

Circulation patterns identified by spatial rainfall and ocean wave fields in Southern Africa

András Bárdossy^{1,2*}, Geoffrey Pegram², Scott Sinclair², Justin Pringle² and Derek Stretch²

¹ Institute for Modelling Hydraulic and Environmental Systems, Universität Stuttgart, Stuttgart, Germany, ² Civil Engineering, University of KwaZulu-Natal, Durban, South Africa

OPEN ACCESS

Edited by:

David Barriopedro,
Universidad Complutense de Madrid
& Instituto de Geociencias (CSIC,
UCM), Spain

Reviewed by:

Daoyi Gong,
Beijing Normal University, China
Lin Wang,
Chinese Academy of Sciences, China
Sancho Salcedo-Sanz,
Universidad de Alcalá, Spain

*Correspondence:

András Bárdossy,
Institute for Modelling Hydraulic and
Environmental Systems, Universität
Stuttgart, Pfaffenwaldring 61,
70569 Stuttgart, Germany
bardossy@iws.uni-stuttgart.de

Specialty section:

This article was submitted to
Atmospheric Science,
a section of the journal
Frontiers in Environmental Science

Received: 04 September 2014

Accepted: 30 March 2015

Published: 28 April 2015

Citation:

Bárdossy A, Pegram G, Sinclair S,
Pringle J and Stretch D (2015)
Circulation patterns identified by
spatial rainfall and ocean wave fields in
Southern Africa.
Front. Environ. Sci. 3:31.
doi: 10.3389/fenvs.2015.00031

This paper presents the application of Fuzzy Rule Based Circulation Patterns (CPs) classification in the description and modeling of two different physical processes: rainfall regimes and ocean waves. Large ocean waves are typically generated over fetches of the order of thousands of kilometers far off shore, whereas rainfall is generated by local atmospheric variables including temperature, humidity, wind speed, and radiation over the area of concern. The spatial distribution of these variables is strongly dependent on regional pressure patterns, which are similar for associated weather and wind behavior on a given day. The choice of the CP groupings is made by searching for those CPs which generate (i) different daily rainfall patterns over mesoscale regions and (ii) wave heights from different directions at chosen shoreline locations. The method used to choose the groupings of CPs is a bottom-up methodology using simulated annealing, ensuring that the causative CPs are responsible for the character of the results. This approach is in marked distinction to top-down approaches such as k-means clustering or Self Organizing Maps (SOMS) to identify several classes of CPs and then analysing the effects of those CPs on the variables of choice on given historical days. The CP groups we define are often different for the two phenomena (rainfall and waves) simply because different details of the pressure fields are responsible for wind and for precipitation. The region chosen for the application is the province of KwaZulu-Natal in South Africa, using the same set of raw geopotential heights to represent the pressure patterns, but selecting from the set those typical patterns affecting ocean waves on the one hand and regional rainfall on the other.

Keywords: circulation patterns, rainfall, waves, classification, South Africa

1. Introduction

Local weather (precipitation, temperature wind) and related phenomena such as floods, storm, and waves are strongly dependent on atmospheric processes. These processes are very complex and highly non-linear. On the other hand due to the continuity of the atmospheric conditions these local phenomena are imbedded into large scale features. It is both of theoretical and practical importance to understand these links. Among the atmospheric variables, pressure is the driver for flow and transport. Air pressure at the land surface is a variable which can be measured simply and with good accuracy. Selected observations and meteorological models provide information on high

altitude pressure conditions. Air pressure and geopotential heights are among the best modeled quantities. Therefore, it is reasonable to use them as a basis to describe atmospheric circulation.

The relationship between local variables and atmospheric circulation described with the help of pressure conditions is a complicated and highly non-linear one. On the other hand it is reasonable to assume that similar circulation conditions cause similar local meteorological conditions. Therefore, an appropriate classification can help to quantify the relationship between local variables, such as amount of rain on a day in a region or wave height and period on a shoreline, and circulation. The intention is to define classes on the basis of the circulation and the local variables via conditional distributions. The behavior of the local variable described for example using the conditional distribution should be different for the different classes from the unconditional distribution of the variable. A comprehensive summary of classification methods can be found in Jacobbeit (2010).

There are different ways one can define groups of circulation patterns (CPs):

1. to define CPs using atmospheric variables only
2. to define CPs using atmospheric variables by taking the local variable into account for setting up the patterns
3. to define CPs using a combination of atmospheric variables and local variables

The first method intends to find typical distinct patterns of the atmospheric variables. The patterns differ in their defining space. Their link to local variables is determined after the classification and might for some variables yield a good distinction in the behavior while for others not.

The second method acknowledges the fact that relatively small differences of the atmospheric variables might lead to very different behavior of the local variables. This means that the classification should not intend to distinguish the patterns by producing very different CPs but to group CPs which to some extent are similar but explain the target variable as well as possible. On the other hand it is important that the classification is done on the basis of the atmospheric variables only so that the classification can be used for time periods lacking the observation of the local variables.

The goals of various classification schemes can differ widely. While the classification of pressure fields on a purely statistical basis might reveal specific features of atmospheric dynamics, it might not provide the best basis for the explanation of the behavior of surface variables such as wind, temperature or precipitation. Under certain circumstances relatively small differences in the atmospheric conditions can lead to very different behavior of the surface variables. If one is interested in explaining the relationship between surface variables and atmospheric conditions then the purpose is to obtain classes of CPs with distinct conditional probabilities or distributions of the selected variables.

The first classifications were developed for regions in Europe on a subjective basis (Lamb, 1972 for Great Britain and Hess and Brezowsky, 1952 for Germany). Automated classifications were

developed for different regions in Europe, North America, and China. In this paper fuzzy rule based classifications are developed for precipitation and waves in South Africa.

2. Methodology

The basic methodology of fuzzy rule-based classification was described for explaining precipitation behavior in Bárdossy et al. (1995). The main ideas are summarized here.

The classification is performed on anomalies of daily air pressure based variables: sea level pressures or geopotential heights g . These data can be obtained from reanalysis products on a regular grid. The anomalies are calculated under consideration of the annual cycle of both the mean and the standard deviation of the observed pressure. The anomaly at gridpoint x and time t is calculated as:

$$d(x, t) = \frac{g(x, t) - \bar{g}(x, J(t))}{s_g(x, J(t))} \quad (1)$$

where $J(t)$ is the Julian date corresponding to day t (which is a value between 1 and 366), $\bar{g}(x, J(t))$ is the mean and $s_g(x, J(t))$ is the standard deviation of g at location x and Julian day $J(t)$. These anomalies d have zero mean and unit standard deviation for each grid location.

The classification is based on the location of certain anomalies. For each gridpoint x of the pressure grid G five different possibilities of anomalies are considered to define a classification class. A triangular fuzzy number is assigned to each of these possible anomalies:

1. positive anomaly $(0, 3, +\infty)_T$
2. negative anomaly $(-\infty, -3, 0)_T$
3. non-positive anomaly $(-4, -0.85, 0.25)_T$
4. non-negative anomaly $(0.25, 0.85, 4)_T$
5. non-representative anomaly $(-\infty, 0, +\infty)_T$

where the subscript T denotes a triangular fuzzy number with the membership function:

$$\mu_{(a,b,c)_T}(d) = \begin{cases} \frac{d-a}{b-a} & \text{if } a \leq d \leq b \\ \frac{d-c}{b-c} & \text{if } b < d \leq c \\ 0 & \text{else} \end{cases} \quad (2)$$

This means that for the classification of each gridpoint one of the above possible classes is assigned. Thus, a class k can be defined as a vector $j_k(x) \in G$ where each $j_k(x)$ is an integer such that $1 \leq j_k(x) \leq 5$.

For a given day t and a given set of rules $j_k(x)$ $x \in G$ and $k = 1, \dots, K$ the classification is performed as follows:

1. For each rule k and each location x the membership value of $d(x, t)$ in the fuzzy set corresponding to $j_k(x)$ is calculated. These values are $\mu_{j_k(x)}(d(x, t))$
2. These individual membership values are combined to an overall degree of fulfillment of the rule k by calculating:

$$\text{DOF}(k, t) = \prod_{m=1}^4 \left[\frac{1}{\#\{x \mid j_k(x) = m\}} \sum_{x \mid j_k(x) = m} \mu_{j_k(x)}(d(x, t))^{q_m} \right]^{\frac{1}{q_m}} \quad (3)$$

3. The k_0 index corresponding to the maximal $\text{DOF}(k, t)$ is selected as class for day t

For further details the readers are referred to Bárdossy et al. (1995).

The fuzzy rule based classification is performed to explain the behavior of one or more selected variables. These are denoted with the variable V . Two types of objective functions are considered. The first one relates the CPs to the exceedance of certain thresholds of V . The objective is to define classes with frequencies of exceedances which differ from the unconditional frequency of occurrences. The first is O_1 :

$$O_1(\theta) = \left(\sum_{t=1}^T (h_{CP(t)}(V > \theta | CP(t)) - h(V > \theta))^2 \right)^{\frac{1}{2}} \quad (4)$$

where h stands for the frequency of an event, V is the variable under interest and θ is a prescribed threshold. This objective expresses that a classification should help to decide if a threshold is exceeded or not.

The second objective type is related to the mean magnitude of the variable V with respect to the CPs. It is formulated as:

$$O_2 = \frac{1}{T} \sum_{t=1}^T \left| \frac{(V | \bar{CP}(t))}{\bar{V}} - 1 \right| \quad (5)$$

The second objective function evaluates the ability of the algorithm to derive classes with average values of V that are different from the unclassified average. In other words this objective function measures the separability of the classes from the mean.

These objective functions can be used in combination—for example for different thresholds θ and even for different variables. A weighted combination of them is used as the classification objective function.

The value of the objective functions depends on the classification, which itself depends on the selected rules $\{j_k(x) | x \in G, k = 1, \dots, K\}$, where K is the number of rules. Once objective functions for the classification are defined, different classifications can be compared with respect to their ability to explain the variability of the surface variable investigated. The higher the objective function values the better the classification is. Thus, one can use an optimization procedure to find rules $j_k(x)$ which maximize the objective functions. The number of possible rule systems for a given number of patterns is given by the combinatorial:

$$N_K = \binom{5^{|G|}}{K}$$

This is usually an extremely large number (for example for a small pressure grid 10×10 with $|G| = 100$ nodes and $K = 10$ rules the number of possible rule systems is $> 10^{340}$) thus the best rules cannot be found by trying all possibilities. Instead an optimization method has to be applied. Simulated annealing provides a reasonable alternative as described in Bárdossy et al. (2002).

2.1. Classification Quality

The quality of the classification can be measured by the above defined objective functions. Besides that other measures can be defined, which can be used for comparison, for example to find the optimal number of classes. For this one can use thresholds for the variable V . The entropy of the conditional distributions can be used to measure the binary quality of a classification:

$$H(K) = \sum_{k=1}^K h_k (p_k \log_2 p_k + (1 - p_k) \log_2 (1 - p_k)) \quad (6)$$

where p_k is the probability of the exceedance of the threshold on a day with the k -th CP, and h_k is the frequency of the k -th CP. The quantity $H(0)$ is the entropy in the case no classification was performed:

$$H(0) = p_0 \log_2 p_0 + (1 - p_0) \log_2 (1 - p_0) \quad (7)$$

A classification provides information for the variable V if $H(K) < H(0)$. In general a classification using K classes is better than another with L classes if:

$$H(K) < H(L)$$

2.2. Non-Uniqueness Issues

Different CP classifications can be obtained for the same region. In Philipp et al. (2014) the authors investigated 27 automatic CP classification approaches and found that different classification methods lead to very different classifications. The Fuzzy Rule Based classification method can lead to different classifications as the stochastic optimization obtained using simulated annealing does not necessarily lead to the same classification if another sequence of random numbers is selected. A different choice of stations or a different choice of thresholds or weights for the objective functions can lead to different classifications too. Different classifications can be compared from the viewpoint of their performance or from the viewpoint of the similarity of their classes. While the calculation of the objective functions is straightforward, the similarity of the classifications requires further measures.

Two statistics can be calculated based on the contingency tables of the pairs of classifications. For two different classifications n_{ij} is the number of days in class i for the first and in class j in the second classification. The χ^2 statistics to compare the classifications is calculated as:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij}n - n_i \cdot n_j)^2}{n_i \cdot n_j n} \quad (8)$$

where r is the number of classes in the first classification, s is the number of classes in the second classification, n_i is the number of days in class i for the first classification, n_j is the number of days in class j for the second classification and n is the total number of days. If the classifications are independent then the χ^2 value is small, the bigger it is the more the classifications resemble each

other. Based on the χ^2 values two measures of dependence were calculated, the modified Pearson coefficient C :

$$C = \sqrt{\frac{\min(r, s)}{\min(r, s) - 1}} \sqrt{\frac{\chi^2}{\chi^2 + n}} \quad (9)$$

and the Cramer coefficient (Hartung et al., 2005) V :

$$V = \sqrt{\frac{\chi^2}{n(\min(r, s) - 1)}} \quad (10)$$

Both coefficients are bounded by 1, and the higher they are the stronger the association is. The associations of CP classifications for the same geographical regions using different objective functions or different stochastic optimization settings are usually strong, which is a consequence of the fact that the same variables (geopotential heights) are classified.

2.3. Data Used

The classifications were performed using ECMWF ERA reanalysis data sets. The classification is based on daily normalized anomalies, derived from the 700 hPa geopotential height with a grid resolution of 2.5° (10° – 50° S; 0° – 50° W). Geopotential heights were obtained from the ERA-Interim data set of Dee et al. (2011) for the period 1979–2009 (<http://apps.ecmwf.int/datasets/>).

3. Classification for Precipitation

3.1. Objective Functions

Precipitation has a highly skewed distribution with above 80% probability of a dry day in many parts of South Africa. Individual very high values of precipitation can lead to “random” optima which are not robust. Instead a new variable related to the average wetness of a selected number of stations was considered:

$$W(t) = \frac{\#\{x_i; Z(x_i, t) > 0\}}{\#\{x_i; Z(x_i, t) \geq 0\}} \quad (11)$$

by definition $0 \leq W(t) \leq 1$. This variable is taken as V for the objective functions in Equations (4, 5).

The 24 Climate Regions defined by Kruger (2004), slightly modified by concatenating some of the very small regions (mostly in dry areas) with larger ones, were used for the classification. The map with the locations of the regions is shown in **Figure 1**. A set of representative stations was selected for each region and the CPs were classified using the above defined objective function. The classification obtained using this objective function works well for precipitation in all selected regions both for the calibration and the validation time periods. For variety, we will show some CPs and resulting rainfall in **Figure 2** and specific CPs selected for region 6 in **Figure 3**.

3.2. Spatial Extent of Classification and Non-Uniqueness of CP Set Selection

We explore three sets of classifications driven by precipitation wetness. The first CP set is conditioned on wetness in region 5 and we compare the differences in rainfall distributions from

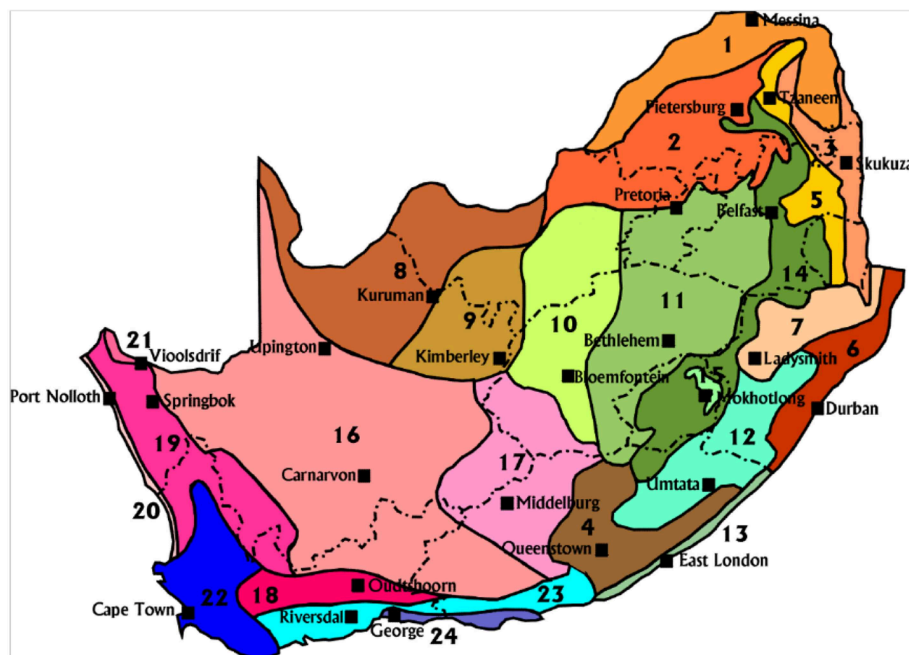


FIGURE 1 | The Kruger climate regions of South Africa after (Kruger, 2004).

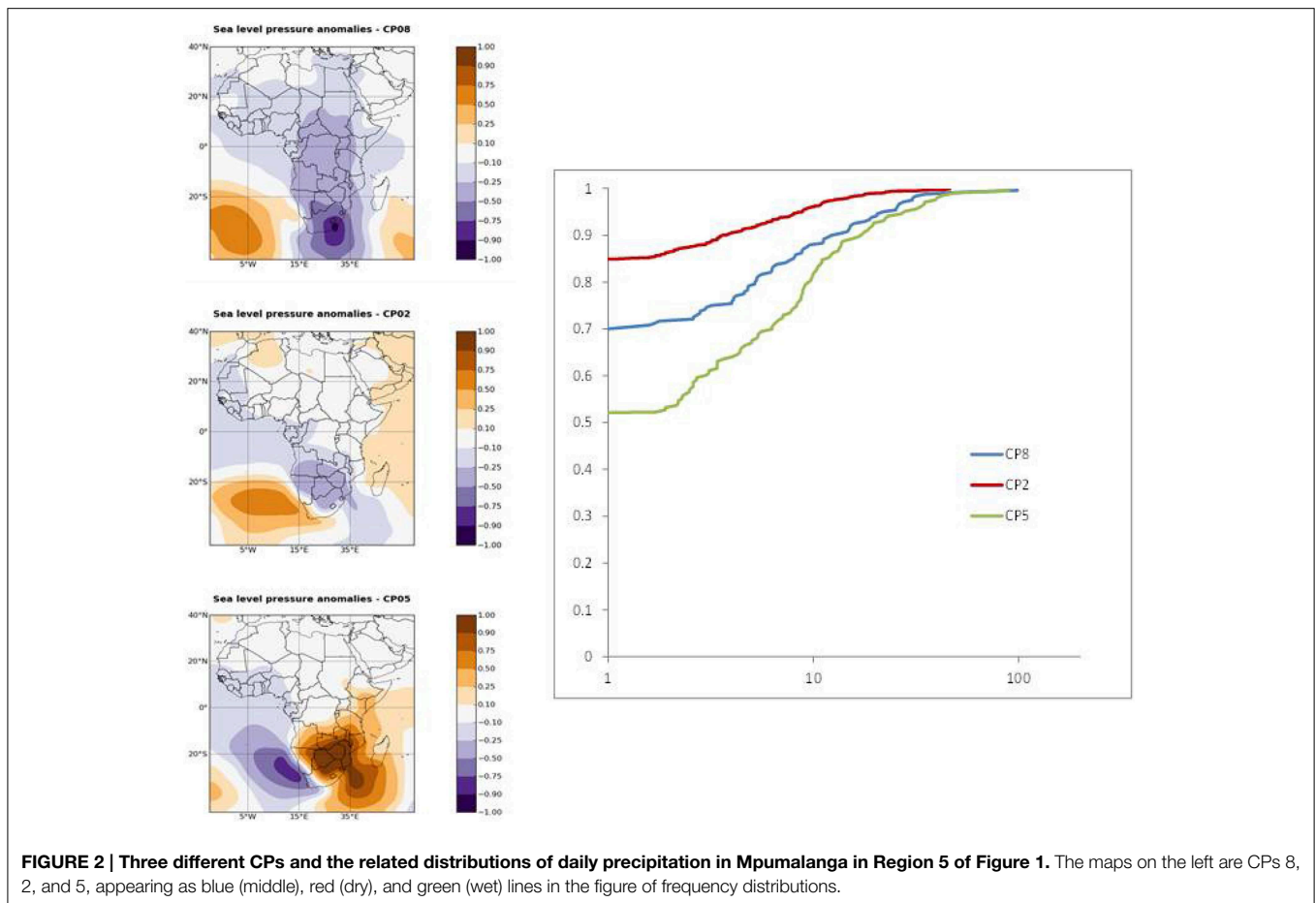


FIGURE 2 | Three different CPs and the related distributions of daily precipitation in Mpumalanga in Region 5 of Figure 1. The maps on the left are CPs 8, 2, and 5, appearing as blue (middle), red (dry), and green (wet) lines in the figure of frequency distributions.

three distinct sets of CPs. The second set involves determining the robustness of the method when different sets of gauges in a region are used for CP classification. For this purpose we use data from region 6 and randomly sample 2 sets for comparison. The third experiment treats CPs defined over three regions in **Figure 1**. These are region 6 which experiences subtropical summer rainfall and the occasional hurricane, 10 which is typically savannah, experiencing summer rainfall dominated by convective systems, and 22 which has a Mediterranean climate and experiences mostly frontal system winter rainfall.

Due to the stochastic optimization no unique best classification can be achieved. It is of interest to see how

- the selection of the stations for the objective function
- the randomness due to stochastic optimizations

influences the results. For this purpose the performance measures were calculated for different classification results.

To introduce the first experiment, which is to demonstrate the link between rainfall regimes and CPs, we offer **Figure 2**, which shows the correspondence between (i) three CPs chosen from a set of 8 in Mpumalanga and (ii) the rainfall distribution at a particular rain-gauge in the region. **Figure 2** shows the 700 hPa anomalies for three selected CPs and the corresponding distributions of daily precipitation. CP2 is the driest with the

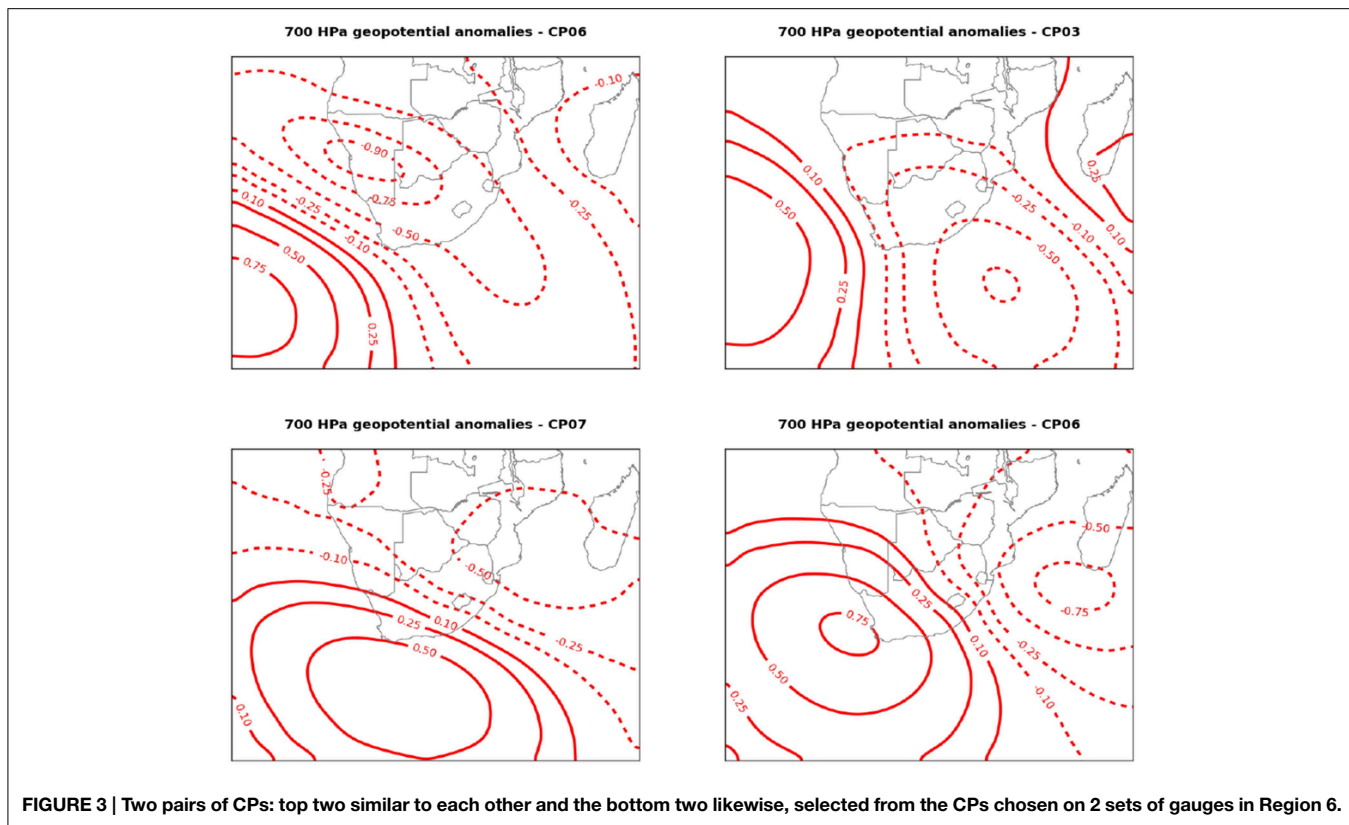
highest probability of a dry day. CP5 is the wettest with a much lower probability of a dry day. CP8 is medium wet with statistics between CP5 and CP2. The distributions are significantly different from each other indicating that classification provides useful information on precipitation behavior.

To illustrate the methodology applied in the second experiment, we selected region 6, then we selected some gauges within the region, in different configurations, to classify the Circulation Pattern anomalies [CPs] which are the cause of different types of rainfall over this region.

We find we get similar CPs, enough similarity of shape to pick a set, [note that the labeling within each set is random, so we match by correlation of shape, not label]. **Figure 3** shows the 700 hPa anomalies for two different classifications.

This similarity allows us to settle on one set of CPs per region (and season) because we have a robust method. In the next section we turn to the Infilling/repair problem an extension of which we will use later in spatial interpolation between gauges.

The third experiment was to make comparisons between CP sets within and between 3 regions. This started by selecting 5 random subsets of 12 gauges from the historical datasets in each of the regions. CPs were conditioned on each of the 5 subsets in each region independently. We then set out to determine whether the similarity of the sets within each region was materially greater



than between regions. Because these are categorical data it does not make sense to compute correlation coefficients between sets, so Pearson and Cramer coefficients (Equations 9, 10) based on the χ^2 were the statistics used.

We chose three climate regions within South Africa to compare sets of CPs. The regions chosen and shown in **Figure 1** are: region 6 (KwaZulu-Natal), region 10 (Free State), and region 22 (Western Cape). We used 5 different random groupings of gauges for classifications in each region, thus for each region we derived 5 independent sets of CPs, making 15 in all.

The purpose of this calculation was to determine if the CPs derived for a given region had a higher inter-association than between regions. These coefficients were calculated for all pairs of classifications for the same regions. Then the averages of the association measures were calculated for classifications corresponding to the same (excluding the comparison of a classification with itself), and corresponding to different regions leading to 2 sets of 3 by 3 matrices.

The average of the 20 within-region statistics are compared with the averages of the 25 between-region statistics and appear in **Tables 1, 2** (the difference in the numbers in each group is because we omitted the diagonal elements of comparing a set with itself). The result is that there is a significant difference between the within against the between coefficients, supporting our exploitation of this CP selection procedure based on daily wetness.

All classifications show significant dependence, which of course is reasonable as the atmospheric conditions on the

TABLE 1 | CP similarities for different classifications using the Pearson coefficient in the regions numbered 6, 10, and 22 in Figure 1.

Region	6	10	22
6	0.699	0.617	0.640
10	0.617	0.674	0.592
22	0.640	0.592	0.698

TABLE 2 | CP similarities for different classifications using the Cramer coefficient in the regions numbered 6, 10, and 22 in Figure 1.

Region	6	10	22
6	0.263	0.213	0.226
10	0.213	0.247	0.201
22	0.226	0.201	0.262

same region are classified in all cases. On the other hand the within block average statistics are larger than the between block counterpart, inferring that there is a stronger within relationship than between blocks. The conclusion is that the choosing of CPs dependent on regions is valuable and worth the effort.

4. Classification for Waves

Atmospheric circulations drive regional wave climates through atmosphere-ocean interactions. In particular they control the

generation of the extreme wave events that cause severe coastal erosion. They are therefore also fundamental drivers of coastal vulnerability. The link between the wave climates and atmospheric circulation is complex. However, statistical models that link synoptic scale atmospheric circulation to regional wave characteristics have recently been shown to give significant insights (Pringle et al., 2014). We propose that the classification of atmospheric drivers can improve coastal vulnerability assessments and the prediction of climate change effects. For example it provides a natural way to identify and isolate the effects of independent storm events, which is required for extremum analysis. For example (Corbella and Stretch, 2012a) identified independent events based on the autocorrelation or by using a 6-h inter-arrival window. The transition of CPs between classes is a physically more meaningful method for defining independent events. Atmospheric CPs also contain important information regarding the distribution of wave height, direction and period, because when a particular CP type occurs the associated wave height, direction, and period can be (statistically) predicted. Linking wave events to CPs can also be used to extend current data sets and to infill missing data (Hewitson and Crane, 2002). Finally the prediction and evaluation of climate change impacts on coastal vulnerability would be more robust if linked to changes in the atmospheric CPs that are the basic drivers of wave climates and extreme wave events.

4.1. Case Study Description

The KwaZulu-Natal (KZN) coastline (Figure 4) is associated with a high energy wave climate. A number of weather types have been cited as the drivers of this wave climate. For example tropical cyclones, mid-latitude (extra-tropical) cyclones and cut off lows (Rossouw et al., 1982; Corbella and Stretch, 2012c; Mather and Stretch, 2012). The location and persistence of tropical cyclones (TC's) are believed to drive large wave events that cause severe beach erosion in KZN (Corbella and Stretch, 2012c; Mather and Stretch, 2012). Cut-off lows are deep low pressure systems that

are displaced from the normal path of west-east moving mid-latitude cyclones (Preston-Whyte and Tyson, 1988). They are caused by instabilities within the westerly zonal flow due to the high wind shear. Vortices can become cut-off and move equatorward (Preston-Whyte and Tyson, 1988). These features lead to seasonality in the wave climate and the occurrence of storms in particular. On average autumn and winter (April to September) are associated with the largest wave energy, while summer (January to March) has the smallest (Corbella and Stretch, 2012c). The significant wave height (H_s) is the key variable of interest for coastal vulnerability applications. Our algorithm considers both the daily average significant wave height and the daily maximum significant wave height. Wave data for the period 1992–2009 were obtained from wave buoys located near Durban and Richards Bay on the KZN coastline (Figure 4).

In this application the goal of the classification is to obtain a set of CP classes which explain extreme wave events. Wave heights larger than 3.5 m have been shown to cause significant erosion along the KZN coastline (Corbella and Stretch, 2013). The objective functions (Equations 4, 5) were used as performance measures.

The wave height threshold θ can be exploited to incorporate various scenarios. Two different thresholds were used. The first relates to the occurrence of extreme events ($\theta_1 = 3.5$ m), while the second relates to midrange wave heights ($\theta_2 = 2.5$ m). The second objective function type is also used to provide a more detailed classification.

To account for the persistence of CPs during extreme events (Equation 4) was modified to include storm durations, defined as the duration of wave height excursions above 3.5 m.

4.2. Dominant CP Classes

Figure 5 shows the average anomaly patterns for all the CP classes. CP99 refers to an unclassified class. Useful statistical parameters for each CP class are their frequency of occurrence, their contribution to extreme events, and the average and maximum significant wave heights (H_s) associated with them. These parameters are shown in Table 3.

The results reflect a number of trends in CP-wave generation. However, only the two most significant trends are discussed herein. Firstly CP01 and CP02 (Figure 5A) occur most frequently (about 17% of the time). Both CP classes resemble a mid-latitude cyclone in its different stages of development. The CP01 resembles the central low pressure region of a mid-latitude cyclone as it moves from west to east south of the country, while CP02 resembles the high-pressure region that follows. Secondly, Table 4 shows a class (CP03) that comprises 60% of all extreme wave events. The CP03 is shown to contribute significantly to extreme events all year round with highest contribution in winter (65%). However, CP03 (Figure 5C) occurs infrequently (9% of the time). Its occurrence is associated with large average and maximum significant wave heights ranging from 2.4 to 3.0 m and 5.0 to 8.5 m, respectively. CP05 and CP06 (Figures 5E,F) contribute approximately to 30% of extreme events in spring and summer, respectively. The CP06 resembles a pattern similar to TC's south of Madagascar. TC's located within this region have been

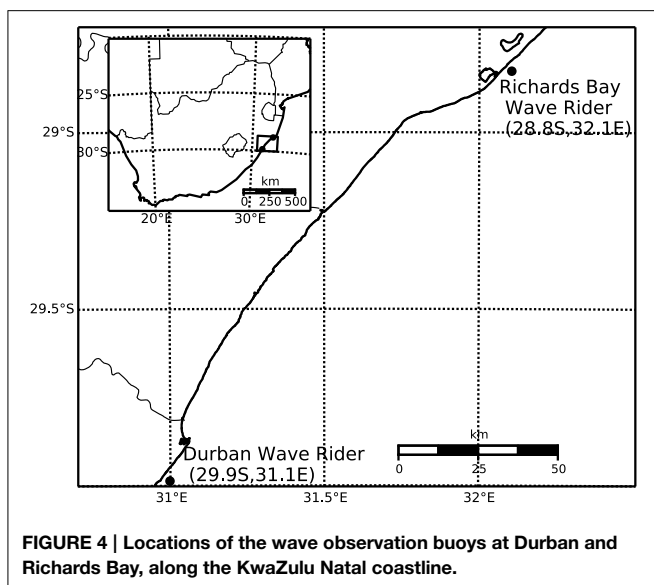


FIGURE 4 | Locations of the wave observation buoys at Durban and Richards Bay, along the KwaZulu Natal coastline.

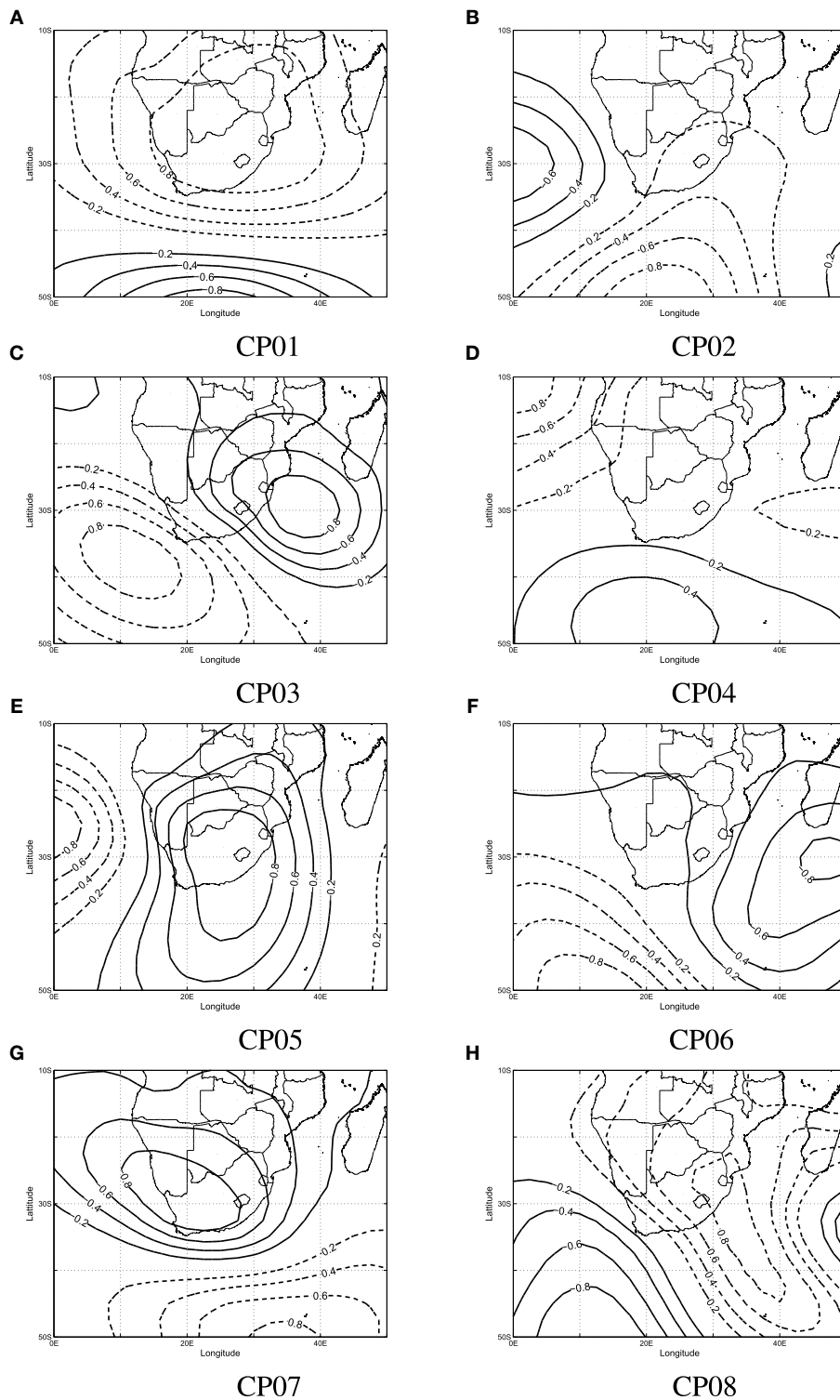


FIGURE 5 | Average anomaly patterns for 8 CP classes derived from the regional wave climate data. Negative (low) pressure anomaly contours are shown as solid lines while positive pressure contours are dashed.

cited to drive large swells toward the KZN Coast (Mather and Stretch, 2012). The CP05 resembles low-pressure systems over the interior.

The algorithm only considered CPs associated with extreme events at the time the event was recorded. In other words no time lags were considered when deriving the CP classes. Therefore, the

TABLE 3 | CP Occurrence frequencies and wave height statistics associated with each CP class.

Statistics	CP01	CP02	CP03	CP04	CP05	CP06	CP07	CP08	CP99 ^a
OCCURRENCE FREQUENCY ($p(\text{CP})\%$)									
Summer	18	18	8.0	13	7.5	8.1	5.6	15	8.3
Autumn	18	19	8.0	11	10	7.2	5.1	13	8.8
Winter	16	17	8.1	12	11	8.4	4.5	14	8.9
Spring	17	16	9.1	12	9.4	7.7	5.0	15	9.2
All Seasons	17	17	8.3	12	9.6	7.8	5.1	14	8.8
THRESHOLD EXCEEDANCE FOR A GIVEN CP ($p(H_s \geq \theta \text{CP})\%$)									
Summer	–	0.4	8.0	0.3	–	3.5	2.6	0.7	–
Autumn	1.2	1.5	12	1.6	2.1	2.0	5.6	0.6	5.0
Winter	0.9	0.8	14	–	0.9	0.4	2.3	0.8	0.4
Spring	0.3	–	4.6	0.6	2.2	–	0.7	–	–
All Seasons	0.6	0.7	9.6	0.6	1.4	1.5	2.8	0.5	1.4
EXCEEDANCE CONTRIBUTION ($p(\text{CP} H_s \geq \theta)\%$)									
Summer	–	5.6	50	2.8	–	22	11	8.3	–
Autumn	7.7	10	33	6.4	7.7	5.1	10	2.6	17
Winter	7.5	7.5	64	–	5.7	1.9	5.7	5.7	1.9
Spring	4.5	–	55	9.1	27	–	4.5	–	–
All Seasons	5.8	7.4	48	4.2	7.9	6.9	8.5	4.2	7.4
AVERAGE H_s (m) FOR EACH CP									
Summer	1.8	1.9	2.5	1.8	1.8	2.2	2.2	1.9	1.9
Autumn	1.8	1.9	2.7	1.9	2.0	2.0	2.1	1.9	2.1
Winter	2.0	2.0	2.9	1.9	2.1	2.0	2.2	2.0	1.9
Spring	2.0	1.9	2.4	1.9	2.2	2.0	2.2	2.0	2.0
All Seasons	1.9	1.9	2.6	1.9	2.1	2.1	2.2	2.0	2.0
STANDARD DEVIATION of H_s (m) FOR EACH CP									
Summer	0.48	0.49	1.1	0.49	0.53	0.76	0.74	0.61	0.49
Autumn	0.58	0.66	1.0	0.70	0.76	0.66	0.90	0.55	1.0
Winter	0.58	0.61	0.94	0.55	0.66	0.58	0.66	0.55	0.67
Spring	0.51	0.49	0.84	0.52	0.71	0.50	0.61	0.50	0.49
All Seasons	0.54	0.57	1.0	0.56	0.69	0.63	0.74	0.56	0.70
MAX H_s (m) FOR EACH CP									
Summer	3.4	4.0	8.5	3.7	3.4	5.0	5.2	5.6	3.3
Autumn	4.0	5.5	5.7	5.5	6.3	4.3	5.1	4.0	5.4
Winter	4.2	3.8	5.6	3.4	3.8	3.5	4.3	4.8	3.6
Spring	3.9	3.3	5.3	4.5	5.4	3.4	3.7	3.5	3.3
All Seasons	4.2	5.5	8.5	5.5	6.3	5.0	5.2	5.6	5.4

^aCP99 is the unclassified class. Blank entries imply zero occurrences in the data set.

algorithm assumes that extreme events are driven by relatively stationary CPs.

4.3. CP Variability

The degree of fit (DOF) describes the membership of a CP for given day to a particular class. The larger the DOF the stronger the belief that the CP belongs to a particular class. **Figure 6** shows the average anomaly pattern for CP03 and the CPs associated with both the strongest and weakest membership for that class. CP03 is associated with a strong low pressure region east/south-east of South Africa. The pattern also shows a strong high-pressure region to the southwest. The coupling

between the strong low and high pressure drive strong winds and subsequently large waves toward the coastline. The CP with the weakest membership to CP03 is shown to be a weak anomaly pattern (refer to **Figure 6C**).

4.4. Variability within Classes

It is expected that in regions of high pressure pattern variation should be low. This is attributed to the stability of high pressures in their positions. In contrast pattern variation in regions of low pressure should be higher primarily due to the movement of low pressure systems. This is reflected in **Figure 6D**, which shows significantly larger variation (standard deviation of 1)

in the vicinity of the low-pressure region. The magnitude of the variation also indicates that CPs driving extreme events are associated with strong low pressures.

4.5. CP Rules and Extreme Events

Daily CP realizations associated with extreme wave events ($H_s \geq 3.5$ m) were compared to the average class pattern to which they belong. **Figure 7** shows the average pattern for CP03 together with selected extreme events corresponding to CP03. The locations of the peak negative anomalies are shown in the plot. Significant pattern variability within the class is apparent. **Figure 8** and **Table 4** show the CPs associated with six of the

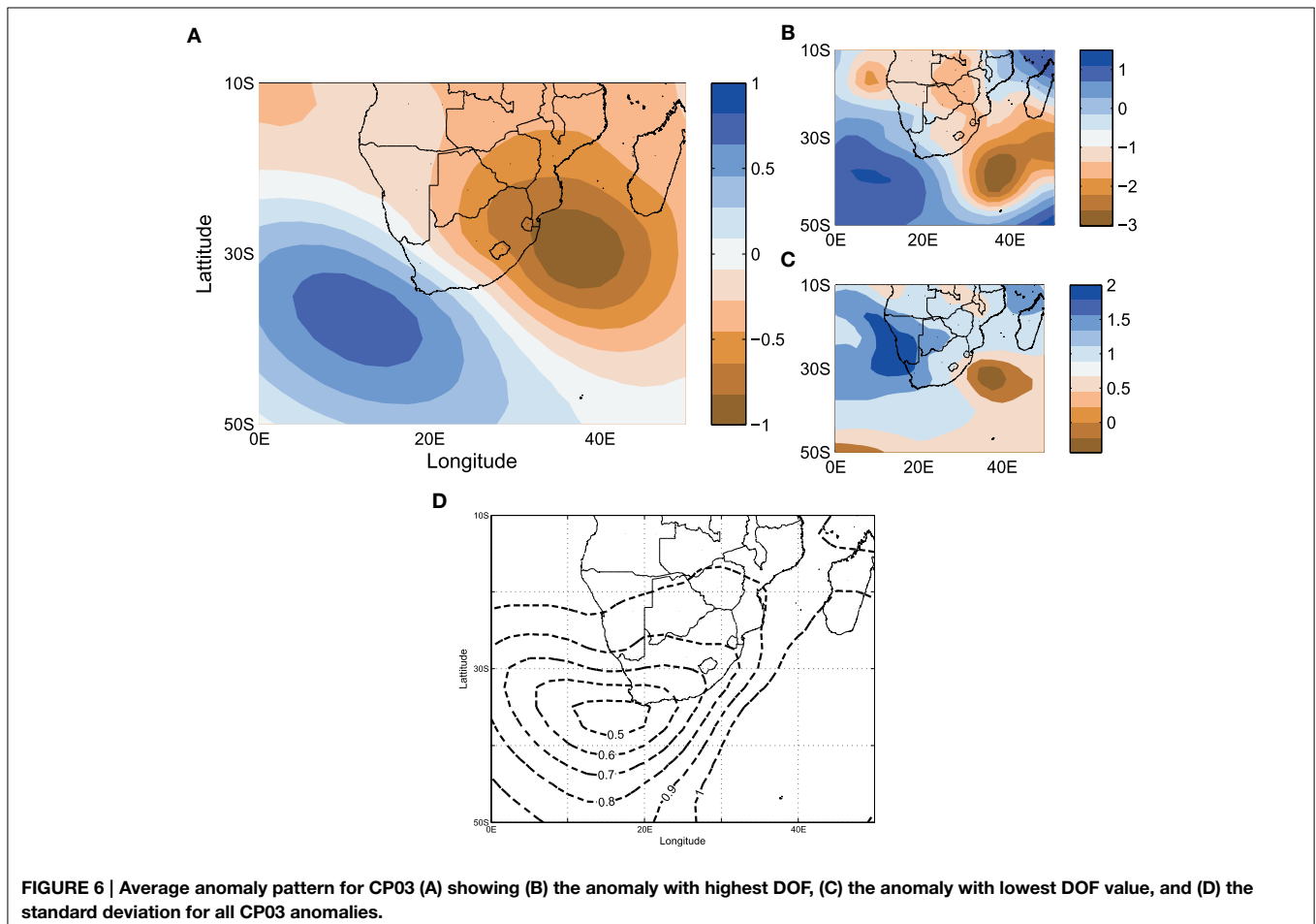
largest significant wave height events. The majority of the six events have been classified as belonging to CP03. **Figure 8F** from visual inspection shows a pattern similar to CP04 and CP08. Both classes are associated with low-pressure regions southeast of Madagascar. However, the CP has been classified as belonging to class CP08 and not CP04. From a visual account it appears to better resemble class CP04. **Figures 8A,C** are the CPs associated with the March 2007 storm which caused severe coastal erosion along the KZN coastline (Corbella and Stretch, 2012b; Mather and Stretch, 2012) with significant wave heights reaching 8.5 m.

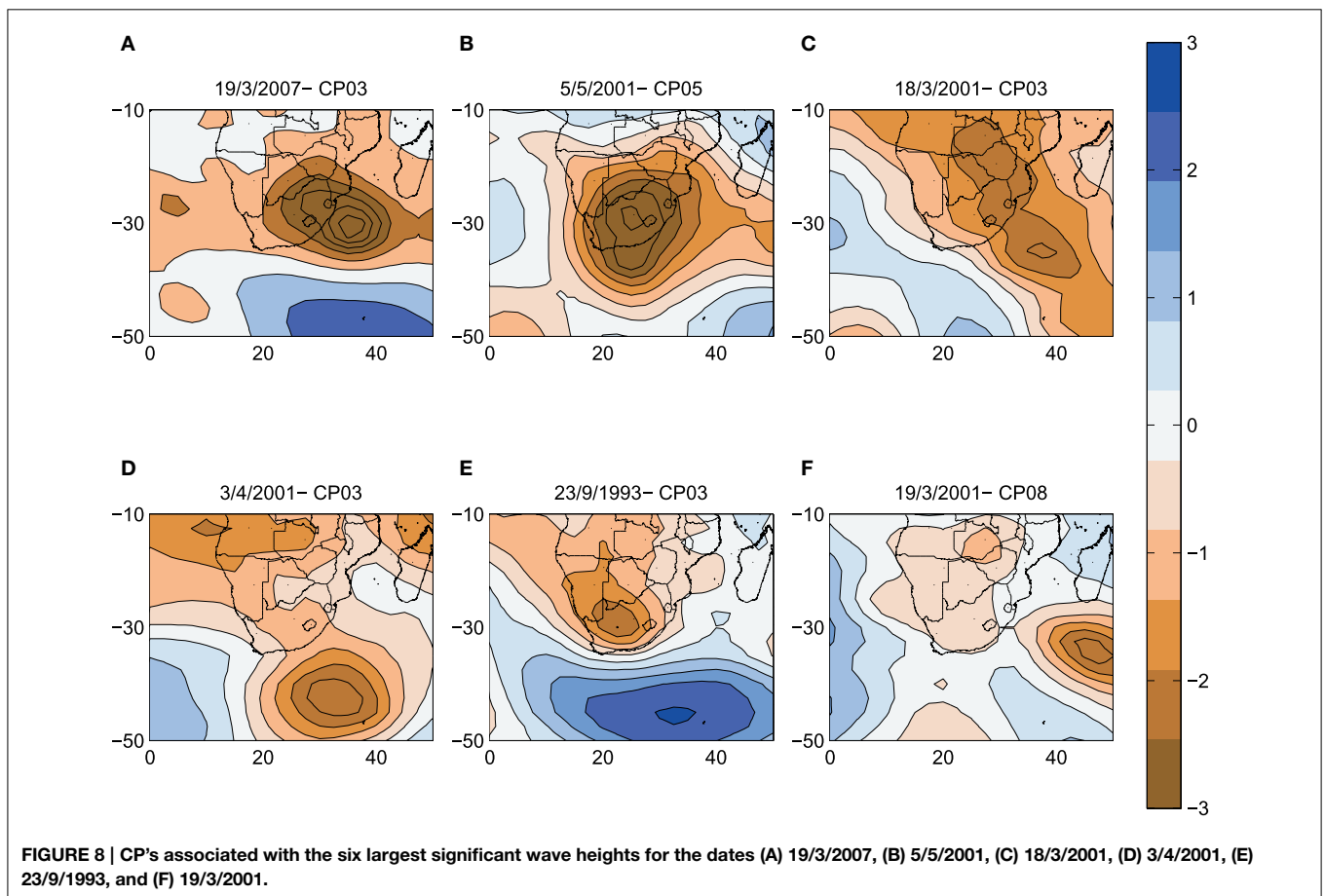
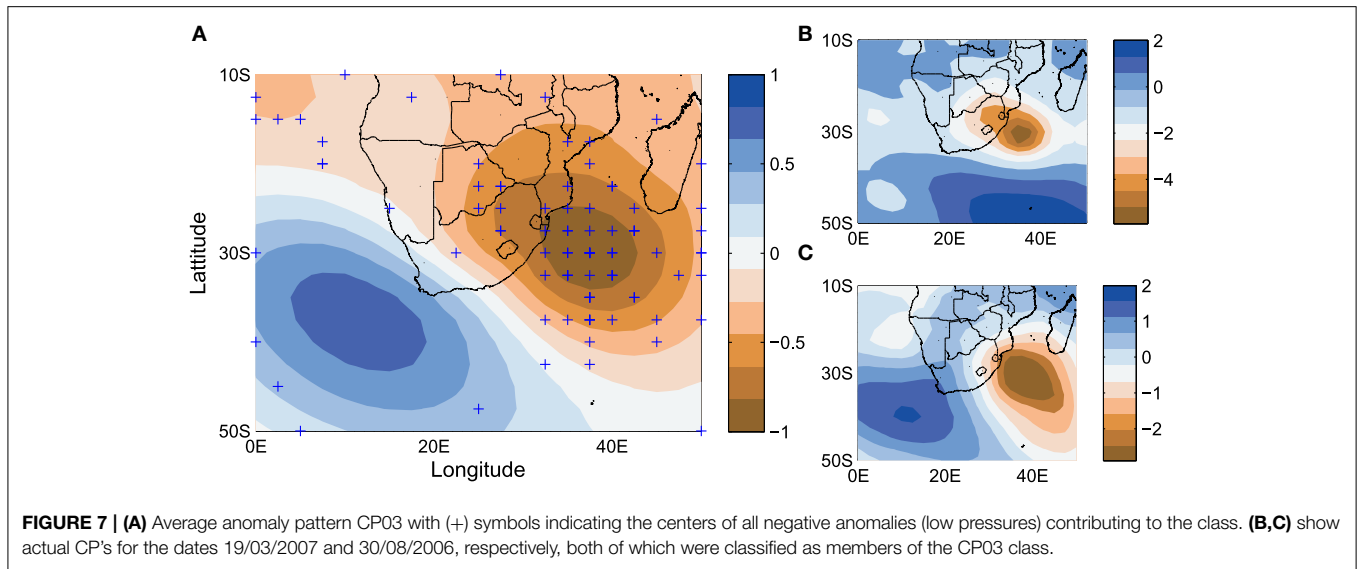
5. Discussion and conclusions

In this paper different problems related to the fuzzy rule based classification of CPs were discussed. It was shown that reasonable classifications can be developed for precipitation and waves in South Africa. We did this by selecting long sequences of CPs in the region surrounding Southern Africa and conditioning the selection of the CPs on selected variables germane to the task; daily wetness for rainfall and wave height frequencies of ocean waves. The classifications yield classes with significantly different behavior depending on the classification goal. It was shown in the

TABLE 4 | Six of the most extreme wave events on record and their associated CPs for the period 1992–2009.

Fig	Date	CP	H_s (m)
(a)	19/03/2007	CP03	8.50
(b)	05/05/2001	CP05	6.30
(c)	18/03/2001	CP03	5.92
(d)	03/04/2001	CP03	5.66
(e)	23/09/1993	CP03	5.64
(f)	19/03/2001	CP08	5.63





precipitation part of the study, that using different stations of a climatic region for the classification objective leads to different classifications, but the performance of these classifications is nevertheless very similar. Different classifications for the same

region are more similar than classifications corresponding to different regions. This justifies the use of different classifications for the different regions and variables, because they add discriminatory power to the modeling procedure.

In the case of the wave study, the emphasis was on the statistical link between atmospheric CPs and extreme wave events. The results show that the classification algorithm is able to identify CPs that drive extreme wave events along the KZN coastline. The most frequent CPs associated with wave generation are revealed as low and high pressure anomalies south of the country. These reflect easterly traversing mid-latitude cyclones and their associated high pressures anomalies. The CP class that drives the majority of extreme wave events (labeled CP03) is associated with strong low pressure anomalies east of the country. A high/low pressure coupling drives strong winds toward the coastline and generates large waves. It is not clear what weather regimes are associated with these events since both mid-latitude cyclones and cut-off lows may belong to the CP03 class depending on their location. The classification algorithm does not currently identify CP anomalies as part of an overall structure developing in space and time: each class could be a snapshot in the temporal evolution of (perhaps

distinct) CPs. Therefore, a particular weather system may be associated with a number of classes during its development, which makes for added complexity and requires further research.

There are similarities in the CP classes driven by the two surface variables because in some instances they are linked to similar synoptic weather systems. However, there are also significant differences in the details of the pressure fields. These differences are expected because large ocean waves are typically generated over fetches of the order of thousands of kilometers far off shore, whereas some rainfall is generated by local atmospheric variables including temperature, humidity, wind speed, and radiation over the area of concern.

Acknowledgments

Research leading to this paper was partly supported by the DFG, project number Ba-1150/18-1.

References

- Bárdossy, A., Duckstein, L., and I, I. B. (1995). Fuzzy rule - based classification of atmospheric circulation patterns. *Int. J. Clim.* 15, 1087–1097.
- Bárdossy, A., Stehlik, J., and Caspary, H.-J. (2002). Automated objective classification of daily circulation patterns for precipitation and temperature downscaling based on optimized fuzzy rules. *Clim. Res.* 23, 11–22. doi: 10.3354/cr023011
- Corbella, S., and Stretch, D. (2012a). Predicting coastal erosion trends using non-stationary statistics and process-based models. *Coast. Eng.* 70, 40–49. doi: 10.1016/j.coastaleng.2012.06.004
- Corbella, S., and Stretch, D. (2012b). Shoreline recovery from storms on the east coast of South Africa. *Nat. Hazards Earth Syst. Sci.* 12, 11–22. doi: 10.5194/nhess-12-11-2012
- Corbella, S., and Stretch, D. D. (2012c). The wave climate on the the KwaZulu Natal coast. *J. S. Afr. Inst. Civ. Eng.* 54, 45–54.
- Corbella, S., and Stretch, D. D. (2013). Simulating a multivariate sea storm using Archimedean copulas. *Coast. Eng.* 76, 68–78. doi: 10.1016/j.coastaleng.2013.01.011
- Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The era-interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* 137, 553–597. doi: 10.1002/qj.828
- Hartung, J., Bärbel, E., and Klösener, K. (2005). *Statistik*. München: Oldenburg Verlag.
- Hess, P., and Brezowsky, H. (1952). *Katalog der Großwetterlagen Europas*. Bad Kissingen: Ber. DT. Wetterd. in der US-Zone 33.
- Hewitson, B., and Crane, R. (2002). Self-organizing maps: applications to synoptic climatology. *Clim. Res.* 22, 13–26. doi: 10.3354/cr022013
- Jacobeit, J. (2010). Classifications in climate research. *Phys. Chem. Earth A/B/C* 35, 411–421. doi: 10.1016/j.pce.2009.11.010
- Kruger, A. C. (2004). *Climate of South Africa: Climate Regions*. Technical Report WS45, South African Weather Service, Pretoria.
- Lamb, H. H. (1972). *British Isles Weather Types and a Register of the Daily Sequence of Circulation Patterns, 1861–1971*. London: HMSO; Meteorological Office. Geophysical Memoir No. 116.
- Mather, A., and Stretch, D. (2012). A perspective on sea level rise and coastal storm surge from southern and eastern africa: a case study near Durban, South Africa. *Water* 4, 237–259. doi: 10.3390/w4010237
- Philipp, A., Beck, C., Huth, R., and Jacobeit, J. (2014). Development and comparison of circulation type classifications using the cost 733 dataset and software. *Int. J. Climatol.* doi: 10.1002/joc.3920. [Epub ahead of print].
- Preston-Whyte, R. A., and Tyson, P. D. (1988). *The Atmosphere and Weather of Southern Africa*. Cape Town: Oxford University Press.
- Pringle, J., Stretch, D., and Bardossy, A. (2014). Automated classification of the atmospheric circulation patterns that drive regional wave climates. *Nat. Hazards Earth Syst. Sci.* 14, 2145–2155. doi: 10.5194/nhess-14-2145-2014
- Rossouw, J., Coetzee, L. W., and Visser, C. J. (1982). “A South African wave climate study,” in *IPCC*. 18, 87–107. doi: 10.9753/icce.v18.%25p

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Bárdossy, Pegram, Sinclair, Pringle and Stretch. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.