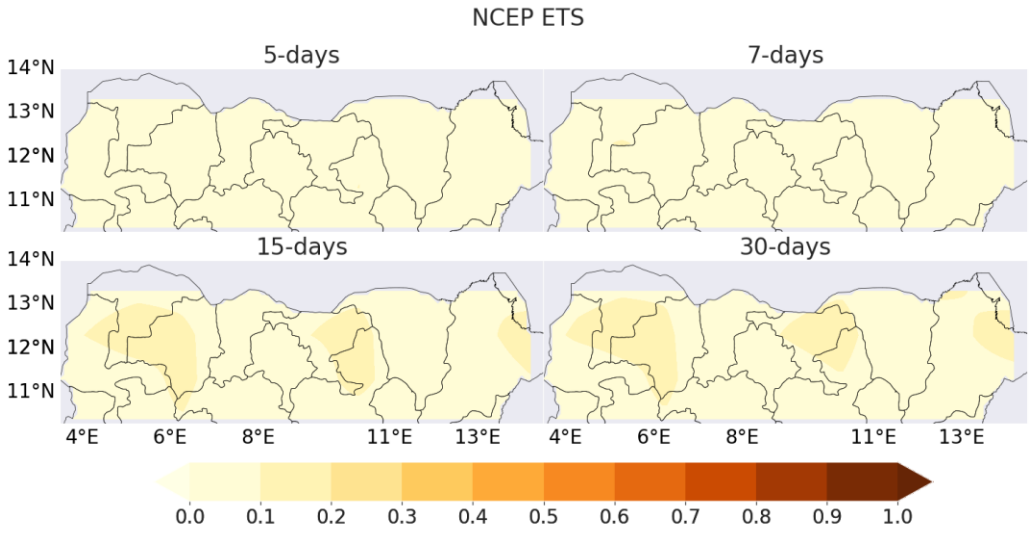

Figure 3.1.3: NCEP CFsv2 Equitable Threat score at 5 -, 7 -, 15 -, and 30 – days lead period

The NCEP's CFsv2 performs excellently in terms of hit rate, which increases as time lead decreases. Hit rate scores near a perfect score of 1 are observed at 5-day and 7-day lead periods with little to no spatial variation. However, the northern part of the region shows an average hit rate score of 0.6 to 0.7 at 30-day lead, while a score greater than 0.8 is observed at 15-day lead. The model's low reliability is due to its high false alarm rate, indicating overestimation of heat wave frequency in the region. A closer examination of false alarm rates reveals that the model overestimates heat wave occurrence to a greater extent at lower lead times. At a 30-day lead time, lower false alarm rates appear along longitude (6.27, 10.27 and 14.27), which gradually decrease at 15-day lead and almost disappear at 7-day and 5-day lead times, likely due to surface characteristics of the area. Additionally, the hit rate tends to increase as lead time decreases. The high hit rate is most likely attributable to shared bias between the MERRA-2 reanalysis CFS reanalysis product, the latter of which is used for initial conditions in the NCEP model.

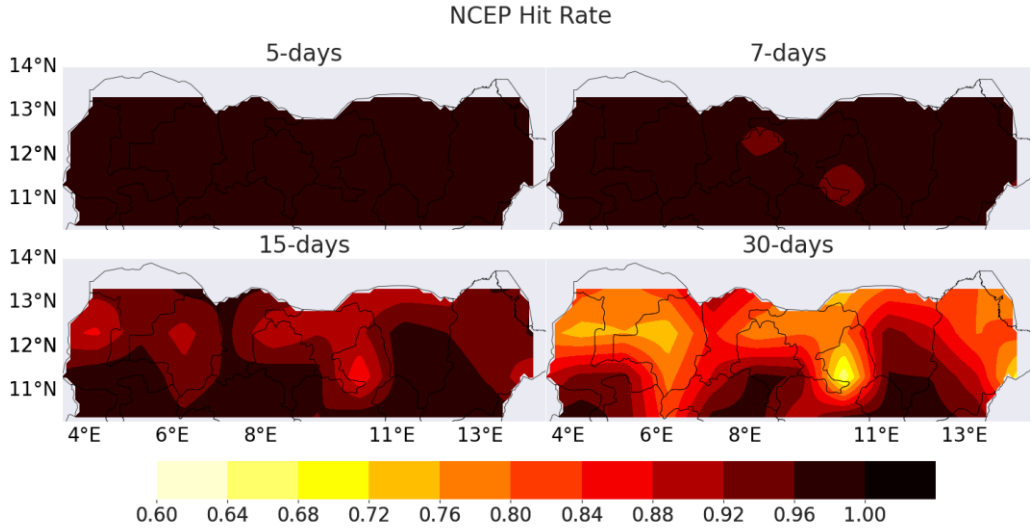

Figure 3.1.4: NCEP CFsv2 Hit Rate at 5 -, 7 -, 15 -, and 30 – days lead period

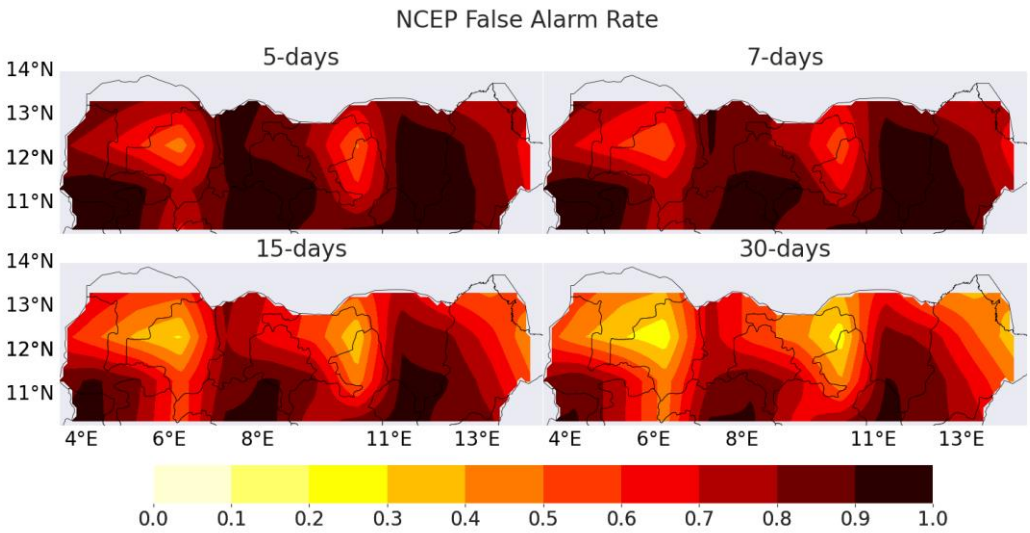

Figure 3.1.5: NCEP CFsv2 False Rate at 5 -, 7 -, 15 -, and 30 – days lead period

The skill of the NCEP's CFsv2 in predicting heat wave in the region is assessed using the SEDI score. Similarly to the ACCs of the indices, both the 30-day and 15-day lead time model data shows larger SEDI score of heatwave prediction than 7-day and 5-day model data showing greater skills. At 30-day lead time, the score is as high as 0.64 and few regions of less than 0.2. Also, the 15-day lead model data has a higher score of up to 0.73 and average of 0.4 in some areas.

The skill vanishes quicker as the forecast lead time decreases such that there is almost no skill (SEDI score below zero means random forecast better than the model) in forecast initialized at 5-day and 7-day lead period. Though some areas at 7-day lead time still have positive SEDI score values but very weak. At 5-day lead period, the model perform woefully with the skill of the SEDI score vanishing in most areas. Also, the effect of surface characteristics still improves the score at places where both water body and vegetation are present.

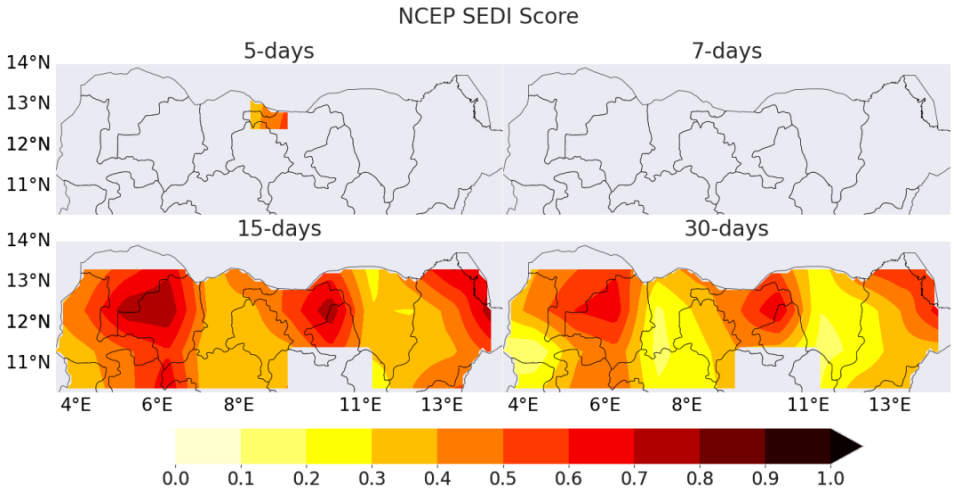

Figure 3.1.6: NCEP CFsv2 SEDI Score at 5 -, 7 -, 15 -, and 30 – days lead period

3.2. UK Met Office GloSea5-GC2-LI

Fig. 6 displays reliability maps for dichotomous heat wave hindcasts with lead times of 5 days, 7 days, 15 days, and 30 days. Notably, the 30-day lead time exhibits the lowest reliability, with a reliability score of 0.3 in much of the region. However, the northwestern part of the study area demonstrates improved reliability with a score of 0.5. Conversely, the reliability score gradually increases for lower lead times. Additionally, a persistent local maximum of reliability score is observed at latitudes 6°E, 10.5°E, and 13°E, while other regions appear to be fairly reliable at lead times of 5, 7, and 15 days.

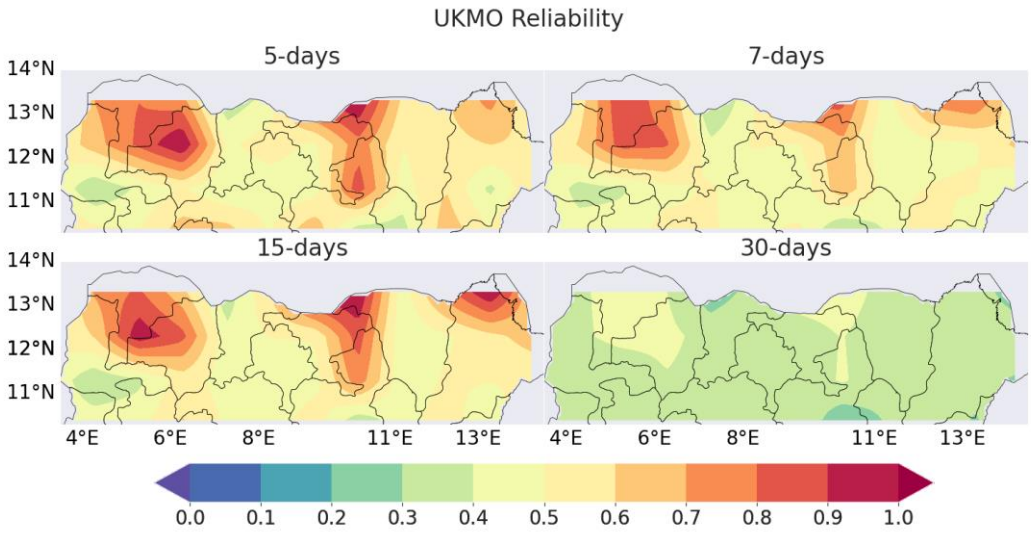

Figure 3.2.1: UKMO Reliability Map at 5 -, 7 -, 15 -, and 30 – days lead period

Temporal Autocorrelation bias maps of dichotomous heat wave hindcasts are plotted for 5-day, 7-day, 15-day, 30-day lead time (Fig 3.2.2). The plots at 5-day, 7-day, 15-day and 30-day lead time indicates that the GloSea5-GC2-LI model overestimates the auto-correlation (i.e persistence of maximum temperature). Although, at 30-day lead time, the over-estimation is relatively low when compared to the other lead time while at 5-day, 7-day and 15-day lead time, the Northern part of the region at longitudes 6.27, 10.27 and 12.27 shows a maximum autocorrelation bias score compared to other region. It is also observed that with decreasing lead time, the model over-estimation of heat wave increases. This same pattern is observed in the reliability plot. In addition, the model as well does not under-estimate the heat wave frequency has no negative value across all the lead times.

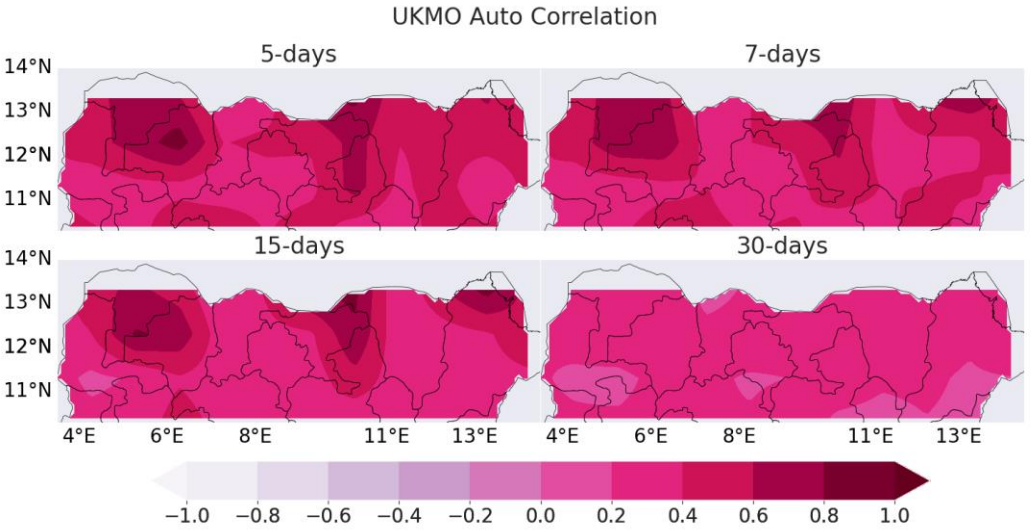

Figure 3.2.2: UKMO Auto-correlation bias at 5 -, 7 -, 15 -, and 30 – days lead period

3.2.2 Equitable Threat Score

The GloSea5-GC2-LI model's heat wave prediction is evaluated using the equitable threat score (ETS) at different time leads, as displayed in Fig 3.2.3. The maps show that the model has low reliability in the region for 15-day and 30-day leads, as evidenced by the low ETS scores. However, the reliability of the model improves for 5-day and 7-day leads, with ETS scores of up to 0.3 and 0.4, respectively, in some areas. Notably, a reverse trend is observed compared to CFsv2, where the reliability increases with increasing time lead. Additionally, the presence of water bodies and vegetation has a positive effect on the ETS score, although not as significant as observed in NCEP's model.

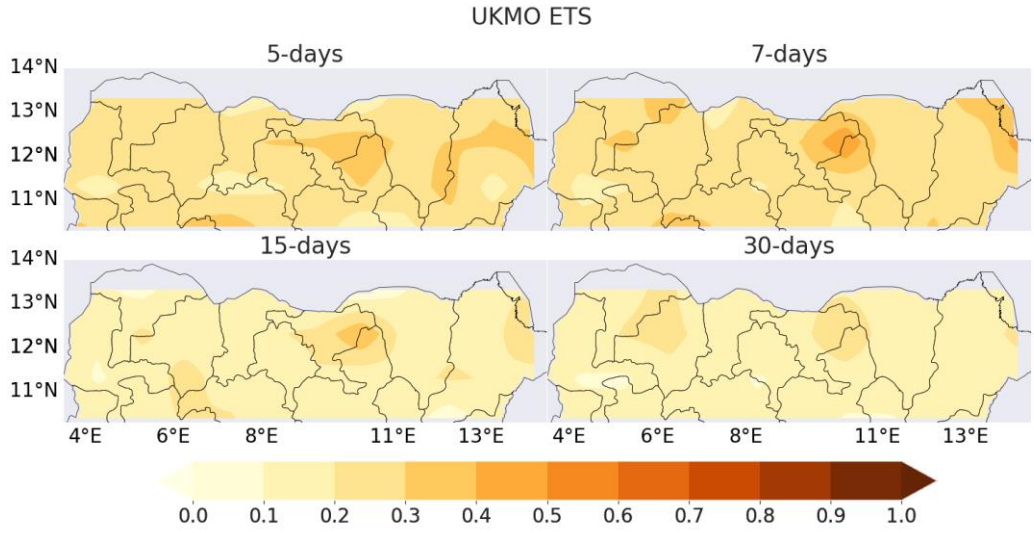

Figure 3.2.3: UKMO Equitable Threat Score at 5 -, 7 -, 15 -, and 30 – days lead period

The hit rate of the GloSea5-GC2-LI model is generally high, except for a time lead of 15 days where the hit rate drops to around 0.2 in the northern region, with a local maximum of 0.6 at 4oE and 8oE, which is relatively low compared to other time leads. The northwest region consistently shows lower hit rate scores than other areas across all time leads. In contrast, the false alarm rate plots indicate low scores (ranging from 0.1 to 0.3) compared to the corresponding hit rate scores. The plots show a maximum false alarm rate score of 0.5 at a 30-day time lead in the same area with a local maximum observed in the hit rate plot (Fig 3.2.4) and tend to decrease with decreasing time lead.

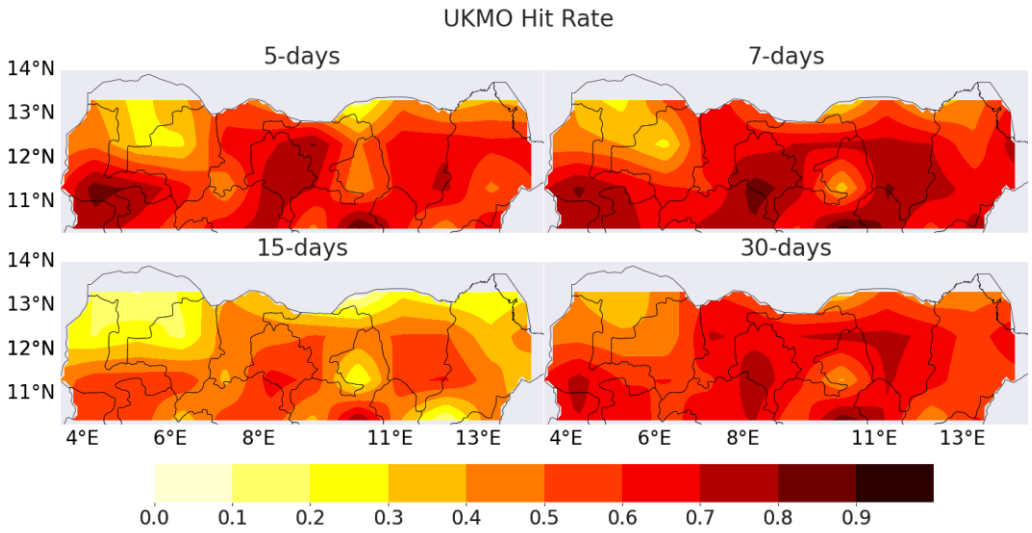

Figure 3.2.4: UKMO Hit Rate at 5 -, 7 -, 15 -, and 30 – days lead period

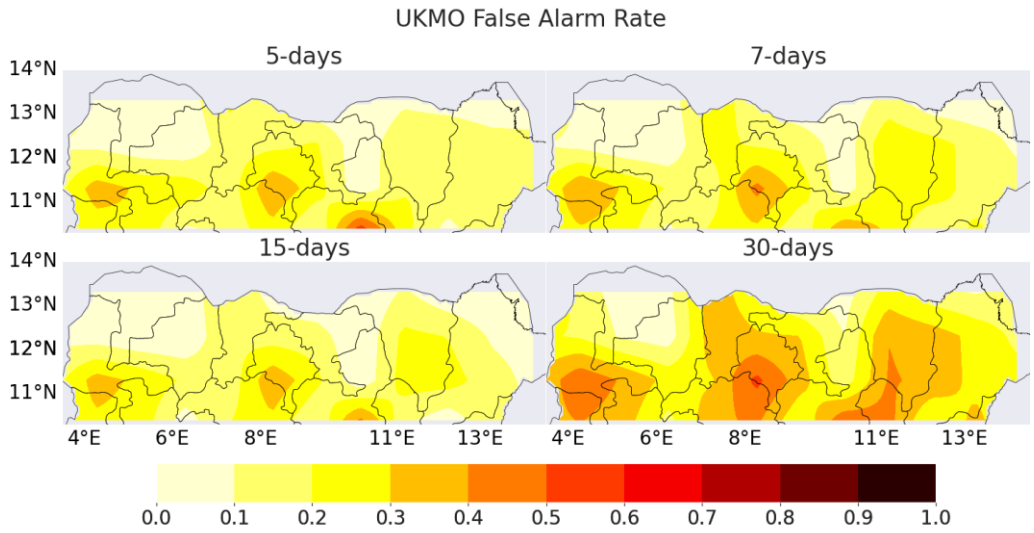

Figure 3.2.5: UKMO False Alarm Rate at 5 -, 7 -, 15 -, and 30 – days lead period

3.2.4 SEDI SCORE

The skill of the UK Met Office's GloSea5-GC2-LI in predicting heat wave in the region is assessed using the SEDI score. The SEDI score map at all lead time shows a high value of the score across all lead time. This shows that the model has better skill at predicting heat wave in the region than the NCEP's CFsv2. The score tends to increase with decreasing lead time. At 15-day lead time, the skill vanishes toward the Northern part of the region while at 5-day and 7-day lead period, the skill vanishes at few points in the region.

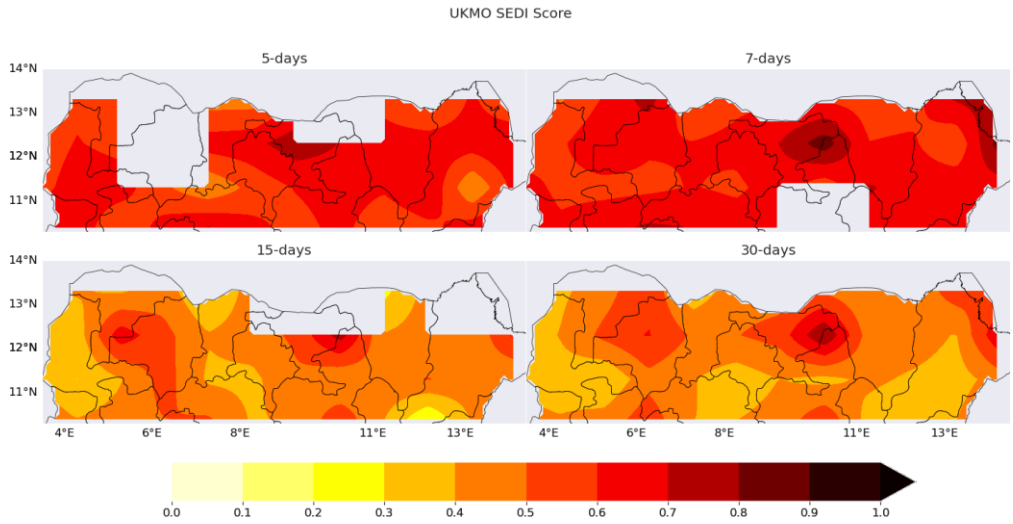

Figure 3.2.6: UKMO SEDI Score at 5 -, 7 -, 15 -, and 30 – days lead period

3.3. ECMWF SEAS5

Unlike the NCEP's CFsv2 and the UK Met Office's GloSea5-GC2-LI, the ECMWF SEAS5 is initiated only once a month, thereby allowing for a maximum evaluation of the model at a 30-day lead time. The reliability map for the 30-day lead time exhibits a localized maximum at latitudes 6°E-7°E, while a significant part of the region indicates lower reliability in detecting heatwave occurrences by the model.

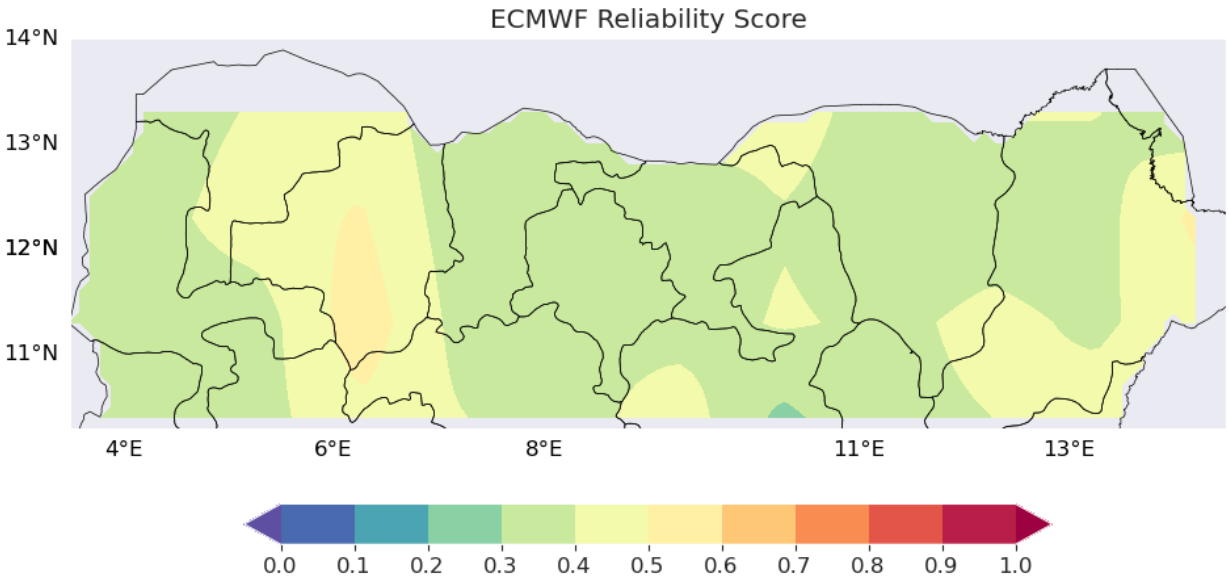

Figure 3.3.1: ECMWF Reliability at 30 – days lead period

The temporal autocorrelation of dichotomous heat wave hindcasts of the SEAS5 model at 30-day lead period shows that the model over-estimates the heat wave frequency at certain locations (longitude 6.27, 10.27 and 14.27 degree). Aside the locations mentioned earlier, the model have a fair estimation of the heat wave frequency in the other areas of the region. The model does not underestimate the heat wave frequency at any point in the whole region.

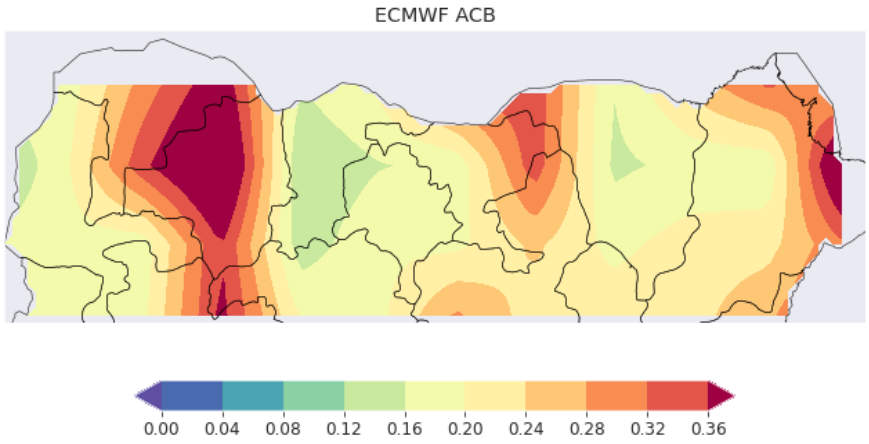

Figure 3.3.2: ECMWF Auto-correlation bias at 30 – days lead period

According to the equitable threat score (ETS) maps for heat wave prediction of the SEAS5 model at a 30-day lead time, the model exhibits low reliability throughout the region. Specifically, the ETS score across most of the region is 0.2, with some localized minima observed at latitudes 7oE and 11oE.

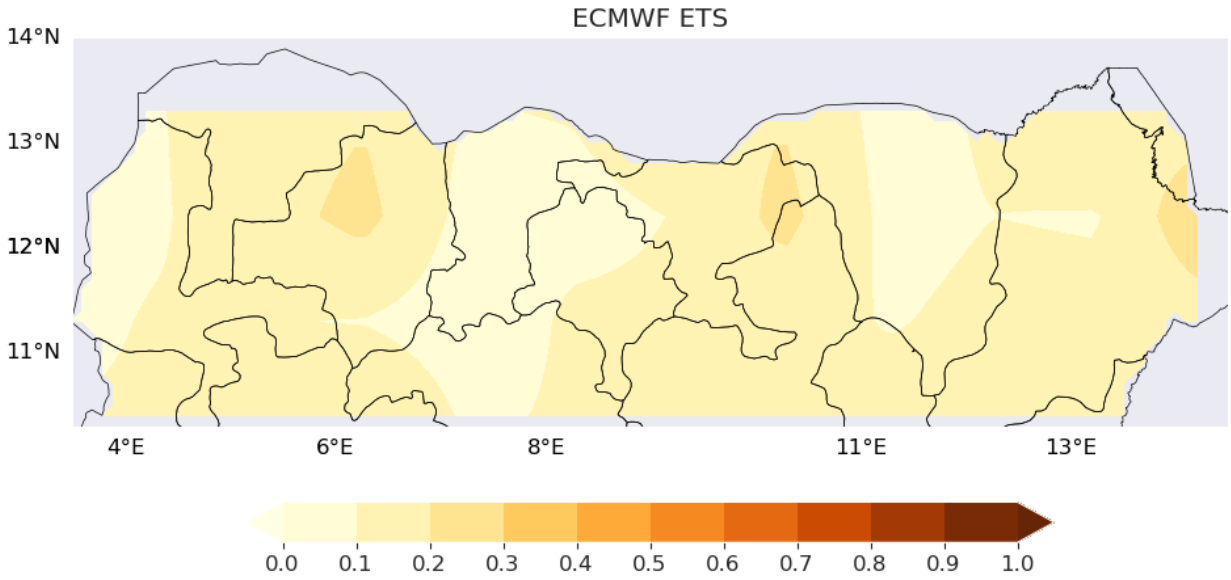

Figure 3.3.3: ECMWF Equitable Threat Score at 30 – days lead period

3.3.2 Hit Rate and False Rate

The SEAS5 model exhibits hit rate scores of 0.4 or greater in most major areas of the region, with only a few areas showing values of 0.2. The false alarm rate scores, on the other hand, are generally low across most areas, with an average score of 0.15. These results suggest that the SEAS5 model has the potential to be a reliable model that does not overestimate heatwave frequency.

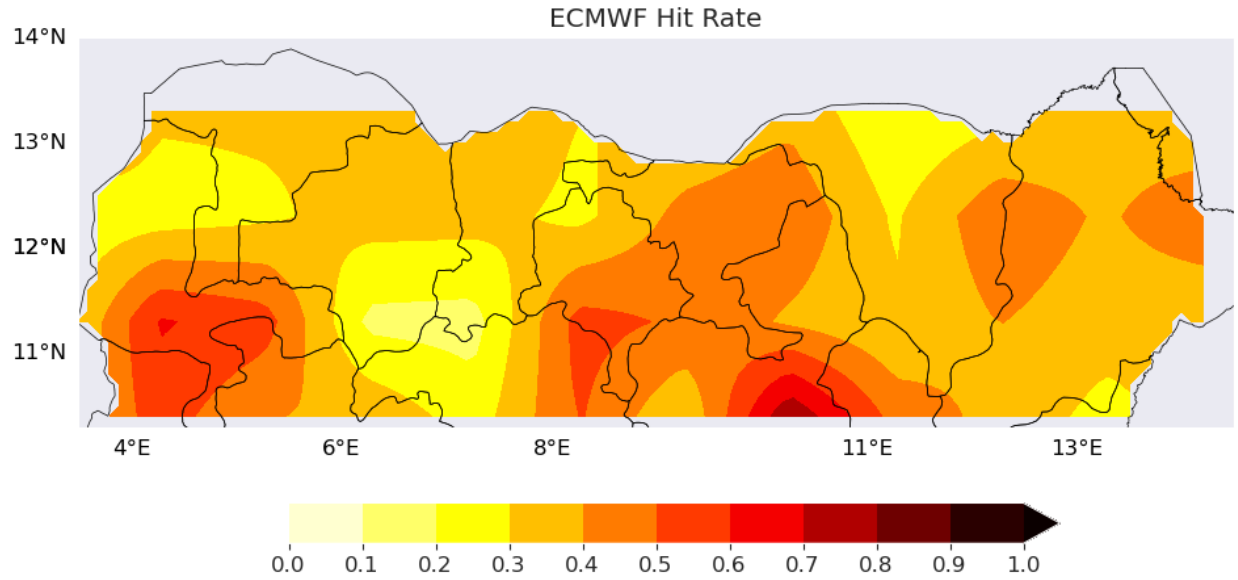

Figure 3.3.4: ECMWF Hit Rate at 30 – days lead period

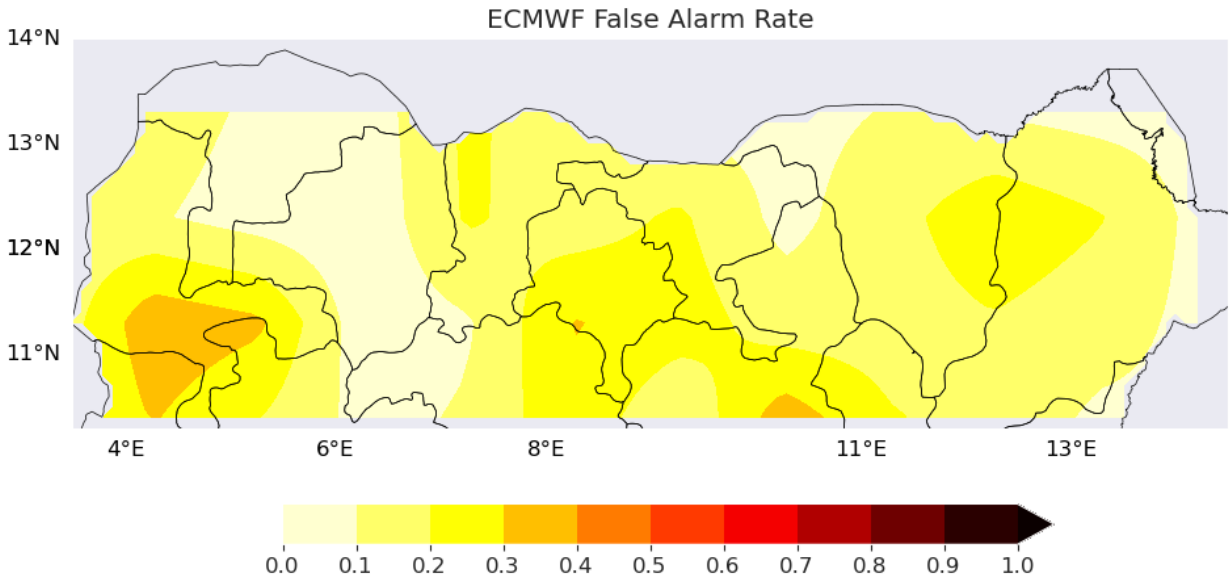

Figure 3.3.5: ECMWF False Rate at 30 – days lead period

The SEDI score of the SEAS5 initiated at a 30-day lead time shows resemblance to the NCEP's CFsv2 model SEDI score initiated at the same time lead (30-day time lead). The skill have its highest scores of up to 0.6 in the region in a pattern that matches that of the NCEP's CFsv2 model. Areas having lower score in the region have a pattern that as well matches that of the NCEP's CFsv2.

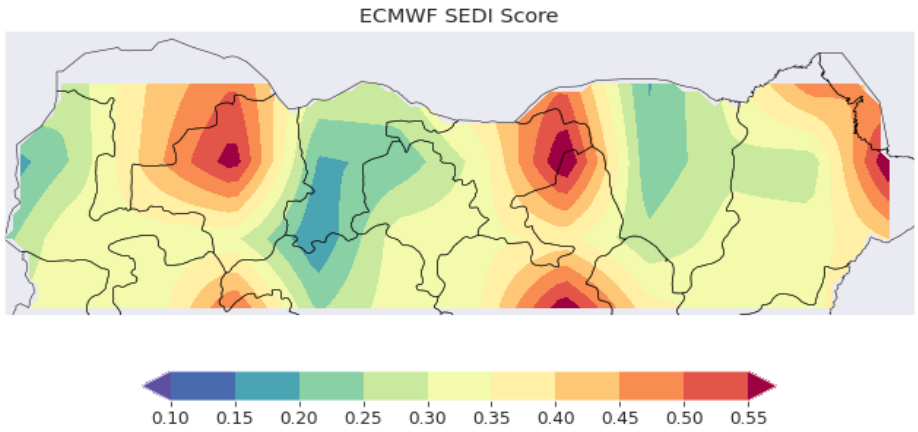

Figure 3.3.6: ECMWF SEDI Score at 30 – days lead period

3.4 DEEP LEARNING POST-PROCESSED OUTPUT

The output of the numerical models are used develop a deep learning model. The model takes the 30-day lead model data (ECMWF's SEAS5, NCEP's CFsv2 and the UK Met Office's GloSea5-GC2-LI) and the time steps as input. The model also takes the analyzed daily heat wave index as output which is derived from the ERA-5 reanalysis data. This, thus form a classification model that determine heat wave occurrence using the set of input data. The Deep learning bias corrected output has produced reliability maps for the dichotomous heat wave at a 30-day lead time, as shown in Fig 3.4.1. These maps reveal that the average reliability score is 0.9 in most parts of the region, indicating an improvement over the individual models. Additionally, the deep learning bias corrected output has identified localized areas with higher reliability scores, which have been consistently reliable.

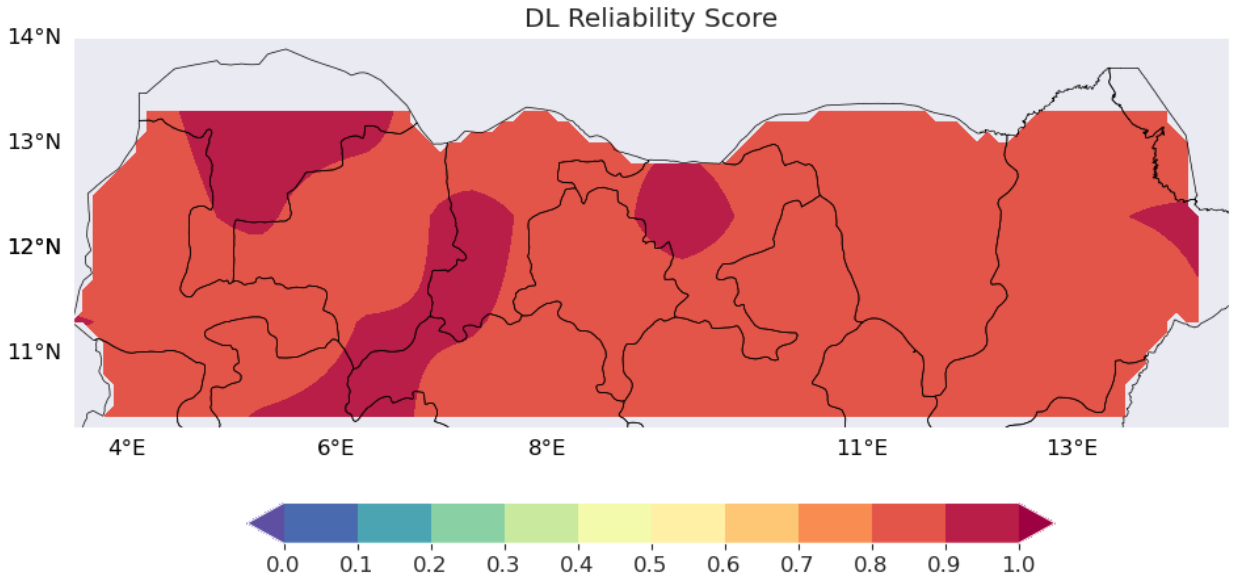

Figure 3.4.1: Deep Learning Model Reliability Score at 30 – days lead period

The temporal autocorrelation of dichotomous heat wave hindcasts of the deep learning model at 30-day lead period shows high value correlation at all region.

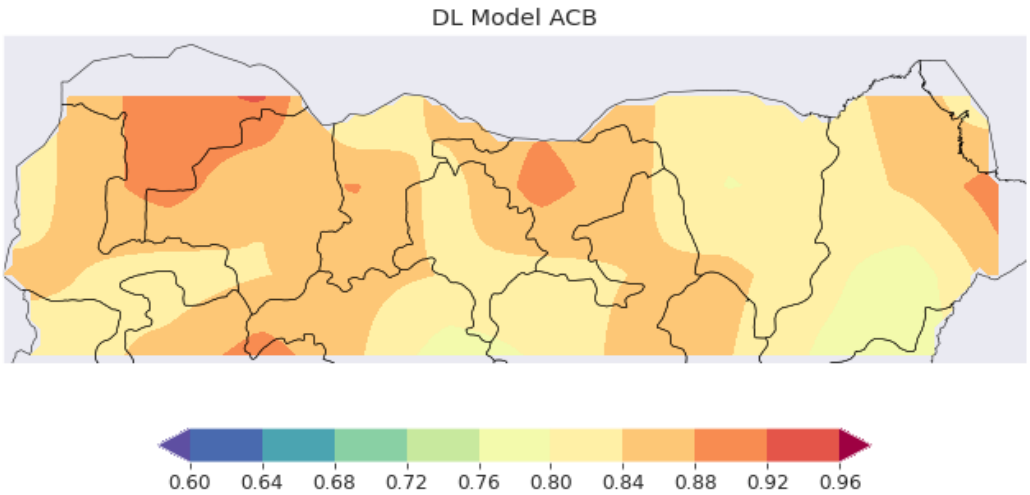

Figure 3.4.2: Deep Learning Auto-correlation bias at 30 – days lead period

The equitable threat score (ETS) maps of heat wave prediction of the deep learning model at 30-day lead time shows that the model generally has high reliability in the region. A high score of the metrics is observed at all areas of the region.

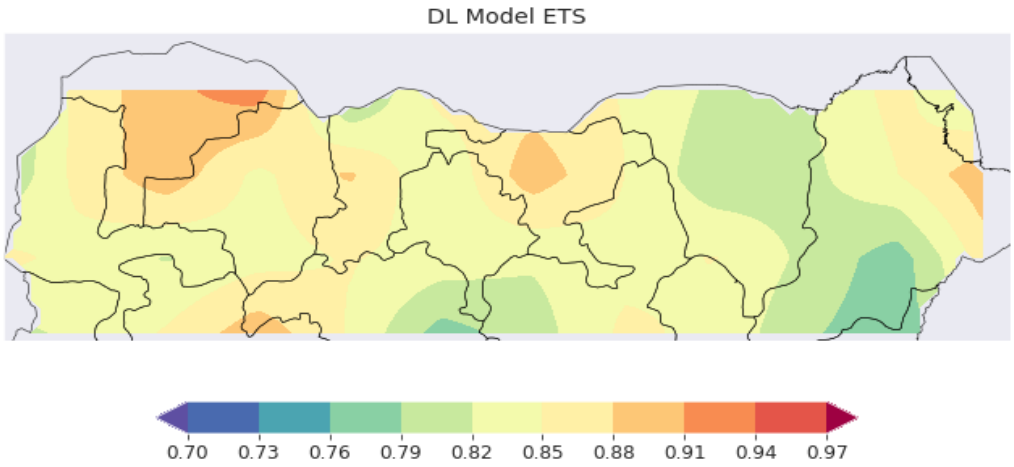

Figure 3.4.3: Deep Learning Equitable Threat score at 30 – days lead period

The hit rate performance of the deep learning model shows values of 0.9 or more in major areas of the region. Few areas are observed to have value of 0.8 within the region. This shows that the bias corrected model have a good skill at predicting the heat wave at exactly when they occur. In comparison with the False alarm rate score which is shown to be relatively low (less than 0.1) in all areas of the region, the deep learning model shows a better skill of reduced false alarm rate when compared to all the numerical model at all time steps.

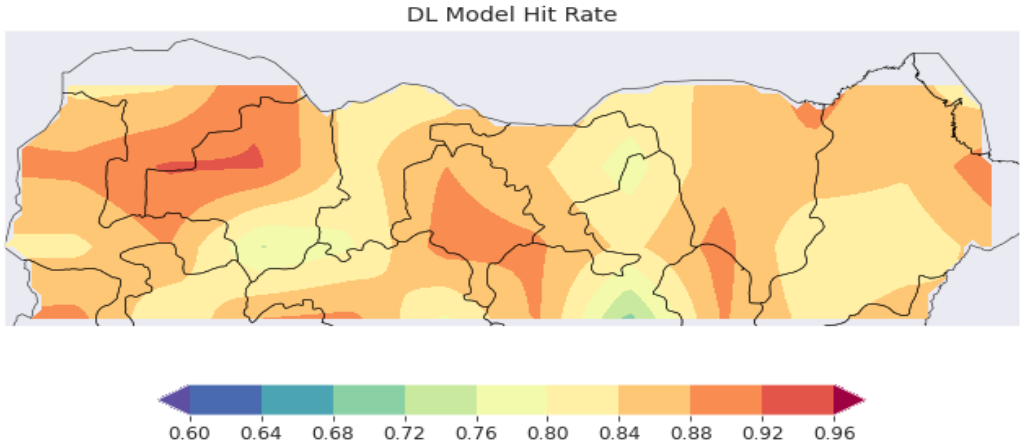

Figure 3.4.4: Deep Learning Hit Rate at 30 – days lead period

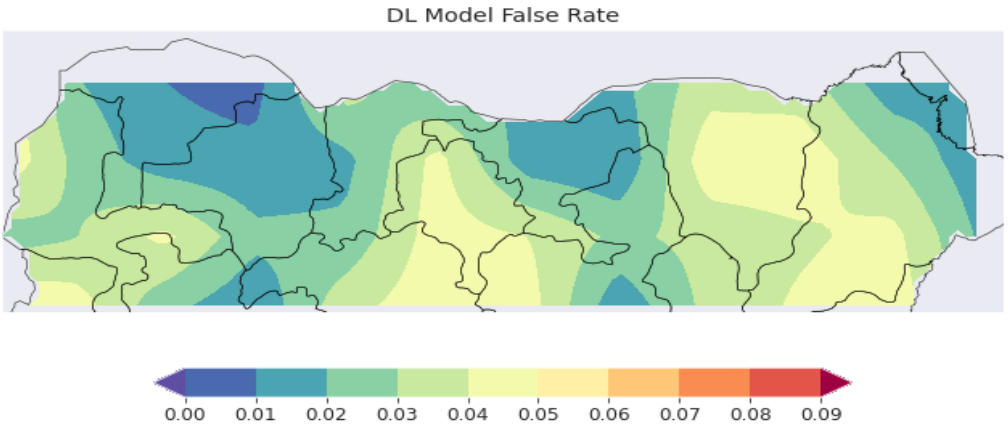

Figure 3.4.5: Deep Learning False Rate at 30 – days lead period

The SEDI score of the deep learning model at a 30-day lead time shows a good performance of the skill with all areas having value between 0.9 and 1.0. This shows a significant increase over the numerical models considered in this research at all the time leads.

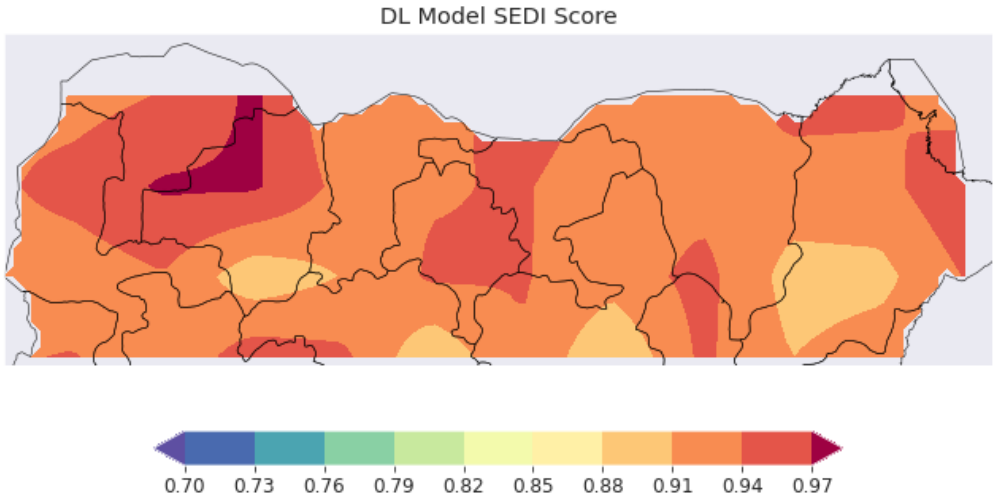

Figure 3.4.6: Deep Learning SEDI score at 30 – days lead period

The models' ability to predict heatwave occurrence in the region is generally observed to be insensitive to the evaluation metrics used, with all metrics showing results that support one another. However, the prediction skills vary spatially, depending on the land cover characteristics, which is specific to each model. For instance, the Argugu River in the western part of the region (longitude 4-50E), Hadejia River/watershed in the central part of the region (longitude 9-100E), and the Lake Chad basin (longitude 130E) have improved the models' heatwave prediction skills. Land-atmosphere feedbacks are a crucial factor in heatwave maintenance in southern latitudes, but the models may misrepresent these feedbacks, resulting in poor skills in the region.

4. Conclusion

This study evaluates the skills of sub-seasonal to seasonal heat wave prediction skills of Numerical models over the Northern Nigeria. The data used consists of the Numerical model data (hindcast and forecast data) of three different models namely, NCEP's CFsv2, UKMO's GloSea5-GC2-LI, ECMWF's SEAS5 and ERA 5 reanalysis datasets. The data output of both the NCEP's CFsv2 and UKMO's GloSea5-GC2-LI initialized at 5, 7, 15 and 30-day lead between year 2000 and year 2020 (21years data) are used. While the ECMWF's SEAS5 data used are at 30-day lead time of the same year range, the ERA 5 reanalysis was used to compare the skills of the models. The heat index used in this study is the excessive heat index (EHI). The EHI was used to characterize the occurrence of heat wave. This was computed on daily basis at each grid point on both the reanalysis data and the model output data. Grid point specific deep learning classification model was developed using the output of the numerical model and lead-time as input. The deep learning output gives a binary output of either 0 or 1 that represents heat wave occurrence or not respectively.

The heat wave predictability skills of the three numerical sub-seasonal to seasonal models are lower in the entire region. At all lead times, the models tends to overestimate heat wave frequency though it is more pronounced in the NCEP's CFsv2 model. The ECMWF's SEAS5 and UKMO's GloSea5-GC2-LI models also overestimates the heat wave frequency, but at a reduced rate (up to 44% decrease) from the NCEP's CFsv2 model. The overestimation of the heat wave frequency in the region contributes to the lower reliability scores and forecast skills of the models in the region.

The skills of the model at different time lead shows different behavior. The NCEP's CFsv2 model shows decrease in reliability of the model with decreasing time leads. This makes the model to have a fair performance at 30-day lead period when compared to other time leads. The equitable threat score show similar pattern while the SEDI score show the same pattern at only 15-day lead and 30-day lead. The hit rate and the false alarm rate shows a different trend, they both tend to increase with decreasing lead time. Thus, the maximum hit rate and false alarm rate are observed at 5-day lead period. The overestimation of heat wave due to bias towards maximum temperature was due to correlations between surface soil moisture and evaporative fraction during summer are generally high, indicating soil moisture's control on surface fluxes and the potential for land–atmosphere coupling. The UKMO's GloSea5-GC2-LI model on the other hand, shows increased reliability with decreasing time leads. This is shown from the pattern observed in the ETS metrics where the metrics shows higher score at 5-day and 7-day time lead while the score at 15-day and 30-day leads appear to be relatively low when compared to the first two time lead period. While the SEDI metrics shows similar pattern, the number of areas where the skill vanishes tends to be more at 5-day time lead. The hit rate shows no significant pattern but the false rate is observed to reduce with decreasing time lead. At all the time lead period, the UKMO's GloSea5-GC2-LI model is observed to be more reliable than the NCEP's CFsv2 model, the forecast skill of the UKMO's GloSea5-GC2-LI model surpasses that of NCEP's

CFsv2 model at all the time step with higher equitable threat score (ETS) than NCEP's CFsv2 model. While the NCEP's CFsv2 model SEDI score skill are lower or vanishes at depending on the time step, the UKMO's GloSea5-GC2-LI model still have a relatively high score in most areas at tall time step. The hit rate and false alarm rate on the other hand is highest in the NCEP's CFsv2 model while the UKMO's GloSea5-GC2-LI model have lower hit rate and false alarm rate.

At 30-day time lead, the ECMWF's SEAS5 model's forecast skill also beats the NCEP's CFsv2 model forecast skill but average the same skill as the UKMO's GloSea5-GC2-LI model. Although, the NCEP's CFsv2 model have higher value in terms of both hit rate and false alarm rate, the hit rate and false alarm rate value of ECMWF's SEAS5 model is lower than that of NCEP's CFsv2 model and UKMO's GloSea5-GC2-LI model. The deep learning model shows a more uniform score of all the evaluation metrics and still perform better and satisfactorily than the three numerical model at the 30-day time lead. The model averages 0.90 value for hit rate at all point, 0.02 value for false alarm rates, 0.94 SEDI score, 0.87 reliability and 0.75 ETS score. From the extensive analysis above, the UKMO's GloSea5-GC2-LI model performs best of the three numerical model at 5-day, 7-day and 15-day time lead. While the deep learning classification model performs better than all the model at 30-day time lead [1-25].

Statements & Declarations

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose."

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Ibraheem Raji and Ademola Akinbobola. The first draft of the manuscript was written by Ibraheem Raji and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data Availability

The datasets generated during and/or analysed during the current study are available in the Copernicus Climate Data store (CDS) repository,https://cds.climate.copernicus.eu/

References

1. Abatan, A. A., Abiodun, B. J., Lawal, K. A., & Gutowski Jr, W. J. (2016). Trends in extreme temperature over Nigeria from percentile-based threshold indices. International Journal of Climatology, 36(6), 2527-2540.
2. Abdussalam, A. F. (2015). Changes in indices of daily temperature and precipitation extremes in northwest Nigeria. Science World Journal, 10(2), 18-26.
3. Seun, A. I., Ayodele, A. P., Koji, D., & Akande, S. O. (2022). The potential impact of increased urbanization on land surface temperature over South-West Nigeria. Current Research in Environmental Sustainability, 4, 100142.
4. ADEFISAN, E. A., & AHMAD, S. (2018). ASSESSMENT OF HEAT WAVE EVENTS IN A CHANGING CLIMATE OVER NIGERIA. International Journal of Research Science and Management, 5(8), 115-133.
5. ADEFISAN, E. A., & AHMAD, S. (2018). ASSESSMENT OF HEAT WAVE EVENTS IN A CHANGING CLIMATE OVER NIGERIA. International Journal of Research Science and Management, 5(8), 115-133.
6. Easterling, D. R. et al. (1997). Maximum and minimum temperature trends for the globe. Science, 277, 364 - 367.
7. Ferro, C. A., & Stephenson, D. B. (2011). Extremal dependence indices: Improved verification measures for deterministic forecasts of rare binary events. Weather and Forecasting, 26(5), 699-713.
8. Field, C. B., & Barros, V. R. (Eds.). (2014). Climate change 2014–Impacts, adaptation and vulnerability: Regional aspects. Cambridge University Press.
9. Gbode, I. E., Akinsanola, A. A., & Ajayi, V. O. (2015). Recent changes of some observed climate extreme events in Kano. International Journal of Atmospheric Sciences, 2015.
10. Guirguis, K., Gershunov, A., Tardy, A., & Basu, R. (2014). The impact of recent heat waves on human health in California. Journal of Applied Meteorology and Climatology, 53(1), 3-19.
11. Hogan, R. J., Ferro, C. A., Jolliffe, I. T., & Stephenson, D. B. (2010). Equitability revisited: Why the "equitable threat score" is not equitable. Weather and Forecasting, 25(2), 710-726.
12. Solomon, S. (Ed.). (2007). Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC (Vol. 4). Cambridge university press.
13. IPCC. (2007a). Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK. Cambridge University Press.
14. IPCC. (2007b). Climate Change 2007: Mitigation of Climate Change. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK. Cambridge University Press.
15. IPCC. (2007c). Climate change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Geneva, Switzerland:IPCC.
16. IPCC. (2007d). Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK. Cambridge University Press.
17. Janiga, M. A., Schreck, C. J., Ridout, J. A., Flatau, M., Barton, N. P., Metzger, E. J., & Reynolds, C. A. (2018). Subseasonal

forecasts of convectively coupled equatorial waves and the MJO: Activity and predictive skill. Monthly Weather Review, 146(8), 2337-2360.
18. Kunkel, K. E., Pielke, R. A., & Changnon, S. A. (1999). Temporal fluctuations in weather and climate extremes that cause economic and human health impacts: A review. Bulletin of the American Meteorological Society, 80(6), 1077-1098.
19. Meehl, G. A., & Tebaldi, C. (2004). More intense, more frequent, and longer lasting heat waves in the 21st century. Science, 305(5686), 994-997.
20. Nairn, J., & Fawcett, R. (2011). Defining heatwaves: heatwave defined as a heat-impact event servicing all. Europe, 220, 224.
21. Perkins, S. E. (2015). A review on the scientific understanding of heatwaves—Their measurement, driving mechanisms, and changes at the global scale. Atmospheric Research, 164, 242-267.
22. Robinson, P. J. (2001). On the definition of a heat wave. Journal of Applied Meteorology and Climatology, 40(4), 762-775.
23. Rosenzweig, C., Iglesius, A., Yang, X. B., Epstein, P. R., & Chivian, E. (2001). Climate change and extreme weather events-Implications for food production, plant diseases, and pests.
24. Stott, P. (2016). How climate change affects extreme weather events. Science, 352(6293), 1517-1518.
25. Ford, T. W., Dirmeyer, P. A., & Benson, D. O. (2018). Evaluation of heat wave forecasts seamlessly across subseasonal timescales. NPJ Climate and Atmospheric Science, 1(1), 20.