Atmospheric and Climate Sciences, 2023, 13, 539-564
https://www.scirp.org/journal/acs
ISSN Online: 2160-0422
ISSN Print: 2160-0414

Evaluation of Candidate Predictors for Seasonal Precipitation Forecasting

Pedro M. González-Jardines, Maibys Sierra-Lorenzo, Adrián L. Ferrer-Hernández, Arnoldo Bezanilla-Morlot

Center for Atmospheric Physics, Institute of Meteorology, Havana, Cuba
Email: pedro.met90@gmail.com, maibyssl@gmail.com

How to cite this paper: González-Jardines, P.M., Sierra-Lorenzo, M., Ferrer-Hernández, A.L. and Bezanilla-Morlot, A. (2023) Evaluation of Candidate Predictors for Seasonal Precipitation Forecasting. *Atmospheric and Climate Sciences*, **13**, 539-564.
https://doi.org/10.4236/acs.2023.134031

Received: September 2, 2023
Accepted: October 22, 2023
Published: October 25, 2023

Abstract

This research proposes to carry out a principal component analysis using the maximum covariance method, with the aim of finding the most robust spatio-temporal relationships between several candidate predictors and the accumulated monthly precipitation recorded in Cuba during the period 1980-2020. This process will make it possible to establish quantitative relationships that, together with theoretical considerations, make it possible to reduce the list of predictors to be used for the purpose of obtaining seasonal predictions. The values of the predictors are represented through monthly averages obtained from ERA5 reanalysis, while monthly accumulated precipitation data were obtained from a national-scope grid with 4 km of spatial resolution, used as predictand. The results obtained reflect the highest spatio-temporal correlation values with the first variability mode in all cases, indicating that the usual regime conditions are predominant and have a greater coupling with the precipitation variability in the analyzed temporal scale. In addition, they suggest that the candidates that explain the transport of moisture at low levels, as well as the gradients between the middle and lower troposphere, show the most robust associations. In the same way, the surface temperature of tropical Atlantic Sea, the flow related to Quasi-Biennial Oscillation and the thermodynamic indices, K Index and Galvez-Davison Index, present good degrees of association, for which reason they can be considered the most recommendable for carrying out forecasting experiments.

Keywords

Principal Component, Maximun Covariance, Predictors, ERA5

1. Introduction

The variability or change in precipitation patterns in any region has an impact on the functioning of societies, as it directly affects extremely sensitive areas such as water supply or food and industrial security. The Latin American and Caribbean region is not an exception to this reality, which increases the demand of societies for seasonal and sub-seasonal predictions related with the behavior of precipitation regimes, to efficiently plan water resources and mitigate the negative effects that may result from extreme behavior of this variable.

Different forecasting methodologies are used to obtain seasonal or sub-seasonal forecasts. The dynamical methods employ an atmospheric general circulation model (GCM), forced by sea surface temperature or coupled to an ocean model. Besides they have the advantage of not being constrained by purely linear considerations, so they are able to represent rare or infrequent events in weather patterns, in addition to the high computational cost required for their implementation, there are difficulties such as the generation of dry biases characterized by deficiencies in the prediction of the propagation of Madden-Julia Oscillation (MJO). They also tend to generate cold biases in the equatorial Pacific region, which affects the forecast of amplitude and intensity of ENSO events and inadequately represent the troposphere-stratosphere interaction [1] [2] [3] [4].

Another methodology employs statistical relationships between a predictor (variable used for forecasting) and the predictand (variable to be forecasted). The main advantage of these tools lies in the low computational resources that require; additionally, they can generate deterministic and probabilistic predictions. As for the disadvantages, there is the fact that they assume a stationary climate system, which often leads to difficulties in representing the dispersion of the predictands due to the linearization of the system, which implies deficiencies in representing the nonlinear interactions of the climate system [5].

One of the challenges of statistical schemes is the selection of predictors for spatial forecast of precipitation because requires a combination of statistical, physical, and expert knowledge. The most effective approach will depend on the specific application and the available data. Different approaches were encountered in the bibliography revision related to this selection process.

The statistical correlations between precipitation and various meteorological variables such as temperature, humidity, wind speed, and atmospheric pressure can be used to identify potential predictors for precipitation forecasting. Correlations can be computed using historical data and can provide insights into the relationships between different variables. However, correlation is a metric that does not imply cause-effect relationships; therefore weak correlations are not necessarily associated with the absence of spatio-temporal relationship between a potential predictor and the predictand.

The physical models can be used to identify potential predictors for precipitation forecasting, supported by theoretical considerations. These models simulate the physical processes that govern precipitation formation and can identify variables that are likely to influence precipitation patterns.

The expert's knowledge is a common tool used to identify potential predictors for precipitation forecasting. Experts in the field of meteorology can provide insights into which variables are likely to be relevant for precipitation forecasting based on their experience and knowledge.

Finally, data mining techniques such as principal component analysis (PCA) and cluster analysis can be used to identify patterns in large datasets. These techniques can identify potential predictors based on how they are related to precipitation patterns. This is related to machine learning algorithm development. These techniques can be used to identify potential predictors for precipitation forecasting. The algorithms can learn from historical data and identify patterns that are likely to be relevant for predicting precipitation behavior. Examples of methods are Random Forest and Support Vector Machine.

These predictor selection methods have convergency points and therefore, often are employed to combine theoretical knowledge with statistical associations. [6] employs monthly means of sea surface temperature (SST) using monthly precipitation accumulations as predictand, where their selection is because this variable can be associated with the behavior of ENSO and, since ENSO is a large-scale phenomenon, SSTs can be correlated with atmospheric circulation anomaly patterns. However, the results obtained showed that ENSO only partially governed the variability of the predictor in the study area.

For example, [7] used simple correlations to select the predictors. These authors use tropical North Atlantic SST anomalies, sea level pressure (SLP) and vertical shear in the equatorial Atlantic to predict precipitation in the Caribbean region. This selection is because ENSO-associated SST behavior modulates the strength of the Caribbean low-level jet (CLLJ) and favors convergence and tropospheric humidity, associated with the precipitation regime.

In relation to the CLLJ, several authors have identified it as a fundamental element in the seasonal variability of precipitation in the Caribbean region because it is a mechanism for moisture transport in this area [8] [9] [10] [11] and the role it plays in the mid-summer dry period when the expansion of subtropical ridge combines with the expansion of warm pools and the CLLJ enhances divergence towards the Caribbean and convergence towards Central America [12].

More recently [13] proposed the use of SST, SLP, 850 hPa surface zonal wind and vertically integrated mean moisture transport flux (Uq, Vq) applied to different Caribbean subregions. The authors point out that not all the proposed predictors adequately reflect the predictor-predictand relationship in the different subregions, also identifying the mechanisms associated with the NAO and ENSO as the main modulators of the precipitation in selected area.

The use of EOF (empirical orthogonal functions) and CCA (canonical correlation) to a set of candidate predictors has dominated foreign works in the last five years. These techniques are based on the principle that the spatial structure of variability at given time scales can be interpreted as the interaction between different components that determine their variability modes [14] [15] [16].

The zonal wind component at 850 and 200 hPa levels, specific humidity, temperature at 850 hPa surface and the variables that compose the Multivariate ENSO Index (MEI) have been employed interchangeably. All represent the climatic drivers already mentioned [17].

Studies related to seasonal forecasting in Cuba have been cursory. However, research associated with the identification of future scenarios warns that annual precipitation could be reduced by 20% - 50%, increasing the frequency and intensity of drought periods [18], hence seasonal drought forecasting is both a necessary and achievable goal.

However, most national studies are related to the diagnosis or characterization of climatic conditions associated with different patterns [19] [20] [21] [22]. On the other hand, some researches found an increase in the annual frequency of synoptical situation type characterized by marked influence of subtropical anticyclone with first quadrant flow [23] [24], which suggests, in fact, a greater influence of subtropical ridge, which may be related to the expansion of Hadley cell [25].

Despite these efforts, attempts to obtain seasonal precipitation forecasts using purely statistical [26] [27] and dynamic [28] schemes have not been fully developed.

The present research proposes as a first step for generation of a statistical or dynamic-statistical seasonal prediction scheme, to identify the predictors that have the greatest degree of spatio-temporal association with the precipitation that would be considered as predictand. Taking into account the insufficiencies of purely theoretical considerations or the application of certain metrics such as linear correlation, which can limit the spectrum of possible predictors; it is proposed to use the maximum covariance method (MCA). This methodology allows establishing numerical relationships that respond to the joint predictor-predictand dynamics, offering the additional advantage of identifying the predominant modes of variability.

2. Materials and Methods

2.1. Candidate Predictors

The candidate predictors are determined a priori taking into account the physical mechanisms that govern the predictor in the tropical zone, specifically in the Caribbean region where Cuba is located. The variables that describe the main climatic drivers, such as ENSO, NAO, MJO, QBO (Quasi-Biennial Oscillation), are taken into account for empirical selection of potential predictors. In all cases, monthly mean values for the period 1980-2020 are used, employing as primary data source the ERA5 reanalysis (https://cds.climate.copernicus.eu/) with 0.25˚ of spatial resolution. These potential predictors can be subdivided into three groups to facilitate their evaluation. The first group includes the SST of the Niño 3.4 region, the Gulf of Mexico, Caribbean and tropical Atlantic (**Figure 1**).

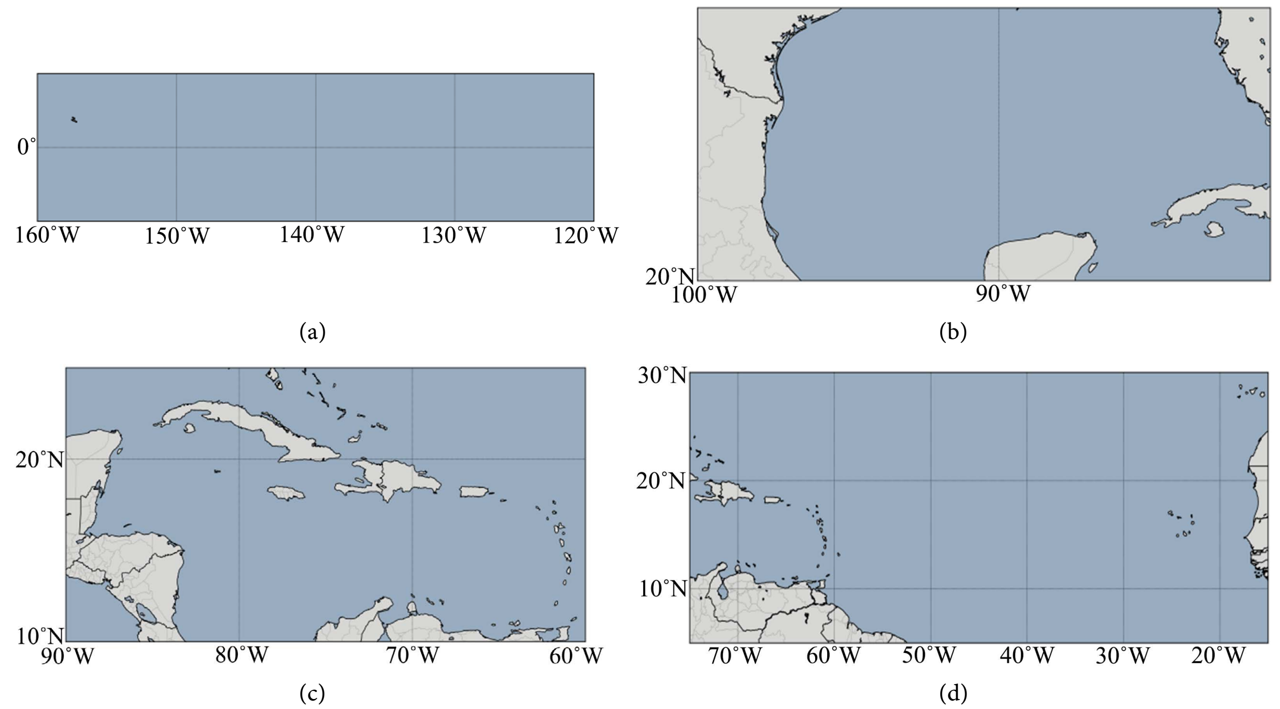


Figure 1. Monthly mean SST selected areas. (a) ENSO 3.4, (b) Gulf of Mexico, (c) Caribbean Sea, (d) Tropical Atlantic.

The second subgroup relates the candidate predictors obtained at different pressure levels. In this case, as well as the third and final subgroup, which relates surface-level candidate predictors and thermodynamic indices, was obtained from the Caribbean subregion (**Figure 2(a)**). It should be noted that the selection of this area is arbitrary and seeks to capture the greatest possible influence of different modes of variability that could modulate the behavior of precipitation patterns in Cuba, without the need to resort to a too-large area that would imply a greater computational cost.

As will be explained in Section 2.3 below, the maximum covariance method decomposes the predictor into its principal components; its EOFs are projected into the space of the predictand to know the thresholds of the spatial association.

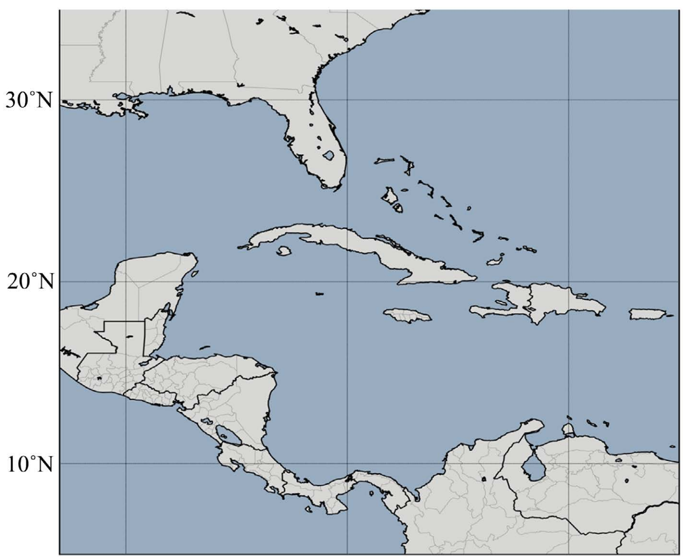
The list of all potential predictors that will be evaluated in this research is reflected in **Table 1** below:

The K Index (KI) and Gálvez-Davison Index (GDI) were included in the list of candidates. Its inclusion was motivated by the results obtained by [29], who evaluate the use of CAPE (Convective Available Potential Energy) transport term to identify spatial patterns of certain circulation regimes associated with the American Monsoon. However, it should be noted that this index has limitations in the tropics since its values can be high because the tropopause is usually higher in this region, so high CAPE values do not always correlate with deep convective regimes, a fact that is more noticeable in the summer months.

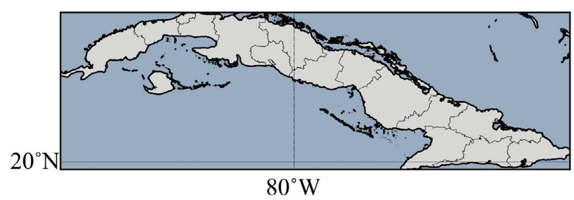
In addition, it should be noted that the traditional indices were designed for environments with strong temperature and humidity gradients, which is not always evident in the tropical zone where these gradients are usually much weaker

Table 1. List of potential predictors.

Potential predictors	Subgroup	Unit of measure
SST ENOS 3.4	SST Areas	°K
SST Gulf of Mexico		°K
SST Caribbean Sea		°K
SST Tropical Atlantic		°K
Temperature in 200 hPa	Selected at different isobaric surfaces	°K
Temperature at 850 hPa		°K
Divergence at 200 hPa		1/s
Divergence at 850 hPa		1/s
Geopotential at 200 hPa		m^2/s^2
Geopotential at 500 hPa		m^2/s^2
Zonal component at 30 hPa		m/s
Zonal component at 50 hPa		m/s
Zonal component at 200 hPa		m/s
Zonal component at 850 hPa		m/s
Pressure at mean sea level	Selected at surface including thermodynamic indexes	Pa
Outgoing Longwave Radiation		w/m^2
K-index		°K
Gálvez-Davison Index		-



(a)



(b)

Figure 2. Caribbean subregion where the monthly mean values of the rest of candidate predictors are obtained (a). Predictand grid area (b).

in relation to mid-latitudes. Of these, the KI (Equation (1)) shows the greatest skill, presumably because its calculation takes into account the moisture content at low levels, which is essential for precipitation processes in the tropics. The fundamental disadvantage of this index, in short-term forecasting experiences, is that it tends to exaggerate the areas where precipitation may occur, but it can be used as an indicator of the average behavior of moisture content at low levels.

$$K = \left(T_{850} - T_{500}\right) + \left(T_{500} - Td_{850}\right) - \left(T_{700} - Td_{700}\right) \tag{1}$$

On the other hand, the GDI is a stability index generated to improve convective forecasting in the Caribbean [30]. It has been validated for tropical areas and in southeastern South America, finding that for Caribbean regions it can explain more than 50% of precipitation's variance. Its formulation considers three physical processes that modulate tropical convection: the simultaneous availability of heat and moisture in the middle and lower troposphere; the stabilizing/destabilizing effects of middle and upper-levels caused by ridges and troughs; and dry air entrainment and stabilization related to inversions (Equation (2)). These considerations suggest that applications of this index need not necessarily be limited to the mesoscale and higher average values may indicate the continued presence of disturbed conditions that generate higher precipitation accumulations.

$$\text{GDI} = \text{ECI} + \text{MWI} + \text{II} \tag{2}$$

where ECI is a stability index, MWI represents the heat content at middle levels and II is the inversion index.

2.2. Precipitation Data

To evaluate the relationship's strength of each proposed predictors with a predictand, precipitation in Cuba generated from a national database is used, which will be referred to as Solvaz in the following, alluding to original authors [31]. This database is built from data collected by conventional meteorological stations and the rain gauge network of the National Institute of Hydraulic Resources, which are subsequently interpolated to generate a 4 km (kilometers) spatial resolution grid. Monthly accumulated are recorded in this grid, which is updated month by month. As in the case of the predictors, the period 1980-2020 is used.

2.3. Maximum Covariance Analysis (MCA)

As described above, the methods for selecting predictors can be purely qualitative, where theoretical associations are made from the physical mechanisms governing the predictor, simple associations such as linear correlations can be established or more complex methodologies such as principal component analysis can be applied.

It should be noted that in the case of qualitative assessments, the selection of predictors may have a high subjective component because, beyond the specialist's theoretical knowledge, its ability to recognize the connection of variables describing large-scale circulations with observable patterns in a given area must be taken into account, which is not always a simple task. On the other hand, linear correlations are measures that do not imply causality, so low correlations do not necessarily imply the absence of a functional relationship between the potential predictor and the predictand.

From this reasoning, the use of the EOF can allow to establish a more robust spatial and temporal association between the predictors and the predictand to determine which the most robust predictors are.

For this reason, it was decided to use the MCA in this research [32]. This method allows the identification of coupled fields of EOF and PCA (Principal Components), which count a covariance fraction of two variables analyzed simultaneously (candidate-predictor predictand). One of the advantages of the method is that the spatial dimensions of the matrices do not need to be equal, just the time step.

Each mode of covariance is determined by a pair of spatial patterns (one for each field), a pair of principal components showing how their respective spatial pattern evolves over time, and a singular value indicating the quadratic covariance accumulated by each variability mode.

Assuming a dataset consisting of N observations and p predictor variables, $X_1, X_2, \cdots, X_n$, and a target variable Y, the MCA method can be expressed mathematically as follows:

Calculate the covariance matrix between the predictor variables and the target variable:

$$C_{xy} = \operatorname{cov}\left(\left[X_1, X_2, \cdots, X_p\right], Y\right) \tag{3}$$

Compute the eigenvalues and eigenvectors of the covariance matrix C_{xy}:

$$\left[\lambda_1, \lambda_2, \cdots, \lambda_p\right], \left[v_1, v_2, \cdots, v_p\right] = eig\left(C_{xy}\right) \tag{4}$$

where $\lambda_1, \lambda_2, \cdots, \lambda_p$ are the eigenvalues and $v_1, v_2, \cdots, v_p$ are the eigenvectors.

Sort the eigenvalues in descending order and select the k eigenvectors corresponding to the k largest eigenvalues. These eigenvectors represent the k linear combinations of predictor variables that have the highest covariance with the target variable.

Compute the k-dimensional linear combination of predictor variables as:

$$Z = \left[v_1, v_2, \cdots, v_k\right]^{\mathrm{T}} \left[X_1, X_2, \cdots, X_p\right]^{\mathrm{T}} \tag{5}$$

where is the matrix of eigenvectors corresponding to the k largest eigenvalues and is the transpose of the matrix of predictor variables.

The k-dimensional linear combination Z is the set of predictors that has the highest covariance with the target variable Y.

In addition to what is described above, the number of eigenvectors and the corresponding number of predictors to select may also be determined using cross-validation or other model selection techniques.

3. Results

3.1. Precipitation in Cuba

Precipitation in Cuba, as is characteristic of tropical zones, has a dry season (November-April) and a wet season (May-October). The latter usually exhibits two maximums, one in May, which is usually related to the presence of systems such as the May-June seasonal trough and, in the case of the maximum at the end of the wet season, it must be conditioned by the peak of cyclonic activity in the North Atlantic and the minimum of atmospheric pressure in the tropics that usually occurs in October (**Figure 3**).

Some variations in the classic patterns associated with rainfall variability on the island suggest structural changes in these patterns. It is difficult to determine if they are a product of variability or, on the contrary, are linked to permanent changes. But there is no doubt that they have an impact on the seasonal behavior of precipitation. The most significant cases are related to a trend of increasing high geopotential values at mid-levels (**Figure 4**). These increases may be associated with expansion processes and modifications in the morphology of subtropical anticyclone, which have been documented by other national authors [23] [24] [33] [34].

These variations often lead to the creation of a high geopotential belt in the tropical zone, connecting the subtropical ridge with the Mexican ridge, which can affect the exchange with extratropical regions and the appearance of systems that commonly generated precipitation on the island, such as the May-June trough in the absence of meridional gradients to support it.

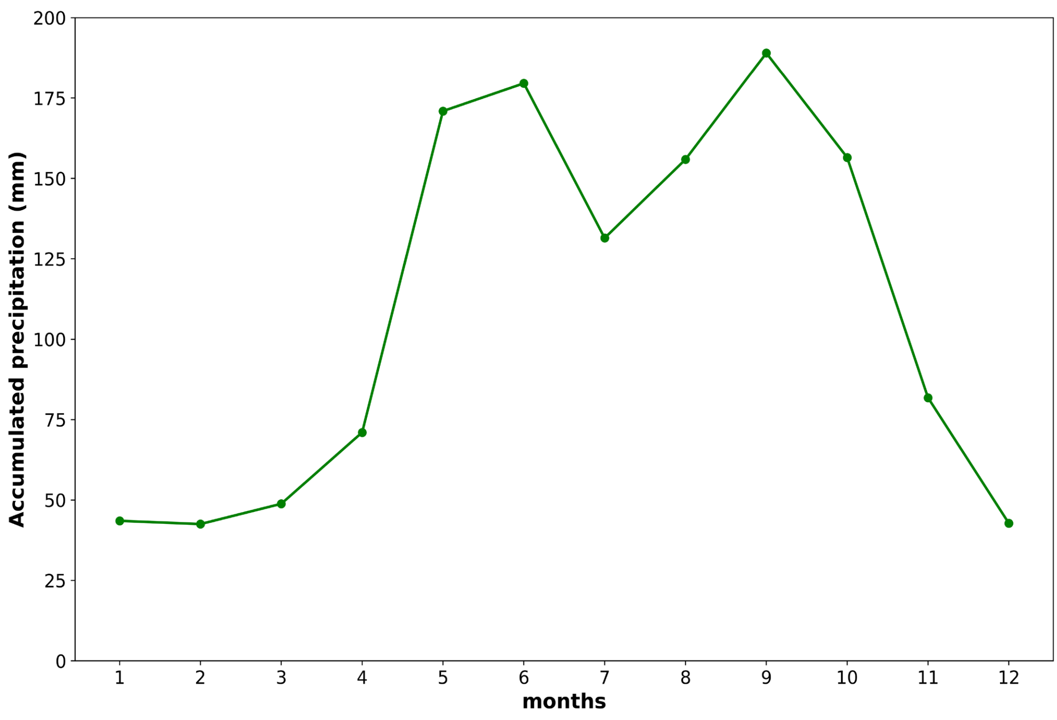


Figure 3. Mean annual cycle of precipitation in Cuba.

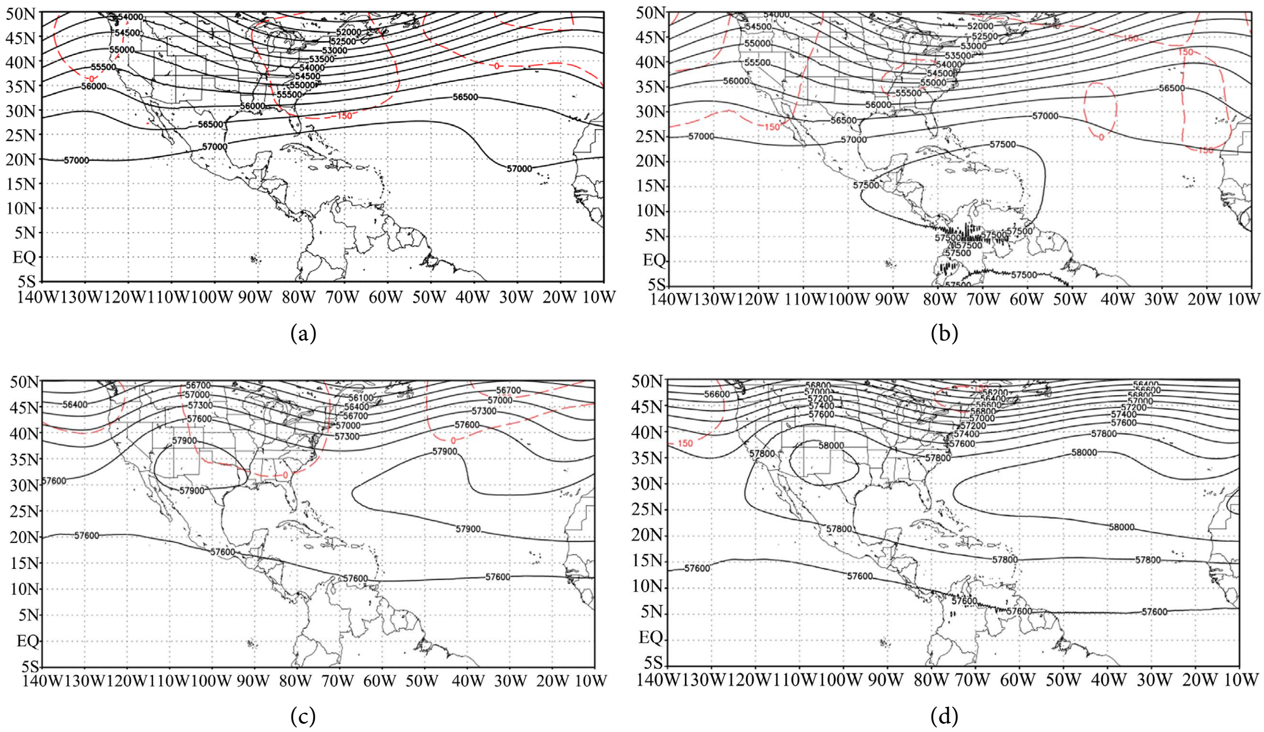


Figure 4. Decadal mean values of geopotential at 500 hPa. (a) January-decade 1980-1989; (b) January-decade 2010-2019; (c) July-decade 1980-1989; (d) July-decade 2010-2019.

Another source of precipitation variability is the TUTT (tropical upper tropospheric trough) (Sadler, 1967; Evans & Laing, 2016). This system seems to have experienced some tendency to contract and to induce an increasingly zonal flow over the Antilles region (**Figure 5**). These variations will have an impact on the precipitation regime since the TUTT cells associated with these structures participate in the convective precipitation cycles in the tropics during the wet season, in addition, the TUTT as such, usually participate in convective organization processes that can derive in tropical cyclogenesis.

As mentioned above, both the impact and the permanence of these variations are not entirely clear, particularly because their permanence over time is not long enough to be dominant over a 30-year climatic period. What is certain is that the measurements on the island suggest that, together with these variations in the large-scale flow, certain modifications appear in the usual precipitation patterns.

These changes seem to be more evident in relation to a drastic reduction in average precipitation towards the eastern half of the island, a fact that becomes more relevant considering that annual accumulations in this region of the country are usually lower than in the western half (**Figure 6**). On the other hand, there is a certain tendency towards an increase in the amplitude of precipitation variability, as a consequence of the simultaneous increase in the number of reports of heavy rainfall (according to national standards) and days without precipitation.

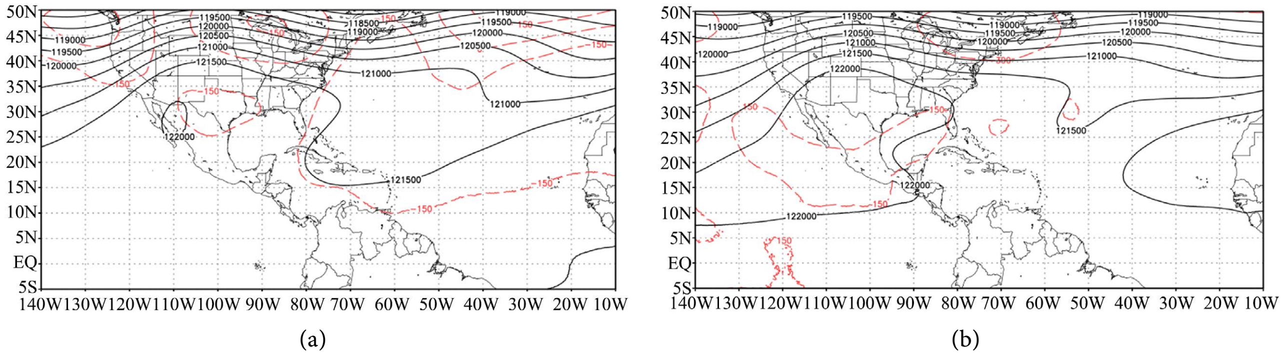


Figure 5. Decadal mean values of geopotential heights at 200 hPa in July. (a) 1980-1989 period showing a classical pattern; (b) 2010-2019 period, with a practically zonal configuration over Cuba.

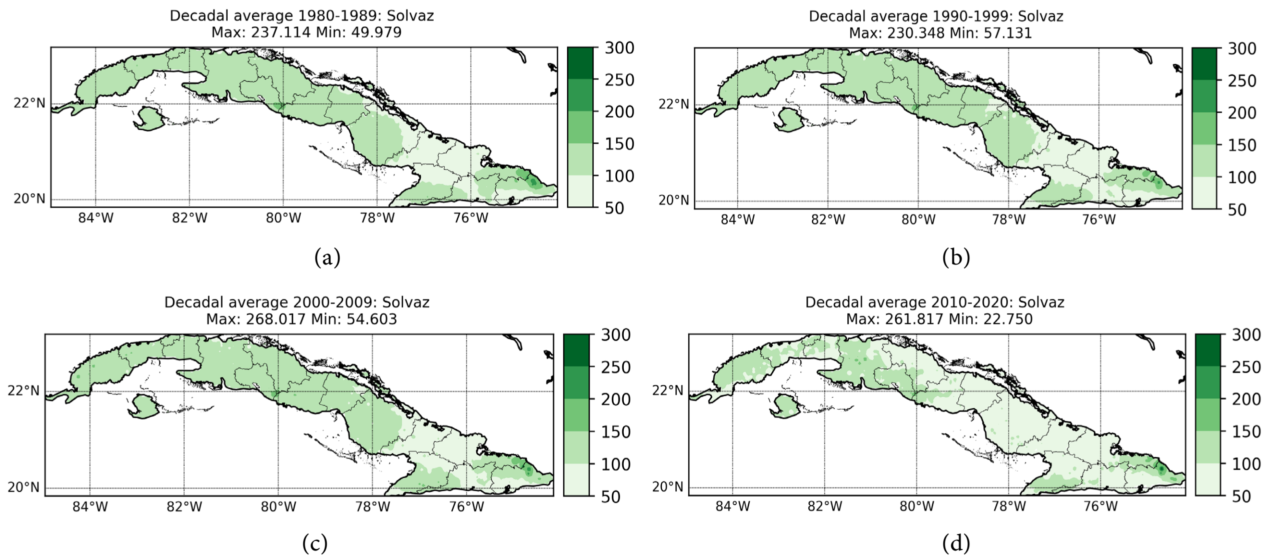


Figure 6. Spatial behavior of decadal mean accumulated for the study period.

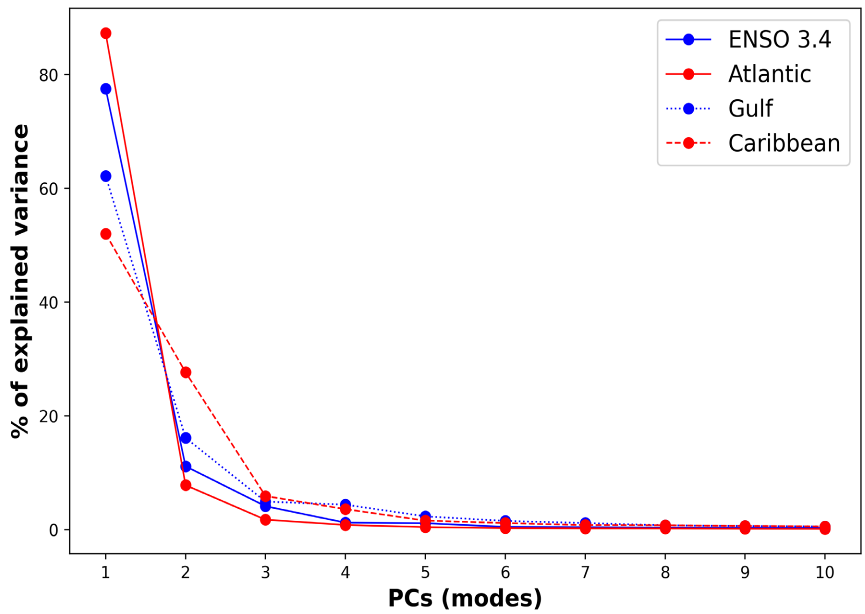
3.2. Evaluation of Candidate Predictors

3.2.1. SST Subgroup

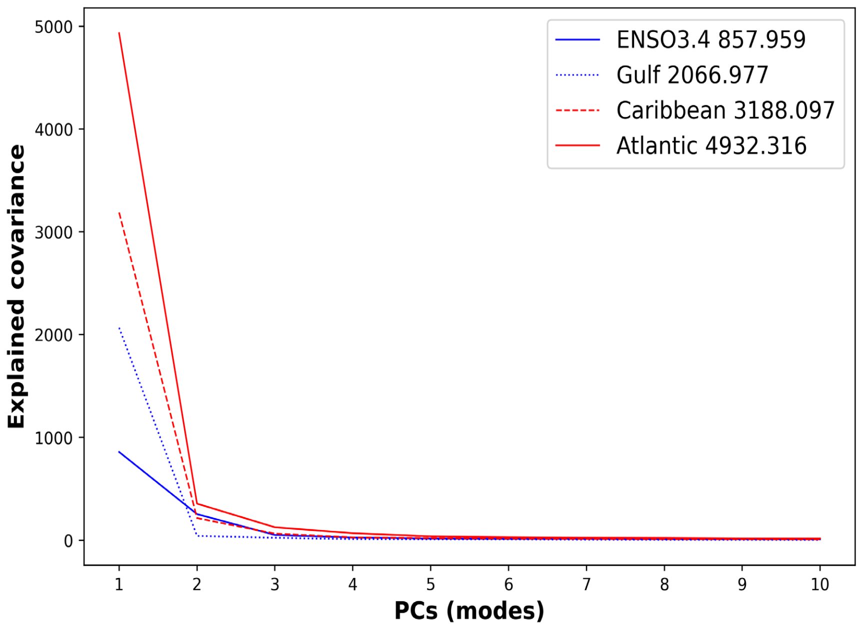
The sea surface temperature is a modulating factor of tropical convection. In the specific case of the Caribbean region, the thermal gradients associated with the presence of hot pools, the circulation of the Loop Current, the Atlantic's meridional circulation, among other processes, influence the transport of energy capable of feeding tropical convection. Small-scale movements such as breezes, which in island regions (as Cuba), are an important mechanism of convective genesis, should not be forgotten.

The variability in distant regions such as the equatorial Pacific, in clear allusion to ENSO, usually disturbs the state of the atmosphere, generating more baroclinic or barotropic tendencies depending on the ENSO signal. Particularly this event has been treated for decades as the main source of climatic variability of the island; however, it should be considered that a process of expansion of the Atlantic subtropical ridge could redistribute the areas of usual influence of the ENSO, reinforcing some and/or weakening others. A priori, it can be predicted that this process could lead to conditions with more barotropic tendencies in the Caribbean region, making it less sensitive to certain ENSO fluctuations in the region.

The analysis made with the SST of ENSO 3.4 region shows that the first mode of variability explains almost 80% of the total variance of the SST in that region (Figure 7). This suggests that there has been a predominant signal in the selected study period.



(a)



(b)

Figure 7. Variance (a) and covariance (b) explained by each variability modes of SST predictors.

Regarding the spatial association expressed through the explained covariance, ENSO shows the lowest values, being the first mode, as in all other cases, the one with the highest covariance value, since the rest quickly tends to zero. This result is consistent with the temporal relationships found between the precipitation described by Solvaz and ENSO 3.4, where the relationships are weak (**Figure 8**). These results suggest a decoupling between the SST anomalies of ENSO 3.4 region and precipitation in Cuba, which may suggest a decrease or lag regarding the influence of processes that condition heat and moisture transport and may react to the gradients between the Atlantic and the Pacific, such as the Caribbean low-level jet or the subtropical jet.

In the case of the SSTs of the Gulf and Caribbean regions they show better associations but they are not the most robust. Spatially the first variability mode in both cases coincides with typical conditions where heat is transported by currents in usual regime situations and the convective activity moves in the trade wind flow. This mode shows the highest temporal correlations, with values of 0.71 and 0.69 respectively, whereas the spatial associations behave at a low-moderate threshold.

These results also show little variation over time of magnitudes associated with the Gulf of Mexico region. This may be an indication that this region does not capture the trend of the variations previously mentioned, which could condition the precipitation patterns in Cuba. Its limited spatial coverage and the fact

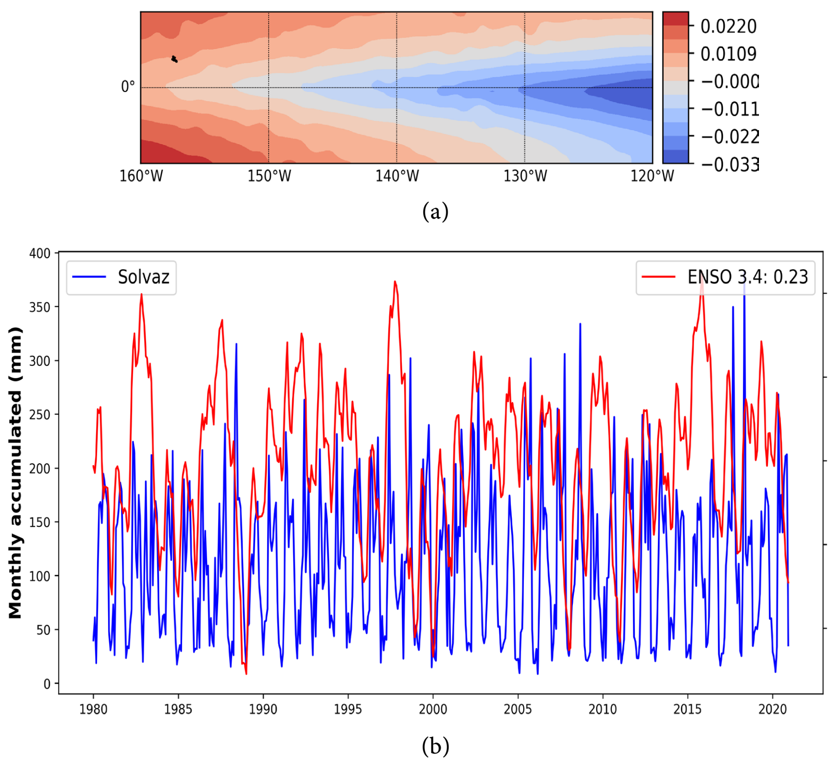


Figure 8. (a) Representative EOF of first variability mode of SST ENSO 3.4; (b) Temporal correlation between the principal component corresponding to the first mode and the original precipitation field.

that it is not a region producing centers of action may be some of the causes of this behavior. On the other hand, the Caribbean region shows a slight tendency to a negative bias in the series corresponding to the first mode of variability, which is indicating a shift in the mean conditions associated with the usual regime and allows it to establish a more robust association in the spatio-temporal scale.

The first variability mode of tropical Atlantic SST shows the highest degree of covariance in relation to the precipitation field (Figure 9), suggesting that this candidate has the highest degree of spatio-temporal association. This mode also represents usual regime conditions, where heat and moisture transport embedded in the easterly flow conditions the lower troposphere and modulates vertical shear by actively participating in tropical convection processes. This impacts on the usual flow of easterly waves where SST values influence large-scale convergence which, together with the aforementioned elements, conditions the activity of these systems, an important source of precipitation, especially in the wet season.

However, the spatio-temporal correlations are in the moderate range. It can be seen that there are certain periods of lag in the fluctuations of the series, which are dominant in the first half of the study period. Towards the second half, which represents the last two decades approximately, the behavior of the series exhibits a greater coupling. Beyond the correlation value that summarizes the association between both series, the observation suggests that the structural

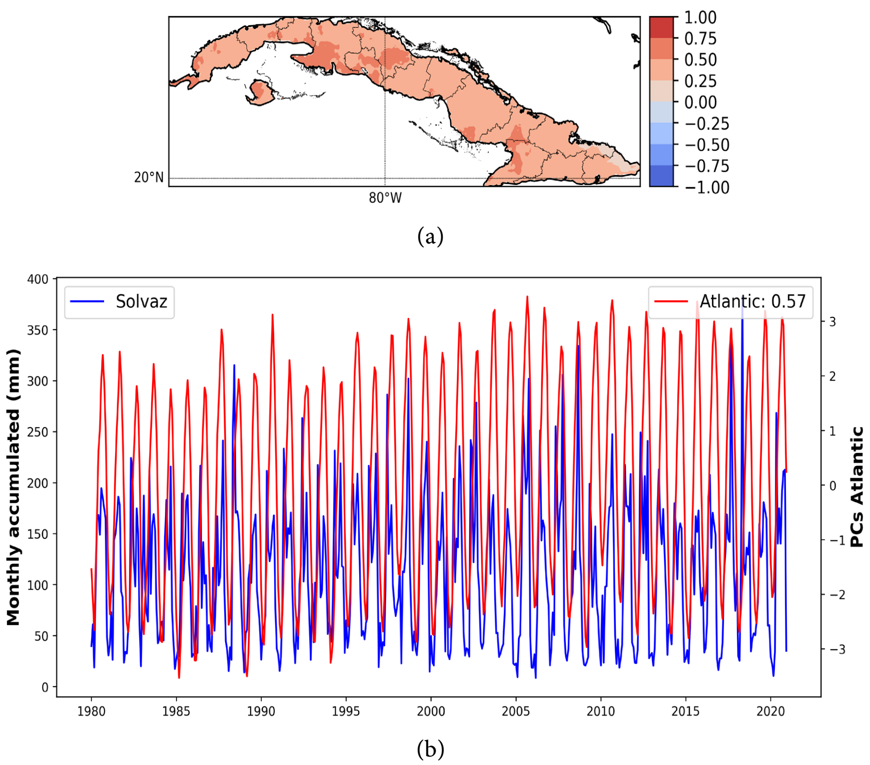
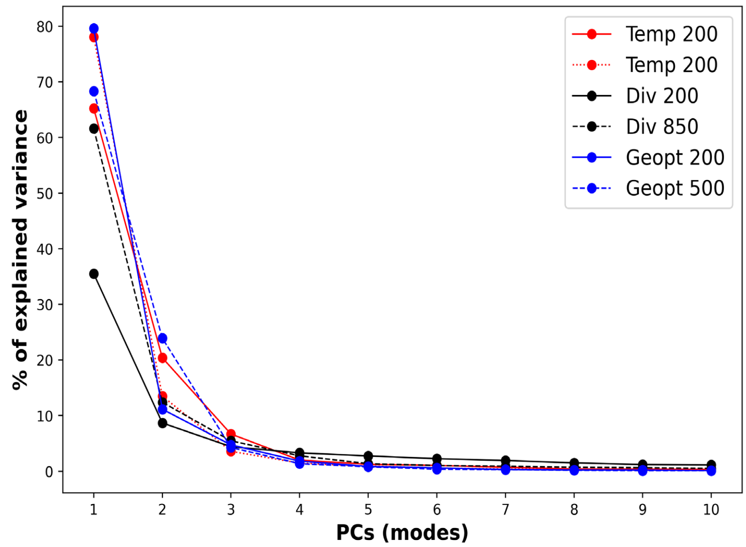


Figure 9. Spatio-temporal relationship between the first variability mode of Atlantic SST and precipitation field. (a) Heterogeneous maps; (b) Principal component correlation.

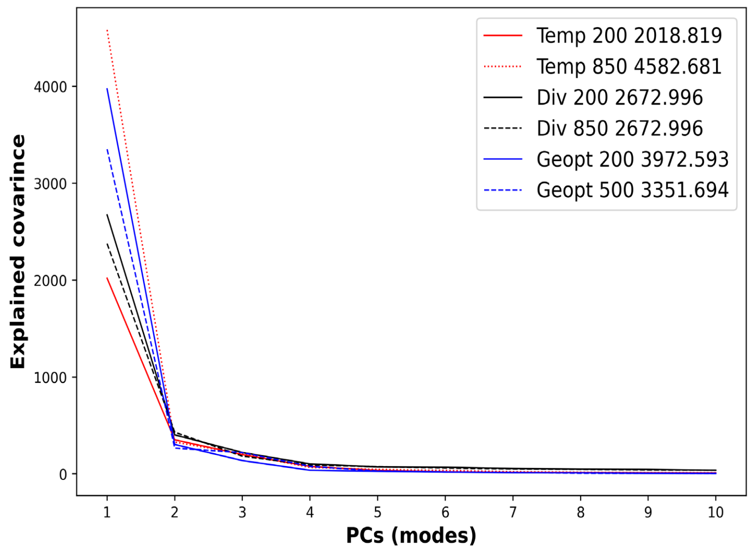
changes in the subtropical ridge and consequently in the meridional gradients of SST seem to lead to a greater influence in the modulating processes of precipitation in the island.

3.2.2. Predictors on Different Isobaric Surfaces

The temperature, divergence and geopotential fields are usually associated with the movement of troughs and ridges, tropospheric moisture content and consequently with atmospheric stability and instability conditions. As in the case of SSTs, the first variability mode explains more than 50% of the variance except for the case of divergence at 200 hPa, hence the first mode largely represents the prevailing conditions in the series estimated by the reanalysis (**Figure 10(a)**).



(a)



(b)

Figure 10. Variance (a) and covariance (b) explained by each variability modes of geopotential, temperature and divergence.

However, the covariances (**Figure 10(b)**) shows that the first modes of the temperature at 850 hPa and geopotential fields are the most robust. In the case of temperature at 850 hPa, it is a variable that defines the moisture holding capacity of the lower troposphere which, as already mentioned, is a crucial mechanism in tropical convection. It is also associated with moisture transport at low levels, where mechanisms such as the American monsoon or the MJO are associated.

In relation to geopotential, the behavior at both levels is very similar. The first mode captures the variability associated with the typical structure resulting from the influence of subtropical ridge. However, other structures such as the TUTT (at 200 hPa level) or the westward expansion of subtropical ridge at the 500 hPa surface are captured in other variability modes (**Figure 11**).

With this behavior of EOFs it is not surprising that the temporal association is only moderate with the first mode and very weak with the rest, since the seasonal structures, although they may persist long enough to decrease or overlap with the usual regime conditions, are not sufficient to explain the variation of precipitation taking into account the whole annual cycle. In other words, the individual correlation of each of these signals with respect to the variability of the original precipitation field is reduced.

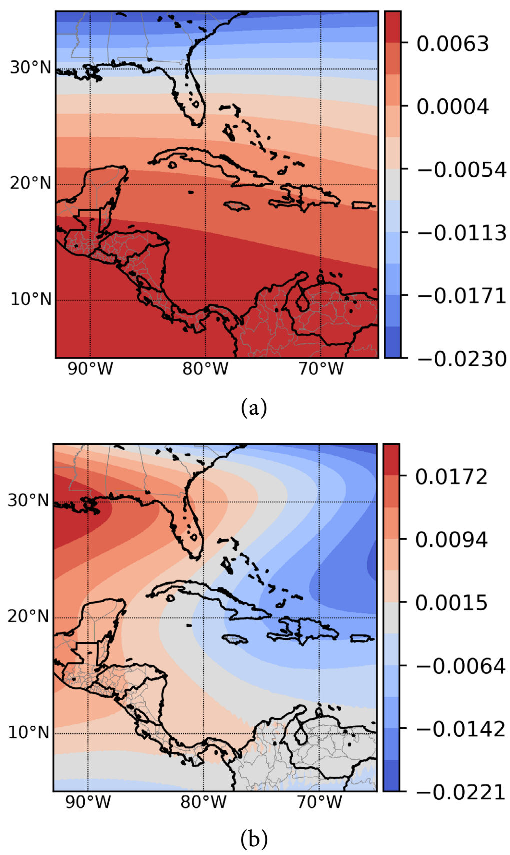


Figure 11. EOF corresponding to 200 hPa geopotential field; (a) first variability mode; (b) second variability mode.

Regarding the temperature at 200 hPa level, the associations corresponding to the first variability mode are moderate, being the highest in relation to the rest of variability modes, whose EOF represent a pattern similar to that shown in the geopotential field, which may be related to the thermal expansion characteristic of Hadley cell. Just as in the case of the geopotential, the following EOFs identify the seasonal structures that tend to prevail in the selected study period, in any case, a single mode of variability does not seem to be sufficient to explain the variation of precipitation with this variable.

The divergence fields, although theoretically a term describing atmospheric stability/instability conditions, presented the weakest associations on the spatial scale, with predominantly moderate correlations. This contrasts with the temporal relationships which turned out to be the strongest with values above 0.80 for both terms analyzed. It can be observed that the representation of the EOFs shows very little variation (**Figure 12**), which could influence the results achieved.

As for the zonal component of the wind, very similar characteristics to those described above are observed (**Figure 13**). In all cases, the first variability mode represents more than 50% of the explained variance and shows relatively high spatial covariance values when compared to other candidate predictors.

In the case of zonal components at 30 and 50 hPa levels, they clearly allude to the Quasi-Biennial Oscillation. Similar to what was observed with the ENSO (Section 3.2.1), the first modes of variability represent the phases of this oscillation where there is one predominant. The decomposition of zonal wind in these levels suggests that the west phase of the oscillation predominates, although this fact is less noticeable in the 50 hPa zonal pattern, where the first two modes explain similar values of variation (**Figure 14**). Nevertheless, the west phase is the most closely related to the precipitation processes described by Solvaz, not only spatially, but also temporally, although it exhibits moderate correlations.

These results are consistent with characteristics of the QBO, where westerly winds tend to last longer than easterly winds. On the other hand, it has been described that the westerly component influences cyclonic activity in the North Atlantic, since by decreasing the shear between the stratosphere and the upper troposphere, they generate more barotropic conditions that favor the increase of deep convection in the region, which also plays an active role in the American monsoon.

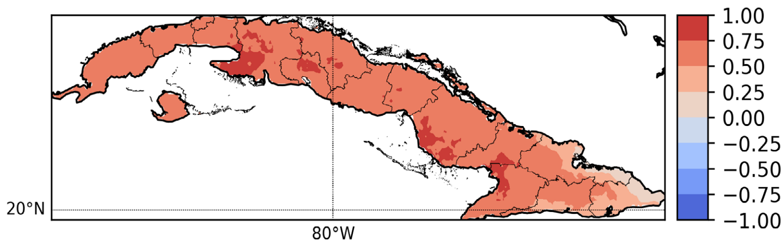
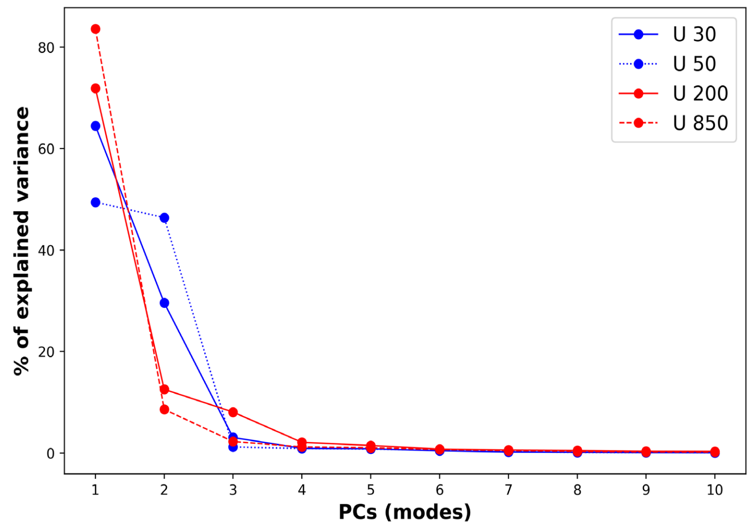
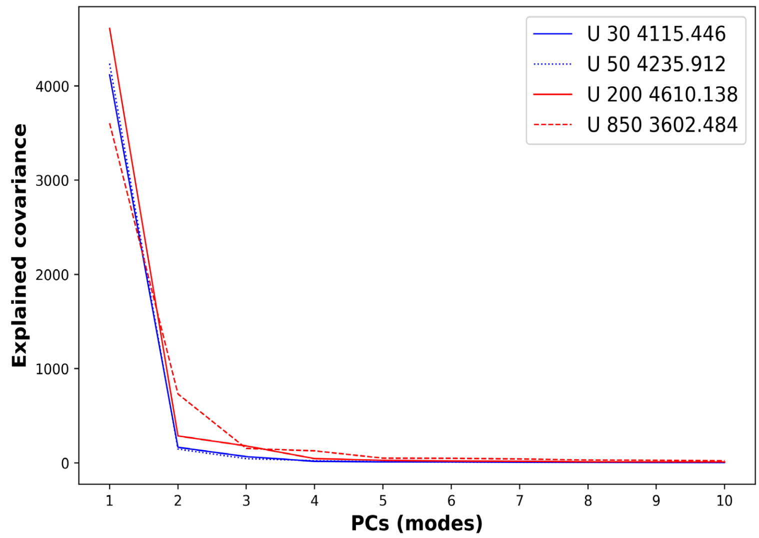


Figure 12. Heterogeneous field of the divergence's first variability mode at 200 hPa level.



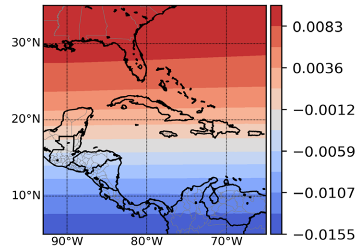
(a)



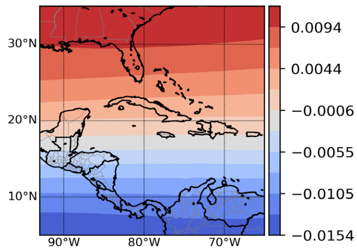
(b)

Figure 13. Variance (a) and covariance (b) explained by each variability modes obtained from the analyzed zonal components.

The zonal components at 200 and 850 hPa show spatial associations comparable to the previous ones, in both cases with a distribution representative of usual regime conditions (**Figure 15**). In the case of the zonal component at 850 hPa, it shows the lowest value of spatial association, which may be conditioned by the fact that moisture transport at low levels is not sufficient to explain the variation of the spatial distribution of precipitation. This may be because the transport mechanisms at this level, as well as the Caribbean LLJ, are present regardless of whether or not precipitation occurs on the island, suggesting that their influence may be limited to certain critical thresholds. On the other hand, precipitation-producing systems such as the easterly waves that usually travel in this flow need additional favorable conditions, mentioned previously (Section 3.1), to activate convergence.

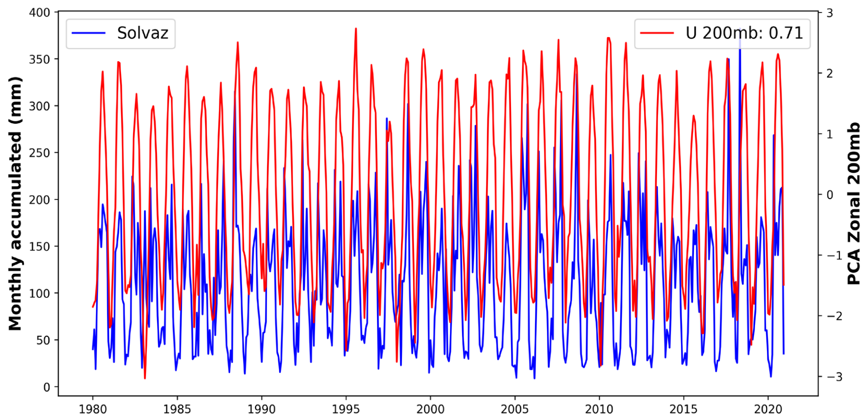


(a)

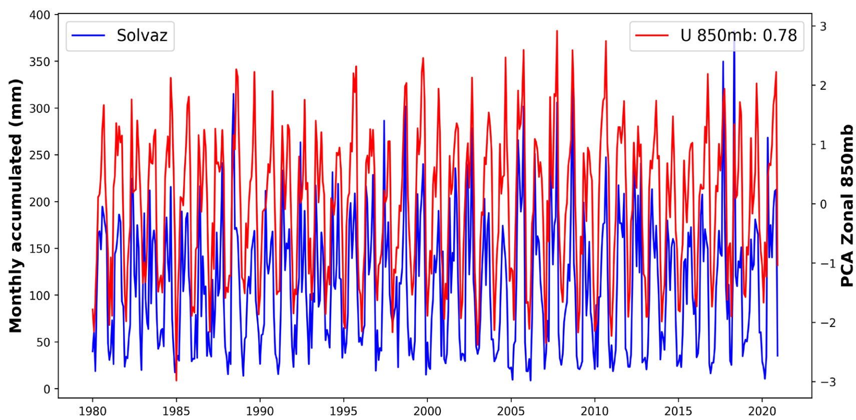


(b)

Figure 14. EOF corresponding to the first variability mode in the case of (a) zonal component at 30 hPa; (b) zonal component at 50 hPa.



(a)



(b)

Figure 15. Principal component analysis showing the temporal association of the first variability mode of zonal wind component at 200 hPa (a) and 850 hPa (b).

In the opposite direction, the flow at high levels is linked to jets and the MJO, capable of generating divergent patterns that favor upward movements and, therefore, atmospheric instability, influencing to a greater extent the spatial distribution of precipitation. It should be noted that both components show a high degree of linear association on the time scale (**Figure 15**), which implies that their seasonal and intraseasonal variability presents a high degree of coupling with precipitation.

3.2.3. Candidate Predictors Represented by Surface Variables and Thermodynamic Indexes

Candidate surface predictors show comparable covariance values with those exhibiting higher values.

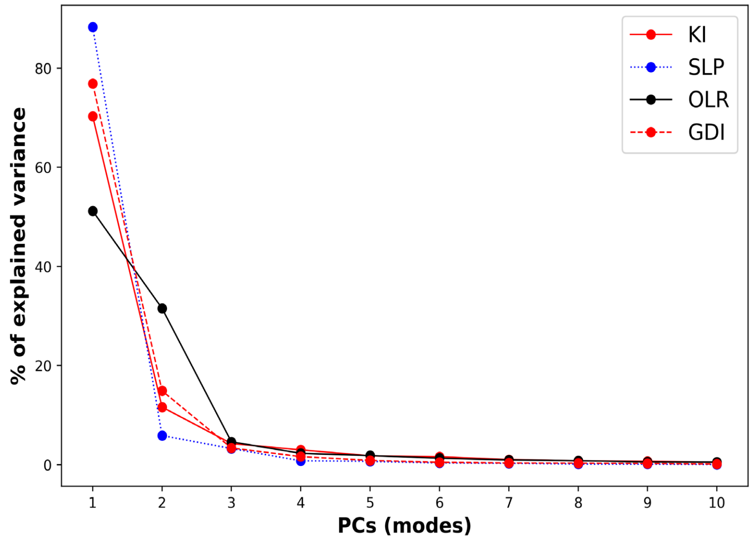
The OLR and the proposed thermodynamic indices are the highest. The OLR, being the outgoing radiation emitted by the clouds, is a variable that must react to the presence of different convection regimes, so its behavior must be very close to an annual precipitation cycle, with the consequent seasonal fluctuations. In fact, once again it is observed how the pattern constructed by the first variability mode responds to seasonal cycle, in which the fluctuations of the intertropical convergence zone (ITCZ), monsoon activity and the MJO are involved (**Figure 16**). The degree of temporal association is high, suggesting a strong spatio-temporal coupling of OLR with precipitation variability.

In relation to the SLP, a moderate spatial association is observed, with the highest correlation in the first variability mode. This mode reflects the mean position of the subtropical ridge that responds to the trade wind flow characteristic of the tropical zone. However, easterly disturbances do not necessarily lead to the triggering of convective processes, which could explain the weakness of the associations in some regions of the country. In contrast, the linear correlation between the first variability mode and the precipitation time series is strong, although it is the weakest of this subgroup (**Figure 17**).

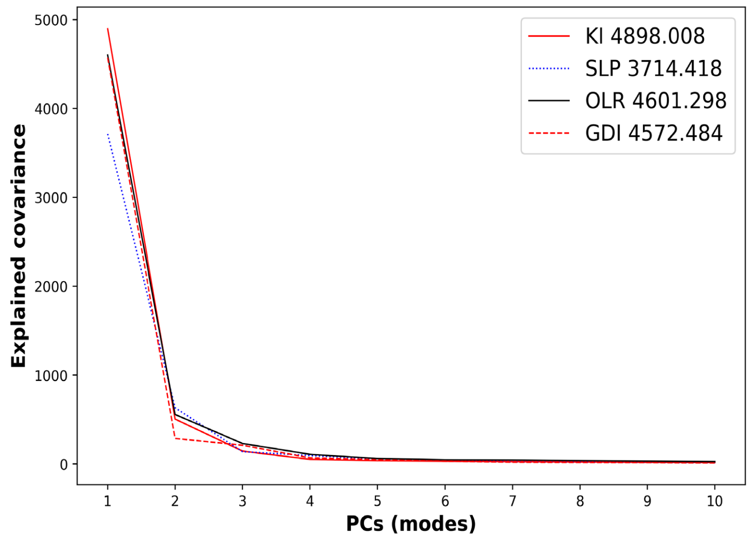
The thermodynamic indices showed good spatio-temporal association, clearing the question of whether they were able to reflect the seasonal and intraseasonal course of precipitation. In the case of KI, reflecting the low-level moisture holding capacity and the gradients between the middle and lower troposphere may reflect the annual cycle of precipitation on seasonal scales. On the other hand, the GDI not only reflects this moisture transport, but also the effects of the successive movement of troughs and ridges, which allows following disturbances embedded in the trade wind flow, even those that generate flat convection [30]; hence, if the first variability mode represents usual regime conditions, this index can successfully identify the conditions of stability/instability generated by this pattern.

3.3. Discussion

The results obtained using the maximum covariance method suggest that the candidate predictors related to low-level moisture transport and gradients



(a)



(b)

Figure 16. Variance (a) and covariance (b) explained by each variability modes obtained from the candidate predictors of surface subgroup.

between the middle and lower troposphere generally are conductive to the strongest associations on spatio-temporal scales, which is consistent with the average characteristics of convective processes that tend to occur in the tropical zone.

This is indicative that the variables related to humidity fluxes at low and medium levels are more skilfull to represent rainfall than others predictors usually used such as SSTs in different subregions.

These results coincide with [13] where it is stated that the evaluations carried out in fields related to humidity flows [29] [35] [36], has been obtained which are more adept at predicting various characteristics of precipitation than other options such as sea surface temperature.

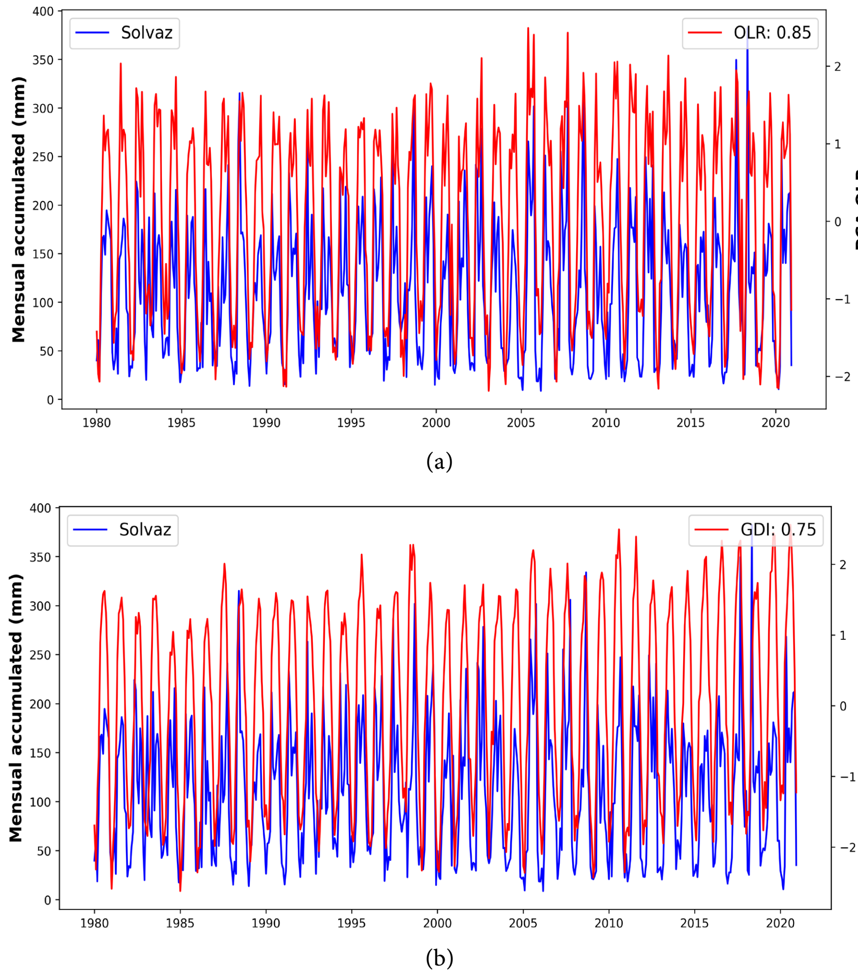


Figure 17. Temporal correlation between the first variability mode and the original precipitation field. (a) OLR; (b) GDI.

However, the results suggest that the tropical Atlantic is an important modulator of the spatio-temporal distribution of precipitation, at least on the island.

An exception to this rule could be the tropical Atlantic, which is an important modulator of the spatio-temporal distribution of rainfall, at least on the island.

Others predictors with high reliability such as the zonal component and temperature at 850 hPa, as well as sea level pressure, coincide with the evaluations made by other authors in the Caribbean region [13] [37], where it is obtained that they are more skilfull explaining the variability of the predictand.

Through the MCA it is obtained that the thermodynamic indices that represent the aforementioned characteristics can constitute good predictors. In this case, the possibility of using GDI for seasonal forecast purposes is introduced, since this predictor is in the group of most robust spatio-temporal associations. This can be an advantage since, in addition to representing the variability of tropical precipitation more skillfully than traditional indices, its non-dependence on convective parameterizations [30] can reduce possible biases originating from these dynamic approximations.

4. Conclusions

The analysis made on the different candidate predictors shows that, as a general average, the first mode of variability is dominant and responds to usual regime conditions imposed by the circulation of subtropical anticyclone, which indicates that this pattern dominates the spatio-temporal variability of precipitation in Cuba. Well-established seasonal structures such as the TUTT can be reflected in other modes but accompanied by random noise, which, combined with their limited permanence in time, leads to weak correlations.

Oscillations such as ENSO and QBO usually present a dominant phase in the analyzed period. In the case of ENSO, its relationships are weak, so its use as a predictor on monthly-time scales does not seem appropriate. Given that the reaction in the Caribbean region to equatorial Pacific SST anomalies may have a certain lag, an evaluation of this oscillation on annual time scales and/or applying a certain lag could generate different results. Regarding the QBO, the results coincide with the theoretical considerations, where its positive phase (westerlies flow) influences the variability of precipitation, with strong spatio-temporal associations.

The principal components analysis corresponding to tropical Atlantic SST, temperature at 850 hPa, zonal wind component at 30 and 200 hPa, the OLR and KI and GDI suggest that these candidate predictors most coherently explain the spatio-temporal variability of precipitation in Cuba and it could be used to build statistical or dynamic-statistical models to obtain seasonal rainfall forecasts.

In the particular case of the K index and temperature at 850 hPa, both candidates explain the moisture content in the lower troposphere and in fact one is contained in the formulation of the other. In order to avoid the risk of multicollinearity errors, it would be appropriate to use one or the other, but not both together.

Acknowledgements

The authors of this research are grateful for the support provided through the project "Building Resilience to Drought in Cuba".

This research was carried out thanks to a grant from the International Development Research Center (IDRC), Ottawa, Canada.

The opinions expressed do not necessarily represent those of IDRC or its Board of Governors.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Vitart, F., Buizza, R., Balmaseda, M.A., Balsamo, G., Bidlot, J.R., Bonet, A., Fuentes, M., Hofstadler, A., Molteni, F. and Palmer, T.N. (2008) The New VarEPS-Monthly

Forecasting System: A First Step towards Seamless Prediction. *Quarterly Journal of the Royal Meteorological Society*, **134**, 1789-1799. https://doi.org/10.1002/qj.322

[2] Molteni, F., Stockdale, T., Balmaseda, M., Balsamo, G., Buizza, R., Ferranti, L., Magnusson, L., Mogensen, K., Palmer, T. and Vitart, F. (2011) The New ECMWF Seasonal Forecast System (System 4). ECMWF, Reading.

[3] Johnson, S.J., Stockdale, T.N., Ferranti, L., Balmaseda, M.A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S.P.E., Mogensen, K., Zuo, H. and Monge-Sanz, B.M. (2019) SEAS5: The New ECMWF Seasonal Forecast System. *Geoscientific Model Development*, **12**, 1087-1117. https://doi.org/10.5194/gmd-12-1087-2019

[4] Robertson, A.W., Vitart, F. and Camargo, S.J. (2020) Subseasonal to Seasonal Prediction of Weather to Climate with Application to Tropical Cyclones. *Journal of Geophysical Research*: *Atmospheres*, **125**, e2018JD029375. https://doi.org/10.1029/2018JD029375

[5] World Meteorological Organization (2020) Guidance on Operational Practices for Objective Seasonal Forecasting.

[6] Barnston, A.G., Thiao, W. and Kumar, V. (1996) Long-Lead Forecasts of Seasonal Precipitation in Africa Using CCA. *Weather and Forecasting*, **11**, 506-520. https://doi.org/10.1175/1520-0434(1996)011<0506:LLFOSP>2.0.CO;2

[7] Ashby, S.A., Taylor, M.A. and Chen, A.A. (2005) Statistical Models for Predicting Rainfall in the Caribbean. *Theoretical and Applied Climatology*, **82**, 65-80. https://doi.org/10.1007/s00704-004-0118-8

[8] Amador, J.A. (2008) The Intra-Americas Sea Low-Level Jet. *Annals of the New York Academy of Sciences*, **1146**, 153-188. https://doi.org/10.1196/annals.1446.012

[9] Muñoz, E., Busalacchi, A.J., Nigam, S. and Ruiz-Barradas, A. (2008) Winter and Summer Structure of the Caribbean Low-Level Jet. *Journal of Climate*, **21**, 1260-1276. https://doi.org/10.1175/2007JCLI1855.1

[10] Duran-Quesada, A.M., Gimeno, L., Amador, J.A. and Nieto, R. (2010) Moisture Sources for Central America: Identification of Moisture Sources Using a Lagrangian Analysis Technique. *Journal of Geophysical Research*, **115**, D05103. https://doi.org/10.1029/2009JD012455

[11] García-Martínez, I.M. and Bollasina, M.A. (2020) Sub-Monthly Evolution of the Caribbean Low-Level Jet and Its Relationship with Regional Precipitation and Atmospheric Circulation. *Climate Dynamics*, **54**, 4423-4440. https://doi.org/10.1007/s00382-020-05237-y

[12] Martínez, C., Goddard, L., Kushnir, Y. and Ting, M.F. (2019) Seasonal Climatology and Dynamical Mechanisms of Rainfall in the Caribbean. *Climate Dynamics*, **53**, 825-846. https://doi.org/10.1007/s00382-019-04616-4

[13] Martínez, C. (2021) Seasonal Climatology, Variability, Characteristics, and Prediction of the Caribbean Rainfall Cycle. https://www.researchgate.net/publication/361266119_Seasonal_Climatology_Variability_Characteristics_and_Prediction_of_the_Caribbean_Rainfall_Cycle

[14] Wallace, J.M. and Gutzler, D.S. (1981) Teleconnections in the Geopotential Height Field during the Northern Hemisphere Winter. *Monthly Weather Review*, **109**, 784-812. https://doi.org/10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2

[15] Barnston, A.G. and Livezey, R.E. (1987) Classification, Seasonality and Persistence of Low-Frequency Atmospheric Circulation Patterns. *Monthly Weather Review*, **11**, 1083-1126. https://doi.org/10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2

[16] Thompson, D.W. and Wallace, J.M. (1998) The Arctic Oscillation Signature in the

Wintertime Geopotential Height and Temperature Fields. *Geophysical Research Letters*, **25**, 1297-1300. https://doi.org/10.1029/98GL00950

[17] Wang, C., Jia, Z., Yin, Z., Liu, F., Lu, G. and Zheng, J. (2021) Improving the Accuracy of Subseasonal Forecasting of China Precipitation with a Machine Learning Approach. *Frontiers in Earth Science*, **9**, Article 659310. https://doi.org/10.3389/feart.2021.659310

[18] Planos, E., Rivero, R. and Guevara, V. (2013) Impacto del Cambio Climático y medidas de adaptación en Cuba. Instituto de Meteorología. Agencia de Medio Ambiente, Ministerio de Ciencia Tecnología y Medio Ambiente, La Habana.

[19] Fonseca, C. (2005) Cambios en la posición e intensidad del anticiclón del Atlántico y modificación en el régimen de las lluvias en Cuba. *Revista Cubana de Meteorología*, **12**, 25-34.

[20] Pérez, R., Fonseca, C., Lapinel, B., González, C., Planos, E., Cutié, V. and Vega, R. (2009) Actualización del conocimiento sobre variaciones, cambios y tendencias del clima en Cuba (Segunda Comunicación de Cuba a la Convención Marco de Naciones Unidas Sobre Cambio Climático, 62), Informe de Resultado Científico. Instituto de Meteorología.

[21] Acosta, H. (2014) La Oscilación del Atlántico Norte y el comportamiento de la temporada invernal en la región occidental de Cuba. Tesis de Diploma, La Habana: Instituto Superior de Tecnologías y Ciencias Aplicadas, Universidad de la Habana 140 p.

[22] Cedeño, Y. (2015) Oscilación Ártica y frentes fríos en el occidente de Cuba. *Revista Cubana de Meteorología*, **21**, 91-102.

[23] Lecha, L. (2018) Biometeorological Forecasts for Health Surveillance and Prevention of Meteor-Tropic Effects. *International Journal of Biometeorology*, **62**, 741-771. https://doi.org/10.1007/s00484-017-1405-2

[24] Lecha-Estela, L.B., Sánchez-Suárez, L., Verdecia-Naranjo, Y., Soler-Torres, E. and Sánchez-Díaz, A. (2020) Variabilidad de los tipos de situaciones sinópticas influyentes sobre el occidente de Cuba. *Revista Cubana de Meteorología*, **26**. http://rcm.insmet.cu/index.php/rcm/article/view/507

[25] Laing, A. and Evans, J.L. (2016) Introduction to Tropical Meteorology: A Comprehensive Online and Print Textbook. 2nd Edition, University Corporation for Atmospheric Research, Boulder. https://www.meted.ucar.edu/tropical/textbook_2nd_edition/

[26] Cárdenas, P.A., Centella, A. and Naranjo, L. (1995) Teleconnection Pacific Caribbean ENSO and QBO as Forcing Climate Variability Elements. *Proceeding Sixth Internannual Meeting of Statistical Climatology*, Galway.

[27] Cárdenas, P.A. (1999) Pronósticos mensuales de lluvias en Cuba, un modelo con varios meses de adelanto. *Revista Cubana de Meteorología*, **6**, 47-51.

[28] Alvarez-Escudero, L., Mayor, Y.G., Borrajero-Montejo, I. and Bezanilla-Morlot, A. (2021) Assesing the Potential of a Long-Term Climate Forecast for Cuba Using the WRF Model. 4*th International Electronic Conference on Atmospheric Sciences*, 16-31 July 2021, 15 p. http://www.researchgate.net/publication/353321776_Assesing_the_Potential_of_a_Long-Term_Climate_Forecast_for_Cuba_Using_the_WRF_Model

[29] Alfaro, E.J., Chourio, X., Muñoz, Á.G. and Mason, S.J. (2018) Improved Seasonal Prediction Skill of Rainfall for the Primera Season in Central America. *International Journal of Climatology*, **38**, e255-e268. https://doi.org/10.1002/joc.5366

[30] Gálvez, J.M. and Davison, M. (2016) The Gálvez-Davison Index for Tropical Con-

vection. https://www.wpc.ncep.noaa.gov/international/gdi/GDI_Manuscript_V20161021.pdf

[31] Solano Ojeda, O., Vázquez Montenegro, R. and Martín Padrón, M.E. (2006) Estudio de la extensión superficial anual de la sequía agrícola en Cuba durante el período 1951-1990. *Revista Cubana de Meteorologí*, **13**, 41-52.

[32] Bretherton, C.S., Smith, C. and Wallace, J.M. (1992) An Intercomparison of Methods for Finding Coupled Patterns in Climate Data. *Cover Journal of Climate*, **5**, 541-560. https://doi.org/10.1175/1520-0442(1992)005<0541:AIOMFF>2.0.CO;2

[33] Fernández, M.A., González, C.M. and González, P. (2021) Spectral Changes in Some Meteorological Variables and It's Relationship with NAO and AO Index. *Revista Cubana de Meteorología*, **27**.

[34] Peña, L. and Valderá, N. (2022) Variability of the North Atlantic Subtropical High during the Months of November to April from 1950 to 2019. *Revista Cubana de Meteorología*, **28**.

[35] Muñoz, Á.G., Goddard, L., Robertson, A.W., Kushnir, Y. and Baethgen, W. (2015) Cross-Time Scale Interactions and Rainfall Extreme Events in Southeastern South America for the Austral Summer. Part I: Potential Predictors. *Journal of Climate*, **28**, 7894-7913. https://doi.org/10.1175/JCLI-D-14-00693.1

[36] Saenz, F., Hidalgo, H.G., Muñoz, Á.G., Alfaro, E.J., Amador, J.A. and Vazquez-Aguirre, J.L. (2022) Atmospheric Circulation Types Controlling Rainfall in the Central American Isthmus. *International Journal of Climatology*, **43**, 197-218. https://doi.org/10.1002/joc.7745

[37] Manzanas, R., Lucero, A., Weisheimer, A. and Gutiérrez, J.M. (2017) Can Bias Correction and Statistical Downscaling Methods Improve the Skill of Seasonal Precipitation Forecasts? *Climate Dynamics*, **50**, 1161-1176. https://doi.org/10.1007/s00382-017-3668-z