Abstract: Since the 90s, several studies were conducted to evaluate the predictability of the Sahelian rainy season and propose seasonal rainfall forecasts to help stakeholders to take the adequate decisions to adapt with the predicted situation. Unfortunately, two decades later, the forecasting skills remains low and forecasts have a limited value for decision making while the population is still suffering from rainfall interannual variability: this shows the limit of commonly used predictors and forecast approaches for this region. Thus, this paper developed and tested new predictors and new approaches to predict the upcoming seasonal rainfall amount over the Sirba watershed. Predictors selected through a linear correlation analysis were further processed using combined linear methods to identify those having high predictive power. Seasonal rainfall was forecasted using a set of linear and non-linear models. An average lag time up to eight months was obtained for all models. It is found that the combined linear methods performed better than non-linear, possibly because non-linear models require larger and better datasets for calibration. The $R^2$, Nash and Hit rate score are
respectively 0.53, 0.52, and 68% for the combined linear approach; and 0.46, 0.45, 61% for non-linear principal component analysis.

**Keywords:** rainfall forecasting; neural network; non-linear principal component analysis; Sirba basin; West African monsoon; air temperature

---

1. **Introduction**

The summer rainfall of semi-arid regions of the world is known for its unreliability, which has a large impact on the continental hydrological cycle, water resources and food security. The Sahel, extending across Africa from the Atlantic Ocean to 30°E and from 12 to 17°N is the largest area of these regions, recording between 200 and 800 mm/year from north to south, ~80% of the rain being recorded in July–September the cool rainy season. Temperature in this region ranges from approximately 18 to 36 °C. The recurrent droughts and subsequent famines that struck the Sahel in the 1970s (1972–1974), and the 1980s (1983–1985) and make it unique at the global scale led the scientific community to investigate possible mechanisms responsible for these dramatic events and to develop forecasting models to help coping with such phenomena. This area experienced severe droughts almost every two to three years. The devastate drought in 2012 affected more than 18 million people in nine countries with food insecurity, high grain prices and environmental degradation. This 2012 crisis came after the severe drought in 2010 [1–18].

The Sahelian rainfall pattern is season dependent and is directly related to the West African Monsoon (WAM).

Although, the WAM’s dynamic is better understood nowadays, the main challenge with regards to its variability and predictability comes from its varying teleconnections. A teleconnection is the linkage of climate variables between two different areas, which may be close or far to each other [19,20]. Teleconnections between Sahelian rainfall and the oceanic basins have changed quickly much since the 60’s: the tropical Atlantic had the strongest influence during the years 60–70, then the equatorial Pacific (El Nino/Southern Oscillation) during the 80–90’s, and the Mediterranean now [21–25]. Thus, the absence of well-established predictors that can be used to predict seasonal rainfall as well as streamflow in the Sahel partly explains why forecasts at all scales in the Sahel are tricky. Many studies attempted to forecast Sahelian seasonal rainfall and streamflow for the purpose to overcome the droughts impacts [4,7,26–33]. Unfortunately, most of these works focused on only sea surface temperatures (SST) over years [24,34–36]. Nevertheless, few studies attempted to use the atmospheric dynamics for carrying out seasonal forecasts. This prediction relies on the explicit simulation of major atmospheric processes [37–44]. Garric *et al.* [42] used the ARPEGE atmospheric model forced by SST anomalies and multivariate linear regression using SST and rainfall predictors observed before the monsoon season. They showed that the ARPEGE model did not give better seasonal rainfall predictions than simple regression systems.

Thus, the objective of this paper is to develop a method that would identify new skillful predictors for seasonal rainfall in the Sahel, and to compare a set of linear models to non-linear ones for forecasting JAS (July to September) rainfall amounts. This would provide the community with actionable seasonal
information that would constitutes a major tool for farmers, decision makers and water resources managers in this region.

For such purpose, a pool of predictors is built by analyzing the physical influence on the WAM. Each predictor is tested as an input to a linear rainfall forecasting model as in [4]. At the end of the process, an optimal lag time and an optimal season are obtained to extract the predictor. Retained predictors are afterward also tested in new developed non-linear models. Finally, the forecast skills of the two groups of models are discussed.

2. Review of the Main Drivers of the Sahelian Rainfall Variability

In West Africa, the rainfall pattern is firmly related to the seasonal movement of the inter-tropical convergence zone (ITCZ) and consequently to the development of the WAM circulation [45].

The SST constitutes a key factor in the variability of the WAM and therefore the Sahelian precipitations as shown in several studies [12,34,37,46–48]. The tropical Atlantic Ocean is considered as the principal source of moisture for West Africa. Its impact on the WAM system was shown since 1970s when the Sahelian rainfall deficit was associated to colder SST in the north tropical Atlantic and warmer SST in the south and at the equator which promotes a southernmost ITCZ than the normal. These results were confirmed, and then extended to longer time scales by many studies [13,49–52]. However, Janicot et al. [53] noticed that the relationship between Sahelian rainfall and Atlantic SSTs considerably decreased to a point to be statistically not significant during the dry period (i.e., post 1970). This teleconnection changed from Atlantic SST to the SST over eastern and center Equatorial Pacific, in agreement with the work of [54].

A rainy season with below average rainfall in West Africa is usually associated to a warm period of El Niño-Southern Oscillation (ENSO). Janicot et al. [55] explained this link by a strengthening of the Walker circulation and a weakening of both the monsoon and the southern cell of the Hadley circulation. This situation leads to an increase in trade winds over the northern tropical Atlantic and a reduction in the water vapor in West Africa. Recently, a teleconnection with the Mediterranean Sea has been highlighted [56–58]. It seems to impact the WAM system in addition to the Atlantic and Pacific [58]. Rowell [56] found that the influence of temperature anomalies in the Mediterranean on the Sahel (for 1947–1996) is of a similar magnitude with that in the Pacific. From numerical simulations, he showed that a warm Mediterranean sea promotes excess rainfall in the Sahel. Additionally, Gaetani et al. [57] pointed out that a positive precipitation response to warmer than average conditions in the Mediterranean Sea is found in the Sudano-Sahelian belt in August to September.

Several recent papers analyzed the interactions of the different oceanic basins and the resulting impact on the Sahelian rainfall variability. Shaman and Tziperman [59] found that the interannual rainfall variability over the Mediterranean region is related to ENSO variability in the eastern Pacific via an eastward-propagating atmospheric stationary barotropic Rossby-wave train. Moreover, Lopez-Parages et al. [60] explained how the teleconnection with the ENSO appears modulated by multidecadal oscillations of the SST over the Atlantic and Pacific basins.

With regards to the role of the land surface on the Sahelian rainfall variability, Webster et al. [61] indicated that the use of the moist static energy (MSE) (its three components: sensitive, latent and potential) could improve the rainfall forecasts in the Sahel. They argued that the variation in temperature between the ocean and the continent is responsible for the monsoon circulation; and this circulation is even better
explained when the moisture gradient is considered. Eltahir [62], Philippon and Fontaine [63], Hall and Peyrillé [64] highlighted the role of these gradients on the dynamics of WAM and the Sahelian rainfall variability. Zheng and Eltahir [65] found that a change in vegetation (e.g., deforestation) on the Guinean coast has a direct substantial impact on atmospheric dynamics associated with the monsoon circulation through MSE gradients. Thus, some authors such as Wang and Eltahir [66] suggested the inclusion of vegetation dynamics in the modeling exercises as it constitutes an important process for simulating Sahelian rainfall variability. In addition, while testing the impact of vegetation in rainfall variability simulation, Zeng et al. [67] found that the decadal variability is best reproduced when interactive vegetation is added to the model. The role of soil moisture on West African rainfall event is also addressed in some studies [62,68]. These authors emphasized that a positive anomaly of soil moisture would strengthen the monsoon circulation through a modification of MSE gradients. Douville et al. [69], Douville [70], and Douville [71] found that any reduction in soil moisture in ARPEGE is associated with low intensities of precipitations. Thus, based on these results they concluded that soil moisture contributes to the interannual variability of rainfall in the Sahel.

It therefore appears that the climate in West Africa (predominantly in the Sahel) is determined by interactions between global processes (e.g., sea surface temperatures) and regional processes (e.g., physiographic characteristics). The use of parameters related to these processes in seasonal rainfall forecasting models would generate more skillful forecasts for the Sahel.

3. Materials and Methods

3.1. Study Area

The study area considered in this work is the Sirba watershed. This watershed, shared by Burkina Faso and Niger, is situated between latitudes 12°55’S–14°23’30”N and longitudes 1°27’W–1°23’42”E with an area of 38,750 km² (Figure 1) [72]. This choice is motivated by the fact that the Sirba basin is central in the Sahel region, and there are many climate stations inside and around the basin that have been collecting climate data daily for more than 40 years. Another reason is that, locally, the Sirba tributary plays an important role in the hydrological regime of the Niger river at Niamey, as it participates in its Sudanian flood in September. The Sirba extends over three sub-climate zones based on the amount of rainfall decreasing from south to north: a southern Sudanian zone with mean annual rainfall between 700 and 800 mm, a northern Sudanian zone with mean annual rainfall ranging from 550 to 650 mm and a Sahelian zone with mean annual rainfall of 300 to 500 mm [73]. The largest quantities of rainfall are observed during the months of July to September (JAS), regardless of the climate zone. The climate is generally characterized by the presence of two seasons: a dry season (October to April) due to the Harmattan (dry wind) and a rainy season (May to September) influenced by the WAM (wet wind). The hydrographic network is relatively dense and consists of three main tributaries (Sirba, Faga and Yeli) as well as some water reservoirs from dams [71]. Based on the description of the rainfall pattern, the hydrological regime in the Sirba watershed is of Sahelian type, as it is characterized by non-sustainable flows with an exoreic operation pattern. At the upper bed of the Sirba, there is a series of depressions with intermittent flow. However some sections of the Sirba reaches have water constantly during the wettest years. Its vegetation formation is thorny, lightly-wooded savannah.
3.2. Climate and Atmospheric Data

Climate data used in this study include rainfall and atmospheric data. *In situ* daily rainfall data span the period 1960–2008, and are obtained from national meteorological offices of Burkina Faso and Niger. Five of these stations are located within the watershed while the remaining six stations are located at most 25 km from the watershed boundary (Figure 1). The Thiessen polygon method is a standard method widely used to calculate average areal precipitation [74]. It was implemented to estimate average rainfall over the watershed from the 11 rainfall time series (Table 1). Though there is a lack of good climate data throughout African Sahel, only a few gaps were found in these records, and they do not alter the quality. The *in situ* rainfall time series have less than 10% missing data as this ratio varies from 0 to 7% for all 11 stations over the period. Moreover, monthly precipitation time series of Climatic Research Unit (CRU TS 3.21 0.5° global) with a spatial resolution of 0.5° × 0.5° and defined on latitudes 10°N–15°N, longitudes 2°W–2°E are used. They are sourced from the British Atmospheric Data Centre (BADC, http://www.cru.uea.ac.uk/cru/data/hrg/) and span the period 1901–2012 [75].

The atmospheric data considered are Sea Level Pressure (SLP), Sea Surface Temperature (SST), Relative Humidity (RHUM), Air Temperature (AirTemp), Meridional Wind (VWND), and Zonal Wind (UWND). These variables are monthly NCEP-DOE Reanalysis data sourced from the National Oceanic and Atmospheric Administration (NOAA; http://www.esrl.noaa.gov) except the SST data series (NOAA NCDC ERSST version3b sst) obtained from the IRI data library (International Research Institute for Climate and Society: http://iridl.ldeo.columbia.edu) [76]. These data span from January 1979 to August
2013, except for SST data which cover the period January 1960 to December 2013. The relationship of these predictors with the WAM is briefly described in the following paragraphs.

### Table 1. Specifics of rainfall stations.

<table>
<thead>
<tr>
<th>Station number (code)</th>
<th>Station name</th>
<th>Longitude (degrees: °)</th>
<th>Latitude (degrees: °)</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>320006</td>
<td>Torodi</td>
<td>1.80</td>
<td>13.12</td>
<td>Niger</td>
</tr>
<tr>
<td>320002</td>
<td>Tera</td>
<td>0.82</td>
<td>14.03</td>
<td>Niger</td>
</tr>
<tr>
<td>320004</td>
<td>Tillaberi</td>
<td>1.45</td>
<td>14.20</td>
<td>Niger</td>
</tr>
<tr>
<td>320005</td>
<td>Gotheye</td>
<td>1.58</td>
<td>13.82</td>
<td>Niger</td>
</tr>
<tr>
<td>200082</td>
<td>Boulsa</td>
<td>-0.57</td>
<td>12.65</td>
<td>Burkina Faso</td>
</tr>
<tr>
<td>200026</td>
<td>Dori</td>
<td>0.03</td>
<td>14.03</td>
<td>Burkina Faso</td>
</tr>
<tr>
<td>200085</td>
<td>Bogande</td>
<td>0.13</td>
<td>12.98</td>
<td>Burkina Faso</td>
</tr>
<tr>
<td>200048</td>
<td>Dakiri</td>
<td>-0.27</td>
<td>13.30</td>
<td>Burkina Faso</td>
</tr>
<tr>
<td>200024</td>
<td>Gorgadji</td>
<td>-0.52</td>
<td>14.03</td>
<td>Burkina Faso</td>
</tr>
<tr>
<td>200086</td>
<td>Piela</td>
<td>-0.13</td>
<td>12.70</td>
<td>Burkina Faso</td>
</tr>
<tr>
<td>200047</td>
<td>Tougouri</td>
<td>-0.52</td>
<td>13.65</td>
<td>Burkina Faso</td>
</tr>
</tbody>
</table>

(a) Zonal Wind and Meridional Wind

Gallée et al. [77] have highlighted the importance of the meridian gradient of MSE on the movement of the ITCZ northward. According to their studies, MSE led the WAM and creates an environment favorable for deep convection over the Sahel. Thus, the spread of WAM to north is associated with strong gradient of MSE in the lower layers, the convergence result in the triggering of convection over the Sahel. These authors emphasized the relationship between the WAM, precipitation, the MSE, the meridian wind and sensible and latent heat in the Sahel. The strong north-south gradient (weak) of the meridian wind seems to be the result of a strong south-north gradient (weak) of MSE. The first maximum MSE gradient develops, followed by the maximum meridian wind. Then the MSE meridian gradient leads meridian wind speed and precipitation to a secondary maximum. In general, changes in precipitation appear to be a consequence of MSE with a major role of the meridian wind.

The zonal and Meridian winds play a very important role in the flow in the middle and upper troposphere. Indeed, in the middle and upper troposphere, the zonal wind profile is dominated by three jets: AEJ (African East Jet), TEJ (Tropical East Jet) and WSJ (West South Jet). AEJ is at the level of 600 hPa and TEJ at 200 hPa. These two jets, which are located above the Sahel and Guinea, are important for the atmospheric dynamics in West Africa. The variation of these winds regulates the position of the AEJ, which, in turn, explains why during wet years (dry) for the Sahel [68,78]. Grist and Nicholson [79] showed evidence of AEJ above the Sahel (10°N to 15°N). JET looks very bound to the African monsoon. Indeed, through the Walker circulation, the intensity of the jet effect of the monsoon which forms the lower part of the cell [80]. Sultan [81] shows that the jet's installation date is a good indicator of the development of the monsoon. Finally, Nicholson and Grist [82] suggested that the jet is a response to precipitation but is not a cause of the variability of rainfall.
(b) Air Temperature

The impact of Air temperature on the WAM occurs through a link with atmospheric dynamics over West Africa. According to Fontaine et al. [83], the air temperature at two meters presents its maximum before the wet period (during the month of May) because of the maximum exposure at the top of the atmosphere and is followed by a maximum equivalent potential temperature (\(Oe\)) in August because of the maximum of the zenith angle of the sun. \(Oe\) in the lower troposphere is equivalent to MSE whose transport is due to the circulation of large-scale (the Hadley). But Flaounas et al. [84] and Guiavarch et al. [85] like other studies have clearly shown the relationship between this surface energy and WAM. Thus, it appears that the relationship between air temperature and the WAM is not a direct link but rather a role on the dynamics of WAM in terms of anomalies reflections in atmospheric circulation like Hadley types.

(c) SST

Several studies have analyzed the relationship between WAM and SSTs. Coëtlogon et al. [86] studied the link between WAM and the contrast between SST and the temperature at the coast of Guinea. According to their results, from spring to summer, a band of cold water settles between Ecuador and the coast of Guinea and enhances the temperature gradient at the surface meridian. This causes the acceleration of the WAM which then moves further north. The appearance of this band of cold water is attributed to the process of recovery in deep and cold ocean masses “upwelling”, mostly due to surface winds. Thus, the acceleration of WAM resulted from a positive feedback system since the acceleration intensifies upwelling increasing itself from WAM that spreads further north. Moreover, Peyrillé and Lafore [87] developed a two-dimensional idealized model to reproduce the monsoon system in West Africa. Their study reveals the importance of SST in the Mediterranean. The lack of moisture transport or transport by zonal eddies above the African continent requires forcing external moisture advection in the Mediterranean to get realistic monsoon in West Africa. Thus, they show that warm SST in the Mediterranean entails strengthening moisture advection in the lower layers. SST of West Mediterranean seems to have a stronger impact on the variability of precipitation in the Gulf of Guinea (Sahel) [88]. Several studies suggested some links between the variability of SST during the season of coastal rain and precipitation, especially through the installation of the equatorial upwelling. Gu and Adler [89] describe, using satellite observations and reanalysis, the seasonal evolution of the tropical Atlantic. They show that, in the Gulf of Guinea, convection is modulated by seasonal forcing of the ocean and the SST gradient meridian. In addition, it was shown that the Pacific Ocean [34], the Atlantic Ocean [90], the Mediterranean [88] as well as the phenomenon El Niño-Southern Oscillation [10] generate atmospheric disturbances and, in this way, affect the African monsoon.

(d) SLP

A climatological analysis by Baldi et al. [91] suggests that the West Africa monsoon influence the central-western [92,93]. Using NCEP/NCAR global reanalysis [91] analyzed in detail the events characterizing summer 2002 over Mediterranean, Europe and North Atlantic, in particular the anomalous SST and Sea Level Pressure (SLP) fields relatively to the mean climate patterns Mediterranean summer, and specifically the SLP (weakly), the temperature and the rainfall. They also found that the overall
pattern of the WAM changed in July, when a lower pressure developed from Iceland to central Mediterranean along a northwest to south-east axis, with anomalously high pressures in the south-west and north-east. Moreover, the summer average SLP field was similar to the pattern observed in July. The surface air temperature field over Mediterranean closely follows the sea level pressure patterns in summer (e.g., Maheras and Kutiel, [94]). They also run sensitivity analysis which shows how the SST anomalies can produce quantitatively significant anomalies in the sea level pressure patterns over North Sahel (positive).

(e) RHUM

The interaction between the flows of heat from the surface of ground water and in the atmosphere has been studied by Lafore [95] and Fontaine et al. [88] over the period 1979–2001. Four phases of the ITCZ (inter tropical convergence zone) were identified: early March, mid-April, May and late June (establishment of WAM). These phases appear to be sensitive to the relative humidity of last year. The interaction between these phases is as follows: positive anomalies in this humidity entail an increase in the humidity of the atmosphere, and the convergence of humidity flux- and a decrease in surface albedo. Therefore, the net solar radiation is strengthened on the surface but the air temperature in the lower layers decreases. Net radiation on the surface increases as well as the flow of heat to the atmosphere. This process results in the strengthening of MSE in the lower layers and the strengthening of the circulation of WAM. On the other hand, Fontaine et al. [83] showed that in the three regions in West Africa (Guinea, 6°N–10°N, Sudan 10°N–15°N, the Sahel 15°N–20°N), the convergence integrated moisture flux in the entire atmospheric column is significantly correlated with the precipitation at different scales. It is interesting that in the Sahel, unlike the rest of West Africa, the relationship between precipitation and soil evaporation (consequently the RUM) in the Sahel is almost linear. Moreover, Broman et al. [96] performed A K-means cluster analysis to identify spatially coherent regions of relative humidity variability during the two periods over West African Sahel. They found that correlating the cluster indices with large-scale circulation and SSTs indicates that the land-ocean temperature gradient and the corresponding circulation, tropical Atlantic sea surface temperatures (SSTs), and to a somewhat lesser extent tropical Pacific SSTs all play a role in modulating the timing of the monsoon season relative humidity onset and retreat.

Thus, it is clear that the RUM is a link because it is connected to the mainland that can exert forcing on the atmospheric dynamics through which it is a reflection of abnormalities in the circulation of WAM.

From the above explanation, it is clear that despite that the atmosphere has no inertia the predictors (SLP, AirTemp, and RUM) are connected either to the ocean or continent that can have a forcing on atmospheric dynamics and they are the reflections of anomalies in the atmospheric circulation types Hadley or Walker. Other predictors are related to the Pacific or the Mediterranean to the tropical Atlantic. Table 2 summarizes the atmospheric data and their geographical locations.

3.3. Selection of Predictors and Optimal Lag Time

Monthly CRU precipitation time series cover an area larger than the extent of the Sirba basin. They were initially used as predictand for selecting a pool of potential predictors having a known relationship with the WAM, and to avoid those with no effect on the monsoon dynamics.
Table 2. Description of atmospheric data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Units</th>
<th>Level</th>
<th>Reference Data</th>
<th>Spatial coverage</th>
<th>Regions of the Predictors</th>
<th>Temporal Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sea level pressures (SLP)</td>
<td>Pa/s</td>
<td>1000 hPa</td>
<td>NCEP 2</td>
<td>2.5° × 2.5° grid</td>
<td>Atlantic ocean</td>
<td>1979/01/01 to 2013/08/31</td>
</tr>
<tr>
<td>Air temperature (AirTemp)</td>
<td>°K</td>
<td>1000 hPa</td>
<td>NCEP 2</td>
<td>2.5° × 2.5° grid</td>
<td>Pacific ocean</td>
<td>1979/01/01 to 2013/08/31</td>
</tr>
<tr>
<td>Meridional wind (VWND)</td>
<td>m/s</td>
<td>1000 hPa</td>
<td>NCEP 2</td>
<td>2.5° × 2.5° grid</td>
<td>Sahel (Easterly jet)</td>
<td>1979/01/01 to 2013/08/31</td>
</tr>
<tr>
<td>Zonal wind (UWND)</td>
<td>m/s</td>
<td>1000 hPa</td>
<td>NCEP 2</td>
<td>2.5° × 2.5° grid</td>
<td>Sahel (Easterly jet)</td>
<td>1979/01/01 to 2013/08/31</td>
</tr>
<tr>
<td>Relative humidity (RHUM)</td>
<td>%</td>
<td>1000 hPa</td>
<td>NCEP 2</td>
<td>2.5° × 2.5° grid</td>
<td>Mediterranean basin</td>
<td>1979/01/01 to 2013/08/31</td>
</tr>
<tr>
<td>Sea surface temperature (SST)</td>
<td>°C</td>
<td>Surface</td>
<td>NOAA NCDC</td>
<td>2° × 2° grid</td>
<td>Atlantic ocean</td>
<td>1854/01/01 to 2013/08/31</td>
</tr>
<tr>
<td>Climatic research unit rainfall (CRU)</td>
<td>mm</td>
<td>Surface</td>
<td>CRU</td>
<td>0.5° × 0.5° grid</td>
<td>January 1901 to December 2012</td>
<td></td>
</tr>
</tbody>
</table>

These monthly precipitations time series were averaged over the season July–September (JAS), which is the core of the rainy season in the Sahel. As a preliminary test, different time periods 1960–2010, 1970–2010, 1980–2010, 1990–2010, 2000–2010 are considered to check the most favorable periods in terms of signals within the pool of potential predictors. The reason for testing these sub-periods resulted from the previously mentioned studies (Section 2) which showed that the teleconnections of the Sahelian rainfall have evolved since the 60’s. Based on these defined periods, 13 groups (with 89 sub-components) of predictors were considered for the preliminary test to check the presence of significant correlations (R > 0.5) between each predictor and CRU rainfall.

After this first selection which relies on the CRU precipitations dataset as predictand, predictors were selected using the in situ rainfall (from rain gauge stations) as predictand. This selection is done using the method developed in [4]. This method was employed to link the observed rainfall and each predictor through some statistical techniques. The candidate predictor was aggregated over all possible time windows (where a time window’s length in months is an integer) during the 18 months prior to the rainy season onset and each of the obtained time series was used as an explanatory variable in a linear model having the seasonal rainfall on the Sirba watershed as explained variable.

The choice of the period over which the predictor is averaged will impact the performance of the forecast. Since the best period is not known a priori, predictor data sets were aggregated over various periods with different lengths and different start dates. The periods were restricted to start at the
beginning of a calendar month and finish at the end of a calendar month. The beginning of a period has to be later or equal to January 1st of the previous year (year Y-1, where Y is the year containing the rainy season for which the forecast is issued).

The end of the period must be prior or equal to June 30th of year Y. Figure 2 shows how time windows were systematically generated. The upper bar indicates all months starting from January of the previous year (year Y-1) to June of the year the forecast is issued (year Y). In the first run, for example, only the predictor of January (Y-1) was selected to use as a predictor. Predictor averaged over January-February (Y-1) was used as a predictor in the second run. This process was iterated at one-month increments until June (Y) was reached as the end of the period. The process was repeated until the beginning and the end of the periods were June (Y).

For each time window, a linear model linking the predictor averaged over that time window and seasonal rainfall on the Sirba was built as follows:

(a) For each year Y that the predictor was available,

(i) The predictor of year Y-1 was removed from the predictor grid;
(ii) The rainfall of year Y was removed from the rainfall data set;
(iii) A coefficient of correlation (R) is used to screen the remaining predictor data: a correlation analysis between the predictor at each grid point and the rainfall was computed and its level of significance (P-value <0.05) was assessed. Once the correlation was not significant, the grid point was discarded. The remaining grid points were then ordered decreasingly;
(iv) Afterward, a principal component analysis (PCA) was applied on the retained predictor gridded data from the previous step to reduce the number of predictors;
(v) Since PCA gave rise to more sets of new predictor data, a stepwise regression (5% confidence interval) was used to keep only grid points with high predictive power;
(vi) A linear regression was fitted between the predictors and precipitation time series;
(vii) The fitted linear regression was used to simulate the rainfall of year Y. If predictor and rainfall were in the same year (Year Y), only predictor and rainfall time series for that year were removed in the first step.

(b) Then, the coefficient of determination (R²), Nash-Sutcliffe coefficient (Nash), and Hit-Rate scores (HIT) were computed to estimate the model's performance.

Figure 2. Predictor averaging periods (Adapted from Sittichok et al. [4]).
In summary, all predictors were selected according to their physical link with the WAM based on the hypothesis that only predictors having dynamical link with the WAM seem to give good forecast skills. Each of these predictors was screened through simple correlation test to find its link with the CRU rainfall data defined over a region covering more than the Sirba watershed. The main criteria used to find either the predictor should be retained or rejected is the correlation coefficient (R) that has to be greater or equal to 0.5 (i.e., p-value < 0.05). The CRU rainfall was used at this stage for the purpose to have a good assessment of all possible predictors having impact on the WAM. The retained predictors were further screened based on in situ rainfall using the approach explained previously and summarized in Figure 2. All these steps were summarized in the new flowchart of Figure 3. It should be noted that the figures provided were based on leave-one out cross validation (LOOV) which was also used is selecting the predictor. The LOOV was used due to the small length of data used.

The first three (3) best predictors in terms of high Nash values were used to test the performance of the two sets of methods (linear approach and non-linear approach).

3.4. Linear Approach

Several linear methods are applied successively for selecting the predictors and developing the seasonal rainfall forecast models. They include: correlation analysis, principal component analysis (PCA), stepwise regression, linear regression, and cross validation.

They were used to perform three successive tasks: predictor selection; predictor dimension reduction and linear regression. The first application discarded meaningless predictors from the original data set using the coefficient of correlation as criteria of selection. At this stage, the correlation coefficient between the predictor at each grid point and the rainfall on the Sirba watershed was calculated and its level of significance (p-value < 0.05) was tested. When the correlation was not significant, the grid point was discarded. The remaining grid points were then decreasingly ordered according to the correlation p-value. Only the best grid points were included in the analysis. Afterward, PCA was applied on the retained predictors to reduce their number. A forward stepwise regression method (5% confidence interval threshold) was then applied to keep predictors having only a significant predictive power.

It should be noted that a leave-one-out cross validation was used in the model application to avoid the bias which might occur during the development of empirical equations using statistical models.

3.5. Non-Linear Approach

Two non-linear methods were tested for each of the three best predictors selected based on the correlation analysis. The description of these methods and how they are applied is detailed in the next paragraphs. The $R^2$, Nash, and HIT were calculated to estimate the model’s performance.

3.5.1. Non-Linear Principal Component Analysis

The non-linear principal component analysis (NLPCA) algorithm developed by [97] is generally considered as a non-linear generalization of standard linear PCA and was successfully applied in atmospheric and oceanic sciences [98–102]. The principal components (PCs) are generalized from straight lines to curves, thus the NLPCA helps to extract PCs either linear or not. This could improve
seasonal rainfall forecast skills because it is well known that most of atmospheric/climate relationships are not linear as always assumed. Each predictor was first screened using $R^2$ before being fed into the NLPCA. However, due to the high computational time of NLPCA the number of PCs is narrowed in considering only the three best PCs in the process. Figure 4 presents the entire process of the NLPCA seasonal forecast model. More details on the way NLPCA model works and its difference with the ordinary PCA can be found in Scholz et al. [97].

**Figure 3.** Steps for the selection of predictors.
3.5.2. Feedforward Neural Network

The feedforward neural network (FFNN) is just tested in this work because to show its performance which is somehow poor. For more details on this method, the reader can refer to [103–109].

4. Results and Discussions

4.1. Selected Predictors and Lag Time Period

In the preliminary setting of a pool of predictors, the synchronous correlations between predictors-predictand (i.e., correlations between rainfall and predictors averaged over JAS period) have revealed the presence of the largest correlations for the period 1970–2010. Thus, it was used as the reference study period. For instance, Figure 5 presents the correlation between rainfall and SLP over this time period with a lag time of three months between the two variables.

During the second step dedicated to selecting the predictors, the time window (or season) which yielded the best Nash coefficient (and therefore the optimal lag time) was determined. Tables 3–5 summarize the final selected predictors used to forecast seasonal rainfall using combined linear methods, NLPCA and FFNN, respectively. It is obvious that for all models the best predictors according to the forecast skills are AirTemp, RHUM and SLP. In addition, the lag time (eight months lead time on average) obtained for most predictors is large enough to develop early warning systems for the decision makers and socio-economic actors about the issue of the forthcoming rainy season.
4.2. Seasonal Rainfall Forecast

The performance of the combined linear models showed that AirTemp (from Pacific Tropical North), RHUM (from Mediterranean East) and SLP (from Atlantic tropical South) are the best predictors with respective Nash coefficients of 0.53, 0.52 and 0.46 (see Table 3). They have also the best coefficients of determination (53%, 58%, 48%) and Hit rate scores (67.9%, 64.3%, 60.7%), respectively. While SST (from Atlantic Ocean) obtained 0.34, 43%, and 58.5% as Nash, $R^2$ and Hit rate scores, respectively. These results from the predictors AirTemp, RHUM, and SLP seem to be better than those obtained in [4] who used linear methods to forecast seasonal rainfall in the same area based on Pacific and Atlantic SSTs as predictor. They obtained 0.45, 0.38 and 66.67% respectively for $R^2$, Nash, and Hit rate score. Figures 6 and 7 present the observed and simulated seasonal rainfall for combined linear models.

![Correlation map between seasonal precipitation and sea level pressure (Da Silva analysis, sea).](image)

**Figure 5.** Correlation map between seasonal precipitation and sea level pressure (Da Silva analysis, sea).
Table 3. Combined linear methods for seasonal rainfall forecast.

<table>
<thead>
<tr>
<th>PREDICTOR</th>
<th>NMAX*</th>
<th>R²</th>
<th>Nash coef.</th>
<th>HIT Score</th>
<th>Best period M1-M2**</th>
<th>Lag period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sea Level Pressure (SLP) at 1000hPa</td>
<td>50</td>
<td>0.48</td>
<td>0.46</td>
<td>60.71</td>
<td>17-18</td>
<td>0</td>
</tr>
<tr>
<td>Relative Humidity (RHUM) at 1000hPa</td>
<td>80</td>
<td>0.58</td>
<td>0.52</td>
<td>64.29</td>
<td>10-10</td>
<td>8 months</td>
</tr>
<tr>
<td>Air Temperature (AirTemp) at 1000hPa</td>
<td>10</td>
<td>0.530</td>
<td>0.527</td>
<td>67.86</td>
<td>1-4</td>
<td>7 months</td>
</tr>
<tr>
<td>Meridional Wind (VWND) at 1000hPa</td>
<td>170</td>
<td>0.31</td>
<td>0.28</td>
<td>53.57</td>
<td>5-5</td>
<td>8 months</td>
</tr>
<tr>
<td>Zonal Wind (UWND) at 1000hPa</td>
<td>190</td>
<td>0.33</td>
<td>0.324</td>
<td>71.43</td>
<td>11-11</td>
<td>7 months</td>
</tr>
<tr>
<td>Sea surface temperature (SST)</td>
<td>30</td>
<td>0.43</td>
<td>0.34</td>
<td>58.54</td>
<td>3-6</td>
<td>12 months</td>
</tr>
</tbody>
</table>

(**) M1=1:12 (January to December); M2=M1:18 (considered month of M1 to the next coming June)

(*) NMAX: number of best grid points retained after screening the predictor grid based on R²

Table 4. Seasonal rainfall forecast model skills using non-linear principal component analysis (NLPCA).

<table>
<thead>
<tr>
<th>PREDICTOR</th>
<th>R²</th>
<th>NASH</th>
<th>HIT score</th>
<th>Lag time Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sea Level Pressure (SLP) at 1000hPa</td>
<td>0.32</td>
<td>0.31</td>
<td>53.57</td>
<td>9 months</td>
</tr>
<tr>
<td>Relative Humidity (RHUM) at 1000hPa</td>
<td>0.36</td>
<td>0.36</td>
<td>53.57</td>
<td>7 months</td>
</tr>
<tr>
<td>Air Temperature (AirTemp) at 1000hPa</td>
<td>0.46</td>
<td>0.45</td>
<td>60.71</td>
<td>8 months</td>
</tr>
</tbody>
</table>

Table 5. Feedforward neural network (FFNN) model (single predictor) output for Sirba seasonal rainfall forecast.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>R²</th>
<th>Nash</th>
<th>HIT score (%)</th>
<th>Lag time (months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AirTemp</td>
<td>0.26</td>
<td>0.20</td>
<td>48.24</td>
<td>4</td>
</tr>
<tr>
<td>RHUM</td>
<td>0.18</td>
<td>0.10</td>
<td>29.12</td>
<td>4</td>
</tr>
<tr>
<td>SLP</td>
<td>0.21</td>
<td>0.09</td>
<td>18.03</td>
<td>2</td>
</tr>
<tr>
<td>SST</td>
<td>0.18</td>
<td>0.044</td>
<td>11.49</td>
<td>5</td>
</tr>
</tbody>
</table>

For the NLPCA model, the issued seasonal forecast skills can be judged satisfactory regarding the short study period considered, because non-linear models need longer study periods to over perform the linear ones. Results showed that the predictor AirTemp (R²: 0.46; Nash: 0.45; HIT: 60.7%) was the best, and then followed by RHUM and SLP, respectively. It was also found that this method provides a to larger lag time compared to combined non-linear methods despite of its relative low forecast skills. Table 4 presents the model performance and the lag time, while Figure 8 illustrates some of the rainfall forecast obtained from the NLCPA model using, respectively, the predictors AirTemp and RHUM.

Overall, the set of linear models performs better than the non-linear ones. This suggests that there is less benefit using non-linear methods when dealing with small samples, as found in previous studies [27,96,102].
Figure 6. Combined linear model for seasonal rainfall forecast using SLP (upper panel) and RHUM (lower panel).
Figure 7. Combined linear model for seasonal rainfall forecast using AirTemp (upper panel) and SST (lower panel).
Figure 8. NLPCA seasonal rainfall forecast model using AirTemp (upper panel) and RHUM (lower panel).
5. Conclusions

Two non-linear methods and a combined linear approach were used to forecast JAS (July to September) rainfall on the Sirba watershed, West Africa. Predictors were first screened using a series of steps to isolate those having the highest predictive power. At the end of the process three predictors, air temperature (from Pacific Tropical North), sea level pressure (from Atlantic Tropical South) and relative humidity (from Mediterranean East) were retained and tested as inputs for seasonal rainfall forecasting models. Forecast performances were compared using $R^2$, Nash and HIT. Results showed that the combined linear approach performed better than the non-linear models. The best forecasts were obtained using air temperature as predictor ($R^2 = 53\%$; Nash = 0.53; HIT = 67.9%; Lead-time = 7 months). The next best model uses relative humidity as predictor ($R^2 = 58\%$; Nash = 0.52; HIT = 64.3%, Lead-time = eight months). Nonlinear Principal Component Analysis (NLPCA) was the best non-linear method while FFNN performed poorly.

These new predictors found in this study could lead to better forecasts of seasonal rainfall over West Africa, an issue which has challenged forecasters over many years. This paper also helped understanding that non-linear methods could also be used instead of the usual linear methods. The specificity of this work is the use of other predictors rather than the SST which gave acceptable results than the SST. However, the limit of the approach resides on the small length of data used. Therefore, to generalize the results for other scientists, it is recommended that during the forecast a Sahelian global index must be constructed using CRU data while examining its correlation with the index of the watershed and the skill of models. As a future step of this work, a multi-model approach will be used to compare the resulting skills to that of the best model.

Acknowledgements

The authors would like to thank the International Research Initiative on Adaptation to Climate Change (IRIACC) program through the International Development Research Center (Canada) for funding this research.

Author Contributions

Abdouramane Gado Djibo developed the models, performed analyses and wrote the paper. Nathalie Philippon, Ousmane Seidou, Harouna Karambiri, Hadiza Moussa Saley, Ketvara Sittichok and Jean Emmanuel Paturel contributed to analysis and interpretation of results. Nathalie Philippon, Ousmane Seidou, and Ketvara Sittichok proofread the manuscript and contributed to answer reviewers’ comments.

Conflicts of Interest

The authors declare no conflict of interest.
References


© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).