ECMWF System 4 forecasts for Malaria in Ghana

This analysis employs the European Centre for Medium Range Weather Forecasting (ECMWF) System 4 hindcast ensemble of 15 members starting on the 1st of every month for the years 1981-2010, and run for 7 months. The system consists of an initial ocean analysis to estimate the initial state of the ocean and a global coupled ocean-atmosphere general circulation model to calculate the evolution of the ocean and atmosphere. Daily temperature and precipitation from the System 4 ensemble are then used to drive the Liverpool Malaria Model (LMM), with a one-year spin up period driven by ERA-Interim climatological values for Africa. The output is compared to various sets of gridded climatological products:

- Temperature observations from the Climate Research Unit (CRUT3.1, Mitchell and Jones, 2005) covering the period 1950-2009.
- Mixed satellite and rain gauges observations from the Global Precipitation Climatology Project (GPCP) dataset (Huffman et al, 2001). Monthly values are available for rainfall between 1979 and 2010, however daily values required to drive LMM were only available between 1997 and 2008.
- Mixed satellite and rain gauges observations from the NASA Goddard Space Flight Center Tropical Rainfall Measuring Mission (TRMM) dataset (Huffman et al, 2001) for 1998-2010.
- Temperature and rainfall products based on NCEP-NCAR (1948-2010, Kalnay et al, 1996) and ERA Interim for (ERAI, 1979-2011, Uppala et al, 2008) reanalyses (blend of climate model outputs and various sources of observations using complex assimilation methods).

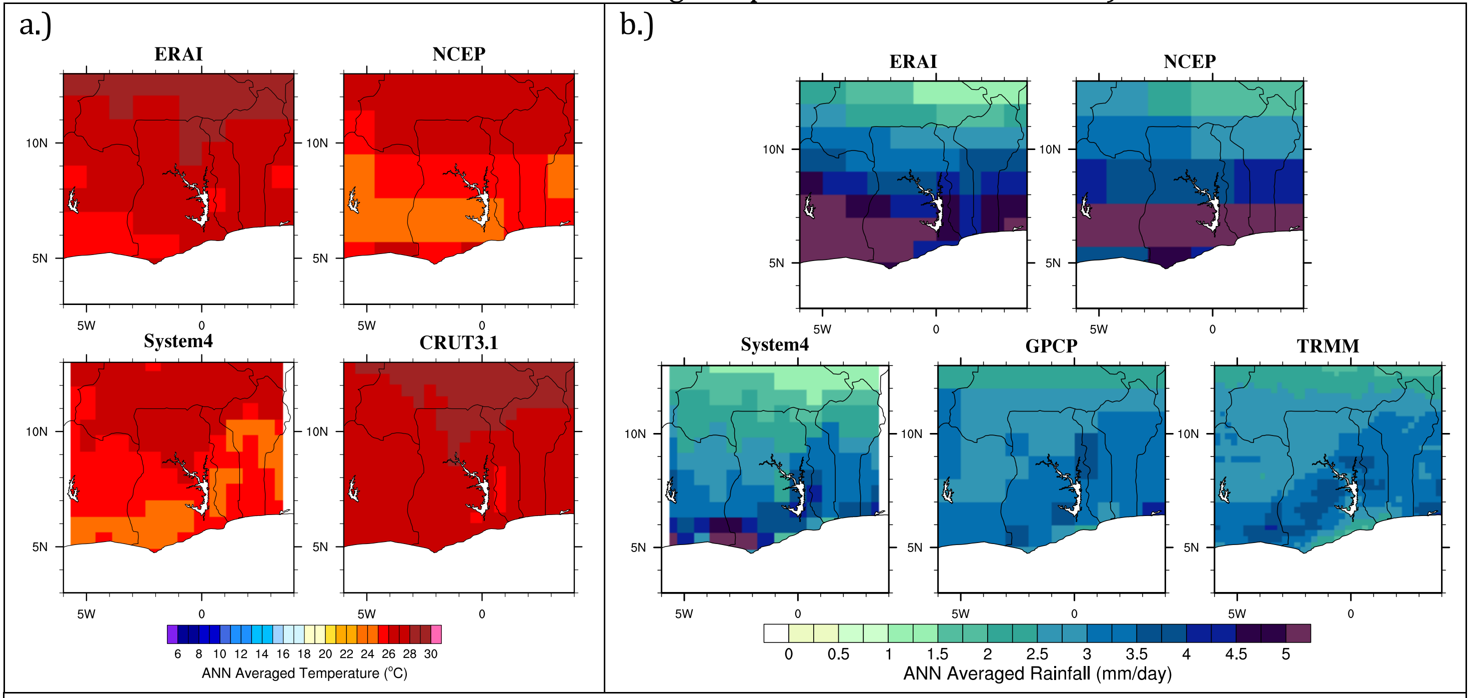

Figure 1: Annual average a.) temperature and b.) precipitation for Ghana. System 4 forecasts take the fourth month forecasted from each start date (see table 1 for explanation).

The annual averages for System 4 (forecast month 4, see table 1 for explanation) compare reasonably well with the climatological values, with cooler coastal temperatures than inland (compare NCEP and System 4 in Fig.1a) but a few degrees cooler overall compared to the observations (CRUTS). Annual averaged rainfall is elevated near the Gulf of Guinea in all data (Fig.1b), but System 4 compares particularly well against the hybrid satellite/model climatologies of GPCP and TRMM, with significant excess rainfall in ERAI and NCEP.

The seasonal cycles of temperature and precipitation (Fig.2, averaged exclusively over Ghana) are considered for northern and southern regions separately to distinguish their different climatological characteristics.

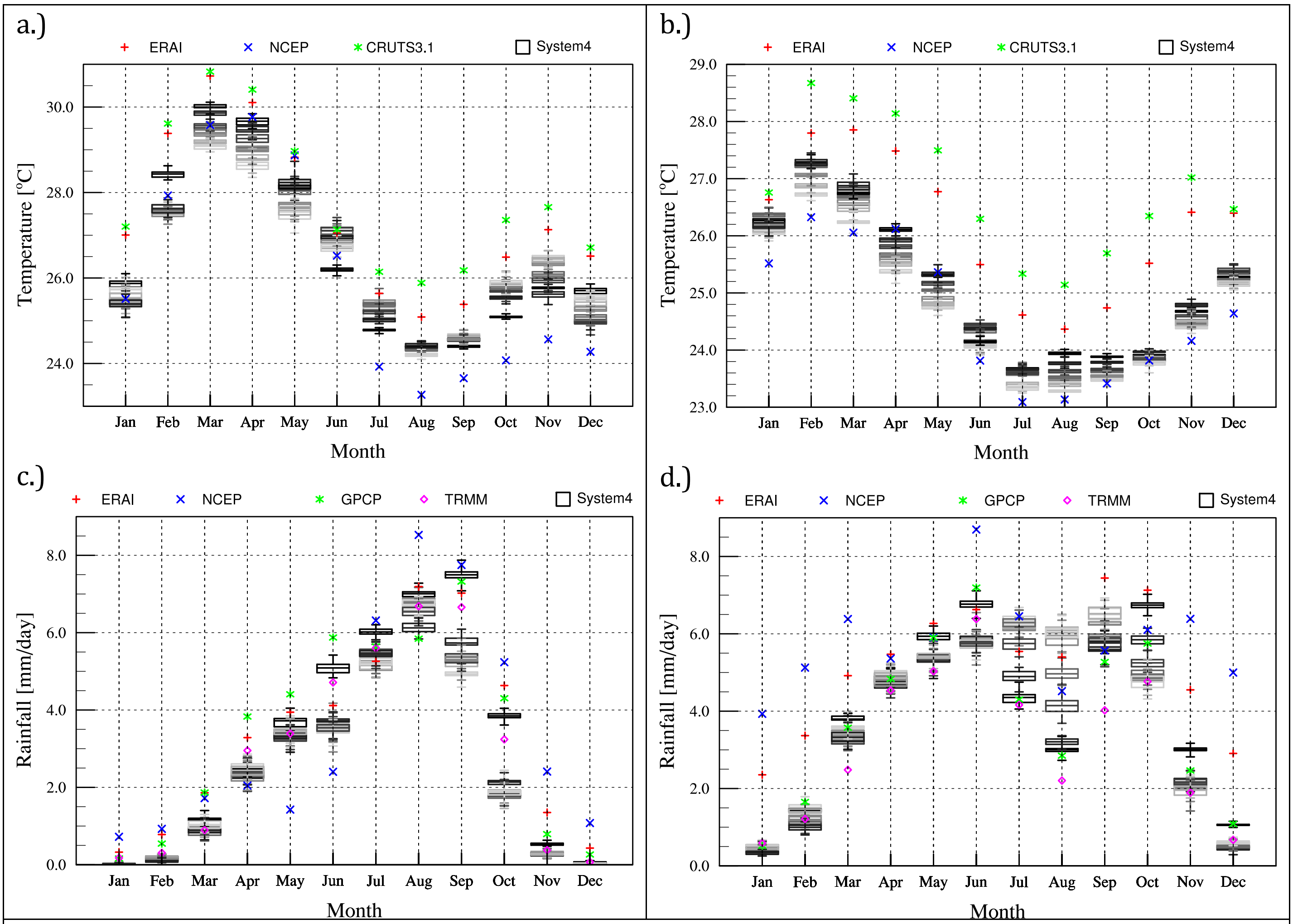

Figure 2: Seasonal cycle of (a. and b.) temperature and (c. and d.) precipitation in North Ghana (left) and South Ghana (right) separated at 8°N. Data were masked using the Database of Global Administrative Areas (GADMv2.1). System 4 forecasts are shown as box and whisker plots for forecast months 1 to 7 (black – light grey) for the lower tercile, mean and upper tercile values.

Table 1: Illustration of the System 4 seasonal forecast scheme. Each System 4 forecast is seven months long, with a start date in each month of the year. The "Forecast month" is the distance in time from the in initial conditions with 1 being the first month of the forecast (same month as the start date) and 7 being the final month of the forecast. To build a timeseries, we chose the value for "Forecast month" and extract that month's data from each start date. For example, for Forecast month = 1, we take Jan from the first start date in 1981, Feb from the second start date in 1981, etc. For Forecast month = 7, we take Jul from the first start date in 1981, Aug from the second start date in 1981, etc. When Forecast month > 1, the timeseries extends into 2011 since the last start date (12) in 2010 gives Dec when Forecast month=1, Jan when Forecast month=2 to Jun when Forecast month=7.

Forecast month	1981												...	2010			2011					
1	J	F	M	A	M	J	J	A	S	O	N	D	...	J	...	D						
2		F	M	A	M	J	J	A	S	O	N	D	...	J	...	D	J					
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
6						J	J	A	S	O	N	D	...	J	...	D	J	F	M	A	M	
7							J	A	S	O	N	D	...	J	...	D	D	F	M	A	M	J

In both regions, temperature shows a distinct seasonal cycle, captured by all three climatologies and the System 4 forecasts, with the warmest temperatures in February and March in South Ghana and slightly later in March and April for North Ghana. Coolest temperatures appear consistently in JAS. Precipitation, on the other hand, shows a single rainfall peak in JAS in northern Ghana, which is fairly well captured by all datasets. Southern Ghana has two rainy seasons in AMJ and September-October associated with the passing of the Intertropical Convergence Zone (ITCZ). There is relatively small spread in System 4 forecast distances of 1 to 7 months from the start date (see Table 1), although the dry interlude in August varies considerably with GPCP, TRMM and early System 4 forecasts suggesting only 2-3mm/day while ERAI, NCEP and later System 4 forecasts suggest ~5-6mm/day.

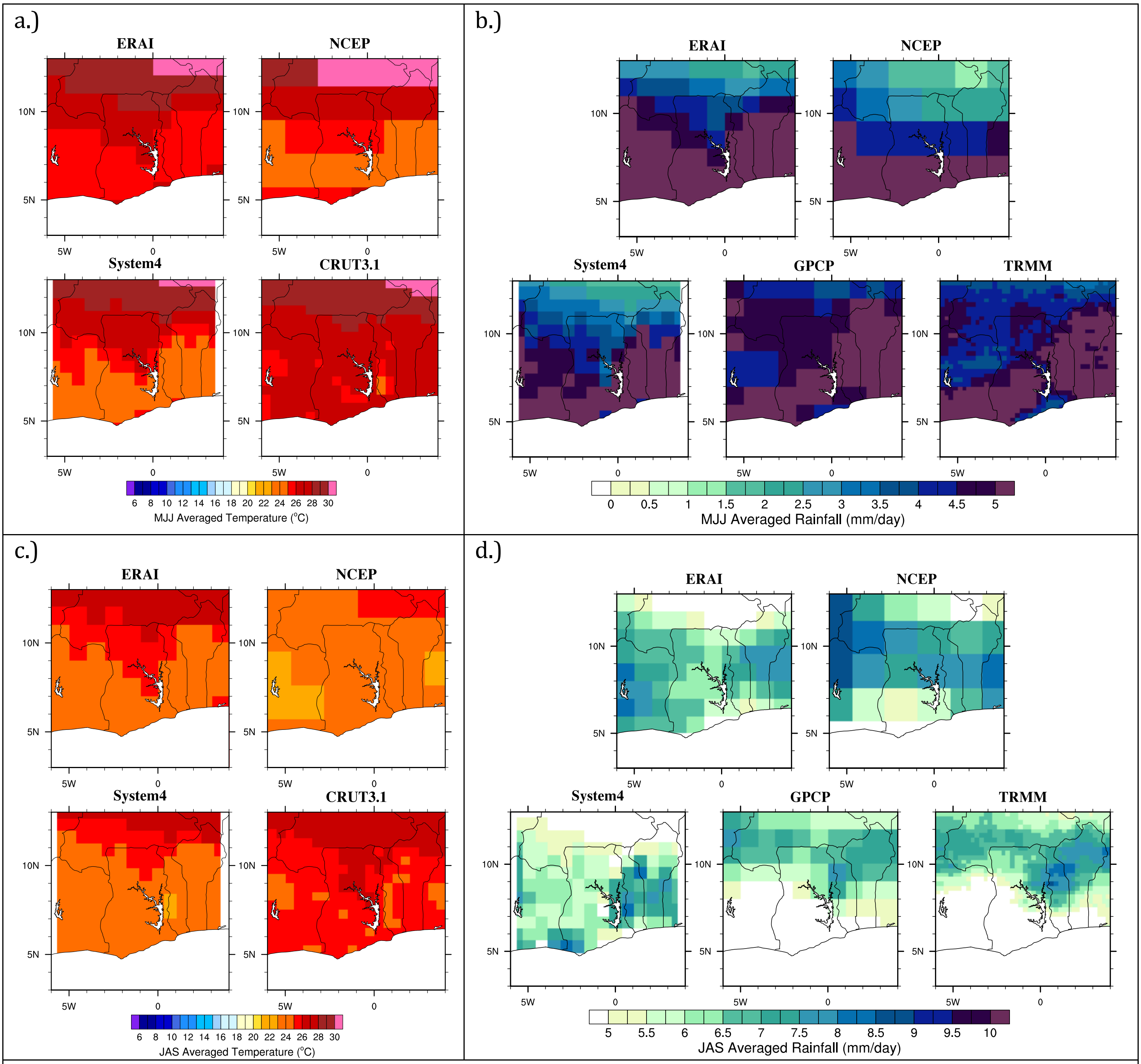

Figure 3: MJJ average a.) temperature and b.) precipitation and JAS c.) temperature and b.) precipitation for Ghana. System 4 forecasts use a four-month lead time (Feb start dates for MJJ, Apr start dates for JAS.

Distributions of average MJJ temperature and rainfall (the South Ghanaian rainfall peak, Fig.3a,b) compare well between datasets, particularly for rainfall, with ERAI, NCEP and System 4 only slightly drier than GPCP and TRMM in the north of the country. Average distributions of temperature and rainfall for the JAS peak in rainfall in North Ghana (Fig.3c,d) compare reasonably well with the other datasets, although the band of elevated rainfall in GPCP and TRMM is noticeably further north than in ERAI, NCEP and the System 4 forecast. There is agreement between precipitation in Togo and East Ghana between all datasets, although rainfall in West Ghana and Côte d'Ivoire is lower in GPCP, TRMM and the System 4 forecast compared to the reanalysis datasets.

The seasonal cycle of malaria in Ghana, simulated by the Liverpool Malaria Model (LMM) varies considerably between the north and south of the country (Fig.4). In North Ghana, simulated incidence peaks in SON, showing high levels of incidence of more than 50% in July and August following the nearly linear increases in rainfall between March and August (Fig.2c). The variability in precipitation predicted by the System 4 forecasts in June and July might well be contributing to the wide range of forecast malaria incidence in July and August. However, this variability is also captured within the range of the climatological data in July, with TRMM and NCEP suggesting low values and ERAI and GPCP suggesting high values. The peak malaria incidence in SON is closely mirrored in all datasets with the exception of NCEP, which appears to lag the other climatologies by one month, being 40% too low in September and ~20% too high in October.

In South Ghana, malaria incidence is particularly variable in the first half of the year largely associated with overly wet conditions provided by ERAI and NCEP, between November and March (Fig.2d). The hybrid satellite/rain gauge climatologies GPCP and TRMM indicate a much drier November to March and therefore lower malarial levels in FMA. System 4 forecasts only a short distance from initialisation (forecast months 1 and 2) are susceptible to this effect with malaria incidences of 20-40% between January and April as a result of the overly wet ERA-Interim initial conditions. However, increasing the forecast lead to months 3-7 greatly reduces the occurrence of malaria to less than 10% in FMA, consistent with GPCP and TRMM. Subsequently, between June and December, incidence rates are consistently between 60-80% with some seasonal variation in GPCP and TRMM associated with drying in August and onset of the second peak in rainfall.

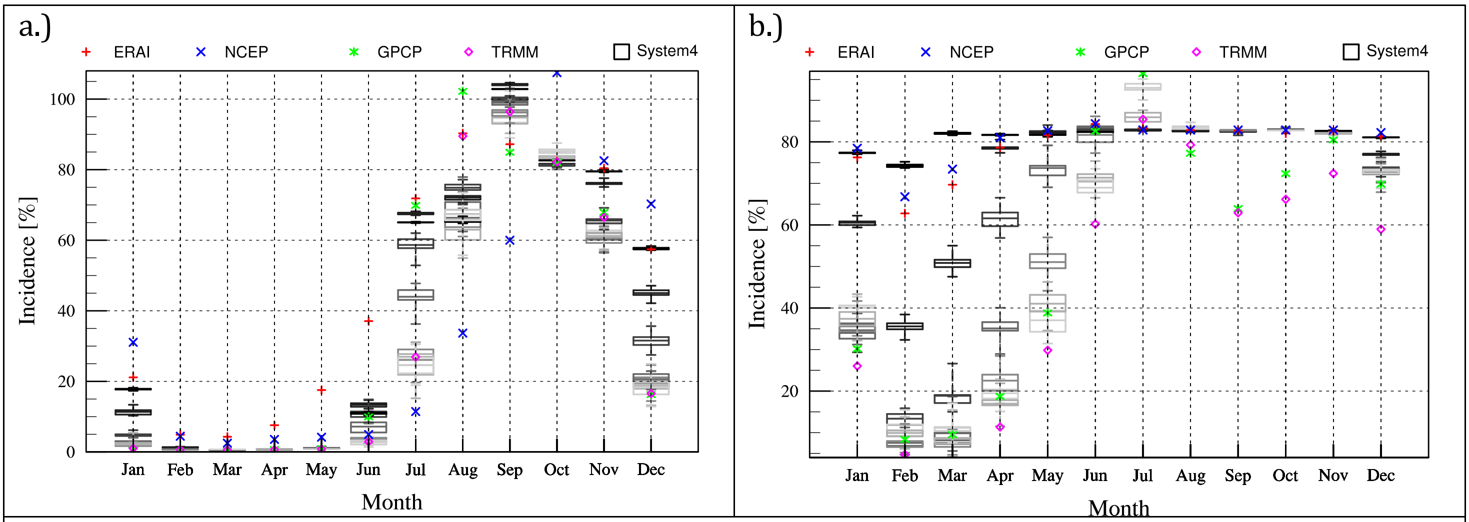

Figure 4: Seasonal cycle of malaria incidence in a.) North Ghana and b.) South Ghana, separated at 8°N. Data were masked using the Database of Global Administrative Areas (GADMv2.0). System 4 forecasts are shown as box and whisker plots for forecast months 1 to 7 (black – light grey) for the lower tercile, mean and upper tercile values.

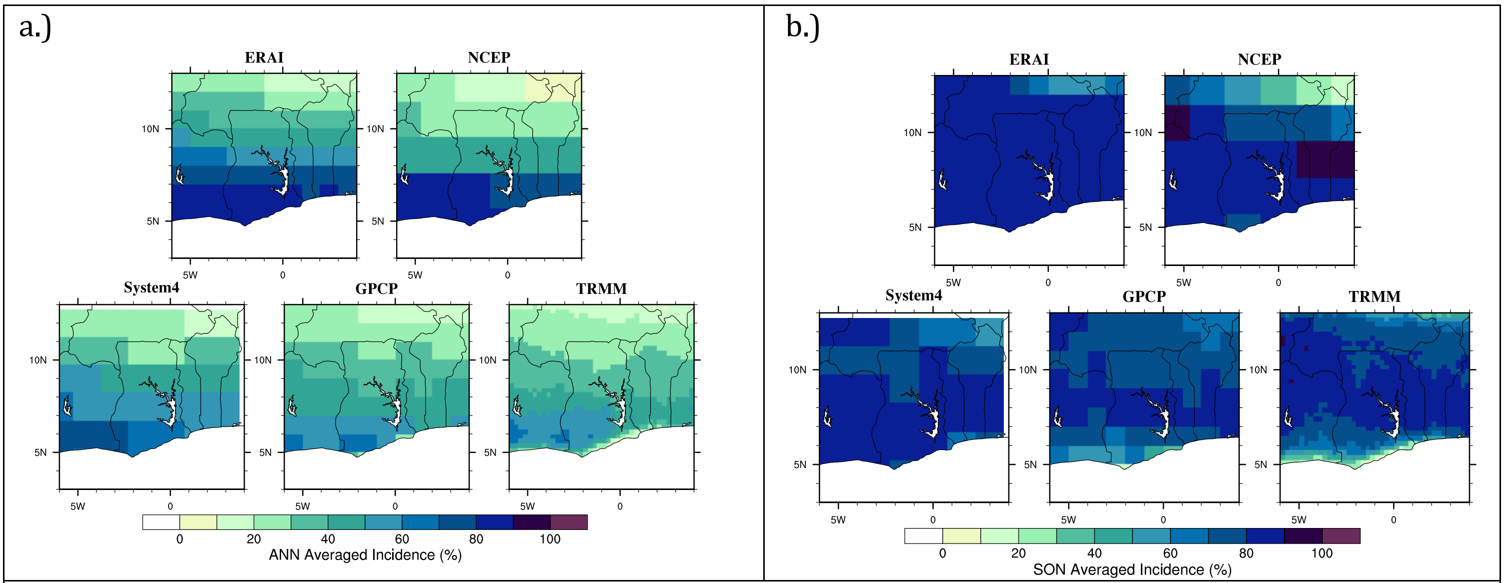

Figure 5: Malaria Incidence a.) annually averaged and b.) averaged over the peak malaria season SON. System 4 forecasts use a four-month lead-time using June start dates for an SON target.

The distribution of malaria incidence is comparable across the range of temperature and rainfall products used to drive the model in the annual average (Fig.5a), with elevated average incidence of 50-60% near the coast of the Gulf of Guinea and reduced values inland. Incidence values for the reanalyses-forced runs (ERAI and NCEP) are systematically higher than those forced by either of the satellite/rain-gauge climatologies in the southern-most region of Ghana due to elevated average rainfall compared to GPCP or TRMM. The System 4 forecast has slightly elevated incidence close to the coast, but is much more comparable with GPCP and TRMM than the reanalyses products ERAI and NCEP. For the SON peak incidence season (Fig.5b), malaria shows a peak band stretching across North Ghana at roughly 10°N in GPCP and TRMM. System 4 and ERAI only partially capture this feature, largely in the neighbouring countries while NCEP actually suggests lower incidence in this location.

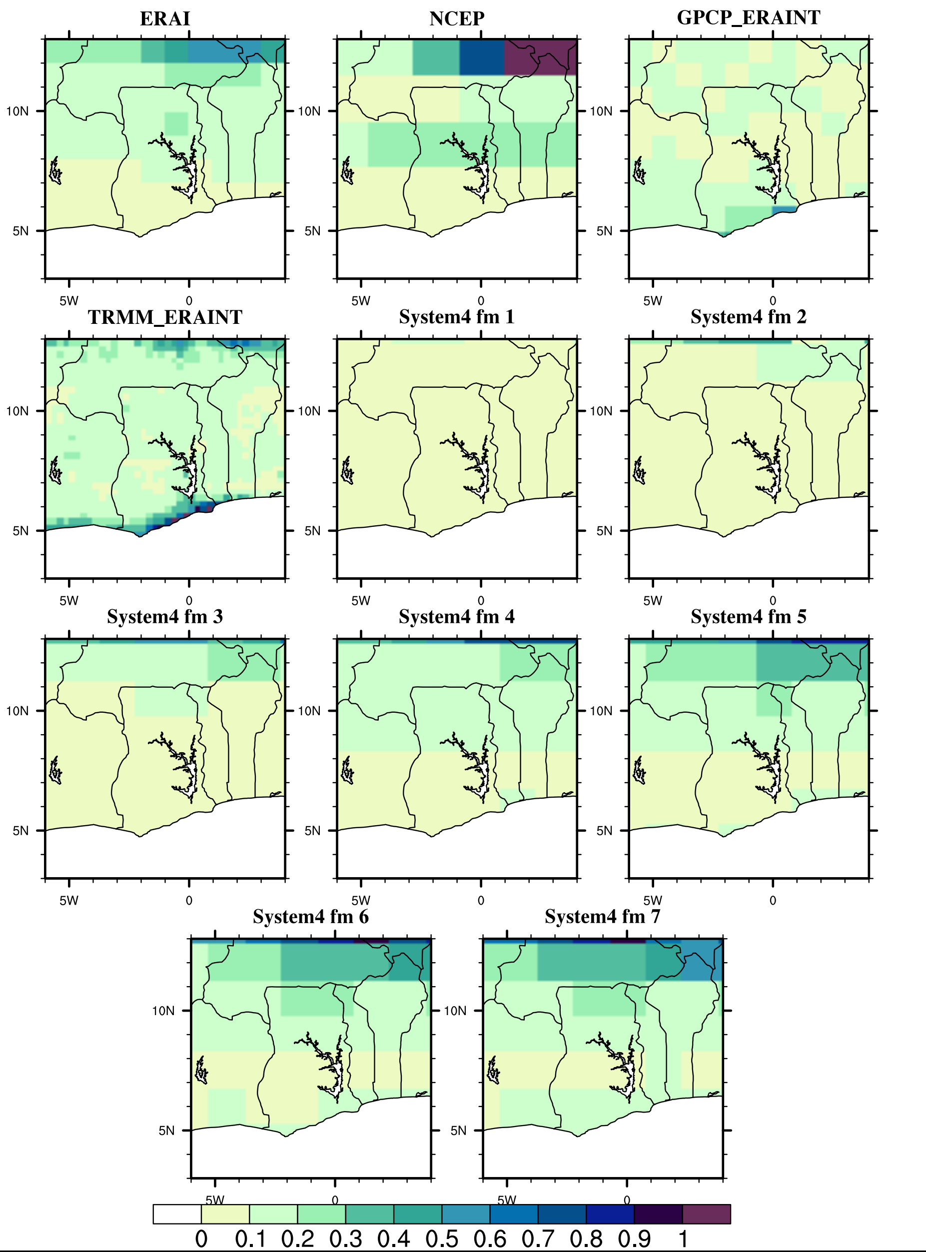

Figure 6: Malaria incidence coefficient of variation (COV) which is the standard deviation divided by the annual mean for the datasets driving LMM. System 4 values are shown for each month of the forecast.

The epidemic fringe to the north of Ghana (Fig.6) is present in ERAI, NCEP and TRMM, whereas only TRMM, and to a lesser extent GPCP produce high malaria variability along the south Ghana coast. System 4 values suggest that long lead times of 4-6 months (forecast months 5-7) capture the high interannual variability of malaria incidence in the North of Ghana.

The same diagnostic, repeated for only the peak malaria season (SON) with a four month System 4 forecast lead time produces a similar result with the epidemic fringe to the north of Ghana highlighted in ERAI, NCEP, TRMM and System 4, as well as high variability in the south of Ghana in GPCP and TRMM (Fig.7). This northern band of high variability in Figures 5 and 6 result from the southward location of the entire West African epidemic fringe, which is also relatively broad in ERAI and NCEP compared to

TRMM, which just encroaches into the focused region around Ghana. For GPCP, the fringe lies just to the north.

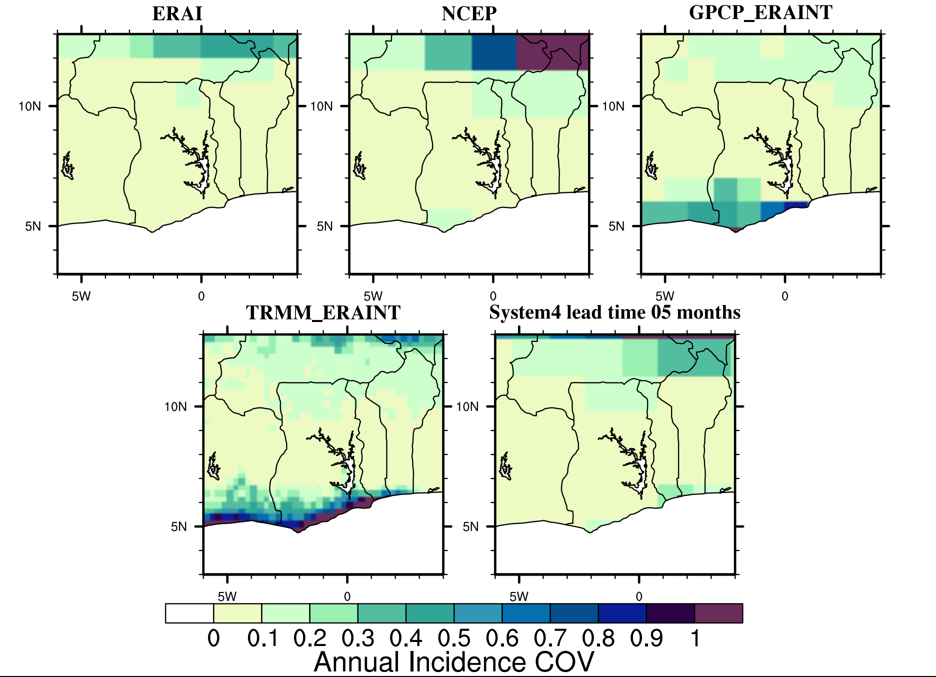

Figure 7: Malaria incidence coefficient of variation (COV) which is the standard deviation divided by the SON mean for the datasets driving LMM. System 4 values use a four-month lead-time using June start dates for an SON target.

Interannual variability of temperature, precipitation and malaria incidence averaged over North and South Ghana for the peak malaria season (SON) and specific precipitation seasons are shown in Fig.8, with the System 4 forecasts taken with increasing lead times from the target season, allowing five realisations (see table 2 for explanation).

Table 2: Illustration of the different System 4 forecast start dates used for the two target seasons to calculate interannual variability and Relative Operating Characteristics (ROC) curve areas at several forecast lead times. Three month seasonal averages could not be computed for lead times of 6 and 7 months.

Forecast Month (fm)	1	2	3	4	5	6	7
MJJ Target Start Date	5 (May)	4 (Apr)	3 (Mar)	2 (Feb)	1 (Jan)	12 (Dec)	11 (Nov)
JAS Target Start Date	7 (Jul)	6 (Jun)	5 (May)	4 (Apr)	3 (Mar)	2 (Feb)	1 (Jan)
SON Target Start Date	9 (Sep)	8 (Aug)	7 (Jul)	6 (Jun)	5 (May)	4 (Apr)	3 (Mar)
Target Months	1-3	2-4	3-5	4-6	5-7	N/A	N/A

Interannual variability for temperature in SON is reasonably well captured by System 4 forecasts, even at short forecast lead times, with particularly good correspondence with the CRUTS observational data. The variability in South Ghana is smaller in magnitude and the pattern is well captured by all datasets although the variability within the ERAI dataset is sometimes missed by the System 4 forecasts. For North Ghana there is a wider spread in variability and forecast lead-time does influence whether the System 4 results are entirely consistent with the climatological data.

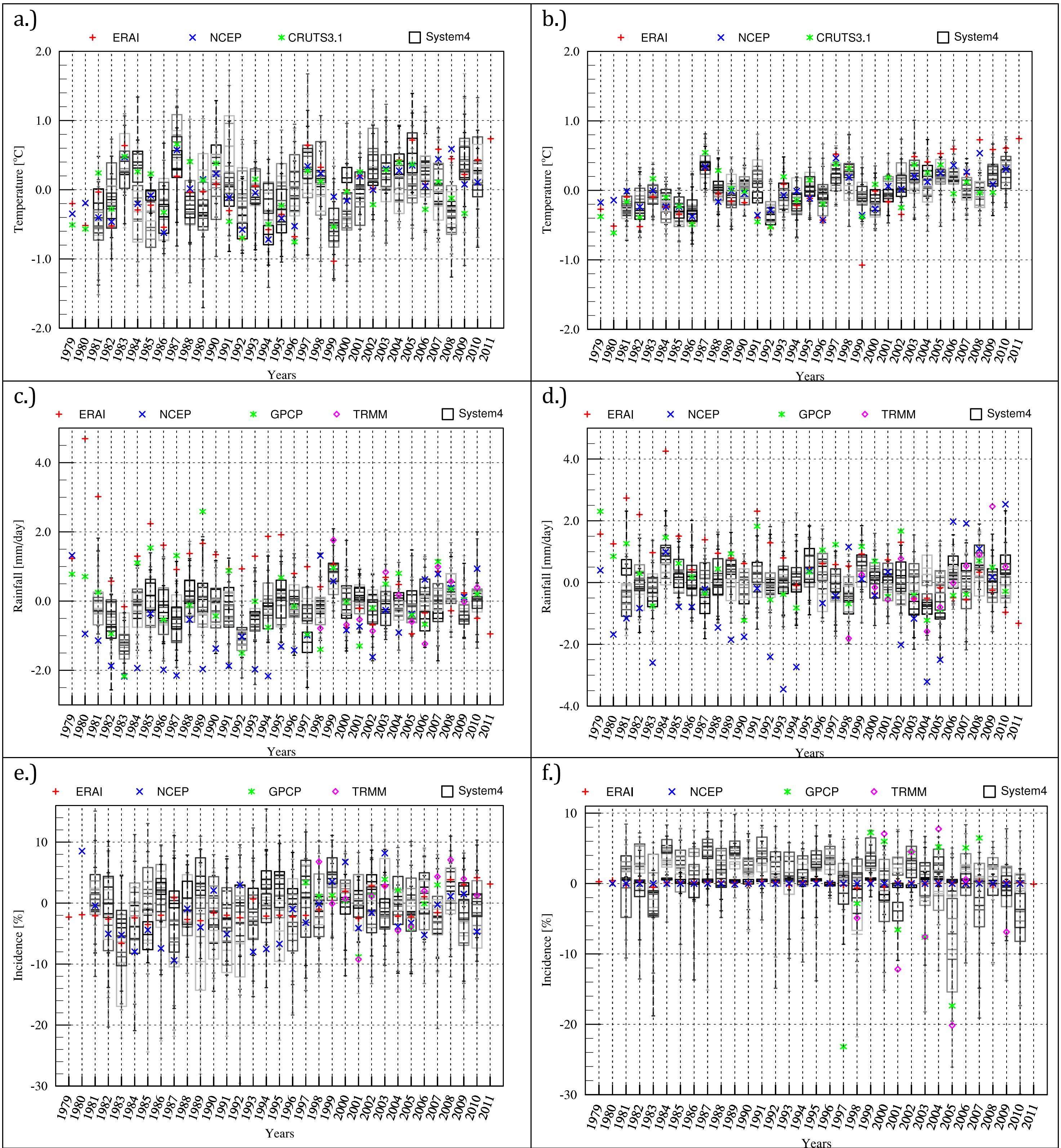

Figure 8: Interannual variability of (a,b.) SON temperature, (c.) JAS precipitation, (d.) MJJ precipitation and (e,f) SON malaria incidence for North Ghana (left) and South Ghana (right). The values are anomalies with a mean value subtracted that is calculated as the multi-year mean of each data product over their common, overlapping window of data availability. System 4 forecasts are plotted with increasing forecast start date lead times from 1 (black) to 5 (light grey) – see table 2 for details.

As was noticed in the Figs.2 and 4 for the seasonal cycles of precipitation and malaria incidence in Southern Ghana, System 4 forecasts with short lead times (black and dark grey boxes) perform poorly against climatological data, with improvement as the forecast moves away from initial conditions. Again, the System 4 forecasts compare better to the hybrid observation datasets GPCP and TRMM, where available, with poor correspondence with the climatological reanalyses ERAI and NCEP. A similar situation exists for Northern Ghana, where ERAI and NCEP values for precipitation and malaria incidence lie outside the ensemble range of the System 4 forecasts. The is improved agreement between the forecasts and GPCP and TRMM datasets, however the forecasts sometimes do not capture their Interannual variability well.

Skill in the System4 forecasts is determined by computing the geometrical area under the Relative Operating Characteristics (ROC) curve at each grid point for below lower tercile events, above median events and above upper tercile (extreme) events compared to the other climatologies. For perfect forecasts, all ensemble members will correctly predict an event (or non-event) in all years with a ROC area under the curve of 1.0, the maximum possible value. Forecasts with little or no skill that perform no better than climatology return ROC areas of less than 0.5. ROC areas were calculated for targets of the three-month peak rainfall season in North and South Ghana (MJJ and JAS, respectively) and the peak malaria season (SON), with 5 increasing forecast lead times (see table 2).

The significance of these values can be determined by performing a two-tailed Mann-Whitney U-statistic test, using a look-up table referencing the number of events and non-events, and calculating a critical ROC area value that must be exceeded. For example, upper tercile events, with a probability of 33%, in a 30-year timeseries, should produce 10 events (N_e) and 20 non-events (N_n). At the 95% confidence level, the critical U value from a reference table is 55, giving a ROC area of $0.5+U/(N_eN_n) = 0.775$. Calculated ROC scores comparing upper tercile events in 30 years of data that exceed 0.775 are therefore significant where $p<0.05$.

There is a significant skill in forecasting SON temperatures in Ghana compared to the three climatological datasets (Fig.9), which declines as forecast lead-time increases although a three to four month forecast still has a relatively high level of skill in the three categories. The System 4 forecasts are noticeably less skillful compared to the observational CRUTS3.1 climatology, presumably because System 4 is consistently cooler in the border region between Ghana and Togo (see Fig.1a, for example). Indeed, Southwest Ghana consistently retains significant forecast skill except for 5 month lead-times compared to observations.

Contrary to this, there appears to be little skill in the System 4 forecasts of MJJ rainfall (Fig.10) compared to climatology. Nevertheless, low-lead time forecasts compared to NCEP and GPCP are significantly skillful across South Ghana. The situation is somewhat improved for forecasts of JAS rainfall (Fig.11), where System 4 compared to NCEP is significantly skillful across the north of the country, as System 4 compared to GPCP (at low lead times) and TRMM (at slightly longer lead times).

ROC skill for malaria incidence in SON (Fig.12) should be interpreted in conjunction with geographical incidence variability (Fig.6 and 7) since the useful skill occurs in regions of high interannual variability where epidemics are possible. With this in mind, the System 4 forecasts are not significantly skillful in North Ghana compared to ERAI or NCEP, although there might be moderate skill in mid-range forecasts when compared to GPCP and TRMM although this is fairly patchy and only significant in isolated regions.

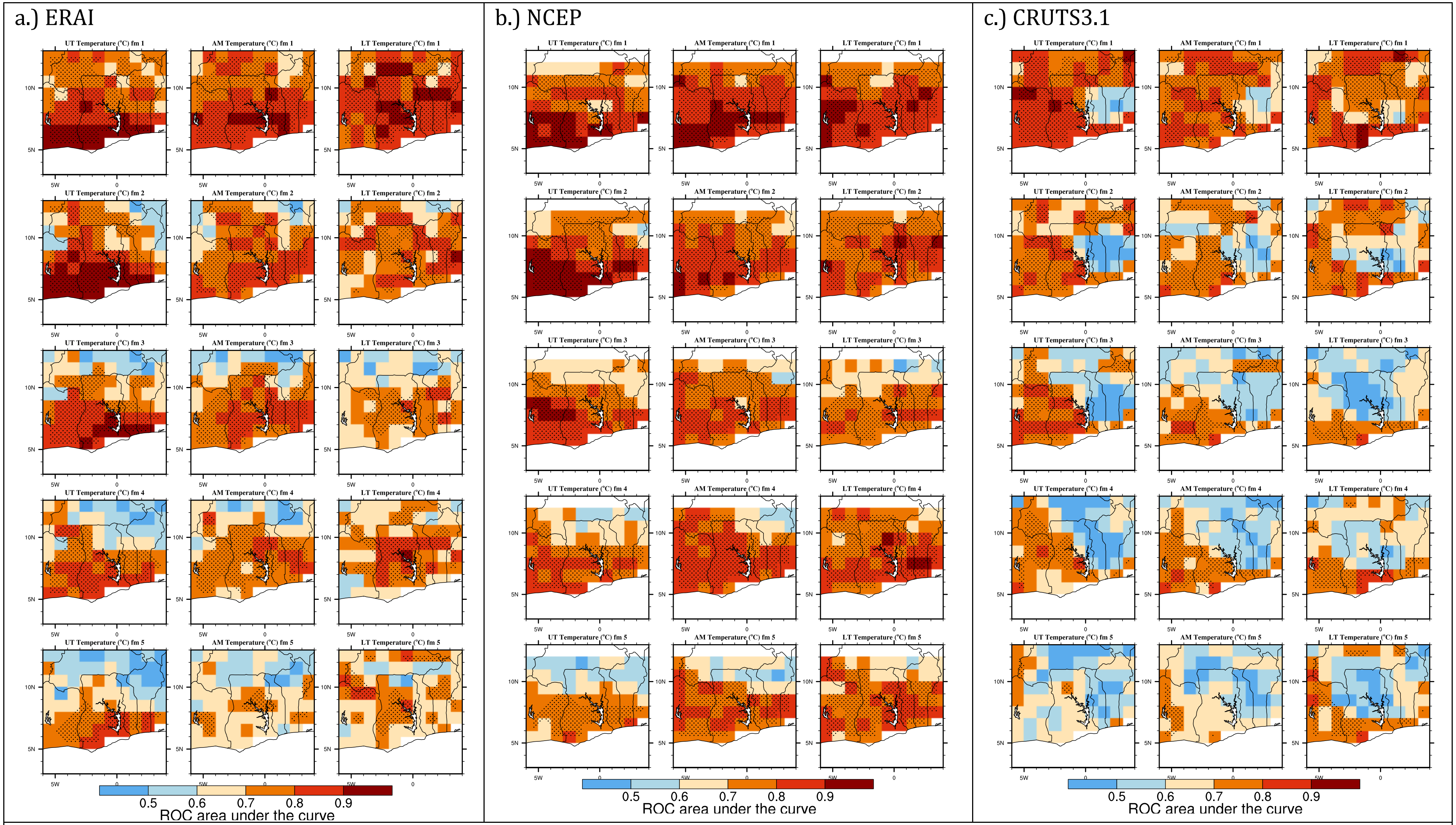

Figure 9: Relative Operating Characteristics (ROC) areas under the curve for SON season temperature forecast skill for above upper tercile (extreme) events (UT, left column), above median events (AM, central column) and below lower tercile events (LT, right column) compared to a.) ERAI, b.) NCEP and c.) CRUTS3.1 climatological data. The five different rows represent increasing forecast lead times from 1 month (top row) to 5 months (bottom row) of the target season (see table 2). Stippling indicates significance at the 95% confidence interval, calculated by performing a two-tailed Mann-Whitney U-statistic test.

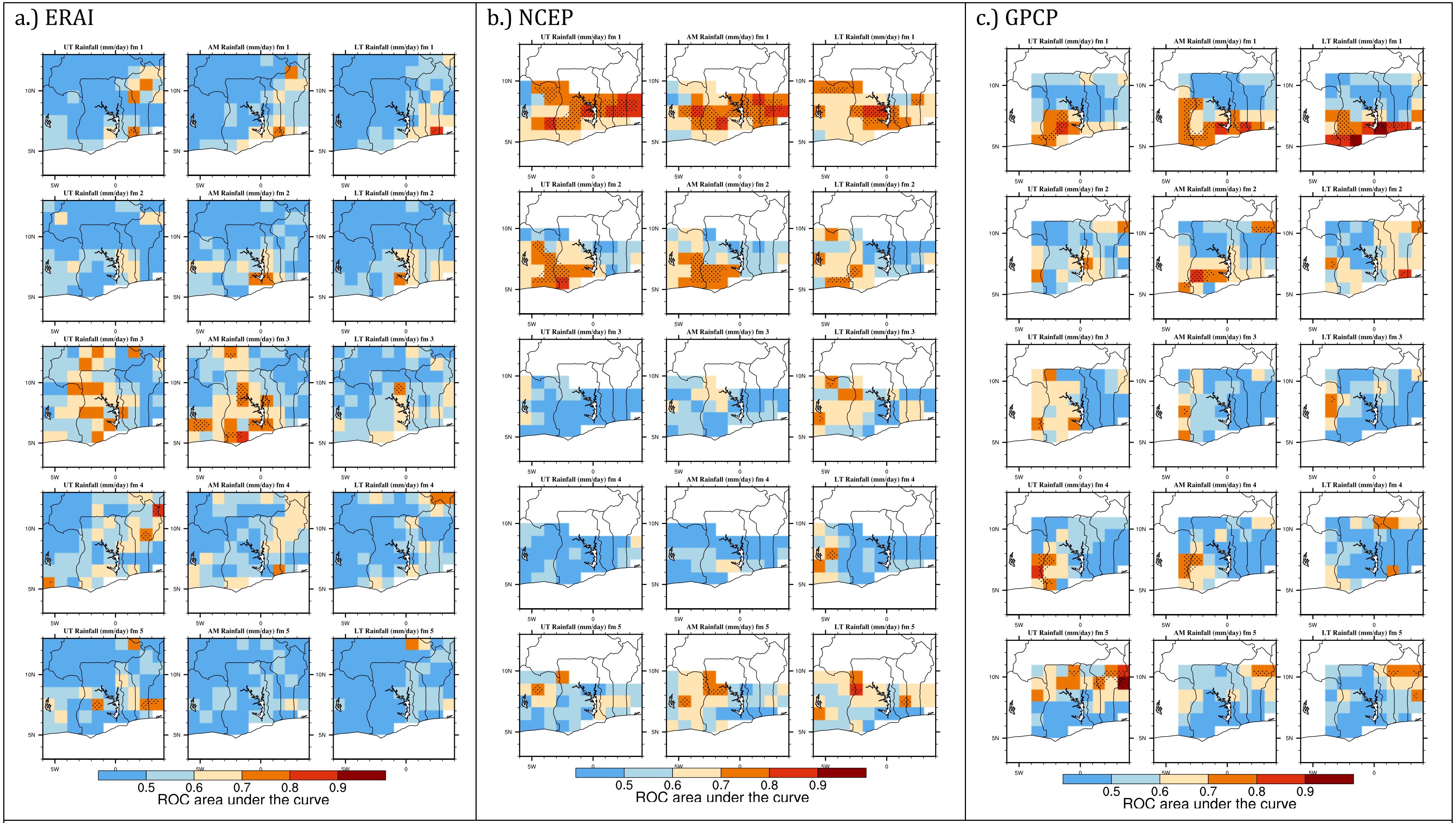

Figure 10: Relative Operating Characteristics (ROC) areas under the curve forecast skill for MJJ season peak rainfall in South Ghana for above upper tercile (extreme) events (UT, left column), above median events (AM, central column) and below lower tercile events (LT, right column) compared to a.) ERAI, b.) NCEP, c.) GPCP and d.) TRMM climatological data. Data were masked by very low perecipitation (<1mm/day). The five different rows represent increasing forecast lead times from 1 month (top row) to 5 months (bottom row) of the target season (see table 2). Stippling indicates significance at the 95% confidence interval, calculated by performing a two-tailed Mann-Whitney U-statistic test.

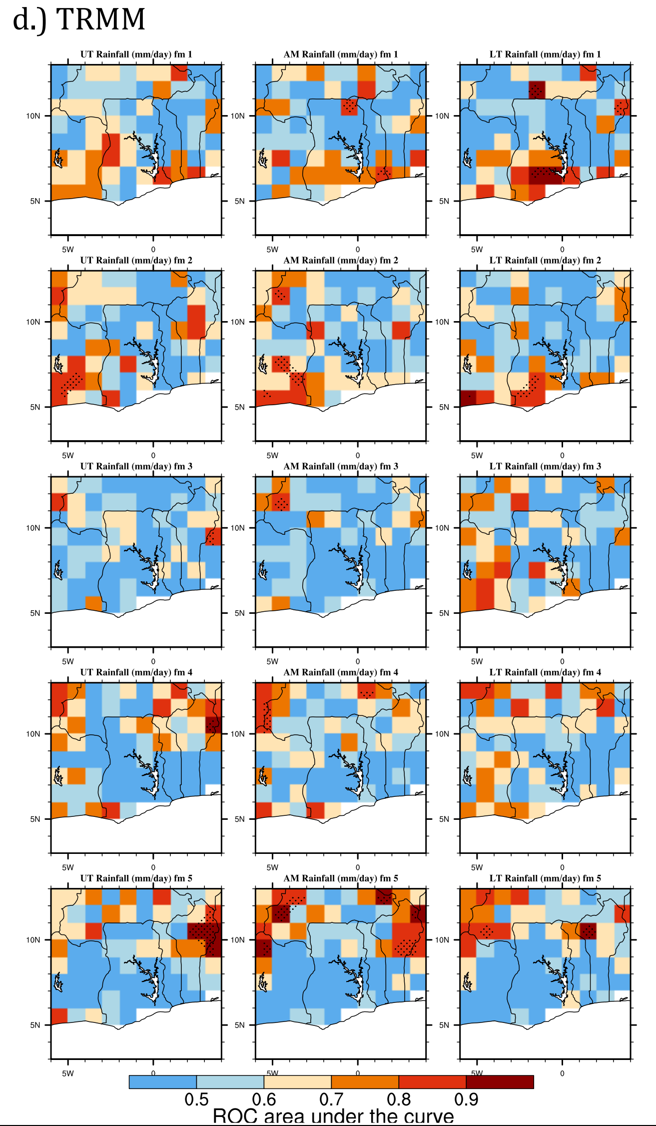

Figure 10, continued

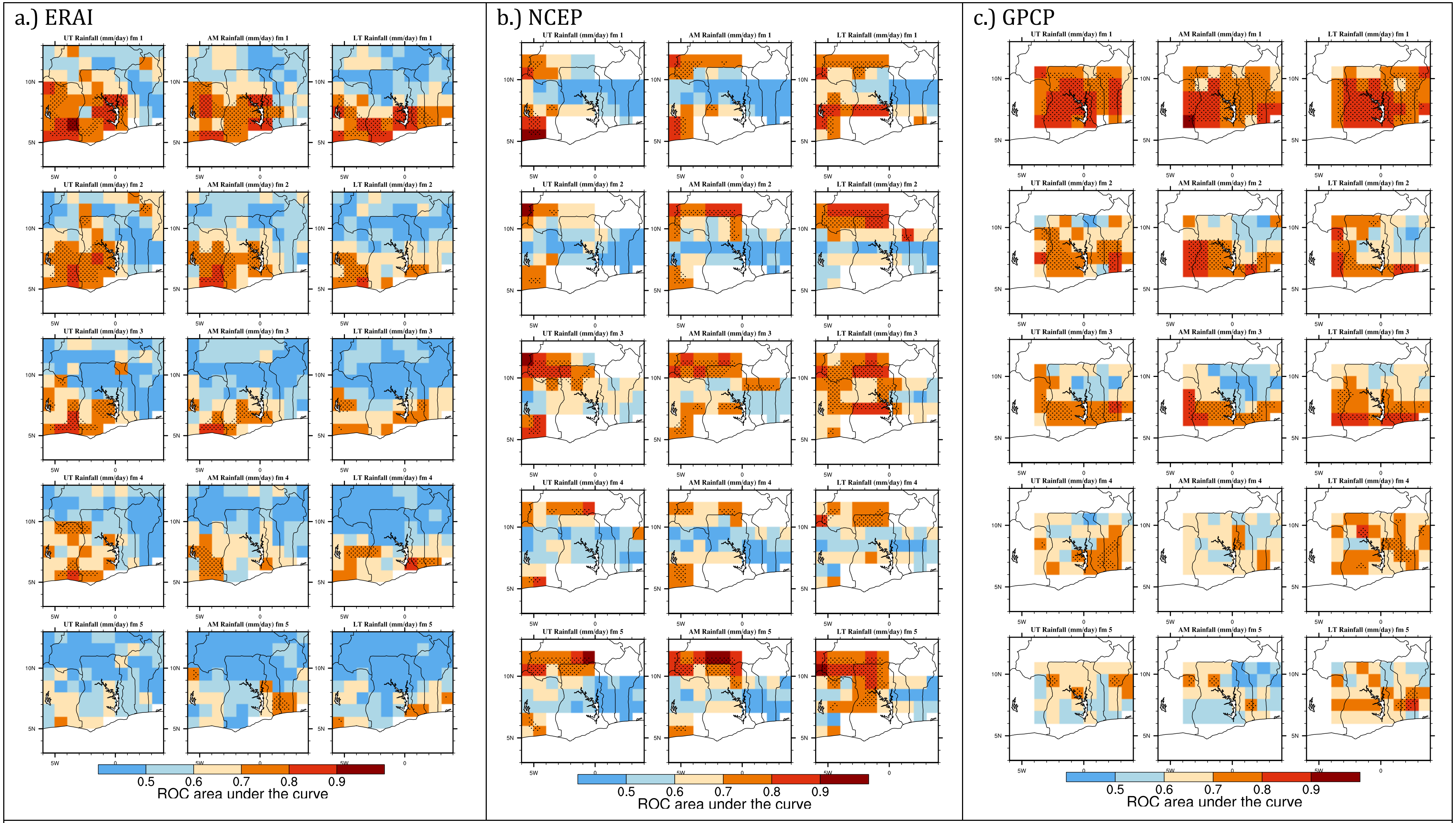

Figure 11: Relative Operating Characteristics (ROC) areas under the curve forecast skill for JAS season peak rainfall in North Ghana for above upper tercile (extreme) events (UT, left column), above median events (AM, central column) and below lower tercile events (LT, right column) compared to a.) ERAI, b.) NCEP, c.) GPCP and d.) TRMM climatological data. Data were masked by very low perecipitation (<1mm/day). The five different rows represent increasing forecast lead times from 1 month (top row) to 5 months (bottom row) of the target season (see table 2). Stippling indicates significance at the 95% confidence interval, calculated by performing a two-tailed Mann-Whitney U-statistic test.

d.) TRMM

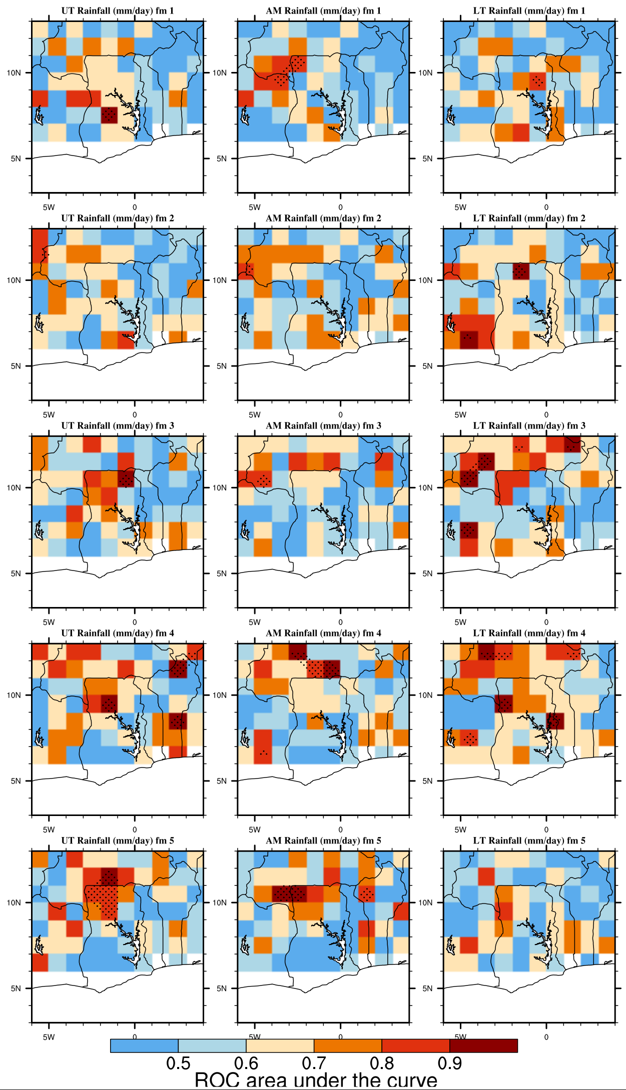

Figure 11, continued

Figure 12: Relative Operating Characteristics (ROC) areas under the curve for SON season malaria incidence forecast skill for above upper tercile (extreme) events (UT, left column), above median events (AM, central column) and below lower tercile events (LT, right column) compared to a.) ERAI, b.) NCEP, c.) GPCP and d.) TRMM climatological data. Data were masked by very low perecipitation (<1%). The five different rows represent increasing forecast lead times from 1 month (top row) to 5 months (bottom row) of the target season (see table 2). Stippling indicates significance at the 95% confidence interval, calculated by performing a two-tailed Mann-Whitney U-statistic test.

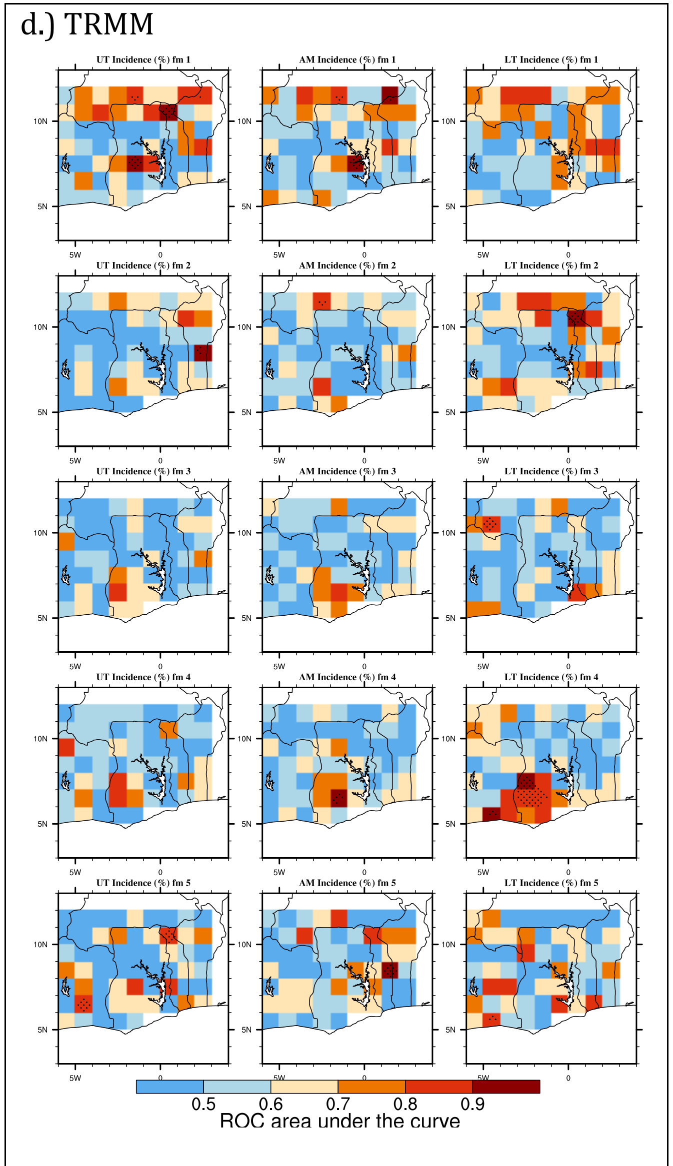

Figure 12, continued