



*Escuela Técnica Superior de Ingenieros de Caminos,
Canales y Puertos.*
UNIVERSIDAD DE CANTABRIA



SEASONAL FORECAST OF TROPICAL CYCLONE ACTIVITY IN THE SOUTH PACIFIC OCEAN

Trabajo realizado por:
Andrea Pozo Estívariz

Dirigido:
Fernando Méndez Incera
Laura Cagigal Gil
Sara Ortega van Vloten

Titulación:
**Máster Universitario en
Ingeniería de Caminos, Canales y
Puertos**

Santander, mayo 2021

TRABAJO FIN DE MASTER



ABSTRACT

TITLE:

Seasonal Forecast of Tropical Cyclone Activity in the South Pacific Ocean.

AUTHOR:

Andrea Pozo Estívariz

DIRECTORS:

Fernando Méndez Incera

Laura Cagigal Gil

Sara Ortega van Vloten

CALL: May 2021

KEY WORDS

Tropical Cyclone, track, forecast, statistical downscaling, predictor, predictand, daily weather type, aggregated expected mean number of Tropical Cyclones.

The objective of this final master's thesis is the development of a tropical cyclones (TCs) seasonal forecast outlook for a target area in the South Pacific basin (latitudes 0°S-30°S and longitudes 160°W-150°E) through climate-based statistical techniques; which is of particular interest for practical application to predict the future risk and occurrence of TCs and to improve the resilience of Small Islands Developing Countries (SIDS) in South Pacific.

TCs develop over extensive surfaces of warm water and around an atmospheric perturbation when atmospheric conditions are favorable. They are storms characterized by a closed circulation around a low-pressure center that produces strong winds and abundant rainfall. They are generated in the area within 5° to 30° North and South from the equator and then they migrate in a direction away from it.

TCs remain the most destructive natural hazard in the tropical South Pacific, being a cause of mortality and injury and greatly and regularly affecting the sustainability of the SIDS' economies. Therefore, delivering accurate TCs forecasts in time is of key importance.

For carrying out the methodology several databases have been employed. The IBTrACs global database for the TCs information, the NOAA 1/4° daily Optimum Interpolation Sea Surface Temperature, the Daily mean precipitation from the TRMM Multi-Satellite Precipitation Analysis TMPA (3B42) and the NCEP Climate Forecast System Reanalysis (CFSR) for Sea Level Pressure and Mixed Layer Depth and for forecast and reforecast data.

Before building the TCs seasonal forecast, the relationship between the TCs genesis and two large-scale predictors: the annual weather type (AWT) as El Niño, La Niña or neutral year; and the Madden-Julian Oscillation (MJO) is firstly explored.

Then, the focus is on the role of the sea surface temperature (SST) and the mixed layer depth (MLD) as predictor variables for TCs occurrence in the target area. A statistical downscaling model that identifies the patterns of large-scale SST – MLD associated with TCs activity in terms of incidence rate is developed. In this way, a tailor-made predictor, with the mean expected number of TCs as the predictand variable, is built and it can be used with forecast data to predict TC activity.

A weather typing method is proposed for the statistical model. Firstly, historical data for the predictor (SST and MLD) and predictand (mean expected number of TCs) variables is collected. Then, this data is pre-processed with the definition of the temporal domain and scale (a calibration period from 1982 to 2019 in a daily basis) and the spatial domain (the target area discretized in a $\frac{1}{2}^\circ$ grid resolution). The historical databases (SST, MLD and the minimum pressure value of the TC track) are then combined into a tailor-made index predictor. This index predictor is partitioned into a certain number of clusters ($N = 49$), named daily weather types (DWTs), obtained combining three data mining techniques (principal component analysis (PCA), k-means algorithm (KMA) and a similarity criterion for an intuitive visualization of the DWTs). The resultant DWTs are deeply analyzed, as well as their relationship with TC genesis and the original predictor variables.

DWTs chronology, seasonality and annual variability and their link with TC activity is explored. Furthermore, additional predictand variables, daily mean precipitation and Sea Level Pressure; are also transferred into the DWTs in order to better understand and explain the oceanic and atmospheric dynamics that are more favorable for TCs genesis and development.

The last step in building the statistical model consists on defining the statistical relationship between the predictor and the predictand. The target area is divided in squared cells of $8 \times 8^\circ$. For each cell, the mean daily number of TCs going through it is computed (from the historical data) conditioned to the DWT. This way each DWT has a predictand variable map assigned, which allows to generate predictand maps in different time scales for any period of time once the DWTs are known.

After this, the built model is validated for year 2020, not included in the calibration period. SST and MLD data are pre-processed, and the statistical model is applied to obtain the DWTs and from them the correspondingly inferred predictand variable maps.

Finally, the TCs seasonal forecast is validated and applied. For the validation, reforecast data from the predictand variables is preprocessed and a TCs seasonal reforecast for each past TCs seasons is built. For the final forecast application, data from the first day of April 2021 has been used to forecast the TC activity for the next 9 months.

The main conclusions obtained after the completion of this work are displayed below.

Firstly, with respect to the relationship between TCs activity and the large-scale predictors it was found that MJO phases 6,7 and 8 have the highest TC activity and that El Niño, which is more unlikely, has the greatest proportion of TCs with respect of its total days and the greatest proportion of TCs reaching at least category 2.

The TCs genesis activity is generally focused under the following conditions. In areas with index predictor values ranging from 0.60 to 0.79, corresponding to positive anomalies. Then, in the warm SST zone, ($28 - 30^\circ \text{C}$), and where MLD values are smaller than 75 m; corresponding to mild positive SST anomalies and negative MLD anomalies respectively. Moreover, in intense precipitation areas, from 9 to 16.5 mm/day, corresponding generally to positive anomalies. Lastly, in low pressure areas, with 1013 mba or lower values, corresponding generally to negative anomalies.

Finally, regarding the performance of the resultant statistical model, a good agreement estimating the expected TC activity (number and intensity of TCs) is found, not underestimating the TC threat in the area. However, for the forecast period, the performance greatly depends on the quality of the forecast data, which unfortunately, has large room for improvement.



RESUMEN

TÍTULO:

Predicción estacional de la actividad de ciclones tropicales en el Pacífico Sur.

AUTOR:

Andrea Pozo Estívariz

DIRECTORES:

Fernando Méndez Incera

Laura Cagigal Gil

Sara Ortega van Vloten

CONVOCATORIA: Mayo 2021

PALABRAS CLAVE

Ciclón tropical, traza, predicción, downscaling estadístico, predictor, predictando, daily weather type, número medio esperado agregado de ciclones tropicales.

El objetivo de este trabajo de fin de máster es el desarrollo de un modelo de predicción estacional de ciclones tropicales (CTs) para un área objetivo en la cuenca del Pacífico Sur (latitudes 0°S-30°S y longitudes 160°W-150°E); a través de técnicas estadísticas basadas en el clima; que es de particular interés para predecir el riesgo y la ocurrencia futuros de TCs y para mejorar la resiliencia de los Pequeños Países Insulares en Desarrollo (SIDS) en el Pacífico Sur.

Los CTs se desarrollan sobre extensas superficies de agua cálida y alrededor de una perturbación atmosférica cuando las condiciones atmosféricas son favorables. Son tormentas caracterizadas por un sistema de circulación cerrado alrededor de un centro de baja presión que produce fuertes vientos y abundantes precipitaciones. Se generan en la zona comprendida entre 5° y 30° al norte y al sur del ecuador y luego migran alejándose de él.

Los CT siguen siendo el desastre natural más destructivo en el Pacífico Sur tropical, siendo una causa de mortalidad y lesiones y afectando en gran medida y de forma regular a la sostenibilidad de las economías de las SIDS. Por lo tanto, es de vital importancia realizar predicciones precisas de CT a tiempo.

Para llevar a cabo la metodología se han empleado varias bases de datos. La base de datos global IBTrACs para los CTs, la temperatura superficial del mar de interpolación óptima diaria de 1/4° de resolución de la NOAA, la precipitación media diaria del análisis de precipitación multi satélite TRMM TMPA (3B42) y el reanálisis del sistema de previsión climática del NCEP (CFSR) para la presión a nivel del mar y la profundidad de la capa mixta y para los datos de predicción y re-predicción.

Primero, se explora la relación entre la génesis de los CT y dos predictores a gran escala: el tipo de tiempo anual (AWT) como El Niño, La Niña o año neutro; y la Oscilación de Madden-Julian (MJO).

A continuación, se centra la atención en el papel de la temperatura de la superficie del mar (SST) y la profundidad de la capa mixta (MLD) como variables predictoras de la actividad de los CT en la zona objetivo. Se desarrolla un modelo de downscaling estadístico que identifica los patrones de SST-MLD a gran escala asociados a la actividad de los CT en términos de tasa de incidencia. De esta manera, se construye un predictor a medida, con el número medio esperado de CTs como predictando, que puede ser empleado para predecir la actividad ciclónica.

Se propone un método de tipificación meteorológica diaria para el modelo estadístico. En primer lugar, se recogen los datos históricos de los predictores (SST y MLD) y el predictando (número medio esperado de CT). A continuación, estos datos se pre procesan con la definición del dominio temporal (un periodo de calibración de 1982 a 2019 a escala diaria) y el dominio espacial (el área objetivo discretizada a $\frac{1}{2}^{\circ}$ de resolución). Las bases de datos históricas (SST, MLD y el valor de presión mínima de la traza del CT) se combinan entonces en un índice predictor. Este índice se divide en un número de clusters ($N = 49$), denominados daily weather types (DWTs), obtenidos combinando tres técnicas de minería de datos (análisis de componentes principales (PCA), algoritmo de k-means (KMA) y un criterio de similitud para una visualización intuitiva de los DWTs). Las DWTs resultantes se analizan en profundidad, así como su relación con la génesis de CTs y las variables predictoras originales.

Se explora la cronología, estacionalidad y variabilidad anual de las DWTs y su relación con la actividad de CTs. Además, variables predictoras adicionales, la precipitación media diaria y la presión al nivel del mar, también se transfieren a los DWTs para entender y explicar mejor la dinámica oceánica y atmosférica que es más favorable para la génesis y el desarrollo de los CTs.

El último paso en la construcción del modelo estadístico consiste en definir la relación estadística entre el predictor y el predictando. El área objetivo se divide en celdas cuadradas de $8 \times 8^{\circ}$. Para cada celda se calcula el número medio diario de CTs que la atraviesan (a partir de los datos históricos) condicionado al DWT. De esta forma cada DWT tiene asignado un mapa del predictando, lo que permite generar mapas de predicción en diferentes escalas temporales para cualquier periodo de tiempo una vez conocidos los DWTs.

A continuación, se valida el modelo construido para el año 2020, no incluido en el periodo de calibración. Los datos de la SST y MLD son preprocesados, y se aplica el modelo estadístico para obtener los DWTs y a partir de ellos los correspondientes mapas inferidos del predictando.

Por último, se valida y aplica la predicción estacional de los CT. Para la validación, se preprocesan los datos de las variables de predicción y se construye una previsión estacional de los CT para cada una de las temporadas pasadas. Para la aplicación final, se han utilizado los datos del primer día de abril de 2021 y se ha creado la predicción de la actividad de CTs para los próximos 9 meses.

Las principales conclusiones que se obtienen tras la realización de este trabajo son las siguientes. En primer lugar, con respecto a la relación entre la actividad ciclónica y los predictores de gran escala; las fases 6,7 y 8 de la MJO tienen la mayor actividad ciclónica y El Niño, que bastante improbable, tiene la mayor proporción de actividad ciclónica con respecto a sus días totales y la mayor proporción de CTs que alcanzan al menos la categoría 2.

La génesis de los CTs ocurre generalmente bajo las siguientes condiciones. En valores del índice predictor entre 0,60 y 0,79, correspondientes a anomalías positivas. En la zona de SST cálida ($28 - 30^{\circ}\text{C}$), y donde los valores de MLD son menores de 75 m; correspondientes a anomalías positivas leves de SST y negativas de MLD respectivamente. Además, en zonas de precipitación intensa, de 9 a 16,5 mm/día, correspondientes generalmente a anomalías positivas. Por último, en zonas de bajas presiones, con valores de 1013 mba o inferiores, correspondientes generalmente a anomalías negativas.

Finalmente, en cuanto al desempeño del modelo estadístico resultante, se encuentra un buen acuerdo en la estimación de la actividad esperada de los CT (número e intensidad), no subestimando la amenaza de los CT en la zona. Sin embargo, para el periodo de predicción, el rendimiento depende en gran medida de la calidad de los datos de predicción, que desafortunadamente, tiene un gran margen de mejora.

Index

1	Introduction and Motivation	1
2	Databases	3
2.1	Tropical cyclones	3
2.2	Sea Surface Temperature	4
2.3	CFS (Climate Forecast System)	5
2.4	Daily mean precipitation	6
2.5	Madden-Julian Oscillation (MJO)	6
3	Methodology overview	6
4	Large scale predictors	9
4.1	The Madden-Julian Oscillation (MJO)	9
4.2	El Niño Southern Oscillation (ENSO)	11
4.3	MJO and AWT relationship with TC genesis.....	13
5	Tailor-made predictor	17
5.1	Index Predictor definition and building.....	17
5.2	Statistical downscaling method	21
5.2.1	DWT classification	21
5.2.2	DWTs seasonality, annual variability and chronology.....	30
5.2.3	Relationship predictor-predictand	36
6	Additional predictand variables	42
7	Model validation	47
8	Forecast	52
8.1	Forecast validation	52
8.2	Forecast application	56
9	Bibliography	58
10	Annex 1, The Jupyter book	61

Figures Index

Figure 1 Damages caused by TCs in Samoa (left) and Tonfa (right) (Source: Samoa Meteorological Services webpage)	1
Figure 2 Target area in the South Pacific basin.....	3
Figure 3 South Pacific TCs tracks included in IBTrACs database.	4
Figure 4 Ocean temperature profile through the ocean layers within the mid-latitude regions. 5	
Figure 5 Statistical model structure.	7
Figure 6 Methodology for building TCs seasonal forecast.....	7
Figure 7 Timeline.	8
Figure 8 Difference from average rainfall for all MJO events from 1979-2012 for November-March for the eight MJO phases (Gottschalck, 2014).....	10
Figure 9 Phase-space diagram of the RMM index (Wheeler & Hendon, 2004) showing daily phase (quadrant) and magnitude (distance from center) of the MJO from 1 Jun 1974 through 31 Mar 2014. Colors indicate thresholds of activity: IA (blue, RMM < 1.0), A (green, RMM >= 1.0 and < 1.5), VA (brown, RMM >=1.5 and < 2.5), and EA (red, RMM >= 2.5)	11
Figure 10 Typical behavior of the couple system of ocean and atmosphere during El Niño, La Niña and Neutral conditions in the equatorial Pacific (Source: Australian Bureau of Meteorology webpage).....	12
Figure 11 Number scheme and color for the ENSO phases.	13
Figure 12 Density histogram of TC genesis according to AWT.....	14
Figure 13 TCs historical tracks transferred to AWT+MJO combinations according to their genesis points.....	14
Figure 14 TCs equal or greater than category 2 historical tracks transferred to AWT+MJO combinations according to their genesis points.	15
Figure 15 MJO 6,7 and 8 phases.....	15
Figure 16 Daily Mean Precipitation transferred to MJO - AWT combinations.	16
Figure 17 Daily Mean Precipitation anomalies transferred to MJO - AWT combinations.....	17
Figure 18 Target area and islands in the South Pacific basin.	18
Figure 19 Predictor grid nodes for the target area.	19
Figure 20 Predictand-predictor datasets.	19
Figure 21 Predictand-predictor datasets discretized in intervals of 0.5 m and 0.5 °C for the MLD and the SST respectively.	20
Figure 22 Index predictor.	21
Figure 23 Number of PCs vs Percentage of variance explained.....	22
Figure 24 First four modes of the data space.	22
Figure 25 Index data in the PC space. The black dots represent the N = 49 clusters obtained using K-means.	23
Figure 26 DWTs bidimensional lattice organization and corresponding colors.....	24
Figure 27 DWTs.	24
Figure 28 DWTs anomalies.....	25
Figure 29 Days associated to cluster DWT = 5.	26
Figure 30 Days associated to cluster DWT = 41.	26

Figure 31 SST and MLD transferred into the DWTs.	28
Figure 32 SST and MLD anomalies transferred into the DWTs.	29
Figure 33 Histogram of TCs genesis per month during the calibration period.	30
Figure 34 DWTs seasonality.	31
Figure 35 DWTs monthly chronology during all the calibration period.	33
Figure 36 DWTs chronology.	34
Figure 37 DWTs according to AWT.	34
Figure 38 TCs transferred into DWTs according to genesis point.	36
Figure 39 TCs transferred into DWTs according to each segment of the track.	37
Figure 40 Absolute number of TCs going through the target area in 8x8° cells during all the calibration period for each DWT.	38
Figure 41 Mean expected number of TCs going through the target area in 8x8° for each DWT.	39
Figure 42 Monthly aggregated mean expected number of TCs for all the calibration period. ...	40
Figure 43 Monthly aggregated mean expected number of TCs for year 2015 of the calibration period.	41
Figure 44 SLP fields transferred into the DWTs.	43
Figure 45 SLP fields anomalies transferred into the DWTs.	44
Figure 46 Precipitation hindcast (TRMM) transferred into the DWTs.	45
Figure 47 Precipitation hindcast (TRMM) anomalies transferred into the DWTs.	46
Figure 48 Index predictor data in the first three PCs space.	48
Figure 49 Validation period (year 2020) DWTs chronology (left) and DWTs bidimensional lattice and colors (right)	48
Figure 50 Cluster comparison between the calibration and the validation period.	49
Figure 51 Monthly aggregated mean expected number of TCs and historical tracks of year 2020, the validation period.	50
Figure 52 Aggregated mean expected number of TCs and historical tracks for the whole validation period (year 2020).	51
Figure 53 Full season forecasts from past TCs seasons.	53
Figure 54 DWTs mean probabilities for the full TCs season.	55
Figure 55 Forecast from day 1/04/2021.	57

1 Introduction and Motivation

The objective of this final master's thesis is the development of a tropical cyclones (TCs) seasonal forecast outlook for a target area in the South Pacific basin through climate-based statistical techniques; which is of particular interest for practical application to predict the future risk and occurrence of TCs and to improve the resilience of Small Islands Developing Countries (SIDS) in South Pacific.

TCs develop over extensive surfaces of warm water and around an atmospheric perturbation when atmospheric conditions are favorable. They are generated in the area within 5° to 30° north and south from the equator and then they migrate in a direction away from the equator. In meteorological terms, TCs are storm systems characterized by a closed circulation around a low-pressure center that produces strong winds and abundant rainfall. TCs can indistinctly be referred as typhoons or hurricanes depending on the place's terminology since there are originated in many regions around the world, generally in intertropical regions.

In terms of both numbers of people affected and overall financial costs, TCs remain the most destructive natural hazard in the tropical South Pacific. Firstly, they pose an annual threat to the livelihoods of the inhabitants of the study zone, being a cause of mortality and injury. Additionally, they greatly and regularly affect the sustainability of the islands' economies, generating massive financial losses.

Damages caused by TCs are due to high wind speeds, and by associated heavy rainfalls and generated storm surges. The costs of recovery and consequent dependency of aid honors have lasting political, economic, and social effects (Barnett & Campbell, 2010).

For instance, TC Ofa, which struck Samoa in 1990 or TC Kina which struck Fiji in 1993, caused over US\$100 million in damage (Campbell, 1997) (Olsthoorn, et al., 1999). Furthermore, in 2007, flooding caused by TC Guba caused 149 deaths in the Oro province of Papua New Guinea and 58,000 people required food relief and other assistance (IFRC, 2008).



Figure 1 Damages caused by TCs in Samoa (left) and Tonfa (right) (Source: Samoa Meteorological Services webpage)

Therefore, delivering accurate TCs forecasts in time is of key importance when it comes to saving human lives and reducing economic loss. Difficulties arise because the geographical and climatological characteristics of the various cyclone formation basins are not similar. Therefore,

the chosen forecasting technique needs to be applied together with basin-specific techniques to be accurate.

As cyclone track is governed by a range of factors variations in weather conditions, wind pressure, sea surface temperature, air temperature, ocean currents, and the earth' rotational force—the Coriolis force, it is a formidable task to combine these parameters and produce reliable and accurate forecasts (Roy & Kvordányi, 2012)

In the recent decade, statistical models for seasonal prediction of TC activity have been developed for the South Pacific basin. Some of them have explored the relationship between large-scale climatic drivers such as El Niño-Southern Oscillation (ENSO) or the Madden-Julian Oscillation (MJO). In (Boucharel, et al., 2016) different controls of TC activity in the Eastern Pacific for two types of El Niño are investigated; in (Diamond, et al., 2012) the linkage between the ENSO and the Southwest Pacific TC activity climatology is explored and in (Stephens & Ramsay, 2014) the influence of the ENSO and projected climate change in the extreme TC wave climate in the South Pacific Ocean is studied. On the other hand, (Diamond & Renwickb, 2015) explored the climatological relationship between TCs in the Southwest Pacific Ocean and the MJO on interannual time scale.

Others have used different large oceanic variables, especially the sea surface temperature (SST) as predictors for the TCs activity. (Balaguru, et al., 2013) concludes that in the Northeast Pacific hurricane activity at interannual timescales the SST and the thermocline depth (TD) and the SST have a very relevant influence. In fact, the SST has been proven to have a key role in modulating the whole life cycle of the TCs (Mandal, et al., 2007) (Mohanty, et al., 2016) or more specifically in their genesis (Tory & Dare, 2015) and specially in their intensity (Evans, 1993) (Lavender, et al., 2018) (Arora & Dash, 2016).

The target area for this work comprises a part of the South Pacific basin. Specifically, from latitude 0°S to 30°S and from longitude 160°W to 150°E and it can be observed in [Figure 2.](#)

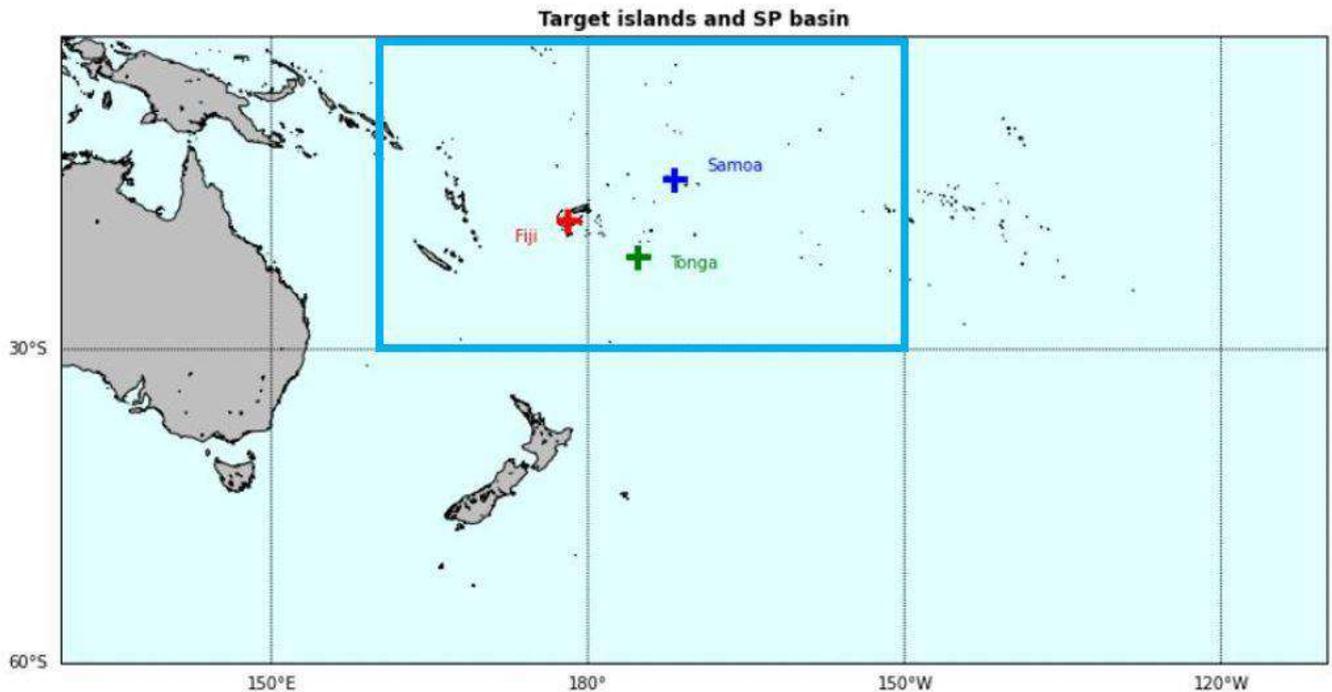


Figure 2 Target are in the South Pacific basin.

2 Databases

2.1 Tropical cyclones

The IBTrACs (International Best Track Archive for Climate Stewardship) database has been used, which compiles an inventory of the TCs reported worldwide with their characteristics (Knapp, et al., 2010). It was fostered as a part of a project whose purpose was to generate a global database through the summary of the best register obtained by different Meteorological Centers: each World Meteorological Organization (WMO) Regional Specialized Meteorological Centers (RSMCs) and Tropical Cyclone Warning Centers (TCWCs); as well as other national agencies.

Among its advantages, it contains the most complete database of historic TCs which is available for the public, it combines as mentioned before the TCs information from different sources, it makes easier data processing by providing the data in only one format in one dataset; and it periodically (each 3 weeks) updates the database including the most recent events.

Some of the variables provided by this dataset at 6-h intervals are the position of the TC along its track (longitude and latitude), the dates of occurrence, the minimum pressure in the low-pressure center, the maximum velocity of the winds, the basin, the name; and the season in which it is included the cyclone. Data for some basins are available as early as 1850.

Generally, TCs are classified according to their intensity, characterized in the Saffir-Simpson scale through the value of the minimum pressure and the velocity of the winds. In [Table 1](#) this classification is shown and, additionally, each category has been assigned a color, which will represent it in all the Figures of this work.

	Mean wind speed (km/h)	Minimum pressure (hPa)	Stormy wave (m)	Damage level
Tropical depression	< 60	990 - 1000	< 1.0	weak
Tropical gust	60 - 120	980 - 990	< 1.0	small
Category 1 cyclone	120 - 150	970 - 980	1.0 - 1.5	small
Category 2 cyclone	150 - 180	965 - 970	1.5 - 2.5	moderate
Category 3 cyclone	180 - 210	945 - 965	2.5 - 4	intense
Category 4 cyclone	210 - 240	920 - 945	4.0 - 6.0	extreme
Category 5 cyclone	> 240	< 920	> 6.0	catastrophic

Table 1 Saffir-Simpson classification scale of tropical cyclones. Each category (corresponds to a row) is colored according to the color assigned to it.

All the TCs tracks going through the South Pacific basin included in this database since the first records, colored with their corresponding highest category reached is shown in [Figure 3](#). The target area is highlighted in blue.

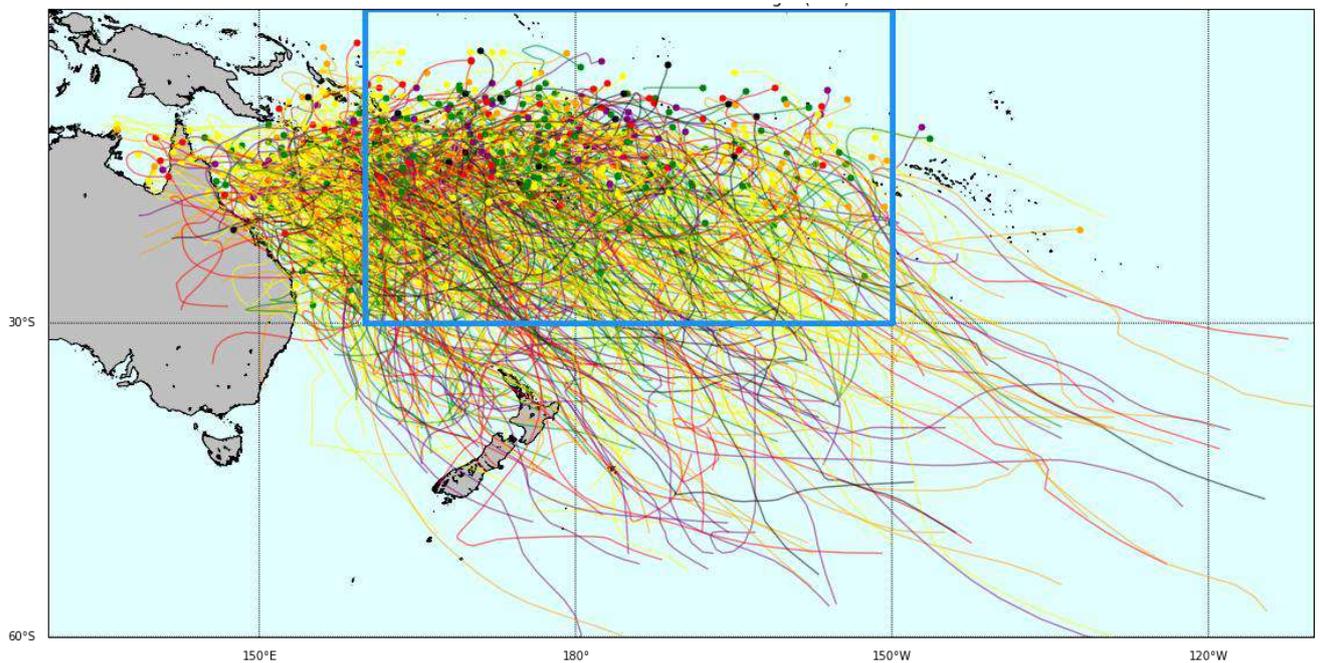


Figure 3 South Pacific TCs tracks included in IBTrACs database.

2.2 Sea Surface Temperature

The NOAA 1/4° daily Optimum Interpolation Sea Surface Temperature or daily OISST is an analysis constructed by combining observations from different platforms (satellites, ships, buoys and Argo floats) on a regular global grid. A spatially complete SST map is produced by interpolating to fill in gaps. The methodology includes bias adjustment of satellite and ship observations (referenced to buoys) to compensate for platform differences and sensor biases.

2.3 CFS (Climate Forecast System)

It is a model which represents the global interaction between Earth's oceans, land, and atmosphere. Produced by several dozen scientists under guidance from the National Centers for Environmental Prediction (NCEP), this model offers hourly data with a horizontal resolution down to 1/2° (approximately 56 km) around Earth for many variables. CFS uses the latest scientific approaches for taking in, or assimilating, observations from data sources including surface observations, upper air balloon observations, aircraft observations, and satellite observations.

The NCEP Climate Forecast System Reanalysis (CFSR) was designed and executed as a global, high-resolution, coupled atmosphere-ocean-land surface-sea ice system to provide the best estimate of the state of these coupled domains over the 32-year period of record from January 1979 to March 2011. It has been extended as an operational real-time product. The CFSv2 Operational Analysis goes from April 2011 onwards. Mixed Layer Depth, Mean Sea Level Pressure have been downloaded from CFSR. From these two variables there is data from 1982, which involves 37 years of hindcast data.

The mixed layer is the layer in which there is active turbulence and mixing of oceanic waters due to winds, heat fluxes, evaporation and salinity fluxes. It changes its depth based on seasons and latitude. Below it there is the thermocline, the oceanic water layer in which water temperature decreases rapidly with increasing depth. (TimeScavengers, 2021).

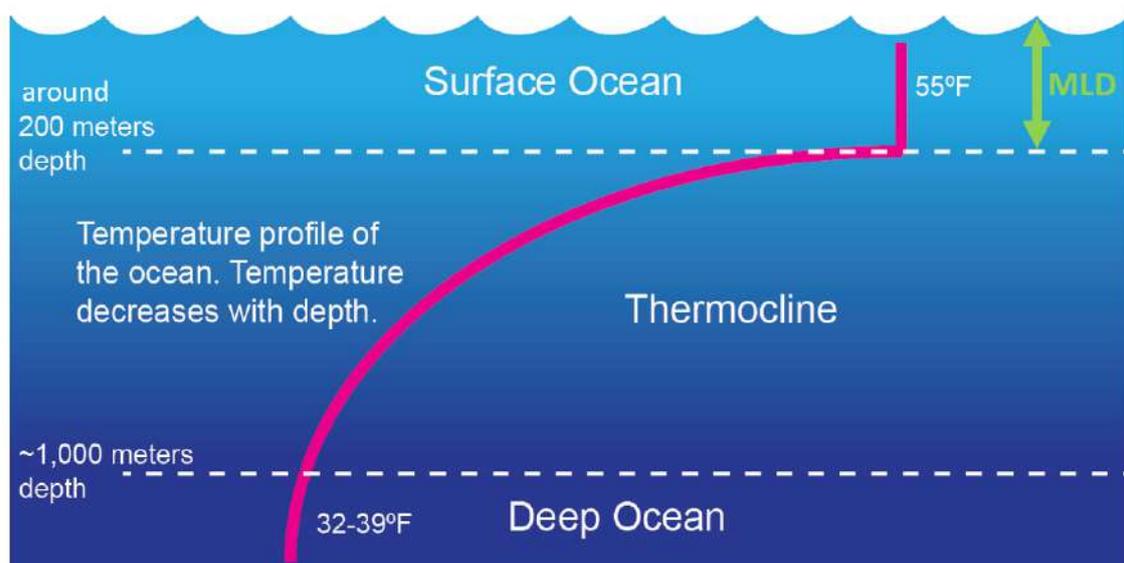


Figure 4 Ocean temperature profile through the ocean layers within the mid-latitude regions

The forecast and reforecast data of the Sea Surface Temperature and the Mixed Layer Depth has been also downloaded from this database, from the CFS operational 9 months forecast section.

2.4 Daily mean precipitation

TRMM was a research satellite in operation from 1997 to 2015, designed to improve the understanding of the distribution and variability of precipitation within the tropics as part of the water cycle in the current climate system. By covering the tropical and sub-tropical regions of the Earth, TRMM provided much needed information on rainfall and its associated heat release that helps to power the global atmospheric circulation that shapes both weather and climate. This daily accumulated precipitation product is generated from the research-quality 3-hourly TRMM Multi-Satellite Precipitation Analysis TMPA (3B42). Simple summation of valid retrievals in a grid cell is applied for the data day. The result is given in (mm).

2.5 Madden-Julian Oscillation (MJO)

MJO data is downloaded from the Australian Bureau of Meteorology webpage. The database goes from 1974 up to actual time. It includes the MJO date (day, month and year), RMM1, RMM2, phase (1 to 8) and amplitude.

3 Methodology overview

In this work, it will first explore in [Section 4](#) the relationship between the TCs genesis and the annual weather type (AWT) as El Niño, La Niña or neutral year and the Madden-Julian Oscillation (MJO); as large-scale predictors.

Then in [Section 5](#), the focus will be on the role of sea-surface temperature (SST) and the mixed layer depth (MLD) as predictors for TCs occurrence in the target area. A statistical downscaling model that identifies the patterns of large-scale SST – MLD associated with TCs activity in terms of incidence rate is developed. This way a tailor-made predictor, with the number of TCs as the predictand variable is built and it can be employed with forecast data to estimate the TC risk. In [Figure 5](#) the scheme representing the statistical model with the predictand and the predictor can be seen:

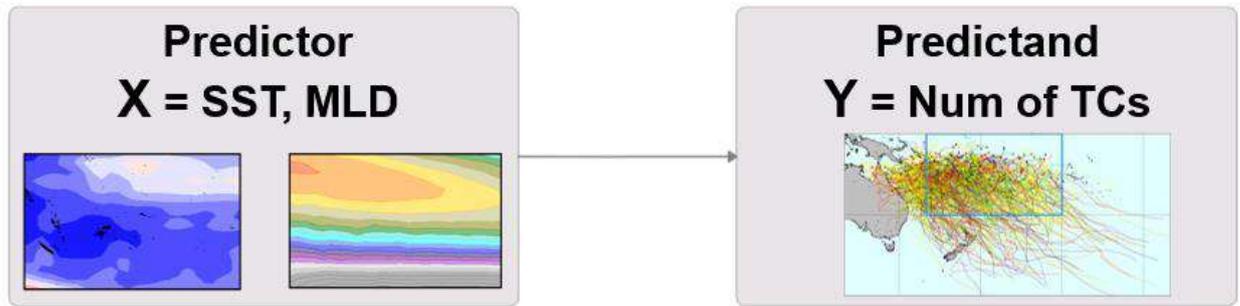


Figure 5 Statistical model structure.

In the following figure a flowchart including the steps involved in the development of the tailor-made predictor is shown.

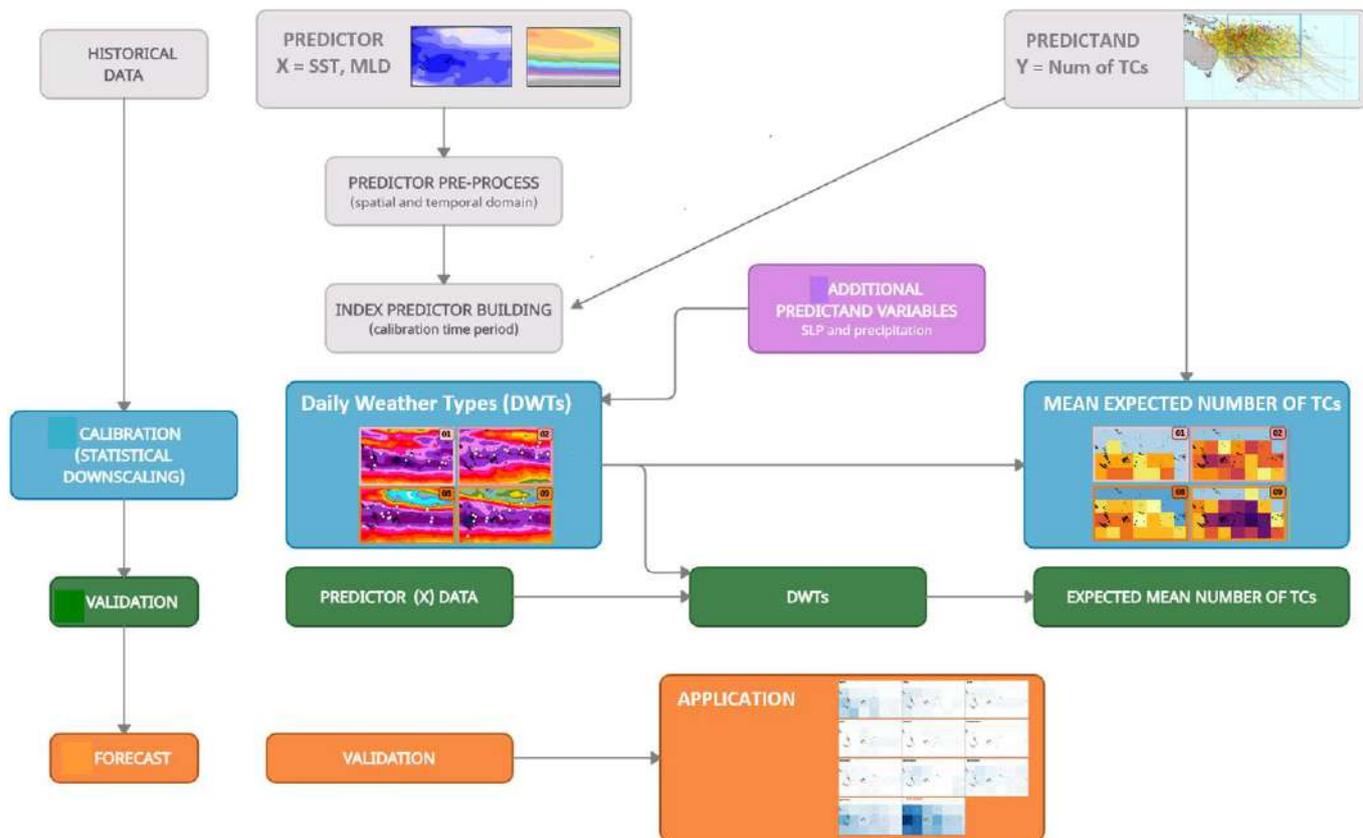


Figure 6 Methodology for building TCs seasonal forecast.

The aim of the statistical downscaling method is to estimate the number of TCs (predictand variable) from SST and MLD (predictor variables), based on a statistical relation. A weather typing method is proposed for this. The main steps of the methodology, included in [Figure 6](#) through a flowchart, are the following ones:

1. Collection of historical data for the predictor and predictand variables.
2. Predictor pre-processing with the definition of the temporal spatial domains.
3. Index predictor definition and building.

4. Classification of the index predictor conditions in 49 clusters named Daily Weather Types (DWTs).
5. Establishing the statistical relationship between the predictor and predictand.
6. Additional explanatory variables are transferred to the DWTs.
7. Validation of the statistical model.
8. Forecast validation and application.

Furthermore, the timeline for this work with the same colors of the flowchart and the steps is included below:

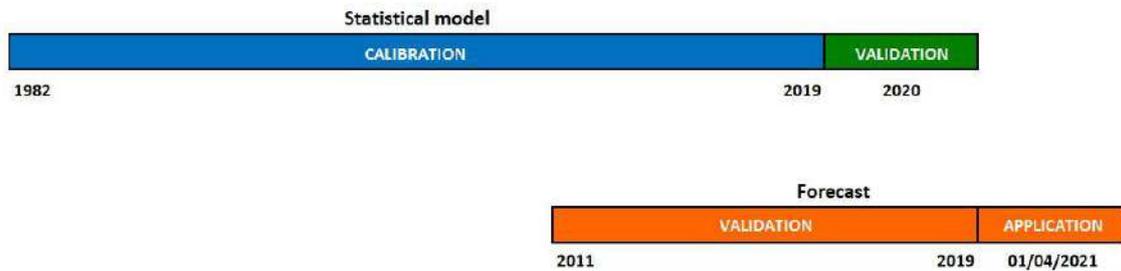


Figure 7 Timeline.

The three two steps are developed in [Section 5.1](#). Firstly, historical data of both predictor (SST and MLD) and predictand (number of TCs) variables is downloaded from the previously mentioned databases. Then the data is pre-processed with the definition of the temporal domain and scale (a calibration period from 1982 to 2019 as seen in [Figure 7](#) in a daily basis) and of the spatial domain (the target area observed in [Figure 2](#), discretized in a grid of $\frac{1}{2}^\circ$ resolution). From the predictand data, only the minimum pressure of the TC track is kept. Then the already processed historical databases, are combined into a tailor-made index, the index-predictor of the statistical model.

In the fourth and fifth steps the statistical downscaling method is deeply developed, which corresponds to [Section 5.2](#).

The seasonal index predictor is first partitioned into a certain number of clusters (in this case 49), named daily weather types (DWTs). Each one represents a daily synoptic pattern of the index. ([Section 5.2.1](#)). This DWT classification is obtained combining three data mining techniques. Firstly, a principal component analysis (PCA) is applied to the index predictor in order to reduce the data dimensionality and simplify the classification process. Then, the predictor in the PCA temporal-spatial domain is clustered using the k-means algorithm (KMA). The KMA divides the data space into a number (N) of clusters (in this case $N = 49$), each one defined by a prototype and formed by the data for which the prototype is the nearest (Camus, et al., 2011). Finally, the set of DWTs is organized in a bidimensional lattice using a similarity criterion so when plotted the similar DWTs remain next to each other, allowing an intuitive visualization.

DWTs chronology, seasonality and intra annual and interannual variability; and its relationship with TCs activity is analyzed through different figures in [Section 5.2.2](#).

The step in the building the statistical model consists of defining the statistical relationship between the predictor and the predictand for the temporal domain chosen (1982 to 2019) This

is done in [Section 5.2.3](#). The predictand is defined as the expected number of TCs going through the target area; which is discretized in square cells of $8 \times 8^\circ$. For each DWT, from the historical TCs tracks data, it is computed the mean expected number of TCs going through the area. This way each DWT has one map of expected number of TCs assigned.

Furthermore, for a more complete analysis, in [Section 6](#) additional predictand variables (Sea Level Pressure and daily mean precipitation) are transferred into the DWTs in order to better understand and explain the oceanic and atmospheric dynamics that are more favorable for TCs genesis and development.

After the development of the seasonal predictor, the statistical model is validated, which is included in [Section 7](#). SST and MLD data from the year 2020 ([Figure 7](#)) is downloaded and preprocessed. Then the statistical model is applied and the corresponding DWTs and expected number of TCs maps are obtained.

Finally, the TCs seasonal forecast is validated and applied in the last section, [Section 8](#). For the validation, CFS reforecast data from the predictand variables SST and MLD from the year 2011 to 2019 is downloaded and processed. A seasonal TCs forecast for each TCs season (which includes from November of the first year to April of the following year) included that time period ([Figure 7](#)) is built applying the statistical model. The outcome from the hindcast and the forecast data is compared. With respect to the application, data from the same forecast source from the first of April of this year (2021) is obtained and the TCs seasonal forecast is obtained in the same way.

4 Large scale predictors

The relationship between large scale circulations, El Niño Southern Oscillation and the Madden-Julian Oscillation, and tropical cyclone climatology has been explored.

4.1 The Madden-Julian Oscillation (MJO)

The MJO is an eastward moving disturbance of clouds, rainfall, winds, and pressure that crosses the planet in the tropics and returns to its initial starting point in cycles of approximately 30 or 60 days. It is the dominant mode of atmospheric intra seasonal variability in the tropics (Hendon & Salby, 1994)

The MJO consists of two parts, or phases: one is the enhanced rainfall (or convective) phase and the other is the suppressed rainfall phase. These two phases produce opposite changes in clouds and rainfall and this entire dipole (what means having two main opposing centers of action) propagates eastward. Strong MJO activity often divides the planet into two halves: one half within the enhanced convective phase and the other half in the suppressed convective phase.

An area of enhanced tropical rainfall is first apparent over the western Indian Oceans, which spreads eastwards into the warm waters of the tropical Pacific. This pattern of tropical rainfall tends to lose its identity as it moves over the cooler waters of the eastern Pacific, before reappearing at some point over the Indian Ocean again. The convective phase is followed by a dry phase, with no rainfall. In each cycle can be divided in 8 phases, that can be observed in [Figure 8](#) through the rainfall anomalies for all MJO events from 1979 to 2012. For example, in phase 6 there is enhanced convection over the western Pacific and suppressed convection over the Indian Ocean. Then in phase 2, it is the opposite way around.

The green color indicates above-average rainfall corresponding to the extent of the enhanced convective phase, and the brown shading shows below-average rainfall denoting the suppressed convective phase. The eastward shifting of shaded areas with each successive numbered phase going from the top to the bottom can be noticed.

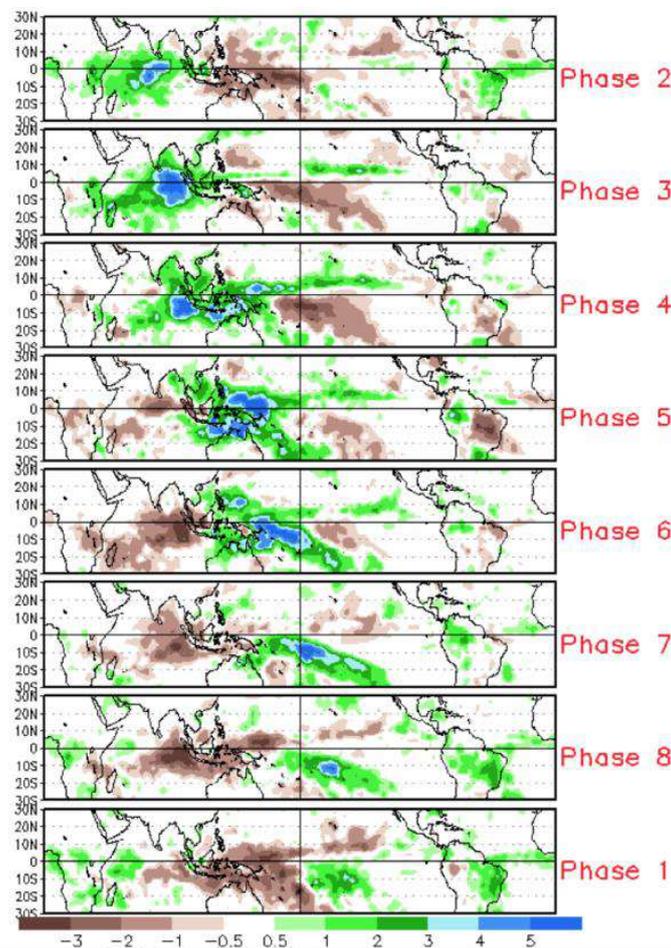


Figure 8 Difference from average rainfall for all MJO events from 1979-2012 for November-March for the eight MJO phases (Gottschalck, 2014)

Finally, in [Figure 9](#) the MJO events from 1974 to 2019 are represented through the RMM index, which is a combined cloudiness- and circulation-based index that has been frequently used for real-time prediction and definition of the MJO (Wheeler & Hendon, 2004).

This diagram shows both geographical location (phase) and amplitude of the MJO (defined as $(RMM1^2 + RMM2^2)^{1/2}$), depending on the quadrant and radial distance from the center of each

point on the diagram. Time progression through it is typically in a counterclockwise sense, representing the eastward propagation of the MJO.

Different studies have quantified the intensity of the MJO. According to (Wheeler & Hendon, 2004) “active” MJO corresponds to amplitude greater than one and “more active” is one where the RMM amplitude is greater than 2. This criterion is commonly adopted.

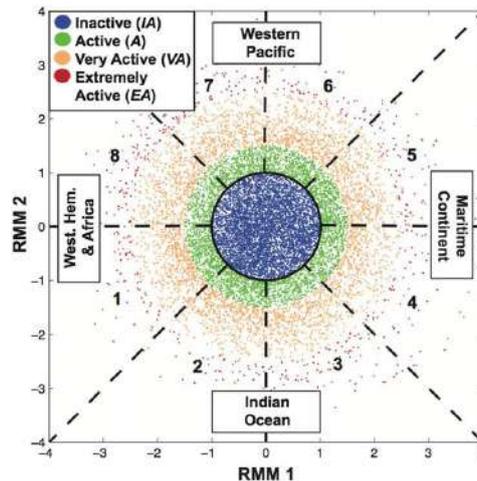


Figure 9 Phase-space diagram of the RMM index (Wheeler & Hendon, 2004) showing daily phase (quadrant) and magnitude (distance from center) of the MJO from 1 Jun 1974 through 31 Mar 2014. Colors indicate thresholds of activity: IA (blue, $RMM < 1.0$), A (green, $RMM \geq 1.0$ and < 1.5), VA (brown, $RMM \geq 1.5$ and < 2.5), and EA (red, $RMM \geq 2.5$)

4.2 El Niño Southern Oscillation (ENSO)

ENSO is one of the most important climate phenomena on Earth due to its ability to change the global atmospheric circulation. It can lead to changes in sea level pressures, sea surface temperatures, precipitation and winds across the globe. ENSO describes the natural interannual variations in the ocean and atmosphere in the tropical Pacific, the zone where it is included the target area. This interaction between the atmosphere and ocean is the source of a periodic variation between below-normal and above-normal sea surface temperatures and dry and wet conditions along the years (International Research Institute for climate society, s.f.). The tropical ocean affects the atmosphere above it and the atmosphere influences the ocean below it. For instance, one layer of the Pacific Ocean that is influenced by ENSO is the thermocline.

This climate phenomenon can be classified in three phases or states. The two extreme phases, “El Niño” and “La Niña,” involve specific changes in the ocean and the atmosphere. Then “Neutral” is in the middle of both.

The term coupled system is generally used to describe the interaction between the atmosphere and the ocean, which is a crucial part of El Niño and La Niña events. During El Niño, sea level pressure tends to be lower in the eastern Pacific and higher in the western Pacific while the opposite tends to occur during a La Niña (International Research Institute for climate society, s.f.).

An **El Niño** state occurs when there is a warming of the central and eastern tropical Pacific Ocean of the ocean surface, or above-average sea surface temperatures. In Indonesia, rainfall tends to decrease, while it increases over the tropical Pacific Ocean. Additionally, the usually present east to west low-level surface winds along the equator weaken and in some cases an anomalous west to east flow (in the other direction) develops.

La Niña involves a cooling of the ocean surface, or below-average sea surface temperatures in the central and eastern tropical Pacific Ocean. The rainfall trend is the opposite of the one happening during El Niño (rainfall increases over Indonesia and it decreases over the tropical Pacific Ocean). The normal easterly winds along the equator not only do they not get weaker, like during el Niño, but they can even intensify.

Finally, during neutral ENSO conditions the sea surface temperature in the tropical Pacific is generally around the average. However, the ocean can look like it is in the El Niño or La Niña phase, but the atmosphere is not playing along (or vice versa). Then, surface trade winds blow westward across the equatorial Pacific Ocean and if they are blowing against the ocean's surface, they cause a westward current.

In [Figure 10](#) it can be observed the common behavior of ocean-atmosphere coupled system during Neutral, El Niño and La Niña conditions in the equatorial Pacific.

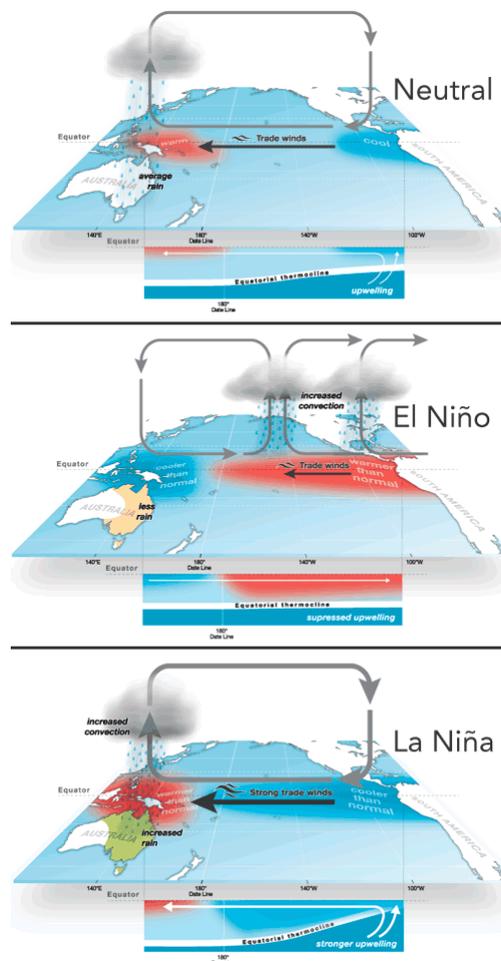


Figure 10 Typical behavior of the couple system of ocean and atmosphere during El Niño, La Niña and Neutral conditions in the equatorial Pacific (Source: Australian Bureau of Meteorology webpage).

Lastly, [Figure 11](#), shows as a summary the SST pattern over the equatorial Pacific Ocean from longitude 120°E to 280°E during the year for each ENSO phase. El Niño is assigned AWT 0 and La Niña AWT 5, the rest are the intermediate phases. Each AWT has also be given a color. These colors and number scheme will be used during this work.

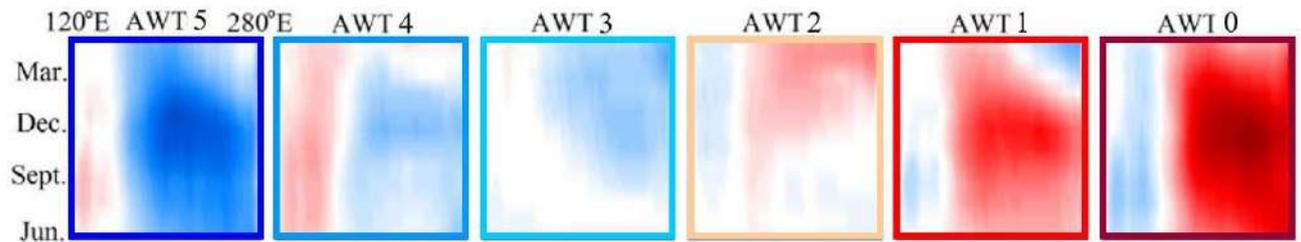


Figure 11 Number scheme and color for the ENSO phases.

4.3 MJO and AWT relationship with TC genesis

To explore the relationship between the TCs genesis and AWT and the MJO, as large-scale predictors; the TC tracks are classified into a AWT - MJO combination according to the date of their genesis. This has been done with all the TC historical tracks from 1979 (when the data collection process in IBTrACs is assumed to be homogenized) to 2019 and then filtering the tracks keeping only the ones reaching category equal or greater than 2. The resulting figures are shown below.

The blue background color corresponds to the probability of having that specific combination of AWT and MJO.

The probability within the same AWT barely changes from one MJO phase to another. However, it varies a lot from AWT to another. Therefore, to help with the analysis of the following figures, the following density histogram of TC genesis according to AWT is displayed.

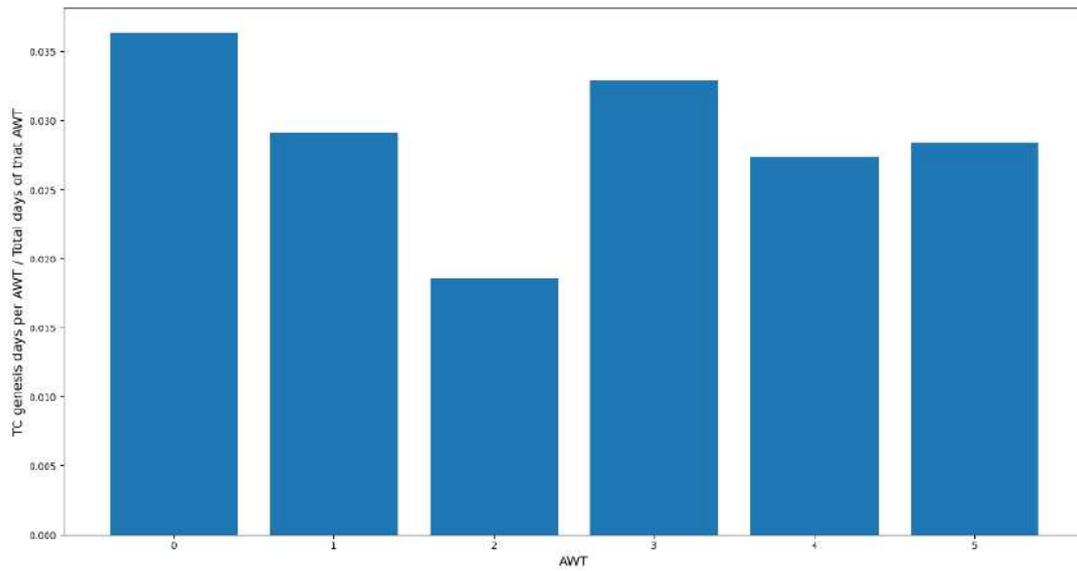


Figure 12 Density histogram of TC genesis according to AWT.

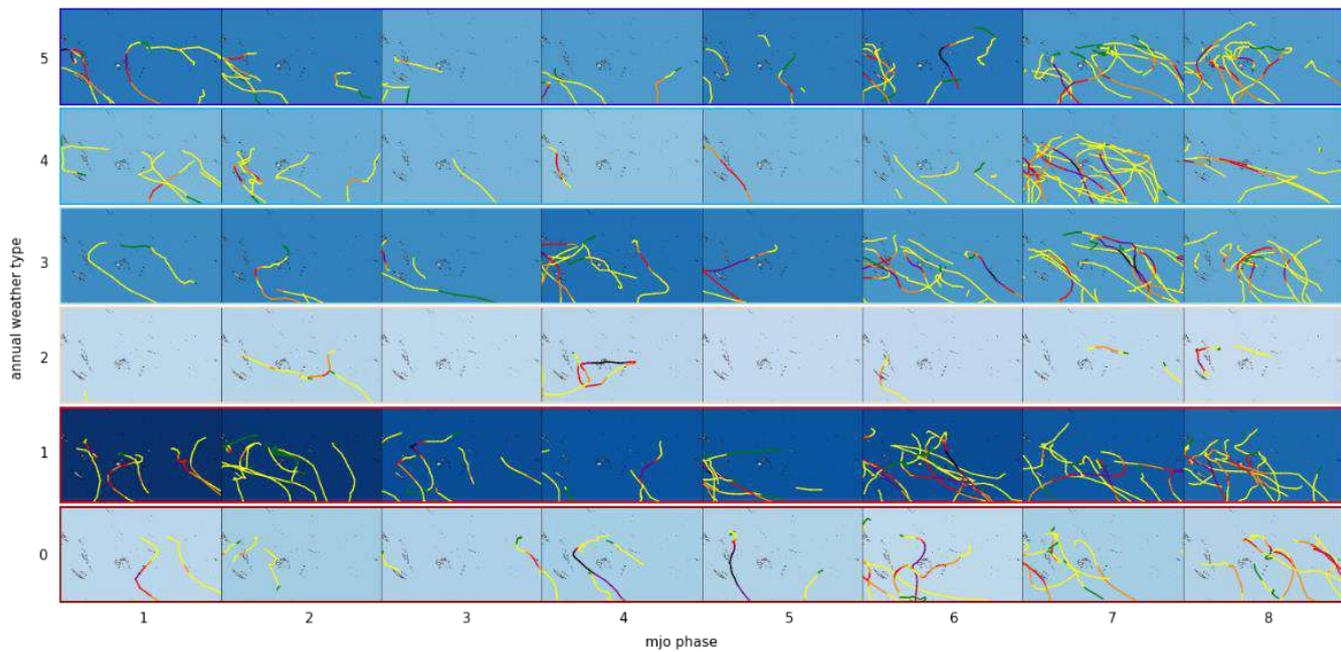


Figure 13 TCs historical tracks transferred to AWT+MJO combinations according to their genesis points.

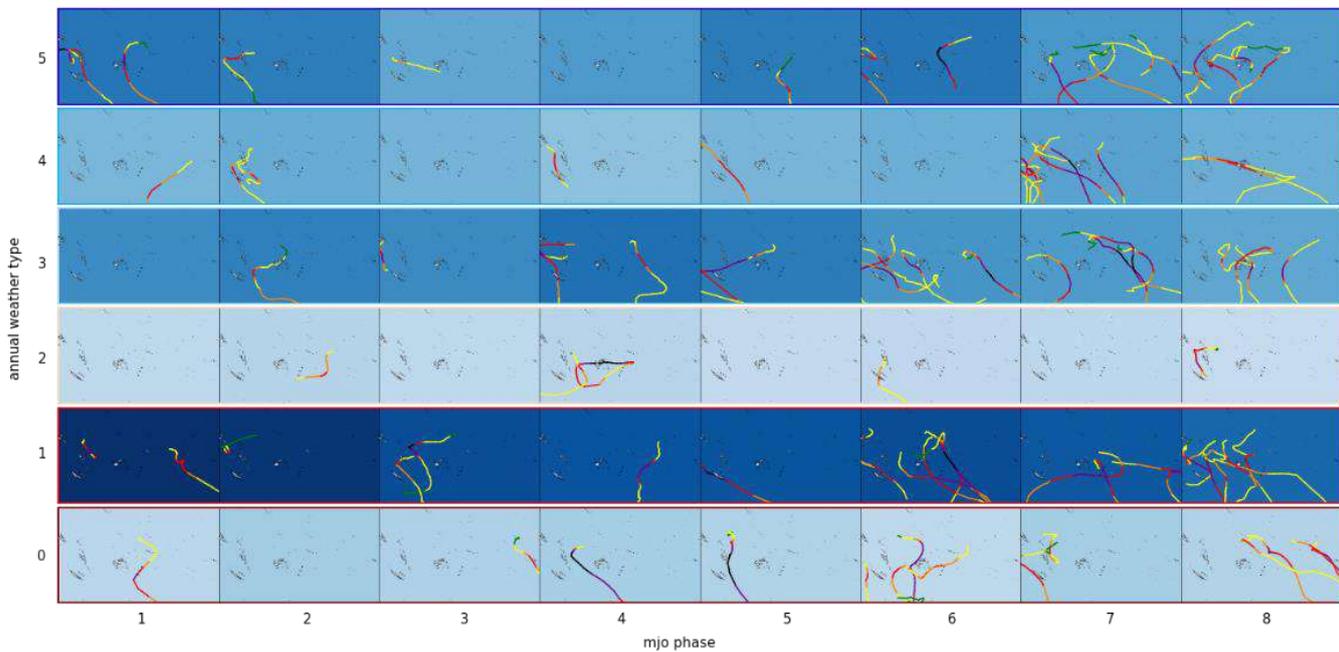


Figure 14 TCs equal or greater than category 2 historical tracks transferred to AWT+MJO combinations according to their genesis points.

In [Figures 13](#) and [14](#) it can be detected a clear pattern, which is that in the MJO phases 6,7 and 8 is when more TCs are generally generated independently of the AWT; and between them phase 7 is the most active one. Nevertheless, AWT 2, has very few TC historical tracks and there is not a particular increase for these MJO phases. Similarly, AWT 1 has generally a lot of tracks, with only a slightly increase observed in phases 6,7 and 8. These three MJO phases correspond to when the Pacific basin is in the enhanced rainfall (or convective) phase. In this phase, as it was explained before, there is an increased condensation and rainfall over the Pacific Ocean, as it can be observed in [Figure 15](#), extracted from [Figure 8](#) through the intense precipitation cloud in green over the Pacific Ocean.

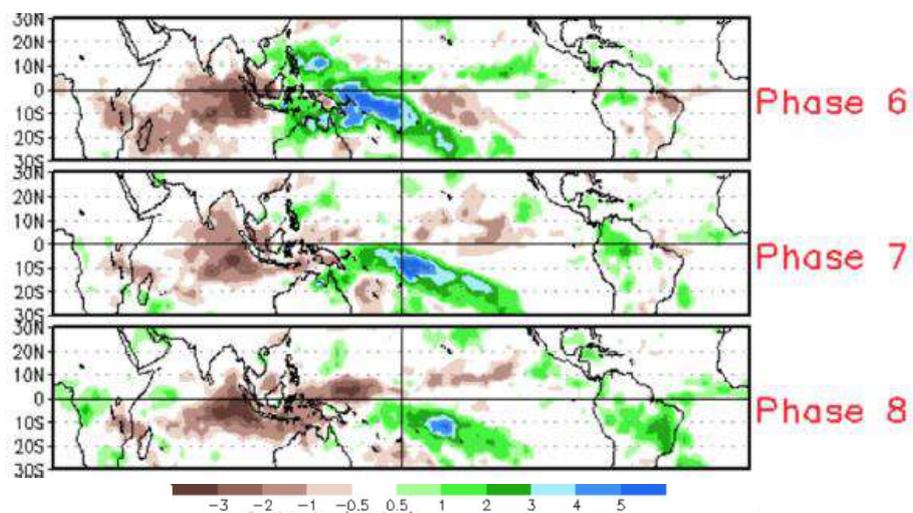


Figure 15 MJO 6,7 and 8 phases.

Additionally, these MJO phases correspond to the Western Pacific more active phases, as it can be seen in [Figure 9](#).

With respect to the AWT, there are more TCs in the AWTs 3 and 1 from an absolute point of view. AWT 1 is especially relevant since, as mentioned before, it shows a lot TCs tracks independently of the MJO phase and it is the most probable AWT. However, there is the commonly belief of associating El Niño years with an increased and more intense TC activity (Diamond, et al., 2012) (Boucharel, et al., 2016) (Stephens & Ramsay, 2014). For this specific target area, El Niño although it is the least unlikely AWT behind AWT 2, still has a remarkable number of TC tracks. This fact makes it outstand as the AWT most active in TC genesis activity since it has the greatest proportion of TC genesis days with respect to its total number of days; as it can be seen in [Figure 12](#). Furthermore, in [Figure 14](#) it can be observed how many of the TCs originated in El Niño years reached at least category 2 and most of them do. Despite El Niño AWT is highly improbable, when it occurs there is a very high risk of intense TC and therefore of potential catastrophic consequences for the SIDS the area. Furthermore, as mentioned before, El Niño does not show the general increase in TC genesis for the MJO phases 6, 7 and 8. Therefore during all the phases of the MJO, except for phase 2, the risk of TC which additionally tends to be more intense than usual; is very high. So, it can be concluded that El Niño years are especially prone to this hazard. The same pattern of generally more intense TCs is also observed for La Niña years.

Additionally, the TRMM mean daily precipitation is also transferred into the MJO and AWT combinations. The resulting patterns are the mean of all the days belonging to that AWT – MJO combination. The resulting figures include the precipitation and their anomalies.

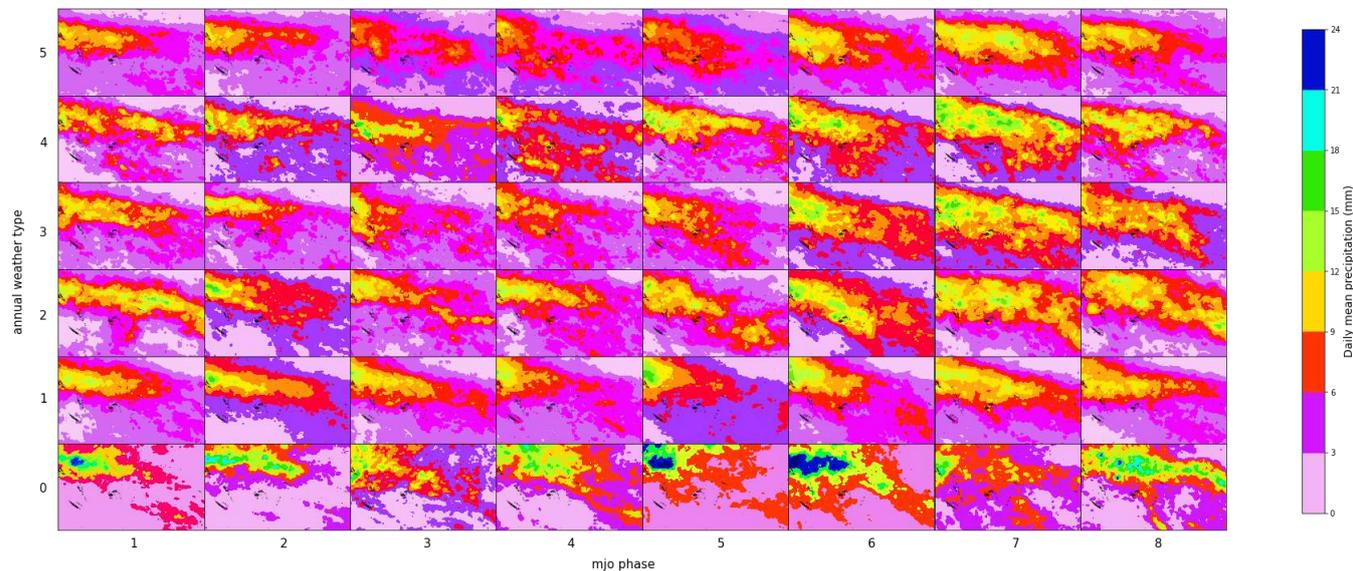


Figure 16 Daily Mean Precipitation transferred to MJO - AWT combinations.

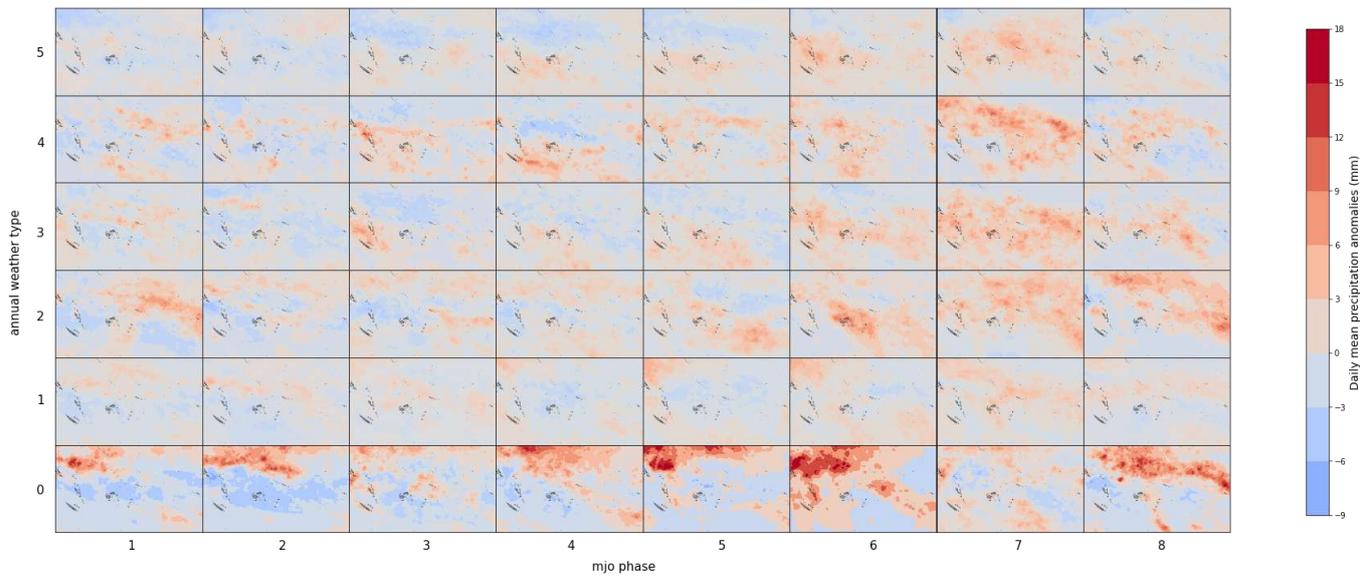


Figure 17 Daily Mean Precipitation anomalies transferred to MJO - AWT combinations.

As it was analyzed, the MJO phases with generally more TCs genesis activity are phases 6, 7, and 8, which correspond to the enhanced rainfall phases over the target area. Consequently, in [Figure 16](#) these MJO phases show more intense precipitation clouds than the rest of the phases. Furthermore, El Niño (AWT 0), shows the most intense precipitation clouds and therefore also the most extreme positive anomalies (intense red clouds). The same pattern is observed in MJO – AWT combinations 3-7, 4-7 and 1-8. These combinations have the highest number of TCs generated from category equal or greater than 2, but also the most intense precipitation clouds behind El Niño. This analysis relating rainfall and TCs genesis leads to the conclusion that higher precipitation values (positive anomalies) is the TCs common genesis zone (bellow the equator around -15 to -10 latitudes). Furthermore, as this cloud becomes more intense (with greater precipitation values and therefore also more extreme anomalies) the TC generated is likely to reach a higher category.

5 Tailor-made predictor

5.1 Index Predictor definition and building

The statistical model is applied in the South Pacific basin, to build a TCs seasonal forecast. The spatial domain (target area) previously defined, spans around the islands of Tonga, Samoa and Fiji, far enough to be able to identify regional as well as local patterns.

The time domain ([Figure 7](#)) is the calibration period, that goes from year 1982 to 2019 and time scale is daily.

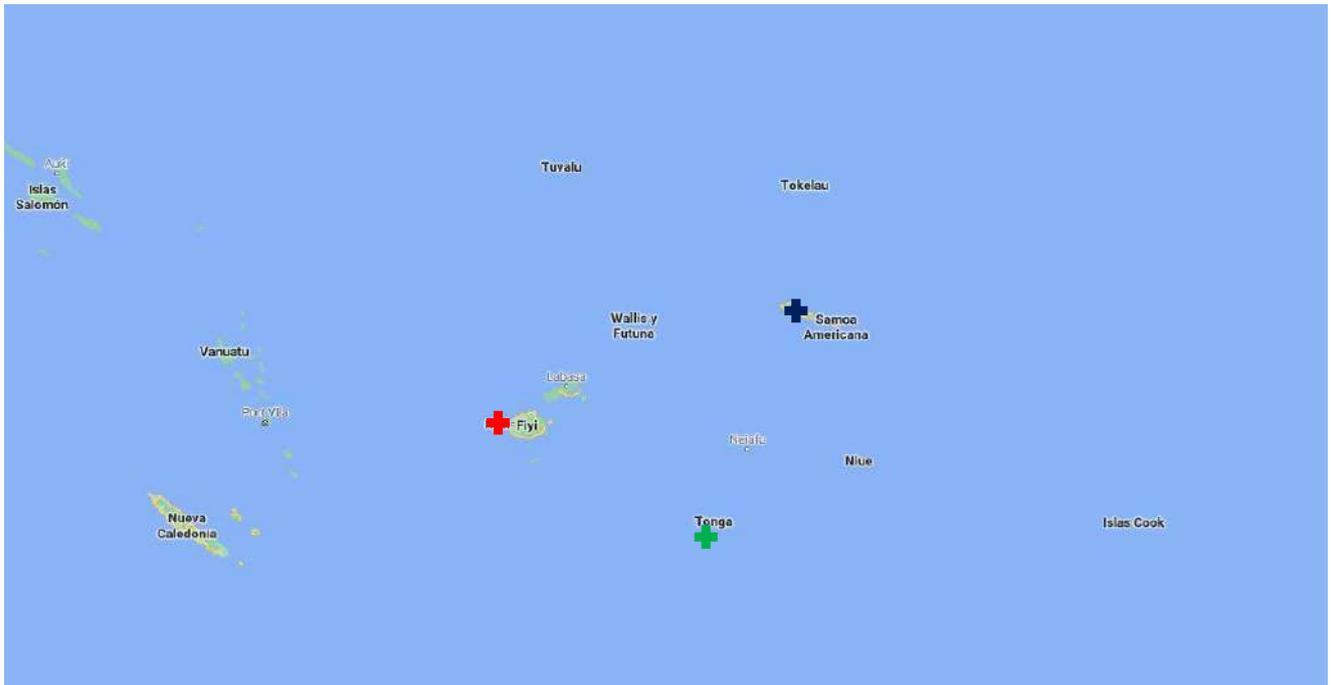


Figure 18 Target area and islands in the South Pacific basin.

Historical data from both the predictor (SST and MLD) and the predictand (number of TCs) is required for building the index predictor. Therefore, for the predictor variables the NOAA 1/4° daily Optimum Interpolation SST and the MLD from the NCEP Climate Forecast System Reanalysis (CFSR) are downloaded. The TCs data is obtained from the IBTrACs database, from which the point of minimum pressure of each TC track is kept.

These historical databases are preprocessed, keeping the data for the corresponding time and spatial domain. They are interpolated into the same 1/2° grid resolution, that is defined as the predictor grid and be seen with all its nodes in [Figure 19](#).

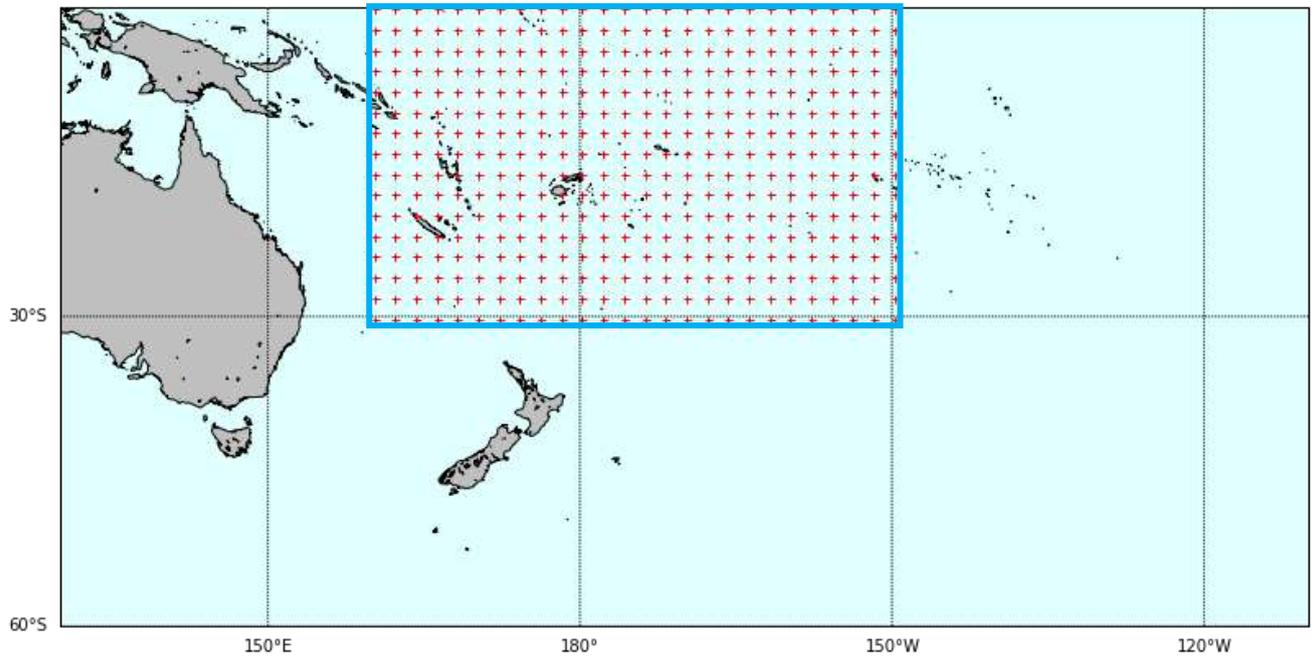


Figure 19 Predictor grid nodes for the target area.

In [Figure 20](#) it can be seen the historical databases after all the preprocessing. The data is plotted in the SST-MLD (predictor) space colored with minimum pressure data from the TCs tracks (predictand). There is not a lot of data available for building the statistical model, so it has been discretized in intervals of 0.5° C and 0.5 m for the SST and MLD respectively. The result is shown in [Figure 21](#).

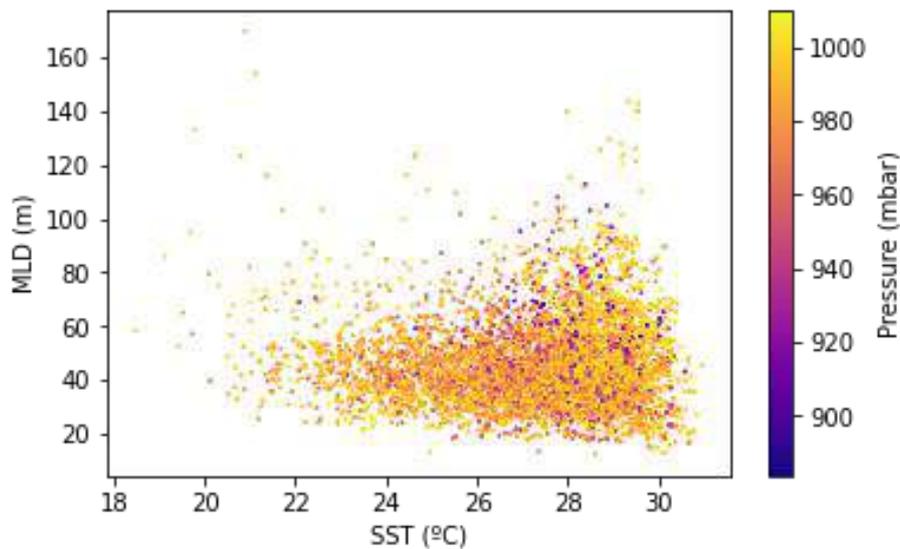


Figure 20 Predictand-predictor datasets.

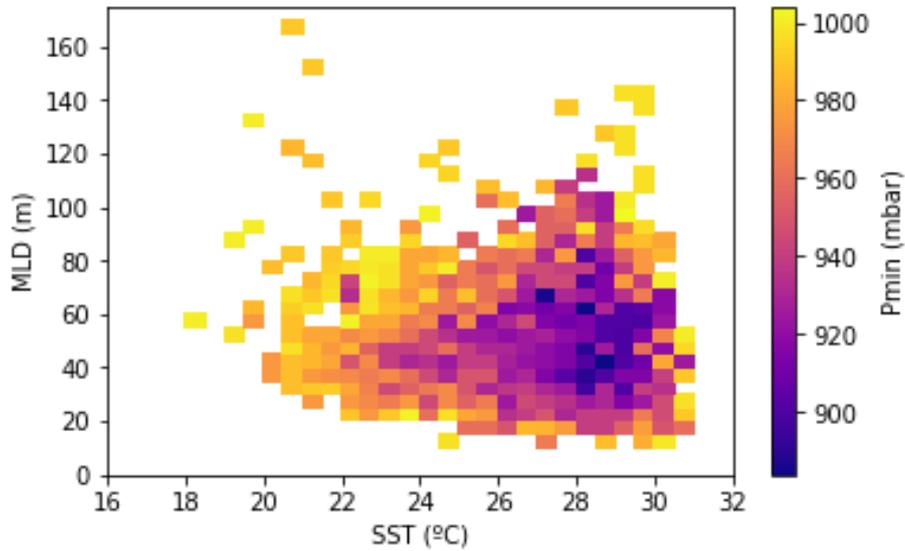


Figure 21 Predictand-predictor datasets discretized in intervals of 0.5 m and 0.5 °C for the MLD and the SST respectively.

The historical data is then combined into a tailor-made index, which will constitute the index predictor; and that for simplicity, it will range from 0 to 1 accounting for the minimum pressure limits of the dataset. So, the pressure is rescaled from 0 to 1 with the following equation, defined as the index function:

$$index_i = \frac{P_{max} - P_i}{P_{max} - P_{min}}$$

Equation 1 Index predictor function.

P_i corresponds to each pressure value of the dataset. P_{min} is the historical minimum pressure value, equal to 884 mba, and P_{max} the historical maximum pressure value, equal to 1004 mba. $Index_i$ is the obtained index predictor for the corresponding P_i .

In [Figure 21](#) all the values for the index predictor are shown. It is clearly noticed that the cells with the highest values of the index correspond to the ones with the lowest pressures ([Figure 20](#), right plot) and therefore reaching higher categories. On the contrary, the lowest index values come from the highest values of the minimum pressure, which belong to the TCs of lower categories (least intense); like for instance the cells in the upper right part of the graph. Furthermore, it is relevant to mention the fact that the points corresponding to the most intense TCs are focused in a SST range from 27 to 29 °C and in a MLD range from approximately 30 to 70 meters.

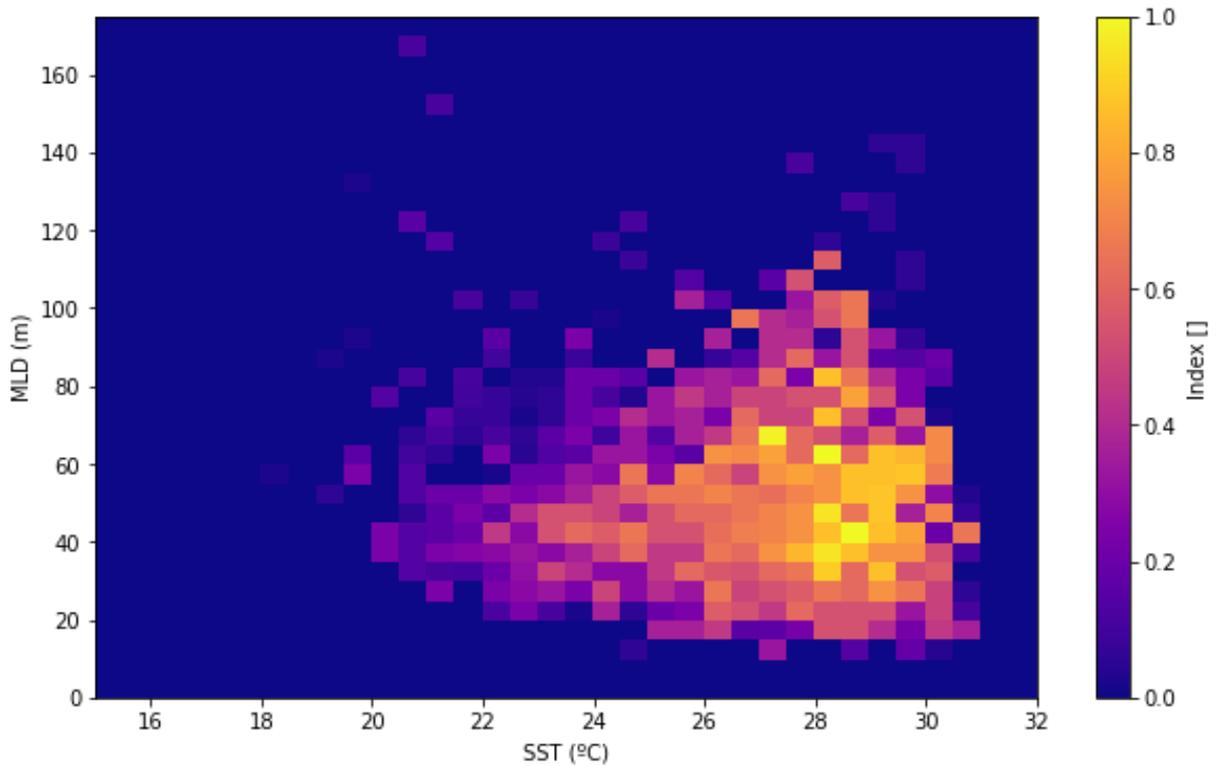


Figure 22 Index predictor.

Once the index predictor is built, it can be computed for each day of the calibration period. From the MLD and SST values of each day (both variables have been resampled by computing the daily mean), the corresponding index predictor value is assigned. This way each day of the calibration period (1982 to 2019, [Figure 7](#)) has its corresponding index predictor value accordingly to its SST and MLD values.

5.2 Statistical downscaling method

5.2.1 DWT classification

Once the predictor index has been defined and generated in a daily basis for the calibration period, the statistical downscaling methodology can be carried out.

Firstly, the principal component analysis (PCA) is employed to reduce the high dimensionality of the original index predictor data space, transforming the index predictor fields into spatial and temporal modes.

The first 237 modes are captured, which explain the 90 % of the variability, as shown in [Figure 23](#). The empirical orthogonal functions (spatial modes), $[EOF_1(x), EOF_2(x), \dots, EOF_{239}(x)]$, and their associated temporal modes or principal components, $[PC_1(t), PC_2(t), \dots, PC_{239}(t)]$ are obtained. The first four EOFs and PCs can be observed in [Figure 24](#). As it can be noticed, the first

mode, captures the greatest percentage of variability. Then the quantity of variance represented by each mode progressively decreases for each consecutive one.

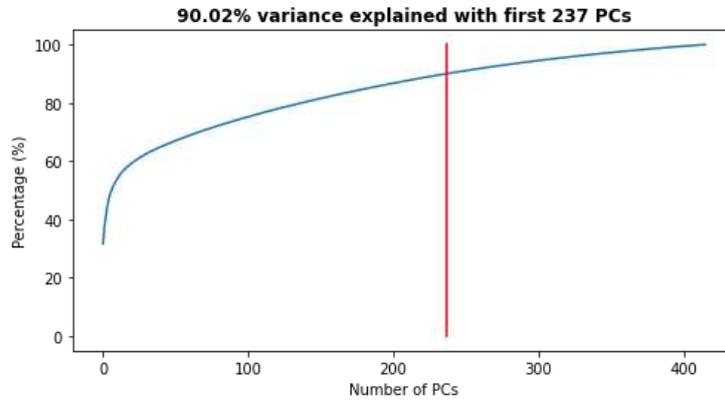


Figure 23 Number of PCs vs Percentage of variance explained

In the EOFs plots the spatial variability for all the years of the period (1982 – 2019) is represented and in the PCs plots the temporal variability for the considered area ([-30,0]N [160,210]E) is shown.

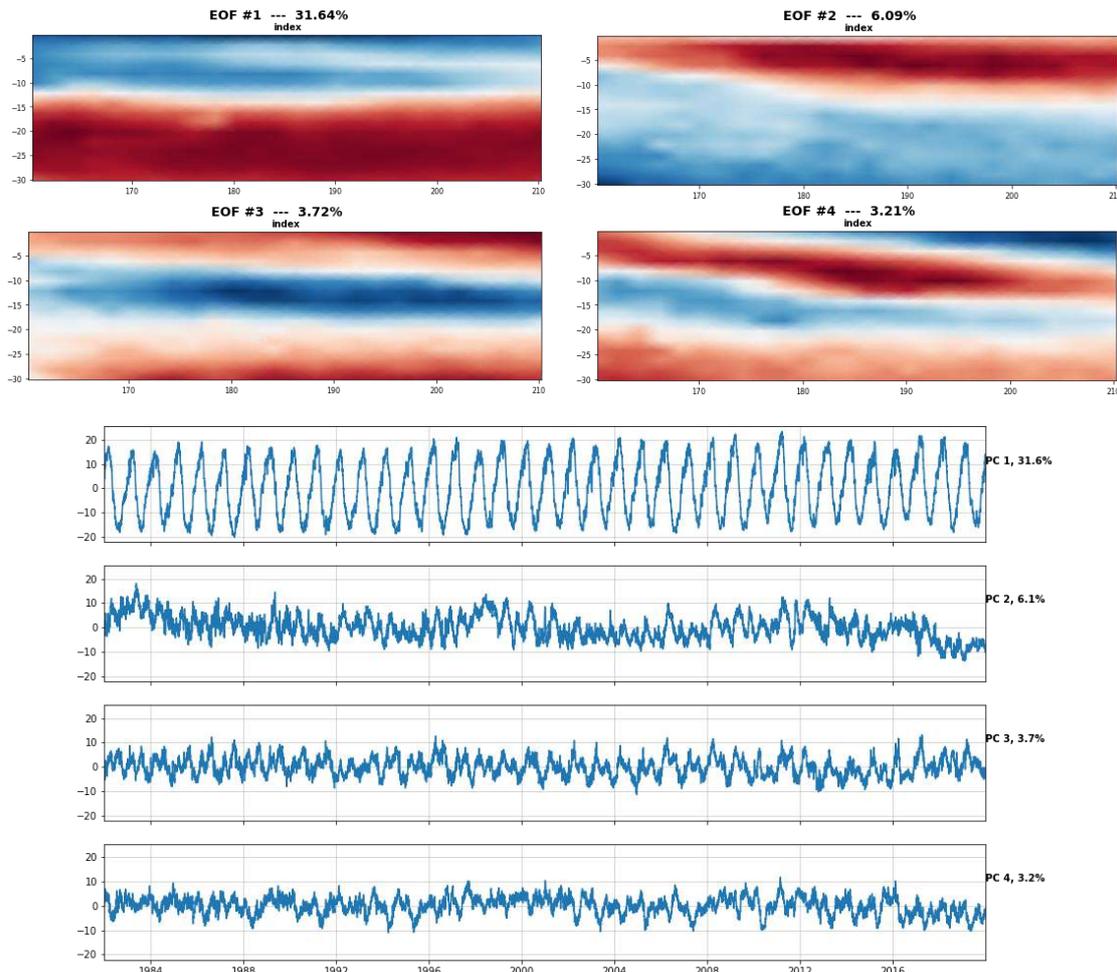


Figure 24 First four modes of the data space.

After the PCA, the k-means algorithm is applied to the multivariate data to obtain $N = 49$ groups defined by a prototype or centroid for each k th-cluster as $Index_k = (PC_{1,k}, PC_{2,k}, \dots, PC_{239,k}); k = 1, \dots, N$. After that, the spatial modes (the EOFs) corresponding to each PC are expressed in the original space for the k th-cluster as $Index_k(x) = [EOF_1(x)PC_{1,k} + EOF_2(x)PC_{2,k} + \dots + EOF_{239}(x)PC_{239,k}]; k = 1, \dots, N$. The number of clusters chosen, $N = 49$, was considered to be representative of the PC space. This number must be a meet compromise between an easy handle characterization of the synoptic patterns and the best reproduction of the TCs activity. A sensitivity analysis could be carried out in future works considering different values for M , but for this work is out of the scope. Nevertheless, previous works with similar analysis confirm that the selection of this number is adequate (Rueda, et al., 2017).

[Figure 25](#) illustrates the result of the classification in the first three PCs space of the tailor-made index data. It can be seen in the figure that the obtained centroids, the back dots, span the wide variability of the data.

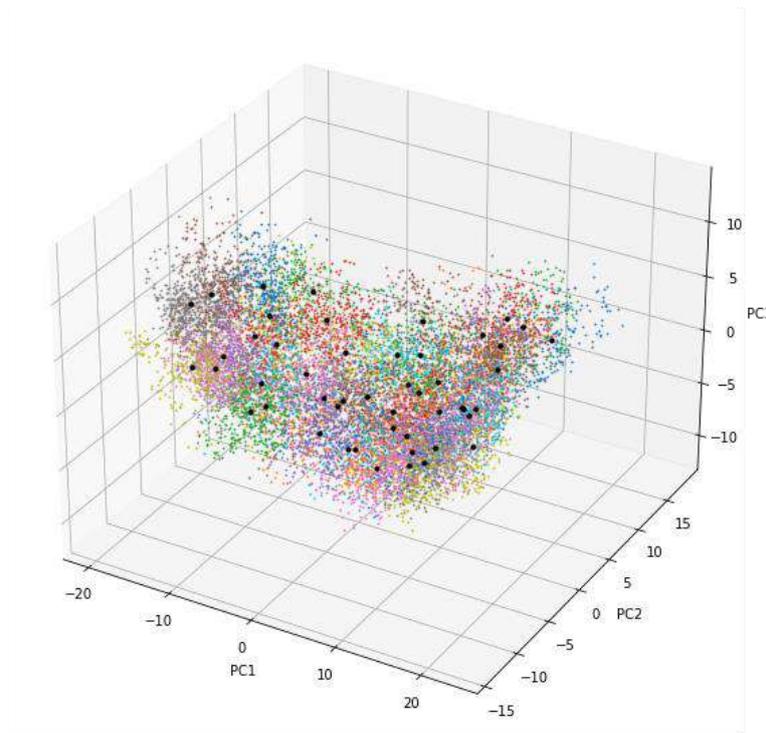


Figure 25 Index data in the PC space. The black dots represent the $N = 49$ clusters obtained using K-means.

The 49 DWTs are organized in a bidimensional lattice and each one has a color associated, which are shown in [Figure 26](#). Note that this color code will be used throughout this document for all the figures.

1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	32	33	34	35
36	37	38	39	40	41	42
43	44	45	46	47	48	49

Figure 26 DWTs bidimensional lattice organization and corresponding colors.

In [Figures 27](#) and [28](#) it can be observed the result of the clustering of the index predictor data into 49 groups, known as daily weather types (DWTs), and their corresponding anomalies. The dots included in the figures, in white in Figure 27 and in green in Figure 28 are the TC genesis points of the TC tracks generate in that DWT.

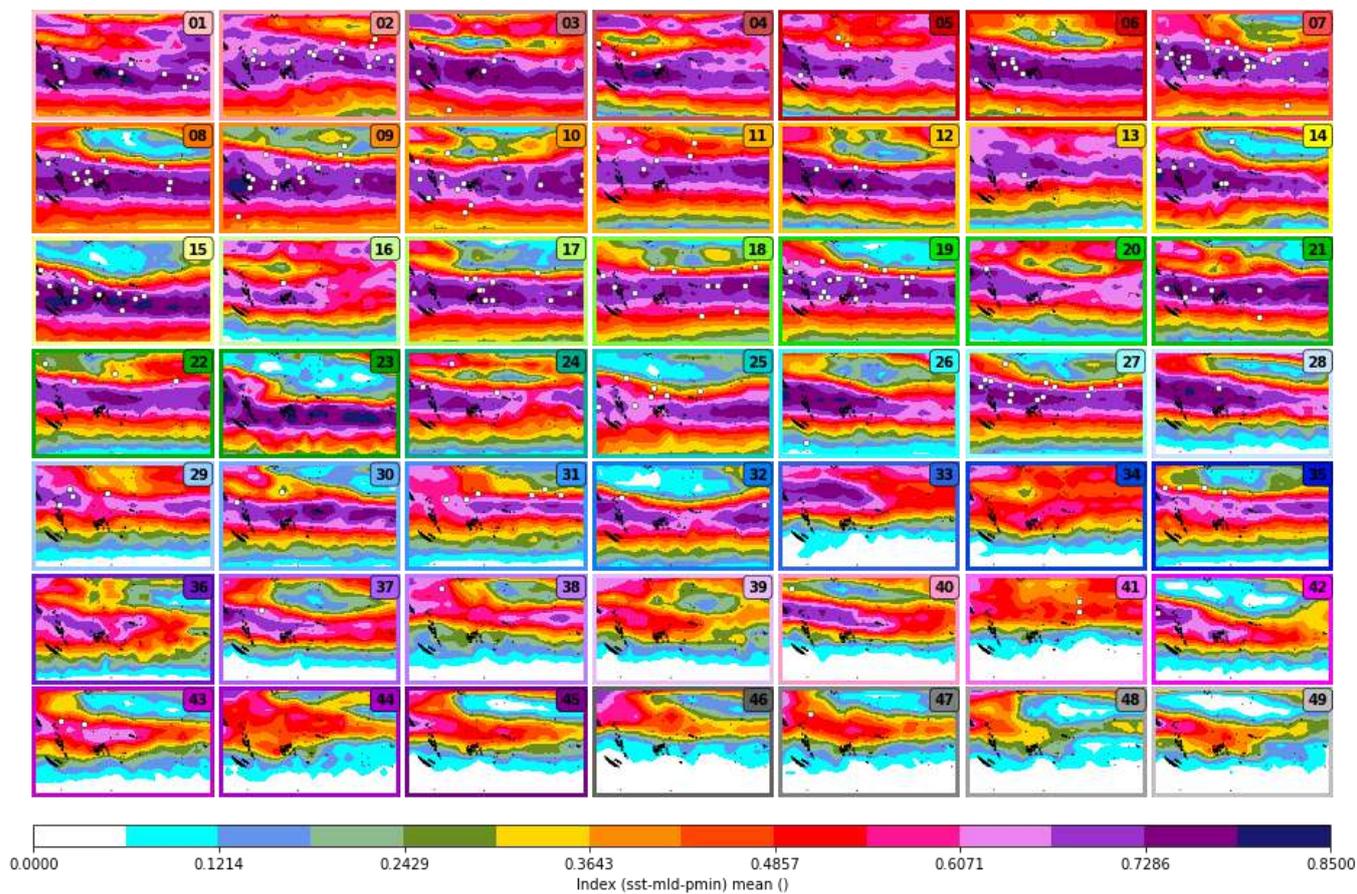


Figure 27 DWTs.

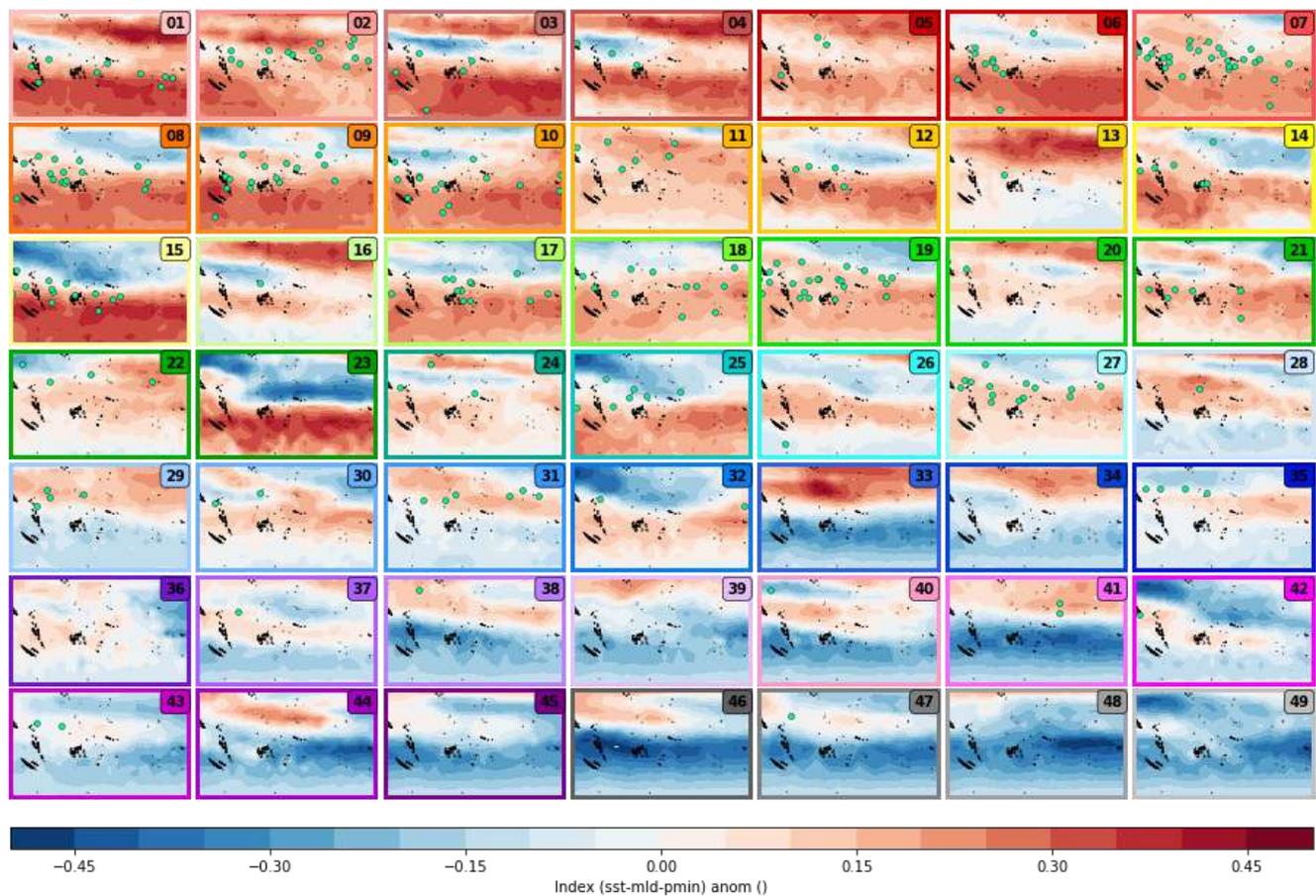


Figure 28 DWTs anomalies.

The DWTs are different to each other, showing the high variability of the data space. However, when looking into each clusters' elements, homogeneity is observed since the patterns included are similar to each other. In [Figures 29](#) and [30](#) all the patterns belonging to DWTs 5 and 41 can be seen and it can be noticed how similar they are amongst them. The average number of elements in each cluster is 283, cluster number 5 and 41 have 186 and 398 elements, respectively. Therefore, it can be confirmed that $M = 49$ is an appropriate choice for this work.

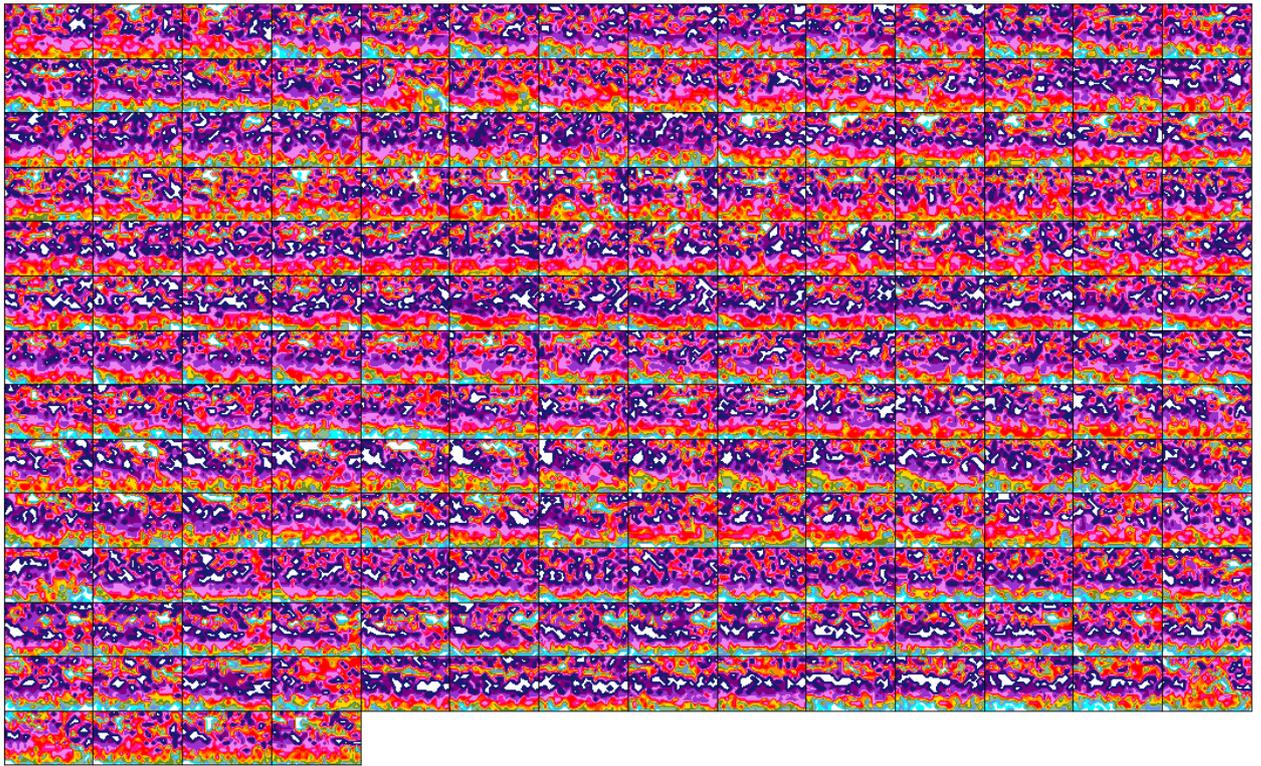


Figure 29 Days associated to cluster DWT = 5.



Figure 30 Days associated to cluster DWT = 41.

As it was previously explained, the bidimensional lattice of the DWTs ([Figure 26](#)) is organized after the k-means so when plotting the similar patterns remain next to each other and there is a logic progression as when moving forward in the mesh.

In the upper part of the 2D lattice the most intense index patters are found. They show in the center of the area a large flame of very high index values, in the range from 0.6 to 0.8. In [Figure 28](#) this flame is represented through a red area of positive anomalies. As going further in the lattice, although the flame remains, its size and intensity decreases progressively and the variability in the range of index values shown increases. This way each DWT presents more contrast, showing the whole range of values for the index, like the DWT 23. As the flame becomes smaller, the patterns become generally less intense (and consequently the anomalies maps turn bluer representing negative anomalies while the red areas of positive anomalies disappear progressively); reaching almost the extinction of the flame in the last patterns. DWT 49 is the opposite of DWT 1, both show very little variability but the first one involves very high index values (full read anomalies map) whereas the last one has 1/3 of the area (the southern part) with index value equal to 0 (full blue anomalies map).

The vast majority of the TCs genesis points are in the first 28 DWTs (upper half of the 2D lattice). Then the frequency drastically diminishes and except in DWT 31 and 35 nearly no genesis activity is found.

The patterns with the highest number of TCs genesis points, like the DWTs 2, 7 or 19 are characterized by the presence of the previously mentioned intense flame in the center of the area. Then the upper and lower parts show up to different index values depending on the DWT, but generally up to 0,06. Most of the genesis points are focused in the purple zone, which corresponds to 0.67 to 0.79 index values. However, in some cases genesis points are found in areas where the index reaches very low values (up to 18), like in the DWTs 14 or 22. Nevertheless, this is not the usual case and consequently, as the flame size and intensity decreases, so do the genesis activity. From the DWTs 28 onwards there are nearly no genesis points, except punctual cases such as DWTs 31 or 35, where some TCs were generated in slightly lower values of in the index, around 0.4 – 0.65.

We can conclude that no TCs is generated generally when the index values is higher than 0.79 or lower than 0.36, and that generally it occurs in the range from 0.67 to 0.79, corresponding to red areas in the anomalies map. Additionally, the red anomalies area has to be extended through the map, occupying most of the area, for increased TC genesis activity.

In the following figures the predictor variables, SST and MLD have been transferred into the DWTs so the index patterns can be better understood.

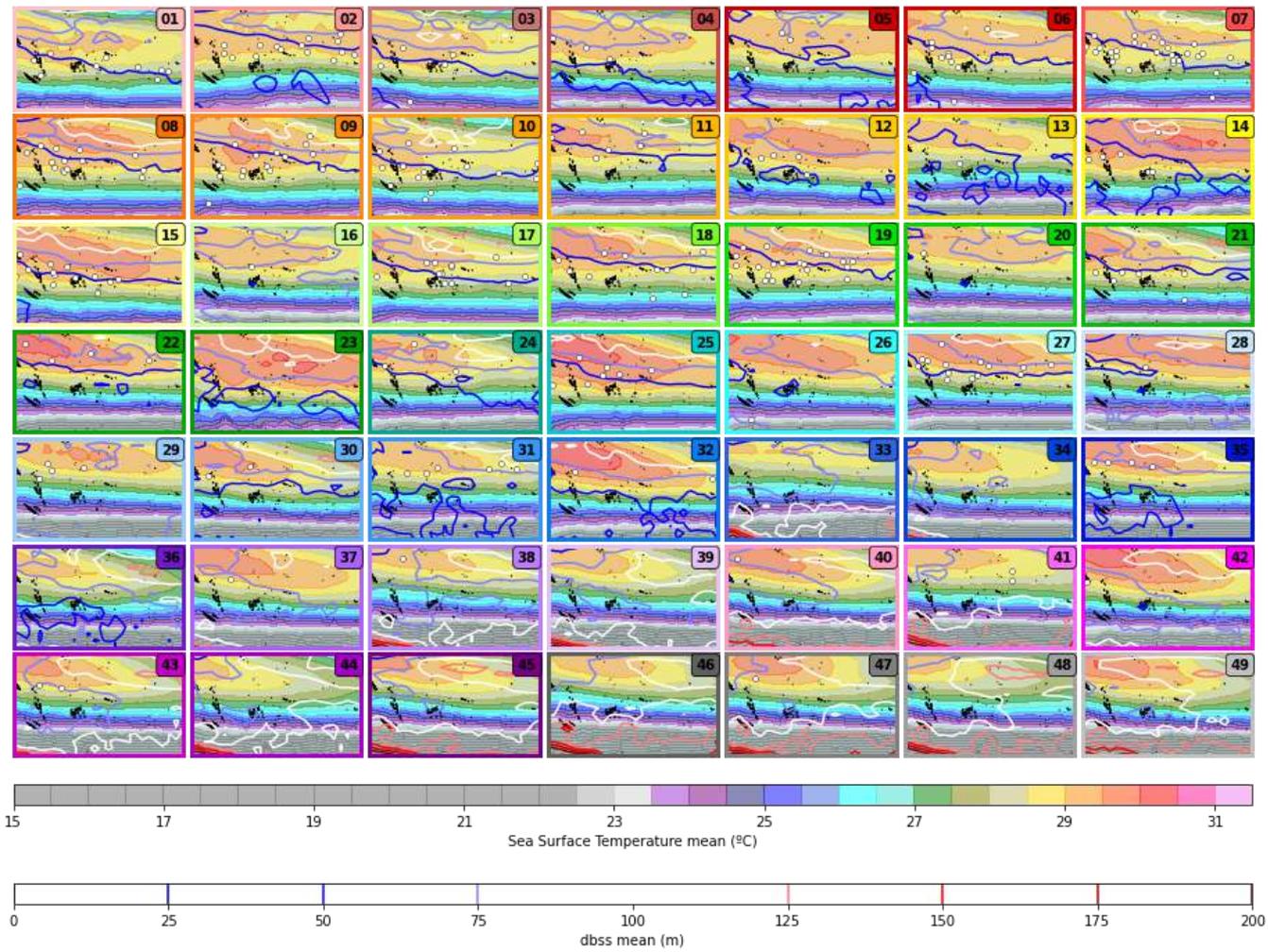


Figure 31 SST and MLD transferred into the DWTs.

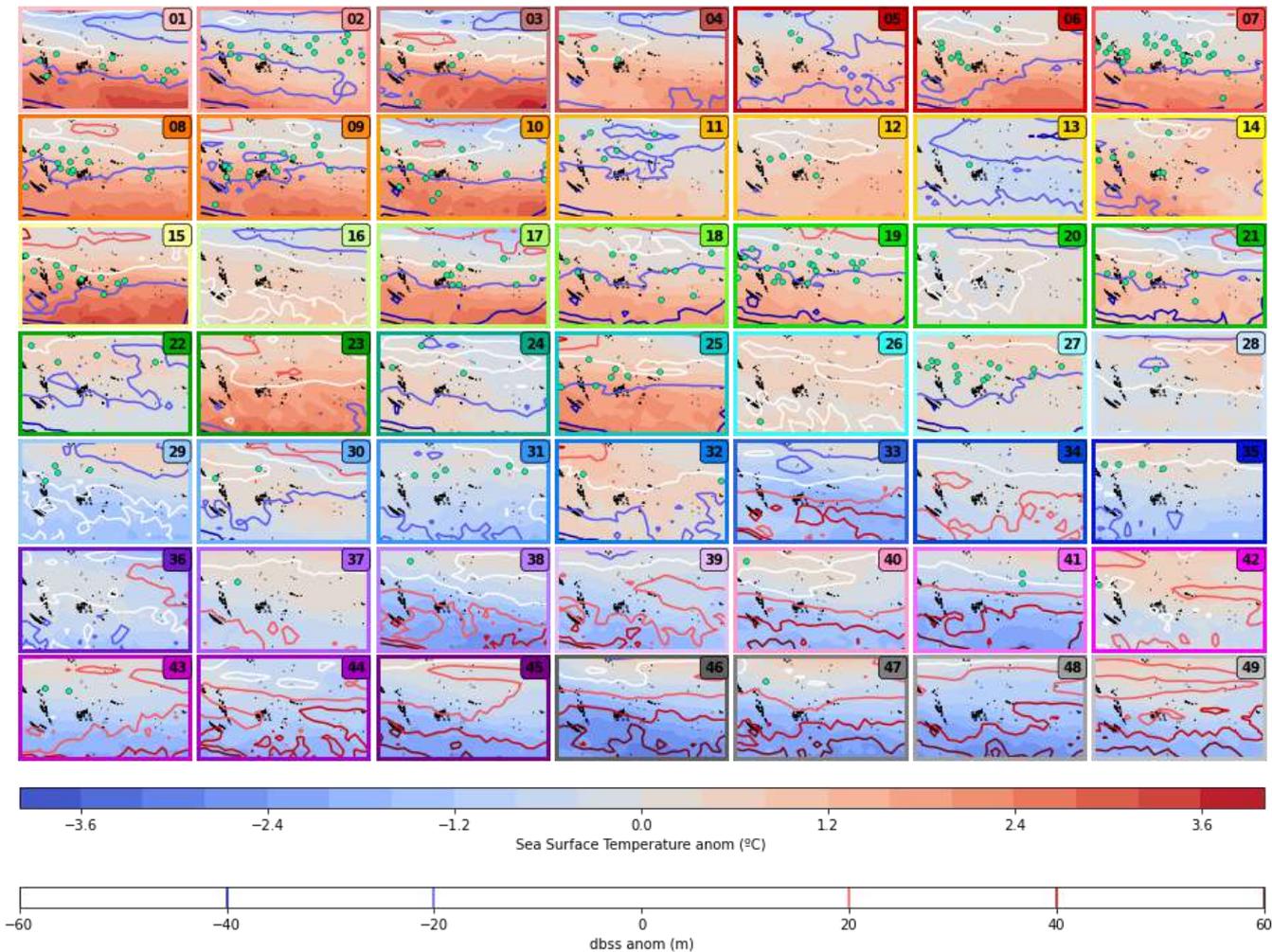


Figure 32 SST and MLD anomalies transferred into the DWTs.

The most intense DWTs correspond to also very intense values of SST, greater than 23°C and to MLD values smaller than or equal to 100 meters. The anomalies maps are generally red, showing positive anomalies of the SST. However, the MLD anomalies show more variability, which are mostly blue (negative) but also red (positive) and white (zero). Most of the area is covered by temperatures from 28.5 to 30 °C, observing again a big and intense flame like the one in the index figures. The hot area becomes more intense from DWT 7 but the variability increases, generating big temperature contrasts like in DWT 22. The anomalies maps become more intensely red. Meanwhile the MLD remains with similar pattern and values.

As we progress in the lattice, the hot area with temperatures from 28.5 to 30.5 °C disappears and increases the area covered by less intense values, generating patterns that are just the opposite of the first ones. From predominant temperatures higher than 26.5 °C, from DWT 37 onwards, nearly half of the area is covered by temperatures lower than this value. This is similar to what happened to the intense index flame values, that was explained for the previous figures. Also, the anomalies maps, show accordingly an increasingly blue tone. At the same time, the MLD had not shown except punctual appearances values greater than 100 m. From DWT 35 on, they appear and remain until the last DWT. This is also reflected in the anomalies, which from blue and white maps turn to maps with red and white ones.

Regarding the TCs genesis points, it can be observed that they are generally located in areas with temperatures in the range from 28 to 30 °C, especially from 28.5 to 29.5 °C. For instance, in patterns 2, 7, 17 and 19. In the DWTs with more intense temperature flames, like the 8, 15 or 25; the genesis points are less dispersed, trying to stick to the area with this favorable temperature. There are a couple of exceptions to this, since some DWTs show genesis points in lower temperature values, like the most southern points of DWTs 18 and 21, in the range from 25.5 to 26.5 °C. What is also very relevant to mention is that the TCs genesis activity is always located in areas with MLD values smaller than 75 m. All these patterns correspond to negative MLD and positive or zero SST anomalies maps, except in DWTs 29, 31 and 35 and punctual cases where the anomalies are very slightly blue (close to null anomalies).

Clear patterns can be extracted from these figures related to TCs genesis occurrence. Most of the TCs genesis take place under the following circumstances except punctual cases:

- SST interval from 28 to 30 °C (specially 28.5 to 29.5 °C) that correspond positive or zero SST anomalies.
- MLD values equal or smaller to 75 m that correspond negative MLD anomalies.

These two characteristics correspond to index values in the range from 0.6 to 0.79.

Therefore, this proves that these two variables are a good choice as predictor variables for the possibility of TCs occurrence.

5.2.2 DWTs seasonality, annual variability and chronology

Several plots have been made to better analyze the distribution of the DWTs, their transition, persistence and conditioning to TCs occurrence.

Firstly, [Figure 33](#) shows total TCs genesis activity during the calibration period for each month of the year. This confirms that the main season of TCs is between November and April.

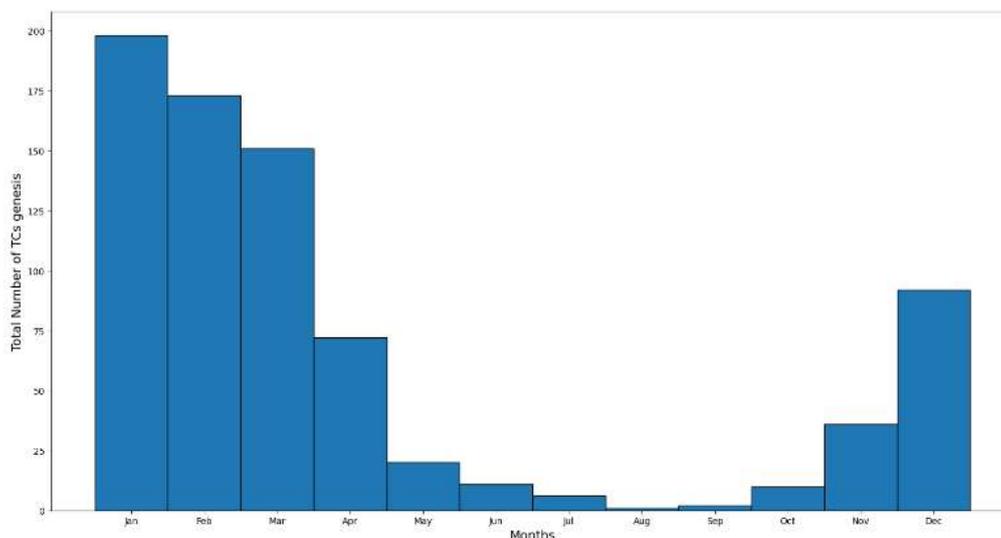


Figure 33 Histogram of TCs genesis per month during the calibration period.

Figure 34 shows the seasonality of the DWTs.

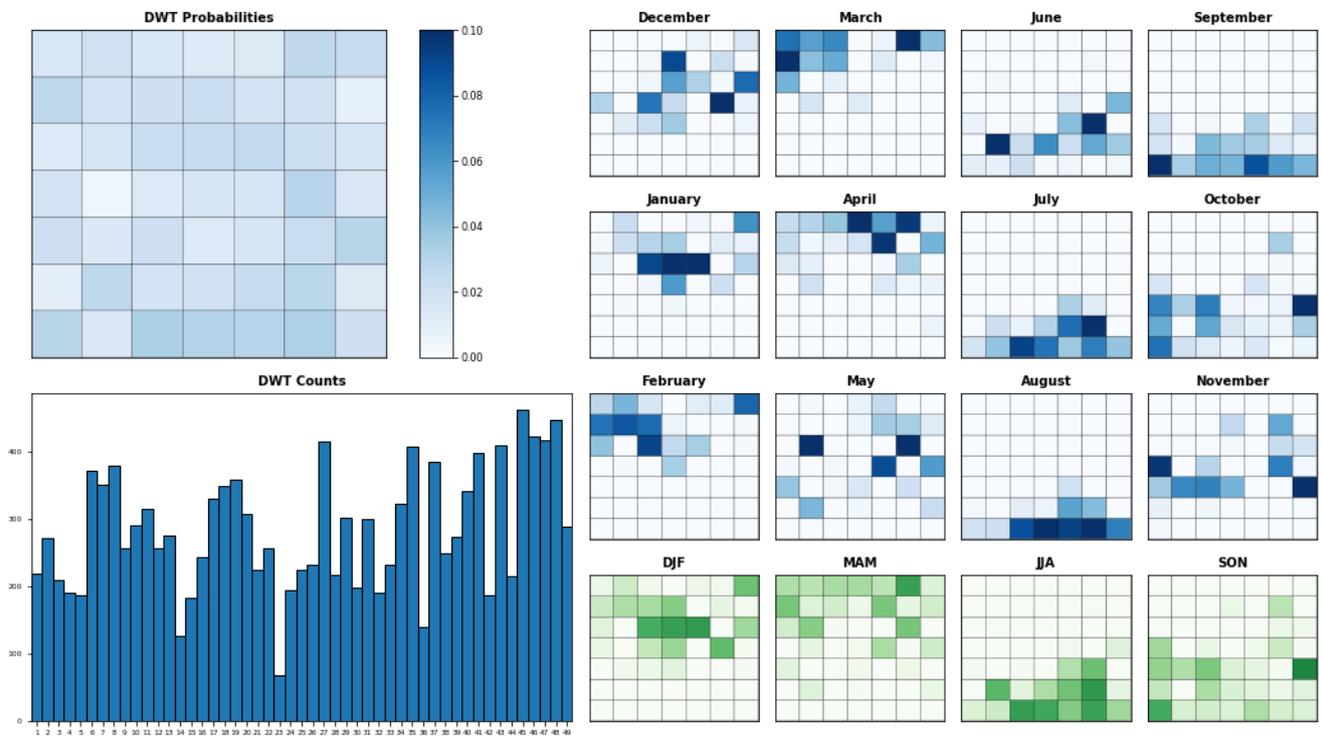


Figure 34 DWTs seasonality.

The mean DWTs probabilities for all the calibration period are displayed in the left upper plot. The probability values are quite homogeneous without major differences and contrasts between them except for some DWTs that are less probable. This can better be observed in the plot below left, where it is shown the total number of days included in each DWT during the whole calibration period in a bar graph. The DWTs less probable have consequently the smallest number of days assigned, like DWTs 14 or 23.

Then the right plot shows the DWT mean probabilities for each month and for the four seasons of a year for the whole calibration time period. It can be clearly noticed a pattern during the months. The TCs season includes the months of November, December, January, February, March and April, where there is a high TCs activity, as shown in [Figure 33](#). The rest of the months generally show no TC genesis.

As it was seen in the previous section, most of the TCs genesis activity is concentrated in the first 28 DWTs (upper part of the lattice). Then, the frequency drastically diminishes and except in DWT 31 nearly no genesis is found. To support the incoming analysis, Figure 27 is included again below, so it can be easily recalled how is the TC genesis activity shared amongst the DWTs.

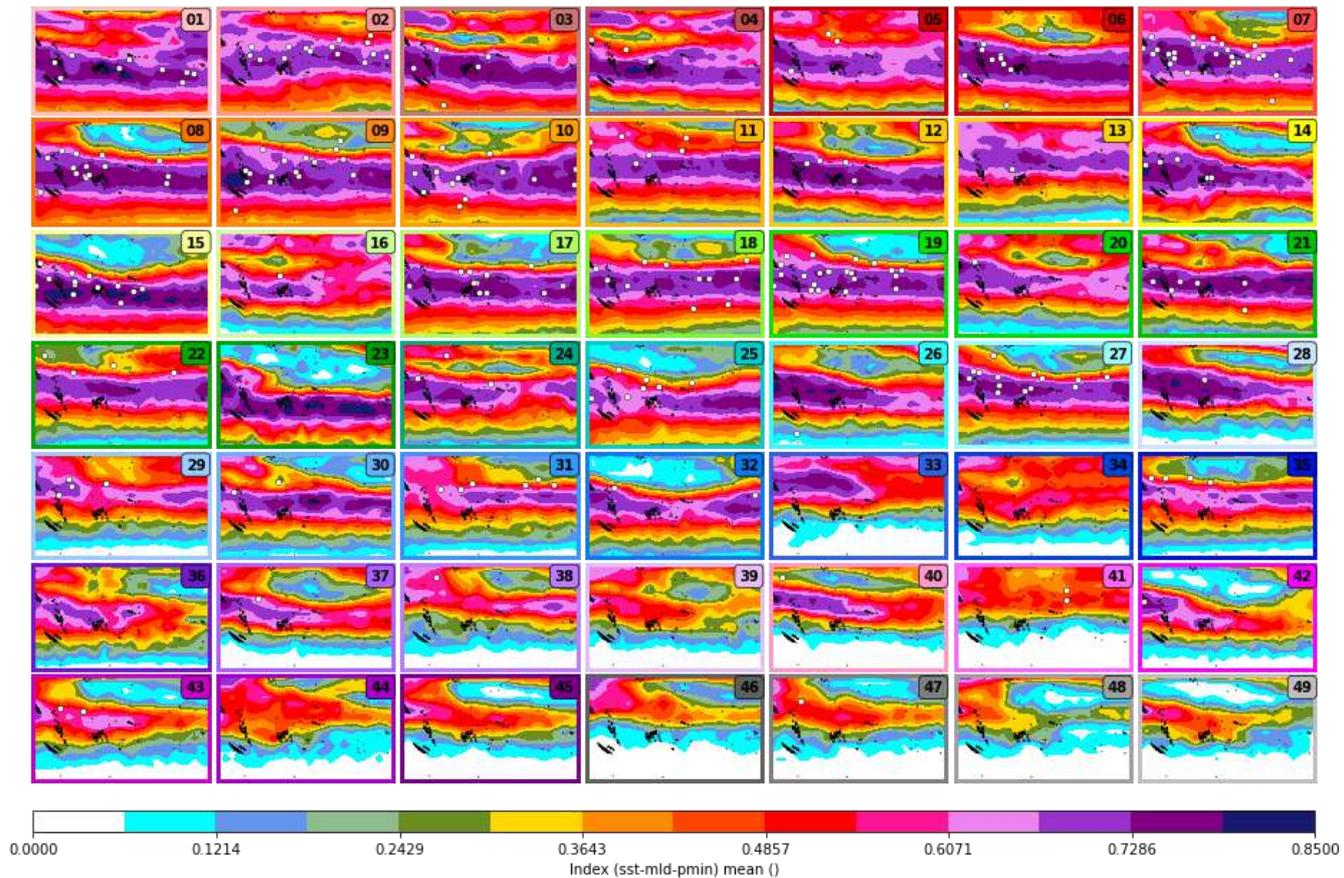


Figure 27 DWTs.

In the DWTs for the months not included in the TCs season it can be observed how the probability is always concentrated in the DWTs from DWT 28 onwards, which are the DWTs with nearly no genesis activity. However, the probabilities patterns start to change in October and it can be noticed how there is a transition of the probability to DWTs higher in the lattice, where it is also more TC genesis activity. Then in November this shift is much clearer, and the probability is now focused only in DWTs from 11 to 35. In December, this tendency continues as we are entering the most intense part of the TCs season, when most of the TCs happen are originated. January, February and March, which are the months with the highest TC activity have the probability located amongst the DWTs in the first three rows (upper half of the lattice). This just the opposite pattern shown from the months from May to November. In May the probability shifts again to restart this cycle and the lower half of the lattice starts to become bluer again.

This pattern is also seen in the seasonal plots. The DWTs probability is shared amongst the last rows of the DWTs, that have nearly no TC genesis during the Winter in the southern hemisphere (June-July-August plot). However, the pattern starts shifting for the September-October - November plot; so although the probability still remains to a large extent in the lower half; it is anticipating the arrival of the TC season by shifting to the upper rows of the lattice. In the December-January-February plot the pattern has changed and reflects the fact that these months are the most intense in TC genesis since the probability is focused on the upper part of the DWTs, where it is mainly located the TC genesis activity (Figure 27). Finally, in March-April-May plot, although the DWTs probability is still focused on the upper half it is more shared amongst the DWTs of the lower half (with less or null TC genesis activity). This is due to the

influence of April and May, months where the TC activity is much lower than in March, although it cannot be disregarded.

In conclusion, during the TC season months (November, December, January, February, March and April) the DWTs probability is focused on the upper half of the DWTs lattice (Figure 26), in the first 28 DWTs, where most of the TC genesis activity is also located (Figure 27). On the contrary, in the rest of the months, the probability is shared amongst the DWTs of the lower half of the lattice (from DWT 28 onwards), where there is very few or null TC genesis activity.

The following figures explore the interannual and intra annual variability of the DWTs.

In Figure 35 it can be seen the perpetual year. This figure includes the aggregated DWTs probability for each day of the calibration period and therefore provides an idea of how it looks the mean year. In Figure 36 it can be observed the chronology of each year. The AWT is indicated through its color code (Figure 11) and the days with TC activity are included with black dots through the years. The colorbar on the right indicates with black dots also which DWTs have more TC genesis activity, since the size of the point is bigger as the TC genesis activity in that DWT is also greater.

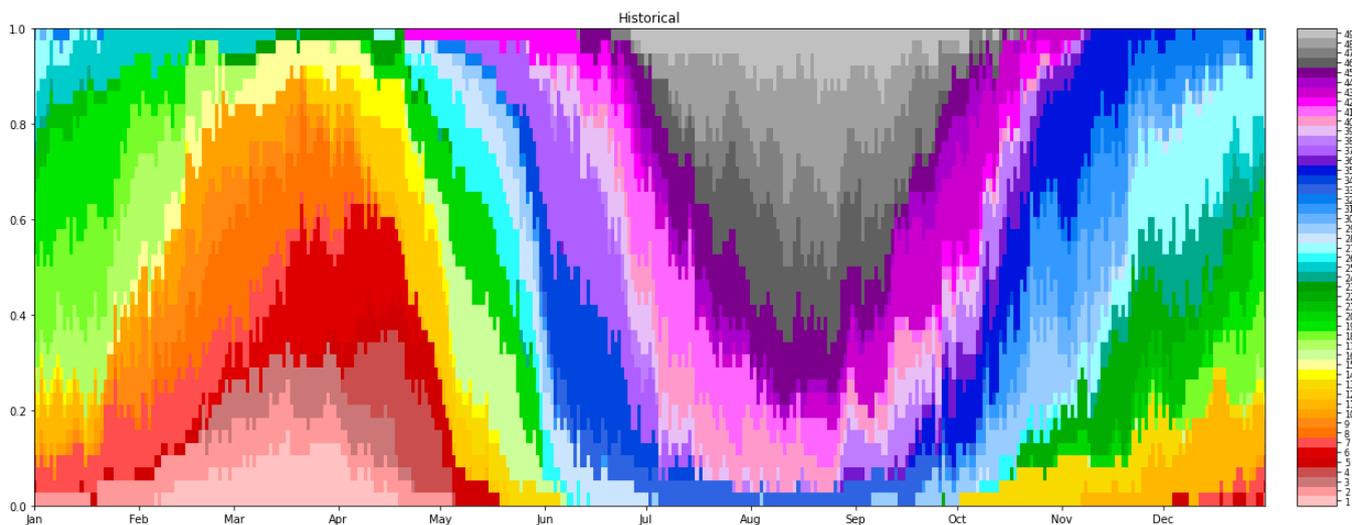


Figure 35 DWTs monthly chronology during all the calibration period.

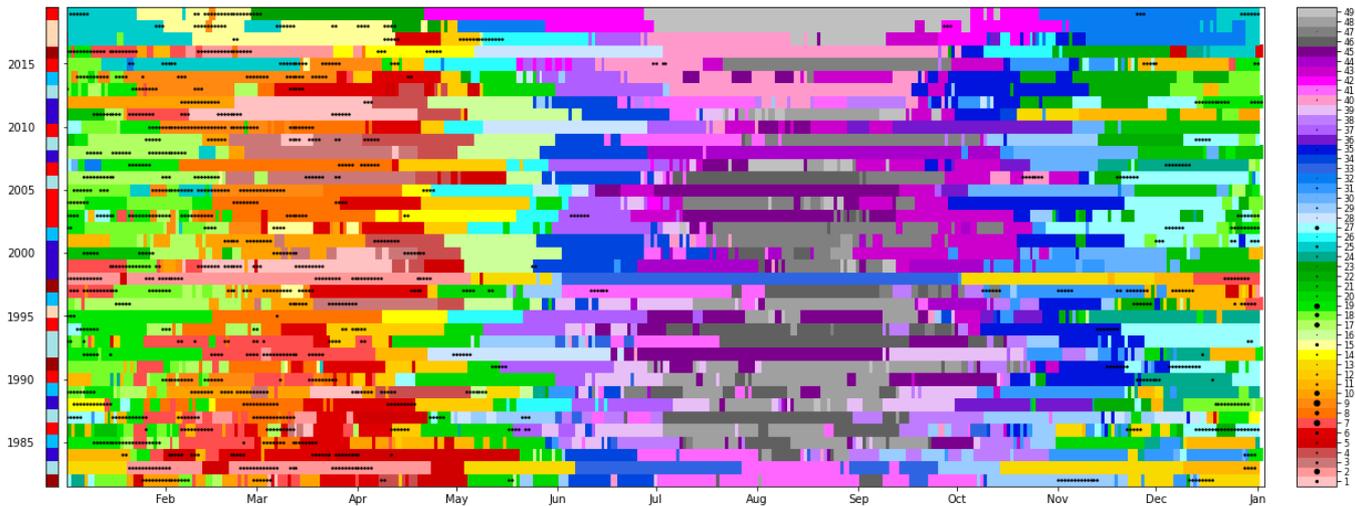


Figure 36 DWTs chronology.

In both Figures 35 and 36 we can firstly analyze the intra annual variability. From June to November there is only purple, pink, gray and blue, colors that corresponds to DWTs from 29 to 49 (the lower half of the DWTs lattice) and thus to the DWTs with nearly or null TCs genesis activity (as it can be noticed from the colorbar and from the Figure 27). However, as the TCs season is entered the trend changes and light blue and greenish blue start to appear, as it can be noticed for November. As the most intense months in TC activity are being approached, more green and orange are introduced; which correspond mainly to the second and third rows of the DWTs lattice. These rows are the ones gathering the highest number of TC genesis points (as it can be noticed from the colorbar and from Figure 27). Then, during January, February and March, the months with most TC genesis activity (Figure 33) DWTs from the first rows are the only ones present, through the red, orange and green colors. Finally, in May, the pattern changes to restart the cycle, and blue, purple and pink colors start to be present once again. This represents the end of the TCs season, as DWTs with no or nearly TC genesis activity are reintroduced.

Focusing in the interannual variability, the same month changes the DWTs depending on the year. Different factors can influence in this interannual variability such as long-term trends, maybe associated to SST warming during this 40-year period. ENSO, represented as the AWT, has been proven to have a very relevant relationship with TC genesis activity in the target area; as previously explained in this section and in Section 4. Due to this reason, and because it explains the natural interannual variations in the ocean and atmosphere in the tropical Pacific; it will be more deeply explored the relationship between the DWTs chronology and seasonality and this climatic phenomenon. For this reason, Figure 37 which includes the DWTs probability according to AWT is displayed bellow. The color of the edge of each lattice comes from the color assigned in the already explained Figure 11. Dark blue is La Niña (AWT 5), maroon is El Niño (AWT 0) and the rest correspond to the intermediate phases.

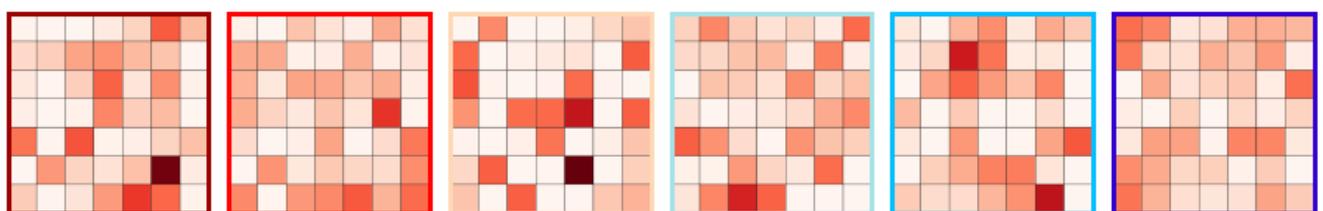


Figure 37 DWTs according to AWT.

Depending on the AWT the DWTs probabilities pattern is different, showing less or more variability.

From the analysis made in [Section 4](#), the following conclusions were extracted with respect to TC genesis and AWT relationship. AWT 1 and 3 show the highest absolute proportion of TC genesis and are the most likely AWTs; however, El Niño which is the least probable AWT behind AWT 2; has comparatively a much higher number of its days with TCs genesis activity and additionally a greater proportion of these TCs reach at least category 2 (this also happens with la Niña). AWT 2 is the least likely and with the least intense TC activity.

Coming back to [Figure 36](#), with respect to El Niño, there are only four years (1982, 1991, 1997 and 2015). However, all of them show relevant TC activity, specially 1997. It is relevant to notice that additionally, this activity extends later in the TCs season. In year 1997 until June and in years 1982 and 1991 until May. Furthermore, in 1997 TC activity is already observed in October. This way, 1997 becomes the year in which the TC activity on a season (96-97) ends the latest and starts the earliest (97-98). In 1982, 1991 and 1997, in May, the TCs days take place during DWT 19, a DWT with very high TC genesis activity ([Figure 27](#)) On 1997 there are some days with TC activity in DWTs 41 (pink), on which TC genesis is likely but very rare. In fact, 1997 is the only year of the calibration period that has TC activity on this DWT.

During el Niño years it can be noticed that there is not much variability in the DWTs chronology comparing to others AWTs. This is reflected in [Figure 37](#) too. For instance, DWT 41 (medium pink) is highly likely, which is reflected in how much pink can be seen during especially the winter months (in the southern hemisphere). DWTs with very high TCs genesis activity ([Figure 27](#)) are very probable, like DWTs 6 or 18.

With respect to AWT 2 (cream color) only 1995, 2016 and 2017 have it and their TC activity is less relevant comparable to the rest of the years. It is the AWT that shows the least variability in [Figure 37](#). The probability is more focused in specific DWTs, especially 15, 40 and 26. In fact, DWT 15, the one with the highest TCs genesis activity of the three ([Figure 27](#)), includes most of the genesis days in 2016 and 2017.

During La Niña years (AWT 5, dark blue), there is a great variability in the DWTs chronology. This is reflected in [Figure 37](#) in how the probability is shared amongst nearly all the DWTs and in [Figure 36](#) in the wide range of colors shown. During the TCs season there are many TC days in DWTs 1, 2 and 5. In fact these DWTs are not likely globally ([Figure 34](#)) but very frequent during this AWT.

With respect to AWT 1 (red color), the one with the highest TC genesis activity from an absolute point of view and the most likely; it shows remarkable variability in [Figure 37](#) but DWT 27 is the most common. Accordingly, this DWT has a very high TC activity ([Figure 27](#)).

Then, years with AWT 3 (light blue color) are fairly spread out and therefore separated in time. Furthermore, they are quite different in DWTs chronology from one to another, except the couple 1992-1993. This may be since the years are more distant one from another and other climatic factors can influence the DWTs pattern. Sometimes close years have more similar DWTs chronology than their analogs in AWT, and this can be seen in how the southern hemisphere winter pattern (months of June, July, August and September) includes widely DWT 40 only in the last decade and DWT 25 from 2014. Similarly, DWT 46, which was very common from 1982

to 2015 during the southern hemisphere winter months, does not barely appear from this year onwards.

5.2.3 Relationship predictor-predictand

To establish the relationship between the predictor and the predictand the TCs tracks going through the target area have been counted.

Firstly, in [Figure 38](#) the TCs tracks have been transferred into the DWTs according to their genesis point.

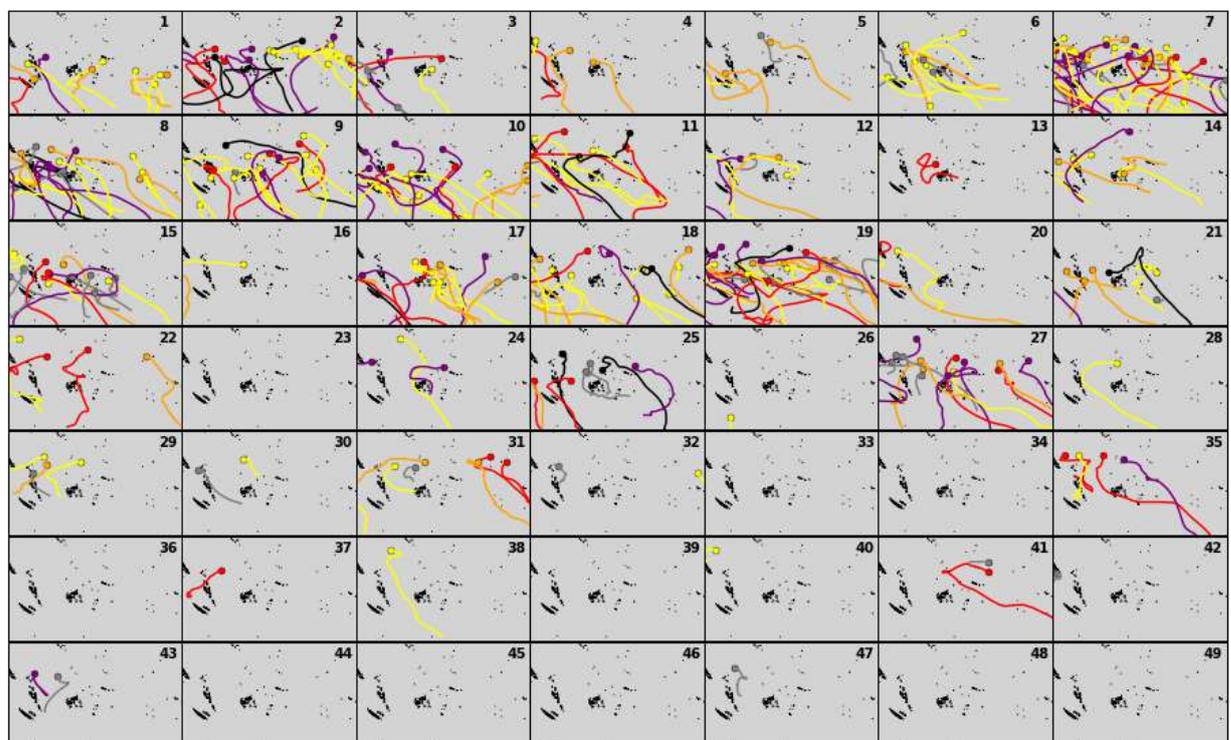


Figure 38 TCs transferred into DWTs according to genesis point.

However, [Figure 38](#) only considers one point of the TC track as the criteria to transfer it to a specific DWT. So, then the TCs tracks have been transferred into the DWTs taking into account each segment that forms the track and the result is shown in [Figure 39](#).

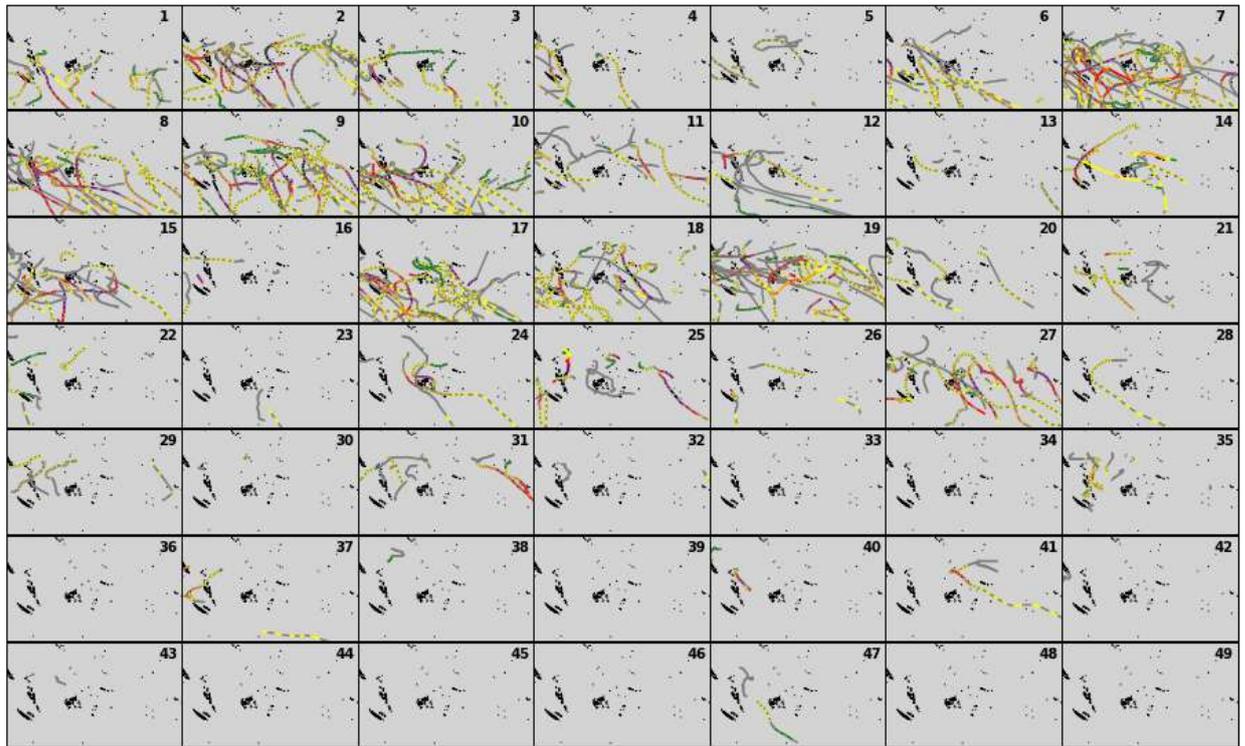


Figure 39 TCs transferred into DWTs according to each segment of the track.

The resultant pattern in [Figure 39](#) is different from the one observed in [Figure 38](#). In some cases, the DWTs show a similar configuration with little or no difference. For instance, DWTs 8, 9, 10 or 16. DWT 16 has only one TC track, and it is completely maintained when transferred by segments. DWTs 8, 9 and 10 have much more TC tracks included. Although it is challenging to visually discern if the pattern is the same, it can be stated that it is very similar and it can be seen how the tracks fully progress from the genesis point to the end of the map.

In other cases, [Figure 39](#) shows maps more crowded than [Figure 38](#), when the TC tracks are transferred according to the genesis point. For instance, DWT 3, 17 and 46. DWT 3 is more crowded in the southern half. Similarly, in DWT 17 many segments have been included on the left down corner. DWT 47 in addition to the gray TC track already present in [Figure 38](#) it includes segments from possibly one or more tracks just below that one.

Finally, some DWTs show maps less crowded when the segments are transferred compared to when the genesis point is used as transfer criteria, like DWTs 5, 11 and 38. In DWT 11 the TC tracks stop at the middle of the map not showing their corresponding tails. In DWT 5 for the two orange tracks and in DWT 38 for the only track it includes; the genesis and just the beginning of the TC tracks are the only parts conserved in [Figure 39](#).

For DWT 31 is missing in [Figure 39](#) most of the TC tracks of the western part and the orange track from the eastern area. DWTs 34 and 22 seem to include the same tracks but not in their full length. Additionally, in DWT 21 not only the tails are not included in [Figure 39](#) but also the segments from the extreme western part (for instance the purple track in [Figure 38](#)) are not included either.

In conclusion, considering only the genesis point as the transfer criteria to the DWTs make us miss some important information. The TCs tracks develop along many days, during which the DWT can change and along a very big spatial domain over the target area. Therefore, to establish the statistical relationship between the predictand and the predictor, the following procedure has been followed.

Firstly, the target area is divided in squared cells of 8×8 grades. For each cell, the total number of TCs going through for all the calibration period (1982 – 2019) has been counted conditioned to each DWT. For each cell each TC track going through it has been identified and the daily segment corresponding to the minimum pressure value of the track part going through that specific cell has been selected. Then the corresponding date of the selected segment is looked at and the TC track is counted in the corresponding DWT to which that daily date belongs. This way, the total absolute number of TCs going through the target area discretized in $8 \times 8^\circ$ cells for each DWT for all the calibration period is computed and the result can be seen in [Figure 40](#).

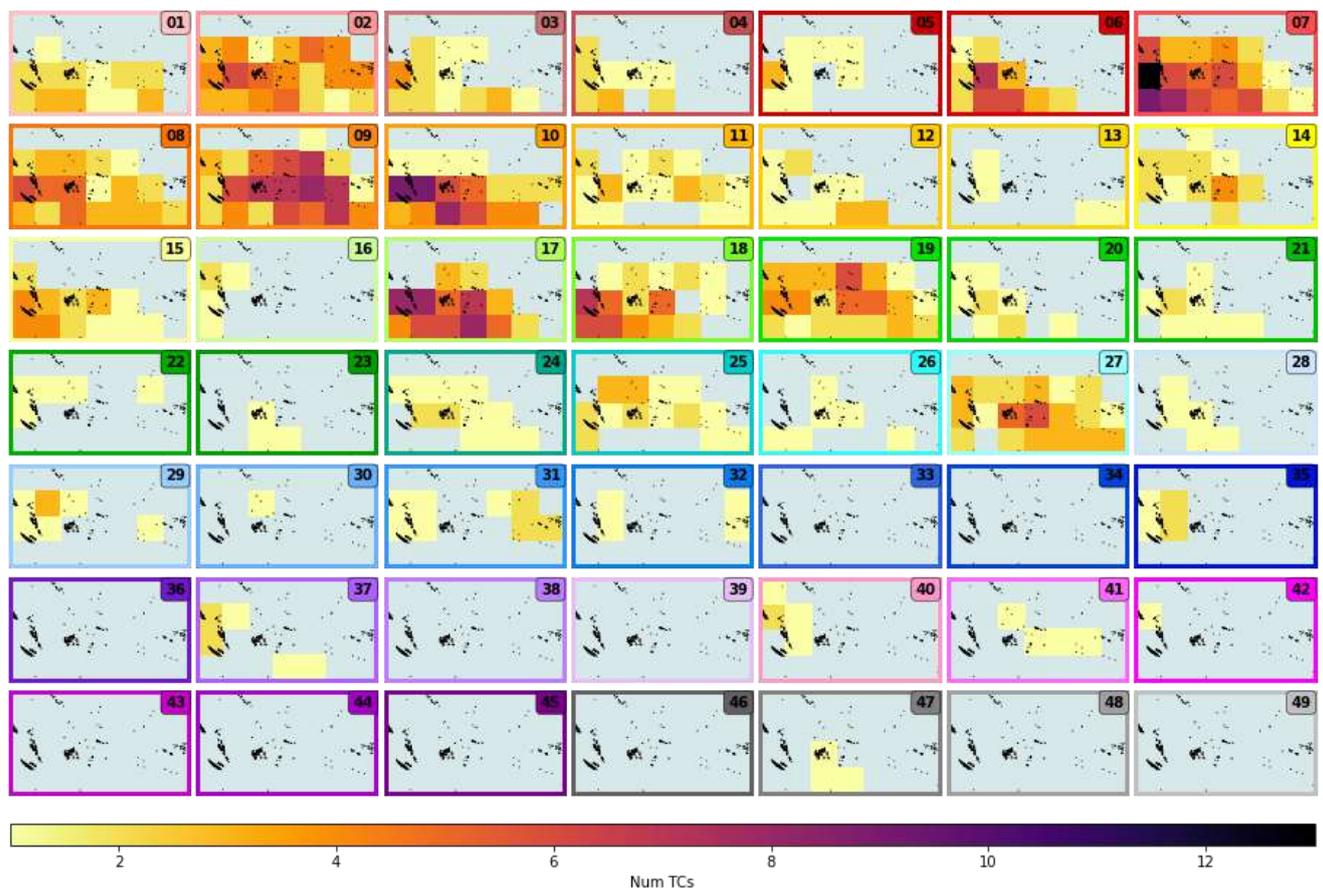


Figure 40 Absolute number of TCs going through the target area in $8 \times 8^\circ$ cells during all the calibration period for each DWT.

Accordingly, to what it has been discussed so far, the first 28 DWTs include most of the TC activity.

However, what is sought is a predictand variable in a daily basis. Therefore, the result from [Figure 40](#) is resampled to obtain the mean expected number of TCs going through the target area in $8 \times 8^\circ$ cells for each DWT. So, the resultant absolute number of TCs for each DWT_i is divided

by the total number of days included in that DWT_i. The resultant predictand is shown in [Figure 41](#), where the probability of each DWT is included as background color in a blue scale.

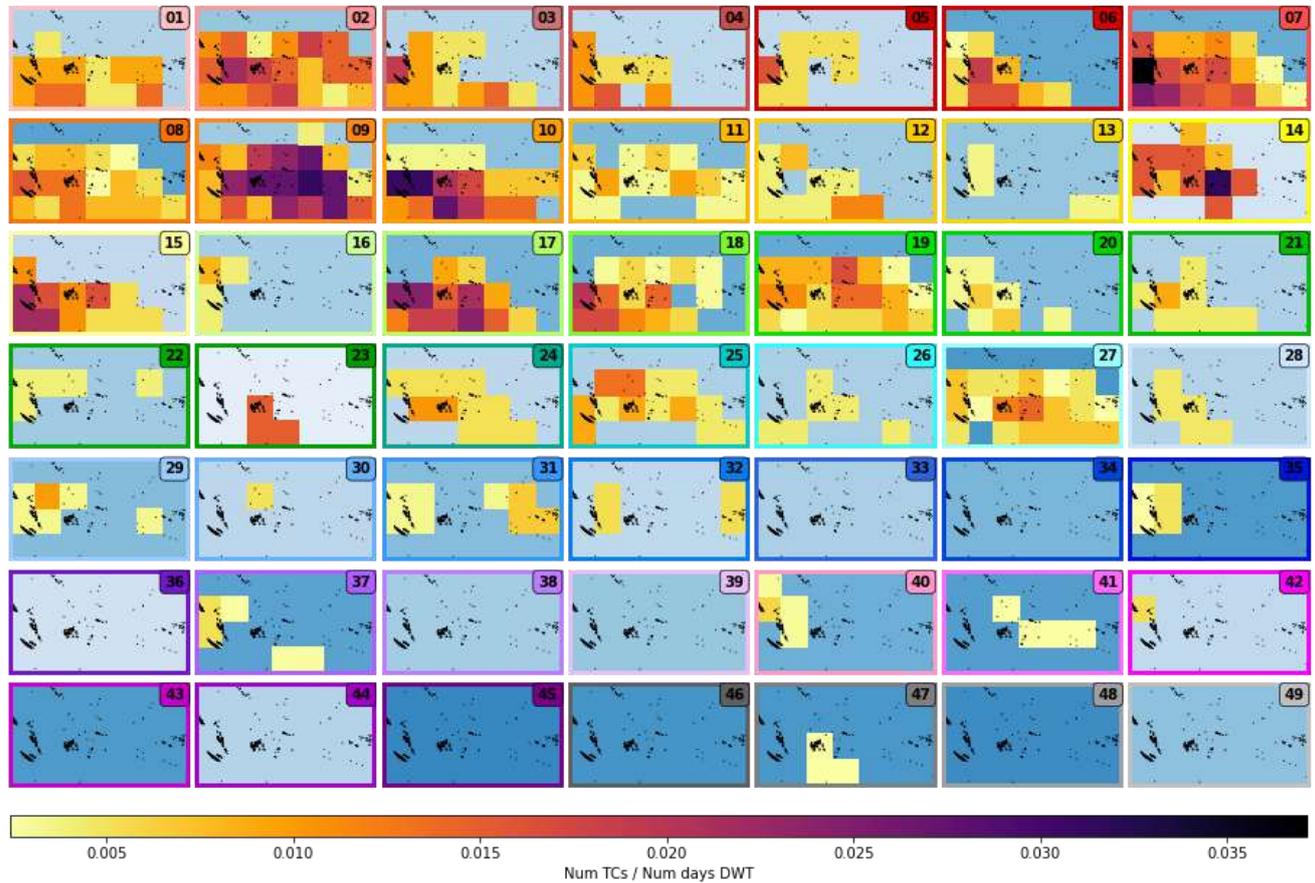


Figure 41 Mean expected number of TCs going through the target area in 8x8° for each DWT.

It must be recalled that in [Section 5.1](#) the index predictor value was generated for all the calibration period. Afterwards, it was partitioned into 49 clusters (DWTs) and therefore each day of calibration period was then included in a DWT. Finally, the mean expected number of TCs going through the target area (which was previously discretized in squared cells of 8x8°) has been computed for each DWT. Consequently, each day of the calibration period has also each mean expected number of TCs (predictand) map.

Two figures have been made, in which the monthly aggregated mean expected number of TCs is shown for the whole calibration period ([Figures 42](#)) and for only the year 2015 ([Figure 43](#)); to be able to observe more in detail a specific year. Therefore, daily mean expected number of TCs of each day have been summed up to compute the plotted variable, in order to have the information organized by months and allow an easy visualization. [Figure 43](#), additionally to mean expected number of TCs it also includes the same information but filtering from TCs reaching category three or greater through dots centered in the 8x8° square cells. The historical tracks from the IBTrACs database have also been included. A track is plotted in a month if it has dates belonging to it, so some tracks can appear in two consecutive months and in [Figure 43](#) they have been colored according to their category.

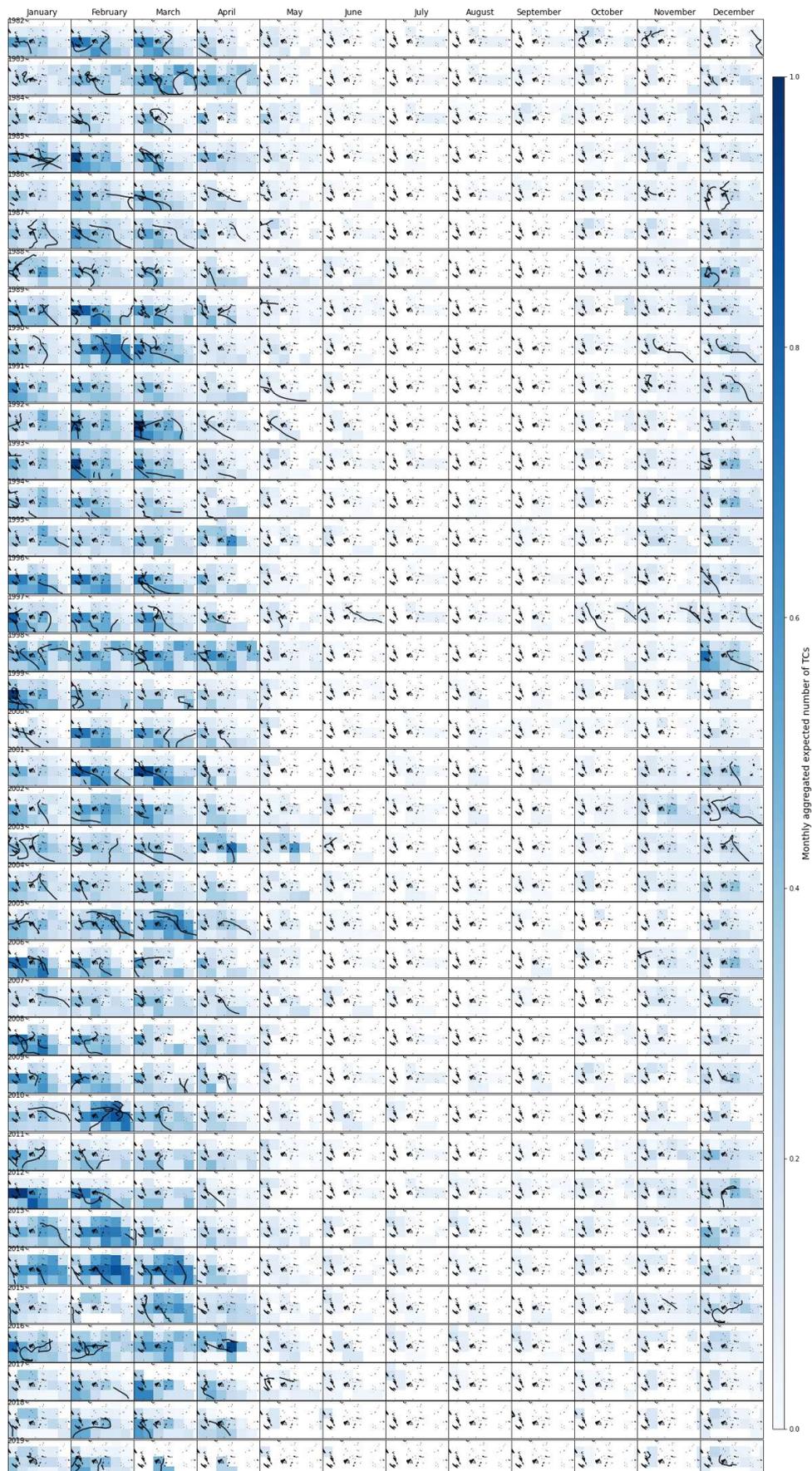


Figure 42 Monthly aggregated mean expected number of TCs for all the calibration period.

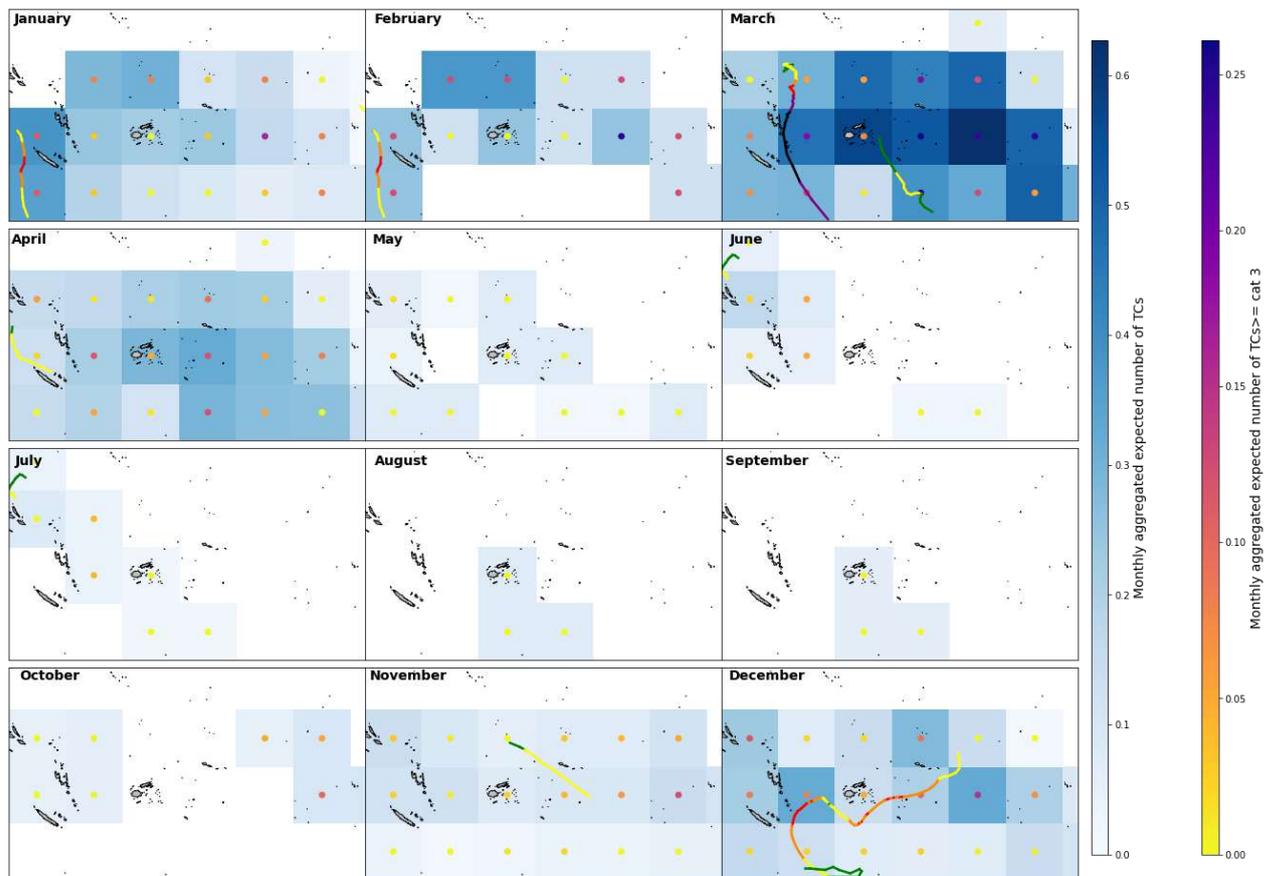


Figure 43 Monthly aggregated mean expected number of TCs for year 2015 of the calibration period.

From the previous figures it can be inferred that the methodology works generally very well.

As it was explained before, the months from November of one year to April of the following year form the TCs season, whereas the rest of the months have very few or no TC activity and do not belong to the season. Consequently, during the months from May to October a null or very low mean expected number of TCs is observed in the target area in general, except some cases where TCs took place. For instance, in October 2006, in June and October 1997 and in May 1992. Also, in the [2015 Figure](#), it can be observed how the mean expected number of TCs is generally null (August, September, May, June or July) or very low (October) out of the TCs season. However, the TC that happened between June and July, in the upper western part of the target area, is accurately predicted since in that zone the maps turn bluer.

Focusing on the TCs season months, relevant conclusions can be extracted. The methodology shows higher mean expected number of TCs, as higher number of TCs are present. For instance, using the Figure of the year 2015 as representative of the calibration period; in the January-February-March months, that include most of the historical TC tracks, the expected mean number of TCs is clearly greater. Additionally, the dots are darker in the zones where the TC tracks reached higher categories, as it can be noticed through the purple points around the TC March TC track. However, for the November and June-July historical tracks the dots remain yellow, agreeing with the fact that these TCs did not reach category 3. Therefore, the methodology is accurate and effective when it comes to reproduce the quantity and intensity of TCs in the area.

Additionally, in some cases it can be noticed the footprint of the TC track in the next or previous month. For instance, in the western track from June 2003 and the Northern vertical track from March 2019 are again reproduce in May 2003 and April 2019 respectively. This can be due to the proximity in dates of the TCs to the following months or because the TC was particularly intense.

In some cases, the model nearly exactly reproduces the TC track, for example, in May of 1992 and 2017 and two tracks in October 1997. In Figure 2015, it also happens in the January and February northern-western TC track, where the methodology follows the path of the TC. However, in many cases rather than accurately reproducing the historical track, the model captures the area through which the TC goes through. This area includes the actual track and the surroundings, sometimes very extended. Examples of this pattern are March 2001 and February 1990 or 2018. It can be better observed in detail this in [Figure 43](#) in November and December months. In these months, the blue color extends through a big area of the map and similarly so do the darker dots representing the category 3 expected TCs. Nevertheless, due to the scarcity of the data, the seasonal forecast of TCs must be analysed in aggregated terms at the end of each TC season.

Finally, in some punctual cases the methodology does not perform well. The resulting maps sometimes show a high expected mean number of TCs when there is no historical TC track. For instance, February and March of 1991, or the year 1995 (except January). From their resultant expected mean number of TCs maps it can concluded that a TC was highly probable. Despite this prediction, no TC track occurred at that moment. The same is observed in [Figure 43](#) in for instance, February. In this month, in the eastern part, what it could seem like a possible track with additionally high risk of being equal or greater to category 3, does not result in any TC. Furthermore, in April for example, there is only one historical TC track in the western extreme part, but the blue color extends through all the target area.

In conclusion, the methodology generally performs very well since the developed model can efectively predict the areas trrough which there was a higher TC activity and additionally, through the coloured dots that represent the expected TCs reaching category equal or greater than 3, it can be concluded that the model can also estimate if the TC will become particularly intense.

6 Additional predictand variables

Additional datasets have been downloaded and preprocessed, including the interpolation into the MLD grid resolution of $1/2^\circ$; in order to better understand and explain the oceanic and atmospheric dynamics that are more favorable for TCs genesis and development.

These additional variables are the Sea Level Pressure (SLP) and the daily mean precipitation.

Firstly, [Figures 44](#) and [45](#) show the result of transferring the SLP fields and their corresponding anomalies into the DWTs.

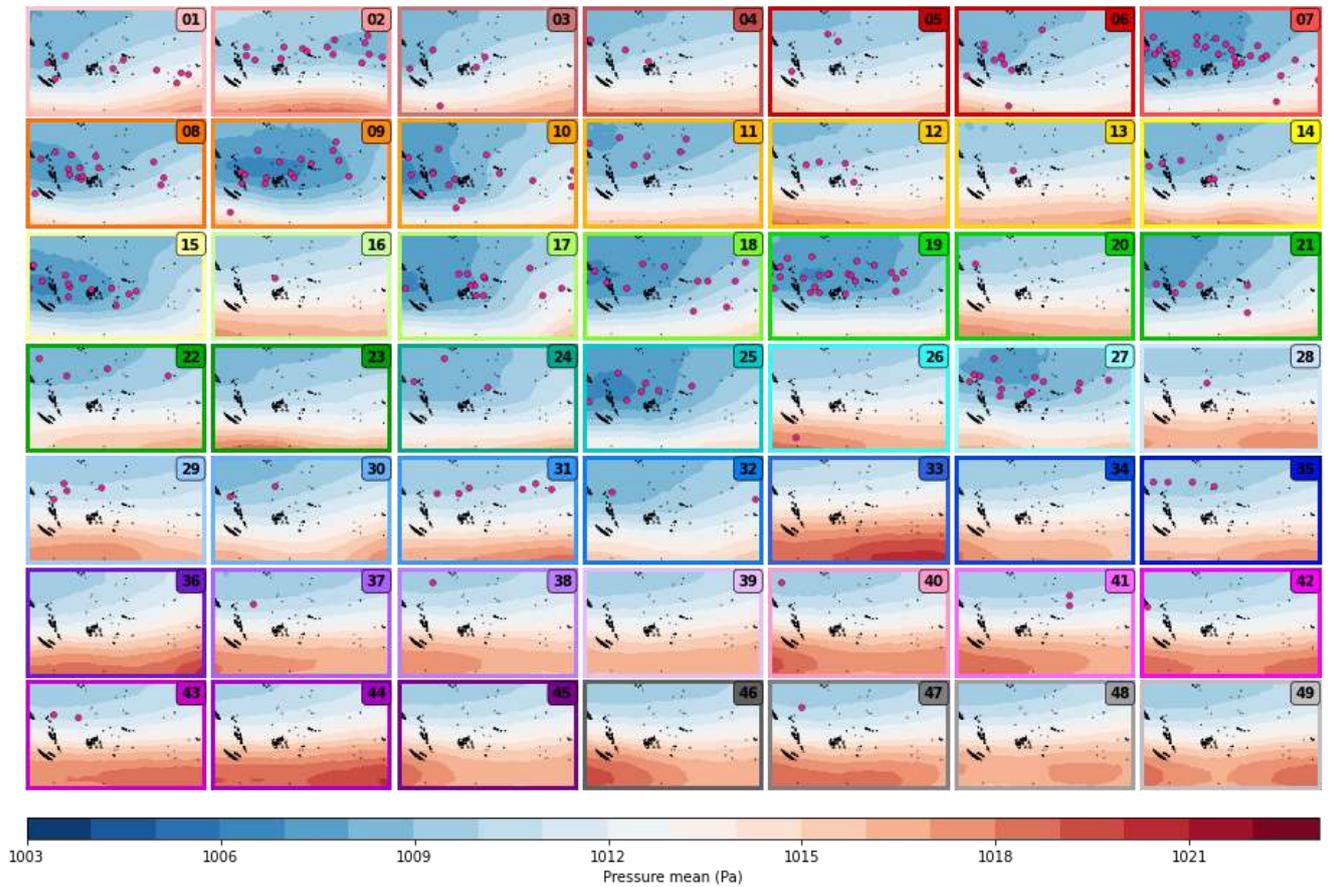


Figure 44 SLP fields transferred into the DWTs.

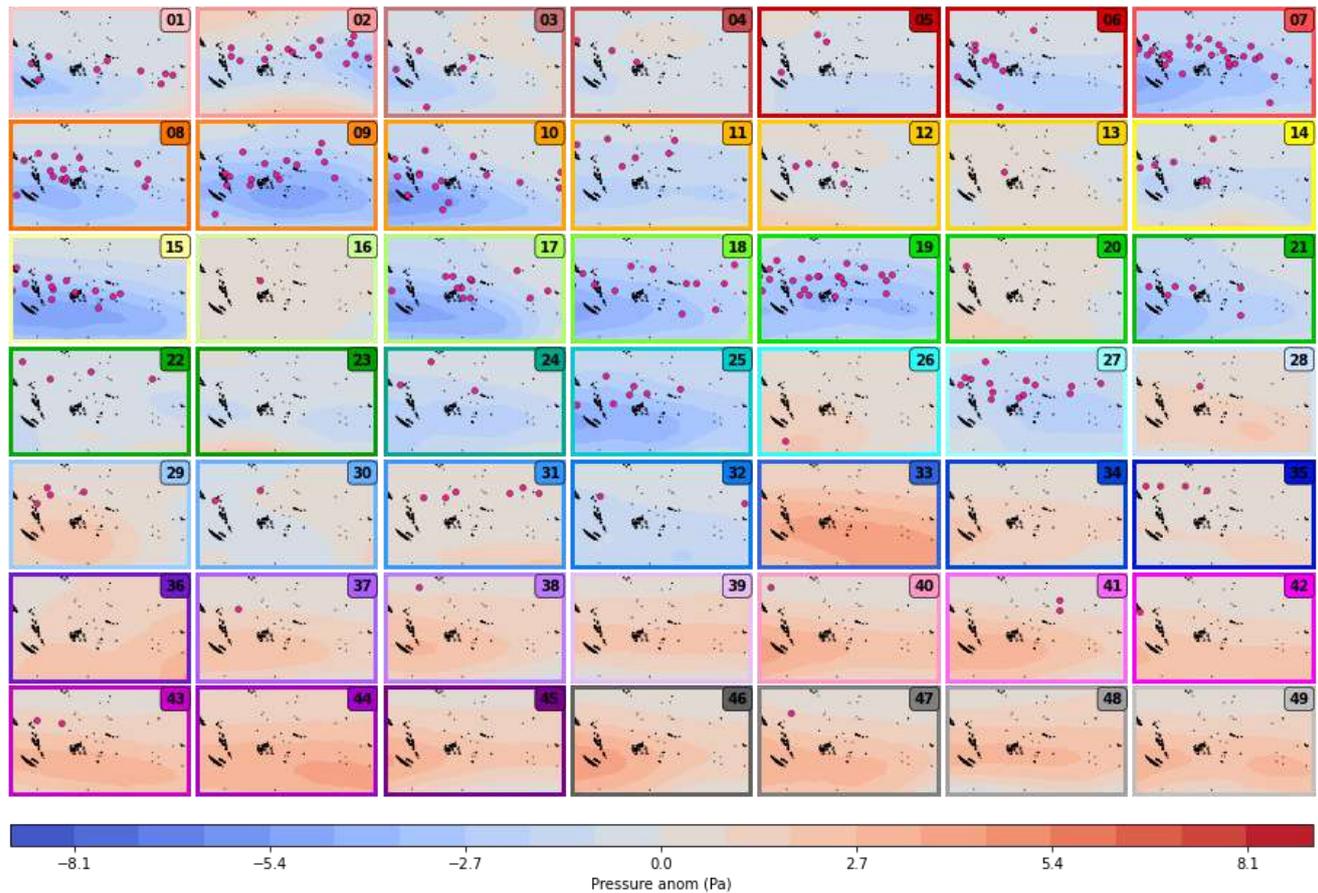


Figure 45 SLP fields anomalies transferred into the DWTs.

The first 28 DWTs, the ones with most of the TCs genesis activity as it is shown in [Figure 27](#); have generally the majority of the area under pressures smaller than 1013 mba, with a small band in the southern part with greater pressure values. This area with bigger pressures gets increasingly larger and intense until it covers nearly half of the maps in the last DWTs. This progression is reflected through the evolution of the anomalies, from completely blue (negative anomalies) maps to full red maps (positive anomalies).

The genesis points are always located (except two southern points in DWT 3 and 6) in areas with pressure equal or less to 1013 mba, that generally correspond to blue anomalies, except in the DWTs 20, 26 and from number 28 onwards, which concentrate a very small proportion of all the genesis activity. Some of the DWTs with the highest TCs genesis activity have all the area under pressures smaller to 1013 mba, like DWTs 7, 9, 10 and 15. However, the TCs signal should be removed in future works to be sure that the low-pressure values in the area are not the TCs themselves.

Then the figures with the TRMM daily mean precipitation transferred into the DWTs are displayed.

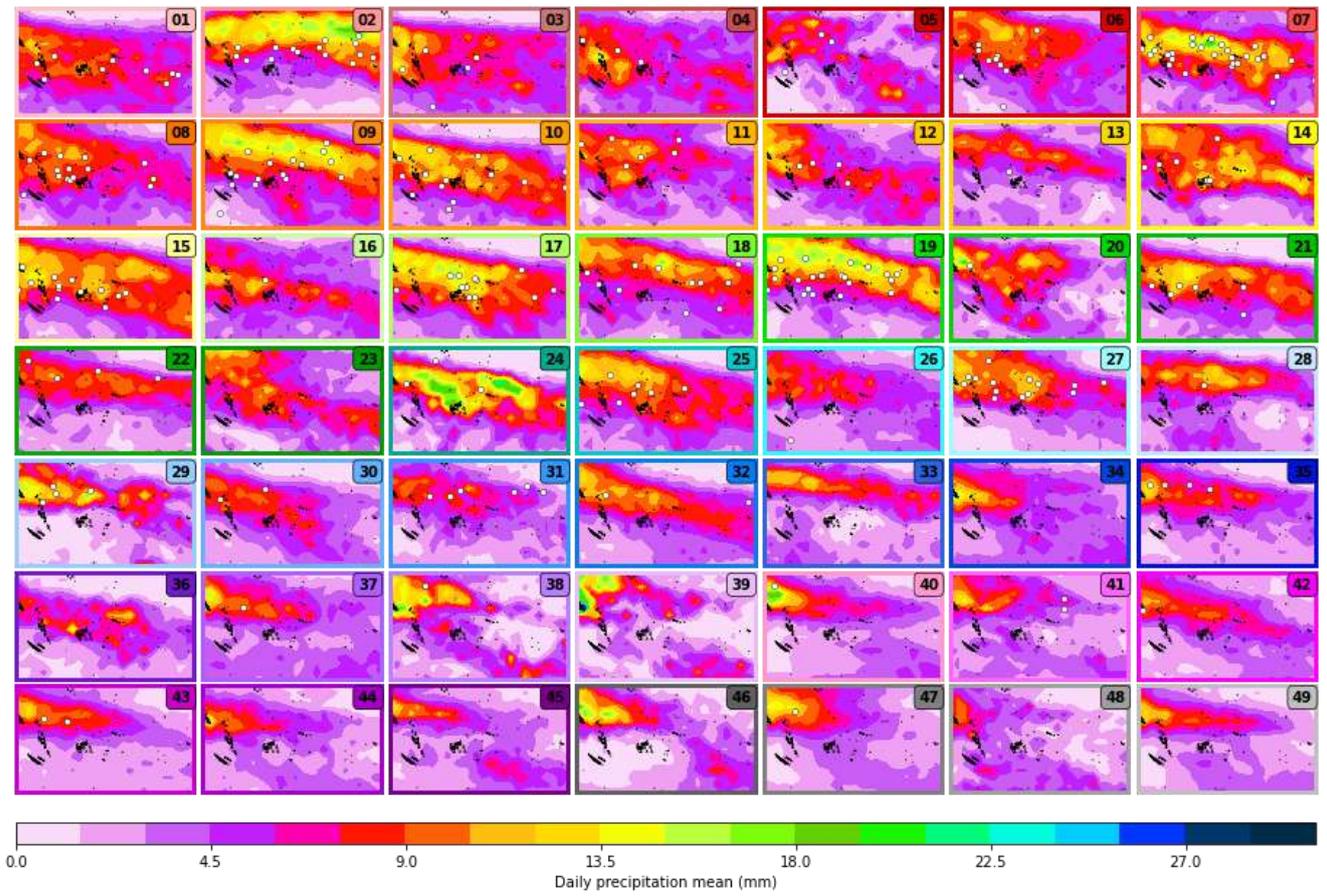


Figure 46 Precipitation hindcast (TRMM) transferred into the DWTs.

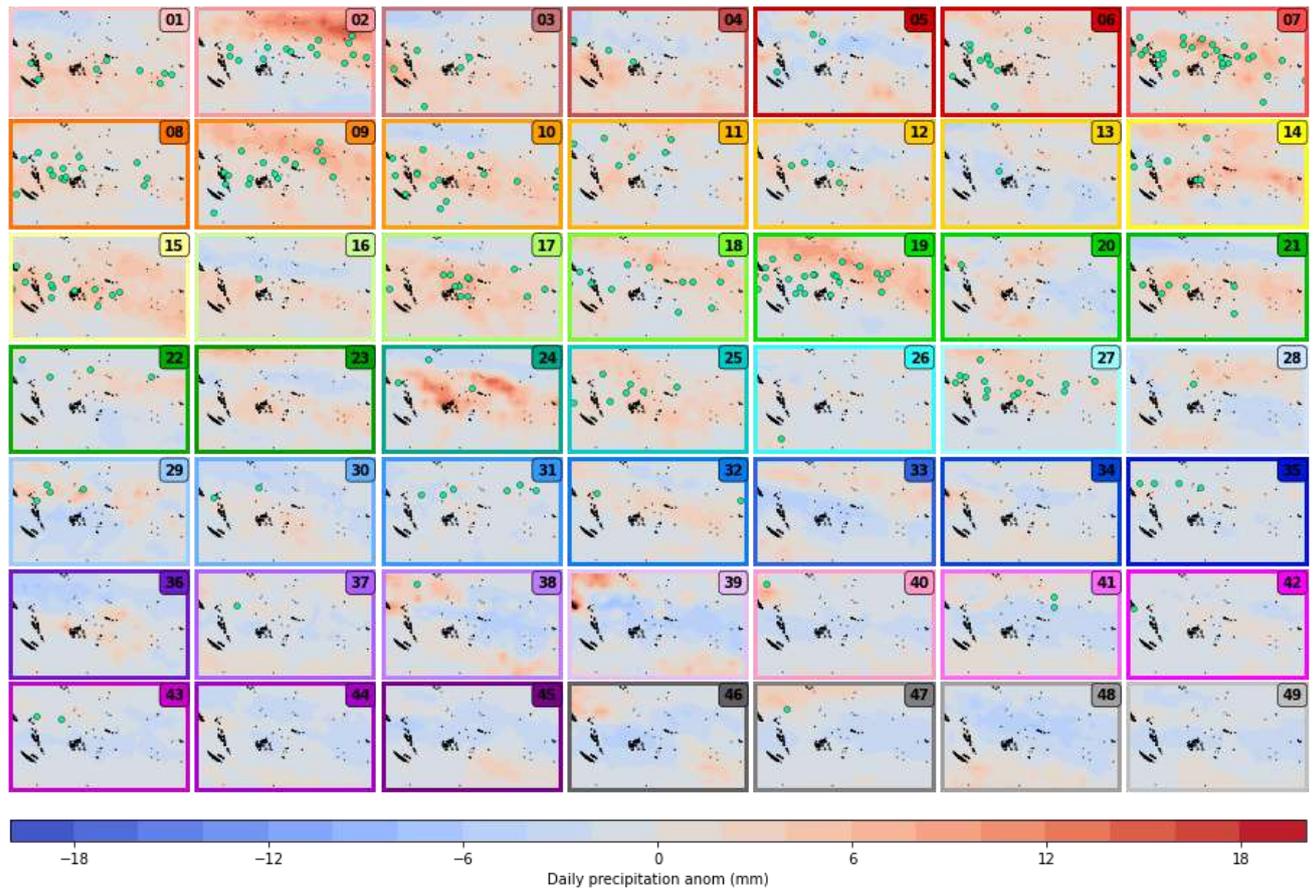


Figure 47 Precipitation hindcast (TRMM) anomalies transferred into the DWTs.

In [Figures 46 and 47](#), the same pattern of the intense flame, observed before in the index and with the SST can be noticed. The DWTs until DWT 28 have a big and intense precipitation cloud, that covers a great part of the southern area and it is represented as a positive anomaly (red). From this DWT onwards, the intense cloud, as it happened with the index and with the SST flames; it becomes smaller and less intense. In this case, the DWTs turn to be more homogeneous, with less contrast since the great differences between such an intense precipitation cloud and the southern part with nearly no precipitation disappears as the intense cloud extinguishes. The DWTs from 33 to 49 are quite similar, with a small medium intense zone in the upper western part of the area and the rest under 7.5 mm/day. Consequently, progressively the anomalies generate bluer (negative anomalies) maps.

The TCs genesis points are generally located in the intense precipitation cloud, in the range from 9 to 16.5 mm/day, which corresponds to red areas in the anomalies maps. However, there are some exceptions to this. Nevertheless, in DWTs from number 28 onwards, which as explained in previous sections concentrate a very small portion of the TCs genesis activity ([Figure 27](#)); the genesis points are generally found in precipitation zones with lower values, in a range from 6 to 12 mm/day, corresponding to blue areas in the anomalies, although not very intense. This is the case of for instance DWTs 29, 30 and 35.

The findings with respect to the precipitation reaffirm the conclusions drawn in [Section 4](#); where TCs genesis was found to generally occur in areas of intense precipitation, above the mean

(positive anomalies) and a higher proportion of TCs reaching the highest categories was linked with the strongest precipitation zones.

After the analysis made in [Sections 4, 5](#) and [6](#), all the conclusions drawn are briefly outlined below.

The TCs are always originated below the equator, most of them in latitudes from -15 to -10. Then they progressed toward the North with different trajectories until going out of the target area. TCs season includes from November of one year to April of the following year.

Regarding the large-scale predictors, MJO and AWT, it has been found that MJO phases 6,7 and 8 and AWT 1 and 3 show the highest TC genesis activity. However, El Niño (AWT 0), which is highly unlikely has the highest proportion of TC genesis days amongst its days and a greater proportion of TCs reaching at least category 2.

Focusing on the tailor-made predictor, most of the TC activity is located in the first 28 DWTs. During TCs season months these first 28 DWTs are highly likely, whereas during the rest of the months DWTs from 28 onwards are the most probable ones.

With respect to the relationship between the predictand and the predictor variables, TC activity takes place generally under the following conditions:

- Index range values from 0.60 to 0.79, corresponding to positive anomalies.
- In the warm SST zone, from 28 to 30 °C, and where MLD values are smaller than 75 m; corresponding to mild positive SST anomalies and negative MLD anomalies.
- In intense but not extreme precipitation areas, from 9 to 16.5 mm/day, corresponding to positive anomalies.
- In low pressure areas, with 1013 mba or lower values, corresponding to negative anomalies.

7 Model validation

The model has been validated applying it to data from the year 2020, not included in the calibration time period from 1982 to 2019 ([Figure 7](#)).

Similarly, to what was carried out for the calibration period, historical data, in this case from the year 2020, has been downloaded and preprocessed from the same databases used for the calibration period. Predictor historical SST data comes from the NOAA 1/4° daily Optimum Interpolation SST and the MLD data is obtained from the NCEP Climate Forecast System Reanalysis (CFSR). First SST and MLD variables were resampled to compute the daily mean data and then they were discretized into the same grid with resolution of 1/2° ([Figure 19](#)).

Afterwards, the predictor index values for the validation period are generated for the target area based on the index function obtained at the calibration period, [Equation 1](#).

Then, the fitted Principal Component Analysis (PCA) for the calibration index predictor is used to predict the index principal components (PCs) in that same temporal-spatial space.

The result is shown in [Figure 48](#), where the index predictor data for year 2020 is shown in the first three first PCs space, including the K-means classification obtained before for the calibration period. It can be seen, as it happened in [Figure 25](#), that the black dots (the centroids of the 49 clusters), span the wide variability of the data; although in this case there is much less data since is only one year.

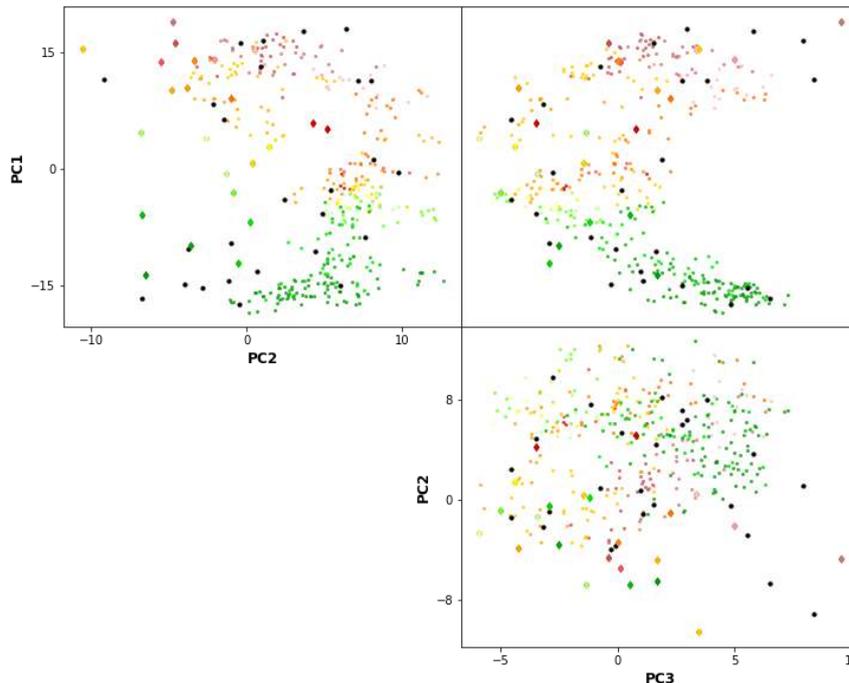


Figure 48 Index predictor data in the first three PCs space.

Then, over the K-means clustering of the index predictor in the PCA space obtained for the calibration period, the N = 49 clusters (the DWTs); new DWTs can be predicted for year 2020 by including its index predictor values in the corresponding cluster.

In the following [Figure 49](#) the resultant DWTs chronology is illustrated, with the DWTs bidimensional lattice and corresponding colors on the right side to recall the color code.

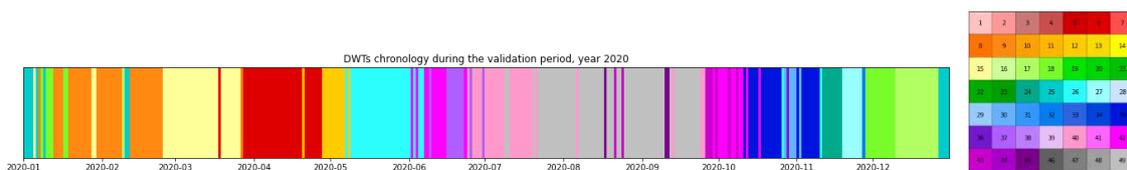


Figure 49 Validation period (year 2020) DWTs chronology (left) and DWTs bidimensional lattice and colors (right)

During the year a high variability is observed, resulting in a quite colorful chronology in some parts of the time period. For instance, during the TCs season, DWT 5 (red), 15 (pale yellow) and 9 (orange) are very common.

Additionally, a cluster comparison between the calibration and validation periods is carried out for the two most common DWTs during the validation period, 9 and 49. The result is shown in in [Figure 50](#). The first row includes the plots from the validation period and the second row from the calibration period. The first two columns include the mean of all the days belonging to each DWT during both time periods. The last two columns include only the first day belonging to each DWT from both time periods.

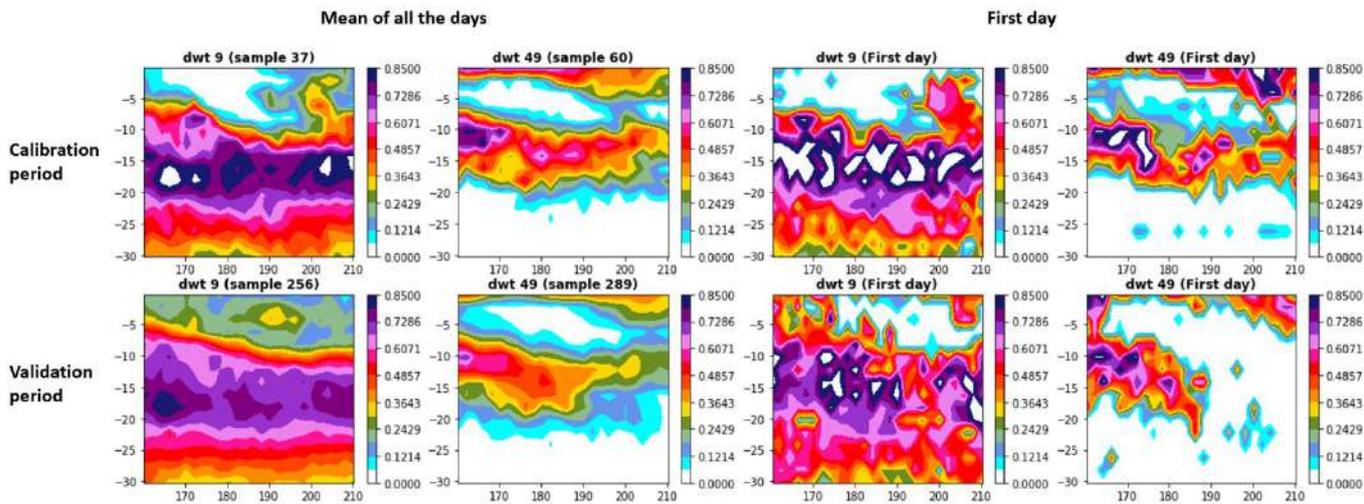


Figure 50 Cluster comparison between the calibration and the validation period.

The computed means for the 2 clusters are quite similar between the calibration and the validation periods. The plots show the same range of index values and the same pattern. The only small difference is that the validation plots reach greater index values in the areas where the index reaches the highest values, in the zone from 0.7286 to 0.85. In fact, the mean of DWT 9 for the validation period becomes white in the central part of the intense flame since it reaches values higher than the plotted range (up to 0.85). In this plot the index also has smaller values in the southern part of the map, reaching 0, whereas in the calibration period remains above it.

The first days belonging to DWTs 9 and 49 are also similar to each other for both time periods. Although they show greater differences than in the first row, they are still very alike with respect to the patterns and the range of index values displayed. For DWT 49, in the validation period the cloud with index value equal to 0 extends covering most of the plot except a cloud with much more intense index values in the central part on the left; whereas in the calibration period plot this intense cloud extends from the left and covers all the central part of the plot until the right border. In the case of DWT 9, the index range values, and the patterns displayed are very alike, except for the upper part. In both plots there is a zone of zero index values. However, in DWT 9 it is located in the whole upper left part and in DWT 49 it is a bit smaller, and it is situated in the right part of the plot.

The similarity between the index predictor plots of both time periods, calibration and validation, confirms the good performance of the statistical downscaling so far.

As it was computed before (Section 5.2.3), each DWT has a map of mean expected number of TCs going through the target area discretized in $8 \times 8^\circ$ cells associated. The resultant daily predictand for each is aggregated in monthly basis, as it was done for the calibration period and also for the whole year aggregated.

The resultant figure with the monthly aggregated mean expected number of TCs is shown for the whole validation period. The monthly aggregated expected mean number of TCs is included in the blue scale and the same information but filtering from TCs reaching category 3 or greater is included through dots in the middle of the $8 \times 8^\circ$ cells. Finally, the historical TCs tracks from the corresponding year 2020 from the database IBTrACs have been included. A track is plotted in a month if it has dates belonging to it, so some tracks appear in two consecutive months.

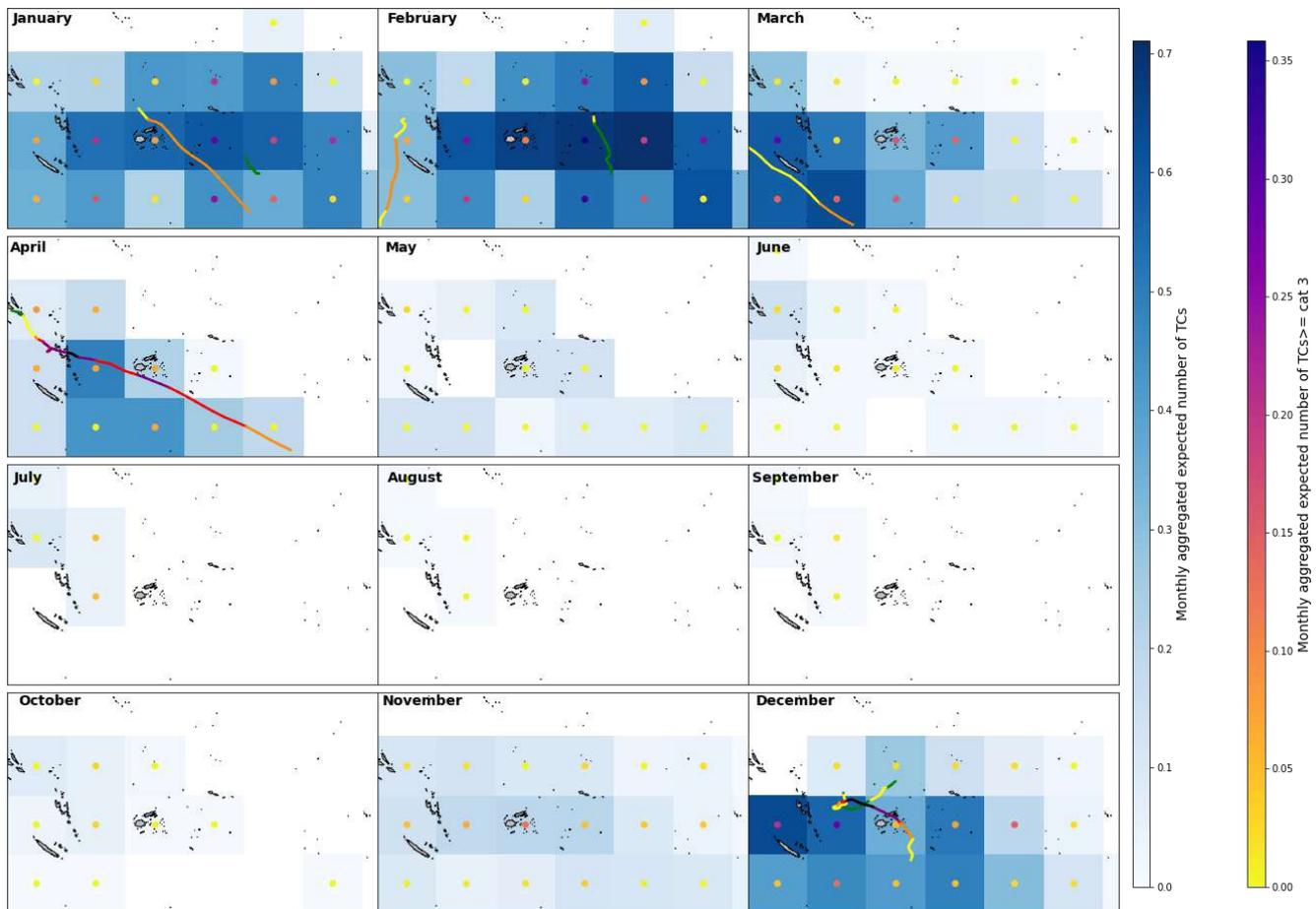


Figure 51 Monthly aggregated mean expected number of TCs and historical tracks of year 2020, the validation period.

Figure 52 shows the result of aggregating all the months of the validation period, the whole year 2020.

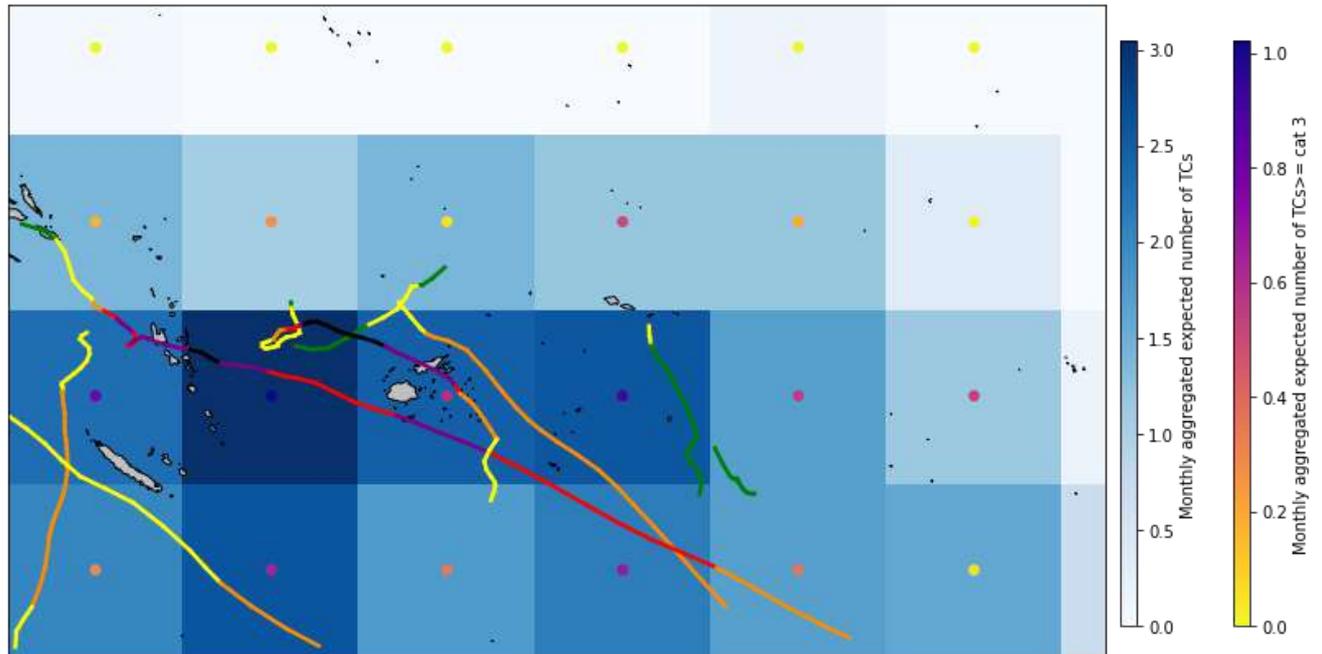


Figure 52 Aggregated mean expected number of TCs and historical tracks for the whole validation period (year 2020).

The months of November, December, January, February, March and April belong to the TCs season. The months from May to October, are not included in the TCs season, show generally very low, like October; or null like June, July, August and September; expected TCs number. However, in May and November, the expected TCs number is slightly here. This can be explained looking at April and December maps. In April there is a very intense TC, named Harold, which happened during the first two weeks of April. Due to the proximity of the dates the footprint of the TC track can be clearly detected in March, especially in expected TCs reaching category 3 or higher. Additionally, although with much less intensity since the dates are further away, Harold footprint could also be detected in May.

Focusing on the TC season months, the model expects more TCs in areas where there are TC historical tracks. Additionally, when the TC reaches a higher category, the expected TCs number equal or greater than category 2 is generally also bigger. For instance, around the orange tracks in January and February the dots are darker. Therefore, the model is accurate and effective when it comes to estimate the TCs number and their potential intensity in the target area.

Nevertheless, regarding the path followed by the TC the methodology is not that accurate, as it was noticed also for the calibration period. In some cases, it approximately reproduces the track followed by the TC. For example, in April or even March. However, in other cases, the methodology captures the area through which the TC passes through, including the surroundings. Despite this fact the expected TCs number is generally higher always in the tracks surroundings although the blue color may extend through a bigger area. This can be observed in January and February. In these months the blue color and darker dots extend through a very big part of map and although they are particularly intense in the TCs tracks zone. So, in these cases, although the model effectively predicts the TC occurrence, it does not give a that accurate prediction of the spatial domain the TCs tracks will cover.

Since the objective of this work is to build a seasonal forecast, analysis will be now focused now on [Figure 52](#), the figure that aggregates the whole the validation period.

[This figure](#) confirms the very good performance of the model. The cells with higher TC activity are also the darker ones (with 2.5 of more TCs expected), not underestimating the number of TCs. The cells including parts of historical TCs tracks reaching category 3 or higher show also the darker dots, like the cell in which Harold reaches category 5. In the surroundings of the TCs tracks the expected number of TCs is overestimated since there is no TC track, however the expected number of TCs is not high (generally around 1 for all categories and around 0.4 or less for category 2 or greater) and still these zones could be considered as relevant areas of influence of the TC, especially if they reach the higher categories. Nevertheless, the model does not underestimate the TCs threat, which is of essential importance given the danger of this natural disaster and its potential catastrophic consequences; and thus, the need to be prepared in advance.

In conclusion, the model performs very well since it can effectively predict the occurrence and intensity of TCs going through the target area.

8 [Forecast](#)

8.1 Forecast validation

The model has been applied to reforecast data (forecast data from the past) to validate this forecast data and to analyze how the resultant DWTs and the expected number of TCs vary from one month to another. Forecast time series of the predictand variables, SST and MLD, have been downloaded from the CFS operational 9 month forecast from the year 2011 to year 2019. A forecast has been made for each TCs season from 2011-2012 to 2018-2019 ([Figure 7](#) timeline). The TCs season, as it was previously explained, includes the months of November and December of one year and the months of January, February, March and April of the following year. Taking the month of November as the starting point since it is the first month of the season, the forecast from the first day of the previous four months has been downloaded. This involves the forecast data from the first day of August, September, October and November of the first year that makes up each season.

Each day has 4 runs of the CFS forecast model, made in 4 different moments of the day, at 00:00, 6:00, 12:00 and 18:00 hours. The methodology is applied to each one of these runs and then the ensemble of the four is computed and shown as the resultant TC forecast from the day.

The statistical model is applied as it was done for the validation period.

Firstly, the SST and MLD downloaded data is preprocessed. These variables have been resampled to compute the daily mean data and then they have been discretized into the same grid of $1/2^\circ$ resolution as it has been done for the calibration and validation time periods.

Afterwards, the predictor index values can be generated based on the index function obtained at the calibration period, [Equation 1](#).

Then, the fitted Principal Component Analysis for the calibration index predictor is used to predict the index principal components in that same temporal-spatial space for the forecast time periods. The predicted PCs are assigned to the best match unit group from the fitted K-means clustering, so each day is assigned a DWT.

Finally, the full season aggregated mean expected number of TCs map is computed for all TC categories and filtering from category equal or greater than 3; as it was done for validation period; based on the daily mean expected number of TCs inferred from the DWTs obtained.

The resultant figure, [Figure 53](#), includes in each row the forecast for each TCs season from the past. The first 4 columns represent the forecast from a different day of the four days being considered (first days of August, September, October and November). The last two columns are the all the forecast mean and standard deviation respectively.

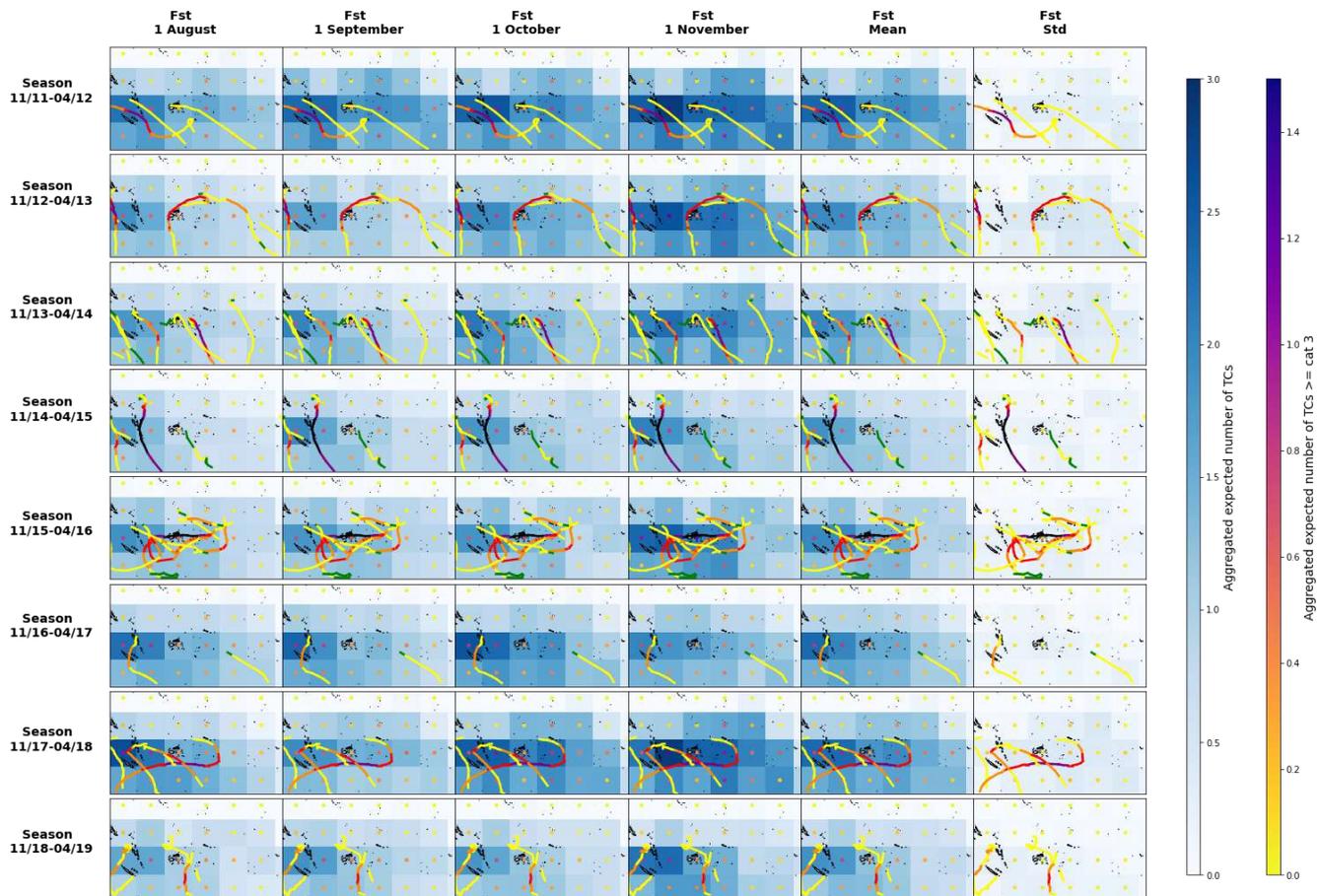


Figure 53 Full season forecasts from past TCs seasons.

It can be observed that as the forecast day gets closer to the start of the TCs season it is predicting the expected number of TCs also increases. In some seasons this tendency can be noticed more clearly, like in season 2012-2013, where it can be seen how the first of November forecast map is much darker than the one from August.

In the cells where there are more TC historical tracks, the expected number of TCs is also higher, effectively estimating the higher TC activity. However, in punctual cases, like season 2015-2016, during which there was a El Niño event; there was an unusually high TC activity, and the forecast expects a smaller number of TCs. Nevertheless, this is not the general trend and the forecast does not generally underestimate the number of TCs going through the area.

Regarding the intensity of the TCs is not that accurately predicted comparing with the conclusions extracted from the validation results, since it generally increases automatically as it increases with the expected number of TCs including all categories. As a result, bluer maps mean usually more purple maps, which it is not always the case, since a higher expected number of TCs does not necessarily mean a higher expected number of more intense TCs. This is clearly seen, for instance, in season 2016-2017, where no TC track reaches category 3 or greater.

With respect to the prediction of the spatial domain the TCs tracks will cover and the overestimation of TC activity in the surroundings of these tracks; trend already present in the calibration and validation figures; here this fact it is enhanced. The blue color tends to extend more through the area despite not including historical TC tracks, like in seasons 14-15 or 16-17. Although generally the expected number of TCs is higher in the areas including TC activity, sometimes the blue color extends more homogeneously through most of the area, like in season 16-17, not giving an accurate prediction.

From the [validation of the methodology section](#), it was concluded that the developed statistical model can effectively predict TC activity. However, the outcome from the forecast data is not the same. It underestimates the TC risk when there is an unusually high TC activity. Furthermore, its performance with respect to the prediction of the spatial domain of the TCs tracks is clearly worse since it tends to generate more extended (bluer) and homogenous maps of expected number of TCs. Finally, it tends to increase the expected intensity when it increases the expected number of TCs of all categories. Therefore, to find the reasons leading to these differences in the performance of the statistical model with respect to the calibration and validation periods, the following [Figure 54](#) with the mean probabilities of the predicted DWTs for the each TCs full season is shown.

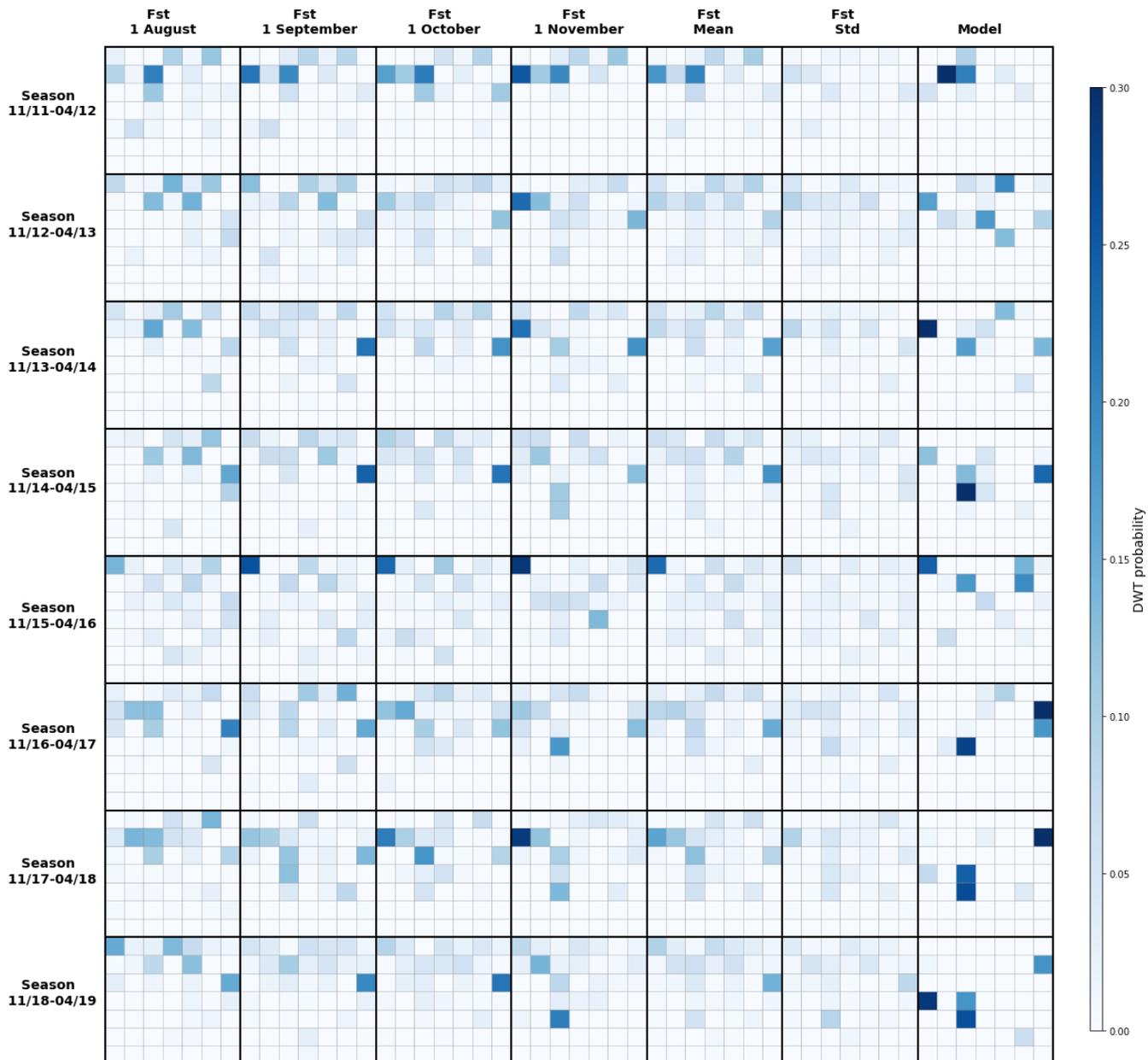


Figure 54 DWTs mean probabilities for the full TCs season.

In [Figure 54](#) it can be noticed the clear difference between the DWTs resultant from the forecast data and the DWTs from the calibration period, so resulting from the hindcast data (model column). The model column has the DWTs probability focused in a much smaller number of DWTs. On the contrary, in the plots of the forecast data it can be seen how the probability it is shared amongst a much greater number of DWTs. As it was explained in the previous sections of this work, the first 28 DWTs concentrate most of the TCs genesis activity of the calibration period. As a result, in the seasonality plot, [Figure 34](#), in the months belonging to the TCs season, the DWTs probability is always focused on the upper half of the DWTs lattice (first 28 DWTs). However, in the forecast data plots the probability extends, in many cases homogeneously, nearly reaching all the rows of the DWTs matrix, like in season 2012-2013. This difference reflects the high level of uncertainty of the forecast data, that results in much more extended mean expected number of TCs maps in the target area.

To conclude, the statistical model has proved its effectiveness and accuracy for the calibration and validation time periods. However, for computing the seasonal forecast its performance greatly depends on the quality of the forecast data, so the resulting expected TCs maps are more reliable and effective in predicting TC risk. Unfortunately, the general performance of the forecast data has ample room for improvement.

8.2 Forecast application

The forecast data from the first day of April of this year, 2021, has been downloaded and preprocessed to compute the seasonal forecast in the target area from this day onwards.

The same procedure used in the [previous part](#) of this [section](#) with the forecast data from the past seasons has been followed here. Thus, the statistical model is applied to the downloaded and preprocessed data from each run of the model of this day (00:00, 6:00, 12:00 and 18:00 hours); obtaining the results for the predictand variable. Then, the ensemble of the four runs is computed as the final outcome.

The resultant TCs prediction maps are displayed in [Figure 55](#). The figure has the same variables and structure of the previous forecast maps, with the mean expected number of TCs in blue and the mean expected number of TCs reaching category equal or greater than 3 represented through dots in the middle of the $8 \times 8^\circ$ cells. The monthly and full season aggregated mean expected number of TCs maps are shown.

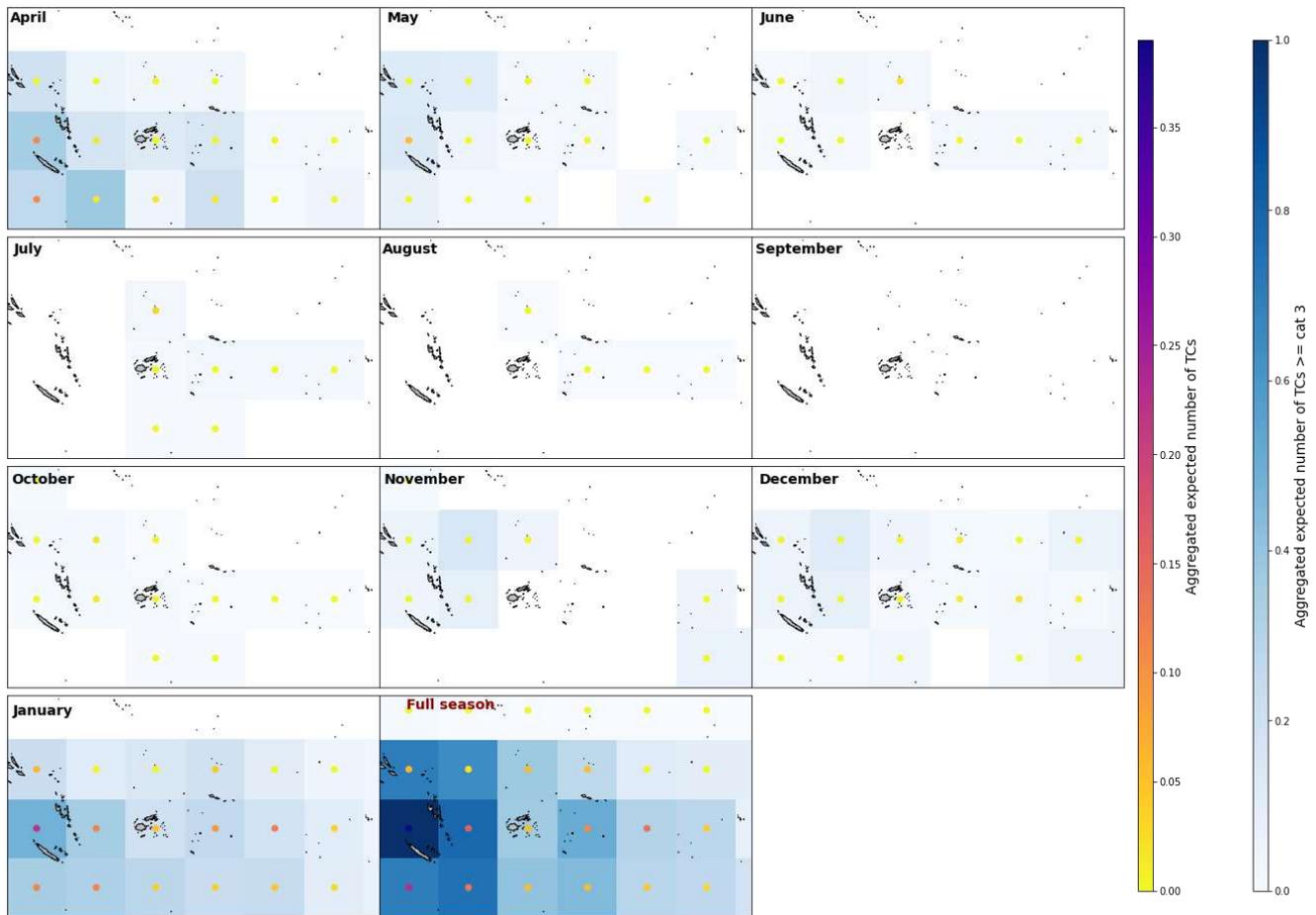


Figure 55 Forecast from day 1/04/2021.

April represents the end of the TCs season. Consequently, the mean expected number of TCs from this month onwards until November is very low or null. In December a clear increase can be noticed as the most intense months of the TCs season are reached. However, these intense months are very far from the day when the forecast has been made and therefore, the prediction is very uncertain yet. As it was analyzed before, the predictions tend to become more intense since as the day of the forecast gets closer to the predicted months the mean expected number of TCs also increases. Therefore, it is expected that the may prediction would give a higher mean expected number of TCs and so on.

Nevertheless, as it was just concluded, although the methodology has been proven to perform very well, the accuracy and reliability of the forecast greatly depends of the quality of the forecast data, which has also been proven to be quite improvable.

9 Bibliography

- Arora, K. & Dash, P., 2016. Towards Dependence of Tropical Cyclone Intensity on Sea Surface Temperature and Its Response in a Warming World. *Climate*, 4(30).
- Balaguru, K., Leung, L. R. & Yoon, J.-h., 2013. Oceanic control of Northeast Pacific hurricane activity at interannual timescales. *ENVIRONMENTAL RESEARCH LETTERS*, Issue 8.
- Barnett, J. & Campbell, J., 2010. *Climate change and small islands states. Power, knowledge and the South Pacific*. s.l.:earthscan.
- Berg, R., s.f. *Tropical Cyclone Intensity in relation to SST and Moisture Variability: a Global Perspective*, Miami, Florida: s.n.
- Boucharel, J. y otros, 2016. Different controls of tropical cyclone activity in the Eastern Pacific for two types of El Niño. *Geophysical Research Letters*, Issue 43, pp. 1679-1686.
- Campbell, J. R., 1997. *Examining Pacific Island Vulnerability to Natural*. South Pacific Programme Office, Suva, s.n.
- Camus, P., Mendez, F. J., Medina, R. & Cofiño, A. S., 2011. Analysis of clustering and selection algorithms for the study of multivariate wave climate. *Coastal Engineering*, Issue 58, pp. 453-462.
- Camus, P. y otros, 2014. A weather-type statistical downscaling framework for ocean wave climate. *Journal of Geophysical Research: Oceans*, 22 09.
- D. Anderson¹, A. R. , L. C. , J. A. A. A. F. J. M. ,., s.f. [En línea].
- Diamond, H. J., Lorrey, A. M. & Renwick, J. A., 2012. A Southwest Pacific Tropical Cyclone Climatology and Linkages to the El Niño–Southern Oscillation. *Journal of Climate*, 15 June.
- Diamond, H. J. & Renwickb, J. A., 2015. The climatological relationship between tropical cyclones in the southwest pacific and the Madden–Julian Oscillation. *International Journal of climatology*, 15 April, pp. 676-686.
- Evans, J. L., 1993. Sensitivity of Tropical Cyclone Intensity to Sea Surface Temperature. *Journal of Climate*, June, pp. 1133-1140.
- Gottschalck, J., 2014. *NOAA Climate.gov, science & information for a climate-smart nation*. [En línea]
Available at: <https://www.climate.gov/news-features/blogs/enso/what-mjo-and-why-do-we-care>
[Último acceso: 04 03 2021].
- Hendon, H. H. & Salby, M. L., 1994. The Life Cycle of the Madden–Julian Oscillation. *Journal of the Atmospheric Sciences*, 1 August, p. 2225–2237.
- IFRC, 2008. *Papua New Guinea: Cyclone Guba, Operations Update No.4*. [En línea].

International Research Institute for climate society, E. I. o. C. U., s.f. *International Research Institute for Climate Society | Earth Institute | Columbia University*. [En línea]

Available at: <https://iri.columbia.edu/our-expertise/climate/enso/enso-essentials/>
[Último acceso: 04 03 2021].

Knapp, K. R. y otros, 2010. International Best Track Archive for Climate Stewardship (IBTrACs).

Lavender, S. L., Hoeke, R. K. & Abbs, D. J., 2018. The influence of sea surface temperature on the intensity and associated storm surge of tropical cyclone Yasi: a sensitivity study. *Natural Hazards and Earth System Sciences*, Issue 18, pp. 795-805.

L'Heureux, M., 2014. *NOAA Climate.gov, science & information for a climate-smart nation*. [En línea]

Available at: <https://www.climate.gov/news-features/blogs/enso/what-el-ni%C3%B1o%E2%80%93southern-oscillation-enso-nutshell>
[Último acceso: 04 03 2021].

Ling, Z., Wang, Y. & Wang, G., 2016. Impact of Intraseasonal Oscillations on the Activity of Tropical Cyclones in Summer over the South China Sea. Part I: Local Tropical Cyclones. *Journal of Climate*, 15 January.

Mandal, M., Mohanty, U. C., Sinha, P. & Ali, M. M., 2007. Impact of sea surface temperature in modulating movement and intensity of tropical cyclones. *Nat Hazards*, Issue 41, pp. 413-427.

Met Office, n. m. s. f. t. U., s.f. *Mett Office*. [En línea]
Available at: <https://www.metoffice.gov.uk/>

Mohanty, S. y otros, 2016. Role of Sea Surface Temperature in modulating life cycle of Tropical Cyclones over Bay of Bengal. *Tropical Cyclone Research and Review*, 8(2).

Olsthoorn, A., Maunder, W. & Tol, R., 1999. Tropical Cyclones in the Southwest Pacific: Impacts on Pacific Island Countries with Particular Reference to Fiji. *Climate, Change and Risk*.

Roy, C. & Kvordányi, R., 2012. Tropical Cyclone track forecasting techniques - A review. *Atmospheric Research*, Issue 40-69.

Rueda, A. y otros, 2017. Multiscale climate emulator of multimodal wave spectra: MUSCLE-spectra. *Journal of Geophysical Research: Oceans*, 21 February, pp. 1400-1415.

Sharma, N. & MM, A., 2014. Importance of Ocean Heat Content for Cyclone Studies. *Journal of Oceanography and Marine Research*, 15 May.

Stephens, S. A. & Ramsay, D., 2014. Extreme cyclone wave climate in the Southwest Pacific Ocean: Influence of the El Niño Southern Oscillation and projected climate change. *Global and Planetary Change*, Issue 123, pp. 13-26.

TimeScavengers, 2021. *TimeScavengers*. [En línea]
Available at: <https://timescavengers.blog/climate-change/ocean-layers-mixing/>
[Último acceso: 2021].

Tory, K. J. & Dare, R. A., 2015. *Sea Surface Temperature Thresholds for Tropical Cyclone Formation*, s.l.: s.n.

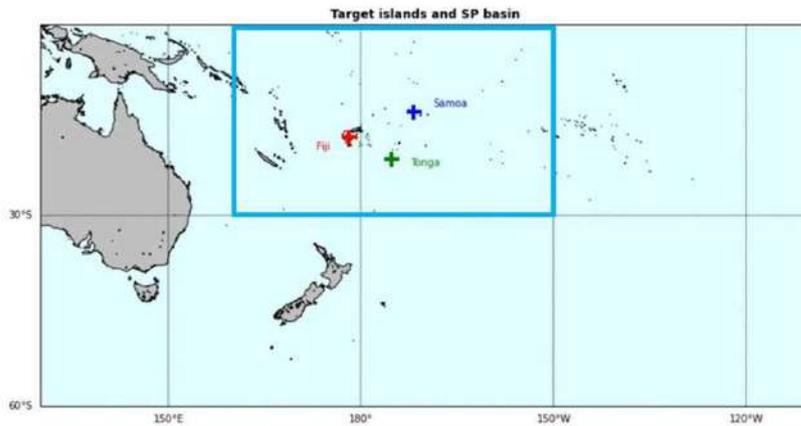


Wheeler, M. C. & Hendon, H. H., 2004. An All-Season Real-Time Multivariate MJO Index: Development of an Index for Monitoring and Prediction. *Monthly Weather Review*, 01 August, 132(8), pp. 1917-1932.

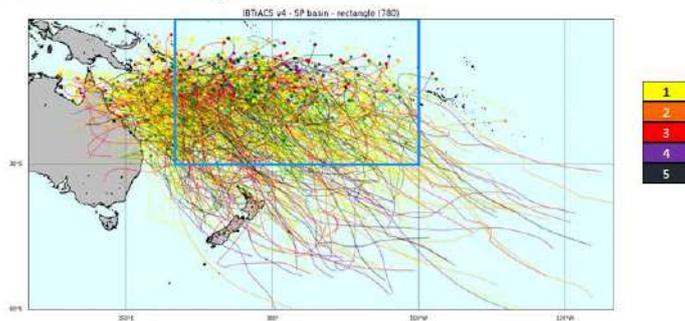
10 Annex 1, The Jupyter book

This work has been carried out with Python in the Anaconda environment, programming through Python language. Several Jupyter Notebooks have been developed, which include the different parts of the work. As a result, the following Jupyter book, which has the notebook as 'chapters' has been created and it is fully displayed just below.

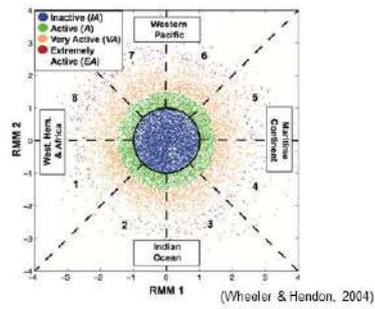
Seasonal Forecast of Tropical Cyclone Activity in the South Pacific Ocean



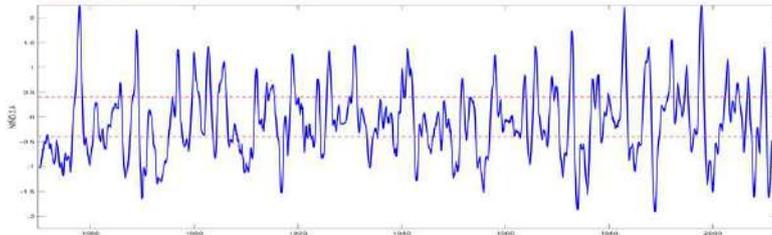
Motivation and objective



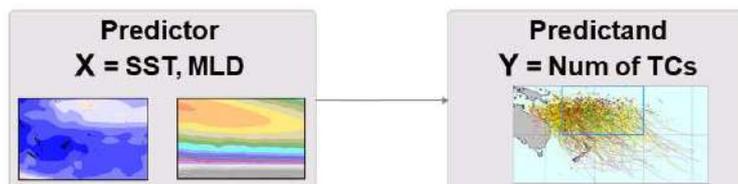
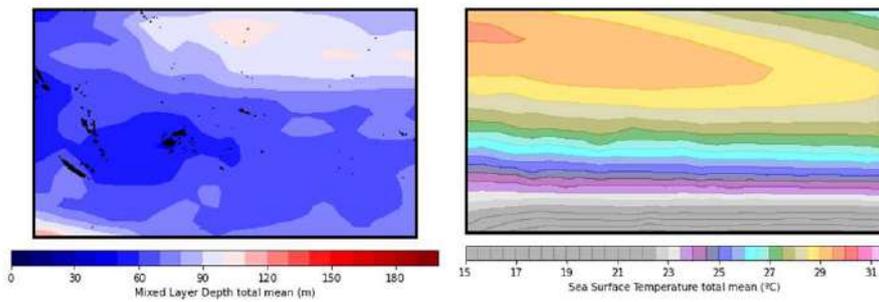
Madden Julian Oscillation

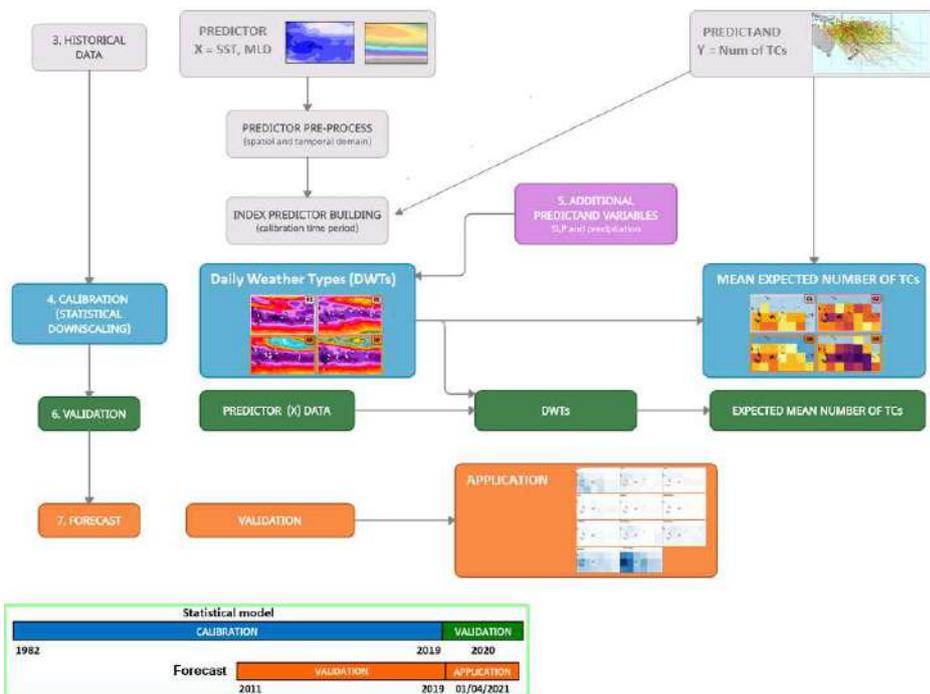


Niño 3.4 Index

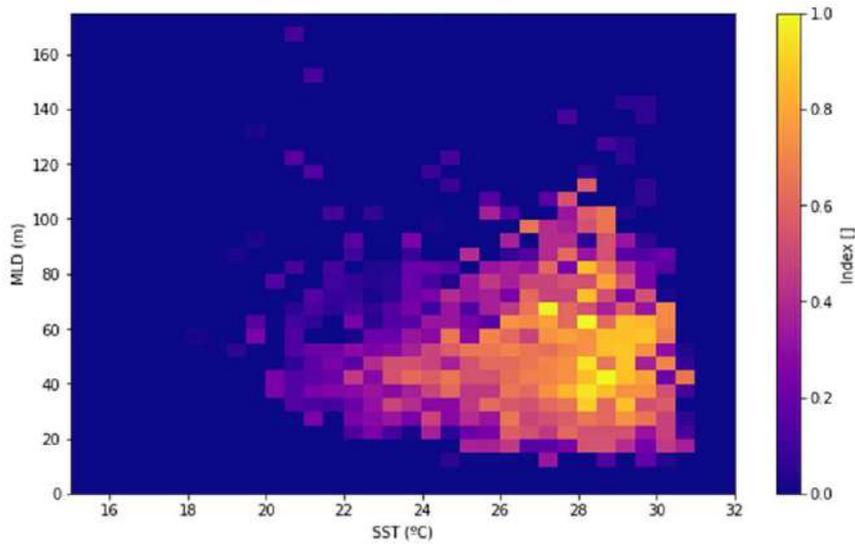


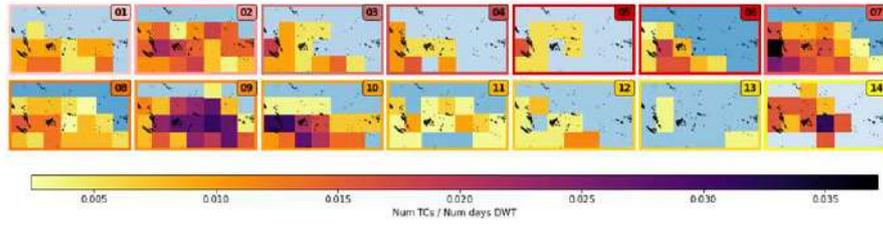
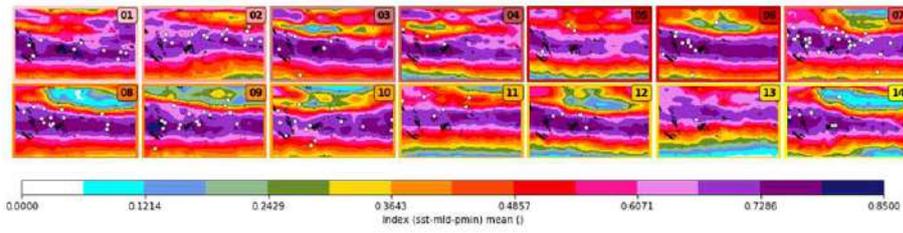
Which large scale oceanic variables can explain TC activity?



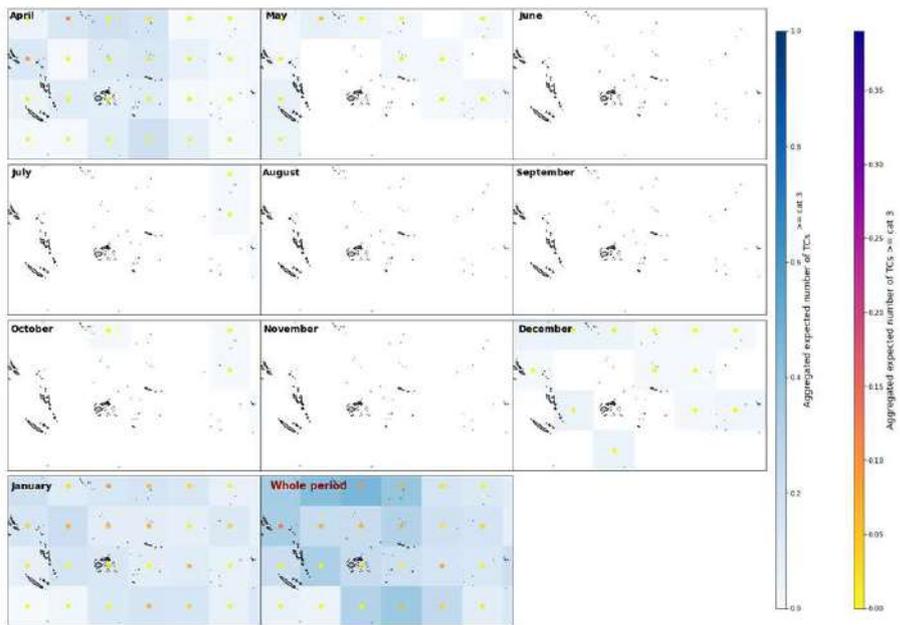


Index predictor based on SST, MLD and Pmin





Forecast from day 1/04/2021

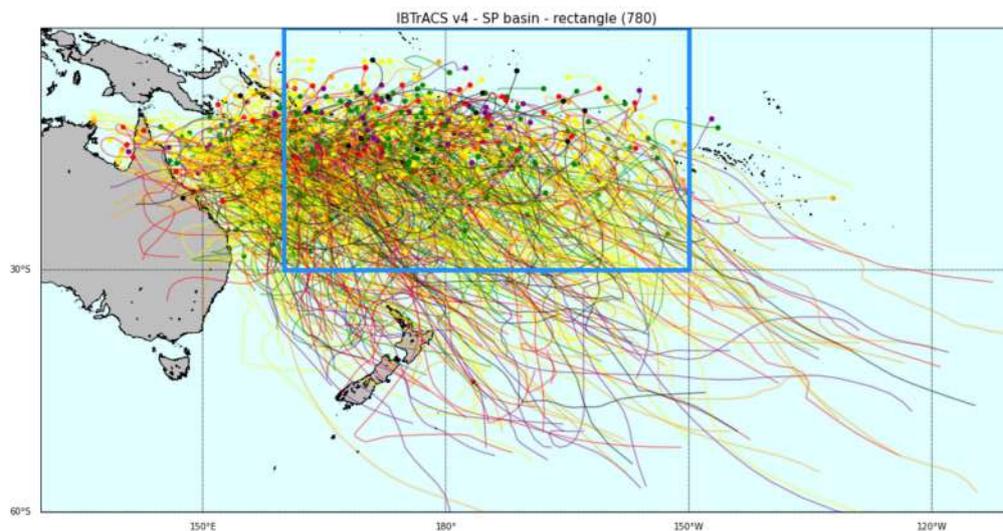


1. Databases

- **Sea Surface Temperature (SST):** The NOAA 1/4° daily Optimum Interpolation Sea Surface Temperature or daily OISST is an analysis constructed by combining observations from different platforms (satellites, ships, buoys and Argo floats) on a regular global grid. More information at the [OISST NOAA webpage](#).
- **CFS (Climate Forecast System):** It is a model produced by several dozen scientists under guidance from the National Centers for Environmental Prediction (NCEP), which offers hourly data with a horizontal resolution down to 1/2°. **Mixed Layer Depth (MLD), Mean Sea Level Pressure (SLP) and forecast data from both SST and MLD.** More information at the [CFS NOAA webpage](#).
- **Daily mean precipitation:** It is generated from the research-quality 3-hourly TRMM Multi-Satellite Precipitation Analysis TMPA (3B42). Simple summation of valid retrievals in a grid cell is applied for the data day. The result is given in (mm). More information at the [TRMM NASA webpage](#) or [TRMM 3B42 dataset product](#).
- **Madden-Julian Oscillation (MJO):** from the Australian Bureau of Meteorology webpage, including MJO date (day, month and year), RMM1, RMM2, phase (1 to 8) and amplitude. More information at the [MJO Australian Bureau of Meteorology](#).
- **IBTrACS v4:** This global dataset compiles an inventory of tropical cyclones (TCs) reported worldwide with their characteristics ([Knapp et al. 2010](#)). Various parameters among which are the best-track position and intensity are provided at 6-h intervals. Data sources are each World Meteorological Organization (WMO) Regional Specialized Meteorological Centers (RSMCs) and Tropical Cyclone Warning Centres (TCWCs), as well as other national agencies. Data for some basins are available as early as 1850. Data can be accessed and downloaded at [IBTrACS NOAA webpage](#).

Saffir-Simpson classification scale of tropical cyclones. Each category has been assigned a color used for all the figures in the work:

	Mean wind speed (km/h)	Minimum pressure (hPa)	Stormy wave (m)	Damage level
Tropical depression	< 60	990 - 1000	< 1.0	weak
Tropical gust	60 - 120	980 - 990	< 1.0	small
Category 1 cyclone	120 - 150	970 - 980	1.0 - 1.5	small
Category 2 cyclone	150 - 180	965 - 970	1.5 - 2.5	moderate
Category 3 cyclone	180 - 210	945 - 965	2.5 - 4	intense
Category 4 cyclone	210 - 240	920 - 945	4.0 - 6.0	extreme
Category 5 cyclone	> 240	< 920	> 6.0	catastrophic




```
import warnings
warnings.filterwarnings('ignore')
from large_scale_predictors_functions import *

#put your path
path = r'D:\forecast_book\large_scale_predictor_data\'
```

```
-----
ModuleNotFoundError                                Traceback (most recent call last)
<ipython-input-1-1e53f605175e> in <module>
      1 import warnings
      2 warnings.filterwarnings('ignore')
----> 3 from large_scale_predictors_functions import *
      4
      5 #put your path

ModuleNotFoundError: No module named 'large_scale_predictors_functions'
```

```
#geopandas as background map
coasts_low=geopandas.read_file(path+'GSHHS_shp\h\GSHHS_h_L1.shp')[['geometry']]
gdf_l=shift_map(coasts_low,180)
```

2. Large Scale Predictors

[Madden - Jullian Oscillation \(MJO\)](#)

[Annual Weather type \(AWT\)](#)

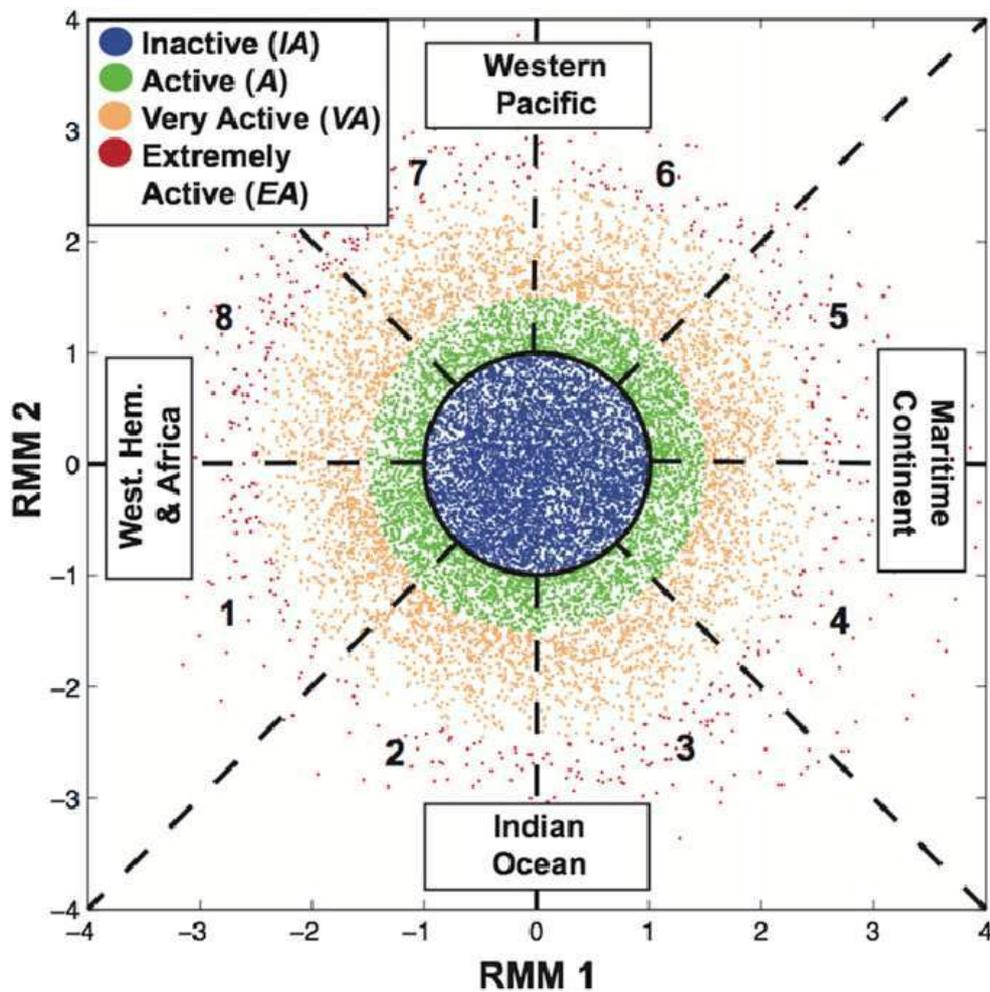
[MJO and AWT relationship with TC genesis](#)

2.1. Madden - Julian Oscillation (MJO)

The MJO is an eastward moving disturbance of clouds, rainfall, winds, and pressure that crosses the planet in the tropics and returns to its initial starting point in cycles of approximately 30 or 60 days. It is the dominant mode of atmospheric intraseasonal variability in the tropics ([Hendon & Salby, 1994](#)):

The MJO consists of two phases: **the enhanced rainfall and the the suppressed rainfall**. They produce opposite changes in clouds and rainfall and this entire dipole propagates eastward. Strongest MJO activity often divides the planet into halves: one half within the enhanced convective phase and the other half in the suppressed convective phase.

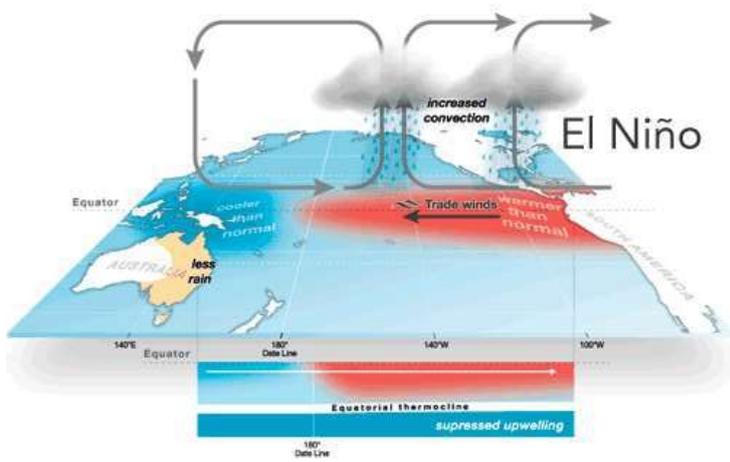
The MJO phases can be observed and represented through the **RMM index**, which is a combined cloudiness- and circulation-based index that has been frequently used for real-time prediction and definition of the MJO ([Wheeler & Hendon, 2004](#)):



2.2. Annual Weather type (AWT)

ENSO is one of the most important climate phenomena on Earth due to its ability to change the global atmospheric circulation; since it can lead to changes in sea-level pressures, sea-surface temperatures, precipitation and winds across the globe. ENSO describes the natural interannual variations in the ocean and atmosphere in the tropical Pacific. This interaction between the atmosphere and ocean is the source of a periodic variation between below-normal and above-normal sea surface temperatures and dry and wet conditions along the years. The tropical ocean affects the atmosphere above it and the atmosphere influences the ocean below it.

Typical behavior of the couple system of ocean and atmosphere during El Niño in the equatorial Pacific ([Australian Bureau of Meteorology](#)):



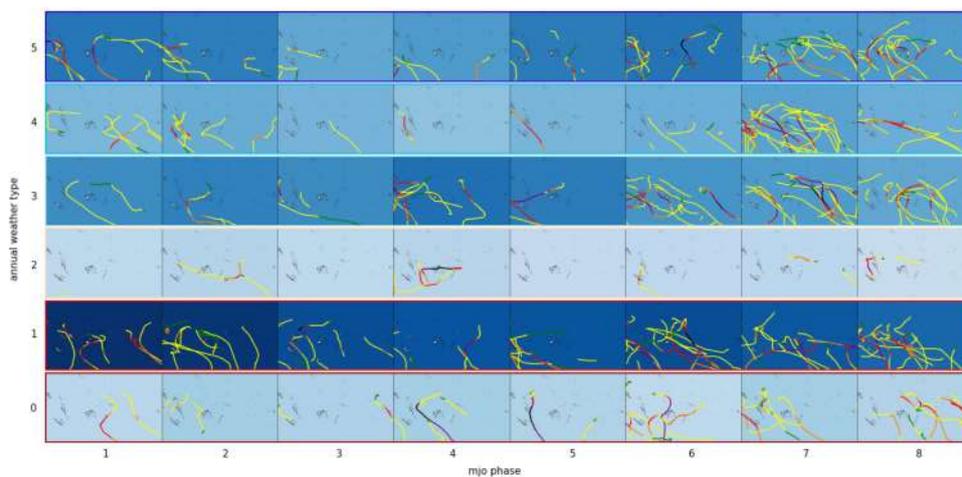
Patterns (Anderson, et al., 2019)



2.3. MJO and AWT relationship with TC genesis

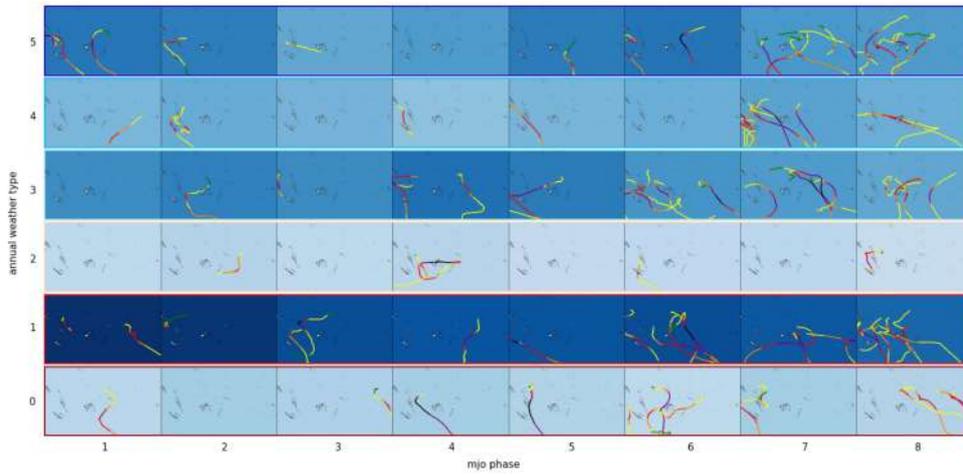
TCs tracks transferred to MJO+AWT combinations according to the genesis point, with the combination probability as the background color

```
fig_tcs = plot_tcs_mjo_awt(path, gdf_1)
```

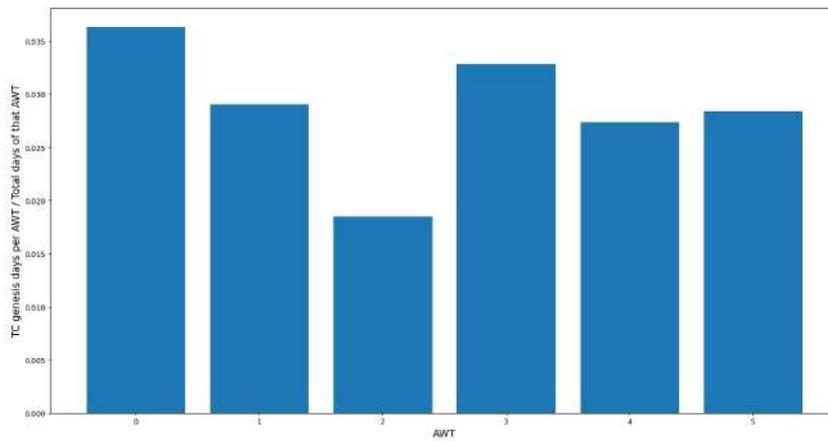


TCs tracks reaching category 3 or greater transferred to MJO+AWT combinations according to the genesis point, with the combination probability as the background color

```
tcs3_mjo_awt = plot_tcs3_mjo_awt(path,gdf_1)
```



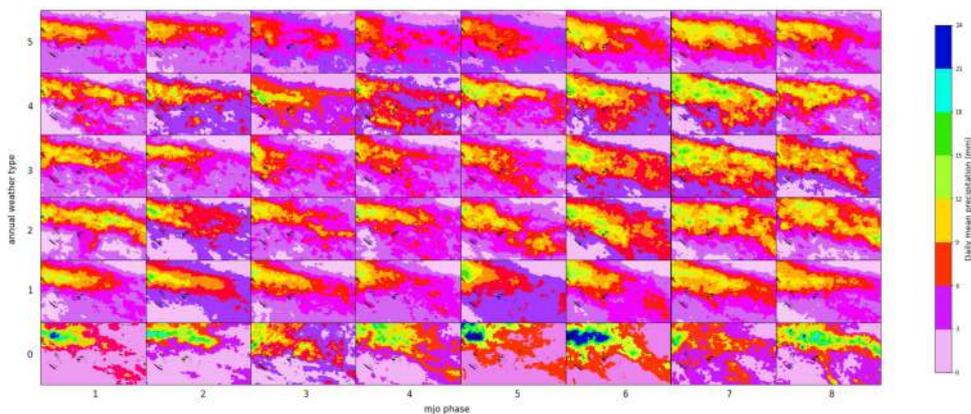
Density histogram of TC genesis according to AWT



Daily mean precipitation (TRMM) transferred to MJO+AWT combinations

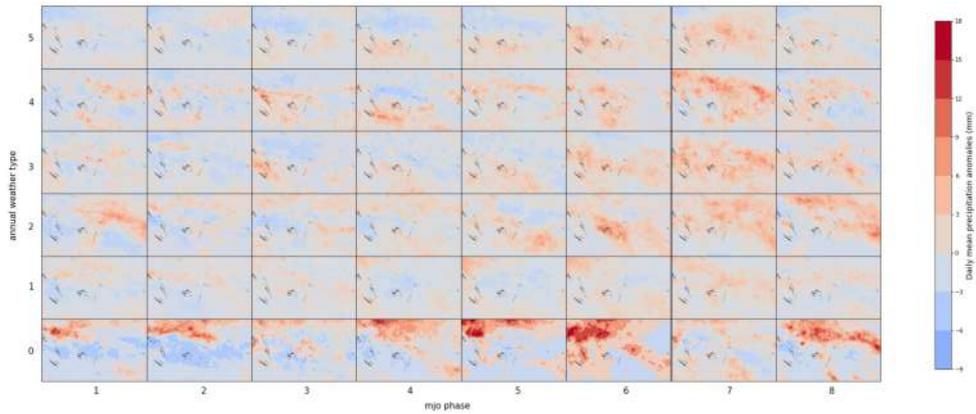
```
path = 'D:\\forecast_book\\large_scale_predictor_data\\'
```

```
fig_trmm = plot_trmm(path,gdf_1)
```



Daily mean precipitation (TRMM) anomalies transferred to MJO+AWT combinations

```
fig_anom_trmm = plot_anom_trmm(path,gdf_1)
```



Conclusions:

- MJO phases 6,7 and 8 and AWT 1 and 3 show the highest TC genesis activity.
- The combination MJO phase 7 + AWT 4 is the most intense one for all TC categories and when filtering from category 2.
- **El Niño, which is highly unlikely has fewer TC genesis but it is the most active in TCs genesis with respect of its total days and a greater proportion of TCs reaching at least category 2.**
- AWT 2 is the least probable and is the one with least TCs genesis activity.
- AWT 1 have same TCs genesis along all MJO phases.
- TCs genesis occurs generally in areas of intense precipitation, above the mean (positive anomalies) and the higher categories of TC are linked with the most extended and amongst the most intense precipitation clouds.

```

# basic
import sys
import os

# common
import numpy as np
import pandas as pd
import xarray as xr
import matplotlib.pyplot as plt
from datetime import datetime, timedelta
import pickle

#Lib
from lib.plots_base import basemap_ibtracs, basemap_var, basemap_scatter, axplot_basemap,
basemap_scatter_both, plot_target_area, plot_predictor_grid
from lib.plots_tcs import get_storm_color, get_category, Plot_DWTs_tracks
from lib.plots_aux import data_to_discret, colors_dwt
from lib.extract_tcs import Extract_Rectangle, dwt_tcs_count
from lib.predictor_definition_building import *

import warnings
warnings.filterwarnings('ignore')
from IPython.display import Image

```

```

-----
ModuleNotFoundError                                Traceback (most recent call last)
<ipython-input-1-9742772fecc0> in <module>
     12
     13 #Lib
--> 14 from lib.plots_base import basemap_ibtracs, basemap_var, basemap_scatter,
axplot_basemap, basemap_scatter_both, plot_target_area, plot_predictor_grid
     15 from lib.plots_tcs import get_storm_color, get_category, Plot_DWTs_tracks
     16 from lib.plots_aux import data_to_discret, colors_dwt

ModuleNotFoundError: No module named 'lib.plots_base'

```

3. Index Predictor Definition and Building

[Spatial and temporal domain](#)

[Predictor grid and data processing](#)

[Index definition and computation](#)

3.1. Spatial and temporal domain

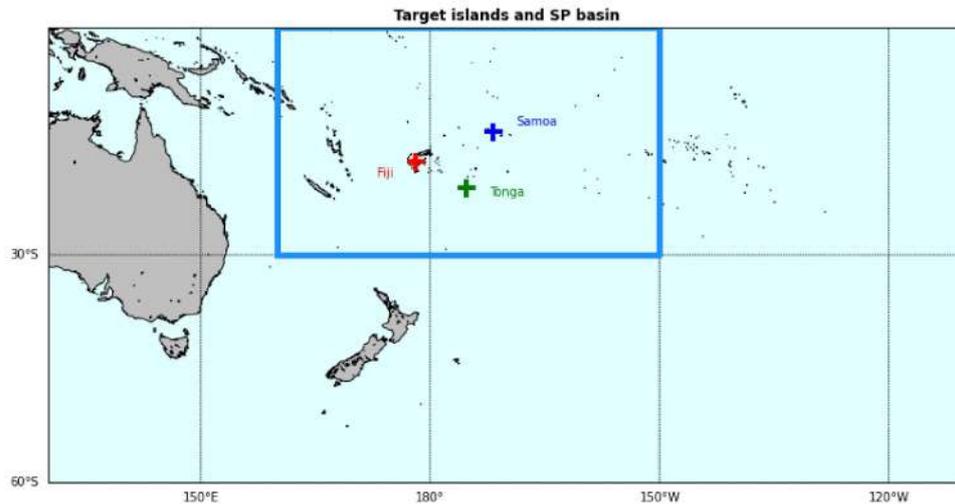
The predictor area spans around the islands of Tonga, Samoa and Fiji, from latitude 0° to 30° and from longitude 160° to 210°; far enough to be able to identify regional as well as local patterns. The calibration period (time domain) is defined from 1982 to 2019.

```
# ibtracs v4 dictionary
d_vns = {
    'longitude': 'lon',
    'latitude': 'lat',
    'time': 'time',
    'pressure': 'wmo_pres',}

lo_SP, la_SP = [130,250], [-60,0]

# predictor area
lo_area = [160, 210]
la_area = [-30, 0]
```

```
fig_target_area = plot_target_area(rectangle=[lo_area[0], lo_area[1], la_area[0],
la_area[1]])
```



The variables required for the methodology are downloaded from the databases:

- **Predictand:** Tropical cyclones tracks from IBTrACs, for the minimum pressure point.
- **Predictor:** NOAA 1/4° daily Optimum Interpolation Sea Surface Temperature (SST) and Mixed Layer Depth (MLD) from the NCEP Climate Forecast System Reanalysis (CFSR).

```
path_st = r'/home/administrador/Documentos/'
xds_ibtracs, xds_SP = storms_sp(path_st)
```

```
All basins storms: 13481
SP basin storms: 1130
```

```
# ibtracs v4 dictionary
d_vns = {
    'longitude': 'lon',
    'latitude': 'lat',
    'time': 'time',
    'pressure': 'wmo_pres',}

lo_SP, la_SP = [130,250], [-60,0]

# predictor area
lo_area = [160, 210]
la_area = [-30, 0]
```

```
# extract rectangle, 772 a 780
TCs_rect_hist_tracks = Extract_Rectangle(xds_SP, lo_area[0], lo_area[1], la_area[0],
la_area[1], d_vns)
```

```
df0 = df_pressures(xds_ibtracs)
df0[6000:6010]
```

	st	time	lon	lat	pres
6000	9833	1987-02-12 12:00:00.000040448	139.125000	-15.225000	970.0
6001	9833	1987-02-12 18:00:00.000040448	139.817963	-15.790068	970.0
6002	9833	1987-02-13 00:00:00.000040448	140.160004	-16.719999	970.0
6003	9833	1987-02-13 06:00:00.000040448	140.032623	-17.892567	980.0
6004	9833	1987-02-13 12:00:00.000040448	139.699997	-19.200003	996.0
6005	9833	1987-02-13 18:00:00.000040448	139.480362	-20.108568	998.0
6006	9833	1987-02-14 00:00:00.000040448	139.399994	-21.100000	1001.0
6007	9835	1987-02-07 12:00:00.000040448	161.800003	-13.050001	990.0
6008	9835	1987-02-07 18:00:00.000040448	163.449997	-13.150000	995.0
6009	9835	1987-02-08 00:00:00.000040448	166.242844	-14.371429	997.0

```
#path to your daily mean SST and MLD data
path_sst = r'/media/administrador/SAMSUNG/seasonal_forecast/data/SST/'
path_mld = r'/media/administrador/SAMSUNG/seasonal_forecast/data/CFS/ocnml/'
path_p = r'/home/administrador/Documentos/seasonal/seasonal_forecast/new/'
```

For the calibration period the points with pressure, SST and MLD data in the target area are kept.

```
df = df_p_sst_mld(df0,path_sst,path_mld)
df_cali = df.drop(df.index[5184:]) #years of the calibration period
```

```
Start time: 2021-04-13 13:37:32.513946
End time: 2021-04-13 13:41:41.488278
```

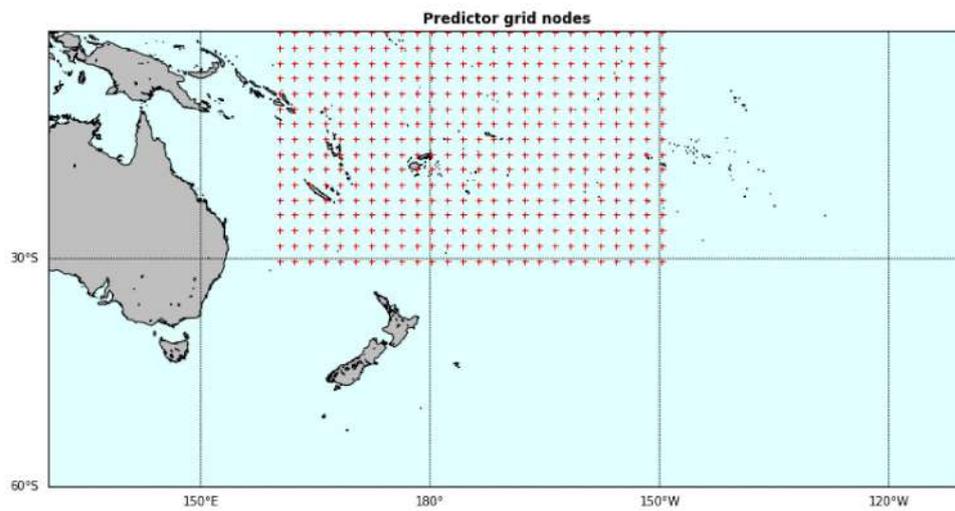
```
# Load data
path_p= r'/home/administrador/Documentos/seasonal/seasonal_forecast/new/'
df = pd.read_pickle(path_p+'df_coordinates_pmin_sst_mld_2019.pkl')
df.tail()
```

	st	time	lon	lat	pres	sst
14274	13348	2019-12-31 06:00:00.000039936	185.616379	-20.368277	999.0	25.750000 40.666
14275	13348	2019-12-31 09:00:00.000039936	186.182968	-20.301870	999.0	25.199999 43.166
14276	13348	2019-12-31 12:00:00.000039936	186.849609	-20.312431	1000.0	26.049999 42.666
14277	13348	2019-12-31 15:00:00.000039936	187.731110	-20.379351	1000.0	27.029999 35.708
14278	13348	2019-12-31 18:00:00.000039936	188.699997	-20.499998	1001.0	26.430000 42.999

3.2. Predictor grid and data processing

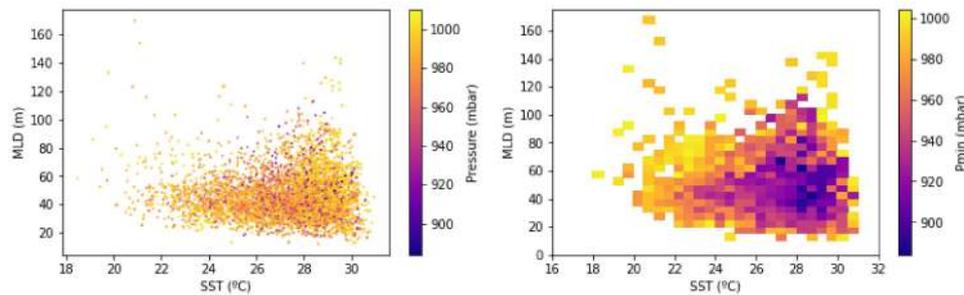
The historical datasets are interpolated into the a 1/2° grid resolution, defining this way the grid for the predictor in the target area.

```
fig_predictor_grid = plot_predictor_grid()
```



MLD, SST and Pmin data plotted after preprocessing (left) and the same data after discretization in intervals of 0.5 m and 0.5°C respectively (right):

```
plot_sst_mlp_pmin_cali(df)
```



3.3. Index definition and computation

The historic datasets are combined into the tailor-made index predictor. It is built from the combination of SST-MLD-Pmin of the coordinate dataset previously generated. For simplicity the index will range from 0 to 1, so the pressure range is rescaled in this range.

```
# discretization: 0.5°C (SST), 5m (MLD)
xx,yy,zz = data_to_discret(df['sst'].values, df['mld'].values, 0.5, 5, df['pres'].values,
15, 32, 0, 175, option='min')

# index function
index = zz
fmin, fmax = np.nanmin(zz), np.nanmax(zz)
index_zz = (fmax - index) / (fmax-fmin)

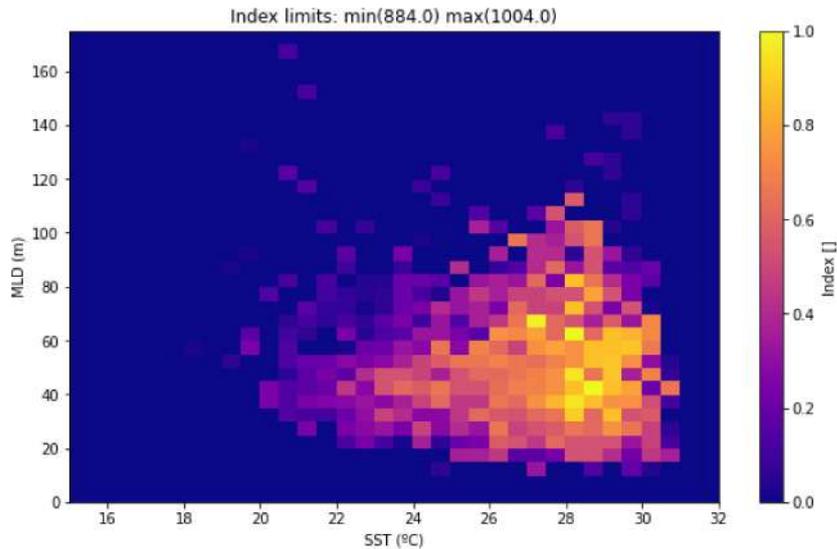
# remove nan
index_zz[np.isnan(index_zz)] = 0
```

The index predictor function:

$$\text{index}_i = (P_{\max} - P_i) / (P_{\max} - P_{\min})$$

The index plotted in the MLD-SST space:

```
fig_index = plot_index(xx,yy,zz,index_zz)
```



Final dataset including all the variable and the index predictor values in the predictor grid.

```
xs = ds_index_sst_mld_calibration(path_sst,path_mld,df)
```

```
Start time: 2021-04-14 08:31:02.952802
End time: 2021-04-14 08:31:14.467662
Merge time: 2021-04-14 08:38:58.435711

<xarray.Dataset>
Dimensions: (lat: 16, lon: 26, time: 13879)
Coordinates:
  * time      (time) datetime64[ns] 1982-01-01 1982-01-02 ... 2019-12-31
  * lat       (lat) float64 -0.25 -2.25 -4.25 -6.25 ... -26.25 -28.25 -30.25
  * lon       (lon) float64 160.2 162.2 164.2 166.2 ... 204.2 206.2 208.2 210.2
Data variables:
  index      (time, lat, lon) float64 0.3667 0.05833 0.0 ... 0.1167 0.2417 0.0
  sst        (time, lat, lon) float32 30.77 30.95 30.03 ... 23.1 22.85 22.9
  dbss       (time, lat, lon) float32 19.38 28.42 76.08 ... 24.33 23.88 19.75
  mask       (lat, lon) float64 0.0 0.0 0.0 0.0 0.0 0.0 ... 0.0 0.0 0.0 0.0 0.0
```

```
path_slp = r'/media/administrador/SAMSUNG/seasonal_forecast/data/CFS/'
path_pp = r'/home/administrador/Documentos/pratel/'
path_trmm = r'/home/administrador/Documentos/TRMM_daily/'
```

```
xs = ds_index_sst_mld_slp_pp_calibration(path_pp,path_slp,xs)
```

```
Start time: 2021-04-14 09:21:34.904732
End time: 2021-04-14 09:21:37.685073
Merge time: 2021-04-14 09:25:09.111452
```

```
xs_trmm = ds_trmm(path_trmm)
```

```
Start time: 2021-04-19 15:21:32.682371
End time: 2021-04-19 15:21:32.855106
Merge time: 2021-04-19 15:21:46.715566
```

```

# basic
import sys
import os

# common
import numpy as np
import pandas as pd
import xarray as xr
import matplotlib.pyplot as plt
from datetime import datetime, timedelta
from sklearn.cluster import KMeans
import pickle
from mpl_toolkits.basemap import Basemap

#Lib
from lib.calibration import *
from lib.predictor_definition_building import *
from lib.PCA_predictor_sst_mld import PCA_EstelaPred_sea_mask, Plot_EOFs_EstelaPred,
standardise_predictor_prep
from lib.mda import Normalize, DeNormalize
from lib.plots_kma import plot_3D_kmeans, plot_scatter_kmeans, plot_grid_kmeans
from lib.plots_base import basemap_ibtracks, basemap_var, basemap_scatter, axplot_basemap,
basemap_scatter_both
from lib.plots_dwts import colorp, custom_colorp, Plot_DWTs_Mean_Anom, Plot_DWTs_totalmean, \
Plot_Probs_WT_WT, Plot_Probs_WT_WT_anomaly, Plot_Probs_WT_WT_WT, Plot_DWTs_Probs, \
Report_Sim_oneyear, Report_Sim, Plot_DWTs_counts, Chrono_dwts_tcs, Chrono_probs_tcs,
Plot_dwts_colormap
from lib.plots_tcs import get_storm_color, get_category, Plot_DWTs_tracks
from lib.plots_aux import data_to_discret, colors_dwt
from lib.extract_tcs import Extract_Rectangle, dwt_tcs_count, dwt_tcs_count_tracks

import warnings
warnings.filterwarnings('ignore')
from IPython.display import Image

```

```

-----
ModuleNotFoundError                                Traceback (most recent call last)
<ipython-input-1-fba630d070b3> in <module>
    14
    15 #Lib
--> 16 from lib.calibration import *
    17 from lib.predictor_definition_building import *
    18 from lib.PCA_predictor_sst_mld import PCA_EstelaPred_sea_mask, Plot_EOFs_EstelaPred,
standardise_predictor_prep

ModuleNotFoundError: No module named 'lib.calibration'

```

4. Statistical Downscaling Method

[Daily Weather Types \(DWTs\) classification](#)

[Principal Component Analysis \(PCA\)](#)

[K-means clustering](#)

[DWTs plotting](#)

[Cluster homogeneity](#)

[DWTs plotting with predictand variables](#)

[DWTs storm frequency and track counting](#)

[Calibration time period predictand plotting](#)

4.1. Daily Weather Types (DWTs) classification

A weather type approach is proposed.

The index predictor is first partitioned into a certain number of clusters, DWTS, obtained combining three data mining techniques.

4.1.1. Principal Component Analysis (PCA)

The PCA is employed to reduce the high dimensionality of the original data space and thus simplify the classification process, transforming the predictor fields into spatial and temporal modes.

```
path_p = r'/home/administrador/Documentos/seasonal/seasonal_forecast/new/'
xs = xr.open_dataset(path_p+'xs_index_vars_19822019_2deg_new.nc')
df = pd.read_pickle(path_p+'df_coordinates_pmin_sst_mld_2019.pkl')
xs_trmm = xs_trmm = xr.open_dataset(path_p+'xs_trmm_1982_2019_2deg_new.nc')
```

```
# predictor area
lo_area = [160, 210]
la_area = [-30, 0]
```

```
# PCA parameters
ipca, xds_PCA = PCA_EstelaPred_sea_mask(xs, ['index']) #ipca son Las componentes
principales, Las 416 que salen
xds_PCA
```

xarray.Dataset

► Dimensions: (lat: 16, lon: 26, n_components: 416, n_features: 416, n_points: 416, time: 13879)

▼ Coordinates:

lon	(lon)	float64	160.2 162.2 164.2 ... 20...		
lat	(lat)	float64	-0.25 -2.25 -4.25 ... -28...		

▼ Data variables:

PCs	(time, n_components)	float64	6.68 -0.6377 ... -0.5949 ...		
EOFs	(n_components, n_features)	float64	-0.01639 -0.008938 ... -...		
variance	(n_components)	float64	131.6 25.35 15.47 ... 0.1...		
pred_mean	(n_features)	float64	0.54 0.5337 0.52 ... 0.15...		
pred_std	(n_features)	float64	0.2995 0.3035 ... 0.189 ...		
pred_time	(time)	datetime64[ns]	1982-01-01 ... 2019-12-...		

pred_data_pos (n_points)

bool True True True ... True Tr...  

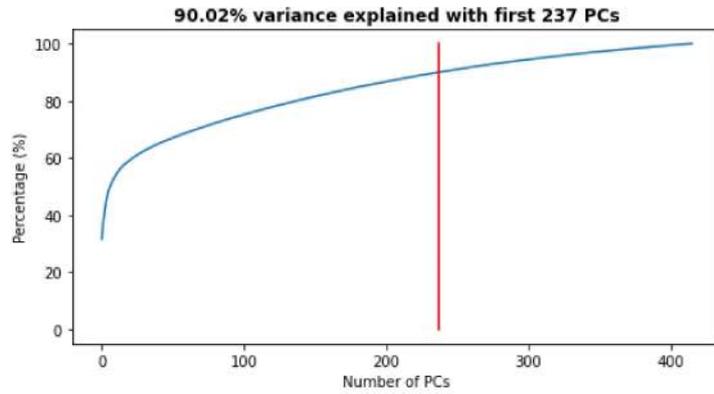
▼ Attributes:

method : gradient + estela + sea mask

pred_name : ['index']

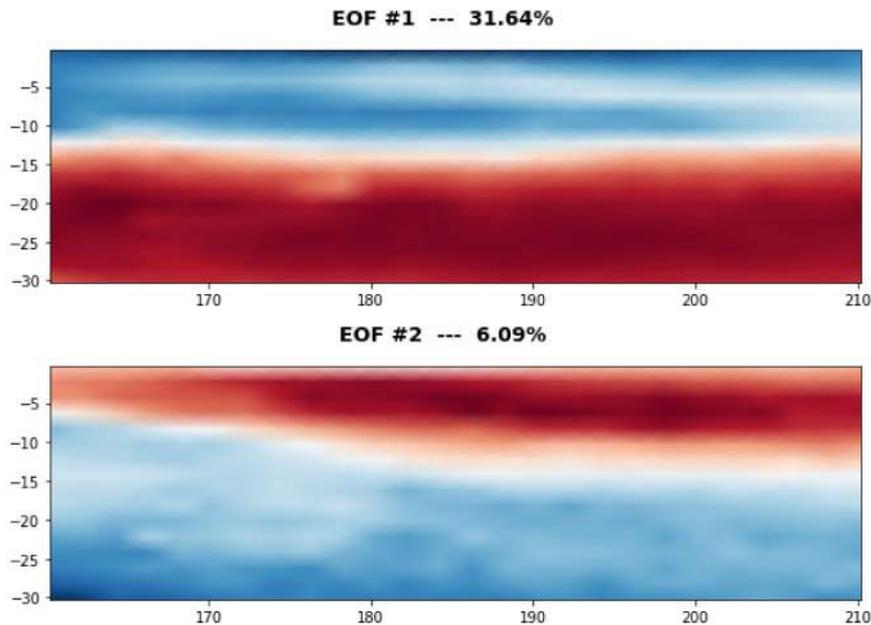
PCA projects the original data on a new space searching for the maximum variance of the sample data. The first 237 PCs are captured, which explain the 90 % of the variability as shown:

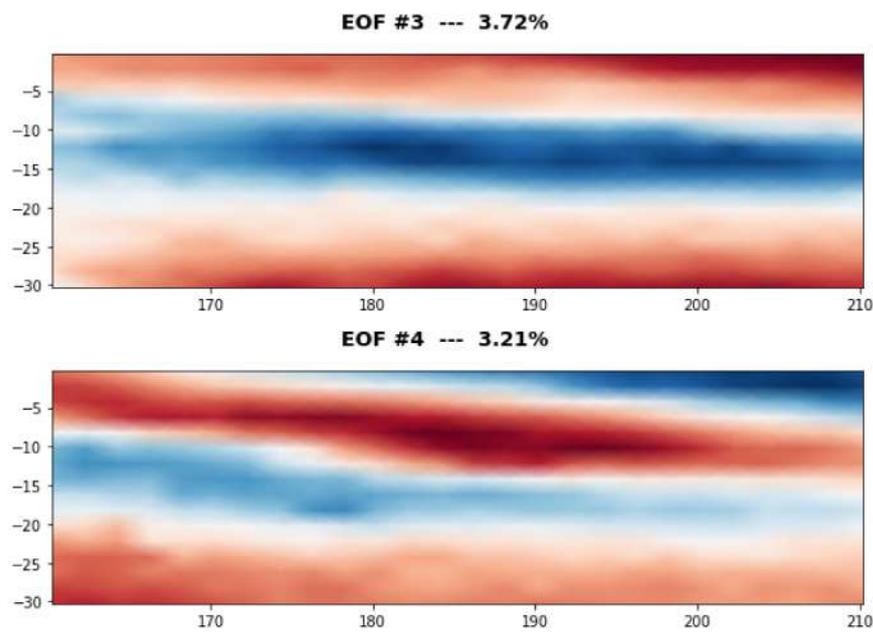
```
variance_PCA = plot_graph_variance_PCA(xds_PCA)
```



The eigenvectors (the empirical orthogonal functions, EOFs) of the data covariance matrix define the vectors of the new space. They represent the spatial variability.

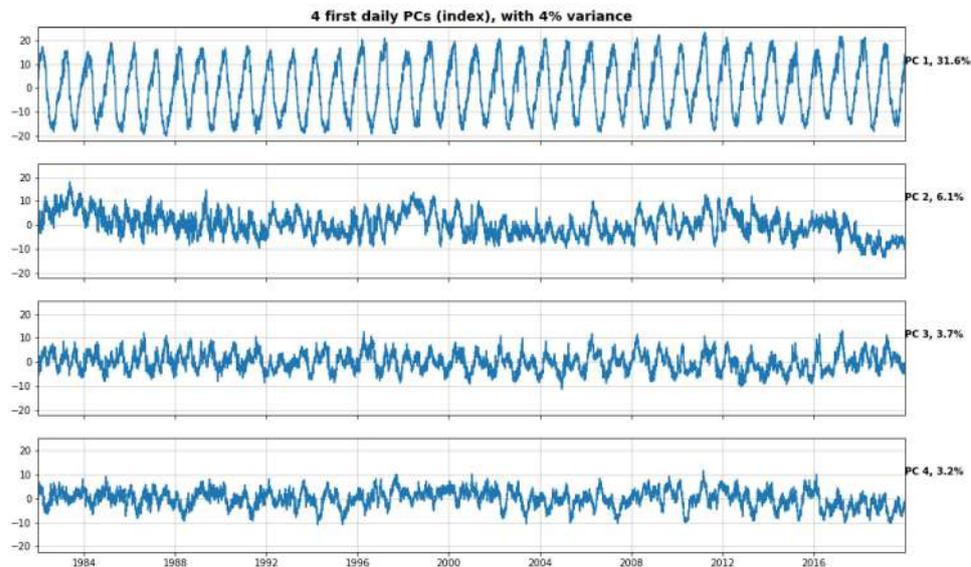
```
fig_eofs = Plot_EOFs_EstelaPred(xds_PCA, 4);
```





The PCs represent the temporal variability.

```
fig_PCA = plot_PCs(xds_PCA,4)
```



4.1.2. K-means clustering

Daily synoptic patterns of the index predictor are obtained using the K-means clustering algorithm. It divides the data space into 49 clusters, a number that which must be a compromise between an easy handle characterization of the synoptic patterns and the best reproduction of the variability in the data space. Previous works with similar analysis confirmed that the selection of this number is adequate ([Rueda et al. 2017](#)).

Each cluster is defined by a prototype and formed by the data for which the prototype is the nearest.

Finally the best match unit (bmus) of daily clusters are reordered into a lattice following a geometric criteria, so that similar clusters are placed next to each other for a more intuitive visualization.

```

# PCA data
variance = xds_PCA.variance.values[:]
EOFs = xds_PCA.EOFs.values[:]
PCs = xds_PCA.PCs.values[:]

var_anom_std = xds_PCA.pred_std.values[:]
var_anom_mean = xds_PCA.pred_mean.values[:]
time = xds_PCA.time.values[:]

variance = xds_PCA.variance.values
percent = variance / np.sum(variance)*100
percent_ac = np.cumsum(percent)
n_comp_95 = np.where(percent_ac >= 95)[0][0]
n_comp_90 = np.where(percent_ac >= 90)[0][0]

# plot
n_comp = n_comp_90

nterm = n_comp_90 + 1 #n_comp_90 es el número de PC que explican el 90% de la varianza, que
en este caso son 237
PCsub = PCs[:, :nterm]
EOFsub = EOFs[:nterm, :]

# normalization
data = PCsub
data_std = np.std(data, axis=0)
data_mean = np.mean(data, axis=0)

```

```
xds_kma_ord,xds_kma = func_kma_order(path_p,xds_PCA,xs)
```

```

Kma order obtained: [11 27 28 45 16 23 47 19 7 14 0 31 46 30 34 35 15 2 8 33 37 41 5
10
48 4 40 21 3 20 42 29 38 26 32 43 13 17 18 39 25 22 44 9 1 24 6 12
36]

```

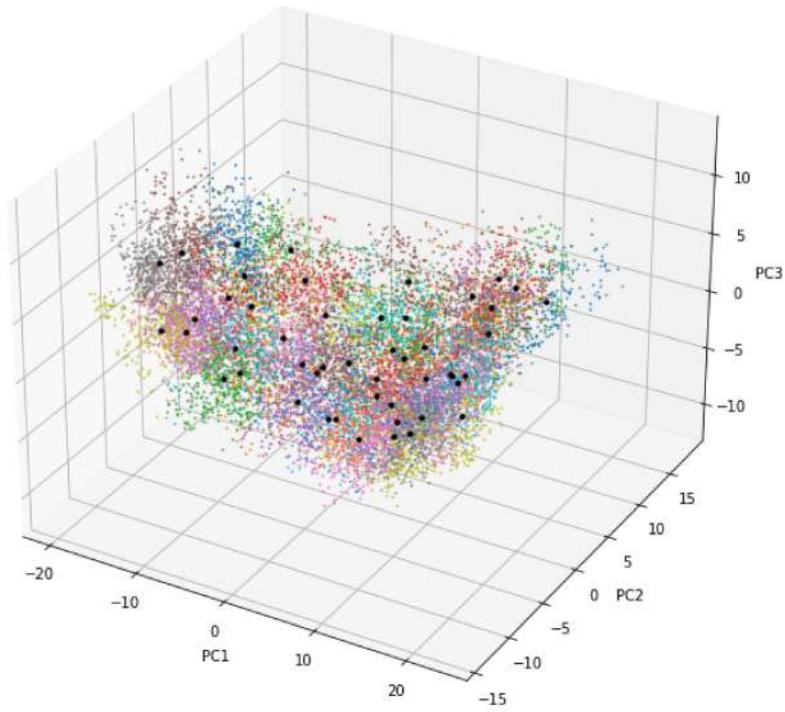
```
xds_kma_sel = trmm_kma(xds_kma,xs_trmm)
```

The resulting classification can be seen in the PCs space of the predictor index data. The obtained centroids (black dots), span the wide variability of the data.

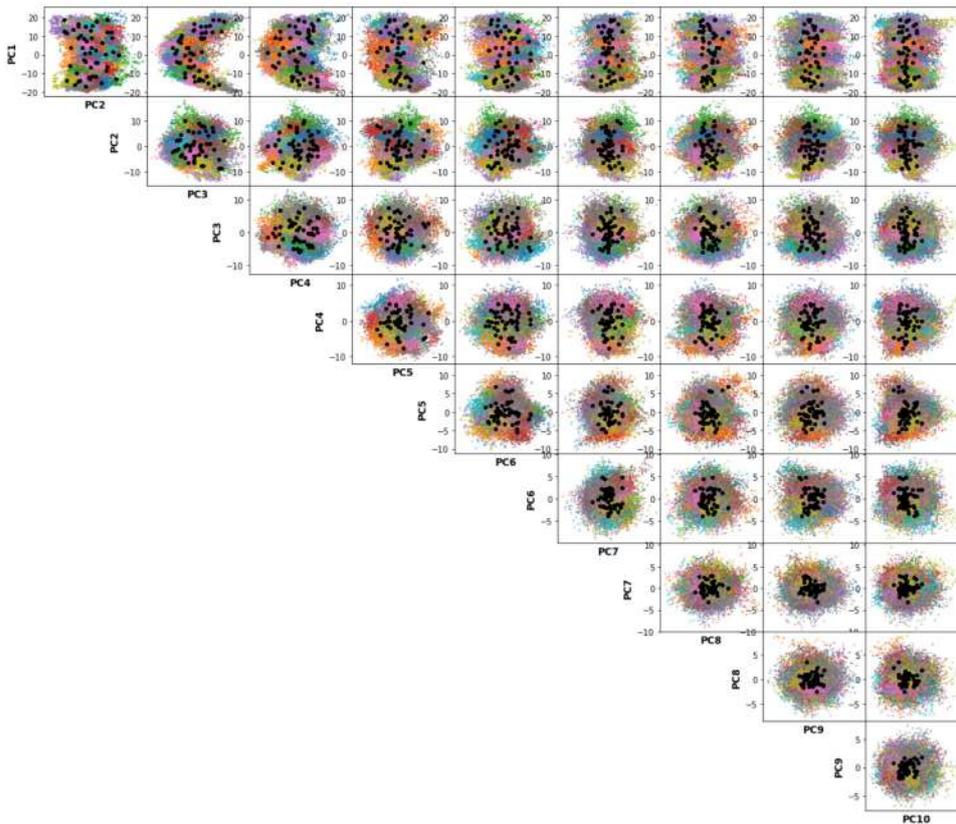
```

# CHECKUP PLOT 3D
%matplotlib inline
fig = plot_3D_kmeans(xds_kma_ord, xds_kma_ord.bmus.values, xds_kma_ord.cenEOFs.values,
size_l=12, size_h=10);

```



```
fig = plot_grid_kmeans(xds_kma_ord, xds_kma_ord.bmus.values, xds_kma_ord.cenEOFs.values, 9,
ibmus=range(49), size_l=20, size_h=18);
```



4.1.3. DWTs plotting

```
#Load
xds_kma = xr.open_dataset(path_p+'kma_model/xds_kma_index_vars_1b.nc')
```

```
path_st = r'/home/administrador/Documentos/'
xds_ibtracs, xds_SP = storms_sp(path_st)
```

```
All basins storms: 13481
SP basin storms: 1130
```

```
st_bmus,st_lons,st_lats, st_categ = st_bmus(xds_SP,xds_kma)
```

```
# custom colorbar for index
color_ls =
['white','cyan','cornflowerblue','darkseagreen','olivedrab','gold','darkorange','orangered',
'red','deeppink','violet','darkorchid','purple','midnightblue']
custom_cmap = custom_colorp(color_ls)
```

DWTs lattice and corresponding colors:

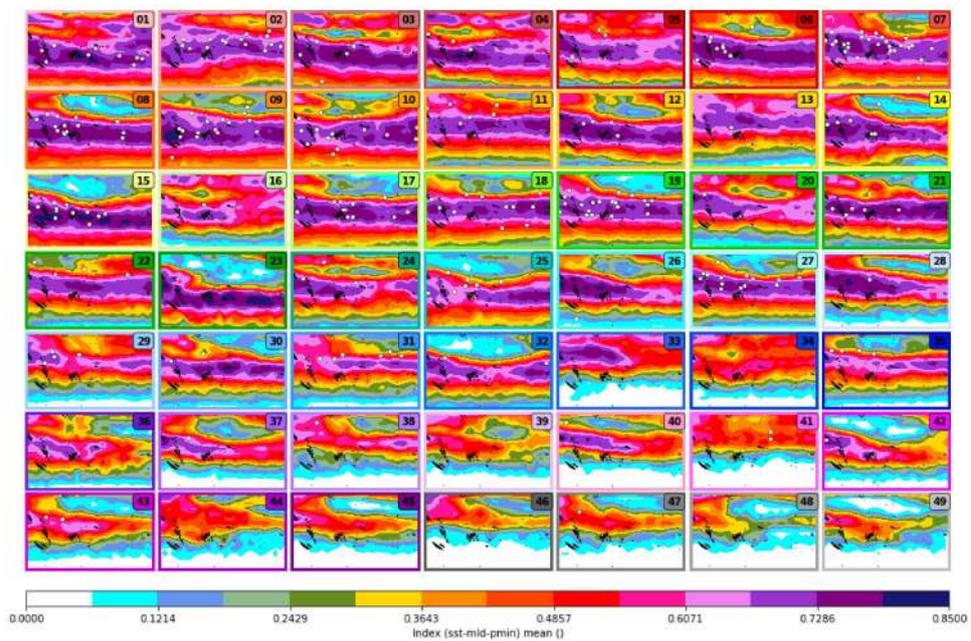
```
fig_dwt_lattice = Plot_dwts_colormap(xds_kma.n_clusters.size)
```



The resulting clustering of the index predictor, each cell is the mean of all the patterns of the corresponding cluster:

```
fig = Plot_DWTs_Mean_Anom(xds_kma, xs, ['index'], minis=[0], maxis=[.85],levels=
[len(color_ls)], kind='mean', cmap = [custom_cmap],genesis='on', st_bmus=st_bmus,
st_lons=st_lons, st_lats=st_lats, markercol='white',
markeredge='k');
```

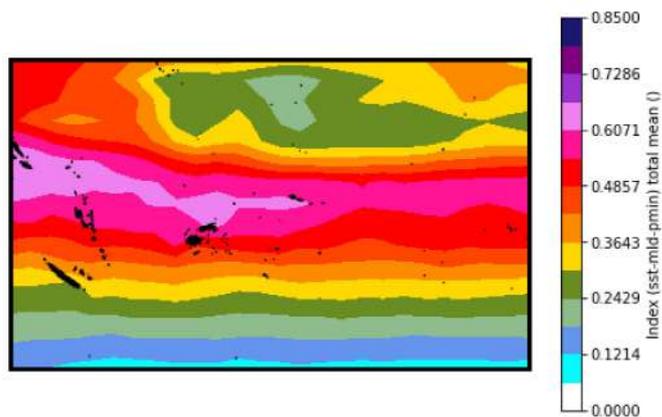
```
index: min(0.0) max(1.0)
```



DWTs (index) total mean:

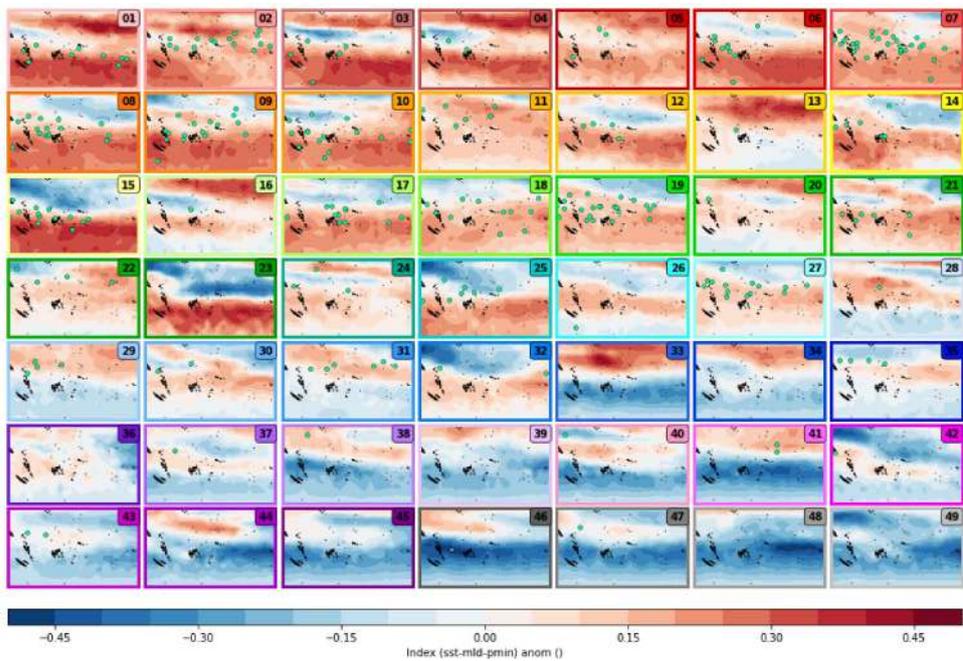
```
fig = Plot_DWTs_totalmean(xs, ['index'], minis=[0], maxis=[.85], levels=[len(color_ls)], cmap=[custom_cmap]);
```

index: min(0.1) max(0.7)



DWTs (index) Anomalies:

```
fig = Plot_DWTs_Mean_Anom(xds_kma, xs, ['index'], minis=[-.5], maxis=[.5], levels=[20],
kind='anom', cmap=['RdBu_r'],
genesis='on', st_bmus=st_bmus, st_lons=st_lons, st_lats=st_lats,
markercol='mediumspringgreen', markeredge='k');
```



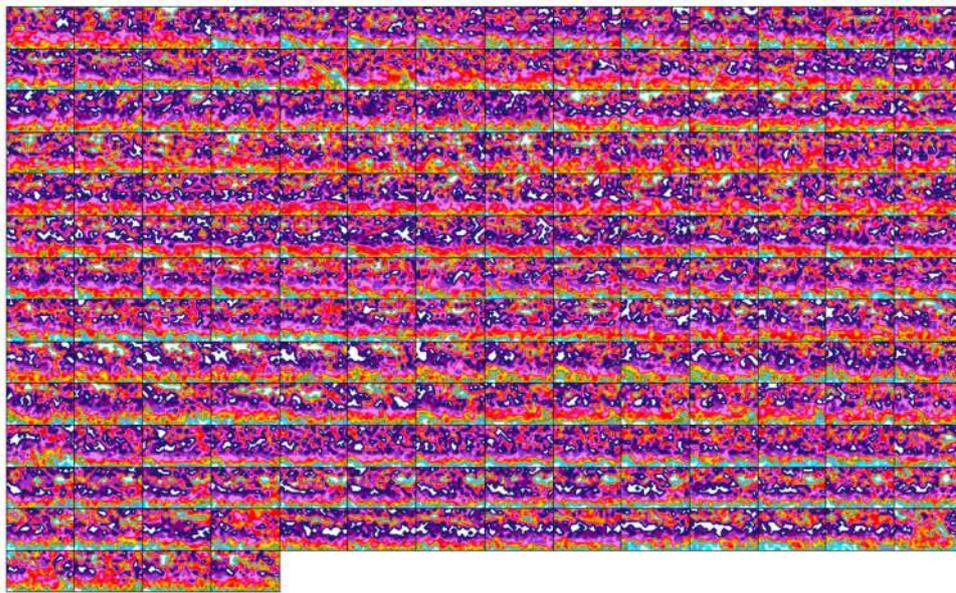
4.1.4. Cluster homegeinity

The DWTs are different to each other, showing the high variability of the data space. The clusters are also very homogenous inside. This confirms 49 as a good choice, which additionally is a manageable number.

```
fig_41 = plot_41(xs,xds_kma)
```



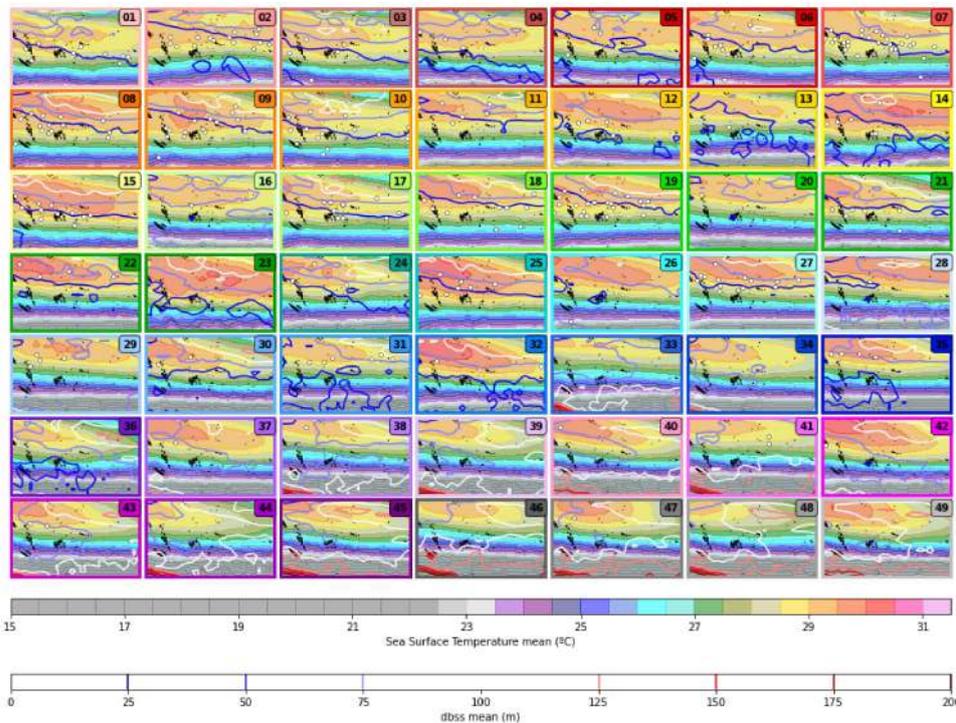
```
fig_5 = plot_5(xs,xds_kma)
```



4.1.5. DWTs plotting with predictand variables

```
fig = Plot_DWTs_Mean_Anom(xds_kma, xs, ['sst', 'dbss'], minis=[22, 0], maxis=[32, 200],
levels=[(32-22)/0.5, 8], kind='mean',
cmap=[colorp(), 'seismic'], genesis='on', st_bmus=st_bmus,
st_lons=st_lons, st_lats=st_lats, markercol='white', markeredge='k');
```

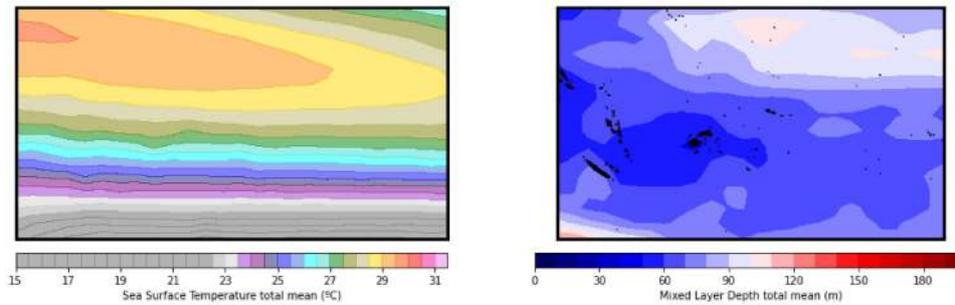
```
sst: min(14.899999618530273) max(33.0)
dbss: min(7.5) max(330.0)
```



DWTs - SST and MLD total mean:

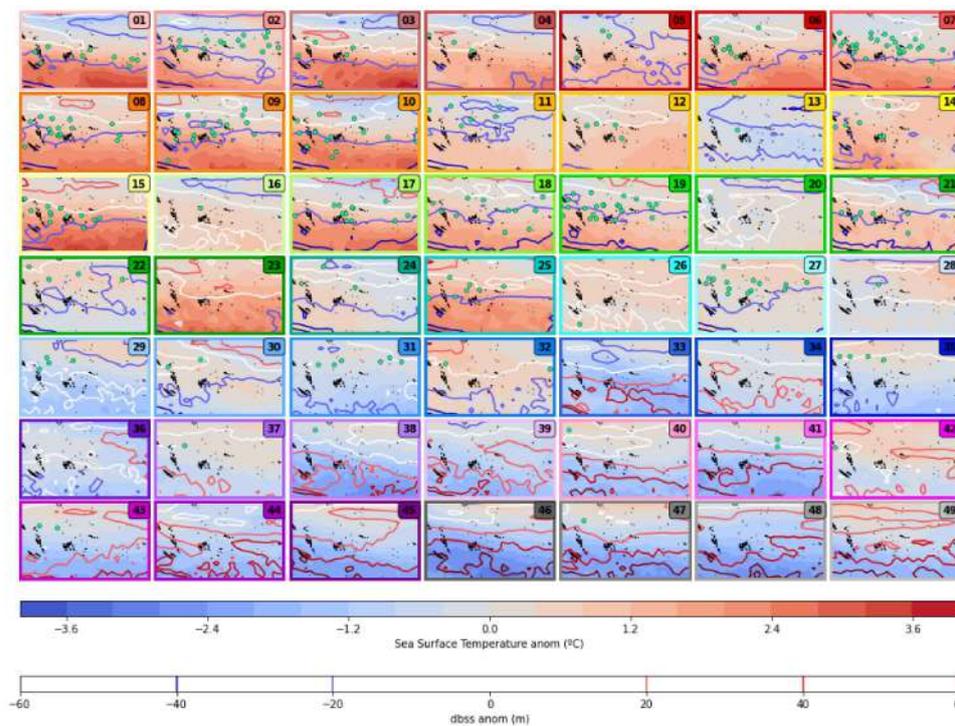
```
fig = Plot_DWTs_totalmean(xs, ['sst', 'dbss'], minis=[22,0], maxis=[32,200], levels=[(32-
22)/0.5,20], cmap=[colorp(), 'seismic']);
```

sst: min(20.299999237060547) max(29.5)
 dbss: min(46.20000076293945) max(120.0)



DWTs - SST and MLD Anomalies:

```
fig = Plot_DWTs_Mean_Anom(xds_kma, xs, ['sst', 'dbss'], minis=[-4, -60], maxis=[4, 60],
  levels=[20, 6], kind='anom', cmap=['coolwarm', 'seismic'], genesis='on',
  st_bmus=st_bmus, st_lons=st_lons, st_lats=st_lats,
  markercol='mediumspringgreen', markeredge='k');
```



Clear patterns can be extracted from these figures related to TCs genesis. Most of it takes place under the following conditions:

- SST interval from 28°C to 30°C (specially 28.5 to 29.5 °C) that correspond to positive or zero SST anomalies.
- MLD values equal or smaller to 75 m that correspond to negative anomalies.

4.2. DWTs seasonality, annual variability and chronology

Several plots are shown to better analyse the distribution of DWTs, their transition, persistence and conditioning to TCs occurrence and to AWT.

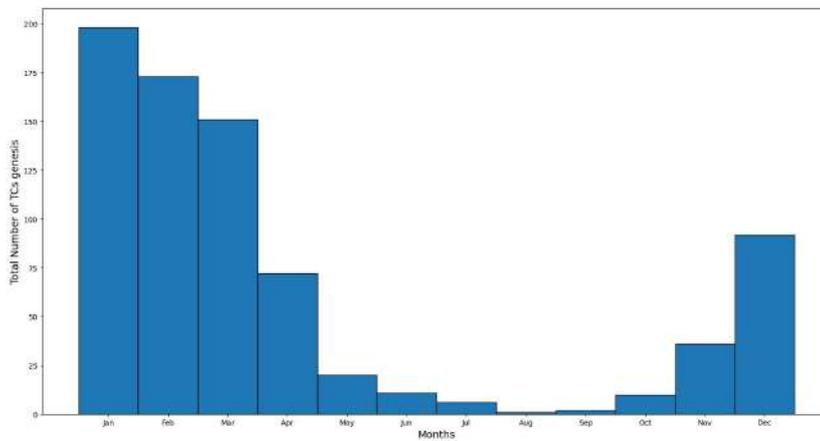
```
path_mjo_aws = r'/home/administrador/Documentos/STORMS_VIEWER/'
```

```
aws,mjo,aws0 = aws_mjo_ds(path_mjo_aws)
```

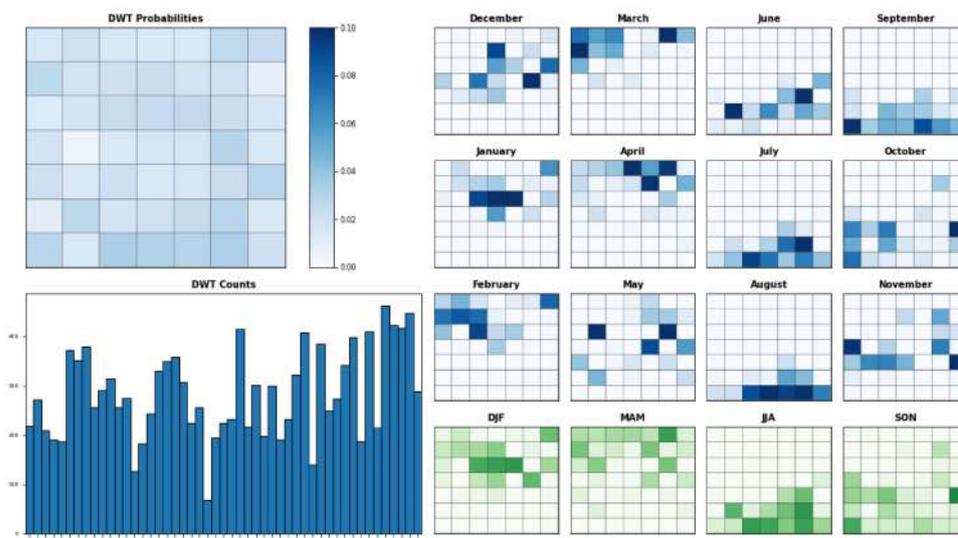
Seasonality:

```
bmus_DWT, bmus_time,aws0_sel_bmus,bmus_AWT,bmus_MJO = bmus_dwt_mjo(mjo,aws,aws0,xds_kma)
```

TCs genesis according to month:



```
fig = Plot_DWTs_Probs(bmus_DWT, bmus_time, 49, height=10, width=18);
```



```

#all categories
xs_dwt_counts = dwt_tcs_count_tracks(xds_kma, df, dx=8, dy=8, lo0=lo_area[0],
lo1=lo_area[1], la0=la_area[0], la1=la_area[1])
xs_dwt_counts.to_netcdf(r'/home/administrador/Documentos/seasonal/seasonal_forecast/new/kma_
model/xds_count_tcs8.nc')

#category 3
xs_dwt_counts_964 = dwt_tcs_count_tracks(xds_kma, df, dx=8, dy=8, categ=965,lo0=lo_area[0],
lo1=lo_area[1], la0=la_area[0], la1=la_area[1])
xs_dwt_counts_964.to_netcdf(r'/home/administrador/Documentos/seasonal/seasonal_forecast/new/
kma_model/xds_count_tcs8_964.nc')

# category 2
xs_dwt_counts_979 = dwt_tcs_count_tracks(xds_kma, df, dx=8, dy=8, categ=979,lo0=lo_area[0],
lo1=lo_area[1], la0=la_area[0], la1=la_area[1])
xs_dwt_counts_979.to_netcdf(r'/home/administrador/Documentos/seasonal/seasonal_forecast/new/
kma_model/xds_count_tcs8_979.nc')

```

```

xs_dwt_counts = xr.open_dataset(path_p+'kma_model/xds_count_tcs8.nc')
xs_dwt_counts_964 = xr.open_dataset(path_p+'kma_model/xds_count_tcs8_964.nc')
xs_dwt_counts_979 = xr.open_dataset(path_p+'kma_model/xds_count_tcs8_979.nc')

```

```

xds_timeline = ds_timeline(df,xs_dwt_counts,xs_dwt_counts_964,xds_kma)

```

```

mask_bmus_YD, mask_tcs_YD = variables_dwt_super_plot(xds_kma, xds_timeline)

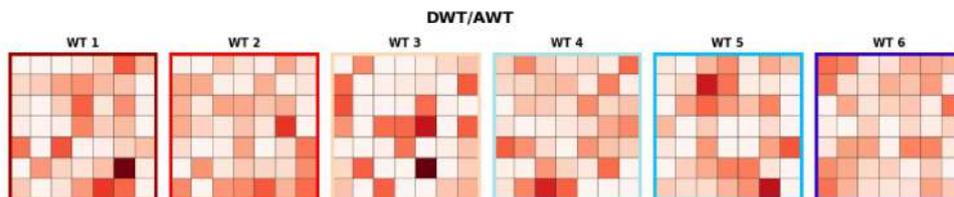
```

AWT transferred to the DWTs:

```

n_clusters_AWT = 6
n_clusters_DWT = 49
n_clusters_MJO = 8
fig = Plot_Probs_WT_WT(bmus_AWT, bmus_DWT, n_clusters_AWT, n_clusters_DWT, ttl =
'DWT/AWT',height = 15, width = 3, wt_colors=True)

```

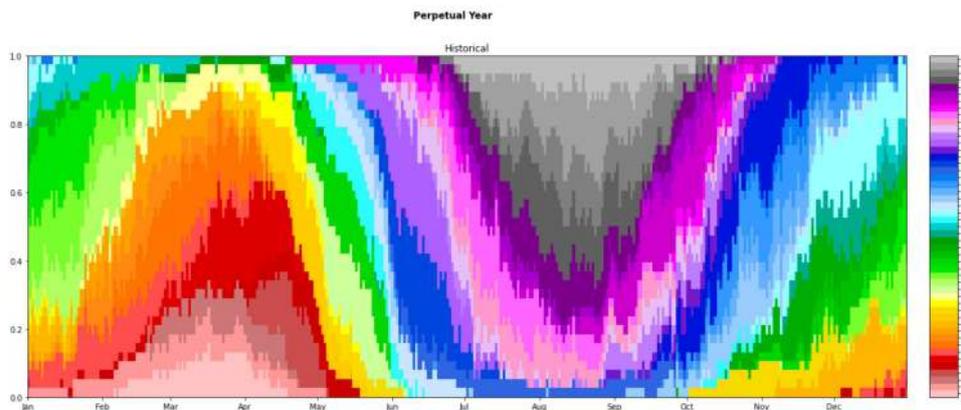


Chronology:

```

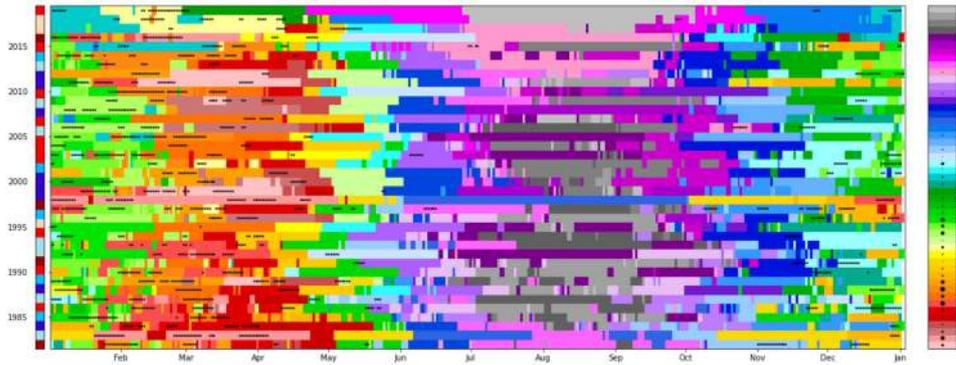
fig = Report_Sim(xds_kma, py_month_ini=1);

```



Chronology during all the calibration period, with the AWT on the left and the TC days included as black dots:

```
fig = Chrono_dwts_tcs(xds_kma, mask_bmus_YD, mask_tcs_YD, awt0_sel_bmus);
```



During the TC season months (November, December, January, February, March and April) the DWTs probability is focused on the upper half of the upper part of the lattice, where most of the TC genesis is also concentrated. In the most intense months (January, February and March) DWTs with the highest number of TCs genesis points are especially likely. On the contrary, in the rest of the months, the probability is shared amongst the DWTs of the lower half of the lattice, where there is very few or null TC genesis activity.

Intra annual variability:

- Months out of the TCs season: purple, pink, gray and blue -> DWTs from 29 to 49 -> low or null TCs genesis activity.
- Months out of the TCs season: green, orange, red, yellow -> DWTs from 1 to 28 -> high TCs genesis activity.

Interannual variability:

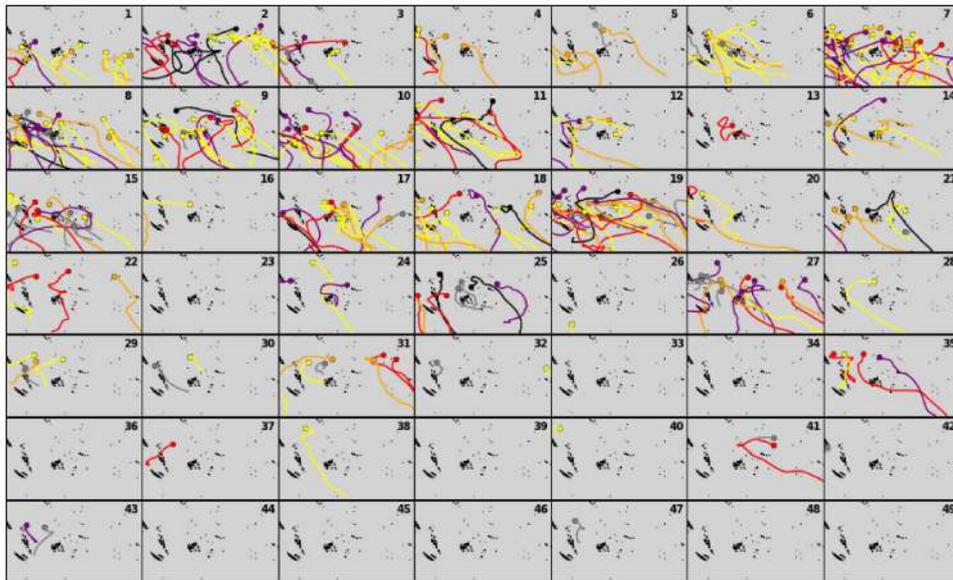
- **El Niño:** not much variability, high TC activity (DWTs like 6 or 18 are very probable), 1997 is the year when the season starts earlier (October) and ends later (June)
- Other factors influences such as long-terms trends, maybe associated to SST warming during this time period.

4.3. Relationship predictor-predictand

4.3.1. DWTs storm frequency and track counting

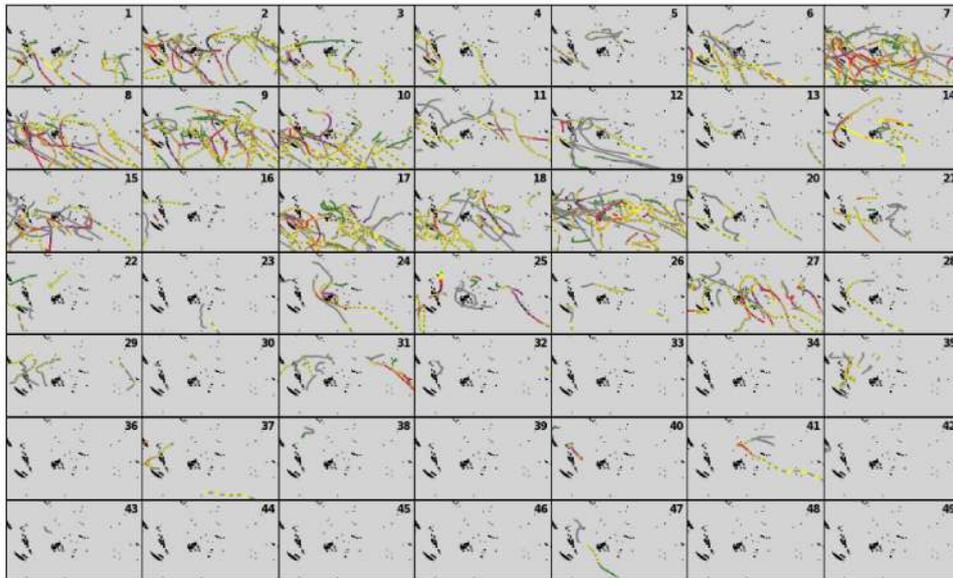
DWTs - TCs tracks according to genesis

```
fig = Plot_DWTs_tracks(xds_kma, xs, st_bmus, st_lons, st_lats, st_categ, mode='genesis',  
cocean='lightgray');
```



DWTs - TCs tracks segments

```
fig = Plot_DWTs_tracks(xds_kma, xs, st_bmus, st_lons, st_lats, st_categ, mode='segments',  
cocean='lightgray');
```



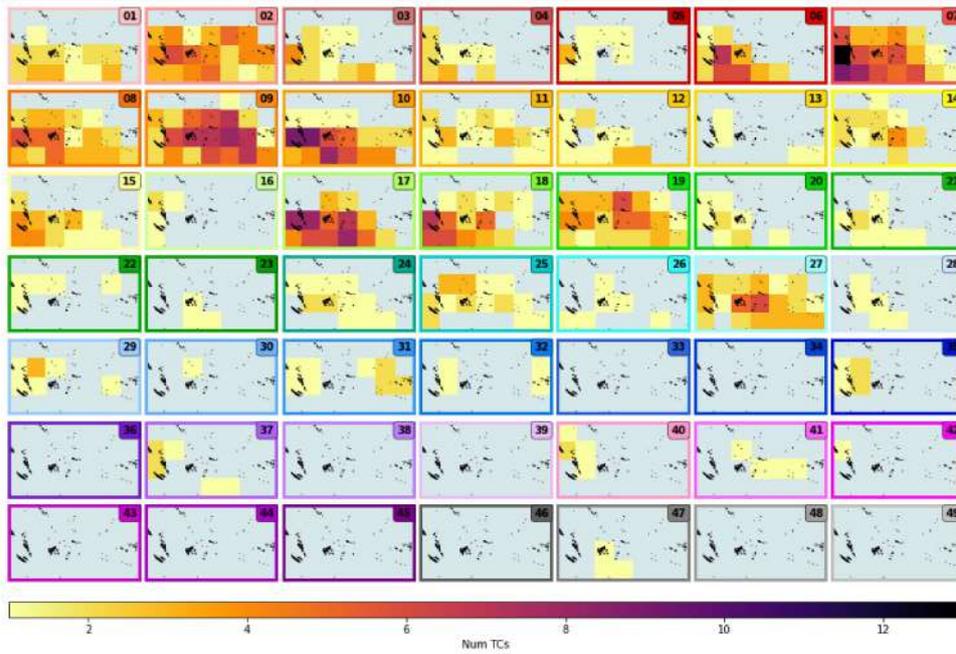
Considering only the genesis point as the transfer criteria to the DWTs make us miss some important information since the TCs tracks develop along many days, during which the DWT can change.

The predictor area is discretized in squared 8° cells to compute the storm frequency per DWT.

Absolute number of TCs per DWT:

```
fig_8 = Plot_DWTs_counts(xs_dwt_counts, mode='counts_tcs');
```

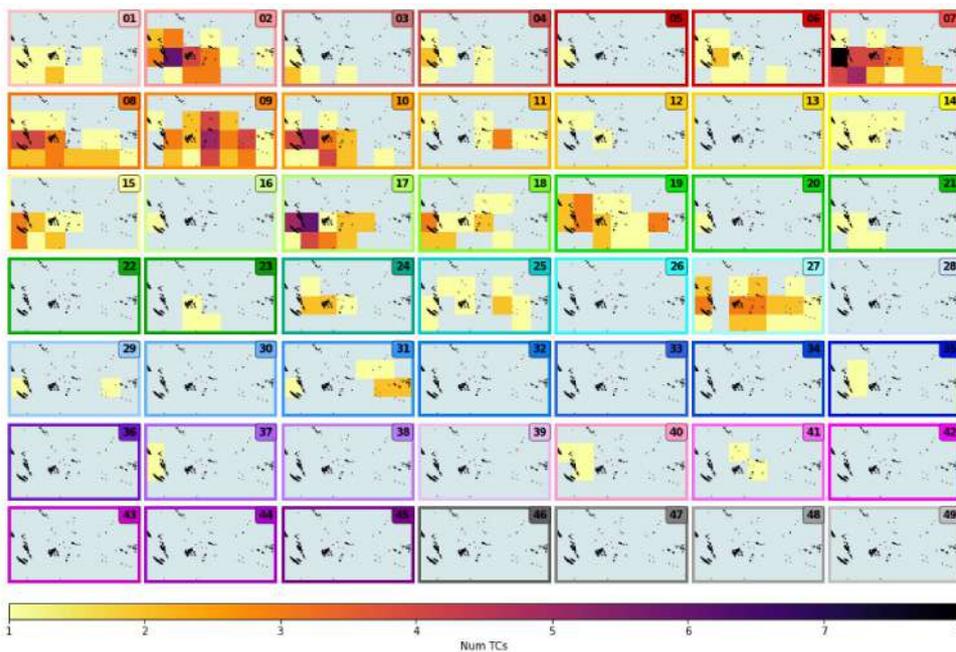
counts_tcs: min(1.0) max(13.0)



Absolute number of TCs reaching category 2 or greater per DWT:

```
fig_8 = Plot_DWTs_counts(xs_dwt_counts_979, mode='counts_tcs');
```

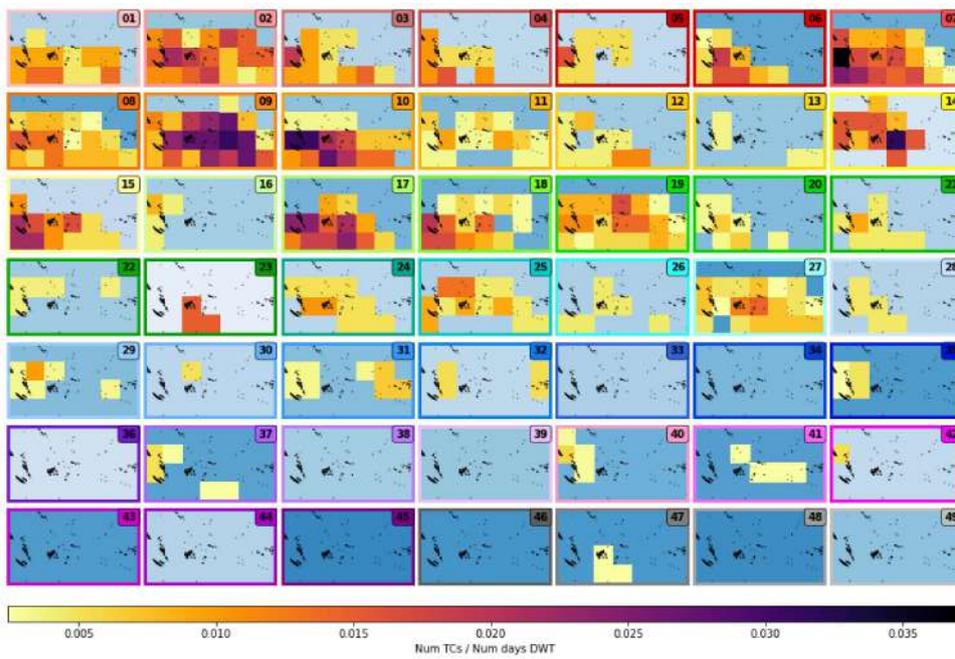
counts_tcs: min(1.0) max(8.0)



Number of TCs per day conditioned to each DWT, with DWT probability as background color:

```
fig_8 = Plot_DWTs_counts(xs_dwt_counts, mode='tcs_dwts');
```

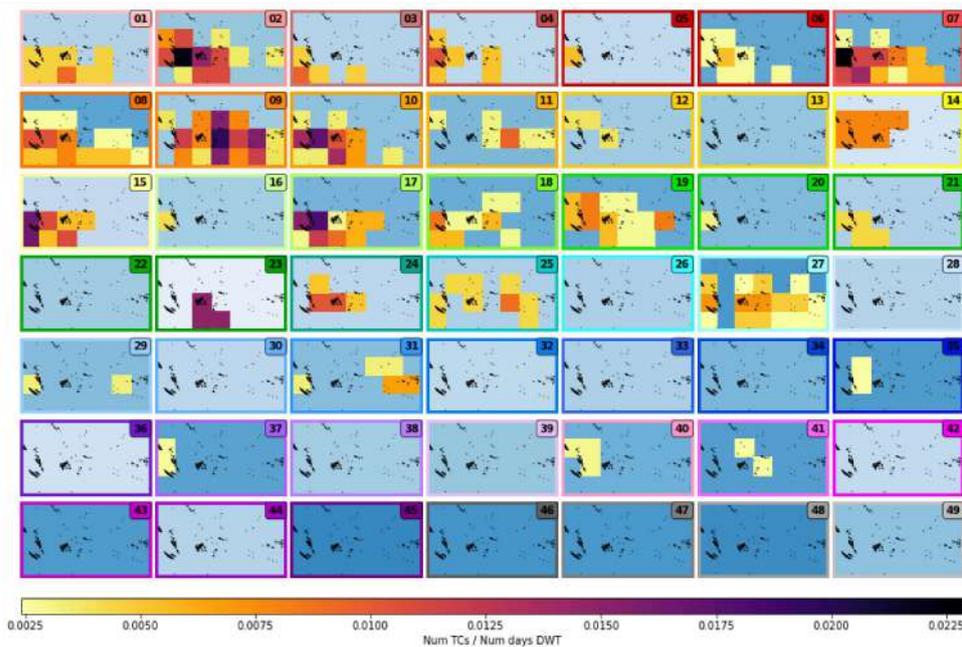
tcs_dwts: min(0.002403846153846154) max(0.037142857142857144)



Number of TCs reaching category 2 or greater per day conditioned to each DWT, with DWT probability as background color:

```
fig_8 = Plot_DWTs_counts(xs_dwt_counts_979, mode='tcs_dwts');
```

```
tcs_dwts: min(0.0024096385542168677) max(0.022857142857142857)
```



4.3.2. Calibration time period predictand plotting

Recall:

- Generation of the index predictor.

- Clustering of the index predictor in 49 synoptic patterns named DWTs.
- Calculation of the number of TCs per day conditioned to each DWT in 8x8° cells in the target area.

Each day of the calibration period has therefore its expected number of TCs map.

```
xds_timeline = ds_timeline(df,xs_dwt_counts,xs_dwt_counts_964,xds_kma)
```

This mean expected daily number of TCs in 8x8° cells is aggregated in monthly basis for an easy management and visualization.

```
# resample months
xds_timeM0 = xds_timeline.resample(time='MS', skipna=True).sum()
del xds_timeM0['bmus']
xds_timeM = xds_timeM0.where(xds_timeM0.probs_tcs > 0, np.nan)
#xds_timeM.to_netcdf(path_p+'xds_timeM8.nc')
xds_timeM
```

xarray.Dataset

► Dimensions: (lat: 5, lon: 8, time: 456)

▼ Coordinates:

time	(time)	datetime64[ns]	1982-01-01 ... 2019-12-01	 
lat	(lat)	int64	-30 -22 -14 -6 2	 
lon	(lon)	int64	160 168 176 184 192 200 208 216	 

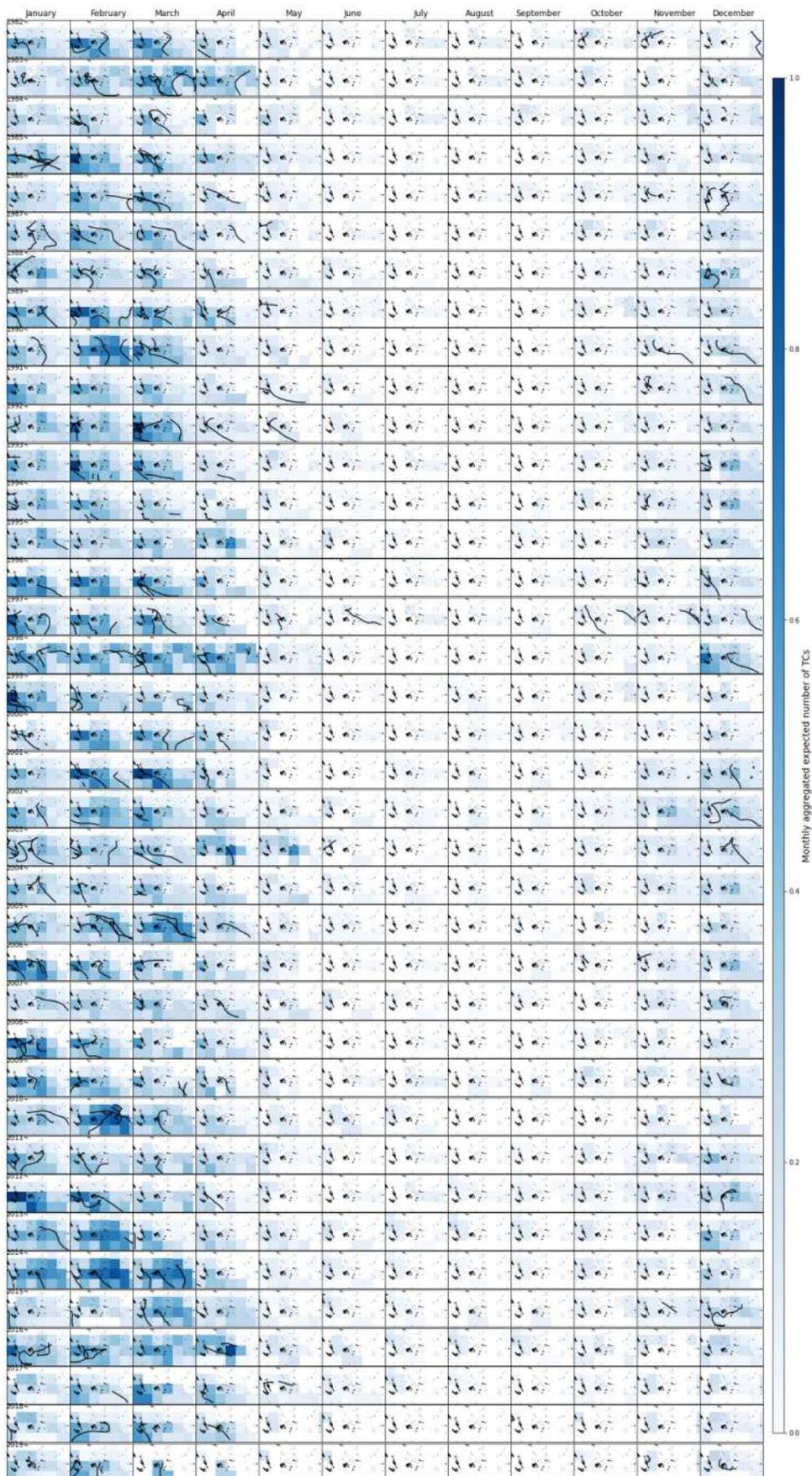
▼ Data variables:

mask_tcs	(time, lat, lon)	float64	8.0 8.0 8.0 8.0 ... nan nan nan nan	 
counts_tcs	(time, lat, lon)	float64	99.0 100.0 120.0 ... nan nan nan	 
counts_tcs_964	(time, lat, lon)	float64	10.0 19.0 22.0 17.0 ... nan nan nan	 
probs_tcs	(time, lat, lon)	float64	0.3036 0.3154 0.3822 ... nan nan	 
probs_tcs_964	(time, lat, lon)	float64	0.02865 0.05969 0.07171 ... nan nan	 

► Attributes: (0)

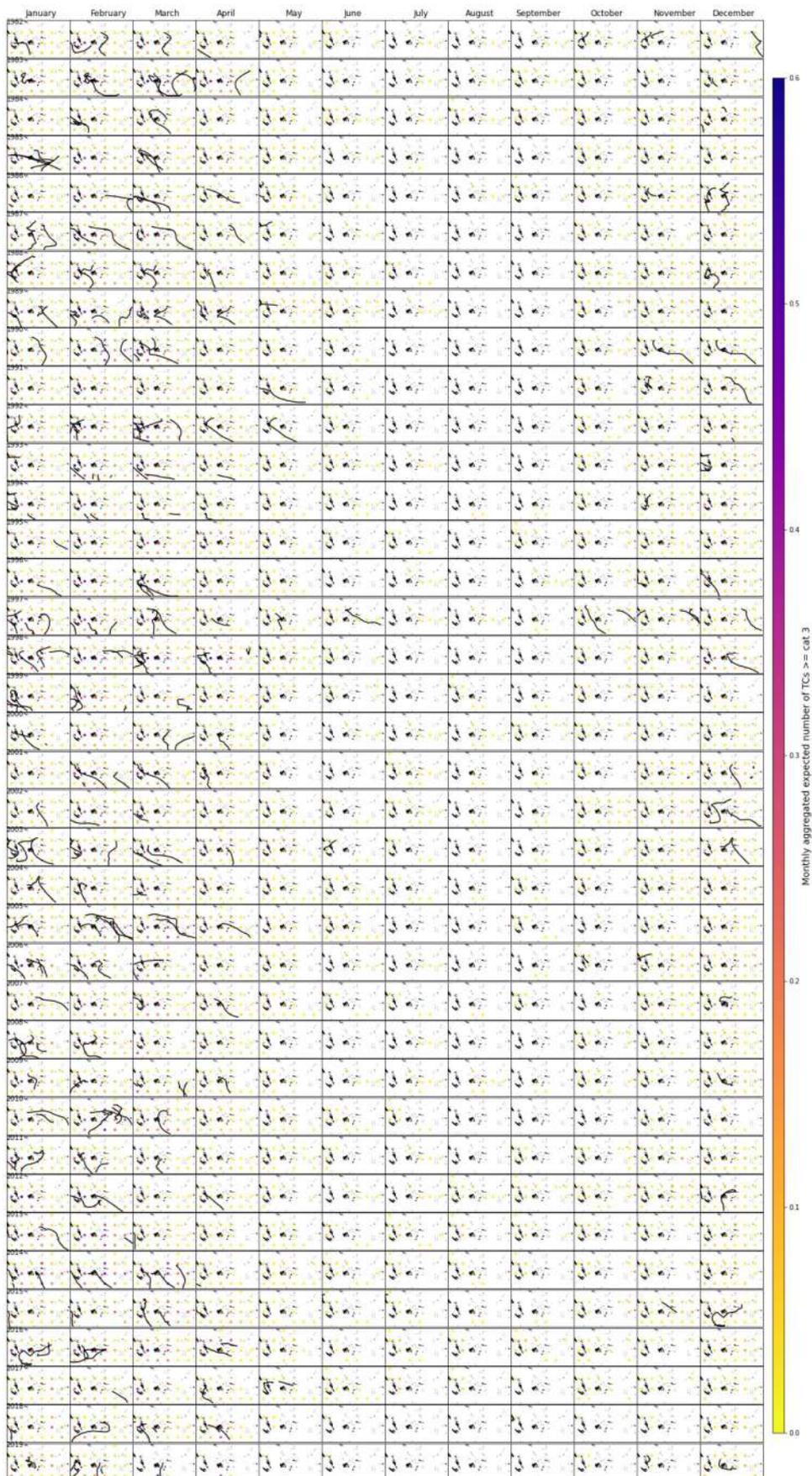
Expected mean number of TCs in 8x8° cells in the target area for the calibration time period (1982-2019), including the historical TC tracks:

```
fig_cali_8 = plot_calibration_period(xds_timeM,xds_timeline,df,1)
```



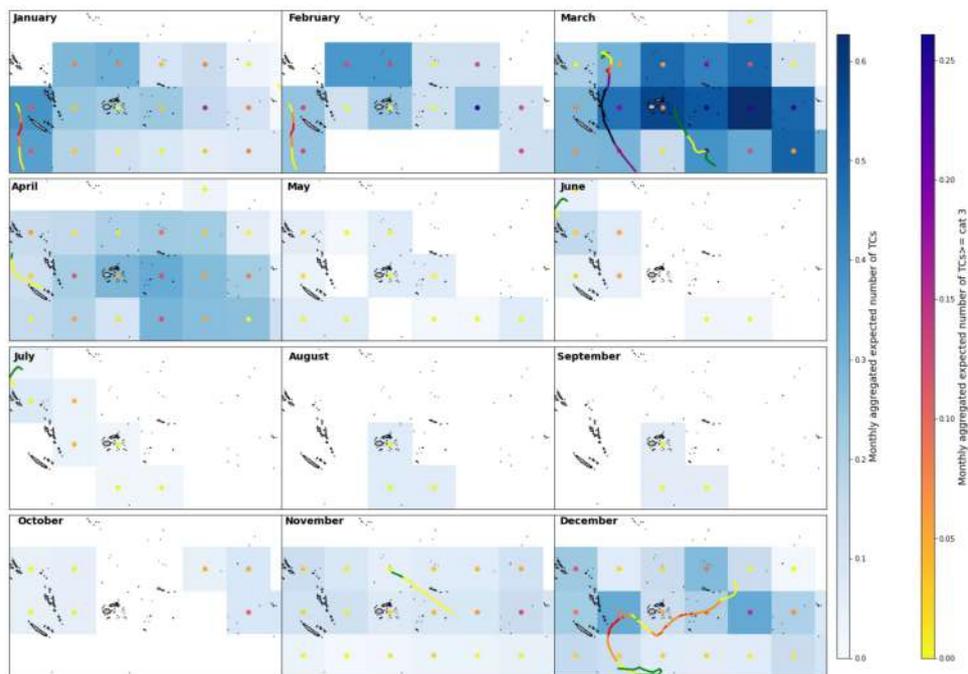
Expected number of TCs reaching category 3 or greater in 8x8° cells in the target area for the calibration time period (1982-2019), including the historical TC tracks:

```
fig_cali_8 = plot_calibration_period_cat3(xds_timeM,xds_timeline,df,0.6,8)
```



Expected number of TCs in 8x8° cells in the target area for year 2015 (El Niño) of the calibration time period (1982-2019) including the historical TC tracks coloured according to their category:

```
fig_year_8 = plot_cali_year(2015,xds_timeline,xds_timeM,df,35)
```



- The model generally performs very well since it effectively and accurately estimates the TC activity (quantity and intensity of TCs).
- In some zones (depending on the month) the model overestimates the number of TCs since rather than predicting the exact path followed by the TC track it predicts the track and its area of influence (surroundings).
- When a TC is very intense or very close in dates to the previous or following month it leaves its footprint (like Pam from 07/03/2015 in February).

```

#Lib
import xarray as xr
from lib.plots_dwts import colorp, custom_colorp, Plot_DWTs_Mean_Anom, Plot_DWTs_totalmean, \
Plot_Probs_WT_WT, Plot_Probs_WT_WT_anomaly, Plot_Probs_WT_WT_WT, Plot_DWTs_Probs, \
Report_Sim_oneyear, Report_Sim, Plot_DWTs_counts, Chrono_dwts_tcs, Chrono_probs_tcs,
Plot_dwts_colormap

from lib.predictor_definition_building import SP_genesis_cat, storms_sp

import warnings
warnings.filterwarnings('ignore')
from IPython.display import Image

```

```

-----
ModuleNotFoundError                                Traceback (most recent call last)
<ipython-input-1-75e1722da373> in <module>
      1 #Lib
      2 import xarray as xr
----> 3 from lib.plots_dwts import colorp, custom_colorp, Plot_DWTs_Mean_Anom,
      4 Plot_DWTs_totalmean, \
      5 Plot_Probs_WT_WT, Plot_Probs_WT_WT_anomaly, Plot_Probs_WT_WT_WT, Plot_DWTs_Probs, \
      6 Report_Sim_oneyear, Report_Sim, Plot_DWTs_counts, Chrono_dwts_tcs, Chrono_probs_tcs,
      7 Plot_dwts_colormap

ModuleNotFoundError: No module named 'lib.plots_dwts'

```

5. Additional Predictand Variables

[Sea Level Pressure \(SLP\)](#)

[Daily mean precipitation](#)

5.1. Sea Level Pressure (SLP)

```

path_p = r'/home/administrador/Documentos/seasonal/seasonal_forecast/new/'

xs = xr.open_dataset(path_p+'xs_index_vars_19822019_2deg_new.nc')
xds_kma = xr.open_dataset(path_p+'kma_model/xds_kma_index_vars_1b.nc')
xs_trmm = xr.open_dataset(path_p+'xs_trmm_1982_2019_2deg_new.nc')
xds_kma_sel = xr.open_dataset(path_p+'kma_model/xds_kma_index_trmm_1b.nc')

```

```

path_st = r'/home/administrador/Documentos/'
xds_ibtracs, xds_SP = storms_sp(path_st)
st_lons = xds_SP.lon.values
st_lats = xds_SP.lat.values

```

```

All basins storms: 13481
SP basin storms: 1130

```

```

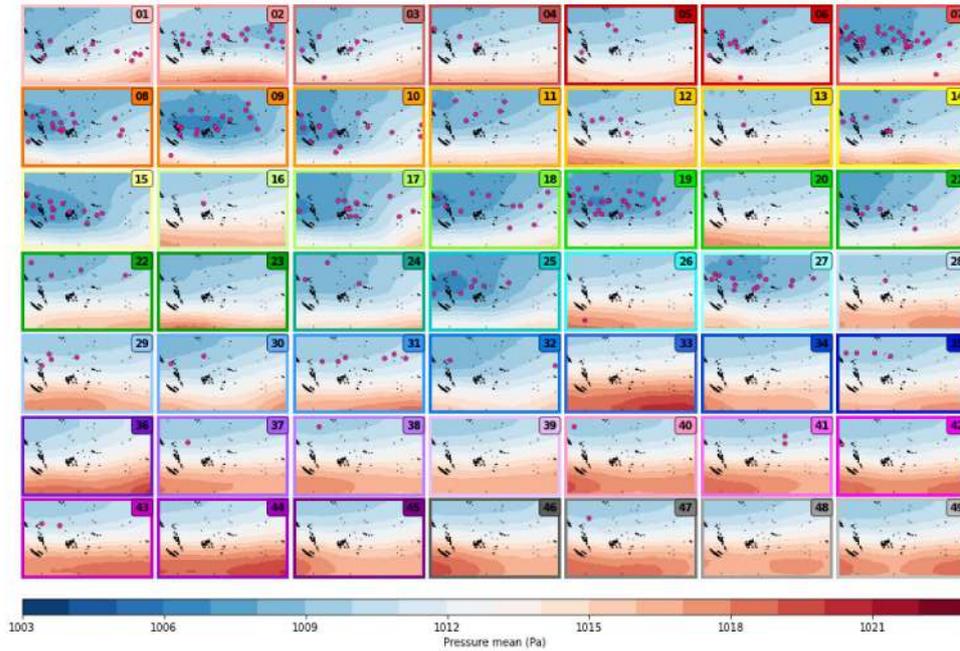
st_bmus = SP_genesis_cat(xds_SP, xds_kma)

```

DWTs - SLP Mean:

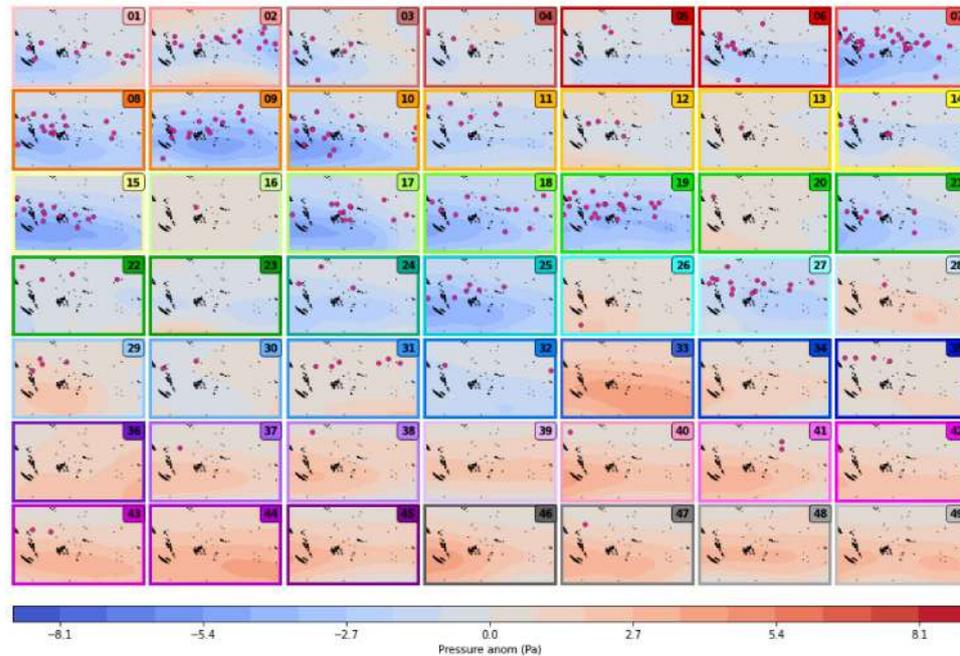
```
# pressure > 1013mbar (anticyclon)
fig = Plot_DWTs_Mean_Anom(xds_kma, xs, ['slp'], minis=[1003], maxis=[1023], levels=[20],
kind='mean', cmap=['RdBu_r'],
genesis='on', st_bmus=st_bmus, st_lons=st_lons, st_lats=st_lats,
markercol='deeppink', markeredge='k');
```

slp: min(96041.796875) max(103900.296875)



DWTs - SLP Anomalies

```
fig = Plot_DWTs_Mean_Anom(xds_kma, xs, ['slp'], minis=[-9], maxis=[9], levels=[20],
kind='anom', cmap=['coolwarm'],
genesis='on', st_bmus=st_bmus, st_lons=st_lons, st_lats=st_lats,
markercol='deeppink', markeredge='k');
```

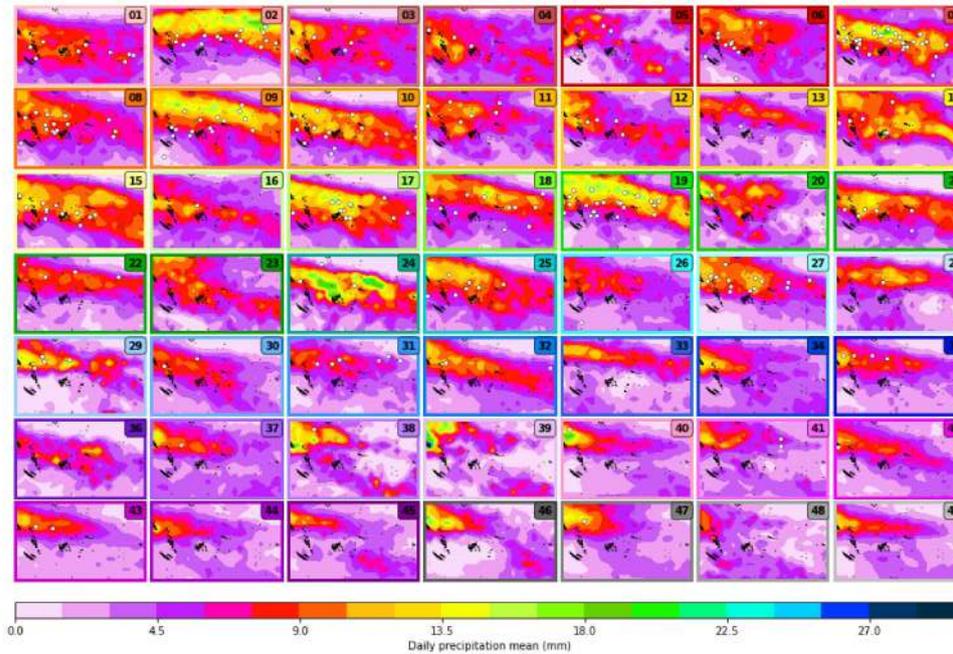


5.2. Daily Mean Precipitation

DWTs - Daily Mean Precipitation Mean

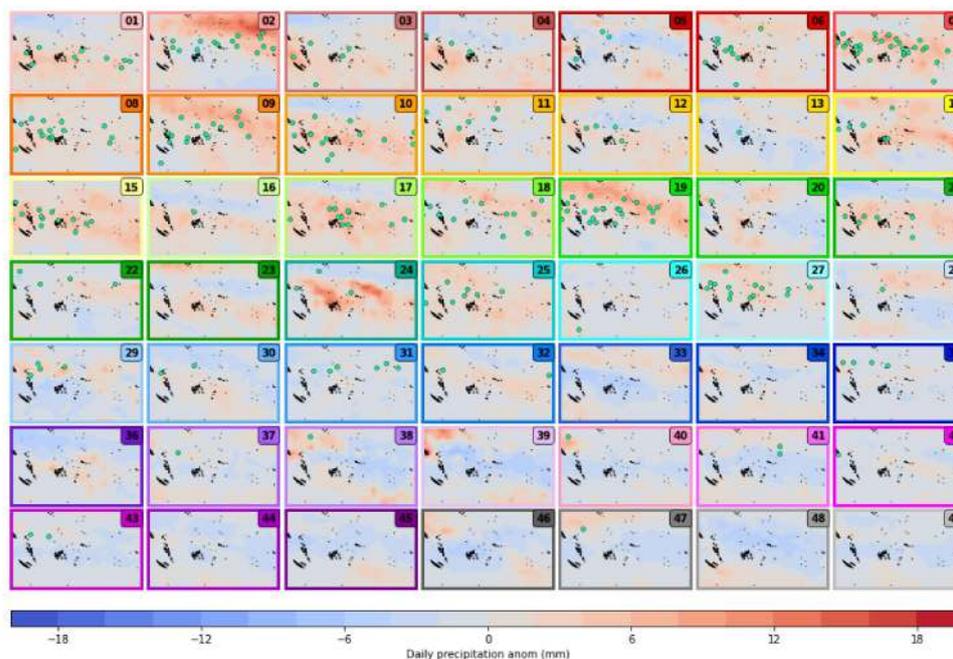
```
fig = Plot_DWTs_Mean_Anom(xds_kma_sel, xs_trmm, ['precipitation'], minis=[0], maxis=[30],  
levels=[20], kind='mean', cmap=['gist_ncar_r'],  
genesis='on', st_bmus=st_bmus, st_lons=st_lons, st_lats=st_lats,  
markercol='white', markeredge='k');
```

precipitation: min(0.0) max(635.4000244140625)



DWTs - Daily Mean Precipitation Anomalies

```
fig = Plot_DWTs_Mean_Anom(xds_kma_sel, xs_trmm, ['precipitation'], minis=[-20], maxis=[20],  
levels=[20], kind='anom', cmap=['coolwarm'],  
genesis='on', st_bmus=st_bmus, st_lons=st_lons, st_lats=st_lats,  
markercol='mediumspringgreen', markeredge='k');
```



Clear patterns can be extracted from these figures related to TCs genesis. Most of it takes place under the following conditions:

- Low pressure areas, with 1013 mba or lower values, corresponding generally to negative anomalies.
- Intense but not extreme precipitation areas, from 9 to 16.5 mm/day, corresponding generally to red anomalies.

These conclusions add to the summary relationship predictand - predictors:

- **TC activity is focused in the first 28 DWTs (index clusters).**
- **The TCs genesis activity is generally focused under the following conditions:**
 - Index range values from 0.60 to 0.79, corresponding to positive anomalies.
 - In the warm SST zone, 28 - 30 ° C, and where MLD values are smaller than 75 m; corresponding to mild positive SST anomalies and negative MLD anomalies.
 - In intense but not extreme precipitation areas, from 9 to 16.5 mm/day, corresponding generally to red anomalies.
 - In low pressure areas, with 1013 mba or lower values, corresponding generally to negative anomalies.

By Andrea Pozo

© Copyright 2020.

```

# basic
import sys
import os

# common
import numpy as np
import pandas as pd
import xarray as xr
import matplotlib.pyplot as plt
from datetime import datetime, timedelta
from sklearn.cluster import KMeans
import pickle
import warnings
warnings.filterwarnings('ignore')
from IPython.display import Image

#Lib
from lib.validation_methodology_plots import *

```

```

-----
ModuleNotFoundError                                Traceback (most recent call last)
<ipython-input-1-e8e2325033f1> in <module>
     16
     17 #Lib
--> 18 from lib.validation_methodology_plots import *

ModuleNotFoundError: No module named 'lib.validation_methodology_plots'

```

```
path_p = r'/home/administrador/Documentos/seasonal/seasonal_forecast/new/'
```

```

df_2021 = pd.read_pickle(path_p+'df_coordinates_pmin_sst_mld_2021.pkl')
xs = xr.open_dataset(path_p+'xs_index_vars_19822019_2deg_new.nc')
xds_kma = xr.open_dataset(path_p+'kma_model/xds_kma_index_vars_1b.nc')
xs_dwt_counts = xr.open_dataset(path_p+'kma_model/xds_count_tcs8.nc')
xs_dwt_counts_964 = xr.open_dataset(path_p+'kma_model/xds_count_tcs8_964.nc')
xds_timeM = xr.open_dataset(path_p+'xds_timeM8.nc')
xds_PCA = xr.open_dataset(path_p+'xds_PCA.nc')
xds_kma_ord = xr.open_dataset(path_p+'xds_kma_ord.nc')

```

6. Model Validation

[Index predictor](#)

[Cluster comparison](#)

[Predictand computation and plotting](#)

After analyzing the tailor-made predictor along the hindcast data for the calibration period (1982-2019), the performance of the model will be validated for year 2020, which has not been included in the predictor calibration process.

Steps:

- 1. Download and preprocess (file conversion and resolution interpolation) SST and MLD data for the validation time period.

- 2. Generation of the index predictor based on the index function obtained at the calibration period.
- 3. The fitted Principal Component Analysis for the calibration is used to predict the index principal components in that same temporal-spatial space.
- 4. The predicted PCs are assigned to the best match unit group from the fitted K-means clustering -> based on the index predictor a DWT is assigned to each day.
- 5. From the DWT the expected daily mean number of TCs in 8x8° cells map in the target area is known.

6.1. Index predictor and DWTs

Download and preprocess (file conversion and resolution interpolation) SST and MLD data for the validation time period.

```
path_val = r'/home/administrador/Documentos/seasonal/seasonal_forecast/validation/'
year_val = 2020
```

```
change_sst_resolution_val(path_val,year_val)
```

```
Start time: 2021-04-12 09:57:46.179185
End time: 2021-04-12 10:01:47.651739
```

Generation of the index predictor based on the index function obtained at the calibration period.

```
xs_val = ds_index_over_time_val(path_val,path_p,year_val)
xs_val
```

xarray.Dataset

► Dimensions: (lat: 16, lon: 26, time: 366)

▼ Coordinates:

time	(time)	datetime64[ns]	2020-01-01 ... 2020-12-31	 
lat	(lat)	float64	-0.25 -2.25 -4.25 ... -28.25 -30.25	 
lon	(lon)	float64	160.2 162.2 164.2 ... 208.2 210.2	 

▼ Data variables:

index	(time, lat, lon)	float64	0.0 0.05833 0.0 0.0 ... 0.0 0.0 0.0	 
sst	(time, lat, lon)	float32	30.49 29.95 29.31 ... 23.79 24.04	 
dbss	(time, lat, lon)	float32	98.42 106.0 109.8 ... 13.92 12.21	 
mask	(lat, lon)	float64	0.0 0.0 0.0 0.0 ... 0.0 0.0 0.0	 

► Attributes: (0)

The fitted Principal Component Analysis for the calibration is used to predict the index principal components in that same temporal-spatial space and the predicted PCs are assigned to the best match unit group from the fitted K-means clustering -> based on the index predictor a DWT is assigned to each day.

```
val_bmus = PCA_k_means_val(path_p,path_val,xs_val)
```

K-means order previously obtained in the calibration period:

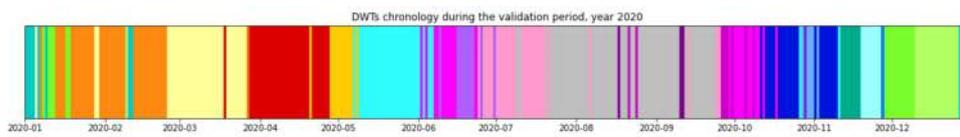
```
[11 27 28 45 16 23 47 19 7 14 0 31 46 30 34 35 15 2 8 33 37 41 5 10
 48 4 40 21 3 20 42 29 38 26 32 43 13 17 18 39 25 22 44 9 1 24 6 12
 36]
```

DWTs for the validation period and their counting:

```
(array([ 6, 8, 9, 12, 15, 17, 18, 24, 25, 26, 27, 30, 32, 35, 36, 37, 40,
        42, 43, 45, 49]), array([[31, 1, 37, 11, 33, 18, 18, 8, 13, 30, 8, 5, 1, 21,
        1, 10, 27,
        22, 8, 3, 60]]))
```

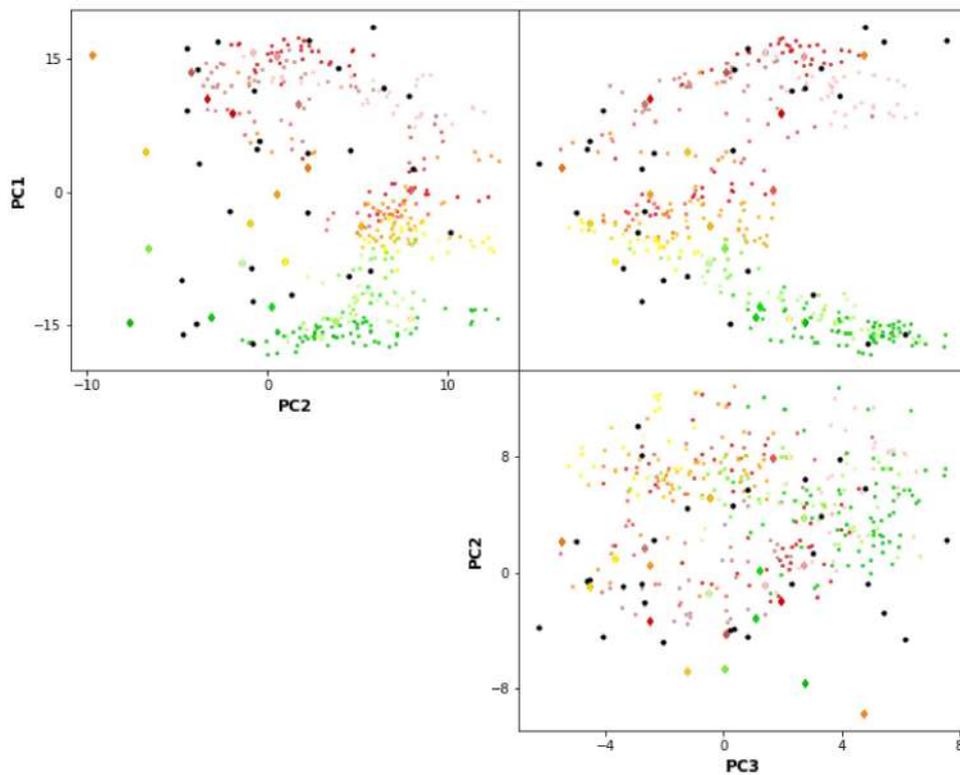
Chronology of the DWTs:

```
fig_bmus = plot_bmus_chronology(xs_val,val_bmus,year_val)
```



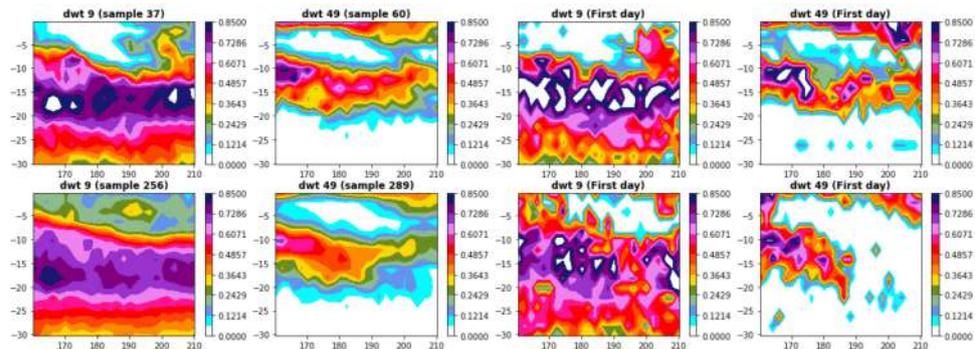
The resulting classification can be seen in the PCs space of the predictor index data. The obtained centroids (black dots), span the wide variability of the data.

```
fig = plot_scatter_kmeans(xds_kma_ord, val_bmus, xds_kma_ord.cenEOFs.values, size_l=12,
size_h=10);
```



6.2. Cluster comparison

```
fig = plot_bmus_comparison_validation_calibration(xs,xds_kma,xs_val,val_bmus,9,49)
```



6.3. Predictand computation and plotting

From the DWT the daily expected mean number of TCs in $8 \times 8^\circ$ cells in the target area is known for each day and thus maps at different time scales can be computed.

Daily mean expected number of TCs

```
xds_timeline_val,xs_M_val =
ds_monthly_probabilities_val(df_2021,val_bmus,xs_val,xs_dwt_counts,xs_dwt_counts_964)
```

Monthly aggregated mean expected number of TCs

```
xs_M_val
```

xarray.Dataset

► Dimensions: (lat: 5, lon: 8, time: 12)

▼ Coordinates:

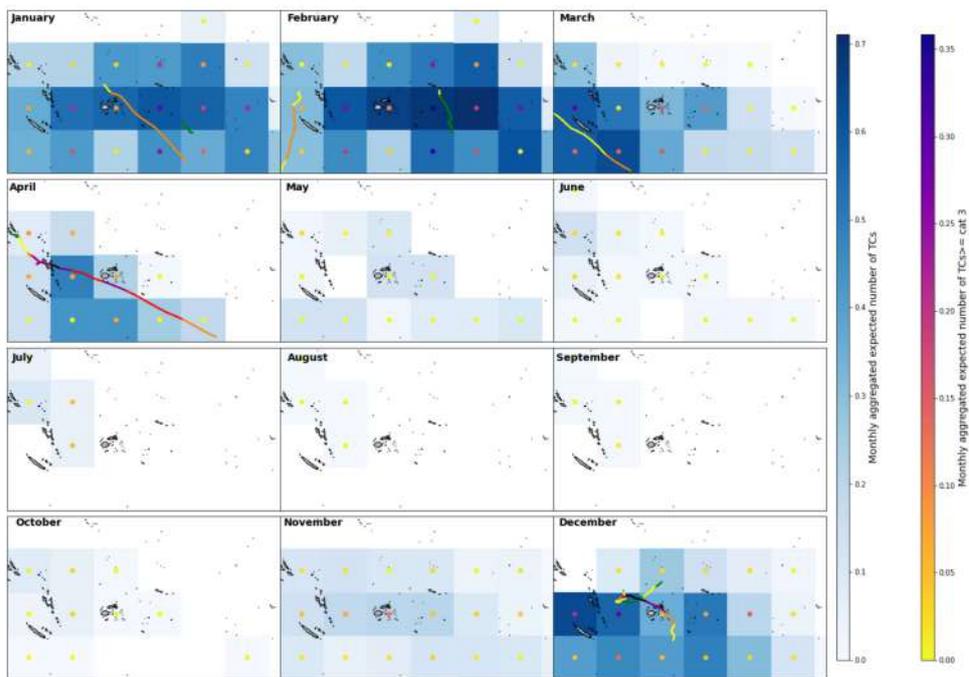
time	(time)	datetime64[ns]	2020-01-01 ... 2020-12-01	
lat	(lat)	int64	-30 -22 -14 -6 2	
lon	(lon)	int64	160 168 176 184 192 200 208 216	

▼ Data variables:

mask_tcs	(time, lat, lon)	float64	7.0 7.0 7.0 7.0 ... nan nan nan nan	
counts_tcs	(time, lat, lon)	float64	92.0 112.0 62.0 ... nan nan nan	
counts_tcs_964	(time, lat, lon)	float64	15.0 41.0 3.0 64.0 ... nan nan nan	
probs_tcs	(time, lat, lon)	float64	0.3473 0.4187 0.2266 ... nan nan	
probs_tcs_964	(time, lat, lon)	float64	0.06037 0.1586 0.01639 ... nan nan	

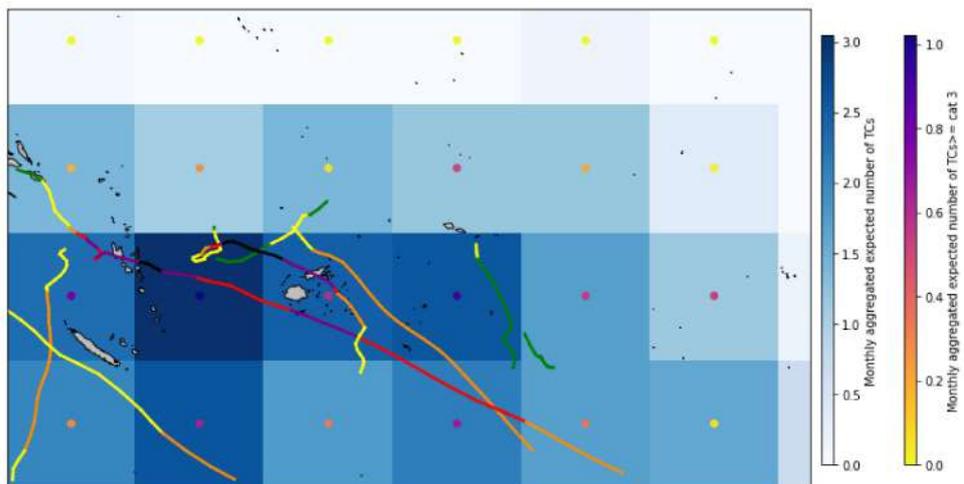
► Attributes: (0)

```
fig_val_year_8 = plot_validation_year(df_2021,xs_M_val,xds_timeline_val,35)
```



Whole period aggregated mean expected number of TCs

```
fig_val_year_8 = plot_validation_full_season(df_2021,xs_M_val,xds_timeline_val,35)
```



- The model performs very well when estimating the expected TC activity (number and intensity of TCs), not underestimating the threat.
- In some cells adjacent to the cells including TC tracks it overestimates TC activity.

```

# basic
import sys
import os

# common
import numpy as np
import pandas as pd
import xarray as xr
import matplotlib.pyplot as plt
from datetime import datetime, timedelta
from sklearn.cluster import KMeans
import pickle
import warnings
warnings.filterwarnings('ignore')
from IPython.display import Image

#Lib
from lib.forecast_methodology import *
from lib.forecast_plots import *
from lib.calibration import ds_timeline
from lib.extract_tcs import tcs_count_season, counting_ds

```

```

-----
ModuleNotFoundError                                Traceback (most recent call last)
<ipython-input-1-e10a6add1a0> in <module>
    16
    17 #Lib
--> 18 from lib.forecast_methodology import *
    19 from lib.forecast_plots import *
    20 from lib.calibration import ds_timeline

ModuleNotFoundError: No module named 'lib.forecast_methodology'

```

```

path_p = r'/home/administrador/Documentos/seasonal/seasonal_forecast/new/'
path_f = r'/media/administrador/SAMSUNG/seasonal_forecast/data/CFS/forecast_past_data/'

```

```

df = pd.read_pickle(path_p+'df_coordinates_pmin_sst_mld_2019.pkl')
xs = xr.open_dataset(path_p+'xs_index_vars_19822019_2deg_new.nc')
xds_kma = xr.open_dataset(path_p+'kma_model/xds_kma_index_vars_1b.nc')
xs_dwt_counts = xr.open_dataset(path_p+'kma_model/xds_count_tcs3.nc')
xs_dwt_counts_964 = xr.open_dataset(path_p+'kma_model/xds_count_tcs3_964.nc')
xds_timeM = xr.open_dataset(path_p+'xds_timeM3.nc')
xds_PCA = xr.open_dataset(path_p+'xds_PCA.nc')

```

7. Seasonal Forecast

[Forecast data validation](#)

[Forecast application](#)

[Forecast outlook from different agencies](#)

7.1. Forecast data validation

After building and validating the model it will be applied to forecast TCs seasons from previous years to validate this data.

The forecasts from the first day of August, September, October and November. There are four runs per day of the model (00:00,06:00,12:00,18:00 hours).

Steps:

1. Download [from CFS 9-month operational forecast](#) and preprocess (file conversion and resolution interpolation) SST and MLD data.
2. Generation of the index predictor based on the index function obtained at the calibration period.
3. The fitted Principal Component Analysis for the calibration is used to predict the index principal components in that same temporal-spatial space.
4. The predicted PCs are assigned to the best match unit group from the fitted K-means clustering -> based on the index predictor a DWT is assigned to each day.
5. From the DWT the expected daily mean number of TCs in 8x8° cells in the target area is known.

```
#your path to the datasets containing the MLD and SST data of the season
path = path_f+'season_18_19/'

#year when the season starts
y = 2018

process_season_with_file(path,path_p,y,xds_PCA,xs_dwt_counts,xs_dwt_counts_964)
```

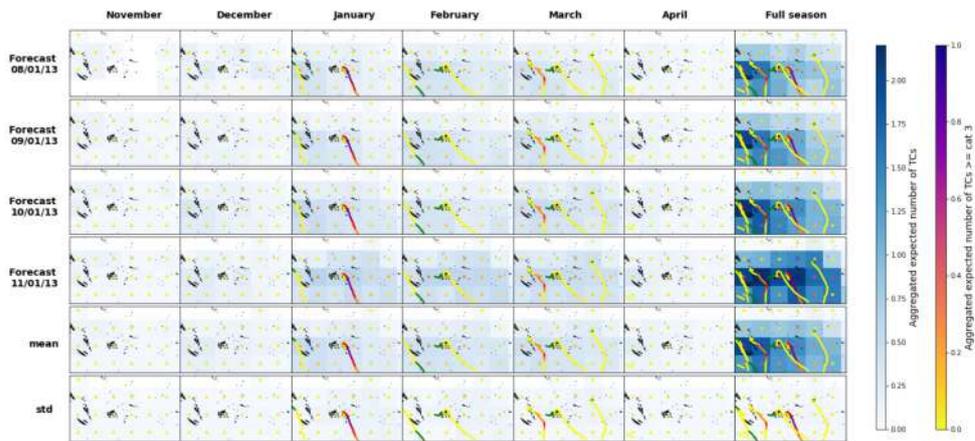
```
Forecast month 8
Hour 0
Unique DWTs(array([ 1,  2,  3,  4,  5,  6,  8,  9, 10, 11, 13, 16, 22, 24, 29, 33, 34,
                   39, 45, 48, 49]), array([ 3, 34, 15,  5,  9, 11,  5,  2,  8,  2, 65,  3, 18,  2,
                   14, 10, 15,
                   3,  2, 17, 62]))
```

SEASON 2013-2014

```
path2 = path_f+'season_13_14/'
y = 2013
mean_8, mean_8_c3, std_8, std_8_c3, mean_9, mean_9_c3, std_9, std_9_c3, mean_10, mean_10_c3,
std_10, std_10_c3, mean_11, mean_11_c3, std_11, std_11_c3, lmean,
lmean_c3, lstd, lstd_c3, mean_fs, mean_fs_c3, mean_mean, mean_mean_c3, std_mean, std_mean_c3, ds4 =
variables_plot_season(path2,y,5,8)
```

```
fig8_13_14 = plot_season(ds4,13,2013,2.2,1, mean_8, mean_8_c3, std_8, std_8_c3, mean_9,
mean_9_c3, std_9, std_9_c3, mean_10, mean_10_c3, std_10, std_10_c3, mean_11, mean_11_c3,
std_11, std_11_c3, lmean,
lmean_c3, lstd, lstd_c3, mean_fs, mean_fs_c3, mean_mean, mean_mean_c3, std_mean, std_mean_c3, xds_tim
eM,12)
```

Forecast for season 11/13 to 04/14

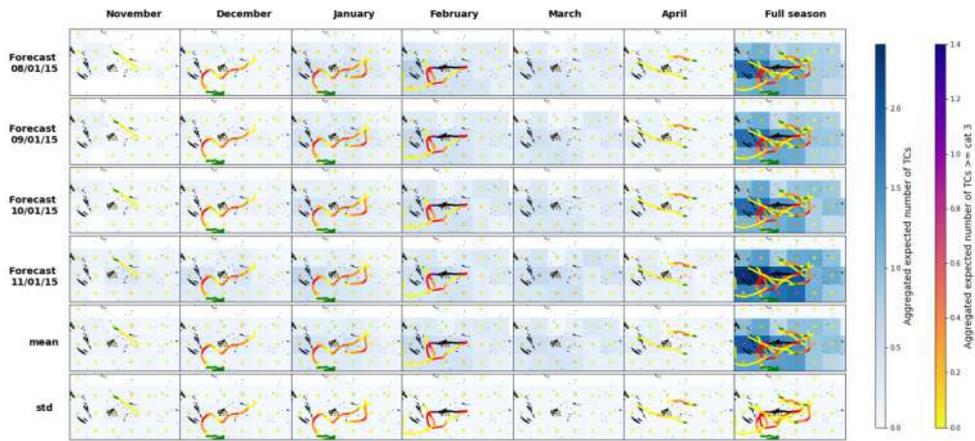


SEASON 2015-2016 (El Niño event)

```
path2 = path_f+'season_15_16/'
y = 2015
mean_8, mean_8_c3, std_8, std_8_c3, mean_9, mean_9_c3, std_9, std_9_c3, mean_10, mean_10_c3,
std_10, std_10_c3, mean_11, mean_11_c3, std_11, std_11_c3, lmean,
lmean_c3, lstd, lstd_c3, mean_fs, mean_fs_c3, mean_mean, mean_mean_c3, std_mean, std_mean_c3, ds4 =
variables_plot_season(path2, y, 5, 8)
```

```
fig8_15_16 = plot_season(ds4, 15, 2015, 2.4, 1.4, mean_8, mean_8_c3, std_8, std_8_c3, mean_9,
mean_9_c3, std_9, std_9_c3, mean_10, mean_10_c3, std_10, std_10_c3, mean_11, mean_11_c3,
std_11, std_11_c3, lmean,
lmean_c3, lstd, lstd_c3, mean_fs, mean_fs_c3, mean_mean, mean_mean_c3, std_mean, std_mean_c3, xds_tim
eM, 12)
```

Forecast for season 11/15 to 04/16

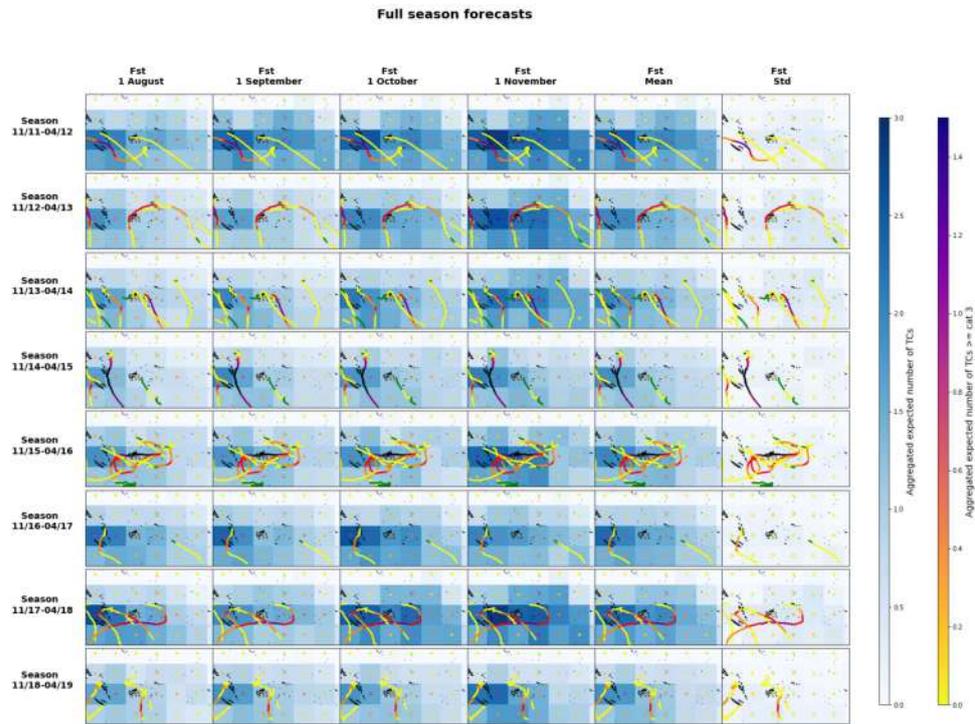


FULL SEASONS

```
seasons = [11,12,13,14,15,16,17,18]
list_fs, list_fs_c3, list_mfs, list_mfs_c3, list_stdfs, list_stdfs_c3 =
variables_plot_season_means(path_f, seasons, 5, 8)
```

11
12
13
14
15
16
17
18

```
fig8_full_seasons = plot_season_means(ds4,3,1.5,list_fs, list_fs_c3,list_mfs,list_mfs_c3,
list_stdfs, list_stdfs_c3,xds_timeM,12)
```



```
xds_timeline = ds_timeline(df,xs_dwt_counts,xs_dwt_counts_964,xds_kma)
```

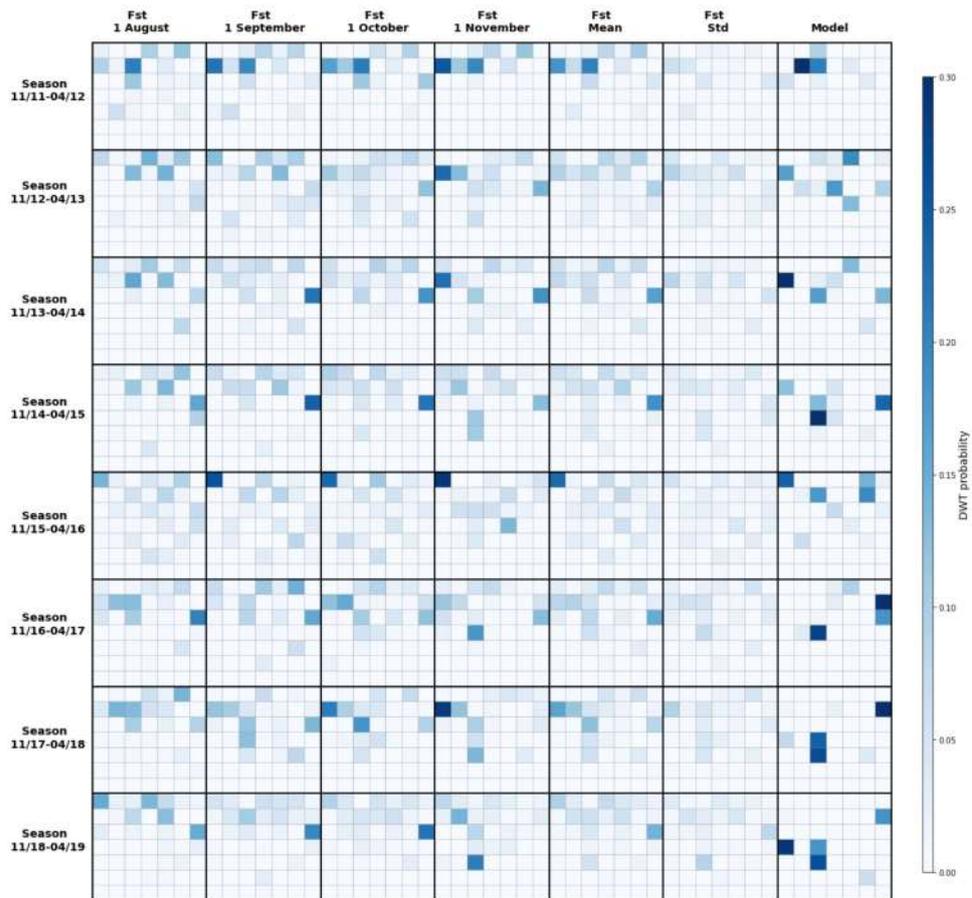
FORECAST vs MODEL DWTs probabilities

```
#bmus from the calibration period of the methodology, from 01/1982 to 12/2019
ds_bmus = xr.Dataset({
    'bmus': (('time'), xds_kma.bmus.values),
}, {
    'time': xds_kma.time.values,
})
print(ds_bmus)
```

```
<xarray.Dataset>
Dimensions: (time: 13879)
Coordinates:
  * time      (time) datetime64[ns] 1982-01-01 1982-01-02 ... 2019-12-31
Data variables:
  bmus       (time) int64 26 26 23 23 23 23 10 26 26 ... 24 24 24 24 24 24 24 24
```

```
s11,m11,std11,s12,m12,std12,s13,m13,std13,s14,m14,std14,s15,m15,std15,s16,m16,std16,s17,m17,
std17,s18,m18,std18,s19,m19,std19,list_metm,list_smet =
variables_dwt_forecast_plot(ds_bmus,2011,path_f+'season_')
```

```
fig_dwt =
dwt_forecast_plot(0.3, s11,m11, std11, s12,m12, std12, s13,m13, std13, s14,m14, std14, s15,m15, std15,
s16,m16, std16, s17,m17, std17, s18,m18, std18, s19,m19, std19, list_metm, list_smet)
```



Recall:

- The model performs very well when estimating the expected TC activity (number and intensity of TCs), not underestimating the threat.
- In some cells adjacent to the cells including TC tracks it overestimates TC activity.

Forecast performance:

- Much greater uncertainty -> DWTs probability is greatly shared and therefore more extended predictand maps.
- The weakness of the model are enhanced -> more extended (more overstimulation in the surroundings of the tracks) and more homogeneous maps.
- When there is an unusually high TC activity (Season 15-16) the forecast underestimates it in the most active cells.

Conclusion: Although the model has been proven to perform very well, the accuracy and reliability of the forecast greatly depends of the quality of the forecast data, which has also been proven to be quite improvable.

7.2. Forecast application

Predictor data from the first day of April is downloaded and preprocessed, and the statistical model is applied to compute the TCs forecast from this day to next 9 months.

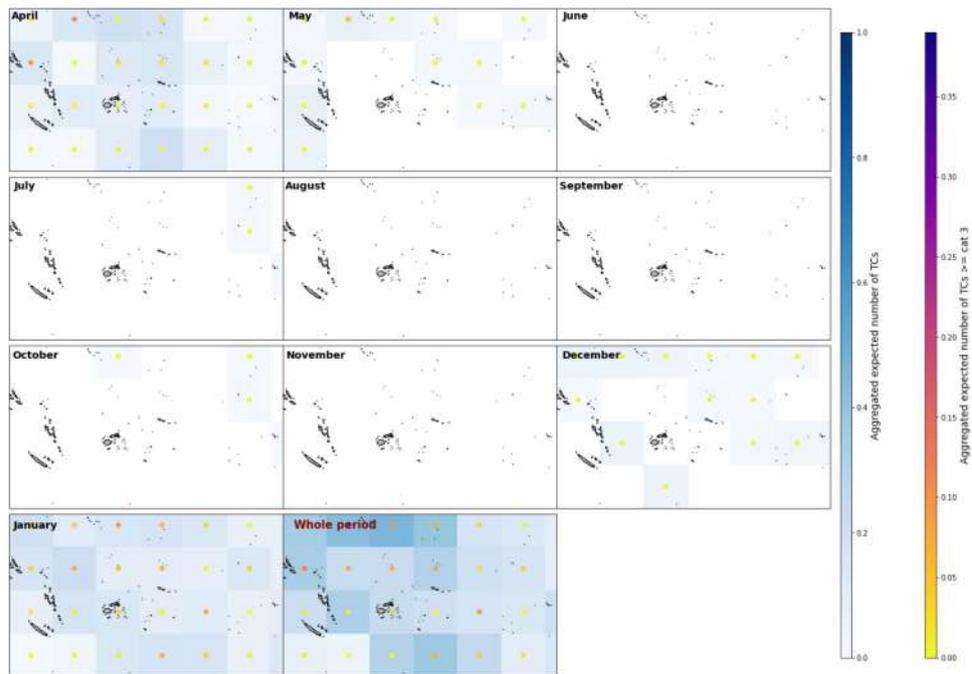
```
#your path to the datasets containing the MLD and SST data
path = r'/media/administrador/SAMSUNG/seasonal_forecast/data/CFS/forecast_01_04_2021/'

#year and month of the forecast day
y = 2021
m = 4
```

```
process_day_with_file(path,path_p,y,m,xds_PCA,xs_dwt_counts,xs_dwt_counts_964)
```

```
Forecast month 4
Hour 0
Unique DWTs(array([ 1,  4,  5,  7, 11, 13, 16, 20, 22, 29, 33, 34, 35, 40, 41, 45, 46,
                    47, 48, 49]), array([ 4, 16, 13, 21, 10, 19, 13,  8, 24, 25,  1, 19,  3, 11, 37,
                    8, 67,
                    3,  4,  1]))
xds_timeline_M4_f1.nc
xs_M4_f1.nc
Hour 6
Unique DWTs(array([ 1,  4,  5,  7, 11, 13, 16, 29, 33, 34, 39, 41, 43, 44, 48, 49]),
array([ 6, 20, 12, 13, 21, 53, 17,  1, 12, 34, 10,  5,  6,  1, 76, 20]))
xds_timeline_M4_f2.nc
xs_M4_f2.nc
Hour 12
Unique DWTs(array([ 1,  4,  5,  6, 11, 13, 16, 20, 22, 29, 34, 38, 39, 40, 41, 46, 47,
                    48, 49]), array([ 3, 13, 19,  1, 14, 33, 15,  4,  6, 36, 19, 11,  7,  2, 16,  9,
                    2,
                    96,  1]))
xds_timeline_M4_f3.nc
xs_M4_f3.nc
Hour 18
Unique DWTs(array([ 1,  4,  5,  7, 10, 11, 13, 16, 18, 20, 22, 25, 29, 32, 33, 34, 38,
                    39, 41, 46, 48]), array([ 3, 24,  9,  4,  1,  2, 17, 10, 15,  7, 13,  1,  5, 19,
                    12, 40, 10,
                    1, 15,  6, 92]))
xds_timeline_M4_f4.nc
xs_M4_f4.nc
```

```
mean_day, mean_day_c3, mean_day_full_season, mean_day_full_season_c3, ds =
variables_plot_forecast(path,2021,4,5,8)
fig8_f = plot_forecast(ds4,1,0.39,mean_day, mean_day_c3,
mean_day_full_season,mean_day_full_season_c3,35)
```



7.3. Forecast outlook from different agencies

```
from IPython.display import IFrame
```

[RMSC Nadi \(Fiji Meteorological Service\), SP>25°S](#)

They generate TCs outlooks for the season, based in other seasons with similar patterns in the index NIÑO3.4.

There is a also a 5 days risk of cyclones outlook, consisting on maps.

They have as well maps of TC tracks prediction when there is TC activity at the moment.

```
IFrame('https://www.met.gov.fj/index.php?
page=climateOutlooks#Detailed_Outlook.pdf',width=1300,height=450,embed = True)
```



Latest Climate Outlooks

**Climate Outlook (index.php?
page=climateOutlooks#May
2021climateOut2021.05.05 08.06.54.pdf)**

**Early Action Rainfall Watch (index.php?
page=climateOutlooks#May
2021ear2021.05.06 15.41.19.pdf)**

**ENSO Outlook (index.php?
page=climateOutlooks#March
2021ensoUp2021.03.26 16.44.14.pdf)**

[RMSC Wellington.\(Metservice\) SP bellow 25°S](#)

They generate a bulletin with the position and forecast of the TC track a few days ahead. They also and upload infrared satellite images of the last day.

```
Iframe('https://www.metservice.com/warnings/tropical-cyclone-activity',width=1000,height=450).
```

[Joint Typhoon Warning Center \(JTWC\)](#)

When there is an alert of TC they published it with downloadable information in text file about the position, maximum sustained winds (1-min), radius at 34/50/64kt and 12-hour and 24-hour forecasts. They also generate annual reports.

[GDACS \(Global Disaster Alert and Coordination System\)](#)

This is not an official center but it is a system developed to centralize warnings and estimates of the impacts of natural disasters (including TCs, using the bulletins issued by the RMSC, TCWC and JTWC). It can be interesting to have it there to consult, they have quite a few maps, it includes the alerts of the last year, but they seem to not be as updated as the other outlooks since the current alerts later.

[48-hr Tropical Cyclone Formation Probability Guidance Product](#)

The NOAA provides maps in the East and West Pacific with the current TC genesis probability and other interesting variables such as SST.

```
Iframe('https://www.ssd.noaa.gov/PS/TROP/TCFP/west_pacific.html',width=1000,height=450).
```



NOAA Satellite and Information Service
National Environmental Satellite, Data and Information Service (NESDIS)

Office of Satellite
Processing & Distrib

DOC / NOAA / NESDIS / OSPO / SPSD

Search OSPO Sites:

Home About US Product Listing Page Special Imagery Geostationary Polar Satellite Services Original SSD Links



Tropical Cyclone Formation Probability Guidance Product



Developed by the Regional and Mesoscale Meteorology Branch at CIRA

Cooperative Research Program | Office of Research and Applications/Center for Satellite Applications and Research

West Pacific Basin

- TCFP Home
- Product Description
- Basins
 - Atlantic

Cumulative 0-48h TC Formation Probability



By [Andrea Pozo](#)
© Copyright 2020.

