Downscaling for Rainfall Prediction in the Aburrá River Valley and its Supplying Watersheds in Antioquia (Colombia) Using Non-Homogeneous Hidden Markov Models

Reducción de escala para la predicción de lluvia en el Valle del Río Aburrá y sus cuencas abastecedoras en Antioquia (Colombia) mediante modelos ocultos de Markov no homogéneos

Redução da escala para a previsão de chuvas no Vale do Rio Aburrá e suas bacias abastecedoras em Antioquia (Colômbia) usando modelos ocultos não homogêneos de Markov

Diego F. Osorio Giraldo[1] ROR

Luis Fernando Carvajal-Serna[2] ROR

Julián D. Rojo Hernández[3] ROR

Resumen

Uno de los principales problemas al proyectar escenarios de cambio climático es la escala gruesa de los productos de los modelos de circulación global. En este artículo se emplea la metodología de *downscaling* estadístico para obtener series de precipitación a escala de estación en tres escenarios de cambio climático en la región del Río Aburrá y sus cuencas abastecedoras, a través de modelos ocultos de Markov no homogéneos, utilizando variables atmosféricas como el viento, la humedad relativa y la presión atmosférica como predictores de precipitación. Se proponen dos enfoques: un modelo anual y uno trimestral. Cada modelo es entrenado mediante estaciones de precipitación para obtener el mejor número de estados ocultos que representen adecuadamente la climatología de la zona a través del criterio BIC. Con base en la climatología representada en los estados, se obtuvieron resultados mucho mejores con el modelo trimestral, cuyas calibraciones se utilizaron para el *downscaling* de los tres escenarios de cambio climático. Se encontraron ligeras diferencias en los promedios mensuales, así como diferencias a escala de estación entre las distribuciones de probabilidad de lluvia diaria para los escenarios analizados, indicando alteraciones locales en la precipitación por efectos del cambio climático.

Palabras clave: escenarios de cambio climático, *downscaling*, modelos ocultos de Markov no homogéneos, modelos de circulación global

1. Magister en Ingeniería de los Recursos Hidráulicos, Universidad Nacional de Colombia, dfosoriogi@unal.edu.co
2. Universidad Nacional de Colombia, Sede Medellín, Facultad de Minas, Departamento de Geociencias y Medio Ambiente, lfcarvaj@unal.edu.co
3. PhD, Universidad Nacional de Colombia, PhMET, jdrojoh@unal.edu.co

Abstract

One of the main issues in projecting climate change scenarios is the coarse scale of global circulation model products. This work employs the statistical downscaling methodology to obtain station-level precipitation series for three climate change scenarios in the region of the Aburrá River and its supplying watersheds, by means of non-homogenous hidden Markov models, using atmospheric variables such as wind, relative humidity, and atmospheric pressure as precipitation predictors. Two approaches are proposed: an annual model and a quarterly one. Each model is trained using precipitation stations to obtain the best number of hidden states that adequately represent the climatology of the area via the BIC criterion. Based on the climatology represented in the states, much better results were obtained with the quarterly model, whose calibrations were used for downscaling the three climate change scenarios. Slight differences were found in the monthly averages, as well as station-scale differences between the probability distributions of daily rainfall for the analyzed scenarios, indicating local alterations in precipitation due to climate change.

Keywords: climate change scenarios, downscaling, non-homogeneous hidden Markov models, global circulation models

Resumo

Um dos principais problemas na projeção de cenários de mudança climática é a escala grosseira dos produtos dos modelos de circulação global. Neste artigo, utiliza-se a metodologia de *downscaling* estatístico para obter séries de precipitação em escala de estação em três cenários de mudança climática na região do Rio Aburrá e suas bacias abastecedoras, por meio de modelos ocultos de Markov não homogêneos, utilizando variáveis atmosféricas como vento, umidade relativa e pressão atmosférica como preditores de precipitação. São propostos dois enfoques: um modelo anual e um trimestral. Cada modelo é treinado com dados de estações de precipitação para identificar o melhor número de estados ocultos que representem adequadamente a climatologia da região de acordo com o critério BIC. Com base na climatologia representada nos estados, os melhores resultados foram obtidos com o modelo trimestral, cujas calibrações foram utilizadas para o *downscaling* dos três cenários de mudança climática. Foram encontradas leves diferenças nos valores médios mensais, bem como diferenças em escala de estação entre as distribuições de probabilidade da chuva diária nos cenários analisados, indicando alterações locais na precipitação em decorrência dos efeitos da mudança climática.

Palavras-chaves: cenários de mudanças climáticas, *downscaling*, modelos ocultos não homogêneos de Markov, modelos de circulação global

INTRODUCTION

Water, a lifeline for developing communities and a cornerstone of ecosystem conservation, is a resource of paramount importance. Understanding and predicting future water availability is a critical mission. Climate change, a key factor influencing water resources, is projected to alter spatial and temporal rainfall patterns in specific locations (IPCC, 2014), with both small- and large-scale implications. As highlighted in Guiza et al. (2020), Rojo et al. (2020), Rojo and Mesa (2018), and numerous other studies on the impact of climate change on Colombia's water resources, the outcomes regarding water balance vary by region, with a direct effect on the water supply.

The most advanced tools currently used to simulate the response of climate systems to the increase in greenhouse gas (GHG) emissions are global circulation models (GCMs). These dynamical models represent the planet's physical processes and can be simulated under different forcings such as gas emissions. The IPCC's

projection includes representative concentration pathways (RCPs), which outline the possible trajectories of GHG concentration and radiative forcing for the year 2100. These RCPs range from a low-emissions scenario with strong mitigation efforts (RCP 2.6) to a high-emissions scenario without additional climate policies (RCP 8.5), including intermediate pathways such as RCP 4.5 and RCP 6.0. These scenarios allow assessing climate change impacts under different socioeconomic and development assumptions. However, due to the coarse spatial scale of their products, the regional and local use of GCMs is not appropriate. Instead, mathematical techniques such as downscaling can be used, which transforms coarse data into a more detailed scale for use at these levels. Statistical downscaling methods are helpful if the historical record is sufficient for generating probability distribution functions and defining statistical relationships.

Weather generators, one form of statistical downscaling, constitute a versatile tool. These stochastic models can generate synthetic climate time series for specific locations, mirroring the statistical properties of observed time series (Yin & Chen, 2020). While they are commonly used for the synthetic reproduction of precipitation series, they can also simulate other variables such as temperature, humidity, solar radiation, and cloud cover. The adaptability of these methods is evident in their ability to accommodate different predictor variables related to the study objectives (*e.g.*, using wind and atmospheric pressure fields as precipitation predictors), as well as in their representation of spatial variability.

The statistical downscaling method known as *non-homogeneous hidden Markov models* (NHMMs) (Hughes & Guttorp, 1994) has been widely used for precipitation modeling (Bellone *et al.*, 2000; Zhang *et al.*, 2020; Rojo *et al.*, 2020). This technique captures daily statistics from precipitation stations and associates them with synoptic variables generated by GCMs. NHMMs decompose spatiotemporal rainfall variability through several weather patterns known as *hidden states*. These states represent the probability of rainfall at each station, describing how precipitation occurs in a region. Defining the states and the transition probabilities between the hidden states allows using GCM projections for predicting rainfall at the station level. This enables the use of precipitation data in hydrological models to obtain more accurate and realistic results of climate change scenarios at larger watershed scales, which can help to improve climate change impact assessments. The objectives of this study included the downscaling of selected GCM outputs in order to obtain daily rainfall series across different climate change scenarios.

Precipitation downscaling has been a key concern in regional climate studies, especially in South America, given the need to adapt global projections to scales useful for environmental management and territorial planning. In this context, Chou *et al.* (2012) stand out for applying dynamic downscaling to represent the present climate of South America based on multiple simulations from the HadCM3 model. Meanwhile, statistical methods such as NHMMs have successfully improved the spatial and temporal resolution of climate projections. These approaches have been implemented in various regions across the continent, including northeastern Brazil (Robertson *et al.*, 2004), Ecuador and Peru (Pineda & Willems, 2016), and northern Chile (Verbist *et al.*, 2010), as well as in broader studies such as those by Fu *et al.* (2013) and Johnson *et al.* (2014), which focused on applying NHMMs to tropical and subtropical contexts.

In Colombia, downscaling efforts have been more limited but are gaining relevance. Arias *et al.* (2021) characterized the country's climatology using projections from the CMIP5 and CMIP6 models, revealing persistent biases and some improvements over previous versions. More specific downscaling studies include those conducted by Posada-Marín *et al.* (2019), who used reanalysis data to enhance the spatial representation of precipitation, and the work by López López *et al.* (2018), who applied spatial downscaling techniques to satellite-based precipitation data in order to assess their impact on hydrological simulations in the Magdalena River basin. These efforts represent significant progress towards generating locally relevant climate information for water resource management and climate change adaptation in Colombia.

This work presents the downscaling of a GCM for the region of the Aburrá River basin at the rain gauge station level. Here, NHMMs are employed to obtain precipitation series for the RCPs and use them as input for hydrological modeling in the Valley of Aburrá. First, the methodology and the information used throughout the research are presented. Afterwards, the results are outlined by means of metrics of efficiency and a graphical analysis of the outcomes of each step in the methodology. In the final section, some concluding remarks are provided.

METHODOLOGY

Study area

The study area is the Aburrá River basin, located in the department of Antioquia, Colombia. This area features a large population, which has evolved along the Aburrá River and its tributaries, with water being a vital resource for its development. Due to the different economic activities carried out in the region, the water produced by the basin is not enough to supply the demand. Therefore, water must be imported from another basin. This area was selected with the aim of assessing its future vulnerability to water shortages caused by climate, demographic, and land use changes.

The study area boasts a comprehensive network of 33 rainfall gauges strategically located to cover the entire region. Figure 1 illustrates the location of the Aburrá River basin, its supplying watersheds, and the precise locations of the rain gauge stations. The daily precipitation data, meticulously collected and spanning from 1970 to 2019, provide a robust foundation for our research.
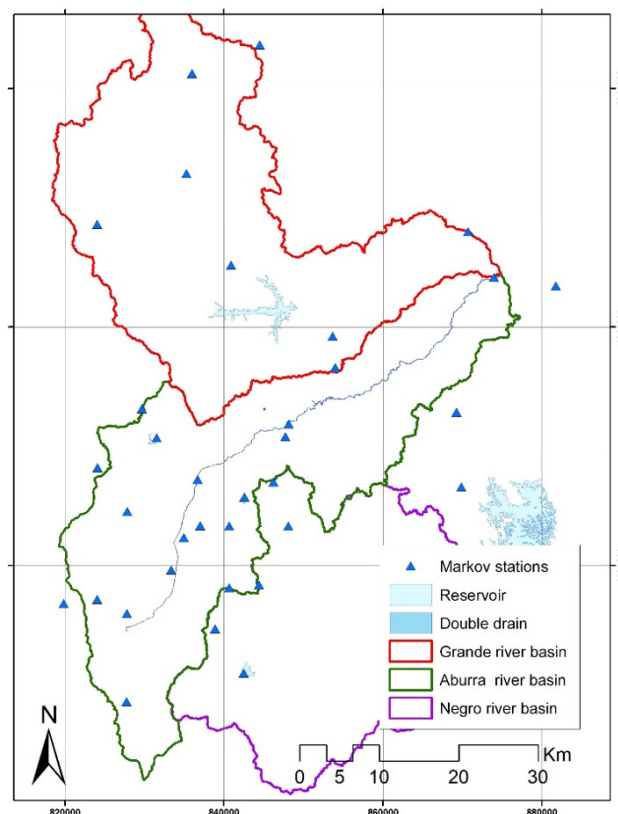

Figure 1. *Study area and rainfall gauge stations*

Non-homogeneous hidden Markov models

In this research, NHHMMs (Liu *et al.*, 2013) assume the existence of 'hidden' stochastic processes in the climate system, which can be summarized as a discrete number of states. These models are non-homogeneous since the transition matrix depends on the values of some atmospheric variables. For instance, in this study, the daily precipitation series Y_t depends on the state of a hidden (or unobserved) Markov process Z_t. Thus, the amount of precipitation at a given location y_t is explained by a set of explanatory variables (predictors) X_t, with $x_t = (x_{1t}, \ldots x_{kt})$. Therefore, with a given z_t, the precipitation series y_t is predicted as $y_t = g(z_t)$, with g being a predetermined function. The hidden process follows a first-order Markov process with *m* states, and the transition probabilities are estimated via Equation (1).

$$P(Z_{t+1} = j | Z_t = i) = p_{ij}, i, j = 1, \ldots., m \tag{1}$$

The following transition probability matrix determines the hidden process at a given time *t*:

$$P = P^{(T)} \begin{bmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & \ddots & \vdots \\ p_{m1} & \cdots & p_{mm} \end{bmatrix} \tag{2}$$

Therefore, the transition matrix, an array that summarizes the probabilities of passing from one state to another or remaining in the same state, varies over time with the atmospheric variables. Figure 2 shows the scheme of a NHMM (Koki, 2022), where the predictors are connected to the transition matrix and, depending on the state, the value of the predicted variable is generated. The model's foundations and parameterization can be consulted in Bellone *et al.* (2000).
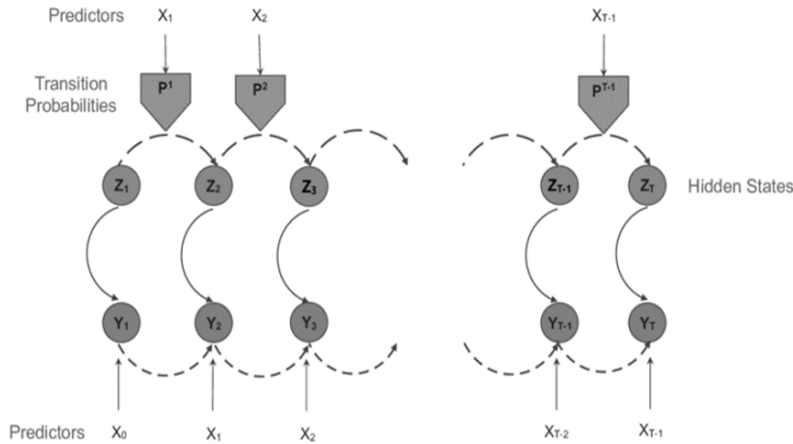

Figure 2. *NHMM scheme*
Source: *Adapted from Koki (2022)*

Several studies have observed that hidden precipitation states depend on observable synoptic variables such as atmospheric pressure, omega, and geopotential heights, among others (Bellone *et al.*, 2000; Liu *et al.*, 2013).

Training a homogenous hidden Markov model is a thorough, meticulous, and rigorous process that involves working with the data to be modeled —in this case, precipitation series— in order to determine the ideal number of states. The BIC (Bayesian information criterion) parameter is used to determine this number; the model with the lowest BIC should be selected. In this work, the physical validity of the selected states was rigorously tested by representing their atmospheric fields, using ERA-INTERIM reanalysis data for Colombia. Different atmospheric fields that could influence the variables of interest were selected and rescaled for use as predictors in the non-homogeneous model.

The method selected to summarize the atmospheric fields of the models was singular value decomposition (SVD) ([Bretherton *et al.*, 1992](#)), which consists of decomposing the correlation matrix between the atmospheric field and the variable to be downscaled (*i.e.*, precipitation), resulting in three matrices: two with singular vectors for each variable and a diagonal one containing the covariance explained by each pair of vectors. Mathematically, given two anomaly matrices, $\mathbf{X}$ (*e.g.*, standardized atmospheric predictor fields) of size $m \times n$ and $\mathbf{Y}$ (*e.g.*, standardized precipitation values) of size $p \times n$, where n is the number of time steps, the cross-covariance matrix $\mathbf{C} = \mathbf{X} \cdot \mathbf{Y}^t$ can be constructed. The SVD of this matrix is given by

$$C = U\acute{O}V^T \tag{3}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices, whose columns are the left and right singular vectors (also referred to as *coupled patterns* or *modes*) of the atmospheric field and precipitation, respectively. In addition, Σ is a diagonal matrix that contains the singular values representing the square roots of the eigenvalues of $C \cdot C^t$ and quantifying the strength of the coupled variability between the two fields.

Each pair of singular vectors defines a mode of covariability, with the corresponding singular value indicating the proportion of shared variance captured by said mode. After the decomposition, the atmospheric field is projected onto the leading singular vector of precipitation by multiplying the standardized predictor matrix $\mathbf{X}$ by the corresponding right singular vector from $\mathbf{V}$, generating a time series of expansion coefficients or principal components. These coefficients can then be used as predictors in downscaling models, as they summarize the most relevant large-scale atmospheric patterns associated with local precipitation variability ([IPCC, 2014](#); [Güiza-Villa *et al.*, 2020](#); [Yin *et al.*, 2020](#)).

The model's adaptability to different scenarios is a key feature. In this vein, different combinations of previously processed atmospheric fields were performed, and the number of selected states was used to calibrate the best non-homogeneous model according to the BIC. The most relevant atmospheric fields were considered, such as wind (zonal and meridional), relative humidity, geopotential height at 850 hPa, and sea-level pressure. The future precipitation was simulated using the IPCC scenarios for different GCMs as well as the best combination of predictor variables and number of hidden states.

Due to the uncertainties introduced by the GCMs and the errors caused by downscaling, bias correction for the estimation errors was necessary. This process corrects the systematic errors in the observations for a given time. Thus, the method known as *quantile mapping* ([Wilcke, 2013](#)) was used, which consists of estimating the empirical cumulative distribution function (ECDF) of the modeled and observed data and, based on this information, determining the value of the modeled precipitation at a given instant, *i.e.*,

$$p_{m,t} = f_o^{-1}\left\{ f_m\left(p_{o,t} \right) \right\} \tag{4}$$

where $p_{m,t}$ is the modeled precipitation at instant t; $p_{o,t}$ is the observed precipitation at t; and f_m and f_O are the ECDFs for the modeled and observed data, respectively. Different metrics have been proposed to evaluate the performance of the Markov-type models and determine the validity of rescaling the data with the observed variables. These include correlation analysis between the observed and generated series, comparisons between annual and monthly precipitation, and the generation of dry and wet days, which are variables of special importance in estimating the water supply.

Downscaling

Following Cadavid Valencia (2015), the GCM selected for downscaling was NORES-M1M from Norway, which stands out among the models that best represent Colombia's climatology and atmospheric dynamics. To apply the NHHMM, 35 precipitation series for the Valley of Aburrá and its supplying watersheds were used, along with a daily time series covering the 1970-2019 period. The rain gauge stations were used for model calibration and for defining the ideal number of states in the Markov model. Before using the series for the model, the missing values were completed using a pondered mean with the data of the best-correlated stations. ERA-INTERIM reanalysis was used to construct the atmospheric fields of the hidden states and verify the coherence of the hidden atmospheric states obtained after model calibration. The following variables were used: zonal and meridional wind, relative humidity, geopotential height, and sea-level pressure (IPCC, 2014; Güiza-Villa *et al.*, 2020).

RESULTS

Two approaches were used to calibrate the model: an annual model and a quarterly one defined by the annual cycle of precipitation in the area (Bellone *et al.*, 2000; Greene *et al.*, 2011; Liu *et al.*, 2013; Robertson *et al.*, 2004). The selected quarters were December, January, February (DJF); March, April, May (MAM); June, July, August (JJA); and September, October, November (SON). Different runs of the model were executed, varying the number of states until the lowest BIC was obtained. Table 1 presents the BIC for each run. Note that, for the quarters, a minimum was reached. However, the annual model calibration showed the BIC kept decreasing in steps of less than 1% between values. In Rojo (2018), the annual model was calibrated for the entirety of the Colombian territory, finding that the best number of states was 5. This number was also used in this research.

Table 1. *BIC for the number of states*

States number	ANUAL	DJF	MAM	JJA	SON
4	1.8208E+06	3.1239E+05	4.9503E+05	4.5461E+05	5.4371E+05
5	1.8146E+06	3.1208E+05	4.9445E+05	4.5354E+05	5.4296E+05
6	1.8091E+06	3.1262E+05	4.9389E+05	4.5252E+05	5.4303E+05
7	1.8044E+06	-	4.9398E+05	4.5253E+05	5.4318E+05
8	1.8008E+06	-	-	-	-

To observe the physical representation of each of the states, maps were generated with their station-level precipitation probability. Figure 3 shows the map for the states of the annual model (5), where three precipitation patterns can be identified: one of low precipitation (state 1), another of high precipitation (states 0 and 4), and a third of intermediate precipitation between repeated patterns (states 2 and 3). State 0 implies precipitation all over the area, while state 4 entails lower precipitation probabilities at some stations. The intermediate states (2 and 4) exhibit a lower precipitation probability in the center and south of the Valley of Aburrá.
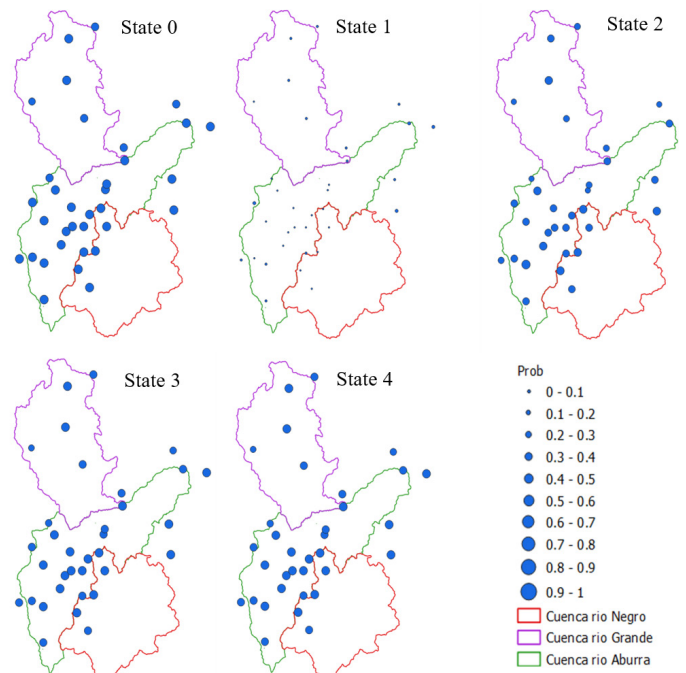

Figure 3. *Precipitation states – annual model*

In the quarterly model, a general pattern of high, intermediate, and low rainfall was maintained, with variations in the spatial distribution of the rainfall probability, as can be observed in Figure 4 for DJF, Figure 5 for MAM, Figure 6 for JJA, and Figure 7 for SON. Table 2 shows the main characteristics of the states for the DJF quarterly model (Figure 4).
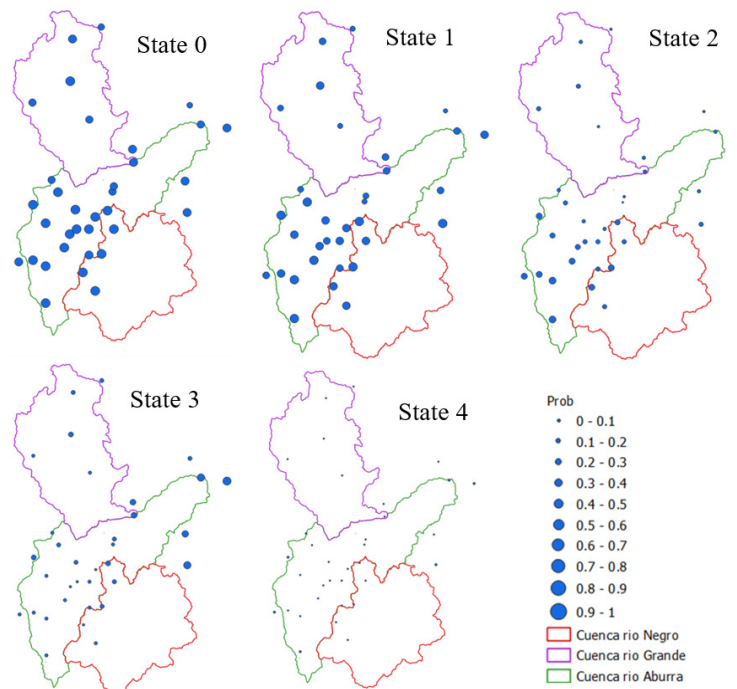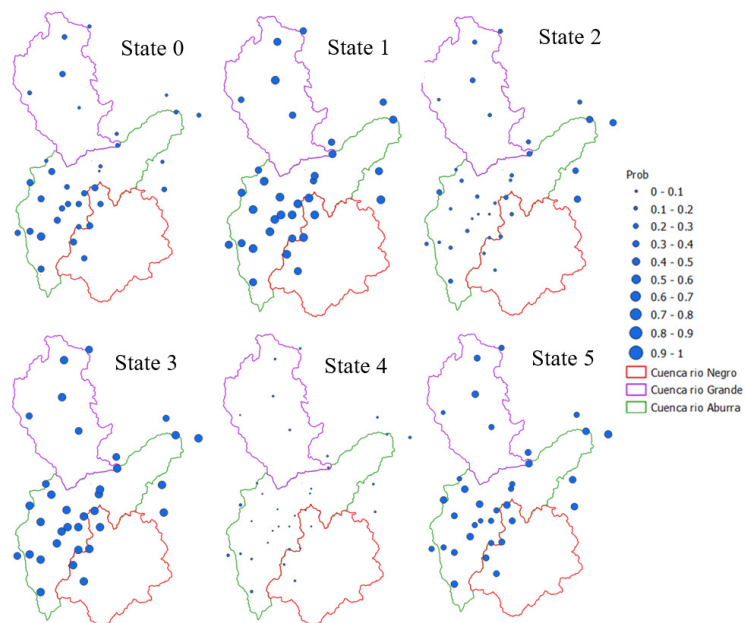
Figure 4. *Precipitation states – DJF quarter*
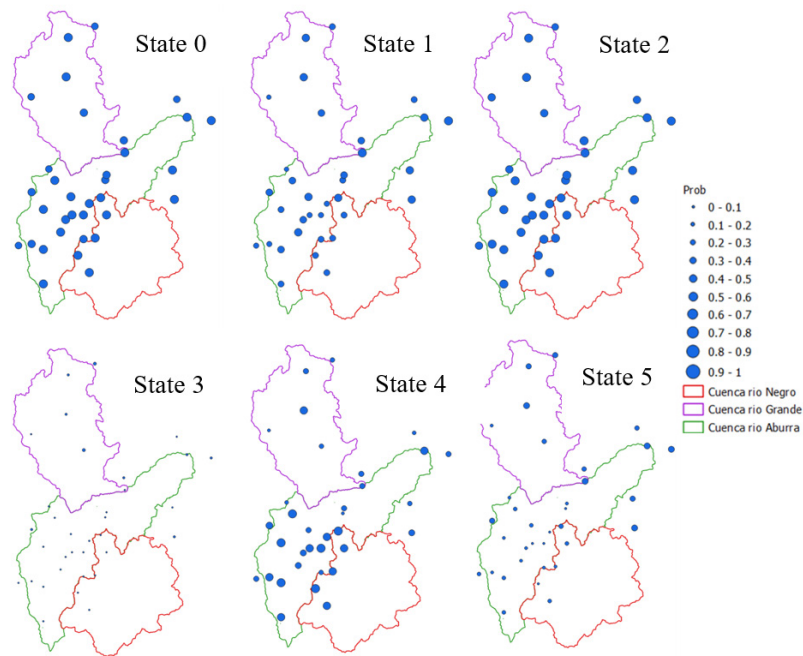

Figure 5. *Precipitation states – MAM quarter*
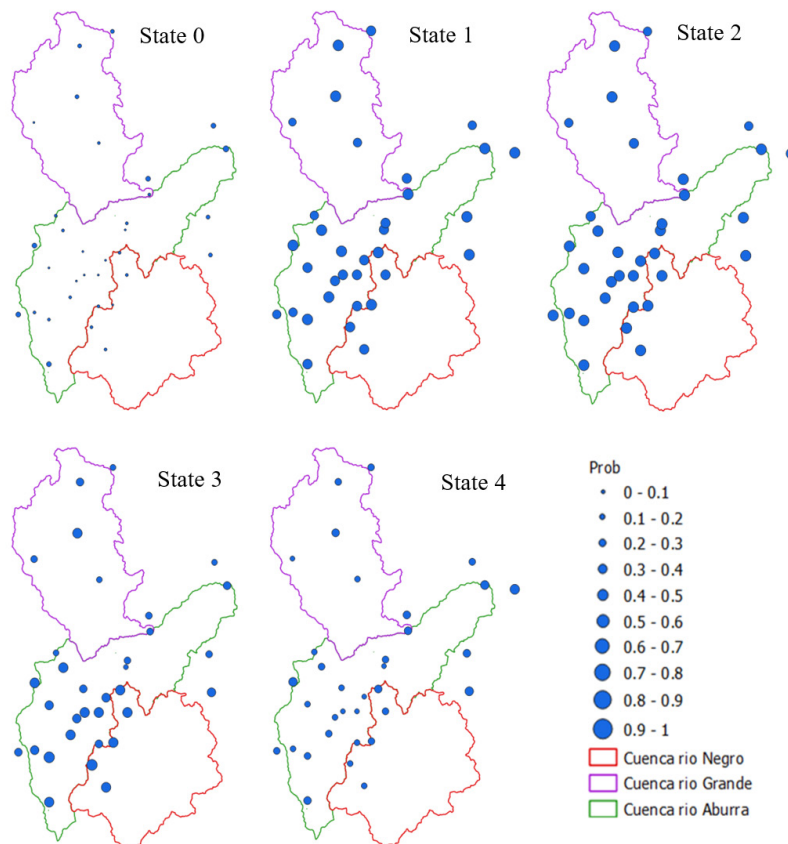
**Figure 6.** *Precipitation states – JJA quarter*

Figure 7. *Precipitation states – SON quarter*

Table 2. *Precipitation states for the DEF and MA quarterly models*

State	DEF	MAM	JJA	SON
0	High precipitation in all areas	Predominantly intermediate precipitation in the south of the area	High precipitation with local variations	Low probabilities for the whole area
1	High precipitation in the south and center of the Valley of Aburrá, except in the north	High precipitation throughout the area	Intermediate precipitation with higher magnitudes towards the northeast of the area	High precipitation, with lower magnitudes at some stations
2	Intermediate precipitation with higher probability in the south of the valley	Intermediate precipitation with higher magnitudes in the northeast of the area	High precipitation throughout the area	High precipitation throughout the area
3	Intermediate precipitation is more likely in the north of the valley and the Río Grande Basin	High precipitation throughout the area, with local variations	Low probabilities for the whole area	Intermediate precipitation with higher magnitudes in the south of the area
4	Low probabilities for the whole area	Low probabilities for the whole area	Intermediate precipitation with higher magnitudes on the edges of the valley	Intermediate precipitation with higher probabilities in the northeast of the valley
5	Not applicable	Intermediate precipitation with magnitudes higher than those of state 3	Intermediate precipitation with less precipitation in the south of the area	Not applicable

We estimated the atmospheric fields of different variables within the different states, in order to validate the physical consistency of the hidden states obtained. ERA-INTERIM reanalysis was used to obtain the following variables: zonal and meridional wind (WD), relative humidity (HR), geopotential height at 850 hPa (ZG), and sea-level pressure (PR). Figures 8 to 12 present the wind patterns for the states found.
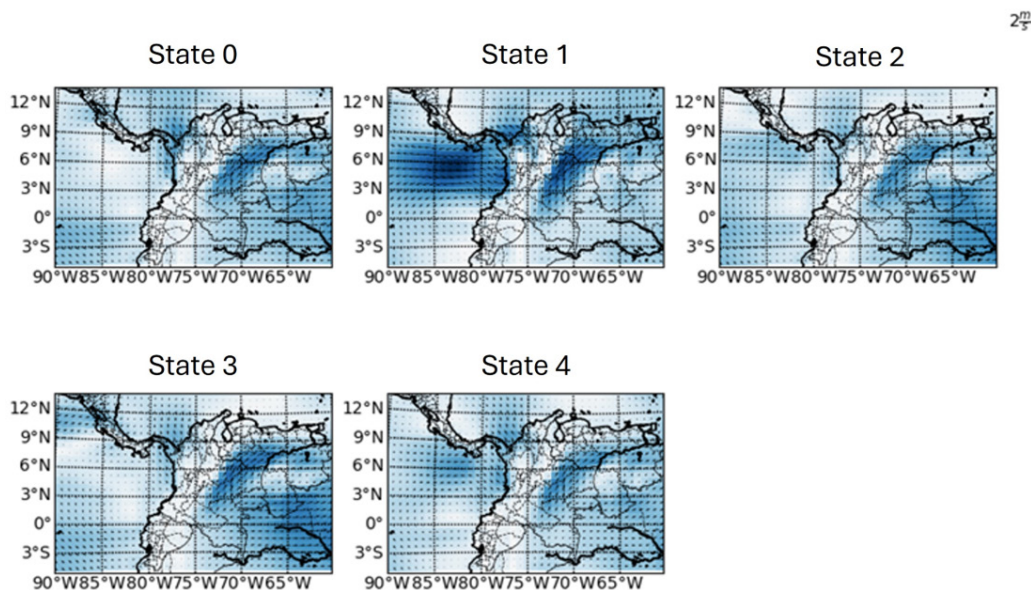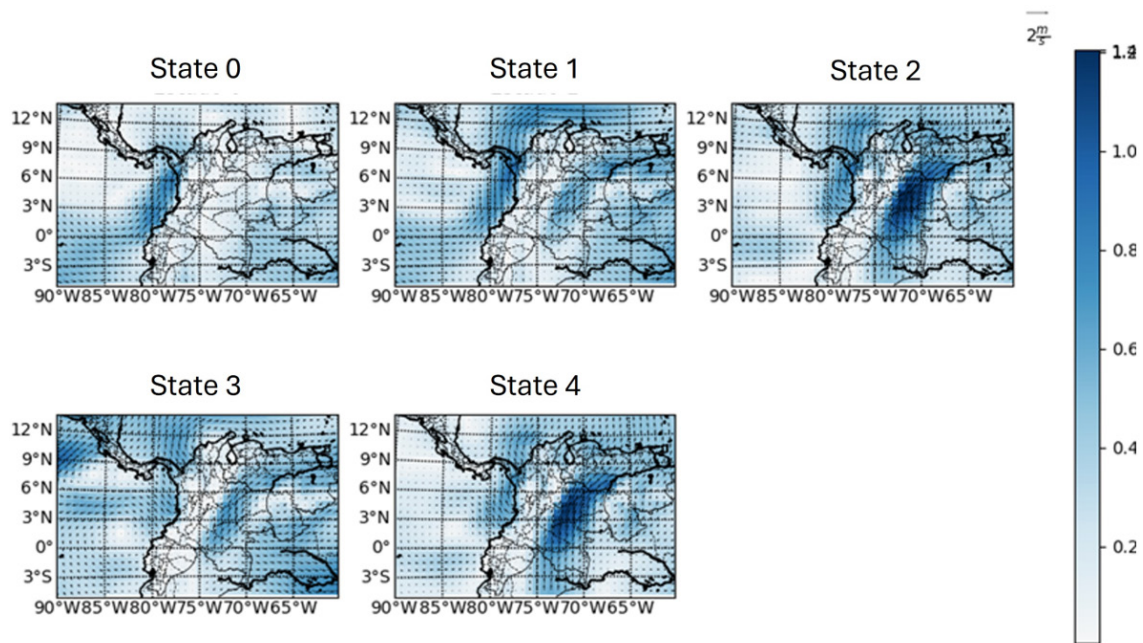

Figure 8. *Wind anomaly for the annual Markov model*

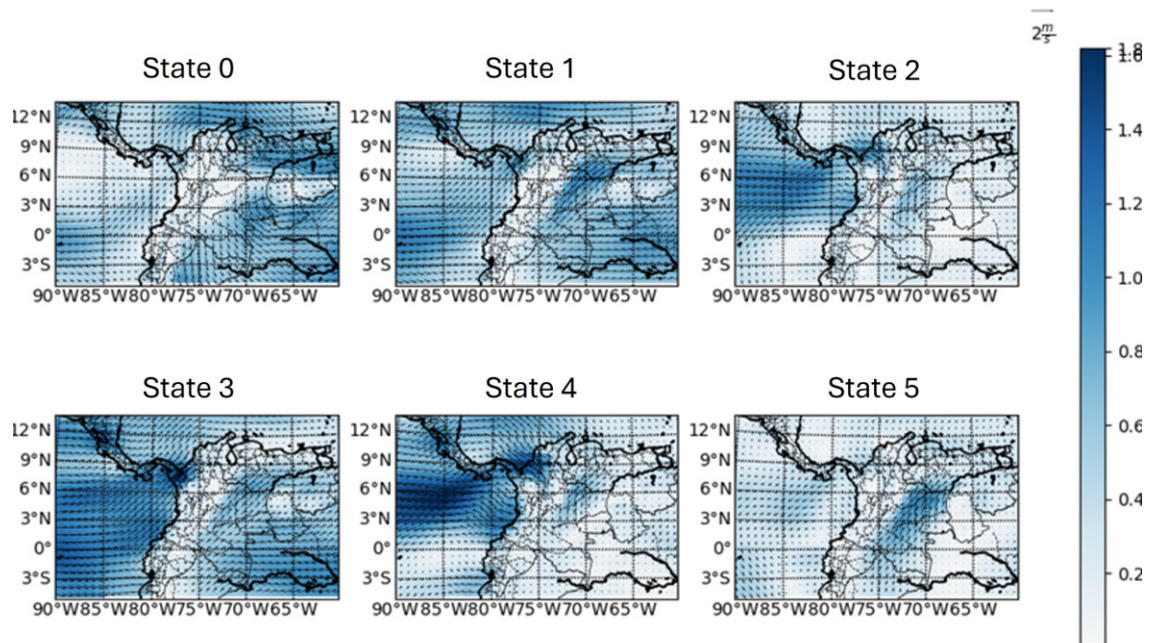**Figure 9.** *Wind anomaly for the DJF quarter*

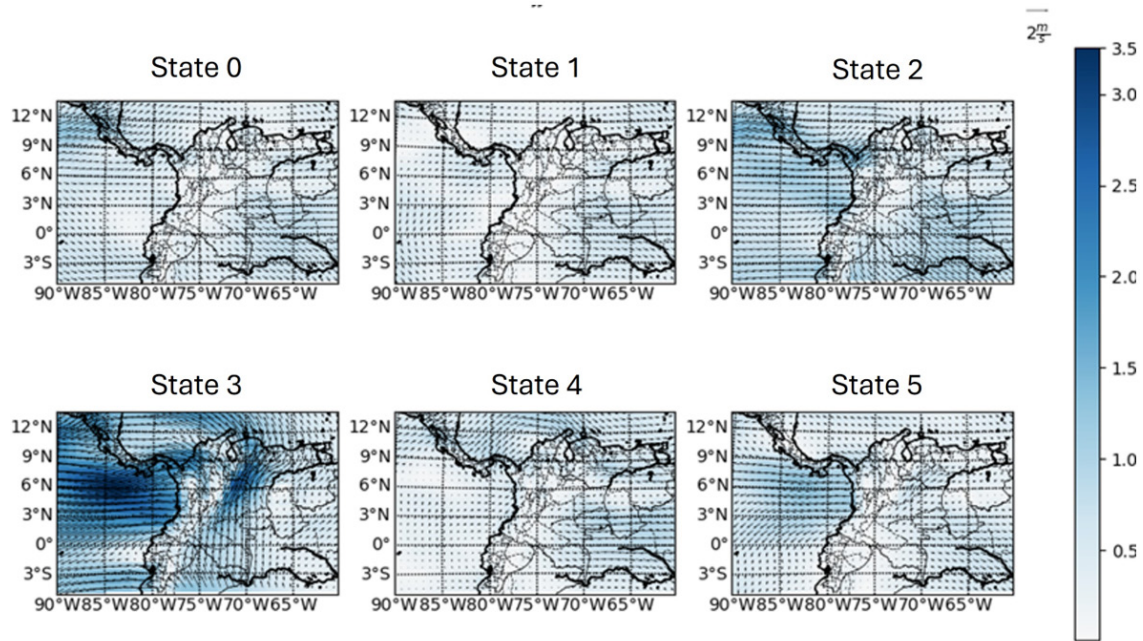**Figure 10.** *Wind anomaly for the MAM quarter*

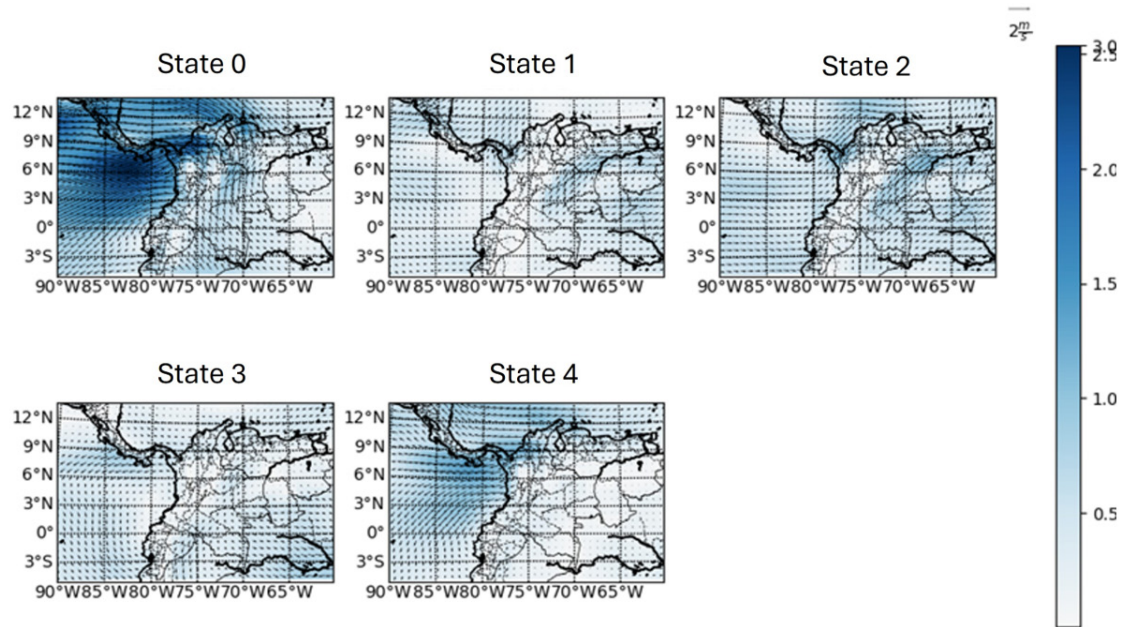**Figure 11.** *Wind anomaly for the JJA quarter*

Figure 12. *Wind anomaly for the SON quarter*

Some inconsistencies were found while analyzing the wind patterns for the annual model, such as the low magnitude of the winds in state 3, which represents high precipitation probabilities. Furthermore, for state 4, which is also characterized by high precipitation, there is also a wind pattern, showing moisture advection towards the center of the country. However, for state 2 (low precipitation), it can be seen how the winds diverge from the Colombian territory, which would mean a negligible moisture advection.

In contrast, regarding the quarterly model, the conditions agree more with the climatology of Colombia and the states found, as can be seen for the first two states. Moisture advection is observed, especially towards Antioquia from the Pacific, while, for state 5 (dry), there is a divergence of winds. For the MAM quarter, similar conditions are observed, *i.e.*, divergence for dry states and advection for humid ones, especially for states 5 and 6, which show no advection from the Pacific —however, there is indeed advection from the Caribbean Sea. As for the JJA quarter, in intermediate states such as 6, there is a wind pattern from the Caribbean towards Colombia. Finally, the wet states of the SON quarter exhibit simultaneous advection from the Pacific and the Caribbean, as well as westward winds from Colombia. Based on the above, it can be confirmed that the number of states represents some features of the circulation patterns that determine rainfall in the study area.

After defining the number of states, we meticulously calibrated the NHMM by introducing the atmospheric variables. This process, crucial for the model, required reducing the size of the atmospheric fields. To this effect, SVD was employed, a detailed account of which is provided in the methodology section. This method effectively captured the greater variability between the precipitation stations and the atmospheric fields, instilling confidence in the robustness of our methodology.

Once again, we used the BIC to determine the variables for the simulation, combining the atmospheric variables with the defined number of states. Table 3 shows the minimum BIC for each model.

Table 3. BIC criterion for atmospheric variables

BIC	Annual	DEF	MAM	JJA	SON
WD	1.7927E+06	310 885	493 891	452 576	542 984
HR	1.7928E+06	310 879	493 896	452 578	543 005
ZG	1.7928E+06	310 884	493 904	452 577	543 017
PR	1.7928E+06	310 884	493 873	452 578	542 999
WD, HR	1.7928E+06	310 909	493 908	452 617	543 012
WD, ZG	1.7929E+06	310 912	493 900	452 624	543 022
WD, PR	1.7929E+06	310 916	493 906	452 583	543 012
HR, ZG	1.7929E+06	310 910	493 918	452 609	543 037
HR, PR	1.7930E+06	310 903	493 925	452 618	543 892
Zg, PR	1.7930E+06	310 912	493 934	452 608	543 276
WD, HR Zg	1.7930E+06	310 942	493 943	452 647	543 215
All	1.7930E+06	310 967	493 952	452 686	543 150

We performed precipitation simulations for the annual cycles of each Markov model after defining the atmospheric variables. The tool generated the same number of stations and lengths as the data used for calibration. The annual cycles and the total annual precipitation were estimated for the annual and quarterly models. Figures 13 to 16 show the validation measures, using observed (obs), simulated (sim), and simulated corrected (bc) data to determine the validity of the calibration.
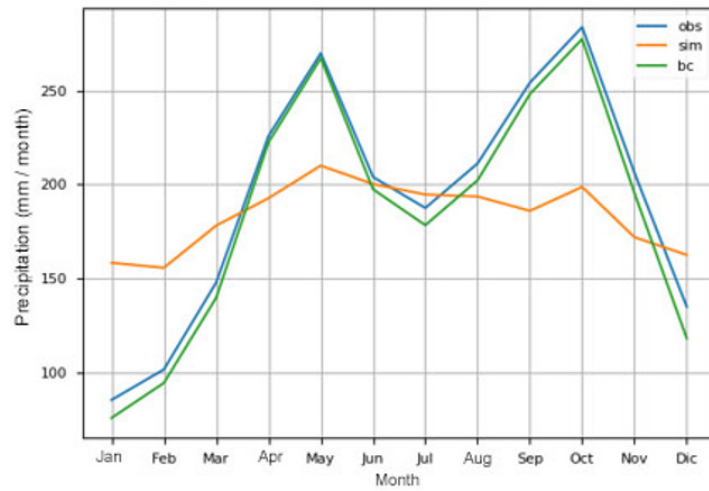
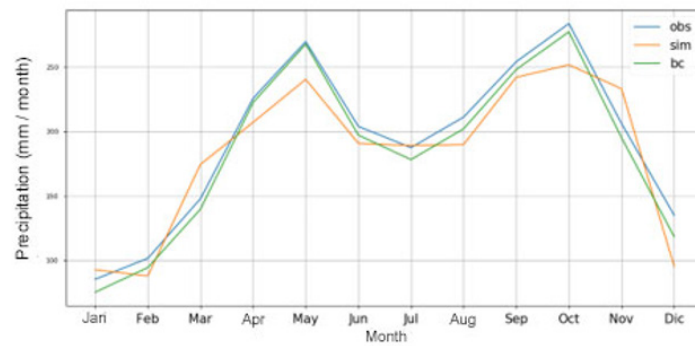**Figure 13.** *Monthly precipitation mean – annual model*

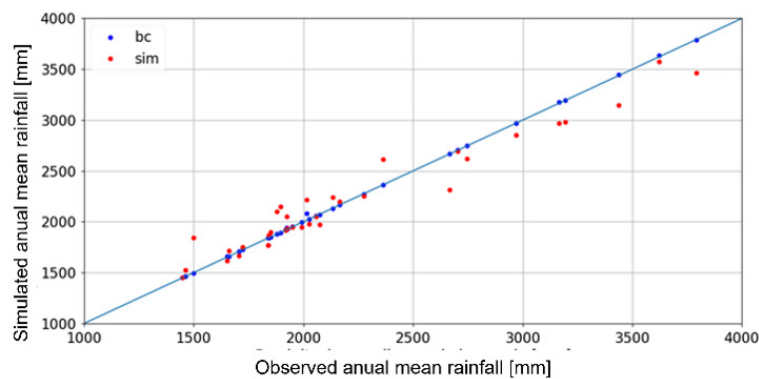**Figure 14.** *Monthly precipitation mean – quarterly model*

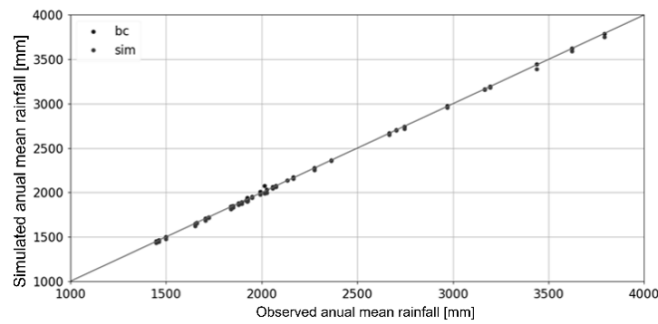**Figure 15.** *Annual precipitation by station for the annual model*

Figure 16. *Annual precipitation by station for the quarterly model*

Figures 13 and 14 show the simulated and observed monthly average precipitation. Note that the quarterly model provides better estimates without quantile mapping correction, in comparison with the annual model's simulated precipitation without bias correction. The simulated and observed annual precipitations of each station are presented in Figures 15 and 16. Here, the annual precipitation without correction is more dispersed in the annual model, while that of the quarterly model practically matches the corrected values.

Finally, downscaling was meticulously performed for the atmospheric variables of each selected GCM, as projected for RCPs 2.6, 4.5, and 8.5. The variable with the lowest BIC for each quarter was used (Table 3). Figure 17 presents the multiannual monthly average of the rescaled series obtained for the GCM scenarios, and Figure 18 shows the average corrected through quantile mapping, reinforcing the thoroughness of our methodology.
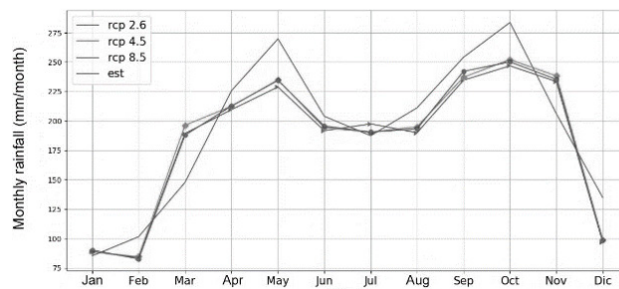

Figure 17. *Multiannual monthly average of the rescaled series obtained for the GCM scenarios (RCP 2.5, 4.5, and 8,5) and the stations (est)*
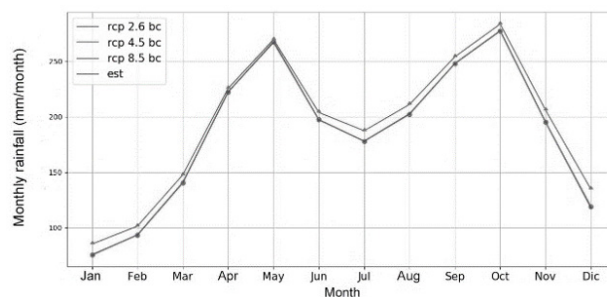

Figure 18. *Bias-corrected monthly mean precipitation for the RCPs, as obtained through quantile mapping, and monthly mean precipitation for the stations (est)*

As shown in [Figures 17](#) and [18](#), the annual average exhibits minimal variability. However, some changes in precipitation can be noted by observing the distributions of some of the stations ([Figure 19](#)), where there is a general trend for scenario RCP 2.6. RCP 8.5 also shows a more extensive rainfall distribution —and in greater magnitudes— than the other scenarios, except for the Olaya Herrera station, where the lowest precipitation corresponds to RCP 8.5. Moreover, RCP 4.5 reports intermediate rainfall magnitudes, unlike La Fe station. This could indicate that the effect of each scenario is local rather than regional.
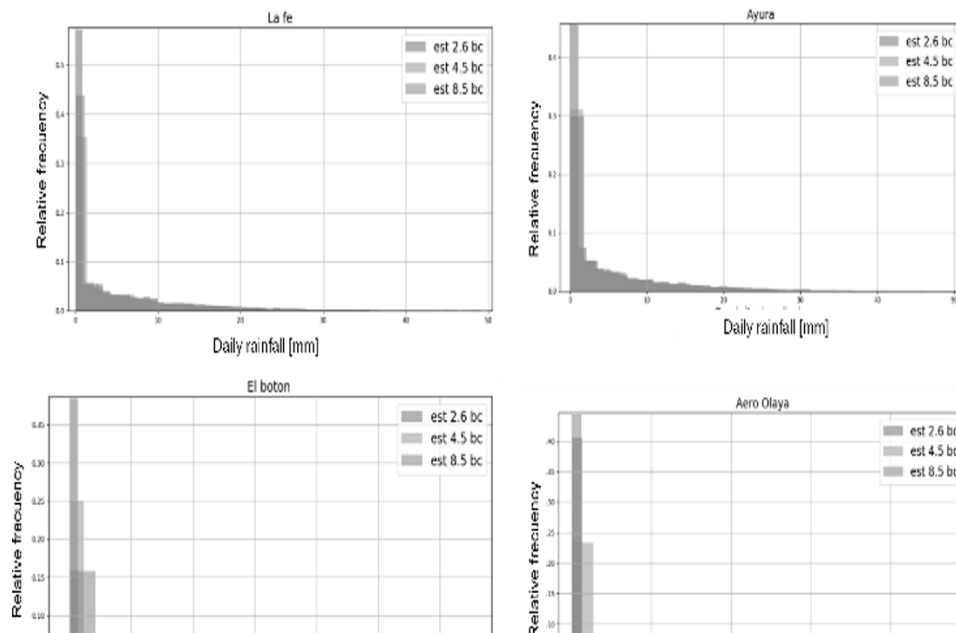

Figure 19. *Precipitation probability distribution in different future scenarios*

CONCLUSIONS

In this work, a NHHMM was applied to the output of a GCM for the Valley of Aburrá and its supplying watersheds in order to generate future station-level daily precipitation forecasts under three different climate change scenarios. This information is valuable for assessing the potential impacts of climate change in the study area. The results indicate that the effects of climate change are not uniform across the region; instead, changes in precipitation distribution vary across different zones and basins.

Both modeling approaches —annual and quarterly— exhibited differences during the validation process, suggesting that the NHHMM can provide reliable predictions even with limited input data. The statistical downscaling method used in this research proved to be a practical and accessible tool for projecting future precipitation under climate forcing scenarios.

The integration of atmospheric fields through SVD also proved to be a valuable enhancement to the statistical downscaling methodology. By identifying the dominant modes of variability between precipitation and large-scale atmospheric patterns, SVD allowed selecting physically meaningful predictors. This approach optimized model calibration and ensured that the hidden states within the NHHMMs aligned with known moisture advection patterns and circulation anomalies in the region. As a result, the model gained greater physical interpretability and robustness in simulating future rainfall.

A comparison between the annual and quarterly models revealed a clear advantage in using the quarterly approach to downscaling precipitation. The quarterly model's seasonal structure enabled a more accurate representation of the intra-annual variability in rainfall, aligning better with observed climatological patterns. Moreover, the quarterly model better fit the observed data without bias correction, highlighting its potential for producing more reliable hydrological inputs under varying climate change scenarios.

An analysis of future rainfall distributions under different RCPs demonstrated a significant degree of spatial heterogeneity across the study area. The magnitude and direction of the changes in daily precipitation varied from station to station, with some locations exhibiting greater sensitivity to low-emission scenarios and others responding more strongly to high-emission projections. This localized behavior suggests that the impacts of climate change on precipitation are modulated by micro-scale factors such as orography, land use, and basin orientation, underscoring the importance of fine-scale modeling for water resource planning and adaptation strategies.

Notably, the quarterly model successfully captured the spatial and temporal variability of rainfall in the study area while retaining key features of the dominant precipitation modes and the spatial patterns of atmospheric variables that drive rainfall in the region. These findings highlight the potential for future research to explore and deepen our understanding of these atmospheric configurations and their influence on regional hydrological behavior.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

Diego F Osorio: data curator, conceptualization, investigation, software, visualization, writing original draft, review.
Luis F Carvajal-Serna: data, conceptualization, investigation, writing original draft, review and edition.
Julian D. Rojo-Hernandez: conceptualization, investigation, writing original draft, review.

REFERENCES

Arias, P. A., Ortega, G., Villegas, L. D., & Martínez, J. A. (2021). Colombian climatology in CMIP5/CMIP6 models: Persistent biases and improvements. *Revista Facultad de Ingeniería Universidad de Antioquia, 100*, 75-96. https://bibliotecadigital.udea.edu.co/entities/publication/0c68b277-9463-400b-a6c1-a81c028e1f0d

Bellone, E., Hughes, J. P., & Guttorp, P. (2000). A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts. *Climate Research, 15*(1), 1-12. https://doi.org/10.3354/cr015001

Bretherton, C. S., Smith, C., & Wallace, J. M. (1992). An intercomparison of methods for finding coupled patterns in climate data. *Journal of Climate, 5*(6), 541-560. https://doi.org/10.1175/1520-0442(1992)005%3C0541:AIOMFF%3E2.0.CO;2

Cadavid Valencia, S. (2015). *Metodología para estimar caudales medios y extremos en escenarios de cambio climático* [Master's thesis, Universidad Nacional de Colombia]. https://repositorio.unal.edu.co/items/0fff47a0-e373-479a-9711-5de1b6941cf9

Chou, S. C., Marengo, J. A., Lyra, A. A., Sueiro, G., Pesquero, J. F., Alves, L. M., kay, G., Betts, R., Chagas, D. J., Gomes, J. L., Bustamante, J. F., & Tavares, P. (2012). Downscaling of South America's present climate driven by 4-member HadCM3 runs. *Climate Dynamics, 38*(3), 635-653. https://doi.org/10.1007/s00382-011-1002-8

Fu, G., Charles, S. P., & Kirshner, S. (2013). Daily rainfall projections from general circulation models with a downscaling nonhomogeneous hidden Markov model (NHMM) for southeastern Australia. *Hydrological Processes*, *27*(25), 3663-3673. https://doi.org/10.1002/hyp.9483

Greene, A. M., Robertson, A. W., Smyth, P., & Triglia, S. (2011). Downscaling projections of Indian monsoon rainfall using a non-homogeneous hidden Markov model. *Quarterly Journal of the Royal Meteorological Society, 137*(655), 347-359. https://doi.org/10.1002/qj.788

Güiza-Villa, N., Gay García, C., & Ospina-Noreña, J. (2020). Effects of climate change on water resources, indices, and related activities in Colombia. In P. T. Chandrasekaran, M. S. Javaid, & A. Sadiq (Eds.), *Resources of Water* (art. 90652). IntechOpen. https://doi.org/10.5772/intechopen.90652

Hughes, J. P., & Guttorp, P. (1994). A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resources Research, 30*(5), 1535-1546. https://doi.org/10.1029/93WR02983

IPCC (2014). *Part A: Global and sectoral aspects (contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*). https://www.ipcc.ch/site/assets/uploads/2018/02/WGIIAR5-FrontMatterA_FINAL.pdf

Johnson, B., Kumar, V., & Krishnamurti, T. N. (2014). Rainfall anomaly prediction using statistical downscaling in a multimodel superensemble over tropical South America. *Climate Dynamics, 43*(7), 1731-1752. https://www.researchgate.net/publication/264157405_Rainfall_anomaly_prediction_using_statistical_downscaling_in_a_multimodel_superensemble_over_tropical_South_America

Koki, C., Leonardos, S., & Piliouras, G. (2022). Exploring the predictability of cryptocurrencies via Bayesian hidden Markov models. *Research in International Business and Finance*, *59*, 101554. https://doi.org/10.1016/j.ribaf.2021.101554

Liu, W., Fu, G., Liu, C., & Charles, S. P. (2013). A comparison of three multi-site statistical downscaling models for daily rainfall in the North China Plain. *Theoretical and Applied Climatology, 111*(3-4), 585-600. https://doi.org/10.1007/s00704-012-0692-0

López López, P., Immerzeel, W. W., Rodríguez Sandoval, E. A., Sterk, G., & Schellekens, J. (2018). Spatial downscaling of satellite-based precipitation and its impact on discharge simulations in the Magdalena River basin in Colombia. *Frontiers in Earth Science*, *6*, 68. https://www.frontiersin.org/journals/earth-science/articles/10.3389/feart.2018.00068/full

Pineda, L. E., & Willems, P. (2016). Multisite downscaling of seasonal predictions to daily rainfall characteristics over Pacific–Andean River Basins in Ecuador and Peru Using a nonhomogeneous Hidden Markov model. *Journal of Hydrometeorology*, *17*(2), 481-498. DOI: https://doi.org/10.1175/JHM-D-15-0040.1

Posada-Marín, J. A., Rendón, A. M., Salazar, J. F., Mejía, J. F., & Villegas, J. C. (2019). WRF downscaling improves ERA-Interim representation of precipitation around a tropical Andean valley during El Niño: Implications for GCM-scale simulation of precipitation over complex terrain. *Climate Dynamics, 52*(5), 3609-3629. https://doi.org/10.1007/s00382-018-4403-0

Robertson, A. W., Kirshner, S., & Smyth, P. (2004). Downscaling of daily rainfall occurrence over Northeast Brazil using a hidden Markov model. *Journal of Climate, 17*(22), 4407-4424. https://doi.org/10.1175/JCLI-3216.1

Rojo, J. (2018). *Spatial and temporal characterization of Colombia's hydroclimatology* [Doctoral thesis, Universidad Nacional de Colombia]. https://repositorio.unal.edu.co/handle/unal/69017

Rojo, J., Lall, U., & Mesa, O. J. (2017). *A hidden model of daily precipitation over western Colombia*. https://meetingorganizer.copernicus.org/EGU2017/EGU2017-18003.pdf

Rojo, J., & Mesa, O. J. (2018). *MJO Influence over northern South American observations and modeling.* https://doi.org/10.13140/RG.2.2.35057.10087

Rojo, J., Mesa, O. J., & Lall, U. (2020). ENSO dynamics, trends and prediction using machine learning. *Weather and Forecasting, 35*(5), 2061-2081. https://doi.org/10.1175/WAF-D-20-0031.1

Verbist, K., Robertson, A. W., Cornelis, W. M., & Gabriels, D. (2010). Seasonal predictability of daily rainfall characteristics in central northern Chile for dry-land management. *Journal of Applied Meteorology and Climatology*, *49*(9), 1938-1955.

Wilcke, R. A. I., Mendlik, T., & Gobiet, A. (2013). Multi-variable error correction of regional climate models. *Climatic Change, 120*, 871-887. https://doi.org/10.1007/s10584-013-0845-x

Yin, S., & Chen, D. (2020, June 30). Weather generators. *Oxford Research Encyclopedia of Climate Science*. https://doi.org/10.1093/acrefore/9780190228620.013.768

Zhang, M., Rojo, J., Yan, L., Mesa, O. J., & Lall, U. (2022). Hidden tropical pacific sea surface temperature states reveal global predictability for monthly precipitation for sub-season to annual scales. *Geophysical Research Letters, 49*(20), e2022GL099572. https://doi.org/10.1029/2022GL099572