Open data and open source software for the development and validation of multi-model monthly-to-seasonal probabilistic forecasts for the Pacific Islands

Nicolas Fauchereau[1,1,1], Doug Ramsay[1,1], Ben Noll[1,1], and Andrew Lorrey[1,1]

[1]National Institute for Water and Atmospheric research (NIWA) Ltd

November 30, 2022

Abstract

In this paper, we leverage open data and open-source software to develop flexible, probabilistic monthly and seasonal (three-month) precipitation forecasts for the Pacific region. We use data from a Multi-Model Ensemble (MME), i.e. a large ensemble of state-of-the-art General Circulation Models (GCMs) and make use of recent developments in the Python open-source software ecosystem allowing the processing of large datasets on standard consumer grade laptops or desktop computers, of particular relevance in the Pacific context. The validation of the deterministic MME forecasts against reanalysis and observational products shows good performance, and confirms that an MME outperforms even the best single GCM. We show that the MME's forecast performance is modulated by the phases and characteristics of the El Nino Southern Oscillation (ENSO), with the longitude of the maximum Sea Surface Temperature anomalies playing a major role. We suggest that these findings could be used to provide additional confidence information along with the operational MME forecasts. Validation metrics for the probability of drought conditions, alternatively defined as seasonal rainfall accumulations below the climatological 1 tercile (percentile 33) or 1st quartile (percentile 25) show that the MME forecasts are reliable enough for most of the region. We provide an example of how this probabilistic forecast information can be integrated with real-time rainfall monitoring, in order to highlight areas in the tropical Pacific region which are at risk of water stress (i.e., where rainfall has recently been in deficit and forecasts indicate a high likelihood of dry conditions to persist or worsen).

Open data and open source software for the development and validation of multi-model monthly-to-seasonal probabilistic forecasts for the Pacific Islands

Nicolas Fauchereau[1], Doug Ramsay[2], Ben Noll[1] and Andrew Lorrey[1]

1. National Institute for Water and Atmospheric research (NIWA) Ltd. 41 Market Place, Auckland 1010. New Zealand

2. National Institute for Water and Atmospheric research (NIWA) Ltd. Gate 10, Silverdale Road, Hamilton 3216. New Zealand

Submitted to climate services

Keywords

Seasonal climate forecasting, drought, Pacific, rainfall, ENSO, Multi-Model Ensemble

Abstract

In this paper, we leverage open data and open-source software to develop flexible, probabilistic monthly and seasonal (three-month) precipitation forecasts for the Pacific region. We use data from a Multi-Model Ensemble (MME), i.e. a large ensemble of state-of-the-art General Circulation Models (GCMs) and make use of recent developments in the Python open-source software ecosystem allowing the processing of large datasets on standard consumer grade laptops or desktop computers, of particular relevance in the Pacific context.

The validation of the deterministic MME forecasts against reanalysis and observational products shows good performance, and confirms that an MME outperforms even the best single GCM. We show that the MME's forecast performance is modulated by the phases and characteristics of the El Nino Southern Oscillation (ENSO), with the longitude of the maximum Sea Surface Temperature anomalies playing a major role. We suggest that these findings could be used to provide additional confidence information along with the operational MME forecasts.

Validation metrics for the probability of drought conditions, alternatively defined as seasonal rainfall accumulations below the climatological 1 tercile (percentile 33) or 1st quartile (percentile 25) show that the MME forecasts are reliable enough for most of the region. We provide an example of how this probabilistic forecast information can be integrated with real-time rainfall monitoring, in order to highlight areas in the tropical Pacific region which are at risk of water stress (i.e., where rainfall has recently been in deficit and forecasts indicate a high likelihood of dry conditions to persist or worsen).

Practical Implications

Tracking and predicting the development of meteorological drought conditions is of paramount importance in the Pacific region, where hydroclimate variability is large, in-situ (station) data is often lacking, and national capacity in seasonal forecasting is limited. In this paper we present data, methods (including software implementation) pertaining to the development of a range of freely available products that utilizes near-realtime monitoring of drought conditions using satellite remote sensing and probabilistic forecasts from a large ensemble of operational forecast General Circulation Models (GCMs), as part of the Island Climate Update (ICU) a climate monitoring and outlook bulletin for Pacific Island nations and regional support agencies. In particular, we provide a flexible product that combines a percentile-based drought index and the probabilistic information for future drought conditions from the Multi-Model Ensemble (MME) to highlight areas at potential for 'water stress', i.e. where drought conditions have been observed and where the forecast information indicates a high likelihood for rainfall deficits to persist or worsen. The full range of the ICU near-realtime and forecast products are operationalized and publicly available and aim to provide the Pacific region with an early alert on island groups which are at risk of developing water stress, allowing resources and assistance to be mobilized and directed ahead of time.

Introduction

Pacific Island countries (PICs) are impacted by large rainfall variability, arising primarily from variations in the position and intensity of the South Pacific Convergence Zone (SPCZ, Vincent 1994, Widlansky et al, 2011, Brown et al, 2020) and the Intertropical Pacific Convergence Zone (ITCZ, Schneider et al. 2014). Extreme phases of El Niño-Southern Oscillation (ENSO, see e.g. Neelin et al, 1998) can lead to multi-year drought. Generally, islands close to the Equator and east of the International Dateline experience dry conditions during La Niña phases of ENSO, while many countries west of the Dateline experience lower rainfall during El Niño (Cottrill et al. 2013). A majority of Pacific Islanders, particularly in rural areas and outer-islands, rely on subsistence agriculture (Geogeou et al, 2022), and can be subject to food security risks arising from a range of weather and climate-related extremes. These extremes also impact water security with reliance on rainwater harvesting and shallow groundwater lenses commonplace on low-lying islands and atolls that are subject to water quality and water shortage issues during prolonged drought or deluge episodes (Iese et al, 2021).

Precipitation variability associated with ENSO is projected to increase in the Pacific in response to climate change (Power and Delage, 2018, Yun et al, 2021), which will further threaten water and food security in the region. As such, better climate forecasts (i.e., one month to season ahead) are becoming increasingly recognized as an important component of successful climate change adaptation strategies. This has been notably the impetus behind the establishment of the World Meteorological Organisation's (WMO) *Global Framework for Climate Services* (Hewitt et al, 2012), and the Pacific Islands Climate Services (PICS) panel, a regional advisory group to the Pacific Meteorological Council (PMC), whose objective is to strengthen the capacity of National Meteorological and Hydrological Services (NMHSs) in observing and understanding weather and climate and in providing related services in support of national needs (WMO, see: https://public.wmo.int/en/our-mandate/how-we-do-it/role-and-operation-of-nmhss). It has also recently motivated the establishment of the WMO Regional Association V Pacific Regional Climate Centre (RCC) Network, a virtual Centre of Excellence that assists National Meteorological and Hydrological Services (NMHSs) in the Pacific Islands region to deliver better climate services and products and to strengthen their capacity to meet national climate information and service delivery needs (See https://www.pacificmet.net/rcc).

Both *statistical* and *dynamical* approaches can be used to produce monthly to seasonal climate forecasts. Statistical approaches harness empirical relationships between target variables (such as time-series of monthly or seasonal precipitation accumulations) and indices representative of known climate modes such as ENSO, while dynamical approaches use initialised General Circulation Models (GCMs, see Meehl et al, 2021). Notably, coupled ocean – atmosphere GCMs provide physically consistent fields of atmospheric and surface climate variables, typically up to six months into the future and aggregated at the monthly time-scale (e.g. average monthly temperature or precipitation rates).

One weakness of a statistical approach is the underlying assumption of stationarity, which is likely to not hold in a rapidly warming climate. The WMO's *Guidance on Operational Practices for Objective Seasonal Forecasting* (WMO, 2020) therefore recommends that regional or national outlooks be based on dynamical approaches. It further indicates that large ensembles of dynamical climate forecasts from different GCMs (Multi-Model Ensembles or MMEs) tend to perform better than a single GCM. An MME forecasting approach can help to better account for the uncertainties that can arise from the initial conditions, the absence of strong climate drivers and the differences between GCMs formulations. Moreover, MME forecasting easily allows forecasts to be expressed in probabilistic terms, which can help communicate uncertainties and be readily translated and communicated in terms of risks.

Several meteorological institutions provide global monthly and seasonal, probabilistic forecasts, typically of tercile categories, i.e. the probabilities for monthly or seasonal aggregated statistics to be below, above or between percentile 33.3 and 66.6: Graphical examples of which can be found at https://climate.copernicus.eu/charts/c3s_seasonal/.

The developments and products presented in this paper are the culmination of a process started in 2000, when NIWA started to provide Pacific Island Countries with the "Island Climate Update" (ICU), a climate bulletin and outlook product suite which adopted a multimodel ensemble approach as early as 2008. This initially developed and utilized a semi-objective ensemble method through the development of the Multimodel Ensemble Tool for Pacific Islands (Lorrey et al, 2009; McGree and Baleisolomone, 2009). This initial effort has grown from ensemble outlooks based on a limited number of rainfall and SST models to now drawing on a much larger model pool and use of more objective methods to create spatially-scaled and seasonally-tuned forecasts for Pacific nations.

The goal of this study is to illustrate how one can help unlock the full potential of MME forecasts by integrating them with other sources of climate or environmental information and allowing for the development of more useful and actionable climate services (i.e., when they are one component in a wider system, be it combining with real-time climate monitoring systems, or inputs to downstream models, such as hydrological, crop or disease models, etc.). This requires the GCM forecast (realtime) and hindcast (retrospective forecasts) data to be openly and freely accessible, allowing the derivation of statistics and diagnostic variables tailored to the system into which monthly and seasonal climate forecasts are integrated.

In this paper, we utilize near real-time satellite precipitation estimates and monthly and seasonal precipitation forecasts from state-of-the-art GCMs to derive an operational system designed to highlight areas at potential for "water stress" in the Pacific *(i.e., where rainfall has recently been in deficit and forecasts indicate a high likelihood of dry conditions to persist or worsen).* The development of this product was initiated in response to feedback from NMHSs and regional institutions who desire the ability to track the development of drought conditions in the region with a focus on placing the forecasts in the context of current hydroclimate anomalies, and to provide improved representation of the confidence of the forecasts.

In order to allow for reproducibility, extensibility and foster the development of further climate service products in the region, we developed a software library, written in Python (van Rossum, 2001) to handle all steps of the data processing and visualisation pipeline, as well as a set of commented, example Jupyter notebooks (Shen, 2014), which will be briefly described in the present paper.

The first section presents the data, the methodological choices made in developing the various near-real time and probabilistic forecast products, as well as a brief overview of the supporting software infrastructure.

The second part of the paper is devoted to the validation of the precipitation forecasts from the MME system, and include some analyses shedding light on the variability in the forecast performance in space and time and its drivers.

The third part provides an example of a regional product based on the above development and designed to communicate the monthly and seasonal probabilistic forecasts in the context of the antecedent conditions from the real time precipitation estimates and to potential trajectories of water stress categories across the tropical Pacific.

1. Material and methods

1.1 Data

1.1.1. Near- real time gridded precipitation estimates: The GPM-IMERG dataset

For monitoring the evolution of various precipitation statistics over different accumulation periods in near –real time, we use the Integrated Multi-satellitE Retrievals for GPM (Global Precipitation Measurement, hereafter GPM-IMERG, see Huffman et al 2014). The GPM-IMERG algorithm combines information from the GPM constellation satellites to estimate precipitation over the majority of the Earth's surface. This product is particularly valuable over the Pacific region, where real time in-situ (surface station) information is sparse and data quality issues are common. We use specifically the Level 3, version 6, daily near- real time product, available at https://gpm1.gesdisc.eosdis.nasa.gov/data/GPM_L3/GPM_3IMERGDL.06/, with two days latency to real time. While there are known biases and deficiencies in the GPM-IMERG product, these are mostly present for daily precipitation and especially extreme precipitation (Silva et al, 2021), it is therefore assumed that these biases are of relatively low impact in the present study, as we use precipitation accumulations (from 30 to 360 days) and are only considering *relative* quantities (such as percentiles of scores) instead of absolute amounts in millimeters.

1.1.2. Monthly and seasonal GCMs forecasts

The monthly and seasonal forecasts and *hindcasts* (retrospective forecasts, also named *reforecasts*) data are sourced from the Copernicus Climate Data Store (CDS), established under the auspices of the Copernicus Climate Change Service (C3S). The CDS collects hindcast and forecast data generated by eight international institutions, namely the European Centre for Medium-Range Weather Forecasts (ECMWF), the United Kingdom Meteorological Office (UKMO, UK), Météo-France (the French Meteorological agency), The Deutscher Wetterdienst (DWD, Germany), the Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC, Italy), the National Centers for Environmental Prediction (NCEP, USA), the Japan Meteorological Agency (JMA, Japan) and Environment and Climate Change Canada (ECCC, Canada). Together, we will refer to the MME constituted from these GCMs as the *C3S MME*. Currently, complete hindcast (1993-2016) datasets for seven GCMs, and forecasts (2017 – present) datasets for 9 GCMs are available.

Some characteristics of the GCMs are summarized in **Table 1**, and more details can be found within the CDS documentation, at URL https://confluence.ecmwf.int/display/CKB/C3S+Seasonal+Forecasts. All GCMs are state of the art, operational, and coupled ocean – atmosphere models.

These data are accessible via an API (Application Programmer Interface, see https://github.com/ecmwf/cdsapi) which allows the user to select variables of interest, initial month, lead-time (in months), and download the resulting files in grib or netcdf, data formats both widely used in the meteorological and climate communities.

Table 1: Some characteristics of the GCMs constituting the C3S MME (more details can be found at https://confluence.ecmwf.int/display/CKB/C3S+Seasonal+Forecasts).

Originating institution	**Forecast system**	**Hindcast ensemble size**	**Forecast ensemble size (as of 15 March 2022)**	**Hindcast complete**
ECMWF	SEAS5	25	51	Yes
UKMO	GloSea6-GC3.2	28	56	Yes
Météo-France	Météo-France System 8	25	51	Yes
DWD	GCFS 2.1	30	50	Yes
CMCC	CMCC-SPS3.5	40	50	Yes
NCEP	CFSv2	20	112	Yes
JMA	JMA/MRI-CPS3	10	140	Yes
ECCC	CanCM4i	10	10	No
ECCC	GEM5-NEMO	10	10	no

The total number of members in the MME for the seven GCMs where the hindcast data is complete (ECMWF, UKMO, Météo-France, CMCC, DWD, NCEP and JMA) is therefore 198, while at the time of writing, operationally the forecast MME comprises 563 members from nine GCMs.

1.1.3. Observational datasets for forecast validation

Due to the GPM-IMERG data starting in 2001, we used the following products to validate the deterministic and probabilistic forecasts from the individual GCMs and the C3S MME: the ERA5 (Hersbach et al., 2020) monthly precipitation (taken from the CDS), the CMAP (CPC Merged Analysis of Precipitation, Xie and Arkin, 1997) and the MSWEP 2.0 dataset (Beck et al, 2019). We present only the results using ERA5 as the conclusions regarding the performance of the individual GCMs and the MME are not dependent on the validation dataset. Note that we make available Jupyter notebooks at (see https://zenodo.org/record/6658577) allowing one to easily select an alternative validation dataset.

1.2. Methods

1.2.1. Quantile-based climatologies

The system developed relies on the calculation of several climatological quantities, notably quantiles, from time-series of satellite precipitation estimates and monthly and seasonal hindcast data.

For the GPM-IMERG satellite estimates, the *percentile of scores* for given accumulation periods (currently 30, 60, 90, 180 and 360 days) are calculated compared to the archived dataset over the period 2001 - 2020 (20 years). More specifically, the latest rainfall accumulation is compared to the corresponding accumulations ending on the target day of year, + / - buffer of three days, so that a 90-day accumulation ending on the 30 September 2021 is compared to the 90-day accumulation ending 27, 28, 29, 30 September as well as 1, 2, 3 October, for each year from 2001 to 2020 (*i.e.,* a total of 7 x 20 = 140 values).

These percentiles of scores are then used as the basis for deriving percentile-based drought monitoring indices such as the "Early Action Rainfall" (EAR) Watch categories developed by the Climate and Oceans Support Programme in the Pacific (COSPPac - http://cosppac.bom.gov.au/) and the US Drought Monitor levels used in the US-Affiliated Pacific Islands (see Heim et al, 2020). In addition, the Standardized Precipitation Index (SPI) is calculated, following the methodology described in Lloyd-Hughes and Saunders (2002). These three indices are widely used by PICs NMHSs to monitor drought conditions, using data from their surface station monitoring network where available, and by regional support and development agencies. The satellite-derived drought indices make it possible to compare and contextualise local to regional drought evolution and provide information for areas where real-time in-situ data is lacking or of poor quality.

The calculation of these indices is done operationally every day, with the corresponding graphical and data products, collectively called the NIWA Island Climate Update (ICU), made available on Amazon Web Services (AWS), respectively at URLs https://s3.ap-southeast-2.amazonaws.com/icu.niwa/gpm/images/images.html and https://s3.ap-southeast-2.amazonaws.com/icu.niwa/gpm/netcdf/netcdf.html

Besides theses indices, the rainfall accumulation and anomalies (in mm) as well as the number of dry days, and number of days since last rain, are also calculated and made available at the URLs above.

Figure 1 presents an example of the maps generated at the regional (Pacific) scale.

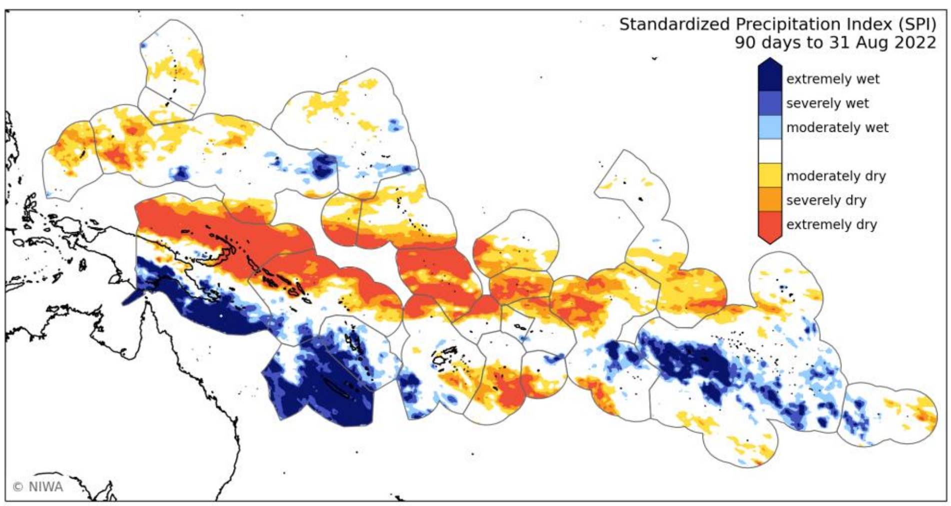

Figure 1: *Standardized Precipitation Index (SPI) over Pacific Island Exclusive Economic Zones (EEZs) calculated for the 90 days precipitation accumulation ending 31 August 2022, according to Lloyd-Hughes and Saunders (2002).*

For the C3S GCMs, we derive lead-time dependent monthly and seasonal (three month accumulation) climatological values from all the corresponding available hindcast data, spanning 1993 to 2016.

The tercile (percentiles 33.3 and 66.6), quartile (25, 50, 75) and decile (10, 20, …, 90) climatologies are calculated using all members of the hindcast's GCM's ensemble: e.g. for the ECMWF GCM, the climatological quantiles are calculated from the 1993 – 2016 hindcast dataset, which gives a total of 600 instances (24 years x 25 ensemble members) for each initial month.

We calculate both the empirical climatological quantiles, as well as parametrized quantiles, whereby a Gamma distribution is first fitted (using the *L*-moments method, Hosking, 1990) to the monthly or seasonal accumulations, the final validations were qualitatively unaffected by the choice of the method, and comparisons with the tercile probabilities displayed on https://climate.copernicus.eu/seasonal-forecasts led us to choose the empirically-derived quantiles as the basis for the derivation of the probabilistic forecasts.

In Section 2, we provide validation information for deterministic forecasts and probabilistic forecasts, Noting the current ICU products are based primarily on the probabilistic information.

Deterministic forecasts are defined as the average of the precipitation anomalies across each member of each GCM ensemble, calculated with respect to the GCM lead-time dependent ensemble mean climatology derived from the whole hindcast period (1993-2016). The deterministic MME forecast is simply calculated as the average of these anomalies across the GCMs.

Probabilistic forecasts for each GCM are calculated as the proportion of ensemble members (see **Table 1**) falling into each quantile category. The MME probabilities are then calculated as the average of the individual GCMs' probabilities, expressed in percentage and summing to 100%.

1.2.2. Forecast verification metrics

To quantify the performance of the deterministic forecasts (average of all GCM anomalies) we use the Anomaly Correlation Coefficient (ACC): The ACC is dimensionless, varies between -1 and 1, and is simply calculated as the correlation between the spatial patterns of forecast and the observed precipitation anomalies, it is therefore a useful measure of the ability of the GCMs to broadly reproduce the spatial distribution of rainfall, and therefore the regional scale hydroclimate.

For the probabilistic forecasts, we use the overall accuracy (or "hit rate") first, then focus on the lower categories for both tercile and quartile probabilistic forecasts, *i.e.,* the forecast probabilities for rainfall being below the 1st tercile (< 33rd percentile) and below the 1st quartile (< 25th percentile), respectively, given the significance of dry conditions in the region for water security.

Verification metrics of particular interest when focusing on one categorical forecast are the *precision* and *recall*.

The *precision* is the number of True Positives (TPs) divided by the number of TPs and False Positives (FPs): in other words, it is the number of "positive" predictions (i.e., when the lower quantile category is the most likely) divided by the total number of "positive" class values predicted. It is also called the Positive Predictive Value (PPV). In the context of this study, it answers the question: out of the months or seasons that were predicted to be in the 'dry' category, how many turned out to be actually dry ? The precision is therefore an informative measure when the costs of a False Positive (predicting dry conditions that fail to materialize) is high.

The *recall* is the number of TPs divided by the number of TPs and False Negatives (FNs). In other words it is the number of "positive" predictions divided by the number of "positive" class values in the observational data. It is also called *Sensitivity* or the *True Positive Rate.* In the context of this study, it indicates what proportion of the months or seasons when rainfall fell in the lower quantile category were correctly predicted by the forecast system. It is therefore a useful measure to determine when the cost of a False Negative (failing to predict dry conditions) is high.

Both precision and recall vary between 0 and 1, with 1 indicating perfect forecasts for the category in question.

The *F1 score* is given as reference, it is calculated from the precision and recall and is a synthetic measure of a categorical forecast's performance. It is calculated as 2 x ((precision x recall) / (precision + recall)), and also varies between 0 and 1.

1.2.3. Calculation of sub-regional time-series

Sub-regional (based on administrative areas or sub-national island groupings , see **Figure 2**) probabilistic forecasts and their validation required the calculation of area-averaged time-series of precipitation from GCMs' hindcasts and forecasts as well as the gridded validation datasets.

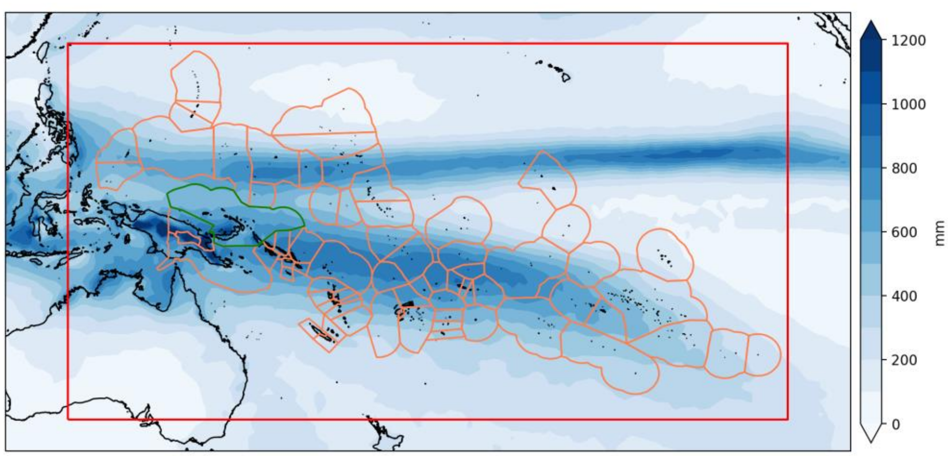

Figure 2: *Domain for the calculation of the ACC and RMSE (red, boundaries are [35°S – 25°N, 125°E - 120°W]), and location of the 73 administrative areas for which regional tercile and quartile probabilistic forecasts are provided (orange), in green is highlighted the 'Islands' administrative area of Papua New Guinea, used to illustrate the derivation of land-sea masks for the calculation of regional time-series (see* ***Figure 3****). The average December – February cumulative precipitation amounts from ERA5 (1993 – 2016) is shown in blue filled contours and displays the typical positions of the ITCZ and SPCZ at this time of year.*

Given the small land area of many Pacific Islands and atolls, the GCM (and all gridded validation datasets such as the ERA5 reanalysis) outputs are first interpolated to 0.2 degree (i.e., five times the typical original resolution for the GCMs). We use shapefiles that delineate the exclusive economic zone boundaries, administrative areas and island coastlines to derive land / sea masks for each of the 73 territorial areas. In order to account for islands and atolls with small land area, we further apply a buffer (0.15°) around the coastlines prior to the mask definition. Given the original resolution of the GCM outputs, results are very much insensitive to the exact extant of the buffer. **Figure 3** illustrates this process for the "Islands region" of Papua New Guinea.

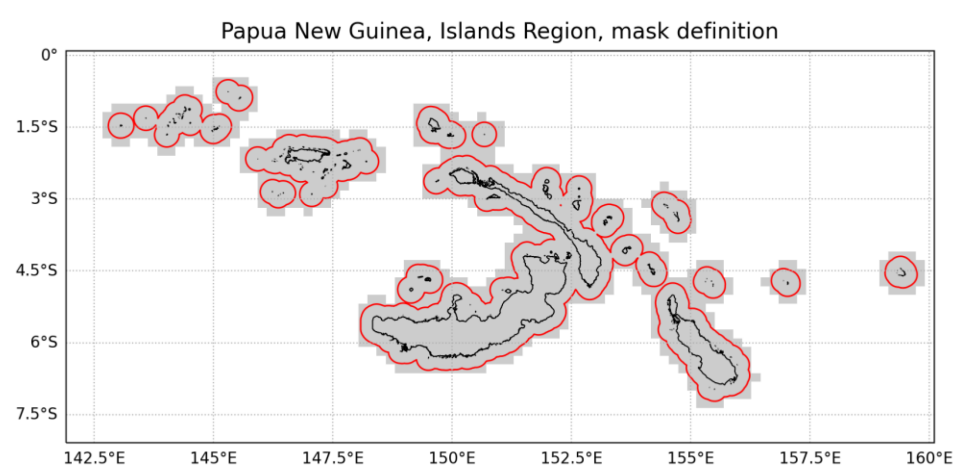

Figure 3: *Example illustrating the derivation of land / sea masks for the Pacific Island countries administrative areas: The black line corresponds to the original coastlines for the 'Islands' region of Papua New Guinea, the red line corresponds to the 0.15º buffer, the gray shading corresponds to the resulting land-sea mask used to derive regional precipitation time-series from the interpolated gridded datasets (GCM hindcasts and forecasts and validation datasets).*

1.2.4. Combining real time rainfall monitoring and monthly to seasonal climate forecasts

The system presented in this paper has been developed in order to ultimately combine real time rainfall monitoring and monthly or seasonal probabilistic rainfall forecasts to alert national and regional institutions around the Pacific of regions that are at potential risk of 'water stress': conceptually, one wants to highlight regions where significant rainfall deficits occurred recently, and at the same time the probabilistic forecasts indicate a high likelihood for dry conditions to persist or worsen. After feedback from potential end-users and several iterations, we established three categories, with criteria based on the most recent 90 days rainfall accumulation percentile of score, and the forecast probability for rainfall being below or above the 25^{th} percentile (1^{st} quartile) for either the next month or the next 3 months accumulation as a whole. **Table 2** presents the detailed criteria use to define these categories; an example will be provided in the results section.

ICU "Water Watch" category	Present situation	Next month outlook	Next 3 months outlook
1) Current "water stress" conditions, potentially easing	Past 90 days rainfall accumulation < 25^{th} percentile	50% chance or more for the next month rainfall accumulation >= 25^{th} percentile	50% chance or more for the next 3 months rainfall accumulation >= 25^{th} percentile
2) Areas moving into "water stress" conditions	Past 90 days rainfall accumulation > 25^{th} percentile and < 40^{th} percentile	50% chance or more for the next month rainfall accumulation < 25^{th} percentile	50% chance or more for the next 3 months rainfall accumulation < 25^{th} percentile
3) Current "Water Stress" conditions getting worse	Past 90 days rainfall accumulation < 25^{th} percentile	50% chance or more for the next month rainfall accumulation < 25^{th} percentile	50% chance or more for the next 3 months rainfall accumulation < 25^{th} percentile

1.2.5. Software implementation

The software infrastructure for downloading and processing the data, the calculation of various quantities and their graphical representations, as well as all the code necessary to reproduce the results and figures presented in this paper are made available freely (https://zenodo.org/record/6658577). It has been developed using the open-source language Python (Van Rossum, 2001), and relies heavily on the Scientific Python Ecosystem (Virtanen et al, 2020) and in particular the foundational libraries of the Pangeo initiative (Hoyer and Hamman, 2017; Abernathey et al, 2017; Brady and Spring, 2021).

The processing of a large amount of data was facilitated by the underlying *dask* library (Rocklin, 2015). It makes it possible to run all steps of the data processing and analysis pipeline on small-scale hardware such as a laptop, even though the complete archive for the C3S MME hindcast datasets (for a surface variable such as the precipitation rate of interest here) exceeds 16 GB, (i.e., too large to fit in memory on typical laptops hardware).

2. Results

2.1. Validation of deterministic forecasts

We first present validation results for the deterministic C3S MME forecasts, calculated as the average of precipitation anomalies across the seven GCMs for which all initial months are available over the 1993 – 2016 hindcast period.

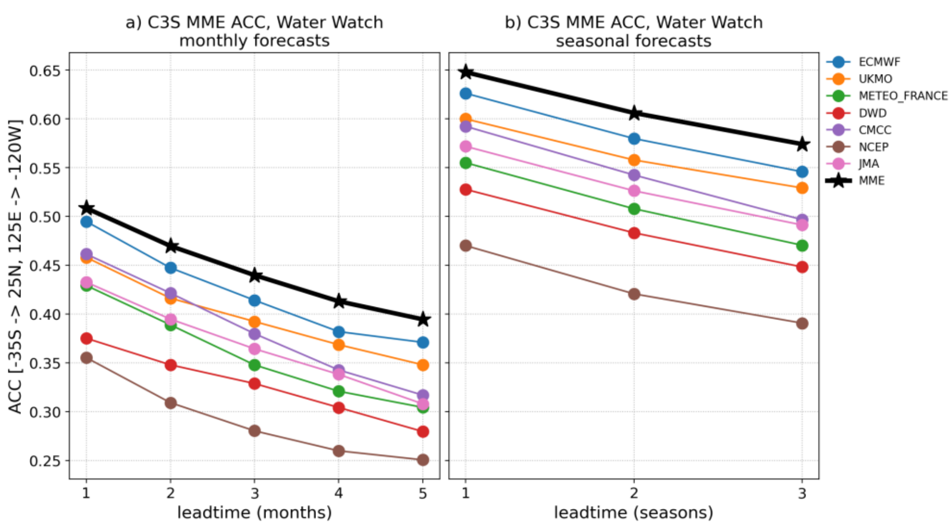

Figure 4*: Anomaly Correlation Coefficient (ACC) between the individual GCMs precipitation reforecasts (1993 – 2016) and ERA5 precipitation for a) monthly accumulations b) seasonal (3 months) accumulations anomalies, over the domain [35°S – 25°N, 125°E – 120°W, see* ***Figure 2****]. The leadtime (x-axis) is given in months (seasons) from the initial month, so that e.g. leadtime 1 for monthly (seasonal) forecasts initialized in January corresponds to February (February – April) accumulations.*

Figure 4 confirms that overall, the C3S MME performs better than even the 'best' GCM (ECMWF in this instance). This is in line with the WMO (2020) conclusions: The average of forecast inputs (the multi-model ensemble approach) is statistically a better predictor of observed climate than a single model's forecast alone and makes combining different climate model predictions advantageous and an advisable approach (See SPECS (2016) for a review on this topic). The ACC for seasonal (three month accumulation) is also significantly larger than for monthly accumulations, and, as expected, the performance degrades as the lead-time increases.

As the next season (three month) period is generally the focus of PICs NMHSs national outlook bulletins, we will mainly focus on this time scale and lead-time in the rest of this paper, however the code allows replication of the following figures for the monthly time-scale and for other lead-times.

As expected, the performance of the individual GCMs and the C3S MME is significantly seasonally dependent: **Figure 5** shows the one season lead-time ACC as a function of the month of the initialization

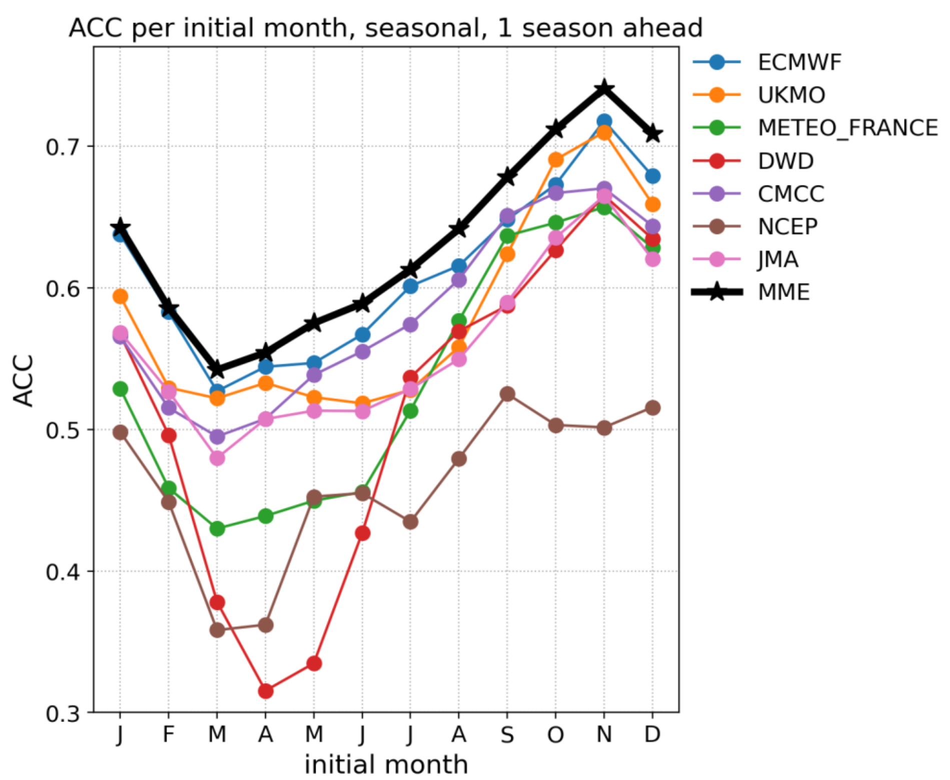

Figure 5: *ACC over the domain [35°S – 25°N, 125°E - 120°W] for one season ahead forecasts, as a function of the initial month (i.e. January initial month corresponds to FMA forecasts and so on)*

Generally speaking, seasonal forecasts for December-February (DJF, initialized in November) have the highest ACC. The lowest ACC is found for forecasts initialized in March (i.e. for the AMJ forecast period). Three GCMs are characterized by low ACC during the Austral autumn period: Météo-France, NCEP and DWD. Removing these three GCMs from the MME however only leads to very marginal improvement on the overall MME's ACC during this period. The maximum difference found is for forecasts initialized in April, where the ACC for the "reduced" MME (4 instead of 7 GCMs) is 0.58 instead of 0.55. On the other hand, the removal of these GCMs from the MME tends to slightly *decrease* the MME's ACC during Austral summer. The same seasonal patterns and conclusions regarding the removal of Météo-France, NCEP and DWD from the MME also hold for longer lead-times.

The ACC variability for seasonal values at one season lead-time is shown in **Figure 6**. There is a considerable amount of variability in the ability of the individual GCMs – and the MME – to reproduce the observed overall pattern of rainfall anomalies over the Pacific domain.

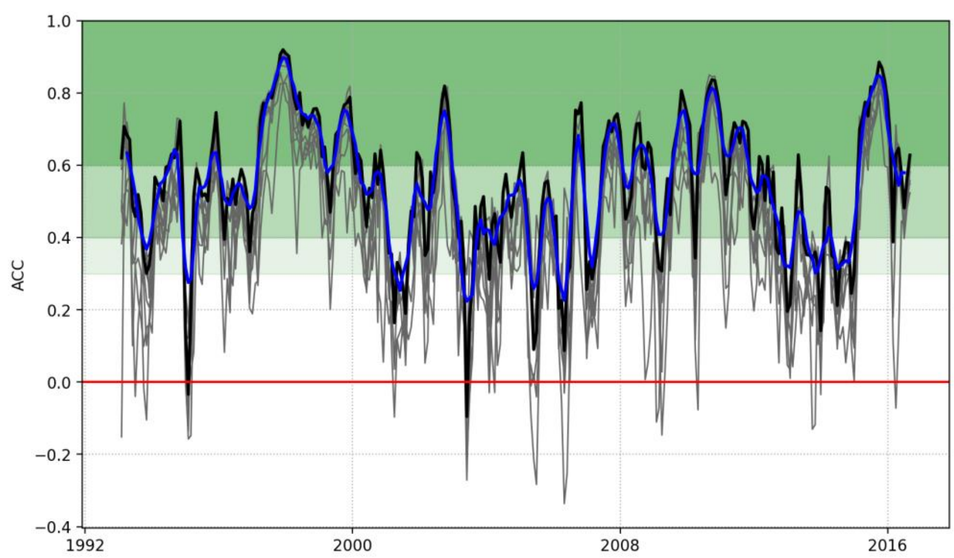

Figure 6*: ACC variability for each target season from March – May 1993 to October – December 2016. Gray lines: individual GCMs, black line: Multi-Model Ensemble, blue line: Centered, 5 points running average of the MME's ACC.*

At one season lead-time, the ACC for the C3S MME exceeds 0.6 40% of the time, and exceeds 0.4 75% of the time, but about 3% of the seasons are associated with ACC <= 0.2.

2.2 Role of El Nino Southern Oscillation

Given the important role of ENSO in controlling the intensity and position of the Pacific Convergence Zones (Widlansky et al, 2011), it can be assumed that the variability in the ability of the GCMs (and MME) to forecast the patterns of precipitation anomalies over the Pacific region is at least partially dependent upon the phase and characteristics of ENSO when the GCMs are initialized. We chose to investigate this potential dependency using three widely used SST (Sea Surface Temperature) ENSO indices (Trenberth and Stepaniak, 2001) : The Niño 3.4 index (190^{o} to 240^{o}E, 5^{o}S to 5^{o}N) representative of the standard 'canonical' ENSO events, The Trans-Niño Index (TNI), calculated as the difference between the Nino1+2 index (270^{o} to 280^{o}E, 10^{o}S to Equator) and the Nino4 index (160^{o}E to 210^{o}E), 5^{o}S to 5^{o}N) and representative of the east – west gradient in SST anomalies, as well as the El Niño "Modoki" index (EMI) used to capture the El Niño Modoki phenomenon (also sometimes referred to as "Central Pacific El Niño") whereby the maximum SST anomalies is located towards the central rather than the eastern Pacific (Ashok et al 2007), the EMI being calculated as:

EMI = $[SSTA]_A - 0.5*[SSTA]_B - 0.5*[SSTA]_C$ (1)

The brackets in equation (1) represent the area-averaged SSTA over each of the region A (165°E–140°W, 10°S–10°N), B (110°W–70°W, 15°S–5°N), and C (125°E–145°E, 10°S–20°N), respectively.

All indices are calculated using the detrended monthly SST anomalies from the ERSST version 5 dataset (Huang et al. 2017). We use a 1993 – 2016 climatology to be consistent with the leadtime-dependent climatologies calculated from the C3S GCMs.

We then use a threshold of +/- 1 standard deviation to define positive (> +1 std), negative (< - 1 std) and neutral phases (>= -1 std and <= +1 std) for each of the above indices.

For reference, the **Figure 7** shows the Tropical (25°S – 25°N) Pacific-wide SST anomalies associated with each index and phase.

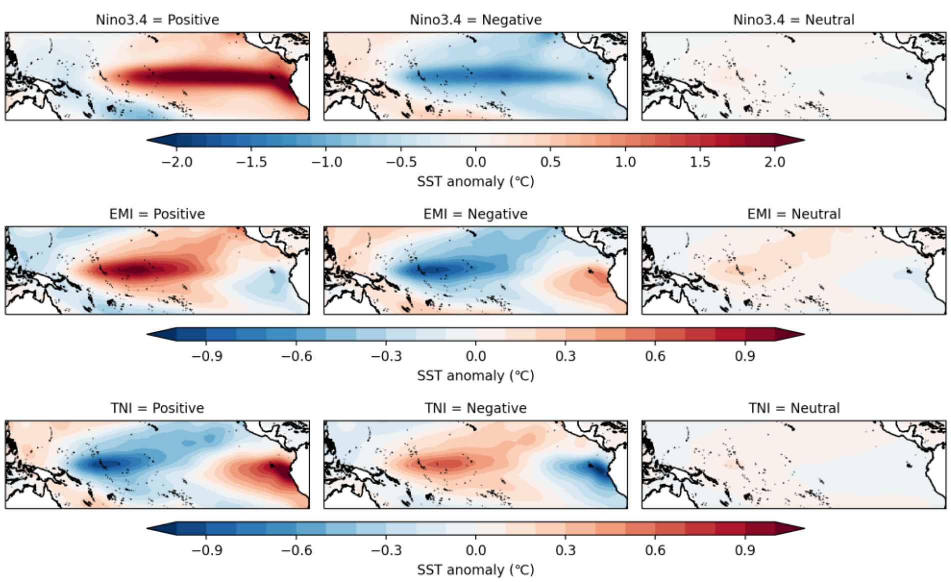

Figure 7: *SST (ERSSTv5) anomalies for the different phases of the Nino3.4, EMI and TNI indices. The detrended anomalies have been calculated with respect to a 1993-2016 climatology. The same threshold of +/- 1 standard deviation has been used to define the positive (> +1 std), negative (< - 1 std) and neutral phases (>= -1 std and <= +1 std) phases for each index.*

As expected, the most prominent difference between 'canonical' ENSO phases and Modoki phases is the location of the maximum SST anomalies along the equator, with canonical positive ENSO phases characterised by maximum positive SST anomalies located east of the International Dateline, towards the South American coast, and 'Modoki' positive phases characterised by maximum SST anomalies located around the International Dateline and negative SST anomalies off the South American coast.

The ACC is then calculated for all GCMs, and the C3S MME for seasonal forecasts initialized during the different phases of each index (**Figure 8**).

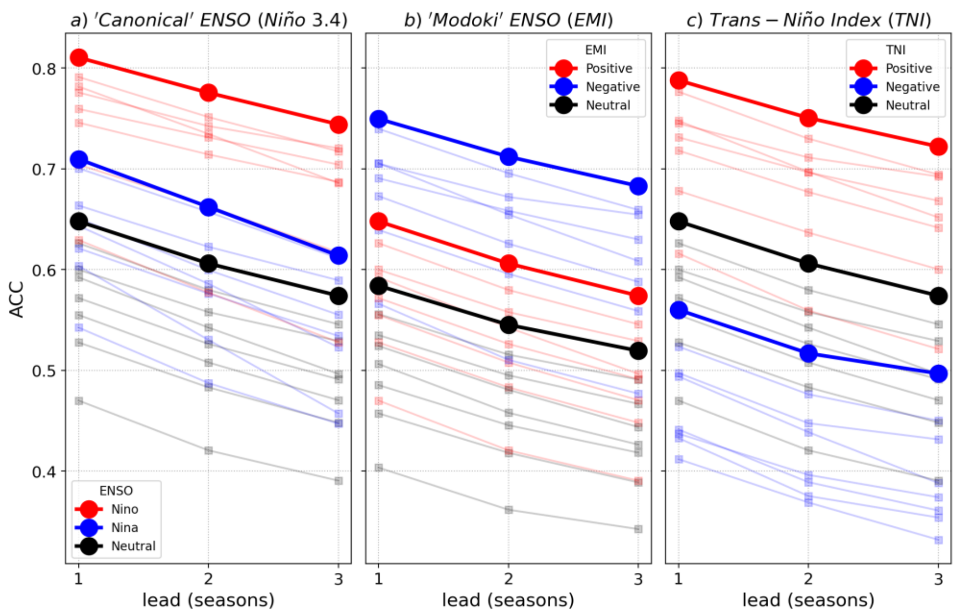

Figure 8: *ACC for seasonal forecasts as a function of leadtime during the different phases of a) the 'Canonical' ENSO mode (as characterized by the Niño 3.4 index) b) the El Niño 'Modoki' index (as characterized by the El Niño Modoki Index) and c) the Trans-Niño Index (TNI). The bold line corresponds to the MME, and the light lines to each individual GCM.*

As expected, there are large differences in the performance of the GCMs and the MME (as measured by the ACC) depending on the phase, and the characteristics of ENSO conditions at initialisation.

The ACC is generally higher during the positive phases of Niño3.4 and TNI, but negatives phase of the EMI, the commonality therefore being the presence of large positive SST anomalies in the far eastern Pacific and the establishment of a strong west-to-east gradient in anomalies.

Conversely, the ACC tends to be lower during ENSO phases and flavors characterized by an inverse gradient in SST anomalies, such as during the negative phases of the TNI.

The patterns displayed for the MME (**Figure 8**) hold true for all individual GCMs: Meaning that for all GCMs at all lead-times (with one exception, see below), the ACC during positive phases of the TNI is larger than for neutral phases, which is in turn larger than for negative phases). The only exception is for Météo-France and for Nino3.4, where at lead 3 (three seasons ahead) the ACC for neutral ENSO phases is 0.47, and 0.46 for negative phases.

These results therefore suggest that the sign and amplitude of SST anomalies specifically in the eastern Pacific plays a major role in determining the skill (as measured by the ACC) of the GCM forecasts: Positive SST anomalies (i.e., during either the positive phase of 'canonical' ENSO events, or the negative phase of the 'Modoki' ENSO events), tend to lead to enhanced skill, while negative SST anomalies (i.e., during either the negative phase of canonical ENSO or the positive phase of modoki ENSO) lead to reduced skill, in comparison to neutral phases of both ENSO flavours.

This result is of operational significance, as this information can be used to convey the level of confidence in the seasonal forecast information in real-time, by monitoring the SST anomaly patterns in the Tropical Pacific.

2.3. Validation of probabilistic forecasts

The MME probabilities for tercile and quartile categories are calculated as the average of the individual GCMs' probabilities. We first present the overall accuracy score (or 'hit-rate'), then focus on the performance of the forecasts for the lower quantile categories, i.e., respectively the lower tercile (probability for rainfall being below the 33rd percentile) and lower quartile (probability for rainfall being below the 25th percentile). This is because an accurate prediction of drought conditions is of primary interest for the region.

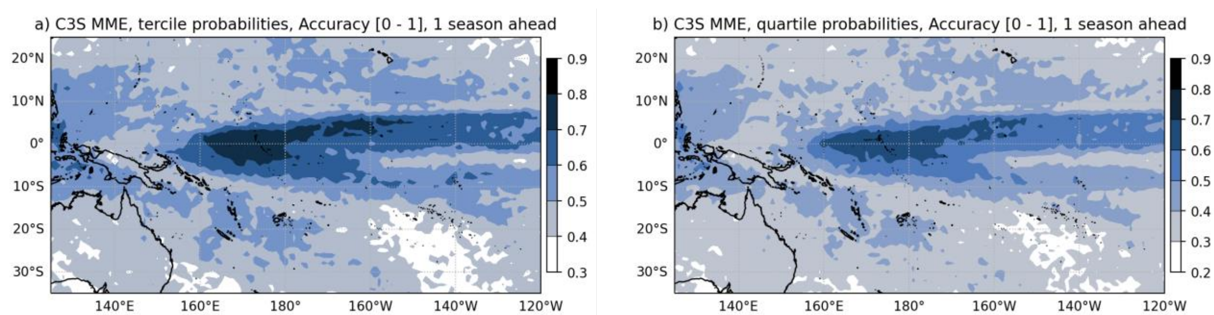

Figure 9: *Accuracy (or 'hit rate') of the MME seasonal tercile (a) and quartile (b) probabilistic forecasts (one season ahead) against the seasonal tercile and quartile categories derived from ERA5. All calculations were performed over the same 1993 – 2016 period.*

Figure 9 presents the respective accuracy for tercile (**Figure 9a**) and quartile (**Figure 9b**) most likely category from the MME, one season ahead. Note that a climatological forecast would result in an accuracy of 0.33 (33%) and 0.25 (25%) respectively for the tercile and quartile forecasts.

The C3S MME is therefore more skillful than a climatological forecast for most of the region, with the notable exception being the southeast Pacific (South of French Polynesia). More precisely, 91% of the grid-points are associated with an accuracy score exceeding 40% for the MME terciles probabilistic forecasts, and 88% of grid points are associated with an accuracy score exceeding 30% for the quartile probabilistic forecasts.

High skill is found in the tropical region between 10S and 10N, and east of ~ 160E as well as for southern parts of Papua New Guinea, the Solomon Islands, Vanuatu and Fiji.

The spatial distribution of the C3S MME accuracy for both terciles and quartiles forecasts can be readily related to the average position of the ITCZ and the SPCZ (see **Figure 2**): The regions with higher accuracy tend to flank the *average* position of the convergence zones (i.e., regions that experience significant rainfall anomalies when there are large variations in the position of the convergence zones, usually associated with ENSO). The predictability of rainfall in the region is therefore clearly linked to the ability of the GCMs to forecast shifts in the position and intensity of the convergence zones.

2.4. Forecasts of drought conditions

Figures 10 and **11** present the precision, recall and F1 scores for the C3S MME forecasts of the lower tercile and lower quartile categories, respectively.

Overall, the general pattern follows the spatial distribution of the accuracy scores in **Figure 9**, but these figures provide insights into the ability of the C3S MME forecast system to specifically forecast dry conditions in comparison with the other rainfall categories.

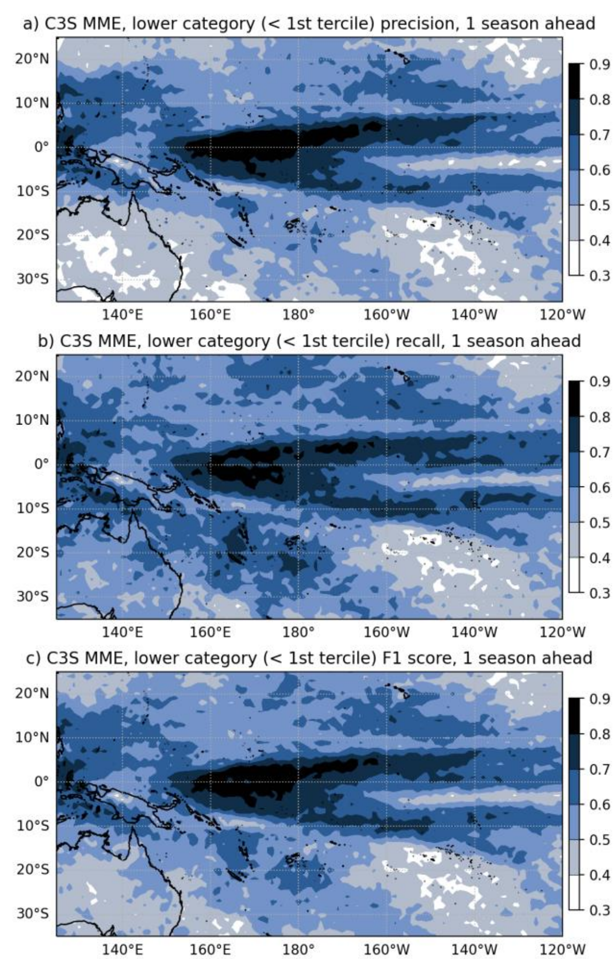

Figure 10: *Precision (a), recall (b) and F1 score (c) for MME seasonal (1 season ahead) forecasts of the lower tercile category (precipitation below the 33rd percentile)*

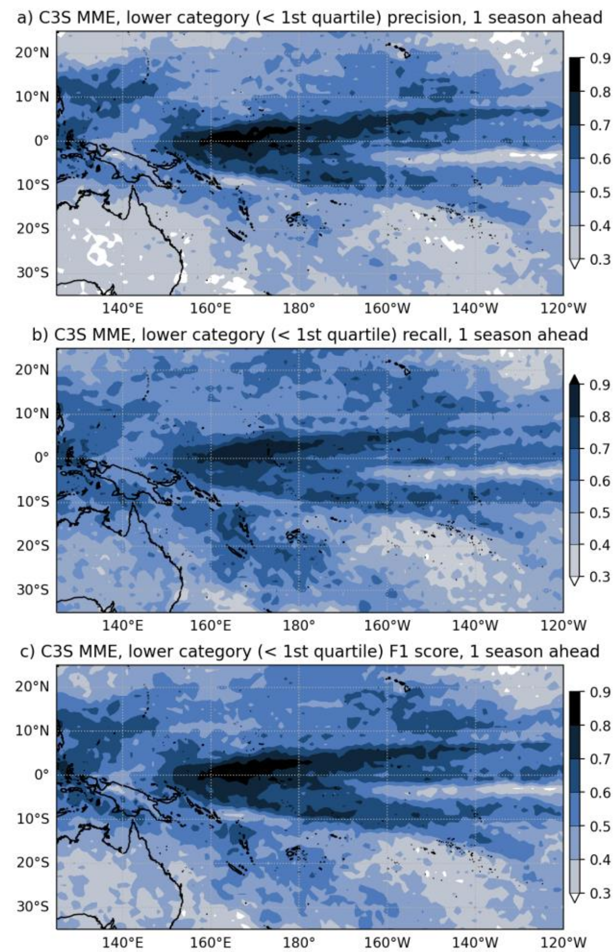

Figure 11*: Precision (a), recall (b) and F1 score (c) for MME seasonal (1 season ahead) forecasts of the lower quartile category (precipitation below the 25th percentile)*

In particular, for any grid point, one can extract the precision and recall statistics and derive insights related to the context-dependent costs of False Positives (predicting drought conditions that failed to eventuate) and False Negatives (failure to predict drought conditions that actually occurred).

2.5. Sub-regional time-series

In the supplementary material (**A1** and **A2**) we provide tables presenting respectively the accuracy scores for the tercile and quartile probabilistic forecasts from the MME for the 73 sub-regional (PICs administrative areas) time-series (see **Figure 2**). These tables do not provide additional information on the performance of the C3S MME compared to **Figure 9**, but are an example of the tailored products that can be derived from the availability of open seasonal forecast data. Tables of probabilistic tercile and quartile seasonal forecasts are derived operationally every month and are part of the suite of products offered by the ICU.

2.6. Combining forecast and near realtime rainfall monitoring

In the previous section we demonstrated that the overall skill of the C3S MME probabilistic forecast system is reasonable (i.e., exceeding the skill of a 'climatological' forecast) for a large proportion of the southwest Pacific, and that, in particular, forecasts for dry conditions, such as rainfall accumulations below the 25th percentile, are associated with reasonable precision and recall statistics (see section 1 and **Figure 11**). The overall performance of the forecast system makes it conceivable to combine it with near-realtime rainfall monitoring information to alert national or regional institutions of potential "water stress" conditions: Defined here when rainfall has recently been in deficit and forecasts indicate a high likelihood of dry conditions to persist or worsen. As one example, based on the empirical conditional statements presented in section 1 (**Table 2**), operationally on the 2nd of each month, we combine the percentiles of scores for the past 90 days (i.e., up to the last day of the previous calendar month) and the probability for the following month or season (three month accumulation) to be below the 25th percentile to produce and map three water stress categories, corresponding to the likely trajectories of drought conditions over the region (**Figure 12**). Further work is underway to assess the validity of combining the satellite derived near real-time three drought indices mentioned above with probabilistic forecasts of these three drought indices.

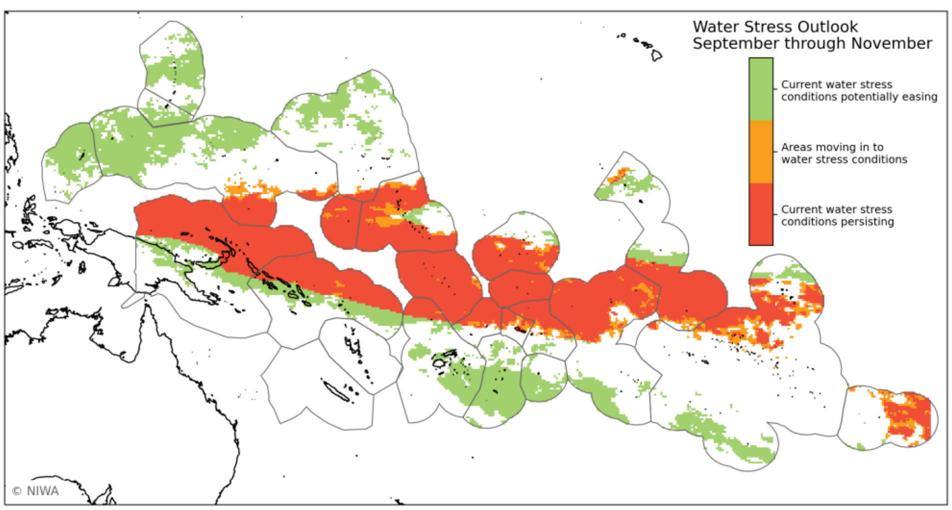

***Figure 12**: Island Climate Update "Water Stress" outlook, which combines near-real-time information and probabilistic, seasonal rainfall forecasts from the C3S MME. This outlook is for the period September – November 2022 and is based on the 90 days GPM-IMERG rainfall accumulations ending 30 August 2022 as well as the C3S MME forecast for the period September – November 2022.*

Summary and conclusions

In this paper, we present a set of products notably aimed at tracking and forecasting drought conditions across the tropical Pacific region. The system employs near–real time satellite rainfall estimates to track the evolution of several drought indices and indicators at different time-scales and a state-of-the-art probabilistic seasonal forecast system based on the forecasts provided operationally for nine coupled ocean-atmosphere GCMs. The validation of the individual GCMs and the MME for the region, both deterministic and probabilistic forecasts, show that the MME out-performs even the best GCM, and is able to forecast the general patterns of rainfall anomalies over the region. The MME's performance varies significantly seasonally (with summer rainfall usually better predicted) and is also a function of the ENSO state (phase and location of maximum SST anomalies). The probabilistic forecasts for dryness, such as rainfall being below the 25^{th} percentile, have reasonable precision and recall for the majority of the region, indicating that this system is associated with a relatively low rate of 'false alarms' or 'misses' for dry conditions. Using empirical conditional statements, the probabilistic information can be combined with the real-time rainfall estimates to highlight regions at risk of "water stress". This product is representative of the kind of climate services that can be developed based on openly available seasonal climate forecast and climate monitoring data. It is enabled by the development of an open-source, flexible software infrastructure, made possible by the growing popularity of the open-source Python programming language in the climate and meteorology communities, the reliance on well-tested, self-described data formats, and the development of an integrated eco-system of third-party Python libraries (packages). This combination of tools can handle all steps of the data processing and analysis pipelines and allows small-scale parallelization, making it possible to run the processing, analysis and visualisation pipeline on small-scale hardware such as a decent laptop.

It was therefore our goal to show that the use of open-data and open-source software, and recent advances in small scale parallelization and out-of-core computation, makes it possible to develop tailored climate services leveraging large-scale ensemble seasonal forecast systems (such as the C3S MME) as a part of a larger, integrated system combining several data streams. While the example presented combines rainfall, and in particular drought, monitoring and forecasting, it could easily be adapted to other variables (such as tracking and predicting the development of marine heatwaves see Jacox et al, 2022) or to develop input fields or time-series to 'downstream' models, either mechanistic, empirical or conceptual. Indeed the data streams generated as part of this project are now being used as input to a range of country-level climate service products, impact forecasting and decision-support systems being developed and operational in a wide range of Pacific Island countries.

References

Abernathey R., Paul K., Hamman J., Rocklin M., Lepore C., Tippett M., Henderson N., Seager R, May R., Del Vento D., 2017. Pangeo: An Open Source Big Data Climate Science Platform (NSF award 1740648)

Ashok, K., S. K. Behera, S. A. Rao, H. Weng, and T. Yamagata, 2007: El Niño Modoki and its possible teleconnection. *J. Geophys. Res.*, **112**, C11007, https://doi.org/10.1029/2006JC003798.

Beck, H. E., E. F. Wood, M. Pan, C. K. Fisher, D. G. Miralles, A. I. J. M. van Dijk, T. R. McVicar, and R. F. Adler, 2019: MSWEP V2 Global 3-Hourly 0.1° Precipitation: Methodology and Quantitative Assessment. *Bulletin of the American Meteorological Society*, **100**, 473–500, https://doi.org/10.1175/BAMS-D-17-0138.1.

Brady, R., Spring, A., 2021. climpred: Verification of weather and climate forecasts. JOSS 6, 2781. https://doi.org/10.21105/joss.02781

COSPPac – Climate and Oceans Support Programme in the Pacific Phase 2, 2018-22, Australian Pacific Climate Partnership. URL: http://cosppac.bom.gov.au/

Heim, R. R., C. Guard, M. A. Lander, and B. Bukunt, 2020: USAPI USDM: Operational Drought Monitoring in the U.S.-Affiliated Pacific Islands. *Atmosphere*, **11**, 495, https://doi.org/10.3390/atmos11050495.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Q.J.R. Meteorol. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hewitt, C., S. Mason, and D. Walland, 2012: The Global Framework for Climate Services. *Nature Clim Change*, **2**, 831–832, https://doi.org/10.1038/nclimate1745.

Hosking, J.R.M., 1990: L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B.* ***52*** *(1): 105–124.*

Hoyer, S., Hamman, J., 2017. xarray: N-D labeled Arrays and Datasets in Python. JORS 5, 10. https://doi.org/10.5334/jors.148

Huang, B., and Coauthors, 2017: Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons. *Journal of Climate*, **30**, 8179–8205, https://doi.org/10.1175/JCLI-D-16-0836.1.

G. Huffman, D. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, P. Xie, 2014: Integrated Multi-satellitE Retrievals for GPM (IMERG), version 4.4. NASA's Precipitation Processing Center, accessed 31 March, 2015, ftp://arthurhou.pps.eosdis.nasa.gov/gpmdata/

Iese, V., Kiem, A.S., Mariner, A. *et al.* Historical and future drought impacts in the Pacific islands and atolls. *Climatic Change* **166,** 19 (2021). https://doi.org/10.1007/s10584-021-03112-1

Lorrey, A.M., Salinger, J., and Renwick, J.A. 2009. The Centennial Issue of the Island Climate Update: Progress and success resulting from a multi-model ensemble forecast. Island Climate Update Bulletin, Issue 100. Published 15 January 2009

McGree, S. and Baleisolomone. 2009. Training Course on Climate Predictions and Applications, 16–20 Feb 2009. Island Climate Update Bulletin, Issue 102. Published 30 March 2009

Meehl, G.A., Richter, J.H., Teng, H., Capotondi, A., Cobb, K., Doblas-Reyes, F., Donat, M.G., England, M.H., Fyfe, J.C., Han, W., Kim, H., Kirtman, B.P., Kushnir, Y., Lovenduski, N.S., Mann, M.E., Merryfield, W.J., Nieves, V., Pegion, K., Rosenbloom, N., Sanchez, S.C., Scaife, A.A., Smith, D., Subramanian, A.C., Sun, L., Thompson, D., Ummenhofer, C.C., Xie, S.-P., 2021. Initialized Earth System prediction from subseasonal to decadal timescales. Nat Rev Earth Environ. https://doi.org/10.1038/s43017-021-00155-x

Neelin, J.D., Battisti, D.S., Hirst, A.C., Jin, F.-F., Wakata, Y., Yamagata, T., Zebiak, S.E., 1998. ENSO theory. J. Geophys. Res. 103, 14261–14290. https://doi.org/10.1029/97JC03424

Power, Scott B., and François P. D. Delage. "El Niño–Southern Oscillation and Associated Climatic Conditions around the World during the Latter Half of the Twenty-First Century." *Journal of Climate* 31, no. 15 (August 2018): 6189–6207. https://doi.org/10.1175/JCLI-D-18-0138.1.

Rocklin, M. (2015). Dask: Parallel Computation with Blocked algorithms and Task Scheduling. Python in Science Conference, 126–132. https://doi.org/gfz6s5

Silva, Nicolas A., Benjamin G. M. Webber, Adrian J. Matthews, Matthew M. Feist, Thorwald H. M. Stein, Christopher E. Holloway, and Muhammad F. A. B. Abdullah. "Validation of GPM IMERG Extreme Precipitation in the Maritime Continent by Station and Radar Data." *Earth and Space Science* 8, no. 7 (July 2021). https://doi.org/10.1029/2021EA001738.

Schneider, T., T. Bischoff, and G. H. Haug, 2014: Migrations and dynamics of the Intertropical Convergence Zone. Nature, 513, 45–53, doi:10.1038/nature13636.

Shen, H., 2014: Interactive Notebooks: sharing the code. *Nature,* volume **515**, 151–152.

SPECS, 2016: "Methodologies for calibration and combination of global and downscaled s2d predictions", SPECS project report, http://www.specs-fp7.eu/sites/default/files/u1/SPECS_D53.1.pdf.

Trenberth, K. E., and D. P. Stepaniak, 2001: Indices of El Niño Evolution. *J. Climate*, **14**, 1697–1701, https://doi.org/10.1175/1520-0442(2001)014<1697:LIOENO>2.0.CO;2.

van Rossum G., 2001. Python/C API Reference Manual, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.211.6702&rep=rep1&type=pdf

Vincent, D.G. 1994: The South Pacific Convergence Zone (SPCZ): a review. *Mon. Wea. Rev.*, 122, 1949–1970, doi:10.1175/1520

Virtanen, P., Gommers, R., Oliphant, and co-authors, 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686

Widlansky, M.J., Webster, P.J., Hoyos, C.D., 2011. On the location and orientation of the South Pacific Convergence Zone. Clim Dyn 36, 561–578. https://doi.org/10.1007/s00382-010-0871-6

WMO, 2020: *Guidance on Operational Practices for Objective Seasonal Forecasting*. World Meteorological Organization, https://library.wmo.int/doc_num.php?explnum_id=10314. ISBN 978-92-63-11246-9.

Yun, K.-S., J.-Y. Lee, A. Timmermann, K. Stein, M. F. Stuecker, J. C. Fyfe, and E.-S. Chung, 2021: Increasing ENSO–rainfall variability due to changes in future tropical temperature–rainfall relationship. *Commun Earth Environ*, **2**, 43, https://doi.org/10.1038/s43247-021-00108-8.

Xie, P., and P. A. Arkin, 1997: Global Precipitation: A 17-Year Monthly Analysis Based on Gauge Observations, Satellite Estimates, and Numerical Model Outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539–2558.

Supplementary material for:

Fauchereau N., Ramsay D., Noll B.E. and Lorrey A.M.: "Open data and open source software for the development and validation of multi-model monthly-to-seasonal probabilistic forecasts for the Pacific Islands" (submitted to *Climate Services*). Hereafter FRNL2022.

A1, Supplementary Table 1: Accuracy (or "hit rate") expressed in percentage for the C3S MME seasonal tercile probabilistic forecasts derived for each of the 73 Island Climate Update (ICU) administrative areas (see **Figure 2** in FRNL2022). The column labels indicate the leadtime (in months) so that 3 refers to 1-season ahead forecast (*e.g.* a forecast initialized on the 1st of January for the February-March period). The validation is based on the 1993-2016 hindcast period, with the MME including 7 GCMs (ECMWF, UKMO, Météo-France, CMCC, DWD, NCEP and JMA, see **Table 1** in FRNL2022). The table is sorted by the 1-season ahead accuracy. Accuracy exceeding 40% is shaded in green, note that for tercile probabilistic forecasts, a climatological forecast would have an accuracy of 33%.

Country	District	3	4	5
Kiribati	Gilberts-South	78%	74%	69%
Kiribati	Gilberts-North	75%	72%	72%
Nauru	Nauru	75%	69%	63%
Kiribati	Ocean Island	74%	68%	64%
Kiribati	Northern Line Islands	71%	69%	70%
Kiribati	Phoenix Islands	66%	63%	60%
Tuvalu	Northern Tuvalu	66%	63%	63%
French Polynesia	Marquesas Islands	61%	55%	56%
Kiribati	Southern Line Islands	59%	57%	55%
Vanuatu	Torba	59%	56%	53%
Tokelau	Tokelau	59%	55%	52%
Cook Islands	Northern Cook Islands	58%	56%	56%
Tuvalu	Southern Tuvalu	58%	55%	55%
Palau	Babeldaob region	57%	57%	51%
Palau	South-west Islands	57%	50%	43%
FSM	Kapingamarangi	57%	54%	49%
Vanuatu	Penama	55%	55%	57%
Papua New Guinea	Southern Region	55%	53%	50%
Fiji	Western	54%	54%	53%
Vanuatu	Sanma	54%	57%	53%
FSM	Pohnpei	54%	54%	54%
New Caledonia	Loyalty Islands Province	54%	55%	54%
FSM	Yap	54%	55%	48%
Papua New Guinea	Highlands Region	53%	46%	47%
Tonga	Vavau	53%	53%	49%
Vanuatu	Tafea	53%	55%	53%
Marshall Islands	Southern Marshall Islands	53%	52%	47%

Fiji	Central	53%	51%	49%
Vanuatu	Shefa	53%	54%	52%
Vanuatu	Malampa	52%	51%	52%
New Caledonia	South Province	52%	49%	51%
Fiji	Eastern	52%	54%	51%
Papua New Guinea	Momase Region	52%	46%	45%
Solomon Islands	Choiseul Province	51%	48%	47%
New Caledonia	North Province	51%	47%	47%
American Samoa	Swains	51%	48%	48%
Tonga	Haapai	51%	52%	52%
FSM	Chuuk	51%	51%	49%
Solomon Islands	Temotu Province	50%	50%	42%
FSM	Kosrae	50%	49%	48%
Guam	Guam	50%	47%	51%
Northern Mariana Islands	Northern Islands	49%	41%	40%
Solomon Islands	Isabel Province	49%	50%	45%
Niue	Niue	49%	48%	49%
Papua New Guinea	Islands Region	49%	45%	46%
Northern Mariana Islands	Southern Islands	49%	45%	46%
Solomon Islands	Makira-Ulawa Province	48%	46%	44%
Tonga	Tongatapu-Eua	48%	48%	49%
Kiribati	Central Line Islands	48%	46%	43%
Fiji	Northern	47%	47%	45%
Solomon Islands	Rennell and Bellona	47%	45%	45%
American Samoa	Manua	47%	43%	42%
Marshall Islands	Northern Marshall Islands	46%	43%	40%
Marshall Islands	Central Marshall Islands	46%	42%	42%
Samoa	Savaii	46%	44%	43%
Solomon Islands	Central Province	46%	44%	42%
Fiji	Rotuma	46%	45%	46%
Tonga	Niuas	46%	46%	46%
American Samoa	Tutuila	46%	44%	42%
French Polynesia	Tuamotu Archipelago	45%	48%	43%
Solomon Islands	Malaita Province	45%	43%	41%
Samoa	Upola	45%	43%	40%
Pitcairn	Ducie	45%	41%	43%
Solomon Islands	Wesgtern Province	44%	44%	43%
Solomon Islands	Guadacanal Province	44%	41%	41%
Pitcairn	Pitcairn, Henderson & Oeno	44%	38%	40%
Wallis et Futuna	Futuna	43%	46%	47%
Wallis et Futuna	Wallis	41%	42%	42%
French Polynesia	Austral Islands	40%	37%	37%
French Polynesia	Gambier Islands	39%	31%	38%

Cook Islands	Southern Cook Islands	38%	41%	37%
French Polynesia	Windward-Society Islands	38%	35%	32%
French Polynesia	Leeward-Society Islands	33%	32%	33%

A2, Supplementary Table 2: Same as **A1** but for seasonal quartile probabilistic forecasts. Accuracy exceeding 30% is shaded in green. Note that for quartile probabilistic forecasts, a climatological forecast would have an accuracy of 25%.

Country	**District**	**3**	**4**	**5**
Kiribati	Gilberts-North	59%	56%	56%
Nauru	Nauru	58%	56%	52%
Kiribati	Gilberts-South	58%	56%	51%
Kiribati	Ocean Island	53%	51%	49%
Kiribati	Northern Line Islands	48%	47%	48%
FSM	Kapingamarangi	47%	44%	42%
Kiribati	Phoenix Islands	45%	41%	38%
Tuvalu	Northern Tuvalu	43%	42%	44%
Palau	South-west Islands	42%	37%	35%
Vanuatu	Torba	42%	41%	39%
Papua New Guinea	Momase Region	41%	35%	40%
Vanuatu	Penama	41%	38%	38%
Tuvalu	Southern Tuvalu	41%	42%	42%
FSM	Chuuk	41%	39%	40%
Palau	Babeldaob region	41%	39%	35%
Papua New Guinea	Southern Region	40%	39%	37%
Papua New Guinea	Islands Region	40%	41%	39%
Tokelau	Tokelau	40%	40%	38%
FSM	Pohnpei	39%	41%	40%
Vanuatu	Sanma	39%	38%	39%
Solomon Islands	Makira-Ulawa Province	39%	37%	33%
French Polynesia	Marquesas Islands	39%	38%	36%
Solomon Islands	Rennell and Bellona	39%	42%	38%
Northern Mariana Islands	Southern Islands	39%	35%	33%
FSM	Yap	39%	38%	36%
Tonga	Vavau	38%	37%	35%
Papua New Guinea	Highlands Region	38%	35%	33%
Cook Islands	Northern Cook Islands	38%	38%	35%
Guam	Guam	37%	34%	33%
Marshall Islands	Southern Marshall Islands	37%	34%	35%
FSM	Kosrae	37%	38%	38%
Fiji	Eastern	37%	35%	35%
Fiji	Central	37%	32%	35%
Fiji	Western	37%	36%	40%

Vanuatu	Shefa	36%	35%	36%
Kiribati	Southern Line Islands	36%	36%	37%
Solomon Islands	Choiseul Province	36%	37%	35%
Tonga	Haapai	36%	37%	37%
Vanuatu	Malampa	36%	36%	38%
Solomon Islands	Isabel Province	36%	38%	36%
Solomon Islands	Central Province	36%	35%	31%
Vanuatu	Tafea	36%	35%	35%
Tonga	Tongatapu-Eua	35%	39%	40%
American Samoa	Tutuila	35%	30%	33%
New Caledonia	Loyalty Islands Province	35%	35%	36%
Fiji	Rotuma	35%	32%	34%
New Caledonia	South Province	35%	34%	36%
Niue	Niue	35%	33%	34%
New Caledonia	North Province	35%	34%	38%
Solomon Islands	Guadacanal Province	34%	33%	32%
American Samoa	Swains	34%	35%	34%
American Samoa	Manua	34%	34%	30%
Solomon Islands	Wesgtern Province	33%	35%	32%
Fiji	Northern	33%	32%	32%
Kiribati	Central Line Islands	33%	30%	32%
Solomon Islands	Temotu Province	33%	37%	35%
Samoa	Savaii	33%	28%	31%
Wallis et Futuna	Futuna	32%	30%	29%
Marshall Islands	Central Marshall Islands	32%	29%	29%
Pitcairn	Pitcairn, Henderson&Oeno	32%	31%	30%
Wallis et Futuna	Wallis	32%	29%	34%
Samoa	Upola	31%	29%	29%
Solomon Islands	Malaita Province	31%	30%	30%
Northern Mariana Islands	Northern Islands	31%	32%	31%
Marshall Islands	Northern Marshall Islands	30%	32%	30%
French Polynesia	Gambier Islands	30%	32%	30%
Tonga	Niuas	30%	32%	29%
French Polynesia	Tuamotu Archipelago	30%	29%	29%
Pitcairn	Ducie	29%	30%	31%
French Polynesia	Austral Islands	29%	25%	25%
French Polynesia	Windward-Society Islands	27%	25%	26%
French Polynesia	Leeward-Society Islands	27%	26%	26%
Cook Islands	Southern Cook Islands	24%	30%	31%