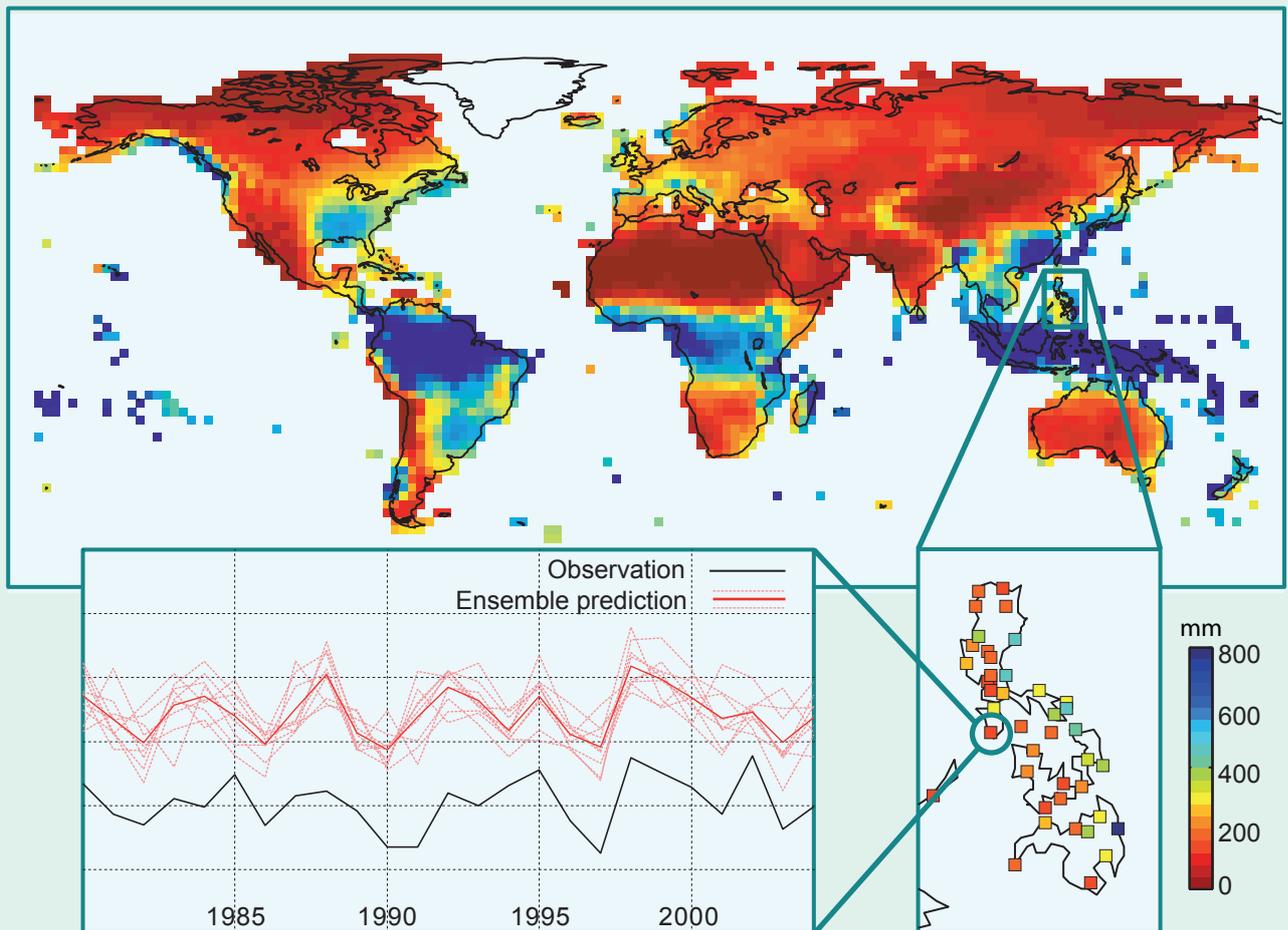


PhD THESIS

STATISTICAL DOWNSCALING OF PRECIPITATION IN SEASONAL FORECASTING: ADVANTAGES AND LIMITATIONS OF DIFFERENT APPROACHES

TESIS DOCTORAL

REGIONALIZACIÓN ESTADÍSTICA DE PRECIPITACIÓN
EN LA PREDICCIÓN ESTACIONAL: VENTAJAS Y
LIMITACIONES DE DISTINTAS ESTRATEGIAS



AUTOR: RODRIGO GARCÍA MANZANAS

DIRECTOR: José Manuel Gutiérrez Llorente

PHD THESIS

Statistical downscaling of precipitation in seasonal
forecasting: Advantages and limitations of different
approaches

PHD PROGRAMME IN SCIENCE, TECHNOLOGY AND
COMPUTATION

Presented by:

RODRIGO GARCÍA MANZANAS

Under the supervision of:

Dr. José Manuel Gutiérrez Llorente



University of Cantabria

June 2016

Statistical downscaling of precipitation in seasonal forecasting: Advantages and limitations of different approaches

Regionalización estadística de precipitación en la predicción estacional: Ventajas y limitaciones de distintas estrategias

*Rodrigo García Manzananas
Santander Meteorology Group
Institute of Physics of Cantabria
(CSIC–University of Cantabria)
Santander, Spain
June 2016*

“Beautiful is what we see, more beautiful what we know, but by far the most beautiful is what we do not know.”

(Niels Stensen. Danish polymath, doctor and anatomist from the XVII century).

*A mis padres,
por dármelo todo.*

Contents

Acknowledgements	1
Acronyms	3
1 Resumen en Español	5
1.1 Contexto y Objetivos de la Tesis	5
1.2 Síntesis	7
1.3 Principales Resultados	10
1.4 Publicaciones Relacionadas	18
1.5 Software Desarrollado: MeteoLab	19
1.6 Líneas Futuras de Trabajo	20
2 Context, Objectives and Structure of the Thesis	23
2.1 Context	23
2.2 Objectives	24
2.3 Structure	25
I Introduction	27
3 Seasonal Forecasting	29
3.1 Introduction to Seasonal Forecasting	30
3.2 Empirical Seasonal Forecasting	30
3.3 Dynamical Seasonal Forecasting: Global Climate Models (GCMs)	33
3.4 Multimodel Ensembles	36
3.4.1 The ENSEMBLES Multimodel Seasonal Hindcast	38
3.4.1-1 ENSO Teleconnections	38

3.4.1-2	Bias and Drift	41
3.5	Verification of Seasonal Forecasts	43
3.5.1	Reliability Categories	47
3.6	State-of-the-Art of Seasonal Forecasting	49
3.6.1	Operational Seasonal Forecasts	49
3.6.2	Towards Climate Services	51
3.7	Beyond the State-of-the-Art in Seasonal Forecasting	53
4	A Regional Case Study: The Philippines	55
4.1	Observed Climate	56
4.2	From Global to Local: The Need for Downscaling	60
II	Methodological Framework	65
5	Statistical Downscaling (SD)	67
5.1	Approaches for Statistical Downscaling	69
5.1.1	Perfect Prog (PP)	69
5.1.2	Model Output Statistics (MOS)	71
5.1.3	PP versus MOS	72
5.2	Cross-Validation	73
5.3	Techniques for Statistical Downscaling	73
5.4	SD Methods Used in this Thesis	75
5.4.1	PP Methods	75
5.4.1-1	Analogs	75
5.4.1-2	Generalized Linear Models	81
5.4.2	MOS-BC (Bias Correction) Methods	86
5.5	State-of-the-Art of SD in Seasonal Forecasting	90
6	Methodological Aspects for SD in Seasonal Forecasting	93
6.1	Predictor Screening	94
6.2	Season-Specific versus Yearly Data for Calibration	97
6.3	Harmonization of GCM Predictor Data	101
III	Main Results	109
7	Validation of Global Seasonal Precipitation Forecasts	111
7.1	Data Used	112

7.2	Validation Methodology	113
7.3	Overall Skill (Worldwide)	115
7.4	Overall Skill (in the Philippines)	117
7.5	ENSO-Driven Skill	121
7.5.1	Sea Surface Temperature in El Niño 3.4 Region	121
7.5.2	ENSO Teleconnections (Worldwide)	122
7.5.3	ENSO Teleconnections (in the Philippines)	126
7.5.4	Contribution of ENSO to the Overall Skill	126
8	The Effect of Reanalysis Uncertainty on Perfect Prog	129
8.1	Data and SD Method Used	130
8.2	Reanalysis Uncertainty (in the Philippines)	130
8.3	Sensitivity to Reanalysis Choice in Cross-Validation Mode	131
8.4	Sensitivity to Reanalysis Choice in Seasonal Forecasts	138
8.5	Sensitivity to Reanalysis Choice in Climate Change Projections	139
9	SD Methods for Seasonal Forecasting: Advantages and Limitations	145
9.1	SD Methods Used	146
9.2	Results for Reanalysis Predictors	147
9.3	Results for GCM Predictors	149
9.3.1	Accuracy	149
9.3.2	Reliability	155
9.4	The Added Value of PP Methods	158
IV	Concluding Remarks	163
10	Conclusions, Achievements and Future Work	165
10.1	Main Conclusions	165
10.2	Related Publications	171
10.3	Developed Software: MeteoLab	172
10.4	Future Work	173
	Bibliography	177

Acknowledgements

This Thesis has been developed at the Santander Meteorology Group (<http://www.meteo.unican.es>), formed by researchers and professors from the Department of Applied Mathematics and Computer Science of the University of Cantabria (UC) and the Institute of Physics of Cantabria (UC and Spanish National Research Council), in the context of the international research projects EUPORIAS (European Provision Of Regional Impacts Assessments on Seasonal and Decadal Timescales) and SPECS (Seasonal-to-Decadal Climate Prediction for the Improvement of European Climate Services), funded by the EU Commission through grant agreements 308291 and 308378, respectively.

First of all, I want to express my most sincere gratitude to all my colleagues at the Santander Meteorology Group. Thank you for giving me the privilege of working in an environment where true cooperation and teamwork is the rule and for your dedication, patience and readiness to solve any doubt at any time, which made possible the accomplishment of this Thesis as well as many other professional achievements. Especially, I am indebted to my director, Dr. José Manuel Gutiérrez Llorente. Thank you for giving me the opportunity to join the group, for supporting and trusting me from the first day and for your constant willingness to help and motivate me; this Thesis would not have been possible without your invaluable supervision and guidance. I am also grateful to the SENAMHI (Servicio Nacional de Meteorología e Hidrología del Perú) team in Lima for hosting me for three months, a very important period of my life, both professionally and personally (*ACyV*).

Besides, I would like to thank all those people who significantly contributed to my education at any point, especially Dra. Concepción Rodríguez Puebla, for introducing me in the climate science and being the key to make me land at the Santander Meteorology Group. Last but not least, I am also grateful to the various anonymous reviewers who helped with their useful comments to substantially improve the articles that form part of this Thesis.

With regard to the data that have been used for the realization of this Thesis, I would like to thank in first place the Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA: <http://meteopilipinas.gov.ph>) for the high-quality observations provided, which have been essential for the development of part of my research during the last years. Likewise, I want to thank the EU project ENSEMBLES, financed by the EU Commission through contract GOCE-CT-2003-505539, for the seasonal simulations provided, which are used throughout the entire dissertation.

At a more personal level, I would like to thank all my friends (they know who they are and what they mean to me) as well as all those people who have been close to me at some moment. To conclude, my deepest thanks to my family, especially to my parents, who have always supported me, in every possible way, to help me achieving my goals. You know I really love you.

To all of you, many thanks.

Rodrigo García Manzananas
June 2016

Acronyms

The following list contains the acronyms most used throughout the Thesis:

ACC Anomaly Correlation Coefficient

BC Bias Correction

CDF Cumulative Distribution Function

CMCC-INGV Euro-Mediterranean Centre for Climate Change

CT Climate Type

DD Dynamical Downscaling

DJF December-January-February

ECMWF European Centre for Medium-Range Weather Forecasts

ENSO El Niño-Southern Oscillation

EUPORIAS EUropean Provision Of Regional Impacts Assessments on Seasonal and decadal timescales

FAR False Alarm Rate

GCM Global Climate Model

GLM Generalized Linear Model

HIR Hit Rate

IFM-GEOMAR Leibniz Institute of Marine Sciences

JJA June-July-August

MAM March-April-May

MF Météo France

MM Multimodel

MOS Model Output Statistics

MOS-BC Model Output Statistics (Bias Correction)

PAGASA Philippine Atmospheric, Geophysical and Astronomical Services Administration

PC Principal Component

PDF Probability Density Function

PP Perfect Prog

q-q Quantile-quantile

r Pearson correlation

rs Spearman correlation

RCM Regional Climate Model

ROCSS ROC (Relative Operating Characteristic) Skill Score

SD Statistical Downscaling

SDM Statistical Downscaling Method

SON September-October-November

SPECS Seasonal-to-decadal climate Prediction for the improvement of European Climate Services

SST Sea Surface Temperature

SVS-LRF Standardized Verification System for Long Range Forecasts

UKMO UK Met Office

CHAPTER 1

Resumen en Español

1.1 Contexto y Objetivos de la Tesis

Las predicciones meteorológicas a corto y medio plazo son un producto muy popular y demandado, tanto a nivel social como comercial. Estas predicciones proporcionan una estimación de las variables meteorológicas de interés (por ejemplo temperatura o precipitación) durante los próximos días, y se elaboran a partir de los datos proporcionados por modelos numéricos de circulación atmosférica que permiten simular la evolución de la atmósfera a partir de una condición inicial dada. Para tener en cuenta las distintas fuentes de incertidumbre que influyen en este tipo de predicción se ha puesto a punto durante las últimas décadas una metodología probabilística basada en la utilización de diferentes predicciones obtenidas a partir de condiciones iniciales ligeramente perturbadas. Esta *predicción por conjuntos* (o *ensembles*) se ha convertido en la herramienta principal para la predicción meteorológica moderna.

Paralelamente, se han producido notables avances en la modelización numérica de la circulación acoplada océano-atmósfera que han permitido desarrollar una nueva generación de modelos para la predicción estacional. A diferencia de la predicción a corto y medio plazo, la predicción estacional no trata de predecir la temperatura o precipitación para un instante de tiempo concreto en unos pocos días, sino las condiciones climáticas promedio para las próximas estaciones (por ejemplo, en forma de anomalías basadas en terciles: frío/normal/cálido, seco/normal/húmedo), con un horizonte temporal de hasta un año. La predecibilidad a esta escala temporal se debe principalmente a la componente oceánica, que tiene una dinámica más lenta que la atmosférica. Dadas las actuales limitaciones de cómputo, la resolución espacial de estas predicciones es todavía demasiado grosera

(en torno a los cientos de km). Asimismo, la naturaleza probabilística de las predicciones estacionales —además de la incertidumbre en las condiciones iniciales es frecuente también considerar la incertidumbre en la formulación del modelo (predicciones multimodelo)— dificulta en gran medida su uso, ya que obliga a utilizar herramientas específicas para su adecuada validación y postproceso.

Pese a ello, las predicciones estacionales tienen un gran número de aplicaciones y ayudan a la toma de decisiones en muchos sectores socioeconómicos importantes como la agricultura, la energía, el transporte, la salud y la hidrología. Por este motivo, su uso no ha dejado de aumentar en las últimas décadas. Entre las variables con mayor demanda (temperatura, precipitación y viento), la precipitación es la más problemática por su carácter mixto (ocurrencia/cantidad) y su alta variabilidad espacial (véase, por ejemplo, Schmidli et al., 2007; Bundel et al., 2011). Por esta razón, la presente Tesis Doctoral se centra exclusivamente en esta variable.

A diferencia del corto y medio plazo, los actuales modelos numéricos para la predicción estacional no son fiables globalmente sino que, para cada variable de interés y estación del año, su utilidad está limitada a ciertas regiones del mundo, principalmente los trópicos (Stockdale et al., 1998). Además, la baja resolución espacial de estos modelos resulta insuficiente para la mayoría de aplicaciones prácticas y sus predicciones no pueden ser utilizadas directamente en estudios de impacto, por lo que es necesario realizar algún tipo de postproceso o regionalización (*downscaling* en inglés) que permita llevarlas a una escala local útil.

Para ello, dos enfoques conceptualmente distintos han sido desarrollado en las últimas décadas: el *downscaling* dinámico y el estadístico. Mientras que el *downscaling* dinámico se basa en el uso de modelos regionales que, anidados a las salidas de baja resolución de los modelos globales, simulan numéricamente el clima a una resolución más alta sobre un área limitada (véase, por ejemplo, Giorgi and Mearns, 1999; Laprise, 2008), el *downscaling* estadístico (SD, por sus siglas en inglés) se basa en modelos estadísticos/algoritmos que relacionan las salidas de baja resolución de los modelos globales (*predictores*) con las observaciones locales (*predictandos*) sobre la zona de interés (véase, por ejemplo, von Storch et al., 1993). En esta Tesis Doctoral se considera únicamente este último, puesto que se ha comprobado su potencial para mejorar la calidad de las predicciones globales de precipitación estacional (Feddersen and Andersen, 2005) y su coste computacional es drásticamente menor. Sin embargo, mientras que el SD ha sido ampliamente utilizado para la modelización del cambio climático, la experiencia hasta la fecha para la predicción estacional es más limitada (véanse, por ejemplo, los resultados del informe D52.1 del proyecto SPECS: *Review of the different statistical downscaling methods for s2d prediction*). Además, las técnicas de SD han sido desarrolladas y aplicadas en su mayor parte para los

extratropicos, ya que varios factores dificultan su éxito en los trópicos (Hewitson et al., 2014), precisamente la zona de mayor predecibilidad a escala estacional.

En base a las consideraciones anteriores, en esta Tesis se plantean los siguientes tres objetivos principales:

1. Realizar una validación exhaustiva de la calidad de las predicciones de precipitación estacional dadas por los modelos globales y estudiar el efecto que el fenómeno *El Niño-Oscilación del Sur* (ENSO, por sus siglas en inglés) —el principal factor de predecibilidad estacional (ver, por ejemplo, Goddard and Dilley, 2005; Doblas-Reyes et al., 2010)— pueda tener sobre la misma.
2. Adaptar las distintas metodologías y técnicas para SD para su correcta aplicación en el contexto de la predicción estacional, contemplando diversos factores que no se han tenido en cuenta antes, tales como el efecto de la incertidumbre en el reanálisis sobre las predicciones regionalizadas.
3. Analizar las ventajas y limitaciones de las distintas metodologías y técnicas para SD en el contexto de la predicción estacional, considerando para ello las medidas de validación adecuadas que permitan evaluar el (posible) valor añadido de las predicciones regionalizadas de precipitación con respecto a las salidas directas de los modelos globales.

Para alcanzar estos objetivos, gran parte de la Tesis se centra en una región de estudio particularmente interesante localizada en los trópicos: Filipinas. Los motivos que sustentan esta elección se exponen en detalle en el Capítulo 4.

1.2 Síntesis

Esta Tesis Doctoral está organizada en cuatro bloques principales: un primer bloque introductorio (Parte I) al que sigue un bloque metodológico (Parte II), un bloque central en el que se presentan los principales resultados (Parte III) y un último bloque (Parte IV) en el que se exponen las principales conclusiones.

La Parte I abre con un capítulo preliminar (Capítulo 3) en el que se introducen en primer lugar las bases de la predicción estacional. A continuación se hace un breve barrido por las diferentes fases que ha ido experimentando la misma, desde la predicción empírica —basada principalmente en el concepto de teleconexión (ver capítulo para más detalles)— hasta el uso de los actuales modelos globales del clima (GCMs, por sus siglas en inglés), complejos modelos numéricos que resuelven las ecuaciones que describen los procesos que ocurren en los diferentes componentes del sistema climático (y las interacciones entre ellos).

Posteriormente se introduce la predicción por conjuntos (la predicción estacional ha de interpretarse en términos probabilísticos), así como el concepto de multimodelo, en el que diferentes GCMs se combinan con la idea de cuantificar la incertidumbre en la predicción debida a diferencias en la formulación de los distintos modelos. En este punto, se presenta el *dataset* de ENSEMBLES (Weisheimer et al., 2009), cuyas simulaciones se utilizan a lo largo de toda la Tesis por ser el multimodelo de predicciones estacionales retrospectivas (*hindcasts*) más largo y completo hasta la fecha. Se estudia como los modelos que componen este *dataset* reproducen la temperatura en la superficie del mar (SST, por sus siglas en inglés) observada en la región El Niño 3.4 y cómo simulan, a nivel mundial, los patrones de teleconexión del Niño con la precipitación. Se analiza también, a nivel mundial, la distribución espacio-temporal de sus patrones de *bias* y *drift* (ver capítulo para más detalles). Después se presenta una breve discusión sobre la verificación de las predicciones estacionales (con especial atención a la *reliability*) y se introducen las métricas que se utilizan para tal fin a lo largo de la Tesis. Para acabar, se da una perspectiva general del estado del arte en el campo de la predicción estacional (sistemas operativos, servicios climáticos, etc.), así como de los avances que pueden esperarse en el mismo en un futuro próximo.

En el Capítulo 4 se presenta el caso de estudio regional considerado para el desarrollo de gran parte de la Tesis, Filipinas, y se discute su idoneidad. En primer lugar, basándonos en el *dataset* de observaciones de alta calidad disponible para esta Tesis —precipitación diaria en 42 estaciones gestionadas por la Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA) y distribuidas uniformemente por todo el país para el período 1981-2005— se hace un estudio descriptivo del clima de la región. A continuación, se lleva a cabo una validación regional de la precipitación de los modelos de ENSEMBLES (se consideran las predicciones deterministas y se toma como referencia las observaciones en las 42 estaciones PAGASA). Los resultados de esta validación —los modelos muestran importantes errores sistemáticos, que cambian sustancialmente entre localidades próximas— ponen de manifiesto la necesidad existente de regionalizar (llevar a la escala local útil) las predicciones globales dadas por los GCMs.

Por tanto, en el primer capítulo de la Parte II (Capítulo 5) se proporciona una descripción completa de las diferentes metodologías disponibles para el *downscaling* estadístico —*Perfect Prog* (PP) y *Model Output Statistics* (MOS)— y se explican las diferencias entre ambos, así como sus relativos ‘pros y contras’. Se introduce también el concepto de validación cruzada, el cual es aplicado con asiduidad durante toda la Tesis. A continuación, se analizan en detalle los diferentes métodos de SD (SDMs, por sus siglas en inglés) utilizados en la Tesis. Entre los métodos PP (que infieren la precipitación local basándose en relaciones estadísticas entre las variables sinópticas de larga escala y las observaciones

locales), se consideran distintas configuraciones de la técnica de análogos y de modelos lineales generalizados (GLMs, por sus siglas en inglés). En cuanto a los métodos MOS, se consideran dos que son comúnmente utilizados para la corrección del *bias* (métodos MOS-BC), que ajustan la precipitación global dada por los GCMs apoyándose en las observaciones, a nivel de distribución. Todos estos métodos se aplican en condiciones ‘perfectas’, es decir, utilizando predictores de reanálisis tanto para la fase de calibración como para la de predicción (ver capítulo para más detalles) sobre cuatro de las estaciones PAGASA —representativas de los cuatro climas diferentes presentes en el país (ver Sección 4.1),— considerando para ello un marco de validación cruzada para el período 1981-2005, y se analizan sus ventajas e inconvenientes para el SD de predicciones estacionales. En base a los resultados de este análisis, se escoge la configuración de las distintas técnicas de SD que será utilizada durante el resto de la Tesis.

Más allá de la configuración de las propias técnicas, en el Capítulo 6 se analizan varios factores que son relevantes para el SD de predicciones estacionales, estableciéndose el marco metodológico bajo el cual se aplicarán las mismas durante la Tesis. En particular, se aborda la búsqueda de predictores (y dominio geográfico sobre el que se definen los mismos) adecuados para la región piloto, Filipinas —necesario para los métodos PP,— el tipo de dato utilizado para la calibración de las distintas técnicas (estacional o anual; ver el capítulo para más detalles) y se detalla el preproceso que se hace de las distintas variables predictoras antes de entrar en los diferentes SDMs.

La mayoría de los resultados centrales de la Tesis se presentan en la Parte III, la cual se estructura en torno a tres capítulos clave.

En el Capítulo 7 se evalúa la calidad de las predicciones estacionales globales de precipitación dadas por los modelos de ENSEMBLES (se toma como referencia una rejilla mundial de observaciones) para el período 1961-2000. Para ello se aplica una metodología de validación probabilística basada en terciles. Esta misma metodología se aplica también para una validación regional detallada sobre Filipinas, tomando en este caso como referencia la precipitación observada en las 42 estaciones PAGASA. Además, puesto que ENSO es el principal factor de predecibilidad estacional (ver, por ejemplo, Goddard and Dilley, 2005; Doblas-Reyes et al., 2010), se analiza también la relación existente entre este fenómeno y la calidad de las predicciones de los modelos globales. En particular, se estudia cómo dichos modelos reproducen la SST en la región El Niño 3.4 y se hace un estudio de las teleconexiones de El Niño/La Niña con la precipitación (se calcula para ello la frecuencia de ocurrencia de cada tercil en los eventos Niño/Niña), determinando hasta qué punto los diferentes modelos son capaces de reproducir los patrones observados.

En la primera parte del Capítulo 8 se cuantifica la incertidumbre del reanálisis, para lo cual se comparan dos reanálisis actuales de referencia sobre una región extensa que con-

tiene Filipinas y se calculan sus diferencias. A continuación, se evalúa el impacto que tiene la elección del reanálisis considerado para la calibración de los métodos PP en las predicciones regionalizadas. Para ello, se considera un método PP ilustrativo y se aplica (teniendo en cuenta los resultados del Capítulo 6) sobre las 42 estaciones PAGASA tanto en condiciones ‘perfectas’ (en modo validación cruzada para el período 1981-2005) como en condiciones ‘no perfectas’ (utilizando predictores GCM). En este último caso, se consideran dos horizontes temporales diferentes: las predicciones estacionales de los modelos de ENSEMBLES en modo *hindcast* (período 1981-2005) y las proyecciones de cambio climático para el modelo del Max Planck Institute (MPI) ECHAM5 (hasta el final del siglo 21).

Como colofón de la Tesis se presenta un capítulo clave (Capítulo 9) en el que se analizan las ventajas y limitaciones de las distintas metodologías (MOS-BC y PP) y técnicas para SD en el contexto de la predicción estacional. En primer lugar, se evalúa el límite de predecibilidad que puede alcanzarse mediante la aplicación de técnicas de SD en Filipinas. Para ello se considera un método ilustrativo PP y otro MOS-BC, los cuales se aplican en condiciones ‘perfectas’ sobre las 42 estaciones PAGASA (en modo validación cruzada para el período 1981-2005). A continuación, y con el objetivo de ver si el SD puede mejorar la calidad de las salidas directas de precipitación de los modelos globales (más allá de reducir sus *biases* sistemáticos), se aplican dos métodos PP y otros dos MOS-BC a los modelos de ENSEMBLES, y las predicciones regionalizadas se comparan con las correspondientes salidas directas sobre las 42 estaciones, considerando para ello métricas de verificación adecuadas que caracterizan el *accuracy* y la *reliability* (para el caso de predicciones deterministas y probabilísticas basadas en terciles, respectivamente).

Por último, en la Parte IV se resumen las principales conclusiones de la Tesis. Además, se enumeran algunos de los logros conseguidos (principalmente publicaciones) y se da una breve perspectiva de las líneas futuras de trabajo a seguir.

1.3 Principales Resultados

A continuación se exponen brevemente los principales resultados obtenidos en cada uno de los capítulos que conforman esta Tesis.

Capítulo 3: Predicción Estacional

Además de presentar las bases de la predicción estacional y dar una visión general del estado del arte en este campo, en este capítulo también se incluyen una serie de contribuciones personales que se comentan brevemente a continuación:

- Se calcularon las teleconexiones observadas entre El Niño/La Niña y la precipitación a nivel mundial, obteniéndose resultados similares a los de otros estudios previos (véase, por ejemplo, Ropelewski and Halpert, 1987; van Oldenborgh et al., 2000; Mason and Goddard, 2001; Kayano et al., 2009; Shaman and Tziperman, 2011; Zhang et al., 2012; Yadav et al., 2013; Zhang et al., 2013). En particular, se encuentran importantes patrones de teleconexión, especialmente en los trópicos.
- Se calculó la correlación entre la SST observada y la simulada por los modelos de ENSEMBLES en la región El Niño 3.4, encontrando valores muy altos para todas las inicializaciones y estaciones, con la excepción de la parte central del año (aproximadamente entre Mayo y Agosto) para la inicialización de Febrero, lo cual está de acuerdo con la barrera de predecibilidad de primavera para el ENSO, que ha sido documentada en estudios previos (véase, por ejemplo, Zheng and Zhu, 2010; Tippett et al., 2011; Yan and Yu, 2012; Duan and Wei, 2013). Estos resultados ponen de manifiesto la predecibilidad del fenómeno ENSO.
- Se calcularon los patrones mundiales de *bias* y *drift* (ver el capítulo para más detalles) de la precipitación de los modelos de ENSEMBLES para cuatro meses ilustrativos (Febrero, Mayo, Agosto y Noviembre) —dada la configuración de los modelos considerados (ver Weisheimer et al., 2009, para más detalles), esta selección de meses permite ver la evolución del *drift* a lo largo de la simulación.— En general, tanto el *bias* como el *drift* se encuentran principalmente en los trópicos y son más importantes sobre los océanos que sobre tierra. Hay que destacar que el *drift* es tan importante como el *bias* para ciertas regiones y modelos. Sin embargo, mientras que el primero se corrige de forma rutinaria en la actualidad, rara vez hasta la fecha se ha tratado la corrección de este último (a pesar de su presencia sistemática en las predicciones estacionales actuales).
- Basándose en la metodología propuesta por Weisheimer and Palmer (2014) —la cual es ligeramente modificada en esta Tesis (ver el capítulo para más detalles),— se calculó la *reliability* de las predicciones probabilísticas de precipitación (se consideraron terciles) para las versiones de 15 y 51 miembros del modelo ECMWF System 4 sobre las 21 regiones de tierra definidas en Giorgi and Francisco (2000), para el período 1981-2010. Los resultados encontrados fueron mejores para la versión de 51 miembros. Sin embargo, se vio que las probabilidades predichas se concentran en torno a $1/3$ (el valor climatológico esperado) en este caso, lo cual da lugar a una mejor *reliability* puesto que, como consecuencia de los pesos aplicados, la pendiente de la recta de *reliability* (que pasa siempre por el punto de intersección climatológico) no

se ve tan afectada por errores en otros tramos de probabilidad.

Capítulo 4: El Caso de Estudio Regional: Filipinas

Estas son algunas de las razones que sustentan la elección de Filipinas como caso de estudio para gran parte del desarrollo de esta Tesis:

- Filipinas es un archipiélago de más de 7000 islas, con una topografía compleja y varios climas diferentes en un área relativamente pequeña, por lo que resulta un banco de pruebas ideal para estudios de SD (Moron et al., 2009). Además, el clima de la región se ve fuertemente influenciado por el fenómeno ENSO (Koide et al., 2012) y los monzones del suroeste (verano) y noreste (invierno) (Wang, 2002), siendo uno de los países más propensos a desastres naturales del mundo (Benson, C., 1997; Israel and Briones, 2012). Por estas razones, el país podría beneficiarse en gran medida de la aplicación de técnicas de SD para la regionalización de las predicciones estacionales.
- Se validó la precipitación global de los modelos de ENSEMBLES (predicciones deterministas) a nivel regional, tomando como referencia la precipitación observada en las 42 estaciones PAGASA y se vio que los *biases* son en general fuertes (por encima de 1000 mm/estación en muchos casos) y varían sensiblemente entre localidades cercanas. Además, la calidad de las predicciones es particularmente mala en verano, cuando el monzón del suroeste tiene lugar y se producen los impactos socioeconómicos más graves. Para superar estas limitaciones, el uso de técnicas de SD se hace necesario. Sin embargo, estas técnicas se han desarrollado casi exclusivamente para los extratropicos —hay múltiples problemas que todavía dificultan su éxito en los trópicos (Hewitson et al., 2014), — por lo que su aplicación en Filipinas resulta un reto.

Capítulo 5: Métodos de Regionalización Estadística

Con el fin de optimizar su configuración y compararlas entre sí, las distintas técnicas de SD que se describen en este capítulo se aplicaron sobre cuatro de las estaciones PAGASA —representativas de los distintos climas de Filipinas (ver Sección 4.1)— en condiciones ‘perfectas’, utilizando una validación cruzada para el período 1981-2005 (ver capítulo para más detalles). Estos son los resultados más importantes que se obtuvieron:

- Se consideraron la técnica de análogos y los GLMs como representativos de la metodología PP (ver capítulo para más detalles). En cuanto a la técnica de análogos,

se utilizó una versión determinista que tiene en cuenta sólo el análogo más cercano y otra estocástica en la que la predicción se obtiene sorteando al azar entre las observaciones correspondientes a los 15 análogos más cercanos. El comportamiento de los dos métodos es muy similar en cuanto a la reproducción de distribuciones —en particular, ambos muestran un déficit de días de lluvia predichos (lo que da lugar a leve *bias* seco) y subestiman ligeramente la varianza.— Sin embargo, la versión determinista produce mejores correlaciones que la estocástica y es por tanto la única considerada en esta Tesis Doctoral.

En cuanto a los GLMs, se consideró una configuración determinista en la que la predicción se obtiene a partir los valores esperados estimados y otra estocástica en la que se introduce un procedimiento de simulación (ver Sección 5.4.1-2 para más detalles). El método determinista no es capaz de predecir precipitaciones pequeñas y además subestima en gran medida la varianza —la mayoría de los valores predichos se concentran en un rango muy pequeño,— lo que da lugar a un *bias* húmedo (superior al 10%) y una fiabilidad distribucional muy deficiente. Por el contrario, el *bias* está centrado alrededor de cero y la varianza predicha se asemeja mucho más a la observada en la versión estocástica —simular permite reproducir todo el rango de valores de precipitación,— mejorando en gran medida la fiabilidad distribucional. Sin embargo, las correlaciones decaen fuertemente como efecto de la componente estocástica introducida, que implica una reducción en la capacidad predictiva del método. En base a su mejor rendimiento en términos de distribuciones, se puede presumir que los GLM estocásticos son necesarios para SD de proyecciones de cambio climático. Sin embargo, en la predicción estacional es clave evaluar el *accuracy*, por lo que es importante aislar la señal determinista de la componente estocástica. Por tanto, los GLM considerados en esta Tesis Doctoral son deterministas. En comparación con los métodos de análogos, las correlaciones mostradas por el GLM determinista son claramente mejores.

- Con respecto a la metodología MOS-BC (ver capítulo para más detalles), se consideraron dos métodos comúnmente utilizados para la corrección del *bias* del tipo *q-q mapping*, uno de ellos paramétrico y el otro empírico. Se vio que ambos (especialmente el empírico) mejoran la baja fiabilidad distribucional mostrada por la precipitación de ERA-Interim. Sin embargo, las correlaciones obtenidas (muy parecidas para los dos métodos) son peores que las correspondientes al reanálisis. Aún así, éstas son más altas que las mostradas por los métodos de análogos y muy parecidas a las obtenidas para el GLM determinista (a excepción de la correlación interanual, que es mejor en este último). En cuanto a fiabilidad distribucional, el *bias* y la va-

rianza predicha (especialmente esta última) son mejores en el método empírico. El paramétrico sobrestima claramente la varianza, lo cual podría reflejar una limitación intrínseca para simular precipitaciones extremas (este método supone que tanto la precipitación observada como la simulada se ajustan a una distribución gamma, lo cual podría no ser cierto). Por su parte, los buenos resultados ofrecidos por el método empírico podrían deberse a cierto sobreajuste (*over-fitting*).

Capítulo 6: Aspectos Metodológicos para la Regionalización Estadística de Predicciones Estacionales

Centrándose en Filipinas, se analizaron una serie de aspectos relevantes para el SD de las predicciones estacionales. Como resultado de las conclusiones que se exponen a continuación, en este capítulo se establece el marco metodológico bajo el cual se aplican todos los métodos de SD que se consideran en la Tesis:

- Aplicando un método PP ilustrativo en condiciones ‘perfectas’ sobre las 42 estaciones PAGASA (en modo validación cruzada para el período 1981-2005) se encontró que la combinación de variables de circulación (velocidad del viento zonal en diferentes niveles verticales) y termodinámicas (humedad y temperatura) sobre un dominio que cubre el país proporciona los mejores resultados, mientras que el potencial predictivo es menor si se utilizan sólo variables de circulación. Por tanto, se considera la combinación de predictores P4 (ver Tabla 6.1) para todos los métodos PP que se aplican en la Tesis.
- Se evaluaron las diferencias entre utilizar para la fase de calibración sólo datos correspondientes a la propia estación a predecir (dato estacional) o todo el conjunto de datos disponibles (dato anual). Para dos configuraciones diferentes de un método PP basado en GLMs (aplicado en condiciones ‘perfectas’ en modo validación cruzada para el período 1981-2005) se encontró que el dato estacional produce mejores resultados que el anual, sobre todo para la parte central del año. En cambio, se vio que los métodos MOS-BC no son sensibles al tipo de dato (estacional o anual) usado para la calibración. En consecuencia, todos los SDMs considerados en esta Tesis se calibran utilizando el dato estacional.
- Se calculó el *bias* de los modelos de ENSEMBLES para los predictores de larga escala incluidos en P4 (ver Tabla 6.1) sobre el dominio usado para SD (Figura 6.1), encontrándose en general errores importantes. Además, más allá de las diferencias esperadas entre modelos, en algunos casos los patrones espaciales obtenidos varían

considerablemente de un mes a otro (dentro de la misma estación), lo que obliga a llevar a cabo un preproceso (o *harmonización*) adecuado de estas variables —que las haga compatibles con el reanálisis utilizado para la calibración— antes de entrar en cualquier método PP (ver capítulo para más detalles).

Capítulo 7: Validación de Predicciones Estacionales Globales de Precipitación: El Papel del Fenómeno ENSO en la Calidad de las Mismas

- Se validaron las predicciones estacionales globales de precipitación de los modelos de ENSEMBLES (se tomó como referencia una rejilla mundial de observaciones) para el período 1961-2000, aplicando una metodología probabilística para terciles en términos del ROC Skll Score (ROCSS). A pesar de que la predecibilidad encontrada varía en función de la región, la estación y el *lead-time*, se encontró que 1) los trópicos son la región donde las predicciones tienen mayor calidad, 2) en general, el invierno (la primavera) es la estación en la que la calidad de las predicciones es mayor (menor) y 3) la calidad de las predicciones se debilita a medida que aumenta el *lead-time*, aunque los patrones espaciales de ROCSS se conservan (sobre todo sobre el norte de Sudamérica y el archipiélago Malayo).
- Para estudiar el papel que el fenómeno ENSO tiene sobre la predecibilidad encontrada a nivel mundial, se llevó también a cabo una validación condicionada (restringida a eventos El Niño y La Niña) y se hizo un estudio de las teleconexiones de El Niño/La Niña con la precipitación (en base a la frecuencia de ocurrencia de cada tercil en los eventos Niño/Niña). Los resultados obtenidos indican que la distribución espacial y temporal de la predecibilidad estacional no sólo está determinada por el efecto directo de ENSO —y por lo tanto por la habilidad de los modelos para predecir la SST en la región El Niño 3.4— sino más bien por su efecto indirecto a través de las teleconexiones asociadas con el Niño y La Niña —y por lo tanto, limitada por la capacidad de los modelos para simular los patrones de teleconexión observados.—
- Siguiendo la misma metodología basada en terciles que se utilizó a nivel mundial, se realizó también una validación probabilística regional sobre Filipinas (tomando como referencia en este caso la precipitación observada en las 42 estaciones PAGASA) para el período 1981-2005. Los resultados varían en función del modelo y la zona del país considerada. Sin embargo, de acuerdo con la validación determinista llevada a cabo en el Capítulo 4, se encuentran ROCSS aceptables en todas las estaciones (especialmente en invierno y primavera) menos en verano, cuando se producen los impactos socioeconómicos más graves como consecuencia del monzón del suroeste.

- También se hizo un estudio regional de las teleconexiones de El Niño/La Niña con la precipitación observada en las 42 estaciones PAGASA para el período 1981-2005 (siguiendo la misma metodología utilizada a nivel mundial). En consonancia con los resultados obtenidos previamente por Lyon et al. (2006), los resultados de este análisis ponen de manifiesto la fuerte influencia del fenómeno ENSO sobre el clima de Filipinas (Koide et al., 2012), dejando lluvias por debajo (por encima) de lo normal en los eventos Niño (Niña), excepto en verano, cuando la señal se invierte (especialmente en los eventos Niño). Las teleconexiones más débiles tienen lugar en verano, la estación en la que la calidad de las predicciones de los modelos es peor.

Capítulo 8: El Efecto de la Incertidumbre en el Reanálisis sobre las Predicciones Regionalizadas

- Una comparación entre los reanálisis de referencia ERA-Interim y JRA-25 sobre un área extensa que incluye Filipinas reveló que mientras las diferencias son mínimas para la circulación (vientos zonales), existe una gran incertidumbre para las variables termodinámicas (humedad y temperatura). Por tanto, la hipótesis de que los predictores de reanálisis reflejan las condiciones atmosféricas ‘reales’ a gran escala no se cumple para la región considerada.
- Se consideró un método PP ilustrativo para evaluar el efecto que tiene en las predicciones regionalizadas la elección del reanálisis considerado para la calibración. En primer lugar, en modo validación cruzada para el período 1981-2005, se utilizaron por separado los reanálisis ERA-Interim y JRA-25 para obtener los coeficientes de regresión que relacionan la precipitación observada en las 42 estaciones PAGASA con los predictores de larga escala incluidos en P4 (ver Tabla 6.1), los cuales fueron seleccionados en base a los resultados del Capítulo 6. Se vio que los resultados son sensibles al reanálisis utilizado si la humedad y/o la temperatura —las variables que muestran la mayor incertidumbre— se incluyen en el campo predictor. Sin embargo, a nivel puntual (en promedio espacial), las diferencias en la correlación diaria son como máximo de 0.1 (0.03), por lo que esta sensibilidad parece ser pequeña en condiciones ‘perfectas’. Posteriormente, los coeficientes obtenidos de la calibración con ERA-Interim y JRA-25 se aplicaron por separado a predictores del modelo MPI-ECHAM5 y de los modelos de ENSEMBLES, con el fin de generar proyecciones de cambio climático (hasta final de siglo) y predicciones estacionales (para el período 1981-2005) locales, respectivamente. En el primer caso, las diferencias inducidas por el reanálisis que se detectan en condiciones ‘perfectas’ (en clima presente) se amplifican considerablemente cuando se incluyen entre los predictores la humedad y/o la

temperatura, las cuales son indispensables para capturar la señal correcta de cambio climático (Goodess and Palutikof, 1998; Wilby et al., 1998). En particular, las *deltas* proyectadas para el final del siglo (2071–2100 menos 1981–2000) difieren en más de un 35% (en promedio para todo el país) entre los dos reanálisis. Por el contrario, en el segundo caso, las predicciones estacionales regionalizadas son muy similares para la mayoría de los modelos, las estaciones y las zonas del país, independientemente del reanálisis utilizado. Estos resultados demuestran que la elección del reanálisis considerado para la calibración de los métodos PP es una fuente de incertidumbre (error) importante para estudios de cambio climático, mientras que no es de especial relevancia en el contexto de la predicción estacional —algo que, hasta nuestro conocimiento, no había sido determinado antes.—

Capítulo 9: Ventajas y Limitaciones de los Métodos de Regionalización Estadística para Predicciones Estacionales

- Se aplicaron dos métodos PP y otros dos MOS-BC a los modelos de ENSEMBLES y las predicciones regionalizadas obtenidas se compararon con las correspondientes salidas directas sobre las 42 estaciones PAGASA para el período 1981-2005, considerando métricas de verificación que caracterizan el *accuracy* y la *reliability* (para las predicciones deterministas y probabilísticas basadas en terciles, respectivamente). En general, los resultados obtenidos varían en función de la estación del año, pero también de la región, el modelo y el método de SD escogidos. Además, son más sensibles a la metodología considerada (MOS-BC o PP) que al SDM usado (para una misma metodología).
- En términos de *accuracy*, ni los métodos MOS-BC ni los PP dan lugar a mejoras relevantes (con respecto a la precipitación dada por el modelo) ni en invierno ni en primavera, lo que sugiere que el valor añadido que puede obtenerse por medio de técnicas SD es limitado para aquellos casos en los que los modelos ya simulan satisfactoriamente la precipitación. Sin embargo, mientras que los métodos MOS-BC no mejoran claramente (o incluso empeoran) la precipitación de los modelos en verano y en otoño, los métodos PP proporcionan en general resultados mejores (peores) que sus salidas directas en verano (otoño). En particular, éstos últimos dan lugar a importantes mejoras de *accuracy* en verano en la parte noroeste del país. Los resultados obtenidos en términos de *reliability* son muy similares a los obtenidos para el *accuracy*, lo que pone de manifiesto la idoneidad de la metodología propuesta por Weisheimer and Palmer (2014) —la cual se modifica ligeramente en este Tesis— para estudios a nivel regional (ver capítulo para más detalles).

- Para los métodos PP se encontró que los casos en los que se mejora (empeora) la calidad —entendida como *accuracy* y *reliability*— de la salida directa, los modelos reproducen mejor (peor) las variables predictoras de larga escala, las cuales se definen sobre un dominio sinóptico, que la precipitación, la cual se ve más afectada por distintos efectos locales. Esto sugiere que los métodos PP podrían aprovechar la habilidad de los modelos para reproducir la larga escala para indirectamente obtener mejores (en comparación con la salida directa) predicciones locales de precipitación.

1.4 Publicaciones Relacionadas

Parte de los resultados centrales de esta Tesis Doctoral (Parte III) han dado lugar a publicaciones en revistas científicas de reconocido prestigio en el campo de las ciencias atmosféricas. En particular:

- El Capítulo 7 se basa en los resultados de “Manzanas, R., M. D. Frías, A. S. Cofiño, and J. M. Gutiérrez, 2014: Validation of 40 year multimodel seasonal precipitation forecasts: The role of ENSO on the global skill. *Journal of Geophysical Research: Atmospheres*, **119** (4), 1708–1719, doi:10.1002/2013JD020680.”
- El Capítulo 8 se basa en los resultados de “Manzanas, R., S. Brands, D. San-Martín, A. Lucero, C. Limbo, and J. M. Gutiérrez, 2015: Statistical downscaling in the tropics can be sensitive to reanalysis choice: A case study for precipitation in the Philippines. *Journal of Climate*, **28** (10), 4171–4184, doi:10.1175/JCLI-D-14-00331.1.”
- El Capítulo 9 se basa en los resultados de “Manzanas, R., J. M. Gutiérrez, and A. Lucero, 2016: Can statistical downscaling and bias correction methods improve the accuracy and reliability of seasonal forecasts? Enviado a *Climate Dynamics*.”

Además, en paralelo al desarrollo de la Tesis han surgido otras publicaciones cuyos resultados no han sido incluidos aquí:

- Haciendo uso de parte del conocimiento metodológico adquirido, algunos de los métodos de *downscaling* utilizados en esta Tesis han sido aplicados sobre España en “San-Martín, D., R. Manzanas, S. Brands, S. Herrera, and J. M. Gutiérrez, 2016: Reassessing model uncertainty for regional projections of precipitation with an ensemble of statistical downscaling methods. En revisión en *Journal of Climate*.”
- Manzanas, R., 2016: Can statistical downscaling improve the skill of global seasonal forecasts in Senegal? En revisión en *Theoretical and Applied Climatology*.

- *Manzanas, R.*, L. K. Amekudzi, K. Preko, S. Herrera, and J. M. Gutiérrez, 2014: Precipitation variability and trends in Ghana: An intercomparison of observational and reanalysis products. *Climatic Change*, **124** (4), 805–819, doi:10.1007/s10584-014-1100-9.
- Gutiérrez, J. M., D. San-Martín, S. Brands, *R. Manzanas*, and S. Herrera, 2013: Reassessing statistical downscaling techniques for their robust application under climate change conditions. *Journal of Climate*, **26** (1), 171–188, doi:10.1175/JCLI-D-11-00687.1.
- Brands, S., *R. Manzanas*, J. M. Gutiérrez, and J. Cohen, 2012: Seasonal predictability of wintertime precipitation in Europe using the Snow Advance Index. *Journal of Climate*, **25** (12), 4023–4028, doi:10.1175/JCLI-D-12-00083.1.

1.5 Software Desarrollado: *MeteoLab*

Para la mayoría de los cálculos realizados en esta Tesis se ha utilizado *MeteoLab*, una *toolbox* de *Matlab*[®] para *downscaling* estadístico desarrollada por el Grupo de Meteorología de Santander que puede descargarse libremente desde <http://meteo.unican.es/trac/MLToolbox>. Mientras que la técnica de análogos y las de corrección del *bias* (ver Capítulo 5) ya estaban implementadas en dicha *toolbox*, los métodos GLMs han sido desarrollados en esta Tesis y se encuentran disponibles en la versión actual. A continuación se muestra el código necesario para reproducir con *MeteoLab* todos los métodos de *downscaling* que se han utilizado en esta Tesis (se sigue la nomenclatura introducida en el Capítulo 5). Para más detalles sobre las distintas opciones de configuración de los mismos se refiere al lector a <http://meteo.unican.es/trac/MLToolbox/wiki/Downscaling>.

```
method.type = 'ANALOGES';           % Metodo
method.properties.NumberOfPCs = 30; % Numero de PCs
method.properties.AnalogueNumber = 1; % Numero de analogos
method.properties.InferenceMethod = 'mean'; % Metodo de inferencia
```

Código *MeteoLab* para definir el método *AN_det*.

```
method.type = 'ANALOGES';           % Metodo
method.properties.NumberOfPCs = 30; % Numero de PCs
method.properties.AnalogueNumber = 15; % Numero de analogos
method.properties.InferenceMethod = 'rand'; % Metodo de inferencia
```

Código *MeteoLab* para definir el método *AN_sto*.

```

method.type = 'GLM'; % Metodo
method.properties.ThresholdPrecip = 0.1; % Umbral dias humedos (mm)
method.properties.NumberOfPCs = 15; % Numero de PCs
method.properties.SimOccurrence = 'false'; % No simular ocurrencia
method.properties.SimAmount = 'false'; % No simular cantidad

```

Código *MeteoLab* para definir el método *GLM_det*.

```

method.type = 'GLM'; % Metodo
method.properties.ThresholdPrecip = 0.1; % Umbral dias humedos (mm)
method.properties.NumberOfPCs = 15; % Numero de PCs
method.properties.SimOccurrence = 'true'; % Simular ocurrencia
method.properties.SimAmount = 'true'; % Simular cantidad

```

Código *MeteoLab* para definir el método *GLM_sto*.

```

method.type = 'GQM'; % Metodo
method.properties.Variable = 'pr'; % Variable
method.properties.threshold = 0.1; % Umbral dias humedos (mm)
method.properties.FreqCorrection = 'true'; % Correccion de frecuencias

```

Código *MeteoLab* para definir el método *QM_par*.

```

method.type = 'EQM'; % Metodo
method.properties.Variable = 'pr'; % Variable
method.properties.threshold = 0.1; % Umbral dias humedos (mm)
method.properties.extrapolation = 'constant'; % Tipo de extrapolacion
method.properties.quantiles = 1:99; % Percentiles corregidos
method.properties.FreqCorrection = 'true'; % Correccion de frecuencias

```

Código *MeteoLab* para definir el método *QM_emp*.

1.6 Líneas Futuras de Trabajo

Por un lado, algunos de los resultados obtenidos durante la realización de esta Tesis han abierto la puerta para el desarrollo de nuevos trabajos que constituyen una continuación natural de algunos de los análisis presentados en esta memoria:

- En la Sección 3.4.1-2 se introdujo una caracterización del *drift* de las predicciones estacionales, considerando para ello la precipitación de los modelos de ENSEMBLES. Para evaluar tanto la posible mejora del *drift* en los modelos globales recientes, como la sensibilidad del mismo al tamaño del *ensemble*, se ha extendido el estudio realizado al System 4 (la nueva versión del modelo del ECMWF incluido en ENSEMBLES, el

System 3). Los resultados preliminares indican que el *drift* sigue siendo importante, encontrándose valores especialmente altos en los trópicos y sobre los océanos. Sin embargo, *ensembles* pequeños (alrededor de 5 miembros) parecen ser suficientes para una caracterización precisa del mismo. Por motivos de tiempo, este trabajo no ha podido ser finalizado durante el transcurso de la Tesis y es una de las primeras actividades que se abordarán una vez finalizada la misma, poniendo especial interés en las implicaciones que puedan derivarse para la corrección del *bias* de las predicciones estacionales.

- Como se explicó en la Sección 3.5.1, en Weisheimer and Palmer (2014) se evaluó la *reliability* de las predicciones estacionales globales de precipitación del System 4 del ECMWF (51 miembros) para las 21 regiones de tierra definidas en Giorgi and Francisco (2000), y los resultados se presentaron en una escala que va desde 1 (peligrosa) a 5 (perfecta). Sin embargo, hemos detectado que esta clasificación es sensible a diversos factores. Por ejemplo, los resultados de la misma pueden verse alterados si el intervalo de confianza considerado para la pendiente de la recta de regresión que determina la *reliability* (obtenido por *bootstrapping*) no se adecua al tamaño del *ensemble* disponible. Asimismo, dichos resultados también pueden cambiar sustancialmente en función de la región de agregación considerada. Por tanto, se pretende analizar con más detalle todos estos factores, en colaboración con los autores de la metodología original.

Por otro lado, y en el marco de iniciativas y colaboraciones internacionales surgidas durante el desarrollo de la Tesis, se plantean también una serie de trabajos y líneas de investigación a desarrollar en el futuro próximo:

- Parte del conocimiento metodológico y las técnicas de *downscaling* desarrolladas durante esta Tesis han sido aplicadas a diferentes regiones dentro de los proyectos SPECS (<http://www.specs-fp7.eu>) y EUPORIAS (<http://euporias.eu>). En particular, en SPECS se ha llevado a cabo un experimento en el que se comparan *downscaling* dinámico y estadístico (incluyendo diferentes metodologías) en Brasil. Por su parte, en EUPORIAS se ha analizado el valor añadido de las predicciones estacionales regionalizadas en Etiopía (en el contexto de un servicio climático para la detección temprana de sequías) y en dos zonas piloto en Europa (Inglaterra e Italia). Durante los próximos meses se van completar todos estos estudios (por ejemplo, en el caso de Etiopía se ha visto que la incertidumbre en las observaciones puede tener efecto sobre las predicciones regionalizadas) y se van a recoger los principales resultados en una serie de publicaciones que, dada la heterogeneidad de las regiones

consideradas (en términos de variabilidad climática, predecibilidad estacional, efecto del ENSO, etc.), supondrán un marco de referencia para estudios de *downscaling* en el contexto de la predicción estacional.

- A raíz de una estancia de investigación en el Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI: <http://www.senamhi.gob.pe>), se dispone de un *dataset* de observaciones de alta calidad que cubre todo el país. Dada la compleja variabilidad climática del mismo (una notable diversidad geográfica hace que la mayoría de climas del mundo estén presentes), así como el fuerte efecto del fenómeno ENSO sobre la región, se plantea una línea futura de trabajo en la que se explorará, en colaboración con el SENAMHI y aplicando la experiencia y el conocimiento metodológico adquiridos durante el desarrollo de esta Tesis, el potencial de la aplicación de las distintas técnicas de *downscaling* para la predicción estacional en Perú.
- Por último, los métodos de *downscaling* desarrollados en esta Tesis han contribuido al *experimento 1a* de la acción COST VALUE (<http://www.value-cost.eu>), que tiene como objetivo realizar una intercomparación sistemática de las distintas metodologías y técnicas de *downscaling* estadístico en el contexto del cambio climático. Como resultado de esta iniciativa, se ha generado una base de datos con las predicciones de más de 40 métodos de *downscaling* diferentes (el *ensemble* más amplio hasta la fecha). Además, el Grupo de Meteorología de Santander ha desarrollado un portal que permite definir índices y medidas de validación y aplicarlos al conjunto de métodos disponibles. Por tanto, otra línea futura de trabajo será incluir nuevas métricas de verificación de interés para la escala temporal de la predicción estacional (como por ejemplo la correlación interanual), con el objetivo de evaluar el potencial de los distintos métodos en este nuevo contexto de aplicación.

CHAPTER 2

Context, Objectives and Structure of the Thesis

2.1 *Context*

Short-to-medium-range weather forecasts are a very popular and demanded product, both socially and commercially. These forecasts provide an estimate of the meteorological variables of interest (e.g. temperature or precipitation) for the next days, and are elaborated based on numerical models which simulate the evolution of the atmosphere from a given initial condition. To take into account the different uncertainty sources which affect these forecasts, a probabilistic methodology based on a number of predictions obtained from slightly different initial conditions (*ensemble forecasts*) has been developed during the last decades. Ensemble forecasts are the main tool for modern meteorological forecasting.

In parallel, important advances in the modelling of the atmosphere-ocean coupled circulation have allowed to develop a new generation of numerical models for seasonal forecasting. Differently from short-to-medium-range weather forecasts, seasonal forecasts do not aim to predict the temperature or precipitation at a given moment within a few days. Instead, seasonal forecasts provide information on the average conditions which can be expected for the next seasons (e.g. based on tercile anomalies: cold/normal/warm, dry/normal/wet), up to one year ahead. Predictability at this longer time-scale is mainly due to the slow dynamics of the ocean. Due to the existing computational limitations, the spatial resolution of current seasonal forecasts is still too coarse (around hundreds of km). Moreover, the probabilistic nature of these forecasts —not only the uncertainty in the initial conditions but also in the model formulation (multimodel predictions) is often considered— makes them difficult to use, since specific tools are needed for their suitable validation and postprocess.

Despite this, seasonal forecasts have a great number of applications and help decision-making in many important socio-economic sectors such as agriculture, energy, transport, health and hydrology. For this reason, their use has been continuously increasing during the last decades. Among the most commonly demanded variables (temperature, precipitation and wind), precipitation is the most problematic due to its mixed (occurrence/amount) character and high spatial variability (see, e.g., Schmidli et al., 2007; Bundel et al., 2011). Therefore, this Thesis focuses exclusively on the latter variable.

Differently from short-to-medium-range weather forecasts, current numerical models for seasonal forecasting are not globally reliable. Instead, for each variable of interest and season of the year, their usefulness is limited to certain regions of the world, mainly the tropics (Stockdale et al., 1998). Moreover, the low spatial resolution of these models is insufficient for most of practical applications so their raw outputs can not be directly used for impact studies. As a result, proper postprocess is needed to translate the coarse global seasonal forecasts to the useful local-scale (*downscaling*).

To this aim, two conceptually different approaches have been developed in the last decades: dynamical and statistical downscaling. Whereas dynamical downscaling is based on the use of regional models which, driven at the boundaries by the coarse-resolution outputs from the global models, numerically simulate the climate at a higher resolution over a limited area (see, e.g., Giorgi and Mearns, 1999; Laprise, 2008), Statistical Downscaling (SD) relies on statistical models/algorithms which link the coarse-resolution outputs from the global models (*predictors*) with the local observations (*predictands*) over the area of interest (see, e.g., von Storch et al., 1993). In this Thesis, only the latter, which has demonstrated potential to improve the skill of global seasonal precipitation forecasts (Feddersen and Andersen, 2005) and is drastically cheaper in terms of computational resources, is considered. Nevertheless, whereas SD has been extensively used for climate change modelling, there is only limited experience regarding its application for seasonal forecasts (see, for instance, the results from deliverable D52.1 of the SPECS project: *Review of the different statistical downscaling methods for s2d prediction*). Moreover, SD techniques have been mostly developed and applied for extratropical regions since manifold problems still hinder their successful application in the tropics (Hewitson et al., 2014), the region with the highest seasonal predictability.

2.2 Objectives

Based on the previous considerations, this Thesis poses the following three main objectives:

1. To assess the skill of global seasonal precipitation forecasts and to study the effect

that *El Niño-Southern Oscillation* (ENSO) —the main driver of seasonal predictability (see, e.g., Goddard and Dilley, 2005; Doblas-Reyes et al., 2010)— may have on it.

2. To methodologically adapt the existing approaches and techniques for SD for their suitable application in the context of seasonal forecasting, taking into account a number of factors which have never been considered before, such as the effect of reanalysis uncertainty on the downscaled predictions.
3. To analyze the advantages and limitations of the different approaches and techniques for SD in the context of seasonal forecasting, considering the proper verification metrics which allow to assess the (possible) added value of local downscaled precipitation forecasts with respect to the raw model global outputs.

To achieve these goals, a great part of the Thesis focuses on a particularly interesting region located in the tropics: the Philippines. The reasons that support this choice are detailed in Chapter 4.

2.3 Structure

The Thesis is organized in four main blocks: an introductory block (Part I) which is followed by a methodological block (Part II), a central block in which the main results are presented (Part III) and a final block (Part IV) in which the main conclusions obtained are given.

In Part I there is an opening chapter (Chapter 3) which introduces seasonal forecasting (as well as many of the concepts which will be later used throughout the Thesis) and provides a general overview of the state-of-the-art in this field. In addition, some personal contributions (see the chapter for details) are also presented.

Next, in Chapter 4, the regional case study considered, the Philippines, is introduced and its appropriateness for the development of the Thesis is discussed.

Part II is formed by two methodological chapters. Chapter 5 introduces the theoretical background needed on the different approaches and techniques available for SD and provides a detailed analysis of the Statistical Downscaling Methods (SDMs) considered through the rest of the Thesis.

Chapter 6 establishes the methodological framework followed for the application of the latter SDMs, covering different aspects —from the selection of predictors to the procedure followed to *harmonize* the model predictor data used.—

Most of the core results of the Thesis are given in Part III, which is structured around three key chapters. In Chapter 7 we carry out a user-oriented validation of global seasonal

precipitation forecasts for a forty-year period (1961-2000), which allows for robustly identifying those regions of the world with significant seasonal skill. Additionally, since ENSO is known to be the major driver of seasonal predictability, the role that this phenomenon plays on the skill found is also analyzed.

In Chapter 8, focusing on the Philippines, we assess the sensitivity of the downscaled results to the choice of reanalysis used for the calibration of the PP methods, both in ‘perfect’ and ‘non-perfect’ conditions (see the chapter for details). In the latter case, two different time-horizons are considered: seasonal forecasts in hindcast mode (for the period 1981-2005) and climate change projections (up to the end of the 21st century).

In Chapter 9 we analyze, also for the Philippines, whether SD can serve to improve the skill of the raw model global precipitation forecasts (beyond reducing their systematic biases). To this, and building on the lessons learnt from Chapters 6 and 8, we apply a number of SDMs representative of the different approaches for SD introduced in Chapter 5 and analyze their relative advantages and limitations by comparing the downscaled results against the corresponding raw model outputs for the period 1981-2005, focusing on accuracy and reliability aspects (for deterministic and tercile-based probabilistic predictions, respectively).

To conclude, Part IV summarizes the main conclusions obtained from the Thesis. Additionally, some of the achievements reached are enumerated and a brief outlook of future work is also given.

Part I

Introduction

CHAPTER 3

Seasonal Forecasting

Whereas short-to-medium-range weather forecasts focus on the daily weather expected over the next two weeks, seasonal forecasts provide information on how average seasonal weather conditions are likely to be from a few months up to one year in advance. These forecasts have a great number of applications for end-users in different socio-economic sectors such as agriculture, energy, health or hydrology, and help decision-making. Typically, temperature and precipitation are the main variables of interest for end-users since they are needed in most impact applications. Moreover, there exist a number of worldwide observational datasets for these variables, suitable to validate the performance of retrospective seasonal forecasts for the last few decades. Therefore, most of the studies carried out so far in this field have focused on these variables —together with Sea Surface Temperature (SST),— although there has been a recent interest in other variables such as wind speed or some climate-related indices, for instance the Fire Weather Index (FWI). This Thesis focuses exclusively on precipitation since previous studies have shown that this variable is less skillfully predicted than temperature over land areas (see, e.g., Kirtman and Pirani, 2009; Barnston et al., 2010; Doblas-Reyes et al., 2010; Bundel et al., 2011) and its downscaling is much more troublesome (see, e.g., Murphy, 1999; Schoof and Pryor, 2001; Schmidli et al., 2007).

This chapter presents the basics of seasonal forecasting, introducing many of the concepts that will be used later along the Thesis, and provides a brief overview of the state-of-the-art in this field.

3.1 *Introduction to Seasonal Forecasting*

Weather changes on a daily basis. While modern forecasting systems can accurately predict weather events a few days into the future, daily predictions beyond two weeks are limited by the non-linear character of the atmosphere dynamics (the complexity and the strong sensitivity to the initial conditions of the processes involved make precise long-term forecasts impossible).

However, to some extent, it is possible to predict deviations from the mean seasonal climatology some months in advance. Seasonal forecasts do not aim to predict the timing of a particular weather event with any accuracy —e.g., how much is going to rain in Madrid on January 2nd;— however, the likelihood of receiving above-, near- or below-normal (i.e., tercile-based categories) precipitation next winter can be predicted in some cases and for some regions, particularly in the tropics.

Much of the skill in predicting departures from normal seasonal averages is derived from some of the slowly varying components of the global climate system, especially the SST, whose slow fluctuations can influence the weather around the world, providing thus a source of predictability at this time-scale. These effects are not easily noticed in day-to-day weather events but become evident in long-term weather averages.

Seasonal forecasting originated in the early 1990s. Initially, operational forecasts were entirely based on empirical/statistical models (typically lagged linear regression methods) which allowed to predict mean seasonal changes based on SST anomalies, several months in advance. However, during the end of the 1990s, atmosphere-ocean coupled models, and later Global Climate Models (GCMs¹) including other climate components such as the soil and the oceanic-ice, started to be used for operational seasonal forecasting. Nowadays, both empirical/statistical and dynamical approaches are used and are often combined.

3.2 *Empirical Seasonal Forecasting*

Anomalies on large-scale, slowly varying variables such as SST and/or sea-ice cover may force changes in atmospheric circulation patterns and hence departures from normal seasonal weather throughout the world (*atmospheric teleconnections*). Until the recent development of GCMs, seasonal forecasts were almost exclusively based on empirical/statistical methods which exploited these teleconnections by relating the observed seasonal rainfall anomalies with observed lagged SST anomalies, typically by means of multivariate linear regression schemes (see, e.g., Nicholls, 1984; Hastenrath, 1995).

¹In this Thesis we will refer to all global dynamical climate models as GCMs, independently of their complexity, i.e., the previous Atmosphere-Ocean Coupled Models (AOCMs) or the modern Earth System Models (ESMs).

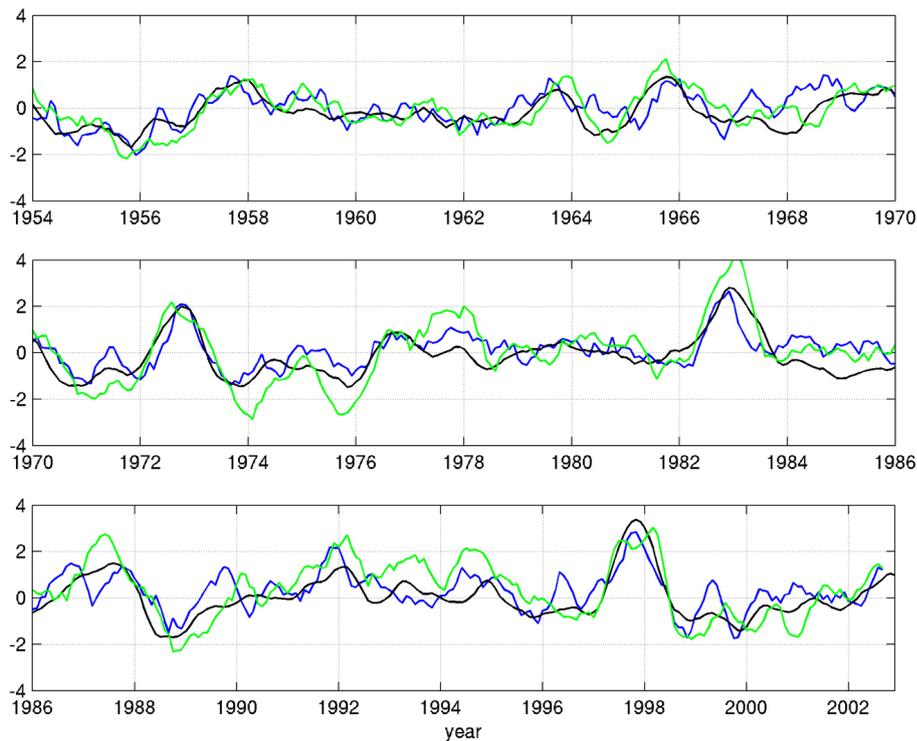


Figure 3.1: Global precipitation anomaly from the extended TRMM-combined dataset (<http://trmm.jpl.nasa.gov/2b31.html>) in blue, the Southern Oscillation Index (SOI: <http://www.ncdc.noaa.gov/teleconnections/enso/indicators/soi>) in green and the Niño-3 index (http://www.cgd.ucar.edu/cas/catalog/climind/Nino_3_3.4_indices.html) in black. Extracted from Haddad (2004).

The strongest links between SST patterns and seasonal weather are those associated with El Niño-Southern Oscillation (ENSO), which is known to be the dominant mode of seasonal variability (Doblas-Reyes et al., 2010). As an illustrative example, Figure 3.1 shows the interannual anomalies of ENSO (as represented by the Southern Oscillation Index and the Niño-3 index) and global precipitation (obtained from the TRMM-extended dataset), which exhibit similar fluctuation patterns, modulated by the extreme ENSO episodes (El Niño and La Niña years). At a regional level, ENSO can disrupt the normal pattern of weather around the globe, leading to large changes in seasonal rainfall that can bring droughts in some regions and floods in others. Figure 3.2 shows the regional teleconnections between El Niño and precipitation, calculated following a tercile-based approach with VASCLimO v1.1² observations, in terms of the frequencies of occurrence of below-, near- or above- normal precipitation categories in El Niño periods. For instance,

²VASCLimO v1.1 (Beck et al., 2005) is a gauge-based dataset providing monthly precipitation totals on a 2.5° resolution grid for the global land areas (except the Antarctica) for the period 1951-2000.

red color for the dry tercile (left column) indicates those grid boxes where dry conditions are dominant in *El Niño* years. This figure illustrates the potential of simple empirical seasonal forecasting, especially in the tropics, the region exhibiting the strongest teleconnections. Noteworthy, the study of ENSO teleconnections is an active research area and a number of studies have focused on the effects of this phenomenon on precipitation (see, e.g., Ropelewski and Halpert, 1987; van Oldenborgh et al., 2000; Mason and Goddard, 2001; Kayano et al., 2009; Shaman and Tziperman, 2011; Zhang et al., 2012; Yadav et al., 2013; Zhang et al., 2013).

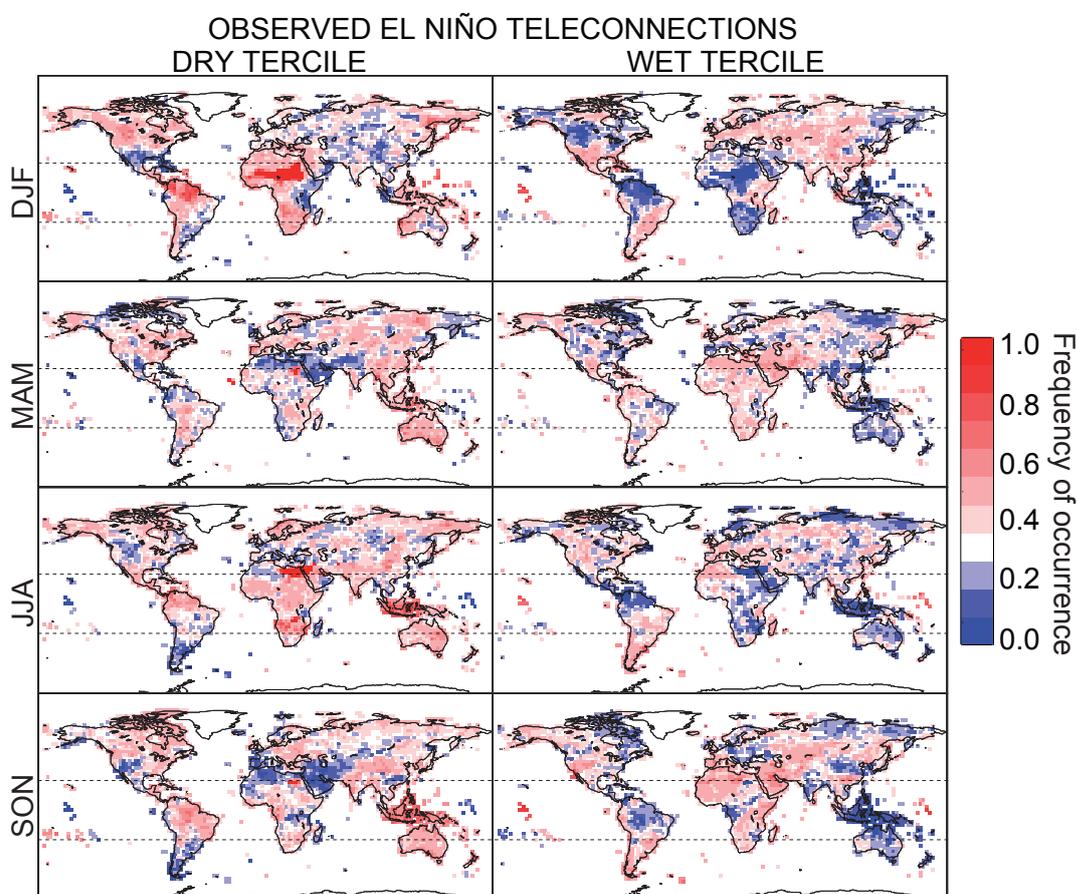


Figure 3.2: Relative frequency of occurrence for the dry and wet terciles (left and right columns, respectively) in the eleven *El Niño Winter* years defined in Pozo-Vázquez et al. (2005) —1964, 1966, 1970, 1973, 1977, 1983, 1987, 1988, 1992, 1995 and 1998— within the period 1960-2000, for the different seasons (in rows). Red (blue) colors correspond to values above (under) $1/3$, the expected climatological frequency. VASClimO v1.1 was the observational dataset considered. Dashed lines indicate the tropics/extratropics division.

Apart from SST, other well-known climate indices representing major climatic features such as the North Atlantic Oscillation (NAO) may be used to estimate their associated regional impacts. For instance, based on the Snow Advance Index (SAI) (Cohen and Jones, 2011) —an index describing the Eurasian snow cover increase in October,— Brands et al. (2012) developed a simple statistical method to predict winter (December-January-February) precipitation totals for the Iberian Peninsula and Norway. Likewise, Eden et al. (2015) presented recently a new empirical system based on multiple linear regression for producing probabilistic forecasts of seasonal surface air temperature and precipitation across the globe, taking as primary predictor the global CO_2 equivalent concentration.

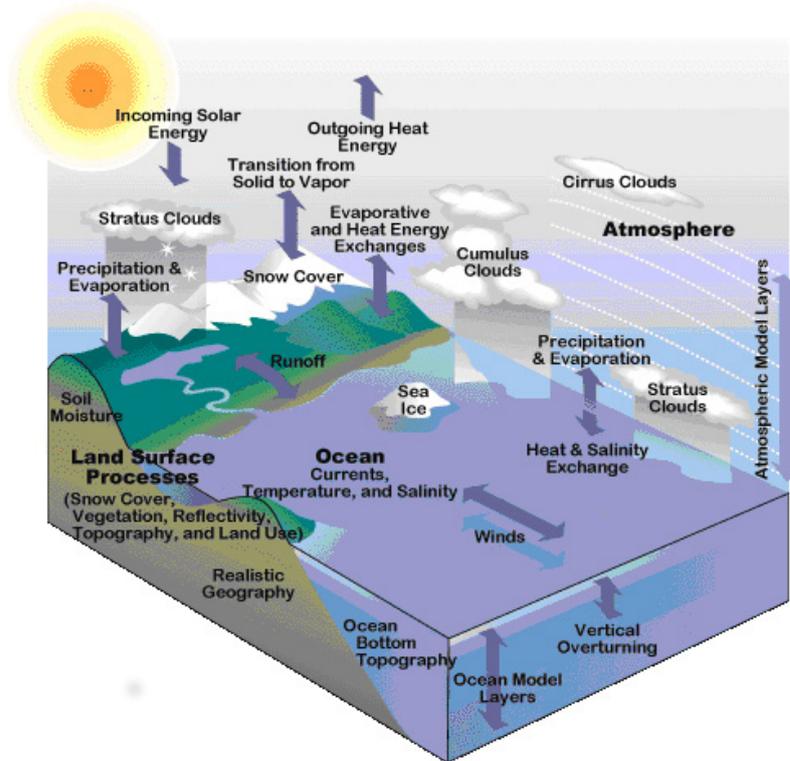


Figure 3.3: Diagram illustrating some of the interactions between the different components of the climate system. Source: UCAR (<http://www2.ucar.edu>).

3.3 Dynamical Seasonal Forecasting: Global Climate Models (GCMs)

Global Climate Models (GCMs) have their origin in numerical weather prediction and are based on the general physical principles of fluid dynamics and thermodynamics. In particular, they solve complicated differential equations that describe the processes that occur in the different components of the global climate system (e.g. the atmosphere and the oceans) and their interactions (see Figure 3.3).

These numerical computations are performed in a discretized three-dimensional space formed by grid boxes (see Figure 3.4) whose typical horizontal resolution is about hundreds of kilometres. This implies a number of simplifying assumptions and requires parametrizing some processes which occur at spatial scales smaller than the grid box size (e.g., convection), using to this aim semi-empirical approaches which change from model to model. At the interfaces, the atmosphere is coupled to the land and oceans through exchanges of heat, moisture and momentum. To this, all GCMs contain some form of coupler module, which tie all components together, interpolating fluxes and controlling their interactions.

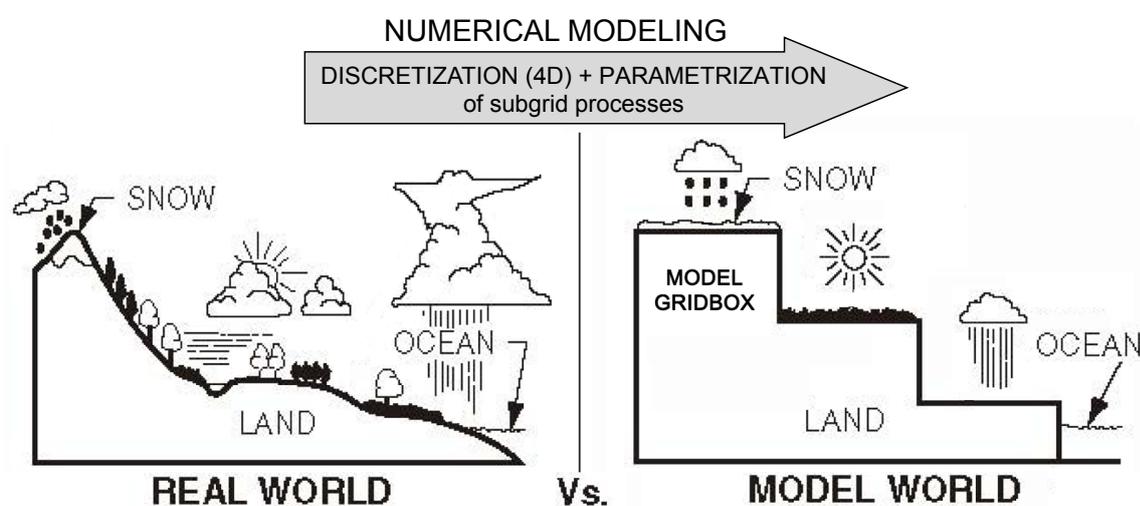


Figure 3.4: Schema of the discretization process followed in a GCM to solve the equations describing the climate system.

As the scientific understanding and the available computational resources increased, GCMs successively became more and more complex. Nowadays, GCMs are being replaced by Earth System Models (ESMs), which incorporate new biogeochemical components (e.g., representation of the carbon cycle) and more detailed descriptions of the physical processes involved in the whole climate system.

Differently from weather forecasting, GCMs used in seasonal forecasting are initialized a few times per month and run forward in time from six months to one year rather than just a few days. Although there is not universally accepted terminology, we will use the terms *initialization time* and *lead-time* in this Thesis. Whereas the former refers to the moment in which the model is initialized, the latter makes reference to the time passed from the initialization moment to the beginning of the target season (season to be predicted). For instance, initializations of August 1st and May 1st could be used to forecast autumn (SON) at one and four-month lead, respectively (see Figure 3.5).

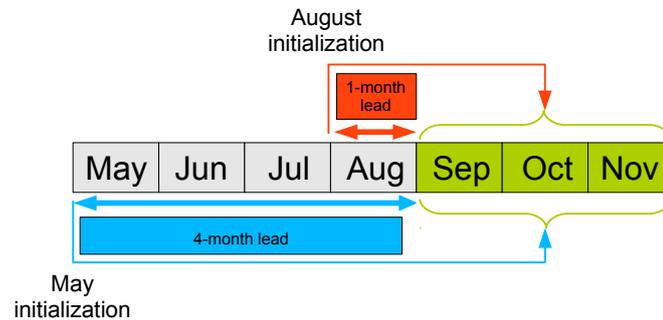


Figure 3.5: Illustration of the terms initialization and lead time for the target season SON.

Initialization is one of the critical stages in dynamical seasonal forecasting. Each component of the GCM needs to be appropriately initialized with the best available observations. However, information about the state of the ocean and sea-ice is scarce in comparison with the data available for the atmosphere (Balmaseda et al., 2007). Thus, the initial experiments for seasonal forecasting were carried out using atmospheric models with persistent SST conditions, like in the PROVOST (PRediction Of climate Variations On Seasonal to interannual Time-scales) project (see, e.g., Branković and Palmer, 2000; Doblas-Reyes et al., 2000). However, it is generally accepted that the predictive skill of atmosphere-ocean coupled models is larger than the skill of models using persistent SST conditions and therefore, state-of-the-art seasonal forecasting models are fully coupled models which use different assimilation approaches (4D-Var, Kalman filters, etc.).

As mentioned in Section 3.2, predictability at seasonal time-scales arises from the slowest components of the climate system, specially the SST. However, even if the slow evolution of SST could be predicted in a deterministic way, seasonal forecast predictions are intrinsically probabilistic. Probabilistic forecasts can be produced in various ways (Stockdale et al., 2010). A common approach is to generate an ensemble of simulations by using a set of slightly different initial conditions. Because of these little initial differences, the forecasts follow different evolutions, providing a variety of atmospheric trajectories—but still compatible with the underlying slow variables.— Rather than issuing a single yes-no prediction, the ensemble members can be used to compute the likelihood of a certain outcome (e.g., receiving above-normal precipitation for the next season). Figure 3.6 illustrates the probabilistic nature of seasonal forecasts, as well as the aforementioned *initialization time*, *lead-time* and *target season* concepts.

In order to properly evaluate their performance and skill (see Section 3.5), and to establish a reference climatology, seasonal GCMs are typically run in retrospective mode for historical periods of decades as if they had been operational during that time. These retrospective forecasts are referred to as *hindcasts*. It is crucial that hindcasts extend for

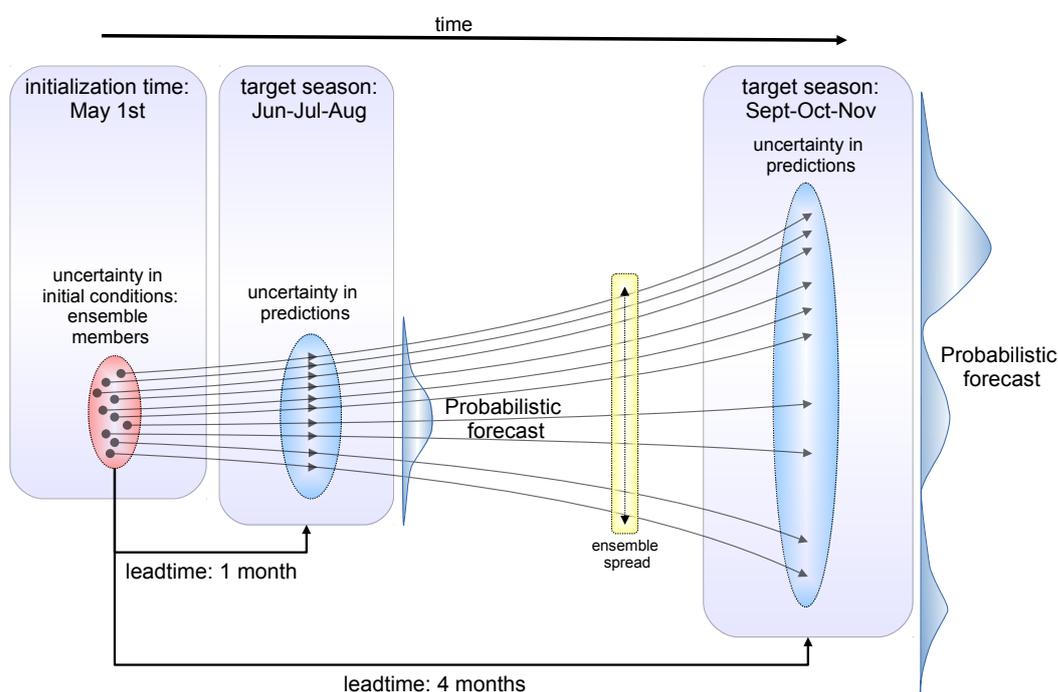


Figure 3.6: Schema illustrating the probabilistic nature of seasonal forecasts obtained from different initial conditions. Black lines show the trajectories of the different ensemble members.

a period long enough to assure statistically significant results. For instance, Shi et al. (2015) have recently shown that sampling uncertainty due to the length of the hindcast period is large. In particular, they found that the skill for forecasting the North Atlantic Oscillation (NAO) varies within a 40-year period, with ‘artificial’ high levels of skill found for some 20-year subperiods. Moreover, their conclusions are in agreement with the results obtained in previous studies (see, e.g., Müller et al., 2005; Kumar, 2009). As a consequence from this, seasonal forecasts from the longest-to-date and most comprehensive multimodel seasonal hindcast, provided by the EU project ENSEMBLES (see Section 3.4.1), are considered in this Thesis. They were retrieved from the Meteorological Archival and Retrieval System (MARS) of the ECMWF (<http://software.ecmwf.int/wiki/display/WEBAPI/Access+MARS>).

3.4 Multimodel Ensembles

Besides the uncertainty in the initial conditions, the particular model formulation (e.g., parametrizations) is a key source of uncertainty which leads to different representation of processes and therefore to different model climatologies.

Multimodel approaches where ensembles from different GCMs are combined allow to quantify the prediction uncertainty due to differences in model formulation. Multimodels generally produce more skillful forecasts than any single model does (see, e.g., Palmer et al., 2004; Wang et al., 2009), due to error cancellation and a reduction of overconfidence (the ensemble spread is widened while the ensemble mean error is reduced). As a result, the latest operational seasonal forecasts are based on multimodel systems (see Section 3.6.1).

In Europe, multimodel ensembles were produced in a series of initiatives such as DEMETER (Palmer et al., 2004) and ENSEMBLES (Weisheimer et al., 2009), and are still the focus in active projects such as SPECS (<http://www.specs-fp7.eu>). The first two projects were the precursors of the operational EUROSIP, which includes the models from the European Centre for Medium-Range Weather Forecast (ECMWF), the Centre National de Recherches Météorologiques (CNRM-Météo-France) and the UK Met Office (UKMO), and has been recently expanded to incorporate the CFSv2 model from NOAA/NCEP.

Besides these European initiatives, other multimodel projects have been developed and are currently operating in other parts of the world, such as the North American Multimodel Ensemble (NMME) (Kirtman et al., 2014); an experimental system including coupled models from several North American modelling centres: NOAA/NCEP, NOAA/GFDL, NCAR, NASA and CMC-Canada. Contrarily to the European multimodels, the NMME provides free retrospective forecasts and real-time seasonal forecasts. An example of the products provided is given in Figure 3.7, which shows the prediction of the 2016 El Niño event, issued on August 2015.

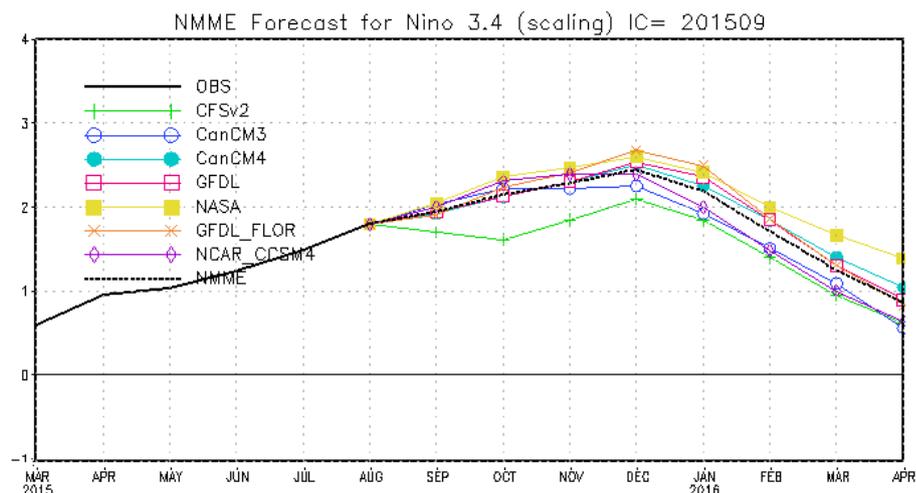


Figure 3.7: One-month lead forecasts of the SST anomaly in El Niño3.4 region from the different models included in the NMME (Kirtman et al., 2014), up to April 2016. Obtained from <http://www.cpc.ncep.noaa.gov/products/NMME/current/plume.html>.

3.4.1 The ENSEMBLES Multimodel Seasonal Hindcast

The ENSEMBLES dataset (Weisheimer et al., 2009) is the longest-to-date and most comprehensive multimodel seasonal hindcast. It comprises five global atmosphere-ocean coupled models from the following centres: The UK Met Office (UKMO), Météo France (MF), the European Centre for Medium-Range Weather Forecasts (ECMWF), the Leibniz Institute of Marine Sciences (IFM-GEOMAR) and the Euro-Mediterranean Centre for Climate Change (CMCC-INGV). Table 3.1 summarizes their main components.

Centre	Atmospheric model and resolution	Ocean model and resolution
ECMWF	IFS CY31R1 (T159/L62)	HOPE (0.3° – 1.4°/L29)
IFM-GEOMAR	ECHAM5 (T63/L31)	MPI-OM1 (1.5°/L40)
CMCC-INGV	ECHAM5 (T63/L19)	OPA8.2 (2.0°/L31)
MF	ARPEGE4.6 (T63)	OPA8.2 (2.0°/L31)
UKMO	HadGEM2-A (N96/L38)	HadGEM2-O (0.33° – 1.0°/L20)

Table 3.1: Main components of the five global atmosphere-ocean coupled models contributing to the ENSEMBLES multimodel seasonal hindcast.

The atmosphere and the ocean were initialized using realistic estimates of their observed states and each model was run from an ensemble of nine initial conditions (nine equiprobable members). For each model, seven-month long runs were issued four times a year within the period 1960-2005, starting the first of February, May, August and November (see Weisheimer et al., 2009, for more details about the experiment).

In this section, we focus on two important aspects which will help us to better understand some of the main characteristics of this dataset, which will be used throughout the entire Thesis. In particular, we analyze the ability of the different contributing models to predict the ENSO phenomenon as well as their patterns of bias and drift.

3.4.1-1 ENSO Teleconnections

During the last two decades, significant amount of research has been devoted to the use of GCMs to simulate ENSO, which is known to be the dominant mode of seasonal variability (see, e.g., Goddard and Dilley, 2005; Doblas-Reyes et al., 2010). Here, we show that the ENSEMBLES models exhibit good skill for this task, being therefore suitable for seasonal forecasting studies. In particular, Figure 3.8 shows the correlation between observed (ERSST v3b³) and simulated SST in El Niño 3.4 region⁴ for the five ENSEMBLES models during El Niño —1964, 1966, 1969, 1970, 1973, 1977, 1978, 1983, 1987, 1988, 1992, 1995 and 1998— and La Niña —1965, 1971, 1972, 1974, 1975, 1976,

³The last version of the Extended Reconstructed Sea Surface Temperature dataset (Smith et al., 2008).

⁴Region of the central equatorial Pacific ocean covering the box 5°N-5°S, 120°W-170°W.

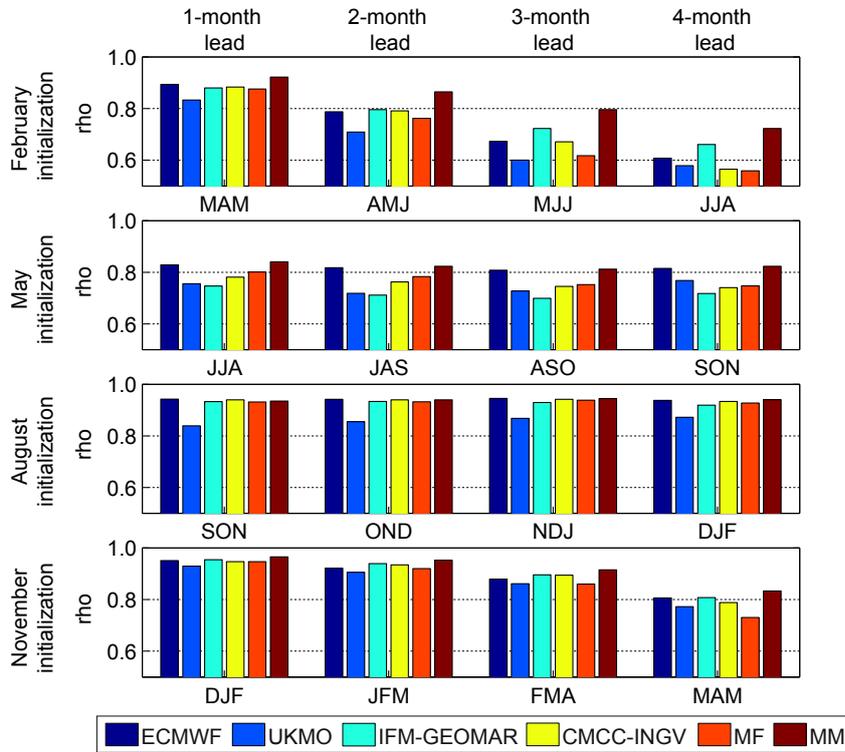


Figure 3.8: Correlation between observed (ERSST v3b) and simulated (five ENSEMBLES models and the corresponding multimodel) SST in El Niño 3.4 region during El Niño and La Niña episodes —defined according to the ONI index (see the text).— For each of the four initialization dates —February 1st, May 1st, August 1st and November 1st— (in rows), results are shown at increasing lead-times from 1 to 4 months (for the corresponding 3-month target seasons).

1984, 1985, 1989, 1996, 1999, 2000 and 2001— years, defined according to the Oceanic Niño Index (ONI: http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_change.shtml).

For each of the four initialization dates available —the first of February, May, August and November (in rows),— results are shown for the five ENSEMBLES models and the corresponding MM at increasing lead-times from 1 to 4 months (for the corresponding 3-month target seasons). For all models, high correlations are obtained for all initialization dates and seasons except for the central part of the year (May-June-July and June-July-August) for the case of the February initialization. This is in agreement with the spring predictability barrier for ENSO, which has been documented in previous studies (see, e.g., Zheng and Zhu, 2010; Tippett et al., 2011; Yan and Yu, 2012; Duan and Wei, 2013) and is still found for recent GCMs. Moreover, except for the latter case, correlations do not decrease substantially as the lead-time increases.

Furthermore, Figure 3.9 shows the El Niño teleconnections reproduced by the multimodel (see Section 3.4.1), which were computed following the same methodology used for Figure 3.2. Despite being smoother, the predicted teleconnection patterns resemble quite well the observed ones —numbers in each map show the spatial correlation between both,— especially in the most skillful seasons, winter (DJF) and autumn (SON) (see Section 7.3 for details). For these seasons, Figure 3.10 —a composite of Figures 7.4 and 3.2— shows the observed teleconnections, but only in those grid boxes showing significant skill at one month lead-time for the ENSEMBLES multimodel. Most of these grid boxes exhibit red/blue rather than white colors, which indicates that skillful regions are appreciable teleconnected with El Niño. Moreover, teleconnections are significant (black dots) over some of the regions showing the highest skill.

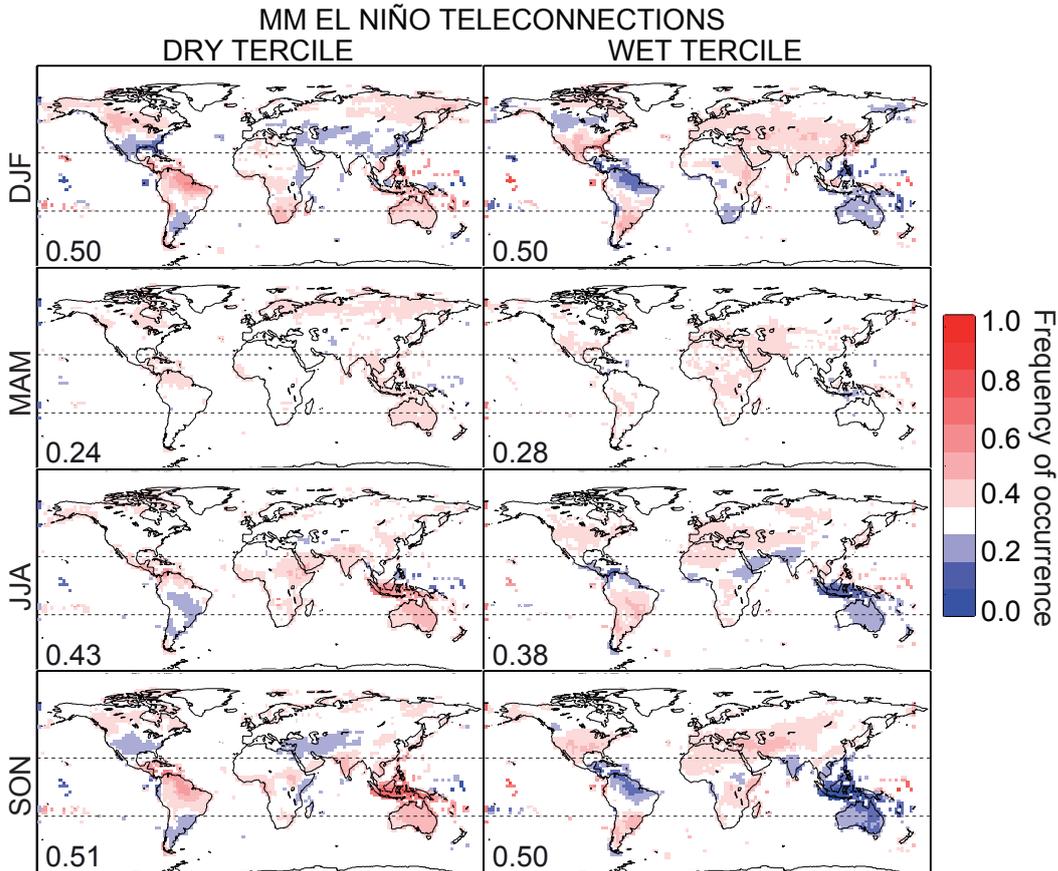


Figure 3.9: El Niño teleconnections reproduced by the ENSEMBLES multimodel for the dry and wet terciles (left and right columns, respectively) at one month lead-time. The numbers in each panel show the spatial correlation with the corresponding observed patterns (shown in Figure 3.2). Dashed lines indicate the tropics/extratropics division.

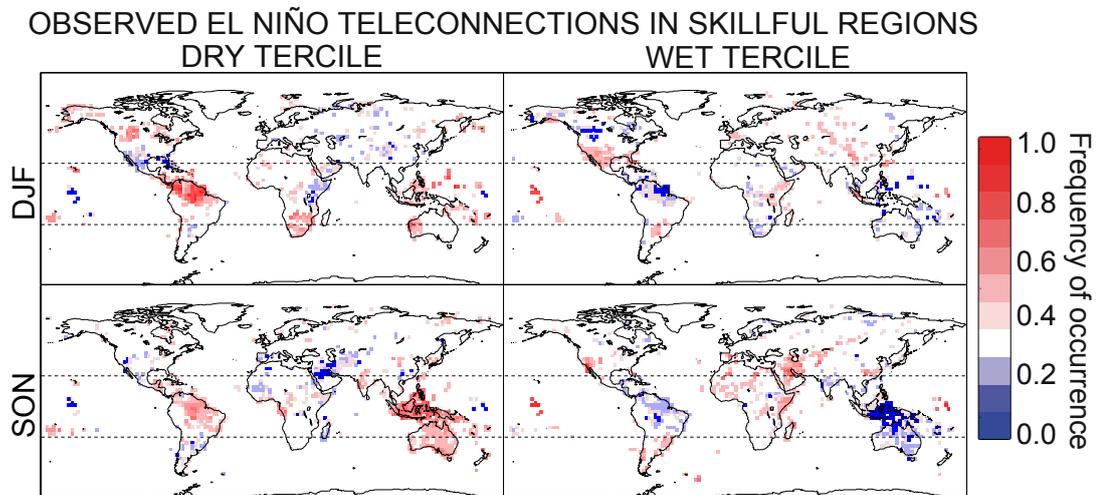


Figure 3.10: As Figure 3.2, but restricted to those grid boxes exhibiting significant skill at one month lead-time for the ENSEMBLES multimodel (see Chapter 7 for details).

These results are in agreement with previous studies which have shown that most of the skill for seasonal precipitation forecasts is found over regions strongly connected with ENSO (see, e.g., Coelho et al., 2006; Barnston et al., 2010; Arribas et al., 2011; Lim et al., 2011; Kim et al., 2012a,b; Landman and Beraki, 2012) and suggest that this skill is determined to a great extent by the ability of the different models to properly reproduce the observed El Niño teleconnections (this will be analyzed in detail in Chapter 7).

3.4.1-2 Bias and Drift

At seasonal time-scales, model mean error or bias (deviation from mean observations) is a function of lead-time. In particular, it arises from the so-called *initial shock* (rapid adjustment processes caused by the imbalance between the initial conditions and the model dynamics) and evolves towards the model climatology, producing a transient trend (or *drift*) which should not be confused with a seasonal climate signal (see Figure 3.11). Analyzing the nature and quantifying the magnitude of these errors has greatly improved the identification of model deficiencies (see, e.g., Vannitsem and Nicolis, 2008; Vannitsem, 2008; Eden et al., 2012).

In order to assess the spatio-temporal distribution of bias and drift of seasonal forecasts worldwide—which can help to identify specific model deficits and offers the possibility of targeted improvement of certain processes formulation, resolution and parametrization (Ehret et al., 2012),— we considered precipitation from four of the ENSEMBLES models for the period 1960-2000. For coherence, all models were bi-linearly interpolated to a

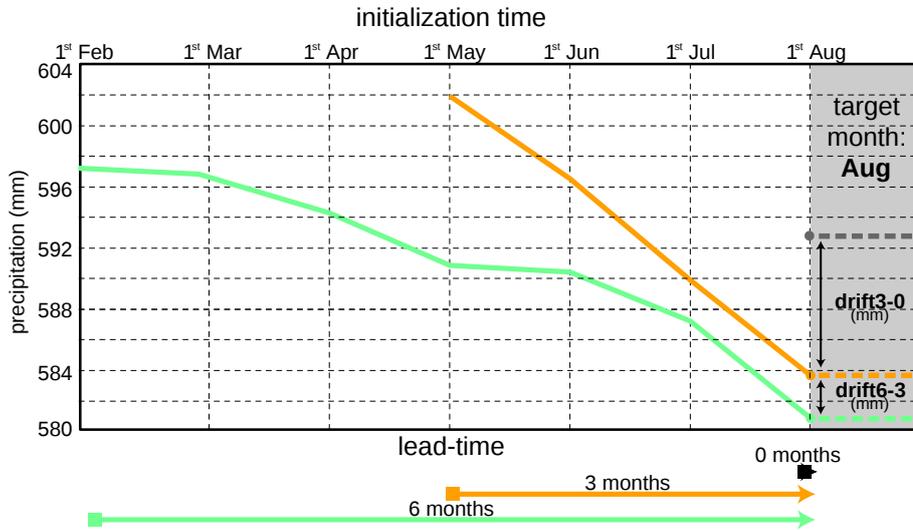


Figure 3.11: Schema illustrating the drift of seasonal forecasts.

common 2° regular grid. Recall that the ENSEMBLES models were initialized four times a year (the first of February, May, August and November) during the hindcast period and run for seven months. Therefore, for a given target month, the corresponding bias can be only computed at certain lead-times. In particular, for the four illustrative months considered here (February, May, August and November), forecasts are available at 0-, 3- and 6-month lead. The first column of Figures 3.12 to 3.15 shows the $bias_0$, i.e., the bias (computed against VASClmO v1.1 observations) corresponding to the 0-lead predictions. Additionally, the second and third columns show the incremental $drift_{3-0}$ and $drift_{6-3}$, which are computed as bias differences for increasing lead months in the following way: $drift_{X-Y} = bias_X - bias_Y$. Note that this allows to assess the evolution of the model drift along the entire run. Note also that whereas $bias_0$ depends on the observations, $drift_{3-0}$ and $drift_{6-3}$ depend exclusively on the model. Only those values significantly different from zero ($\alpha = 0.05$) are displayed. To compute this significance, a bootstrap approach was followed. In particular, 1000 different 9-member ensembles were first built by random selection for each model. Confidence intervals were then computed upon the 1000 bootstrapped results.

The patterns found greatly vary among the different models, especially for particular regions and months. Broadly speaking —analyzing in detail the results obtained is not the aim here,— bias and drift are mainly located in tropical latitudes and are stronger over the oceans than over land. Importantly, the drift is as large as the bias for some particular regions and models. In particular, note that $drift_{3-0}$ is systematically stronger than $drift_{6-3}$, which might be due to the initial shock (Balmaseda, 2012) —moreover, the

initial weeks are usually discarded in seasonal forecasting due to the presence of deterministic predictability related to the initial conditions, which leads to artificial increased skill.— Importantly, if not taken into account, this drift could introduce errors into the forecast that might be large compared to the signal being predicted (Smith et al., 2013). However, whereas bias corrections are nowadays routinely applied in seasonal forecasting, the drift has been seldom studied to-date despite its systematic presence in state-of-the-art seasonal forecasts. In particular, there are very few works on the relative merits of the two procedures commonly considered to deal with it, full-field and anomaly initialization schemes (Magnusson et al., 2013b,a; Carrassi et al., 2014).

3.5 Verification of Seasonal Forecasts

There are two basic aspects that may be assessed in a seasonal forecasting system: predictability and forecast quality.

On the one hand, predictability is concerned with the extent to which a forecast anomaly (*signal*) is large by comparison with relevant sources of variability (*noise*) in the forecast system—it is essentially a property of the system and says nothing directly about whether the forecasts agree with observations or not.— Predictability has commonly been assessed through experiments in which an atmosphere model is forced with prescribed SSTs (see, e.g., Rowell, 1998; Shukla et al., 2000; Straus et al., 2003). In this case, SSTs are typically taken from observations and ensemble integrations with different atmospheric initial conditions are used to sample the internal variability that is generated in the atmosphere. Analysis of variance can then be used to separate the SST forced signal from the internally generated noise (see, e.g., Rowell et al., 1995; Rowell, 1998). Results are usually summarized either by the signal-to-noise ratio, or by the ‘potential predictability’, i.e., the ratio of the SST forced variance to the total variance.

On the other hand, forecast quality can be evaluated by estimating the accuracy, skill (see Chapter 7 for an assessment of skill of seasonal precipitation forecasts worldwide) and reliability of a set of hindcasts (Jolliffe and Stephenson, 2003). Accuracy refers to the precision with which the forecast system tends to match the observed changes that it is trying to predict, while the skill is the relative accuracy of the forecast over some reference prediction (e.g., climatology or persistence). Reliability, instead, measures how well the forecast probability distribution matches the observed relative frequency of the forecast event, i.e., a forecast system will be reliable if predictions of a 80% probability of a dry season correspond to observed dry seasons 80% of the time.

Since seasonal predictability strongly varies with the region and the season being forecast (see, e.g., Halpert and Ropelewski, 1992; van Oldenborgh, 2004; Barnston et al.,

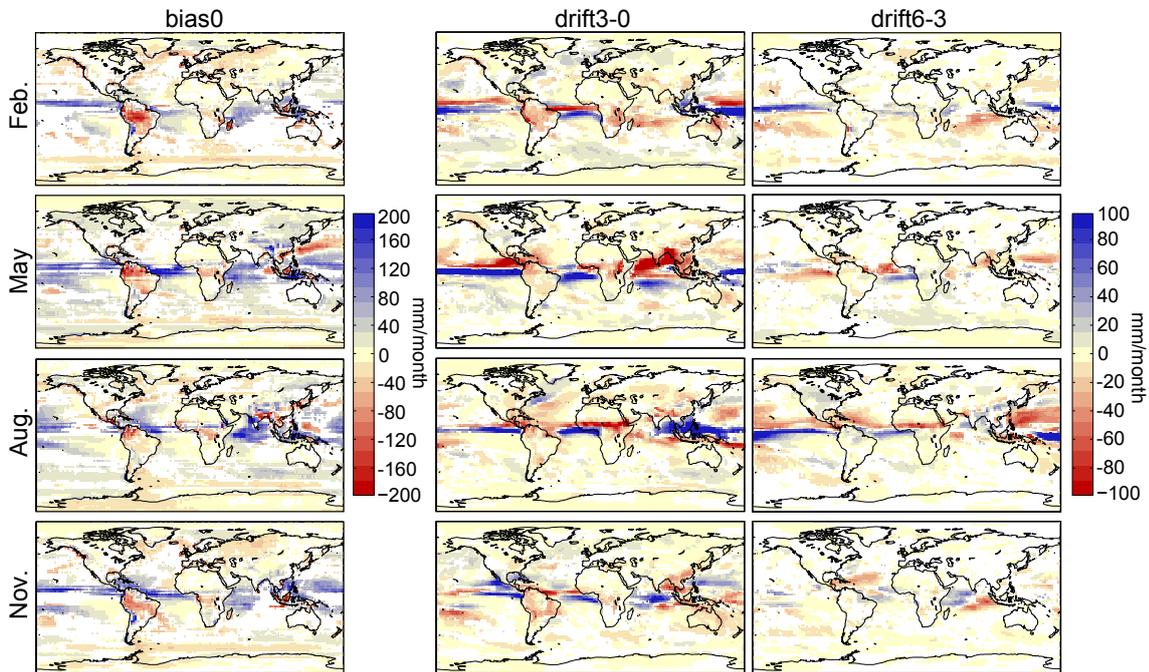


Figure 3.12: Global patterns of significant ($\alpha = 0.05$) bias (i.e., *bias0*) and drift (*drift3-0* and *drift6-3*) for the ECMWF model—the ensemble mean is considered—for four illustrative months (in rows) for the period 1960-2000. See the text for details.

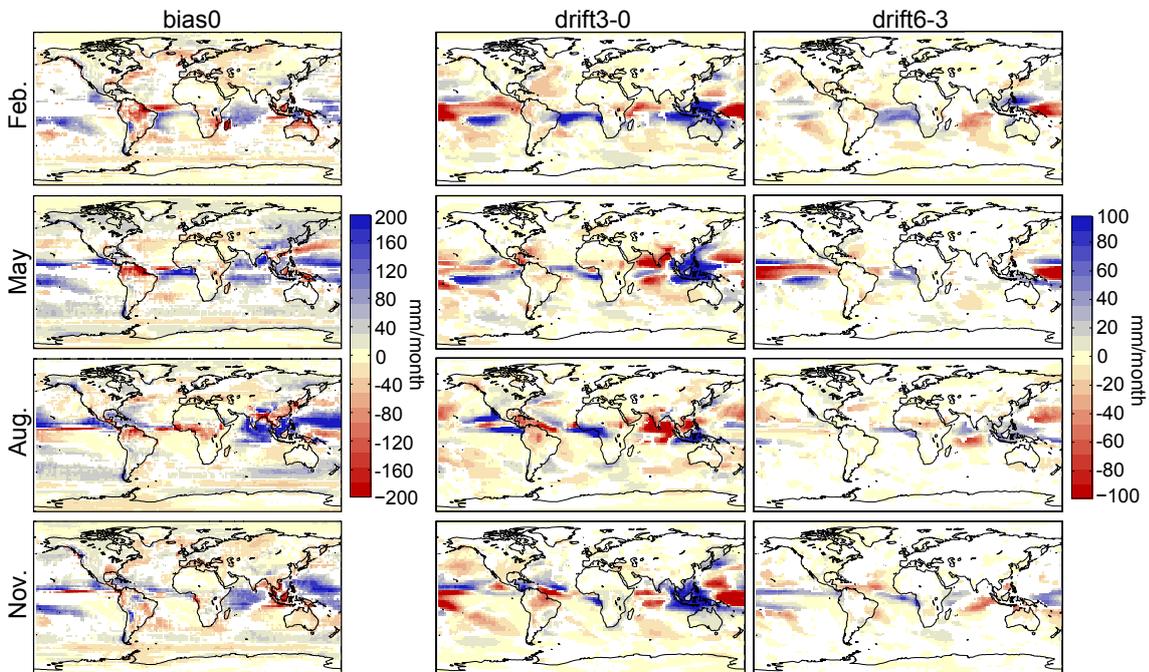


Figure 3.13: As Figure 3.12 but for the IFM-GEOMAR model.

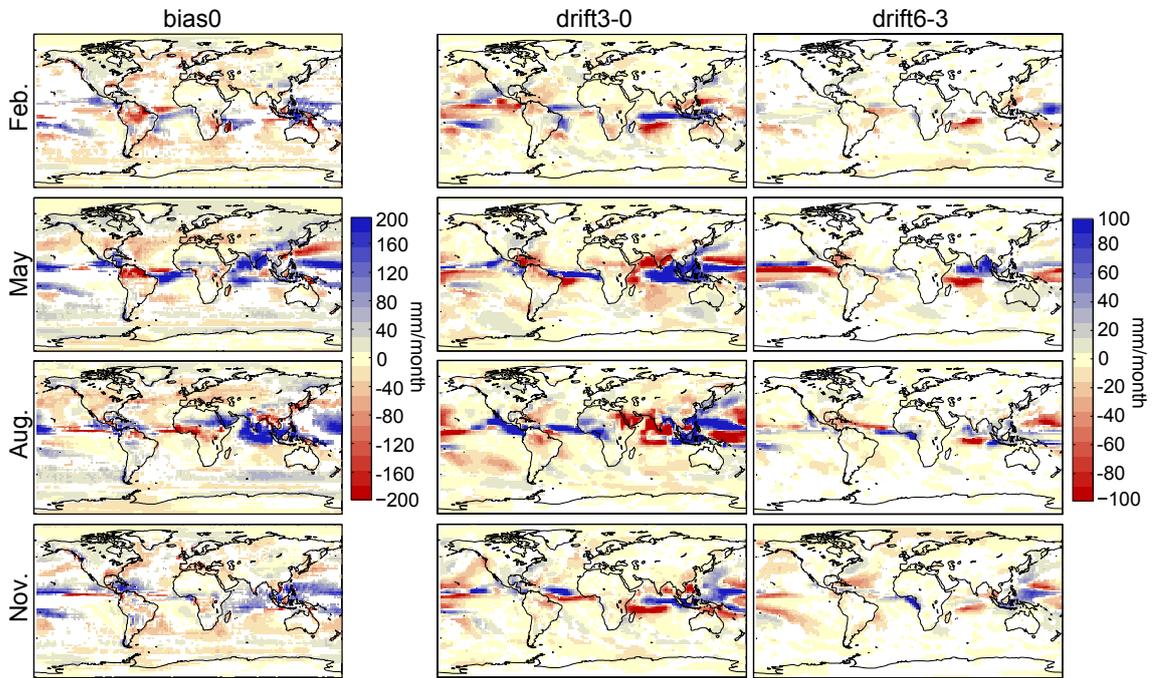


Figure 3.14: As Figure 3.12 but for the CMCC-INGV model.

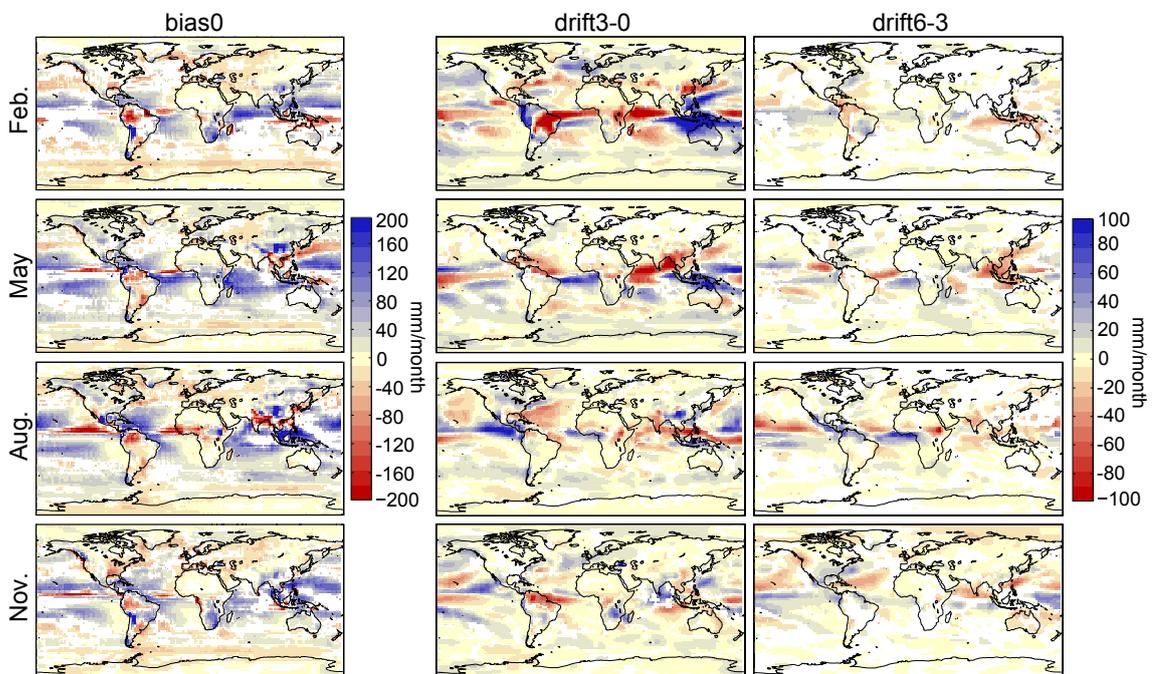


Figure 3.15: As Figure 3.12 but for the MF model.

2010; Doblas-Reyes et al., 2010), forecast quality verification information is essential to enable the end-user to quantify, based on past model performance, the uncertainty associated to a forecast for a particular location and time of the year, allowing thus for a correct use. Although no single metric can provide a complete picture of forecast quality, the Standardized Verification System for Long Range Forecasts (SVS-LRF: <http://www.bom.gov.au/wmo/lrfvs/index.html>) recommends a suite of metrics —see Attachment II.8 (page 122) of WMO (1992) for details— which allows forecasting centres to document the quality of their forecasts according to a common standard. Among the recommended metrics for either deterministic —mean error or bias, Anomaly Correlation Coefficient (ACC), Root Mean Square Error (RMSE), Mean Square Skill Score (MSSS), etc.— or probabilistic —Brier Skill Score (BSS), Ranked Probability Skill Score (RPSS), Relative Operating Characteristic Skill Score (ROCSS), reliability diagrams, etc.— forecasts, which cover different aspects relevant to users (Jolliffe and Stephenson, 2003), we only consider in this Thesis the bias and the ACC for the former and the ROCSS and reliability for the latter. They are described next.

The *bias* (3.1) is the most simple validation metric for deterministic predictions. For a sample size of N , it measures the mean error of the forecasts (f_n), i.e., their mean deviation from observations (o_n). However, this metric does not measure the correspondence between forecasts and observations, i.e., it is possible to get a perfect score of 0 for a bad forecast if there are compensating errors.

$$Bias = \frac{1}{N} \sum_{n=1}^N (f_n - o_n) \quad (3.1)$$

The *ACC* (3.2) is another common validation metric for deterministic predictions which measures the temporal correspondence between the forecasts and the observations, subtracting out their corresponding climatological means (\bar{f} , \bar{o}) at each point. ACC is not sensitive to forecast bias and ranges from -1 to 1 (perfect score).

$$ACC = \frac{\sum_{n=1}^N (f_n - \bar{f})(o_n - \bar{o})}{\sum_{n=1}^N \sqrt{(f_n - \bar{f})^2} \sum_{n=1}^N \sqrt{(o_n - \bar{o})^2}} \quad (3.2)$$

The *ROCSS* (3.3) is a categorical skill score recommended by the SVS-LRF for the verification of probabilistic seasonal forecasts. It assesses how well a forecast discriminates between two alternative outcomes, typically getting/not getting a particular event (e.g. receiving less than normal precipitation). For that event, the ROCSS is computed as

$$ROCSS = 2A - 1, \quad (3.3)$$

where A is the area under the ROC curve (commonly used to evaluate the performance of probabilistic systems). ROC curves are constructed by plotting the Hit Rate (HIR) against the False Alarm Rate (FAR) using a set of increasing probability thresholds (e.g., 0.05, 0.15, 0.25, etc.) to make the yes/no decision. The ROCSS ranges from 1 (perfect forecast system) to -1 (perfectly wrong forecast system). A value zero ($A = 0.5$) indicates no skill with respect to a climatological prediction. Derived from the fact of considering categories (e.g. terciles) instead of absolute values, this metric is not sensitive to the different models' biases (see Section 3.4.1). That is, a biased forecast may still produce a good ROC curve, which means that it may be possible to improve the forecast through calibration. The ROCSS is thus acknowledged to be a reasonable first choice to communicate the value (potential usefulness) of a forecast to the end-users (see, e.g., Thiaw et al., 1999; Kharin and Zwiers, 2003).

3.5.1 Reliability Categories

As previously mentioned, reliability measures how well the forecast probability distribution matches the observed relative frequency of a certain event (e.g., a particular precipitation tercile). Reliability diagrams are used to assess the reliability of probabilistic forecasts for that event. Since they are conditioned on the forecasts (i.e., given that the event was predicted, what was the outcome?), reliability diagrams are a good partner to the ROCSS, which is conditioned on the observations (i.e., given that the event occurred, what was the corresponding forecast?).

Reliability diagrams plot the observed frequencies of the event as a function of its forecast probability, as represented by a determined number of bins (see Doblas-Reyes et al., 2008, for details). For a perfectly reliable forecasting system, the reliability curve would match the diagonal. Points falling within the so-called skill region, i.e., the region contained between the no-resolution line (which indicates the expected frequency of the event; e.g., 1/3 for terciles) and the no-skill line (halfway between the no-resolution line and the diagonal) positively contributes to the forecast skill ($BSS > 0$). Weisheimer and Palmer (2014) proposed a methodology to translate the information provided by the reliability diagram to an easy-to-interpret scale with five reliability categories: *perfect* (green), *still useful* (blue), *marginally useful* (yellow), *not useful* (orange) and *dangerous* (red). In particular, they performed a weighted linear regression as a best-guess estimate on all data points in the reliability diagram, using the number of forecasts in each probability bin as weights. The different reliability categories are defined based on the relative position of the so derived reliability line with respect to the perfect reliability (diagonal), no-skill and no-resolution lines, as well as on the uncertainty range around it (as obtained by bootstrapping). To estimate the confidence interval around the best-guess reliability line

they randomly resampled members, grid boxes and years and 75% of the total range (computed upon 1000 samples) was considered. In this Thesis, we analyzed the sensitivity of the classification to different confidence intervals (the same bootstrapping procedure was used) and found that the ensemble size had a large influence, obtaining higher uncertainty for smaller ensembles. In these situations, the *still useful* (blue) categories may pass to *marginally useful* (yellow) ones due to an enlargement of the confidence region (see Weisheimer and Palmer, 2014, for details on the definition of the different categories). Consequently, we considered as confidence interval the 50% of the total range, which is more suitable for the nine members of the ENSEMBLES models available for this Thesis. Moreover, within the *marginally useful* (yellow) category proposed by Weisheimer and Palmer (2014), we differentiate those cases in which the reliability line falls within the skill region but the uncertainty range around it is not fully contained. These cases are identified with the dark yellow color in this Thesis.

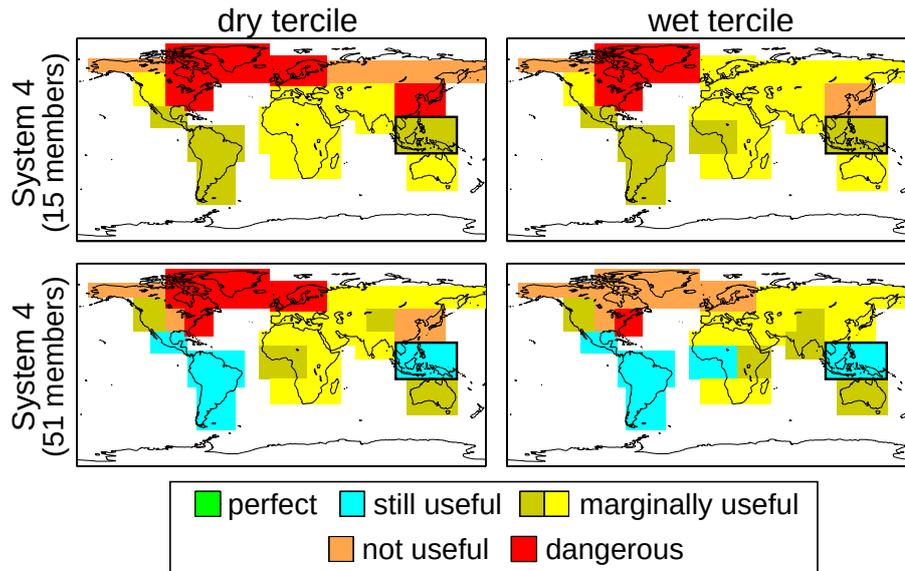


Figure 3.16: Reliability categories (see the text for details) obtained for the one-month lead seasonal precipitation forecasts from the 15 and 51 members version of the ECMWF System 4 (top and bottom row, respectively) over the 21 only-land regions defined in Giorgi and Francisco (2000) for summer (JJA) during the period 1981-2010. Only the dry and wet terciles (left and right column, respectively) are shown. GPCP v2 was considered as reference.

Maps in the top (bottom) row of Figure 3.16 show the results obtained from applying the methodology above described to the one-month lead seasonal precipitation forecasts from the 15 (51) members version of the ECMWF System 4⁵ over the 21 only-land regions

⁵The latest seasonal forecasting system from the ECMWF, System 4 (Molteni et al., 2011), provides

defined in Giorgi and Francisco (2000) for the period 1981-2010. Note that reliability is computed based upon the time-series resulting from joining the predictions for all grid boxes within each region. For brevity, only the dry and wet terciles (left and right column, respectively) for summer (JJA) are shown. Observed precipitation from the Global Precipitation Climatology Project (GPCP) version 2 (2.5° resolution from January 1979 to the present; Adler et al., 2003) was considered as reference. Although this figure suggests that reliability might be improved by using a larger number of ensemble members, there is still an open discussion about the effect on reliability of the ensemble size (see, e.g., Richardson, 2001; Ferro et al., 2008; Kay et al., 2013; Berner et al., 2014).

To gain insight on this, Figure 3.17 shows the reliability diagrams obtained for the southeast Asia region (marked with a black border in Figure 3.16) for the 15 and 51 members version (top and bottom row, respectively). The frequency histograms (also called sharpness diagrams) are also provided inside each plot. These histograms show, for each of the probability bins considered, the frequency of the forecasts. Whereas the probabilities peak at the first bin for the 15 members case, they mainly concentrate around 1/3 (the climatological expected value) for the 51 members case. In consequence, as derived from the applied weights in the regression model proposed by Weisheimer and Palmer (2014), the slope of the best-guess reliability line in the 15 members case would be greatly influenced by small deviations in near-zero probabilities —note that the regression line always passes through the climatological intersection point, i.e., 1/3 for terciles,— which may lead to worse reliability categories.

Differently to reliability, accuracy and skill do not vary sensibly with the ensemble size considered. To illustrate this, Figure 3.18 shows the validation results at one-month lead for the 15 and 51 members version (left and right column, respectively) of the ECMWF System 4 —also validated against GPCP version 2— in terms of bias, ACC and ROCSS for JJA. Results are similar in both cases, which suggests that a few ensemble members might be enough for a robust verification in terms of the aforesaid metrics.

3.6 State-of-the-Art of Seasonal Forecasting

3.6.1 Operational Seasonal Forecasts

Providing operational long-range forecasts (from one month up to two years ahead) on a global scale requires huge amounts of computer resources along with a very specialized knowledge. For this reason, there are only a few centres around the world that are pro-

retrospective forecasts for the period 1981-2010, including a 15 members version with seven-month long simulations initialized on the 1st day of every month and another 51 members version with seven-month long simulations but only for four initialization dates per year (the 1st of February, May, August and November).

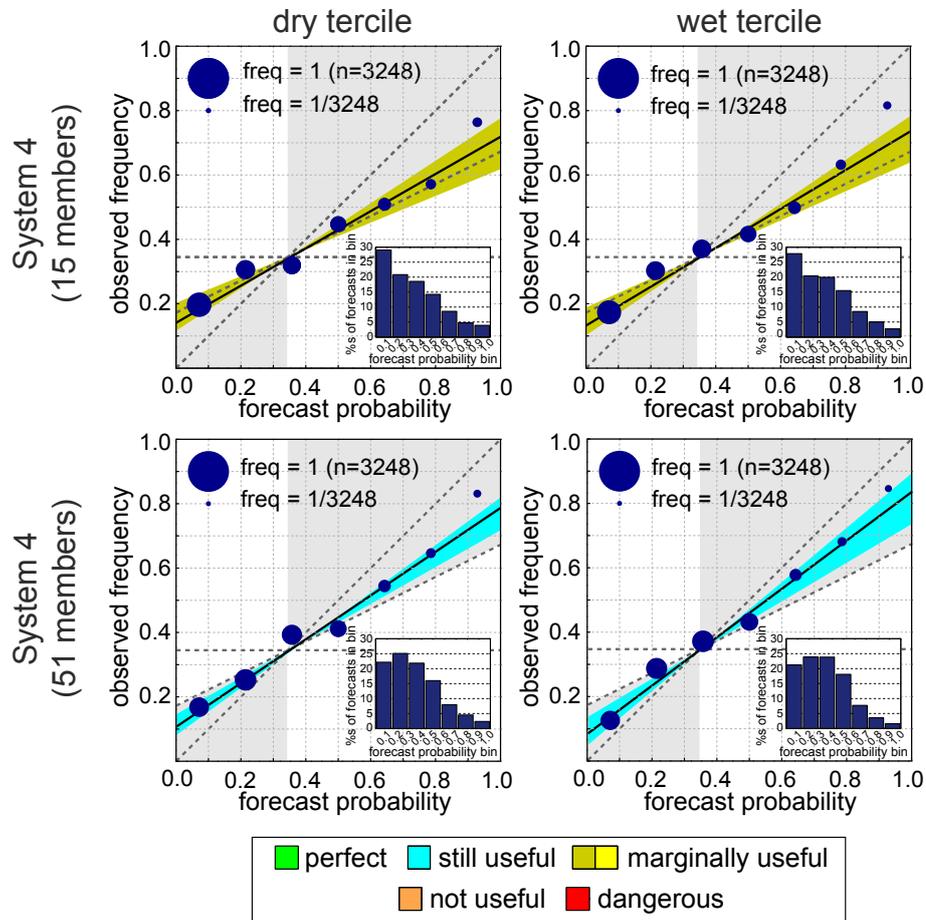


Figure 3.17: Reliability diagrams obtained for the southeast Asia region (marked with a black border in Figure 3.16), for the 15 and 51 members version of the ECMWF System 4 (top and bottom row, respectively) in JJA. Only the dry and wet terciles are shown (left and right column, respectively). The inset shows the corresponding frequency histograms.

ducing these forecasts, the WMO-designated Global Producing Centres for Long Range Forecasts (GPCs: <http://www.wmo.int/pages/prog/wcp/wcasp/gpc/gpc.php>).

The predictions produced by these centres are used to provide consensus-based seasonal forecasts with socio-economic potential by WMO's Lead Centre for Long-Range Forecasts Multimodel Ensemble (LC-LRFMME: <http://www.wmo1c.org>), which brings together prediction providers and local to regional focal points like the Regional Climate Centres (RCCs: <http://www.wmo.int/pages/prog/wcp/wcasp/RCCs.html>), the Regional Climate Outlook Fora (RCOF: http://www.wmo.int/pages/prog/wcp/wcasp/RCOF_Concept.html), National Meteorological and Hydrological Services (NMHSs) and private partners.

In addition to the twelve designated GPCs, other major centres providing global sea-

sonal forecasts are the International Research Institute for Climate and Society (IRI: <http://iri.columbia.edu>) and the Asia-Pacific Economic Cooperation (APEC) Climate Center (APCC: http://www.apcc21.net/eng/service/fore/lmon/japcc030101_1st.jsp).

At present, seasonal forecasts are commonly visualized on maps that show the likelihood of rainfall being below-normal, normal, or above-normal for the coming seasons (see Figure 3.19 for an illustrative example). Other ways of presenting these predictions, taking into account their probabilistic nature, are being developed in the flagships EU-funded initiatives for seasonal forecasting, particularly in the EUPORIAS (<http://euporias.eu>) project (see deliverable D33.3 for details: <http://www.euporias.eu/system/files/D33.3.pdf>).

3.6.2 Towards Climate Services

One of the main barriers of seasonal forecasts is related to the lack of accessibility and understanding by end-users (Lemos et al., 2012). In fact, there is little evidence of their direct use for operational applications nowadays (Coelho and Costa, 2010), which is often ascribed to the users' difficulty in integrating the predictions into existing decision support systems—to be useful, climate information must be tailored to meet the needs of users.—Therefore, it is still needed to improve the manner in which actionable seasonal climate information is disseminated to policy-makers and stakeholders, offering for instance estimates of future risks of occurrence of high-impact extreme events to which society is vulnerable.

To this, *climate services* bridge the gap between science and policy by improving the visualization of the predictions, the public dissemination of the data generated and by elaborating introductory, both targeted and general-public information on the prediction generation methodologies and the regions, variables and times of the year where positive skill can be expected. Moreover, climate services may provide feedback to the operational forecast centres to design strategies to efficiently structure their resources to provide timely, useful and understandable information to a wide range of users—for instance, by means of comprehensive web pages.—

The international community established the Global Framework for Climate Services (GFCS) to promote the development of operational climate services at the national and regional levels, especially in developing countries (see <http://www.gfcs-climate.org/projects-list> for a list), where users need access to expert advice and support to help them select and properly apply climate information.

In Europe, several previous—e.g., QWeCI (<http://www.liv.ac.uk/qweci>), CLIMRUN (<http://www.climrun.eu>), ECLISE (<http://www.eclise-project.eu>)—and active—EUPORIAS and SPECS—projects have contributed to the progress of climate

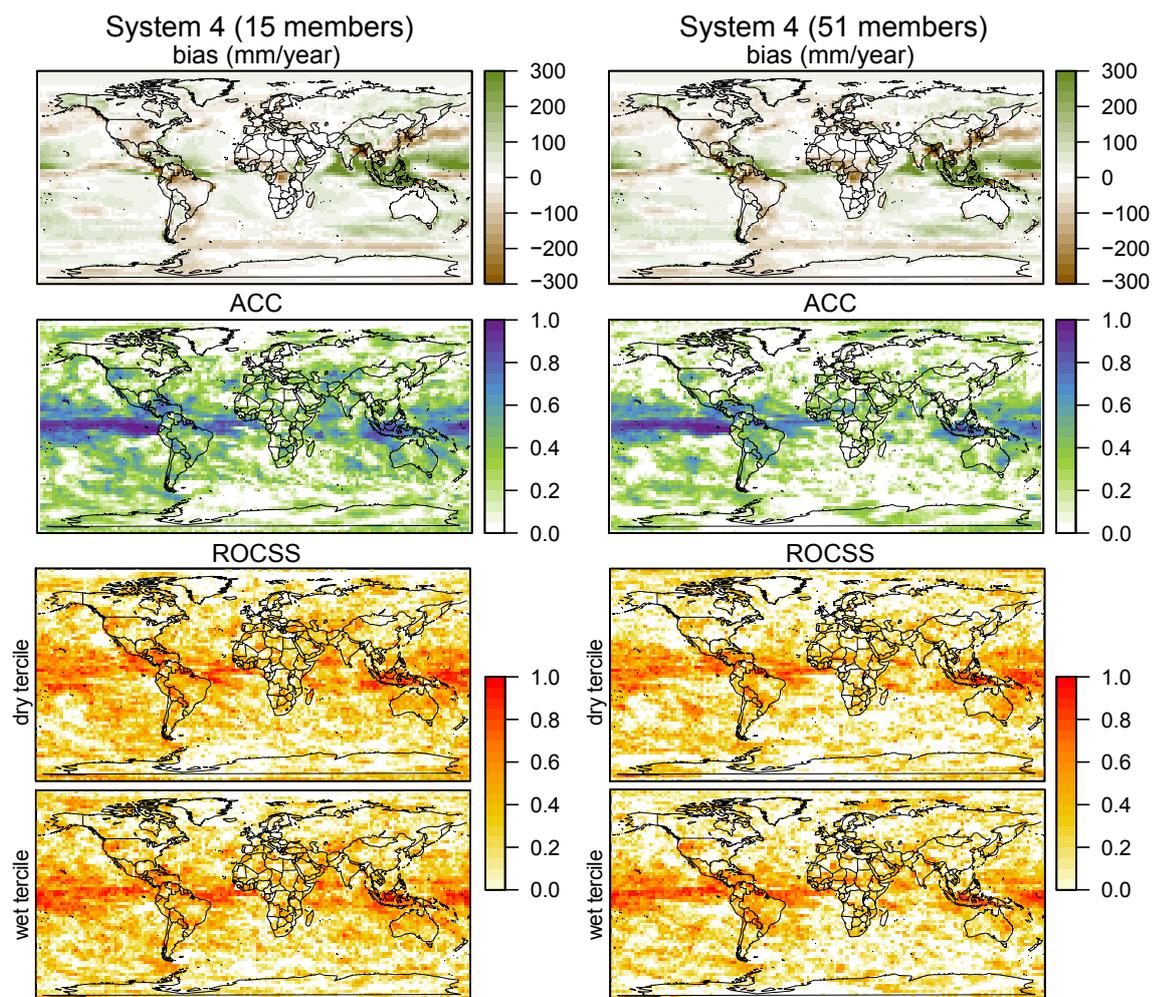


Figure 3.18: Validation results (1981-2010) from the 15 and 51 members version (left and right column, respectively) of the ECMWF System 4 for summer (June-July-August) precipitation forecasts at one-month lead. Bias, interannual correlation (ACC) and ROCSS (dry and wet terciles) are shown in rows. GPCP version 2 was considered as reference.

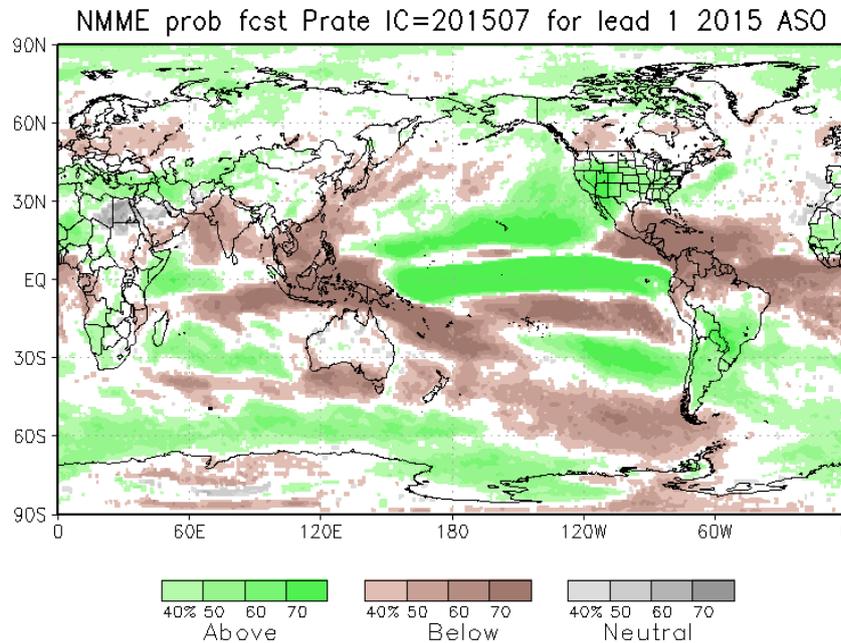


Figure 3.19: NMME (Kirtman et al., 2014) probabilistic forecasts of precipitation for August-September-October 2015, issued in July 2015 (one-month lead). Colours show the probability of the most likely category. Obtained from <http://www.cpc.ncep.noaa.gov/products/NMME/prob/PROBprate.S.html>.

services. In particular, one of the objectives of EUPORIAS is to assess and document the current marketability of climate services in Europe, whereas SPECS aims to enhance the European role on the provision of climate services according to WMO protocols by creating examples of improved tailored forecast-based products for the GPCs and participating in their transfer to worldwide RCCs and NHMSs. Furthermore, the Copernicus Climate Change service (C3S: <http://www.copernicus.eu>) will deliver substantial economic value to Europe by: (1) informing policy development to protect European citizens from climate-related hazards such as high-impact weather events, (2) improving planning of mitigations and adaptation practices for key human and societal activities and (3) promoting the development of new services by providing datasets and tools following an open data policy.

3.7 Beyond the State-of-the-Art in Seasonal Forecasting

Besides the aforementioned lack of accessibility and understanding by end-users (Lemos et al., 2012), the main barriers perceived for seasonal forecasting are linked to the lack of

skill for certain regions/seasons. Despite the huge advances achieved in the last decade, current forecasting systems have limited skill at seasonal time-scales (Lazar et al., 2005; CAIICP, 2010). Among others, this is due to their limited spatial resolution (horizontal and vertical), the use of inaccurate numerical schemes (simplified physics) and the incomplete knowledge of the complex processes involved in the climate system.

Apart from developing new models which can run on finer resolutions and better describe the different components of the climate system —especial focus in being put on the stratosphere— and their coupling, some of the approaches that are being investigated to overcome these issues rely on the use of better initial conditions for the land-surface (Koster et al., 2010), soil-moisture, sea-ice (Holland et al., 2011) and snow cover (Cohen and Jones, 2011).

In this regard, multidimensional observational datasets of the coupled atmosphere-ocean-cryosphere-land surface system will allow to explore the advantages of full-field versus anomaly initialisation methods —there are very few works on this issue to-date (Magnusson et al., 2013b,a; Carrassi et al., 2014),— to assess the impact of the initial shock and to generate sets of initial conditions that efficiently sample the observational uncertainty in the climate system —the estimation of the optimal set of initial conditions to generate the appropriate ensemble given a finite amount of computing resources is far from trivial (Balmaseda and Anderson, 2009)— and help to reduce the typical overconfidence of single-model ensemble forecast systems (Doblas-Reyes et al., 2009).

Additionally, even though ENSO is known to be the main driver of skill at seasonal time-scales (Manzanas et al., 2014b), other processes such as the Madden-Julian Oscillation, the Quasi-Biennial Oscillation, the Indian Ocean dipole, feedbacks between the ocean and the atmosphere and between the land and the atmosphere, and interactions between the stratosphere and lower layers of the atmosphere are expected to bring additional predictability. For instance, the new seasonal forecast system from the UK Met Office, GloSea5, has shown promising skill in predicting the NAO due to a considerable increase in resolution⁶ (Scaife et al., 2014), although this is a controversial issue (Shi et al., 2015).

Finally, exploration of methodologies to efficiently combine multimodel ensembles and statistical downscaling techniques which transfer the coarse forecasts to more local-scales should be continued (CAIICP, 2010).

⁶The climate model at the core of GloSea5 is the Hadley Centre Global Environmental Model version 3, with atmospheric resolution of 0.83° longitude by 0.55° latitude, 85 quasi-horizontal atmospheric levels and an upper boundary at 85 km near the mesopause. The ocean resolution is 0.25° globally in both latitude and longitude with 75 quasi-horizontal levels.

CHAPTER 4

A Regional Case Study: The Philippines

The Philippines is widely recognized as one of the most natural hazard-prone countries in the world (Benson, C., 1997). In particular, typhoons, floods, landslides, droughts and other weather and climate-related natural disasters occur on a regular basis (Israel and Briones, 2012), damaging lives, properties and the economy in general. Some of these hazards are directly related to the ENSO phenomenon, which strongly influences the climate of this region (Koide et al., 2012). The Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA: <http://meteopilipinas.gov.ph>) summarizes the potential impacts of the warm (cold) phase of ENSO in the Philippines as follows: 1) the rainy season is shorter (longer) because of the delayed (normal or early) monsoon onset and the early (normal or late) termination, 2) there is weak (strong) monsoon activity and 3) fewer (more) cyclones pass through the country. As a result, during El Niño (La Niña), generally below (above) normal rainfall is observed.

In the light of the ENSO predictability documented in Chapter 3, its associated adverse impacts (Hilario et al., 2009) could be in part mitigated by taking advantage of suitable seasonal forecasts. In particular, most (if not all) important sectors (e.g., energy, water, transport, industry) could greatly benefit from proper climate information at this particular time-scale, which might serve as basis for strategic planning in risk management and could, in turn, improve decision making from national, regional and local governments. However, given their low spatial resolution, state-of-the-art seasonal forecasts need to be satisfactorily translated to the local-scale required for practical applications —e.g., crop modelling for efficient rice production (Koide et al., 2012), which is crucial for most of the country's population (Lansigan, 2005).— To this, which constitutes the central problem

of this Thesis, statistical downscaling techniques (which are introduced in Chapter 5) can be applied.

Based on these premises, we present in this chapter a general introduction to the climate of the Philippines and argue why this region is selected as the case study of the Thesis.

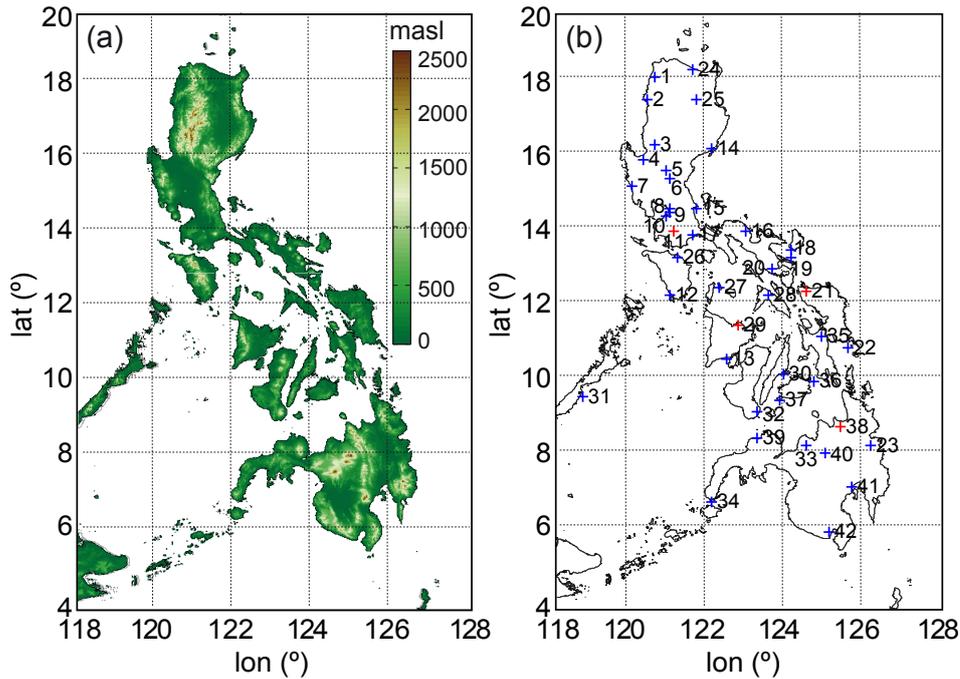


Figure 4.1: Location of the 42 PAGASA stations considered (their details are given in Table 4.1). The four stations used with illustrative purposes later during the Thesis are marked in red.

4.1 Observed Climate

The Philippines is an archipelago of 7107 islands with complex topography (see Figure 4.1a) located between the monsoonal and inner tropics (4°N and 20°N latitudes). Apart from ENSO, the climate of this region is affected by large-scale processes; mainly the southwest summer and northeast winter monsoons of the western North Pacific Ocean (Wang, 2002), but also by local forcings related to the presence of mountains (windward/leeward differences) and coastlines (land-sea breezes) (Robertson et al., 2012). As a result, the country exhibits a variety of climates in a relatively small area. For a good representation of this large variability, daily precipitation from 42 gauges maintained by PAGASA and uniformly distributed across the Philippines (see Figure 4.1b) was available for this Thesis for the period 1981-2005. These stations, whose details are given in Table 4.1, were selected after a rigorous quality-control, minimizing thus the predictand-induced

Number	CT	Name	Longitude (°)	Latitude (°)	Height (masl)
1	1	LAOAG CITY	120.6	18.2	5
2	1	VIGAN	120.4	17.6	33
3	1	BAGUIO	120.6	16.4	500
4	1	DAGUPAN CITY	120.3	16.0	2
5	1	MUÑOZ-NUEVA ÉCIJA (UCLS)	120.9	15.7	76
6	1	CABANATUAN	121.0	15.5	32
7	1	IBA	120.0	15.3	5
8	1	SCIENCE GARDEN	121.0	14.7	43
9	1	PORT AREA (MCO)	121.0	14.6	16
10	1	SANGLEY POINT	120.9	14.5	3
11	1	AMBULONG	121.1	14.1	10
12	1	SAN JOSÉ	121.0	12.4	0
13	1	ILOILO CITY	122.5	10.7	8
14	2	CASIGURAN	122.1	16.3	4
15	2	INFANTA	121.7	14.7	7
16	2	DAET	123.0	14.1	4
17	2	TAYABAS	121.6	14.0	158
18	2	VIRAC (SYNOP)	124.2	13.6	40
19	2	VIRAC (RADAR)	124.2	13.4	233
20	2	LEGASPI CITY	123.7	13.1	17
21	2	CATARMAN	124.6	12.5	50
22	2	GUIUAN	125.7	11.0	60
23	2	HINATUAN	126.3	8.4	3
24	3	APARRI	121.6	18.4	3
25	3	TUGUEGARAO	121.7	17.6	62
26	3	CALAPAN	121.2	13.4	40
27	3	ROMBLON	122.3	12.6	47
28	3	MASBATE	123.6	12.4	6
29	3	ROXAS CITY	122.8	11.6	4
30	3	MACTAN INT L AIRPORT	124.0	10.3	13
31	3	PUERTO PRINCESA	118.7	9.7	16
32	3	DUMAGUETE CITY	123.3	9.3	8
33	3	LUMBIA AIRPORT	124.6	8.4	182
34	3	ZAMBOANGA	122.1	6.9	6
35	4	TACLOBAN CITY	125.0	11.3	3
36	4	MAASIN	124.8	10.1	72
37	4	TAGBILARAN CITY	123.9	9.6	6
38	4	BUTUAN	125.5	8.9	18
39	4	DIPOLOG	123.3	8.6	4
40	4	MALAYBALAY	125.1	8.2	627
41	4	DAVAO CITY	125.8	7.3	18
42	4	GENERAL SANTOS	125.2	6.1	15

Table 4.1: Details of the 42 PAGASA stations considered in this Thesis, which represent the four climate types (CTs) present across the country (see the text for details).

uncertainty (Hewitson et al., 2014), which might be an issue for the statistical downscaling methods applied later during the Thesis —uncertainty in the observations may potentially weaken the predictor–predictand relationship (Maraun et al., 2010).— Hereafter, they are classified into the four precipitation climate types (CTs) defined in Coronas (1920) (see Figure 4.2a), which have been commonly used to-date.

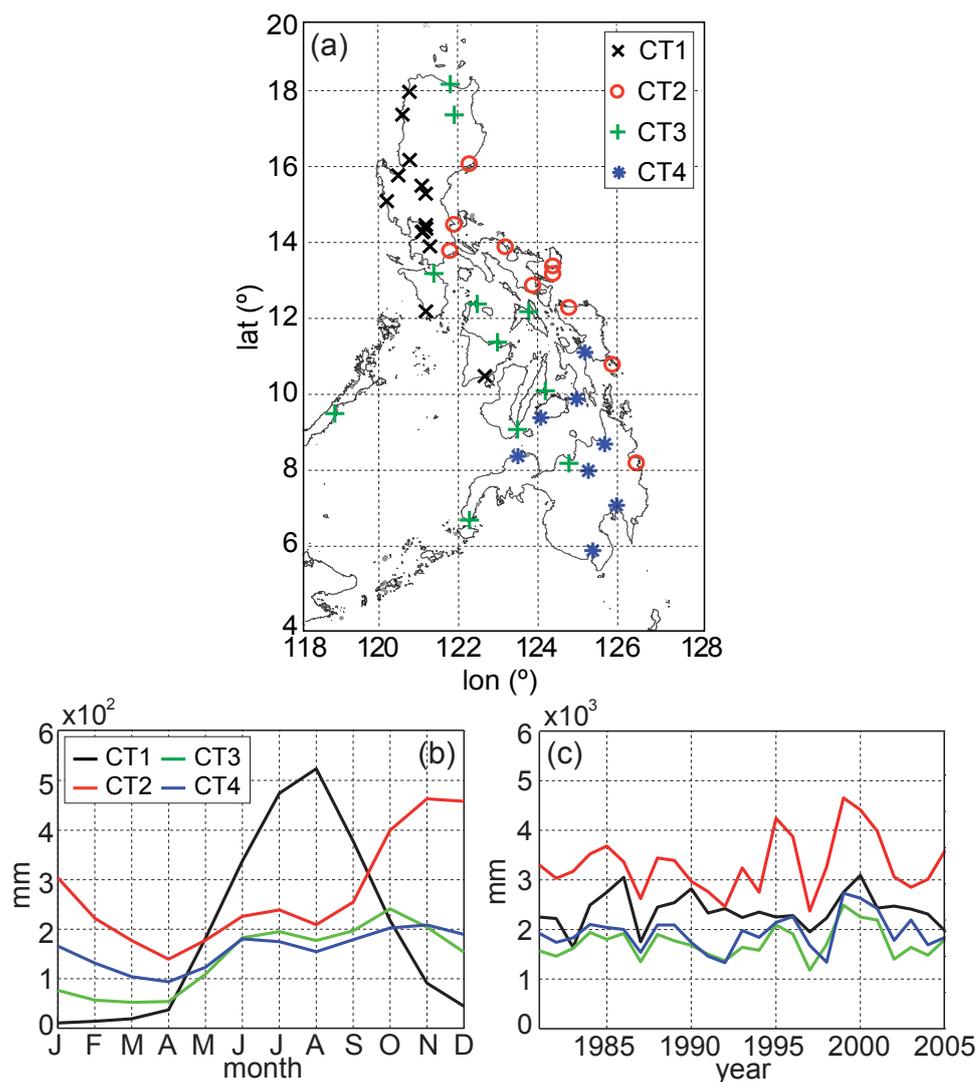


Figure 4.2: (a) Topography of the Philippines. (b) Location of the 42 PAGASA stations considered, classified into the four precipitation climate types (CTs) —in colors— described in the text. (c) Intra and (d) interannual variability of spatial average precipitation amount for each CT for the period 1981-2005.

It can be seen from Figure 4.2*b* that precipitation along the coastlines of the northern part of the archipelago (CT1 and CT2) exhibits a strong seasonal cycle, which is driven by alternating monsoonal winds. In particular, during the southwest monsoon (June-September), precipitation peaks at the stations pertaining to CT1 while CT2 is affected by relative dryness. However, the opposite is the case during the northeast monsoon (October-February). During the dry months (March-May), easterly winds prevail, leading to orographic precipitation along the mountain ranges in the east of the archipelago and to relatively high precipitation amounts for the stations pertaining to CT2. At the stations belonging to CT3 and CT4 (mainly situated in the center and south of the archipelago), precipitation is bounded to meso-scale dynamics and is not directly driven by the monsoons, leading to a weak seasonal cycle. Additionally, interannual variability is larger for CT1 and CT2 than for CT3 and CT4 (Figure 4.2*c*). For further spatial detail of these climatic features, top (bottom) row in Figure 4.3 shows the mean climatology (standard deviation) of the 42 stations for the four standard boreal seasons: winter (DJF), spring (MAM), summer (JJA) and autumn (SON), in columns.

For a more comprehensive description of the climate of the Philippines, the interested reader is referred to Coronas (1920), Flores and Balagot (1969), Kintanar (1984) as well as to the PAGASA website.

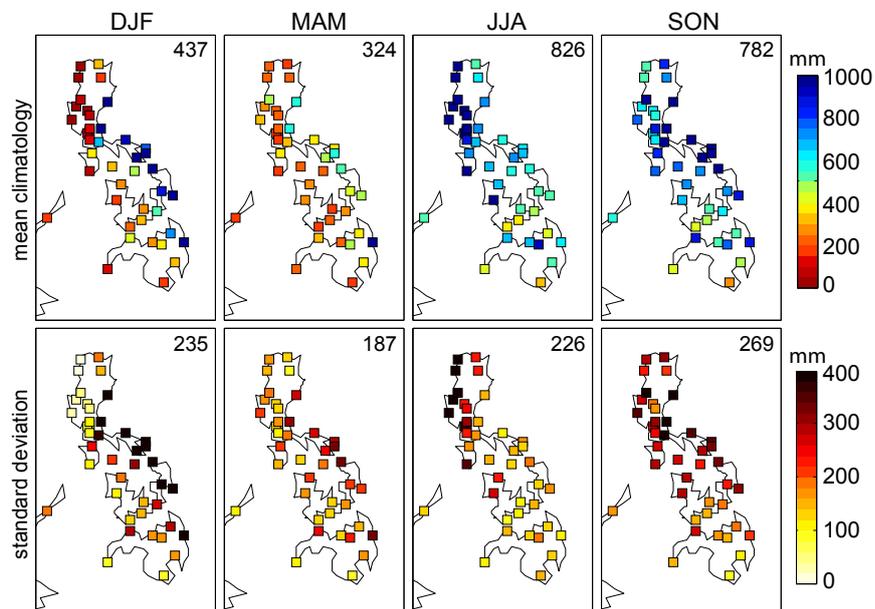


Figure 4.3: Top (bottom) row: Mean climatology (standard deviation) for each of the 42 PAGASA stations (Figure 4.1*b*) for the period 1981-2005, by seasons (in columns). Numbers in the maps correspond to the spatial average values.

4.2 From Global to Local: The Need for Downscaling

As mentioned in Chapter 3, state-of-the-art seasonal GCMs numerically solve the equations characterizing the dynamics of the climate system on a global grid. However, the coarse horizontal resolution of this grid (usually ranging from 0.75° to 1.5°) is insufficient to provide the regional or local climate information needed for practical applications. In particular, GCM outputs are areal-averages representative of the entire grid box and cannot thus represent the local variability occurring at particular points of interest (Luo et al., 2013). For instance, two nearby locations with different climates (e.g., one at the bottom of the valley and the other on top of the mountain) might be represented by the same model grid box, leading therefore to important forecast errors.

To illustrate these limitations, Figures 4.4 and 4.5 show the validation results—in terms of bias (in mm/season) and interannual ACC, respectively—obtained for four of the five ENSEMBLES models for the 42 stations of Figure 4.1*b* (models are bi-linearly interpolated to the points of interest) during the period 1981-2005. In both cases, significant ($\alpha = 0.05$) values are indicated with a black dot. For brevity, only one month lead-time predictions are shown.

These figures show that large differences are obtained for different models and seasons at a regional level. Local biases are in general strong (as compared with the climatologies shown in Figure 4.3). Remarkably, all models show a tendency towards a dry bias for the stations belonging to the CT1 (CT2) in JJA (SON and DJF), when the southwest (northeast) monsoon takes place and thus the most severe socio-economic impacts occur. However, acceptable local interannual correlations are found throughout the year (especially in DJF and MAM) except for JJA. In particular, the low skill found in this season for the CT1 region reflects the models' limitation to predict the interannual variation of the southwest monsoon.

For a more detailed interpretation of the models' performance, Figure 4.6 shows in black the time-series of annual accumulated rainfall for the four illustrative stations marked in red in Figure 4.1*b*: Ambulong, Catarman, Roxas City and Butuan (in rows). Note that each of these stations belongs to a different CT and presents therefore a different precipitation regime. For each GCM (in columns), the solid (dashed) red lines corresponds to the ensemble mean (each of the nine members). Numbers on top of each panel indicate the interannual ACC, the bias and the ratio of variances (RV). The latter is calculated as $\sigma_{model}^2 / \sigma_{observations}^2$.

Biases are strong (above 1000 mm/season) in many cases and can vary substantially among models for a particular location, which evidences that global seasonal forecasts require some kind of *calibration* which satisfactorily translates them to the local-scale useful

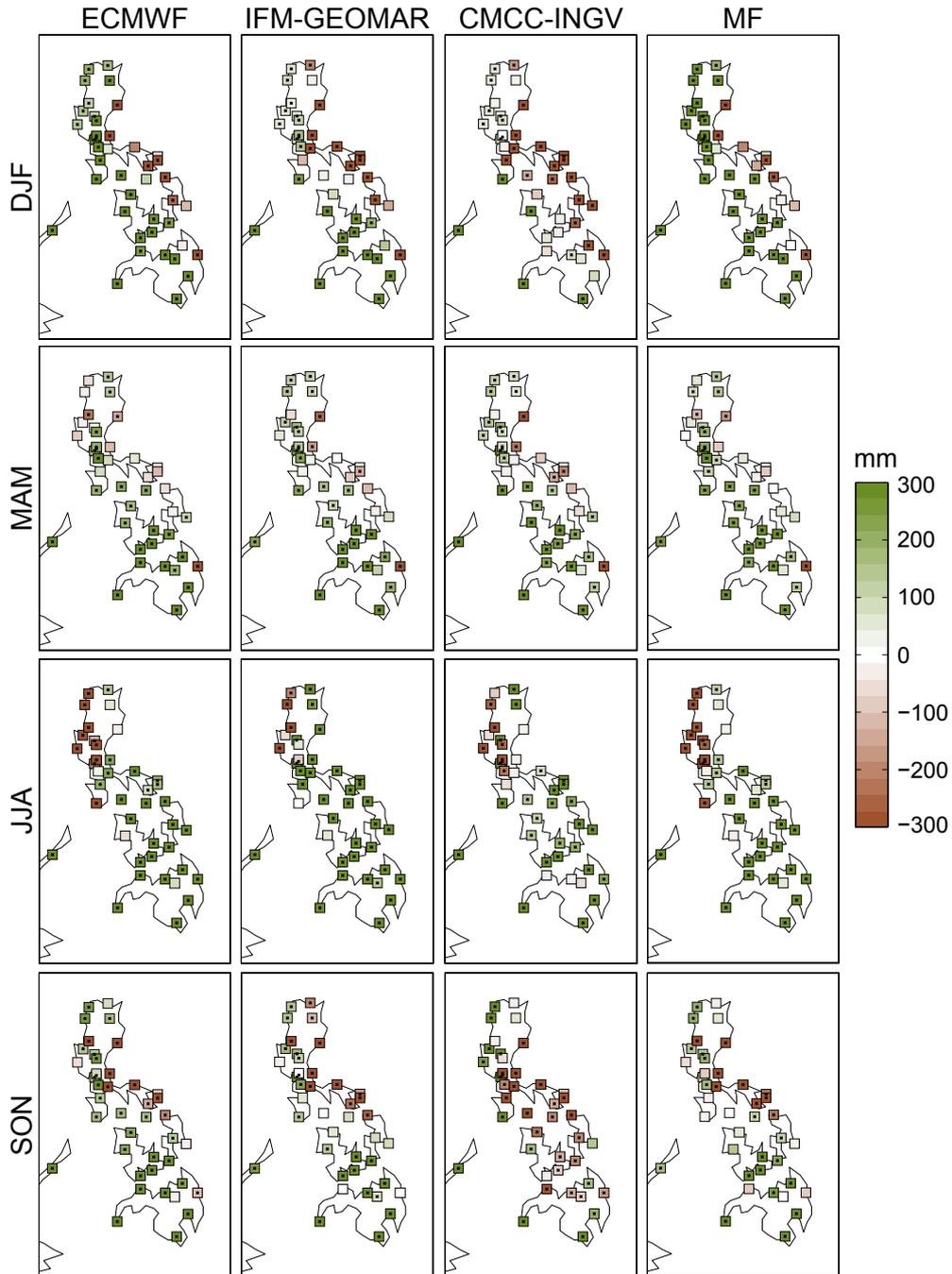


Figure 4.4: Bias (in mm/season) for four of the ENSEMBLES models (in columns) for the 42 PAGASA stations (Figure 4.1b) at one month lead-time for the period 1981-2005, by seasons (in rows). Significant ($\alpha = 0.05$) values are indicated with a black dot. A Student's t-test was applied to compute this significance. The Satterthwaite's approximation for the effective degrees of freedom was considered.

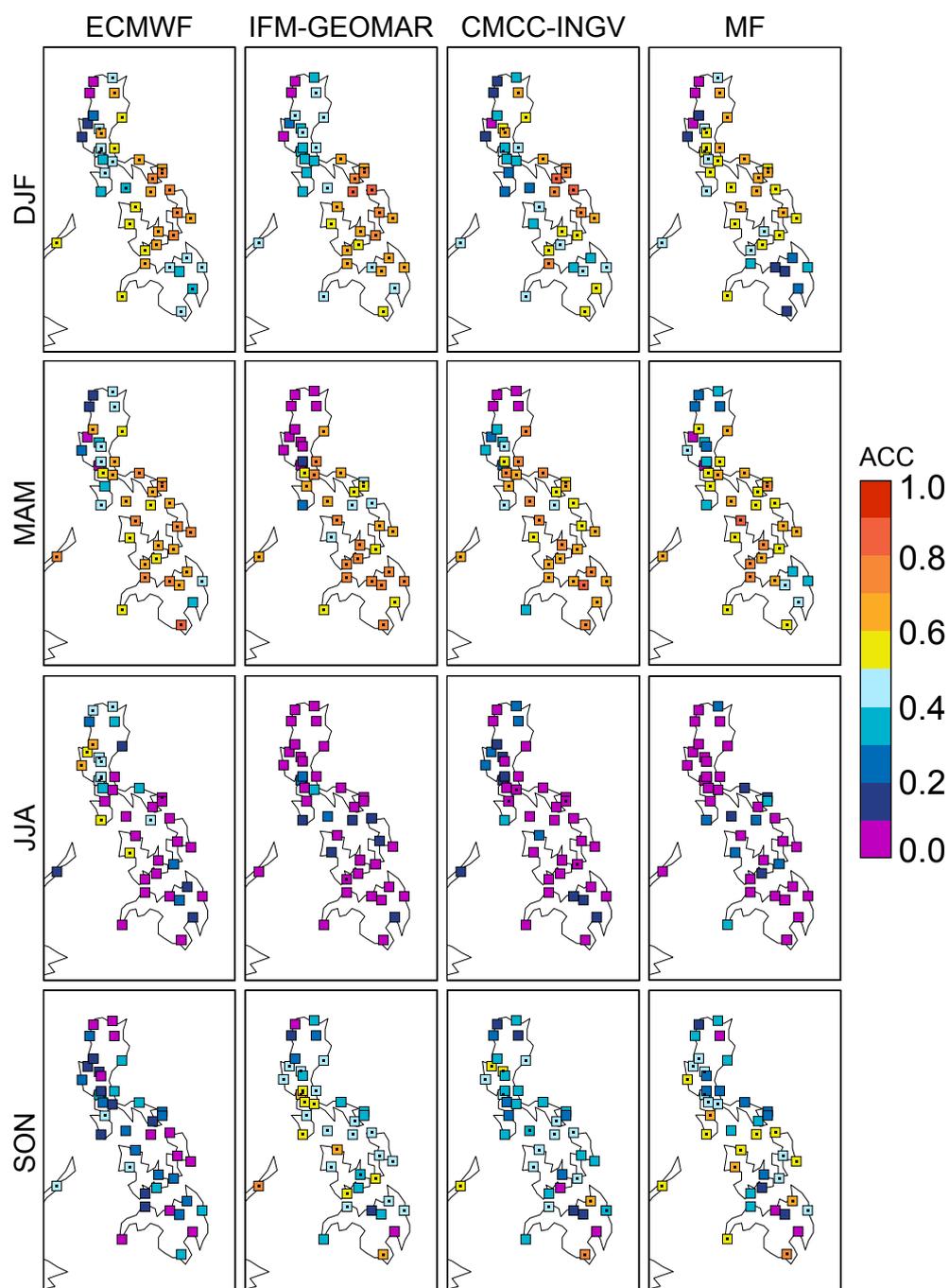


Figure 4.5: Interannual ACC maps for four of the ENSEMBLES models (in columns) for the 42 PAGASA stations (Figure 4.1b) at one month lead-time for the period 1981-2005, by seasons (in rows). Significant ($\alpha = 0.05$) values are indicated with a black dot. A Student's t-distribution with $N - 2$ degrees of freedom ($N =$ number of years) was considered to compute this significance.

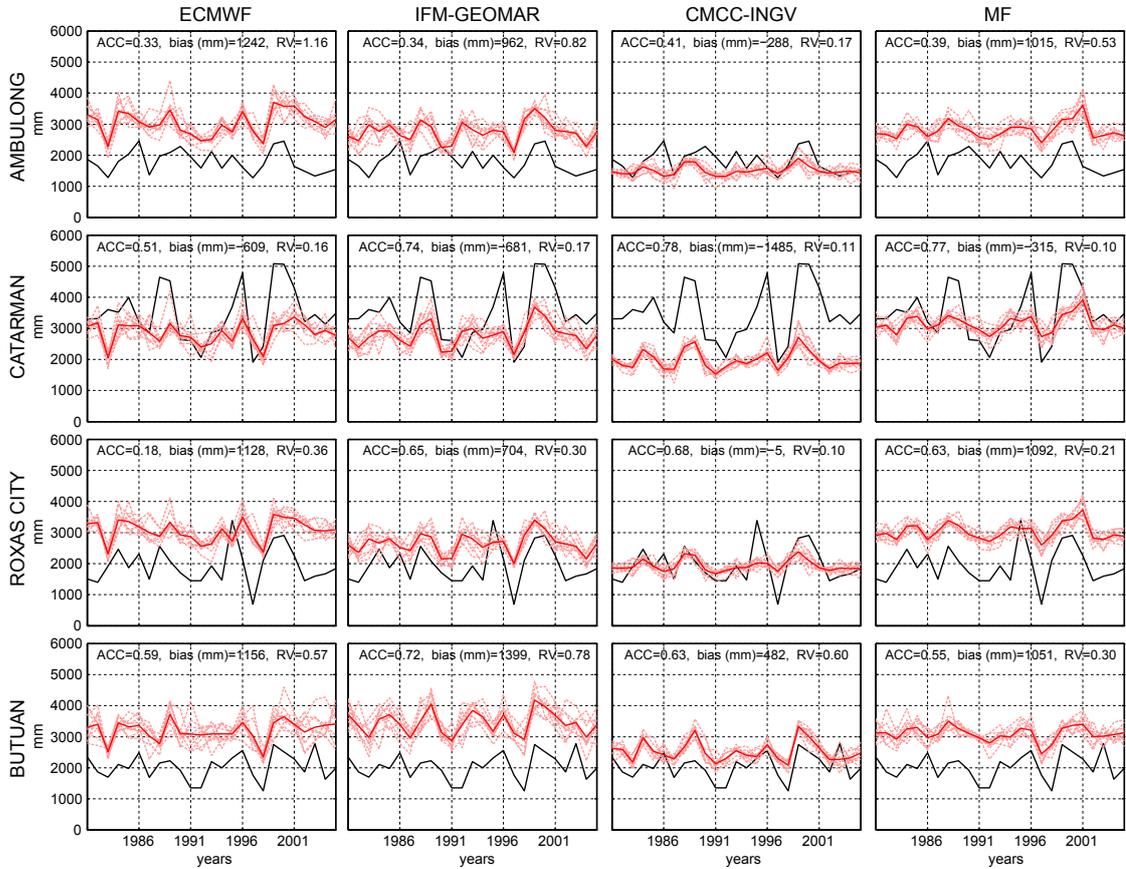


Figure 4.6: Time-series of annual accumulated rainfall at the four representative stations marked in red in Figure 4.1*b*: Ambulong, Catarman, Roxas City and Butuan (in rows). For each model (in columns), the solid (dashed) red line corresponds to the ensemble mean (each of the nine members). Observations are displayed in black. The numbers on top of each panel indicate the interannual ACC, the bias and the ratio of variances (RV).

for the different impact sectors. Moreover, note that beyond a generalized tendency to underestimate the observed variability, RV also varies among models. Therefore, rather than simple factor-scaling corrections, more sophisticated approaches which allow for correcting the bias whilst preserving (if not improving) the skill —as represented by the interannual ACC— should be applied. To this aim, statistical downscaling techniques (which are introduced in Section 5) are used in this Thesis. However, these techniques have been developed and applied almost exclusively for extratropical regions (Hewitson and Crane, 1996; Wilby and Wigley, 1997; Trigo and Palutikof, 2001; Hanssen-Bauer et al., 2005; Fowler et al., 2007; Maraun et al., 2010; Gutiérrez et al., 2013), whereas studies are rare or even non-existent for the tropics since manifold problems still hinder their successful application in this part of the world (Hewitson et al., 2014).

Following from all the previous considerations, it is recognized that the Philippines is an ideal test-bed for studies dealing with statistical downscaling (Moron et al., 2009; Manzanas et al., 2015) of seasonal forecasts and is therefore the region considered to illustrate some of the most interesting results obtained during the realization of this Thesis, which will be presented in the next chapters.

Part II

Methodological Framework

CHAPTER 5

Statistical Downscaling (SD)

As mentioned in Chapter 3, the Global Climate Models (GCMs) used for seasonal forecasting solve the complicated differential equations that describe the global climate system in a three-dimensional space formed by grid boxes, whose typical horizontal resolution is about hundreds of kilometres (Goddard et al., 2003). Therefore, despite their paramount importance to simulate the climate system globally, GCMs are unable to provide information at the local spatial scale required by most of stakeholders (see, e.g., Doblus-Reyes et al., 2013, and references therein). Hence, some form of regionalization (or *downscaling*) is needed in order to improve their usability. To this aim, two conceptually different approaches have been developed in the last decades: dynamical and statistical downscaling. On the one hand, Dynamical Downscaling (DD) is based on the use of numerical Regional Climate Models (RCMs) which simulate regional features of the climate at a higher resolution over a limited area, driven at the boundaries by the coarse-resolution GCM outputs (see, e.g., Giorgi and Mearns, 1999; Laprise, 2008). On the other hand, Statistical Downscaling (SD) relies on statistical models/algorithms which link the coarse-resolution global outputs from the GCMs (*predictors*) with the local observations (*predictands*) over the area of interest (see, e.g., von Storch et al., 1993). The relative merits and limitations of DD and SD —summarized in Table 5.1— have been widely discussed in the literature (see, e.g., Wilby and Wigley, 1997; Fowler et al., 2007; Maraun et al., 2010; Winkler et al., 2011) and it is nowadays recognized that both approaches are complementary in many practical applications. For instance, distributional bias correction techniques (which are considered a form of SD) are routinely applied to calibrate the biased RCM outputs according to the available observations.

In this Thesis we only consider SD, which has demonstrated potential to improve

the skill of global seasonal precipitation forecasts (Feddersen and Andersen, 2005) and is drastically cheaper than DD in terms of computational resources. Noteworthy, the main shortcoming of SD is that it assumes that the predictors-predictand link remains stationary/stable in time, which has been shown to be an important issue in climate change applications (see, e.g., Gutiérrez et al., 2013). However, this is not of special relevance for the case of seasonal forecasts since out-of-sample cases are not expected to be as distinct from historical data as in climate change scenarios.

This chapter provides a comprehensive description of the different approaches and techniques available for SD as well as a detailed description of some of them, which are later applied in this Thesis.

	Strengths	Weaknesses
Dynamical	<ul style="list-style-type: none"> • Individual variables are physically consistent in time and space, and the different variables are internally consistent. • The same fundamental physical principles are used in both a RCM and a GCM. • No specific calibration data is required. 	<ul style="list-style-type: none"> • RCMs are very complex and require substantial computational resources. • Artefacts and spurious effects occur near the boundary of the RCM domain. • RCMs add their own biases to the output data (see, e.g., Christensen et al., 2008), so they need to be statistically calibrated for impact studies.
Statistical	<ul style="list-style-type: none"> • The methods are computationally cheap. • Many different statistical methods are available, allowing for substantial flexibility. • Directly incorporate observations into the model, so no further calibration/correction is needed for the resulting outputs. 	<ul style="list-style-type: none"> • A calibration dataset, typically a long meteorological record of high-quality observations is required. • The models are empirical, not based on physical principles, so temporal, spatial and inter-variable consistency are typically not granted. • Assumes stationary statistical relationships, which has been shown to be an issue for climate change projections (see, e.g., Gutiérrez et al., 2013).

Table 5.1: Summary of strengths and weaknesses of dynamical and statistical downscaling.

5.1 Approaches for Statistical Downscaling

As previously mentioned, SD relies on statistical models/algorithms which link the coarse-resolution global simulated *predictors* with the local observed *predictands* over the area of interest (see, e.g., von Storch et al., 1993). These statistical models/algorithms are first calibrated using historical data of both coarse predictors and local predictands for a representative climatic period (usually a few decades) and then applied to new (e.g., future or retrospective) low-resolution global predictors to obtain the corresponding local downscaled predictands. Therefore, all SD approaches consist of a *calibration/training* and a *prediction* phase (*a* and *b* in Figure 5.1, respectively). According to the nature of predictors in the calibration phase, two different approaches for SD exist, namely Perfect Prog (PP) and Model Output Statistics (MOS) (see Marzban et al., 2006, for an interesting discussion on this). They are described next.

5.1.1 Perfect Prog (PP)

Under the PP approach (see, e.g., Charles et al., 1999; Timbal et al., 2003; Bürger and Chen, 2005; Haylock et al., 2006; Fowler et al., 2007; Hertig and Jacobeit, 2008; Sauter and Venema, 2011; Gutiérrez et al., 2013) *quasi-observed* predictors from reanalysis¹ (*c* in Figure 5.1) are used to calibrate the statistical models/algorithms (*1* and *2* in Figure 5.1*a*). Afterwards, the resulting statistical models/algorithms are applied to GCM predictor data (e.g. the seasonal forecasts from the ENSEMBLES models) in the prediction phase (*1* and *2* in Figure 5.1*b*).

Therefore, large-scale circulation variables well represented by both reanalyses and GCMs are typically chosen as predictors, whereas variables directly influenced by model parametrizations and/or orography (e.g. precipitation) are usually not considered in this approach (see that *c3* and *c4* are ruled out in Figure 5.1). Moreover, it is recognized that suitable predictors for PP should account for a major part of the variability in the predictands (Wilby et al., 2004; Hanssen-Bauer et al., 2005). As a result from these considerations, one of the most time-consuming tasks in PP is the selection of an appropriate combination of predictors for each particular predictand and region of study. Typically, common PP predictors for precipitation are sea-level pressure, specific humidity, geopotential and winds, which describe atmospheric thickness, moisture, stability and circulation.

One of the main shortcomings of PP is that reanalysis data does not necessarily provide a ‘perfect’ representation of the large-scale circulation —for instance, Brands et al. (2012)

¹Reanalyses combine assimilated meteorological observations with a numerical model which generates a synthesized estimate of the state of the climate system and are, therefore, considered as *quasi-observations*. Typically, reanalyses cover the entire globe —from the Earth’s surface to the upper limit of the stratosphere— and extend for several decades. See <http://reanalyses.org> for further details.

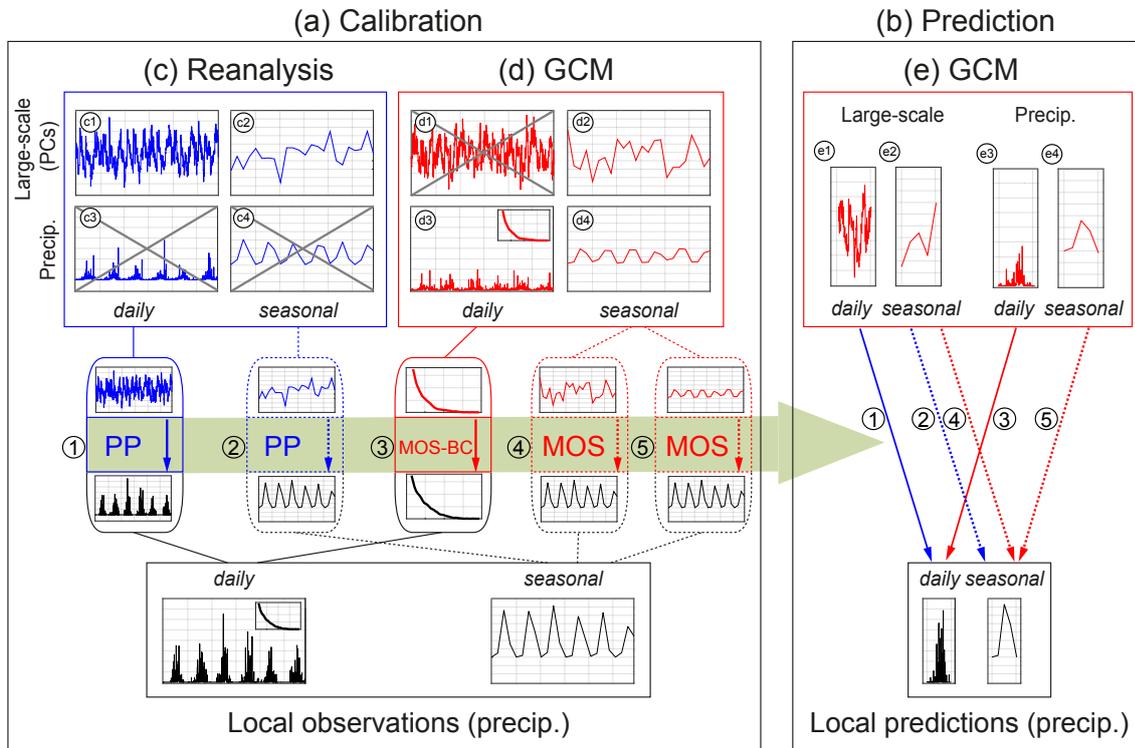


Figure 5.1: Schema of the different approaches for SD according to the type of data used for calibration (reanalysis or GCM) and the temporal scale considered (daily or seasonal). See the text for details.

found significant differences between two distinct reanalyses for key predictor variables, especially in the tropics.— Therefore, the uncertainty due to ‘imperfect’ reanalyses should be taken into account in this approach, in particular in tropical regions (Manzanas et al., 2015). This issue will be analyzed in detail in Chapter 8. In addition, the PP approach is sensitive to the systematic biases of the GCMs (as compared to reanalyses). Thus, appropriate preprocessing (or *harmonization*) of the predictor data considered is required to obtain meaningful results. The harmonization process followed in this Thesis is described in Section 6.3.

Finally, PP techniques can consider point-wise and/or spatial-wise predictors, using the values at nearby grid boxes and/or the Principal Components (PCs) corresponding to the Empirical Orthogonal Functions (EOFs) (Preisendorfer, 1988) of the variables considered over a representative geographical domain (which must be also conveniently determined). The use of predictor values at nearby grid boxes or PCs depends on the application. Usually, the latter are more informative in those cases where the local climate is mostly determined by synoptic phenomena whereas the former may be needed to add some infor-

mation about the local variability in those cases where small-scale processes are important. Sometimes, both type of predictors are combined in order to account for both synoptic and local effects. Working with PCs allows to filter-out high frequency variability which may be not properly linked to the local-scale. Moreover, it prevents from the negative effects related to the different ranges of the different predictor variables considered since it involves the standardization of the raw fields. For this reason, in the PP techniques considering point-wise predictors used in this Thesis, we always use standardized anomalies rather than raw values.

5.1.2 Model Output Statistics (MOS)

Differently from PP, under the MOS approach (see, e.g., Ward and Navarra, 1997; Feddersen et al., 1999; Mo and Straus, 2002; Sokol, 2003; Kang et al., 2004; Marzban et al., 2006; Vannitsem and Nicolis, 2008), predictors are taken from the same GCM for both the calibration and the prediction phase (d and e in Figure 5.1, respectively). Typically², the only predictor variable considered in this approach is the model counterpart of the target predictand, e.g., coarse GCM precipitation for local precipitation ($d3$ and $d4$ in Figure 5.1). Therefore, as opposite to the PP case, the main advantage of the MOS approach is that neither predictor nor domain screening is required.

Distributional Bias Correction (BC) techniques are the simplest case of a MOS scheme. In particular, the BC techniques considered in this Thesis (which will be referred to as MOS-BC hereafter) operate directly on the coarse GCM precipitation (interpolated to the location of interest), correcting it at a distributional level, based on local observations (β in Figure 5.1). Note thus that MOS-BC techniques are specific for the GCM for which they have been calibrated and cannot be used with other models (Maraun et al., 2010). Note also that the relative order of the data time-series (temporal structure) is not explicitly considered in these techniques. Nevertheless, MOS techniques can be also calibrated taking into account the temporal correspondence between simulations and observations considering either large-scale GCM predictors or GCM precipitation (4 and 5 in Figure 5.1, respectively), being the latter case the most popular due to its simplicity. However, the day-to-day temporal correspondence between GCM simulations and observations is negligible³ and, therefore, the application of MOS techniques (excluding MOS-BC) at a

²There are MOS techniques which consider as predictors large-scale variables from GCM instead of precipitation ($d2/e2$ in Figure 5.1). However, these techniques operate on a seasonal basis and are thus not considered in this Thesis. See the text for details.

³Recently, reanalysis-driven RCMs (Turco et al., 2011) and reanalysis-nudged GCMs (Eden et al., 2012) have been successfully applied for SD under the MOS approach on a daily time-scale, due to the temporal correspondence between observed and simulated precipitation. However, these techniques require the availability of the simulations performed with the GCM (or RCM) nudged to (or driven by) reanalysis, what makes them unsuitable for most practical applications.

daily scale is not possible. The advantage of MOS techniques working on longer than daily (e.g. monthly or seasonal) basis is that linear procedures can be applied due to the normality of monthly/seasonal data. Canonical Correlation Analysis (CCA) (see, e.g., Feddersen et al., 1999; Landman and Tennant, 2000; Sinha et al., 2013) and Singular Value Decomposition Analysis (SVDA) (see, e.g., Pavan et al., 2005; Chu et al., 2008; Tung et al., 2013) are typical examples of these kind of techniques. In general, CCA and SVDA provide similar results, and any of them can be claimed to do better than the other (see Feddersen et al., 1999, for a comparison).

5.1.3 *PP versus MOS*

Either in PP or in MOS (excluding MOS-BC) techniques, the statistical models/algorithms are calibrated at an event-wise (time-series) level, i.e., preserving the temporal — e.g. day-to-day/month-to-month/season-to-season — matching between predictors and predictands (*1*, *2*, *4* and *5* in Figure 5.1). As mentioned, large-scale variables from reanalysis are typically considered as predictors in the PP approach (*c1-2/e1-2* in Figure 5.1), whereas GCM precipitation is used in MOS applications (*d3-4/e3-4* in Figure 5.1). Therefore, since the day-to-day correspondence with the observations is preserved in the case of reanalysis, PP techniques can be applied on a daily (or longer) basis, whereas MOS techniques (excluding MOS-BC) require working at longer (e.g. seasonal) time-scales. As a result of considering aggregated data, PP and MOS (excluding MOS-BC) techniques working at a seasonal scale (*2/4* and *5* in Figure 5.1) are typically expected to lose part of the potential predictive capacity of the PP techniques working on a daily basis (which are foreseen to capture more explicative relationships between predictors and predictand). Moreover, PP and MOS (excluding MOS-BC) techniques working at a seasonal scale are prone to poor statistics due to the small sample sizes available for calibration —most of the current seasonal hindcasts have less than 30 years of data— and can lead to artificial skill due to statistical over-fitting if no proper cross-validation (this will be explained in Section 5.2) is applied (Robertson et al., 2012). Finally, some form of weather generator is needed to disaggregate to the daily scale (see, e.g., Paeth and Diederich, 2011) in these techniques.

As a consequence from all the previous considerations, only PP techniques working on a daily basis and MOS-BC ones operating on the distributions of daily values have been considered in this Thesis (*1* and *3* in Figure 5.1, respectively), which allows for robust calibration due to the large sample sizes involved.

5.2 Cross-Validation

When assessing the performance of any SD technique it is crucial to properly cross-validate the results; otherwise, misleading conclusions may be obtained. Leave-one-out (Lachenbruch and Mickey, 1968) and k -fold (Markatou et al., 2005) are the most common approaches for cross-validation (Stone, 1974).

In a k -fold cross-validation framework, the original sample is partitioned into k equal sized subsamples. In each of the k iterations (folds), one of these subsamples is retained for test (prediction phase) and the remaining $k - 1$ subsamples are used for training (calibration phase). The resulting k independent samples are then merged to produce a single time-series covering the whole calibration period, which is subsequently validated against observations. When $k = n$ (being n the number of observations), the k -fold cross-validation is exactly the leave-one-out cross-validation. Gutiérrez et al. (2013) introduced a 5-fold cross-validation framework (see Figure 5.2) which is used throughout this Thesis. Note that both leave-one-out and k -fold cross-validation approaches provide similar results, especially when the size of the available data becomes large (Markatou et al., 2005), as it is the case in this Thesis. However, the latter is computationally cheaper.

PP techniques are first cross-validated using reanalysis predictors in order to evaluate their performance in ‘perfect’ conditions before being applied to ‘non-perfect’ GCM predictors. Therefore, the aim of cross-validation in the PP approach is to properly estimate, given a known predictor dataset (large-scale variables from reanalysis), the performance of the particular technique considered, having an upper-bound for its generalization capability when applied to new predictor data (large-scale variables from GCM).

In the case of MOS-BC techniques, derived from the fact of using the same predictor (GCM precipitation) for both the calibration and the prediction phases, proper cross-validation is needed in order to avoid model-over-fitting (which could lead to artificial skill) if the periods considered for the two phases overlap. Note that, for those techniques working on a monthly/seasonal basis (not considered in this Thesis), cross-validation would be especially important due to the limited amount of available data.

5.3 Techniques for Statistical Downscaling

There have been some previous classificatory attempts in order to give a structured overview of the large number of SD techniques developed in the last decades. Wilby and Wigley (1997) provided a first useful classification which was later adopted without major modifications by Giorgi et al. (2001) and Wilby et al. (2004). In Maraun et al. (2010), a more comprehensive approach was followed in order to encompass all previous classifications in a way adequately suited to the structure of their article. Moreover, a

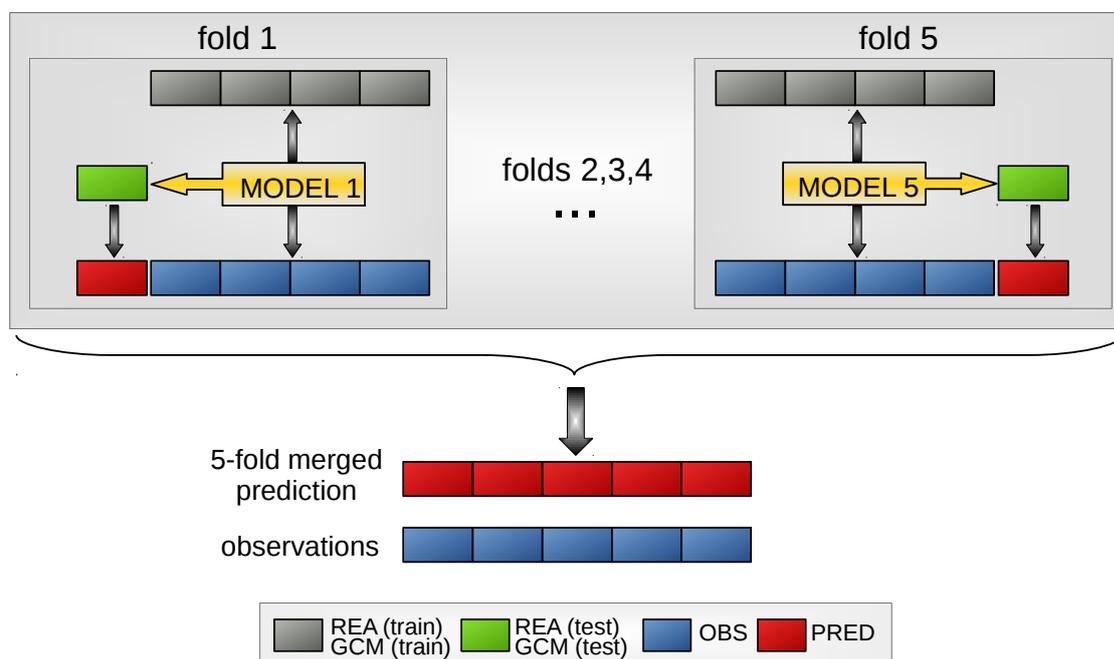


Figure 5.2: Schema of the 5-fold cross-validation used throughout this Thesis, introduced in Gutiérrez et al. (2013). The total sample is divided into five non-overlapping subsamples. In each iteration (fold), the four gray subsamples are used as train/calibration dataset and a prediction is obtained for the green one (test dataset). Finally, the five independent predictions are merged into a unique time-series covering the entire period.

slightly different classification considering *generative* and *non-generative* techniques has been recently proposed in the deliverable D52.1 (*Review of the different statistical downscaling methods for s2d prediction*) of the SPECS project, which is publicly available at http://www.specs-fp7.eu/sites/default/files/u1/SPECS_D52.1.pdf. Basically, whereas generative techniques rely on a mathematical formula (with a parsimonious number of parameters), non-generative ones are based on an algorithm. Alternatively, the *transfer functions*, *weather typing* and *weather generators* categories are also commonly used (Giorgi et al., 2001). Broadly speaking, transfer functions correspond to the generative group, whereas weather typing and weather generators are mainly non-generative. In transfer functions, a quantitative relationship between predictors and predictands is derived, typically based on either linear —e.g., regression, Principal Component Analysis (PCA), SVDA, CCA— or non-linear —e.g., Artificial Neural Networks (ANNs), Generalized Linear Models (Section 5.4.1-2)— models. Weather typing techniques are based on algorithms that infer the local predictions for a particular atmospheric configuration from the observations corresponding to a set of historical similar situations. The catalog

of historical atmospheric configurations is sometimes replaced by a pre-classification into a finite number of possible states (weather types), obtained according to their synoptic similarity (see Huth et al., 2010, for a review). Analogs (Section 5.4.1-1) and Self Organizing Maps (SOMs) (see, e.g., Hewitson and Crane, 2006) are popular examples of this category. Finally, weather generators are stochastic techniques which allow to create random time-series with realistic statistical properties (mean, variability, frequencies of extremes, length of dry and wet spells, etc.), such as the popular Markov Models (MMs) or the Hidden Markov Models (HMMs). In seasonal forecasting, these techniques act often as postprocessing tools for temporal disaggregation to go from monthly/seasonal to daily data (see, e.g., Paeth and Diederich, 2011; Hirschi et al., 2012). Weather generators are not used in this Thesis since we only consider SD techniques operating on a daily basis.

Beyond these common classifications, there are hybrid techniques which combine features from two or more of the above families as well as many intermediate possibilities (see, e.g., Feddersen and Andersen, 2005). Further information on the wide range of different SD techniques introduced in the literature can be found in Maraun et al. (2010) as well as in the downscaling inventory report from the VALUE (Validating and Integrating Downscaling Methods for Climate Change Research) COST action (<http://www.value-cost.eu>).

5.4 SD Methods Used in this Thesis

As justified in Section 5.1.3, PP and MOS-BC techniques working on a daily basis are considered in this Thesis. Hereafter, we will use the term SDM (Statistical Downscaling Method) to refer to a particular implementation of a given SD technique. In the next sections (5.4.1 and 5.4.2), the different SDMs used in this Thesis are described in detail. All of them have been applied using *MeteoLab* (a *Matlab*[®] toolbox for statistical downscaling developed by the Santander Meteorology Group which can be freely downloaded from <http://meteo.unican.es/trac/MLToolbox/wiki/Installation>). The code needed to define each of them is given in Section 10.3.

5.4.1 PP Methods

5.4.1-1 Analogs

The popular non-parametric analog technique (Lorenz, 1969; Zorita and von Storch, 1999) assumes that similar (or *analog*) atmospheric patterns over a given region (predictors) lead to similar local meteorological outcomes (predictands). The performance of this technique to downscale daily precipitation can be influenced by the type of similarity measure considered. In this Thesis we used the Euclidean norm, which has been shown to perform satisfactorily in most cases and hence is a reasonable first choice (Matulla et al., 2008). For

a given atmospheric pattern simulated by the GCM, the corresponding local forecast is estimated (according to a determined criteria) from the observations corresponding to a set of analog patterns within a historical catalog, formed by reanalysis data in a representative climatological period. This is illustrated in Figure 5.3, where a particular geographical domain (marked with blue crosses) is selected to characterize the synoptic phenomena relevant for the predictand variable of interest (precipitation at the 42 PAGASA stations of Figure 4.1*b*, marked in red), as given by a number of informative predictors (winds, humidity and temperature at different vertical levels).

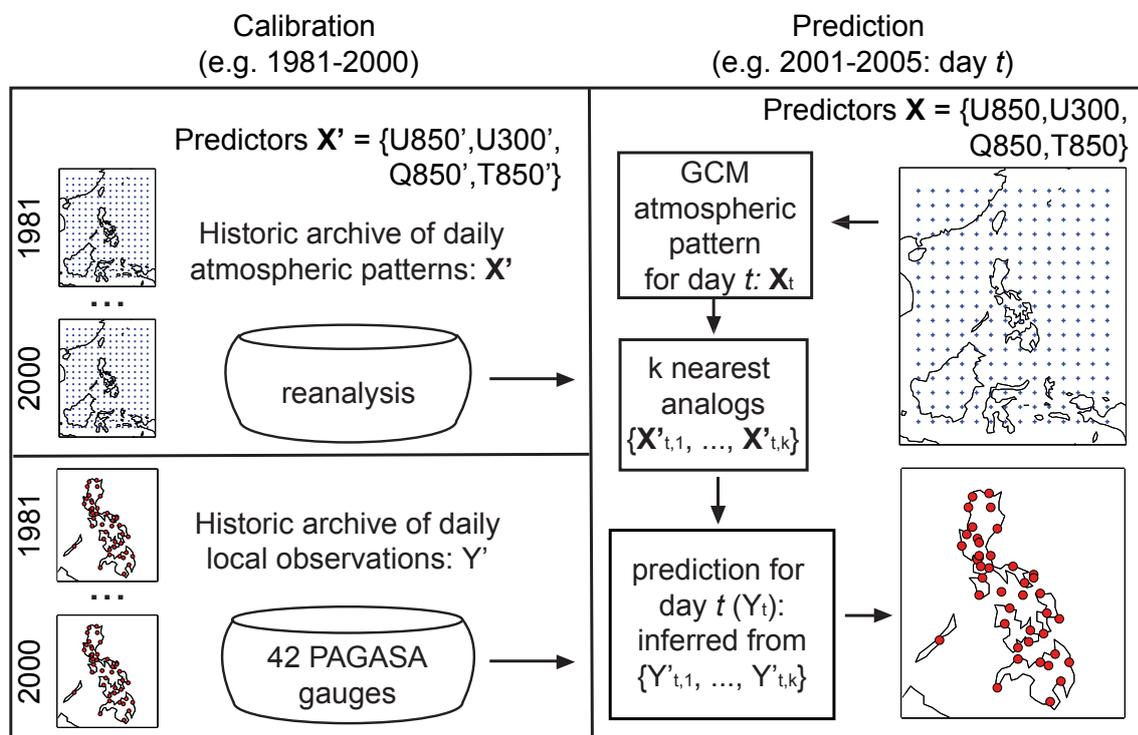


Figure 5.3: Schema illustrating the analog technique.

In spite of its simplicity, the analog technique performs as well as other more sophisticated ones (Zorita and von Storch, 1999) and it is still one of the best to downscale daily precipitation, because it is non-parametric —no assumptions are made about the distribution of the predictors— and non-linear —can take into account the non-linearity of the relationships between predictors and predictands.— Additionally, it is spatially coherent (i.e., preserves the spatial covariance structure of the local predictands).

The main drawback of the analog technique is that it cannot predict values outside the observed range, being therefore particularly sensitive to the non-stationarities arising in climate change conditions (Benestad, 2010; Gutiérrez et al., 2013) —two-step analog

methods (see, e.g., Ribalaygua et al., 2013) have been recently introduced to overcome this problem.— However, this issue is not of special relevance in seasonal forecasting. Hence, analog-based methods have been applied in several studies to downscale precipitation in the context of seasonal forecasting (see, e.g., Frías et al., 2010; Wu et al., 2012; Shao and Li, 2013).

In this Thesis, two different configurations of the analog technique (which will be referred to as *AN_det* and *AN_sto* hereafter) were considered. Whereas the former corresponds to a deterministic (Zorita et al., 1995; Cubasch et al., 1996) version which considers the closest analog, the latter is a stochastic one in which the prediction is given by random selection among the observations corresponding to the 15 closest analogs (Beersma and Buishand, 2003).

Figures 5.4 and 5.5 show the 5-fold cross-validated results (see Section 5.2) obtained for the four PAGASA stations used with illustrative purposes in Chapter 4 (marked with red points in Figure 4.1*b*) for the period 1981-2005 for the *AN_det* and the *AN_sto* methods, respectively. The predictors and geographical domain considered for this analysis are shown in Figure 5.3 (further details on the selection of this predictor-domain configuration are given in Section 6.1). The left column shows, for each station (in rows), a quantile-quantile (q-q) plot of the 1-99 observed (*x*-axis) and downscaled (*y*-axis) percentiles of daily precipitation —only wet (precipitation ≥ 0.1 mm) days are considered.— The Spearman correlation (*rs*) between the observed and downscaled daily complete time-series is indicated inside the graphs. The right column shows the observed (black) and downscaled (red) Cumulative Distribution Functions (CDFs) for wet days (values up to the 99 percentile are considered). The PDF score (Perkins et al., 2007; Maxino et al., 2008), which measures the overlap between the two wet Probability Density Functions (PDFs) is also given inside the plots (0 = no overlap at all, 1 = perfect overlap). In the stochastic method, *AN_sto*, 100 different realizations were carried out. The error bars in the right column correspond to the inter-quartile range (25-75 percentiles) of these 100 realizations. Likewise, the 100 CDFs are plotted. In both cases, a confidence interval (computed upon the 100 realizations) around the median is provided.

Although *AN_sto* yields slightly lower correlations due to its stochastic character, both methods perform alike. For a more robust interpretation of this comparative, the boxplots in the top (bottom) row of Figure 5.6 extend the results from Figure 5.4 (5.5) to the 42 PAGASA stations of Figure 4.1*b*. In the *AN_sto* method, one single realization (chosen at random) was selected —Figure 5.5 shows that the uncertainty introduced by the stochastic character of this method is small, so whichever realization can be considered to be representative of the SDM.— The first column shows the ratio between predicted and observed frequency of wet days. The second column displays the HIIt Rate (HIR, in green) and the

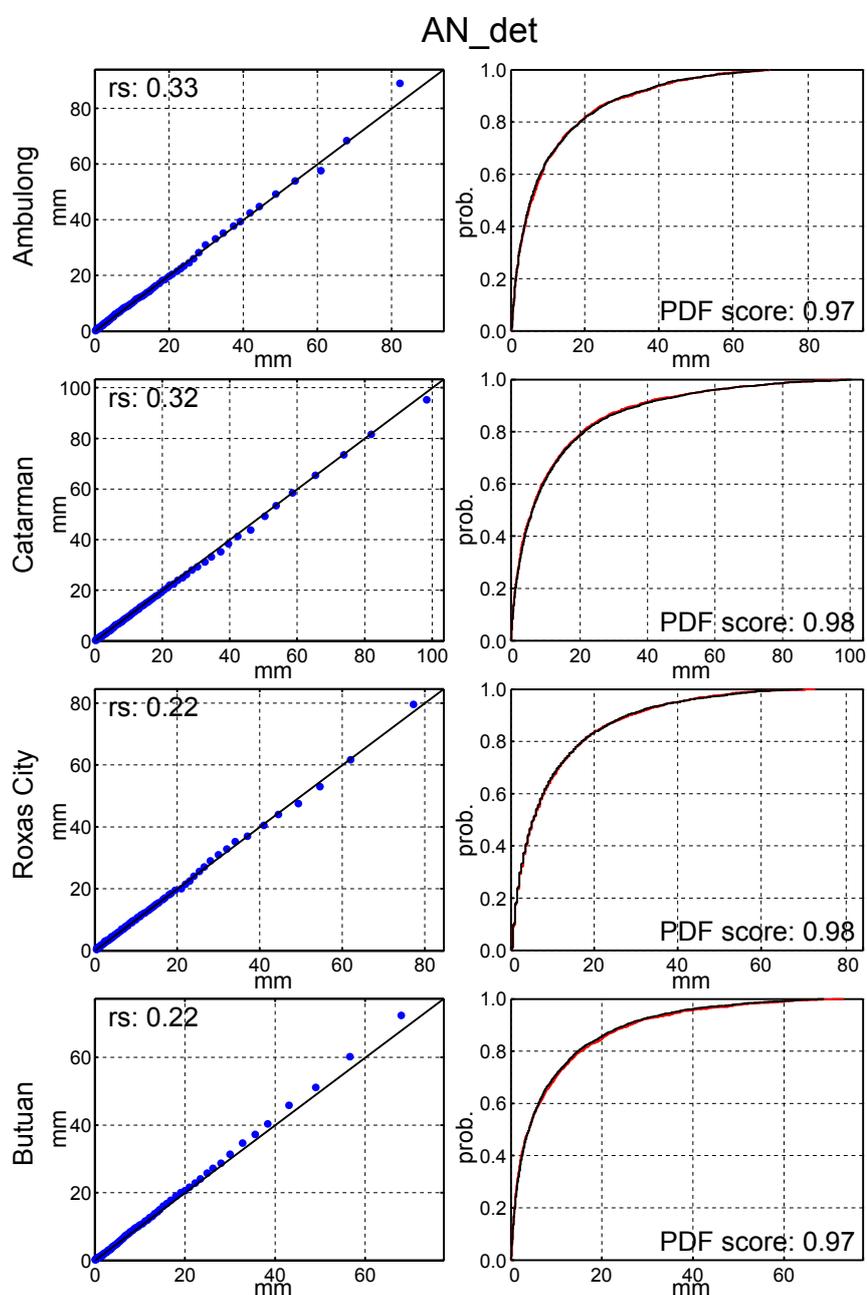


Figure 5.4: Cross-validated results obtained for the AN_{det} method for the four illustrative PAGASA stations marked in red in Figure 4.1b: Ambulong, Catarman, Roxas City and Butuan (in rows). Left column: q-q plot of the 1-99 observed (x -axis) and downscaled (y -axis) percentiles of daily precipitation —only wet (precipitation ≥ 0.1 mm) days are considered.— The Spearman correlation (r_s) between the observed and downscaled daily complete time-series is indicated inside the graphs. Right column: observed (black) and downscaled (red) CDFs for wet days (values up to the 99 percentile are considered). The PDF score is also given inside the plots. See the text for details.

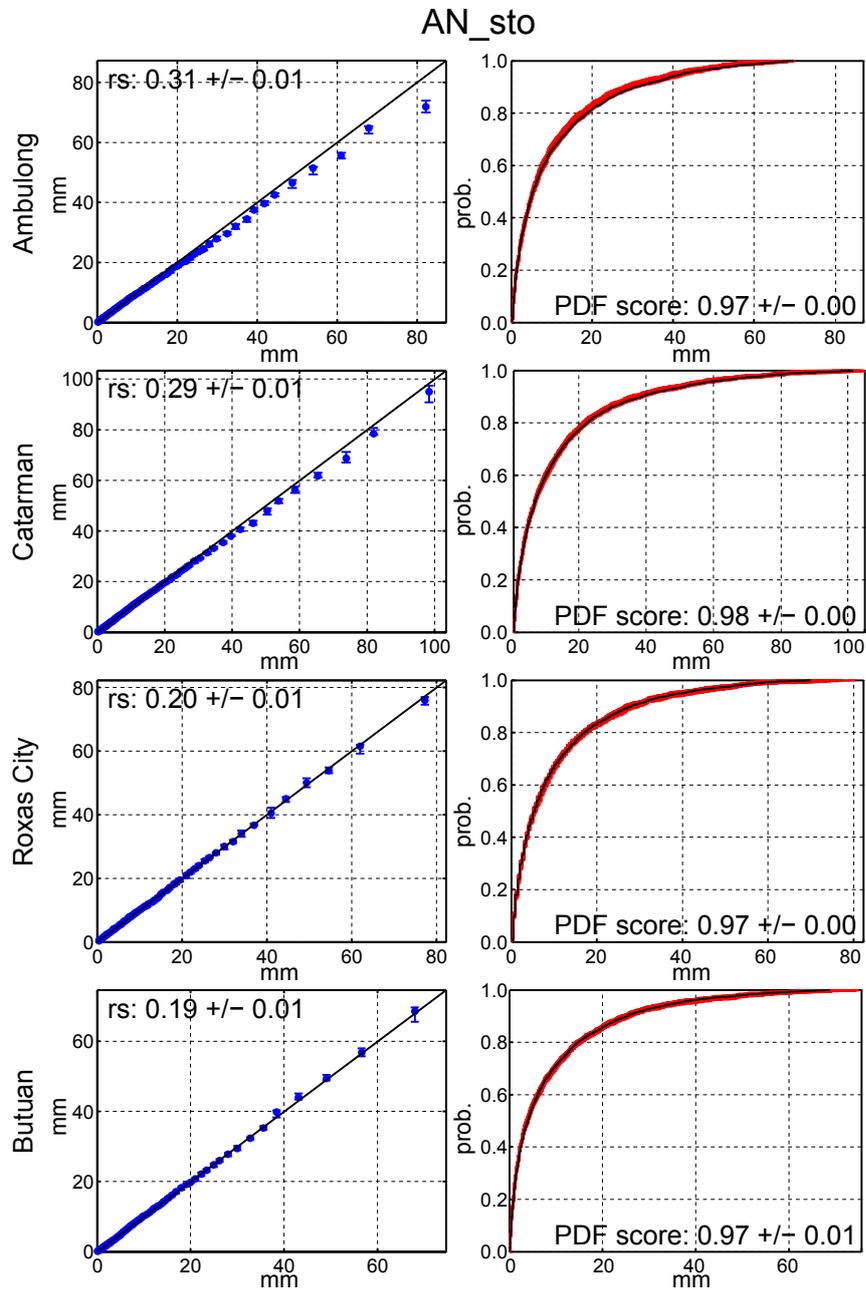


Figure 5.5: As Figure 5.4 but for the AN_{sto} method. Given the stochastic nature of this method, 100 different realizations are carried out. See the text for details.

False Alarm Rate (FAR, in red). These two columns together allow to describe how well the SDM discriminates between wet and dry days. The third column shows the correlation between observed and downscaled daily (blue), 10-daily accumulated (black) and annual accumulated (pink) complete time-series. The fourth and fifth columns plot the bias (expressed in % of the mean observed value) and the ratio of variances (expressed as the variance of predictions divided by the variance of observations) for wet days, respectively. Finally, the last column shows the PDF score between the observed and downscaled wet distributions. In the last three columns all values above 0.1 mm (up to the maximum) are considered.

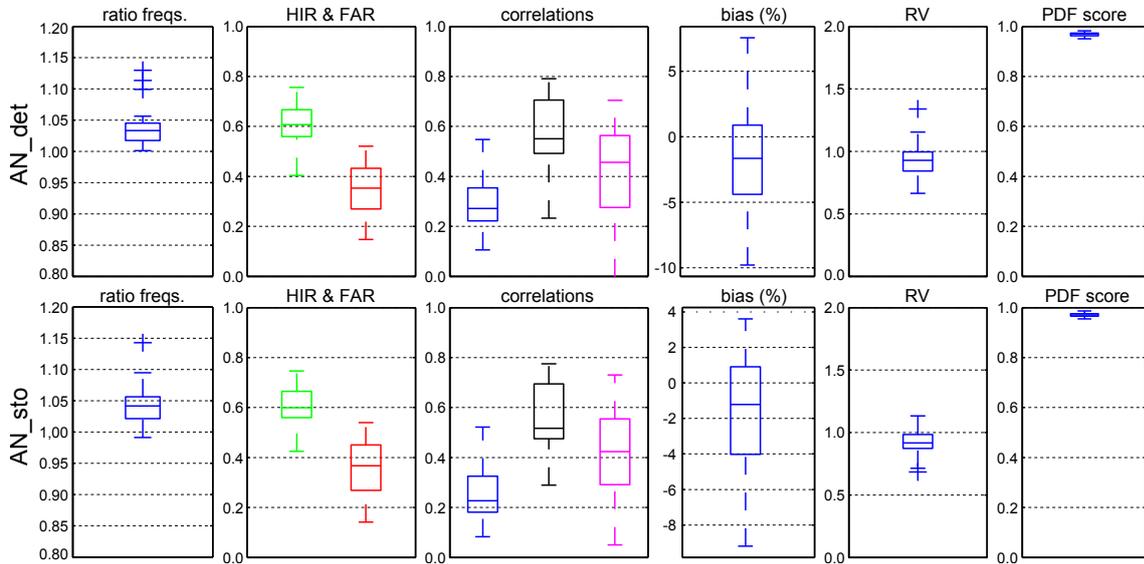


Figure 5.6: Cross-validated results from applying the (top) AN_{det} and (bottom) AN_{sto} methods for the 42 PAGASA stations shown in Figure 4.1b. In the stochastic method, AN_{sto} , one single realization (chosen at random) is considered. See the text for details.

In agreement with the results of San-Martín et al. (2016), both SDMs provide very similar results for all scores. In particular, both predict less rainy days than observed, which in turn leads to a little dry bias. Furthermore, the variance is slightly underestimated by both methods. Therefore, based on the similar performance of both AN_{det} and AN_{sto} , only the former, which yields slightly better correlations is considered in this Thesis. Note that this choice allows also to avoid the uncertainty introduced in the downscaled results (although it is shown to be small) by the random nature of the AN_{sto} method.

5.4.1-2 Generalized Linear Models

Generalized Linear Models (GLMs) were formulated by Nelder and Wedderburn (1972) in the 1970's and are an extension of the classical linear regression which allows to model the expected value of a random predictand variable whose distribution belongs to the exponential family (Y) through an arbitrary mathematical function called *link function* (g) and a set of unknown parameters (β), according to

$$E(Y) = \mu = g^{-1}(X\beta), \quad (5.1)$$

where X is the predictor and $E(Y)$ the expected value of the predictand. The unknown parameters, β , can be estimated by maximum likelihood, considering a least-squares iterative algorithm.

GLMs have been used in numerous previous studies for SD of climate change simulations (e.g., Brandsma and Buishand, 1997; Chandler and Wheeler, 2002; Abaurrea and Asín, 2005; Fealy and Sweeney, 2007; Hertig et al., 2013); however, they have rarely been used for SD of seasonal forecasts. Despite being a powerful tool to downscale precipitation, GLMs are often difficult to apply since they admit a wide variety of configurations, which should be carefully tested as they can lead to significantly different results. Given the dual (*occurrence/amount*) character of precipitation, we followed in this Thesis the common two-stage implementation (see, e.g., Chandler and Wheeler, 2002) in which a GLM with Bernoulli error distribution and *logit* canonical link-function (also known as logistic regression) is used to downscale daily precipitation occurrence (0 = no rain, 1 = rain) and a GLM with gamma error distribution and *log* canonical link-function is applied to downscale daily precipitation amount.

For analogy with linear regression techniques, we considered initially a deterministic configuration in which predictions are obtained from the expected values estimated by both the GLM for occurrence (*GLMo*) and the GLM for amount (*GLMa*). In particular, in the *GLMo*, the continuous expected values $\in [0, 1]$ are transformed into binary ones as 1 (0) if they are equal or greater (smaller) than 0.5, whereas for the *GLMa*, the expected values are directly interpreted as rain amounts. The final predicted value is therefore:

$$Y = \begin{cases} GLMa & \text{if } GLMo \geq 0.5 \\ 0 & \text{if } GLMo < 0.5 \end{cases}$$

Figure 5.7 is the equivalent to Figure 5.4, but for this deterministic method, which will be referred to as *GLM-det* henceforward. Although *GLM-det* yields better correlations than the analogs method, it is not able to predict light precipitation amounts. Furthermore, it greatly underestimates the observed variance, with most of the predicted rain values in

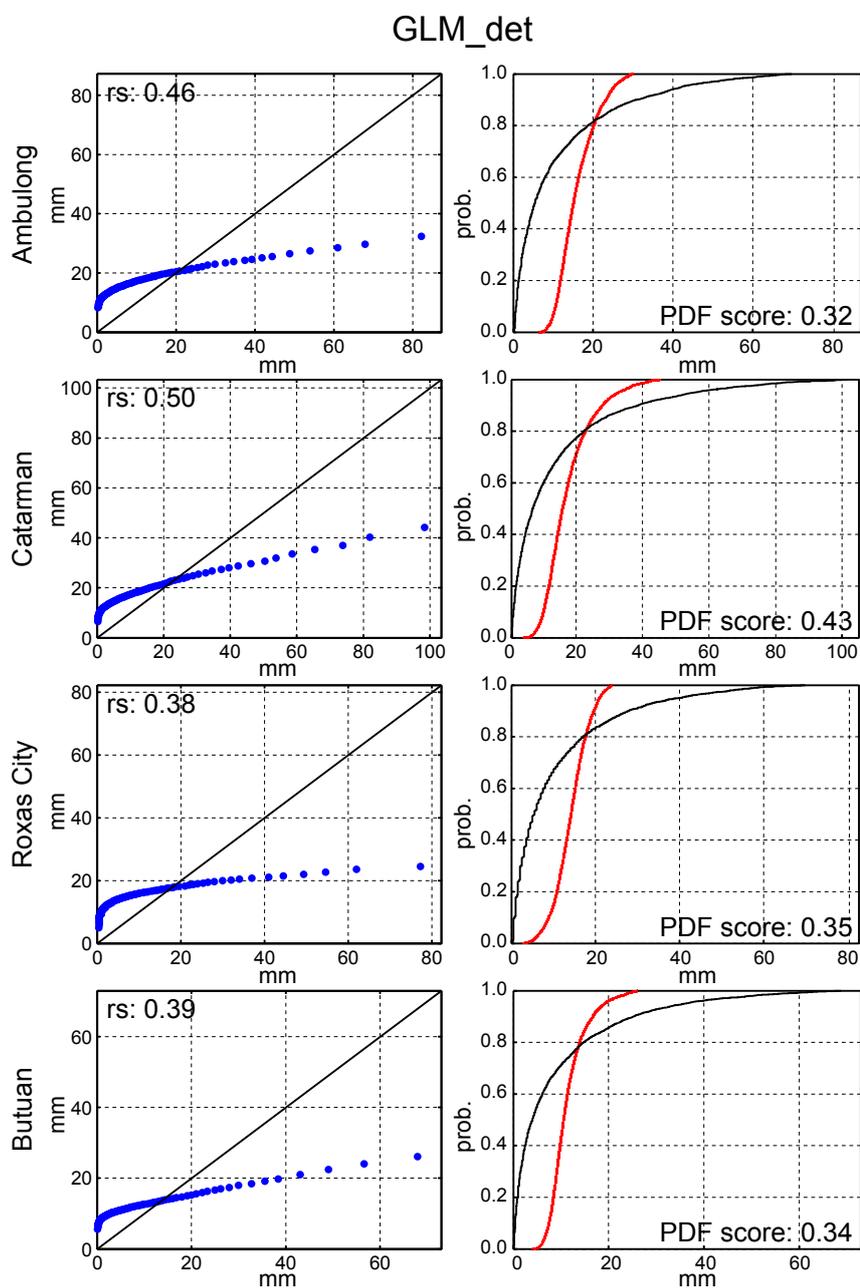


Figure 5.7: As Figure 5.4 but for the *GLM-det* method. See the text for details.

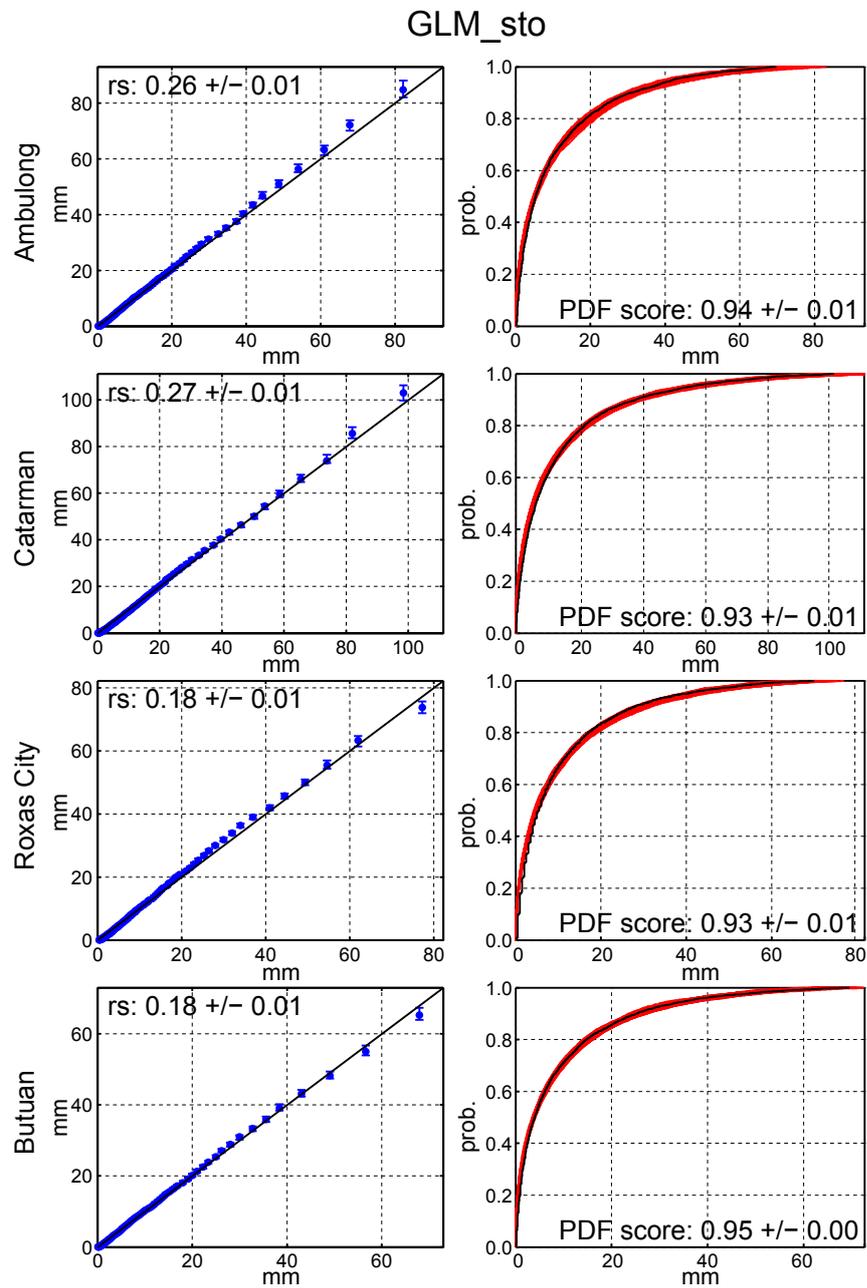


Figure 5.8: As Figure 5.5 but for the *GLM_sto* method. See the text for details.

a small range, which results in a very low distributional similarity. Note that this is not surprising since only the deterministic signal explained by the predictors is considered in this SDM.

Building on this, other configurations introducing a stochastic component⁴ to simulate the unpredicted local-scale variability were implemented and tested. The best results were found by introducing a simulation procedure in both the *GLMo* and the *GLMa*, which could be seen as a dynamic predictor-driven version of the inflation of variance used in some regression-based methods (Huth, 1999). The simulation is as follows: for the *GLMo*, the expected value for each day in the test period is used as probability of occurrence for a Bernoulli distribution from which a random value $\in [0, 1]$ is simulated. If this value is equal or greater (smaller) than the expected one, a 1 (0) is forecast. For the *GLMa*, instead of considering the expected value for each day in the test period (μ_{test_i} ⁵), new daily predictions (μ'_{test_i}) are given by simulating from a gamma distribution whose shape parameter is fitted using the observed wet days in the calibration period (α_{train} ⁶) and whose scale parameter is $\nu_{test_i} = \mu_{test_i}/\alpha_{train}$ ⁷—i.e., for each simulated day in the test period, we keep constant the shape parameter (assuming that $\alpha_{test} \simeq \alpha_{train}$) whilst letting vary the scale one.—

Figure 5.8 displays the results obtained for this stochastic configuration (as in Figure 5.5, 100 different realizations were carried out). Simulation allows the method to predict the full range of observed precipitation amount values, solving thus the discontinuity shown in Figure 5.7, reducing the bias and increasing the predicted variance. As a result, observed and downscaled wet distributions match much better. However, correlations strongly decay as an effect of the stochastic simulation introduced, which implies a reduction of the predictive capacity of both the *GLMo* and the *GLMa*. As in Figure 5.6, the boxplots in Figure 5.9 extend the validation results for the deterministic and stochastic GLM methods considered (top and bottom row, respectively) to the 42 PAGASA stations. For the latter, as for the *AN_sto* method, one single realization, taken at random, was considered—notice from Figure 5.8 that the uncertainty introduced by the stochastic character of *GLM_sto* is very small, so any realization can be considered to be representative of the method.—

Both *GLM_det* and *GLM_sto* yield similar ratios of occurrences, although the latter is worse at reproducing the temporal sequence of rain/no rain events (lower HIR and higher FAR values). The deterministic configuration yields a wet bias (over a 10%) and clearly

⁴For generative techniques, stochastic implementations have the advantage of intrinsically better representing the local variance which is usually underestimated by deterministic ones.

⁵Let's assume a test period of N days, thus $i \in \{1, \dots, N\}$.

⁶A minimum number of wet days is required to make this fitting.

⁷We make use here of the relationship $\mu = \alpha\nu$ —valid for any gamma distribution,— where μ is the expected value and α and ν the shape and scale parameter, respectively.

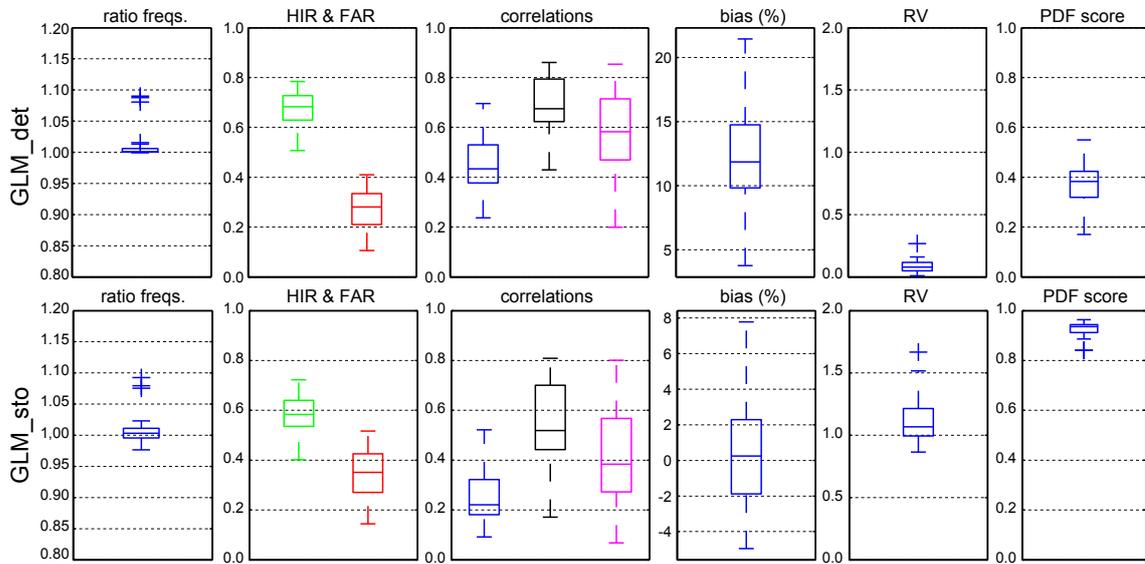


Figure 5.9: As Figure 5.6 but for (top row) the GLM_{det} and (bottom row) the GLM_{sto} methods.

underestimates the variance, which in turn leads to low PDF scores. Differently, bias is centred around zero and the predicted variance matches better the observed one in the stochastic version, greatly improving the distributional similarity (higher PDF scores). Despite this, note that this method overestimates the variance in some cases. We found that this is due to the presence of ‘unrealistic’ (higher than expected) simulated extremes, which are caused by the fixed value of the shape parameter imposed for the simulation process. We have tested the performance of a new type of simulation in which the shape parameter (as the scale one) is let to vary along the test period for those predicted values above the 95 percentile. However, the obtained results were variable, improving the performance in some cases but deteriorating it in others.

Based on their better performance in terms of distributional consistency, stochastic GLMs are needed for SD of climate change projections. However, in seasonal forecasting it is key to assess the accuracy of the predictions (e.g. the interannual correlation) and, hence, it is important to keep the deterministic signal isolated from the stochastic one. Therefore, the GLMs considered in this Thesis are deterministic, i.e., predictions are based on the expected values. Note from the second column of Figure 5.9 that, as compared to the analog methods (Figure 5.6), the correlations exhibited by the stochastic (deterministic) GLM are quite low (clearly higher).

5.4.2 MOS-BC (*Bias Correction*) Methods

As compared to using raw GCM outputs, bias corrected data offers crucial advantages for impact modelling applications (Hempel et al., 2013). For instance, it is well known that GCM precipitation cannot be used to force hydrological models without some form of prior bias correction if realistic output is sought (see, e.g., Feddersen and Andersen, 2005; Hansen et al., 2006; Sharma et al., 2007). Hence, due to their straightforward application and low computational requirements (Li et al., 2010), as well as to the increasing availability of GCM and RCM data, MOS-BC techniques have become very popular in the last years. Nevertheless, and despite these techniques have been shown to present serious drawbacks for climate change projections (Hagemann et al., 2011; Maraun, 2012; Räisänen and Rätty, 2013) —such as the inability to suitably correct the large GCM biases affecting the representation of certain key atmospheric circulation patterns— their problems and limitations for seasonal forecasts have not been explored yet.

Broadly, the existing MOS-BC techniques can be classified into scaling and distributional ones (see, e.g., Maraun et al., 2010; Themeßl et al., 2011). The former, which are the simplest and most used, consist on using an additive or multiplicative scaling factor (Durman et al., 2001; Casanueva et al., 2013) to calibrate the model simulations, i.e, they correct the first moment of the PDF. The latter are the so-called q-q mapping techniques, which adjust higher order moments (different quantiles) of the distribution (Panofsky and Brier, 1968; Piani et al., 2010; Amengual et al., 2012) and have been widely used in impact studies (see, e.g., Quintana-Seguí et al., 2010; Teng et al., 2014). Besides, some extensions to the scaling and distributional techniques such as the multi-variable ISI-MIP (Hempel et al., 2013) method has been recently introduced.

In this Thesis, two widely used distributional (q-q mapping) MOS-BC methods —one parametric and one empirical— have been considered. They are described next. Note that MOS-BC techniques require long series of data from unaltered models (Feddersen and Andersen, 2005). However, this is guaranteed in this Thesis for using the most comprehensive and longest to-date seasonal hindcast from the ENSEMBLES experiment (see Section 3.4.1).

- The parametric q-q mapping (which will be referred to as *QM-par* henceforth) is described in detail in Piani et al. (2010). For each location, normalized daily observed and GCM simulated rainfall intensities (x) are fitted assuming that both are well approximated by a gamma distribution with shape and scale parameters α and ν ,

respectively:

$$PDF(x) = \frac{\exp(-\frac{x}{\nu})x^{(\alpha-1)}}{\Gamma(\alpha)\nu^\alpha} \quad (5.2)$$

This method uses the theoretical instead of the empirical distributions. More sophisticated versions such as the one proposed by Gutjahr and Heinemann (2013) combine a gamma and a Generalized Pareto Distribution to better calibrate the extreme values.

- The empirical q-q mapping (denoted as QM_{emp} hereafter) consists on calibrating the simulated CDF to the observed one by correcting a number of quantiles, according to

$$q_m^* = F_o^{-1}(F_m(q_m)), \quad (5.3)$$

where q_m and q_m^* are the simulated original and corrected quantiles, and F_m and F_o refer to the empirical simulated and observed CDF, respectively. As in Déqué (2007), the 1-99 percentiles are corrected and linear interpolation is used for the values between two percentiles in this Thesis. Constant extrapolation is applied for values outside the calibration range, i.e., the correction function of the last percentile is applied to all values above it.

Both QM_{par} and QM_{emp} incorporate a frequency adaptation which is thought to alleviate the problem that arise when the dry day frequency in the raw model output is larger than in the observations (Thiemeßl et al., 2012; Wilcke et al., 2013), which would lead to a strong positive bias after correction —note that q-q mapping is able to correct automatically the excess of light precipitation frequency (‘drizzle effect’).— In particular, this frequency adaptation consists in randomly sampling the observational distribution into the simulated first bin (0-1 mm) in order to generate dry days.

Figure 5.10 shows in blue (green), for the four stations, cross-validation framework and period considered in the previous examples (Figures 5.4, 5.5, 5.7 and 5.8), the results from applying the QM_{par} (QM_{emp}) method to correct ERA-Interim precipitation (bi-linearly interpolated to the points of interest). Note that, strictly speaking, this scheme would not correspond to the MOS approach since reanalysis precipitation is used here instead of GCM precipitation (see Section 5.1.2); however, it is considered with illustrative purposes and for the sake of comparison with the PP methods presented in Section 5.4.1. Left column displays in red a q-q plot of the 1-99 percentiles —only wet days are considered— of observed against ERA-Interim raw precipitation, whereas blue (green) shows the q-q

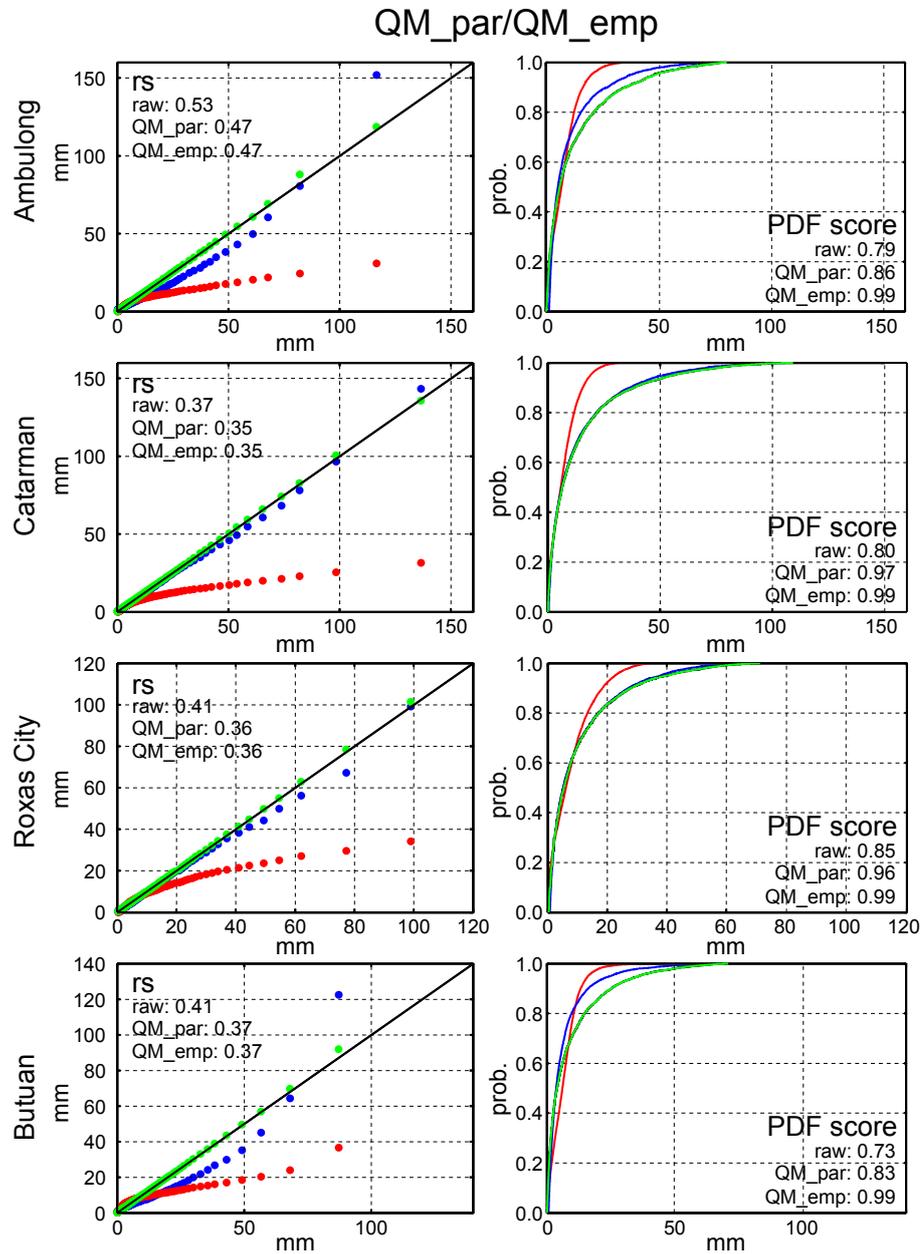


Figure 5.10: As Figure 5.4 but for the MOS-BC methods considered in this Thesis, QM_{par} and QM_{emp} . Red (blue/green) corresponds to raw (parametric/empirical corrected) ERA-Interim precipitation. See the text for details.

plot of observed against ERA-Interim precipitation corrected by means of the QM_{par} (QM_{emp}) method. Right column shows the observed (black), ERA-Interim raw (red) and ERA-Interim corrected (blue/green) precipitation wet distributions.

Although the empirical version performs better, both QM_{par} and QM_{emp} are able to correct to some extent the bad distributional consistency of ERA-Interim (see the improved PDF scores), which exhibits a very scarce variability and is not capable of reproducing high precipitation values. However, neither QM_{par} nor QM_{emp} are able to improve the correlations attained by the reanalysis (which are in general acceptable); on the contrary, they deteriorate them.

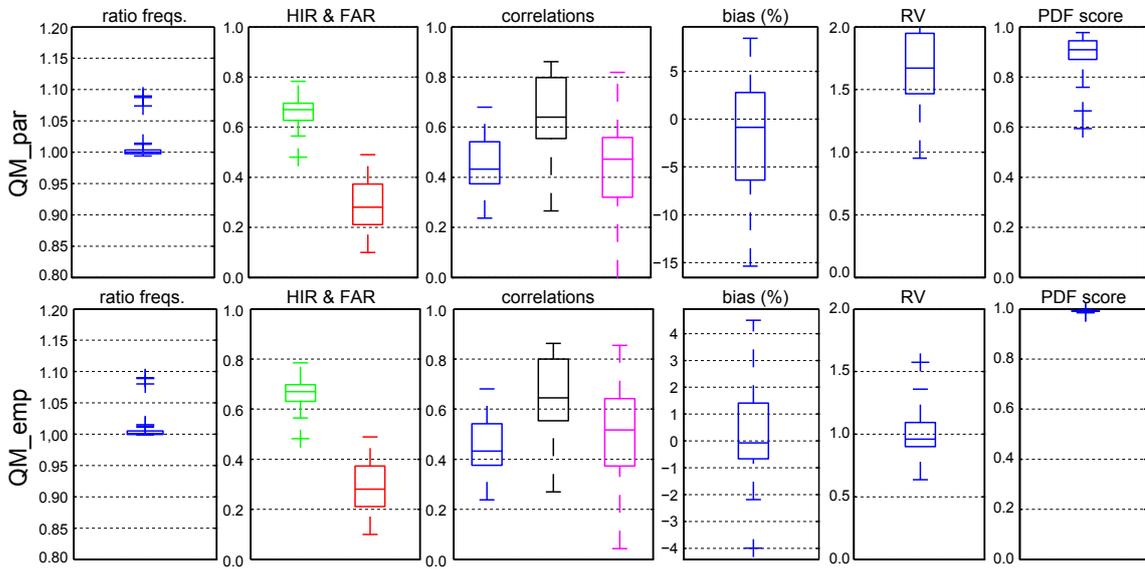


Figure 5.11: As Figure 5.6 but for (top row) the QM_{par} and (bottom row) the QM_{emp} methods.

Top (bottom) row of Figure 5.11 shows the results obtained when applying the QM_{par} (QM_{emp}) method to correct ERA-Interim at the 42 PAGASA stations. Both methods perform similarly in reproducing the occurrence event (first and second column) and also in terms of correlations, yielding better results for all these scores than analogs do (Figure 5.6) and similar ones to those obtained for the deterministic GLM (except for interannual correlation, which is better in the latter; see the top row of Figure 5.9). In terms of distributions, the bias and the predicted variance (especially the latter) are better in the QM_{emp} method, which in turn leads to higher PDFs. On the one hand, the clear overestimation of the variance that occurs for the QM_{par} method could indicate an intrinsic limitation of the method, which, based on Figure 5.10 (only values under the 99 percentile are plotted), might be most likely related to the presence of wrongly simulated extreme

values (recall that this method assumes that both simulated and observed precipitation fit to a gamma distribution, which might not be true). On the other hand, the good performance of *QM.emp* might be related to certain over-fitting.

The performance for SD of seasonal forecasts of all SDMs presented in Sections 5.4.1 and 5.4.2 will be intercompared for the case study of this Thesis, the Philippines, and their relative merits and limitations will be discussed in Chapter 9.

5.5 State-of-the-Art of SD in Seasonal Forecasting

SDMs were first applied in short-range weather forecasts (Klein et al., 1959; Glahn and Lowry, 1972) and later adapted to longer time-horizons, including seasonal forecasts and climate change projections, being the latter problem the one receiving more attention in the literature. For instance, a query on the *Scopus* database (<http://www.scopus.com>) searching for publications including *statistical downscaling* in the “title, abstract and keywords” results in a total of over 700 papers⁸ (see the exponential growth in the last few years in Figure 5.12), 450 of which contained additionally the terms *climate change* or *climatic change*, whereas only 20 included *seasonal forecast* or *seasonal prediction*.

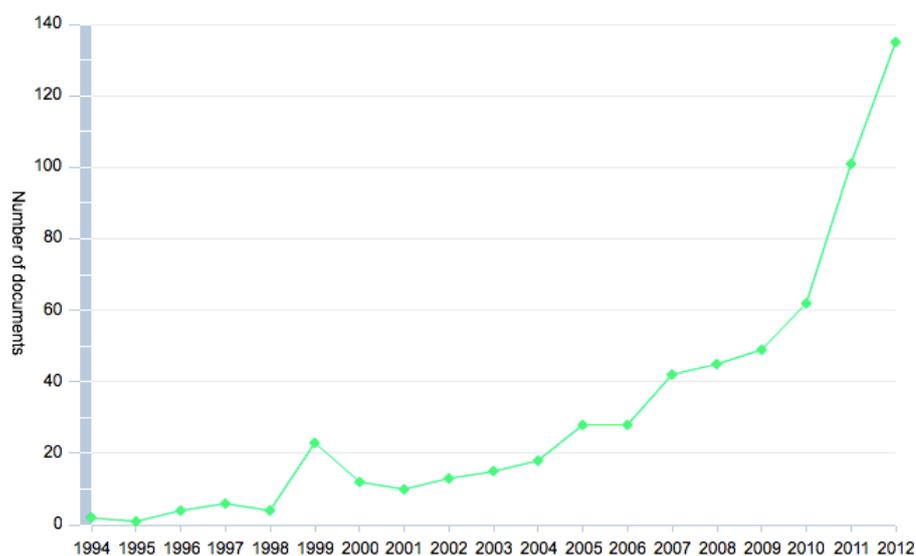


Figure 5.12: Yearly number of papers published which include the term *statistical downscaling* in the “title, abstract or keywords”. Extracted from the deliverable D52.1 (*Review of the different statistical downscaling methods for s2d prediction*) of the SPECS project, which is publicly available at http://www.specs-fp7.eu/sites/default/files/u1/SPECS_D52.1.pdf. Source: *Scopus* publications database.

⁸In journals from the following major subject areas: Earth and Planetary Sciences (48%), Environmental Sciences (27%), Agricultural and Biological Sciences (6%), Engineering (5%) and Mathematics (3%).

Nevertheless, many of the SDMs that have been developed and used for climate change applications have also been applied for seasonal forecasting —due to its low computational cost, SD is especially convenient for seasonal forecasting since it allows to efficiently work with the enormous volume of data of available hindcasts⁹.— In order to provide an inventory of the SDMs successfully applied at this particular time-scale, an extensive electronic bibliographical search was conducted, considering different sources (e.g., *Scopus*, *ISI Web of Knowledge*, *Google Scholar*). Nearly 30 publications —mainly journal papers, but also technical reports and thesis— were identified. Table 5.2 shows that most of the SDMs applied to-date for seasonal precipitation have considered the DEMETER (Development of a European Multimodel Ensemble system for seasonal to inTERannual prediction) models (only three of them have a longer than 40-year hindcast period) and work on a monthly/seasonal basis. Moreover, MOS is the most popular approach (although there is a lack of MOS-BC methods) and ERA-40 (Uppala et al., 2005) the most used reanalysis for PP techniques. Amongst the latter, regression- and analog-based SDMs are the most popular. This evidences the added value and the novelty character of this Thesis, where the ENSEMBLES models (which form the longest to-date and most comprehensive multimodel seasonal hindcast) are considered, all the SDMs (including MOS-BC ones) are applied on a daily basis and the latest reanalysis from the ECMWF, ERA-Interim, is used for the calibration of the different PP techniques. Furthermore, GLM-based methods (rarely used before in the context of seasonal forecasting) are applied, with overall good results.

⁹Typical seasonal hindcast databases contain multi-month (6 to 12 months) long multi-member (10 to 50 members) simulations run for several initializations (often once per month) for a number of years (20 to 40).

Reference	Nationality	Affiliation	Climate model(s)	Approach	Technique	Geographical area(s)	Temporal scale of the statistical model built
Landman and Tennant (2000)	South Africa	South African Weather Bureau	Centre for Ocean-Land-Atmosphere studies (COLA) GCM	MOS	CCA	South Africa	Monthly
Robertson et al. (2004)	USA	Columbia University	ECHAM4.5	MOS	HMM	NE Brazil	Seasonal (FMA)
Díez et al. (2005)	Spain	National Institute of Meteorology	DEMETER (2 models)	PP (ERA-40)	Analog	Spain	Daily
Feddersen and Andersen (2005)	Denmark	Danish Meteorological Institute	DEMETER	MOS	SVDA	Scandinavia, Europe, the contiguous USA and Australia	Seasonal
Gutiérrez et al. (2005)	Spain	University of Cantabria	DEMETER (4 models)	PP (ERA-40)	SOM	Northern Peru	Seasonal
Pavan et al. (2005)	Italy	ARPA-SIM	DEMETER (6 models)	PP (ERA-40)	Regression	Italy	Monthly (DJF)
Chu et al. (2008)	Taiwan	National Taiwan Normal University	SMIP (6 models)	MOS	SVDA	Northern Taiwan	Seasonal (JJA)
Landman et al. (2009)	South Africa	South African Weather Service	ECHAM4.5	MOS	CCA	South Africa	Seasonal (DJF)
Frias et al. (2010)	Spain	University of Cantabria	DEMETER	PP (ERA-40)	Analog	Spain	Daily
Robertson et al. (2012)	USA	International Research Institute for Climate and Society	RegCM3/ECHAM4.5	MOS	CCA	Philippines	Seasonal (AMJ)
Sun and Chen (2012)	China	Institute of Atmospheric Physics	DEMETER (7 models)	MOS	Regression	Global	Seasonal
Wu et al. (2012)	USA	NCAR	CFS	MOS	Analog	SE ranean	Monthly
Ying and Ke (2012)	China	Institute of Atmospheric Physics	DEMETER (3 models)	PP (ERA-40)	Regression	SE China	Seasonal
de Castro et al. (2013)	Brazil	Federal University of Ceara	RSM/ECHAM4.5	MOS	ANNs	Brazil	Monthly
Charles et al. (2013)	Australia	Bureau of Meteorology	POAMA	PP (NCEP-NCAR)	Analog	SE Australia	Daily
Shao and Li (2013)	Australia	CSIRO	POAMA	PP (NCEP-NCAR)	Analog	SE Australia	Daily
Silva and Mendes (2013)	Brazil	University Federal of Rio Grande do Norte	CFS	MOS	ANNs	NE Brazil	Monthly
Sinha et al. (2013)	India	Centre for Atmospheric Sciences, Indian Institute of Technology	In-GIMI (NCMRWF)	PP (NCEP-NCAR)	CCA	India	Seasonal
Sohn et al. (2013)	Korea	APCC	APCC MME (10 models)	MOS	Regression	South Korea	Monthly
Tung et al. (2013)	China	City University of Hong Kong	APCC MME	MOS	SVDA	South China	Seasonal

Table 5.2: Summary of previous studies applying any form of SD to forecast seasonal precipitation. The nationality and affiliation fields refer to the corresponding author. The name of a RCM followed by a “/” indicates the RCM/GCM coupling, e.g., “RegCM3/ECHAM4.5”. In PP techniques, the reanalysis used for calibration is given in brackets, e.g. “PP (ERA-40)”. Adapted from the deliverable D52.1 of the SPECS project.

CHAPTER 6

Methodological Aspects for SD in Seasonal Forecasting

There is a number of methodological aspects which are relevant for the SD of seasonal forecasts and must therefore be carefully analyzed for an appropriate application of the different SDMs introduced in Chapter 5.

First, as mentioned in Section 5.1.1, for each particular predictand variable and region of interest, the most time-consuming task in Perfect Prog (PP) schemes is the selection (*screening*) of a suitable combination of predictors —defined over a proper geographical domain,— which should account for a major part of the variability in the predictand and be realistically reproduced by both reanalyses and GCMs (see, e.g., Wilby et al., 2004). Typically, this is undertaken by assessing the performance of different predictor–domain combinations in a cross-validation framework, using to this aim standard validation scores (e.g., accuracy metrics).

Second, although the calibration phase of the different SDMs is common to seasonal forecasting and climate change studies, the former case presents a number of particularities which typically do not apply for the latter. For instance, in seasonal forecasting, the SDMs can be calibrated using season-specific data, e.g., only JJA training data is considered for JJA predictions. However, typical applications in a climate change context consider the entire available yearly predictor dataset for calibration (see, e.g., Gutiérrez et al., 2013). Both approaches have advantages and limitations and should be tested in order to find the optimum procedure.

Third, in the PP approach, the different SDMs are sensitive to the GCM biases (as compared to reanalysis) in the large-scale predictor variables. Therefore, GCM predictors must be conveniently treated (or *harmonized*) before entering the SDM in order to make them compatible with the reanalysis used for calibration (Maraun et al., 2010); otherwise,

misleading results might be obtained. A number of possibilities exist for this task, which should also be explored.

This chapter addresses all these issues for the case study of this Thesis, seasonal forecasting of precipitation in the Philippines, establishing the framework under which the different SDMs considered (see Chapter 5) will be later applied.

6.1 *Predictor Screening*

Atmospheric variables describing circulation, temperature and moisture are generally considered to be among the most informative predictors for SD of precipitation under the PP approach (see, e.g., Charles et al., 1999; Timbal et al., 2003; Bürger and Chen, 2005; Haylock et al., 2006; Fowler et al., 2007; Hertig and Jacobeit, 2008; Sauter and Venema, 2011; San-Martín et al., 2016). However, whereas it is known that suitable predictors for PP should be realistically reproduced by both reanalyses and GCM (Hewitson and Crane, 1996; Wilby et al., 2004), it is also known that GCMs generally perform better for circulation and temperature variables than for moisture ones (Räisänen, 2007; Brands et al., 2011, 2013). Yet, moisture information should be included into the predictor field in order to improve the statistical link-function, i.e., the predictive potential of the SDM. Moreover, for climate change applications, measures of humidity are necessary to capture changes in the water-holding capacity of the atmosphere under global warming, i.e., to capture the ‘correct’ climate change signal (Goodess and Palutikof, 1998; Wilby et al., 1998). Note that, although climate change is not the main concern of this Thesis, we deal with it in Section 8.5.

With these precepts in mind, and after consulting the expertise from local meteorologists as well as the results from previously published studies (Kang et al., 2007; Chu et al., 2008; Paul et al., 2008; Chu and Yu, 2010), a set of different predictor combinations were considered for the screening process for our case (Table 6.1). These combinations consist of circulation variables alone (in particular the zonal wind component at different vertical levels; P1: U850, U300), circulation and specific humidity (P2: U850, U300, Q850), circulation and temperature (P3: U850, U300, T850) and circulation, specific humidity and temperature (P4: U850, U300, Q850, T850). Note that Q850 is used instead of column integrated water vapour or precipitable water since the latter variables are usually not provided by the common GCM databases. All the variables listed in Table 6.1 were obtained from the ERA-Interim reanalysis (Dee et al., 2011) and are also available for the GCMs which will be used later in this Thesis (e.g. the ENSEMBLES models). To keep consistency with the latter, daily instantaneous values at 00 UTC were chosen in all cases.

As opposite to the case of MOS-BC, PP methods also require the determination of the geographical domain where the predictors are considered. Typically, this domain should

Code	Predictor Variables
$P1$	U850, U300
$P2$	U850, U300, Q850
$P3$	U850, U300, T850
$P4$	U850, U300, Q850, T850

Table 6.1: Predictor combinations considered. U/Q/T corresponds to zonal wind component/specific humidity/temperature. The numbers refer to the vertical level considered, in hPa. For instance, U850 stands for the zonal wind component at 850 hPa.

be large enough to resolve the relevant large-scale circulation patterns affecting the local meteorology (Feddersen and Andersen, 2005). Moreover, the spatial scale considered for the predictors usually varies according to the particular time-scale considered. Often, large-scale quasi hemispheric-like patterns are considered for PP methods operating on a monthly/seasonal basis, whereas smaller domains (a few thousands of kilometres) are considered for those operating on a daily basis (see, e.g., Gutiérrez et al., 2004, for further details). In the former case, ocean-derived variables (mainly SST) and climate oscillation indices (e.g., ENSO, NAO, etc.) are often additionally considered as predictors. Following from these considerations, the geographical domain considered for SD in this Thesis is shown with blue crosses in Figure 6.1 (the 42 PAGASA stations are also displayed in red). In particular, a regular 2° grid is considered, to which ERA-Interim was bi-linearly interpolated from its native resolution. This domain, which contains 238 grid boxes, is very similar to the one used by Robertson et al. (2012) for SD in the Philippines and is large enough to encompass the effect of the ENSO teleconnection affecting the climate of the region (see Section 7.5.3).

A deterministic GLM-based method¹ considering the 15 leading Principal Components (PCs) was applied to all the predictor combinations listed in Table 6.1. As explained in Section 5.2 a k -fold cross-validation framework with $k = 5$ non-overlapping test periods of five years each covering the entire period 1981-2005 was considered. To circumvent spurious trend effects, the five years forming each test period were randomly chosen.

Figure 6.2 shows the results obtained. In particular, it displays the Spearman correlation coefficient (rs) between daily observed and downscaled precipitation time-series for different predictor combinations: in addition to P1-P4, Q850 and T850 alone are also provided for comparison purposes. In each panel, the results for a specific CT (see Section 4.1) are shown. Along the x -axis, stations are sorted by decreasing latitude (from left to right). On the right hand side, the CT-averaged rs are indicated with points.

¹Deterministic GLMs were shown to provide the best interannual ACC among the different SDMs considered in this Thesis (see Section 5.4).

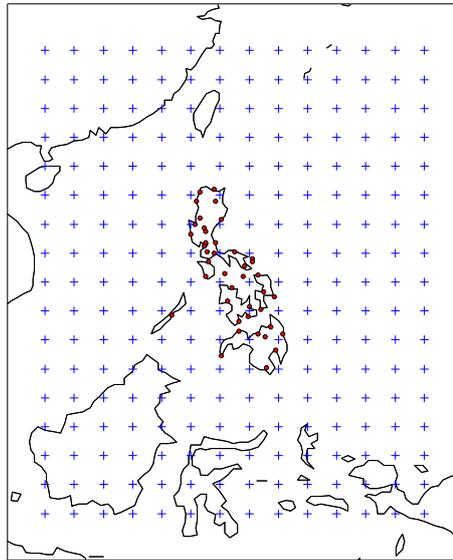


Figure 6.1: Geographical domain considered for SD (blue crosses) and the 42 PAGASA stations (red points).

In general, the combination of circulation and thermodynamic variables (P4) yields the highest correlation coefficients in all CTs, whilst the predictive potential is lower for the case of using circulation variables alone (P1). Moreover, P2 (P3) provides very similar results to P4 (P1), which points out the scarce importance of T850 to explain the local-scale climate variability over the area of study, once circulation is taken into account. Contrarily, adding Q850 to circulation variables improves the statistical link-function (compare P1 and P2). The little contribution of T850 (as compared to Q850) is also manifest from comparing the red and blue lines —the former variable alone brings in the worst results in all CTs.— This in agreement with previous studies which have shown that the linkage of precipitation with moisture is generally stronger than with temperature (see, e.g., Beckmann and Buishand, 2002).

In the light of these results, the P4 predictor combination is considered for the PP methods later applied in this Thesis. Despite providing similar results whilst being simpler, P2 is discarded in favour of P4 since including some temperature in the predictor field is expected to add some predictive capacity for particular seasons. For instance, in companion to humidity, temperature describes as well the rush of warm humid air from maritime equatorial air, which characterizes the southwest summer monsoon. Besides, the strength of convection is also controlled by temperature (Brandsma and Buishand, 1997). For spatial detail, Figure 6.3 shows the performance found for P4, when applying the aforementioned deterministic GLM considering the 15 leading PCs (right column). For

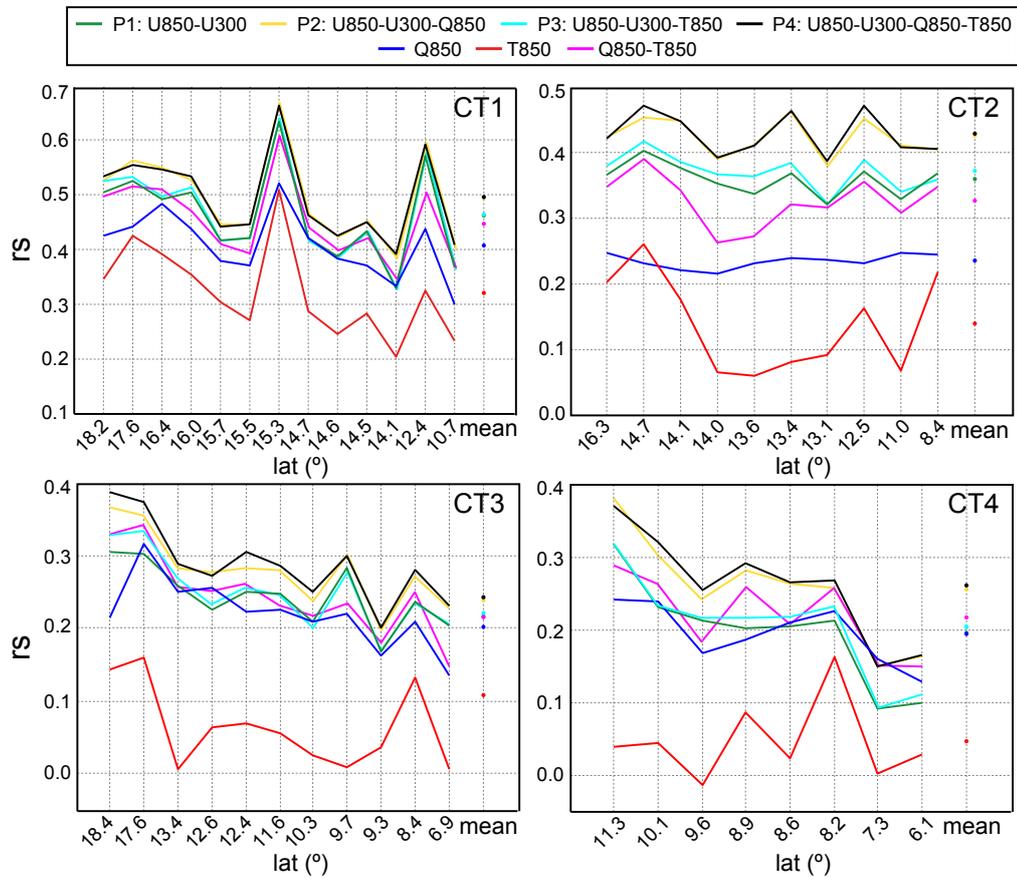


Figure 6.2: Cross-validated results —as measured by the Spearman correlation coefficient (rs) between observed and downscaled daily precipitation time-series for the period 1981–2005— for each CT. Different colors correspond to different predictor combinations (see the legend) from ERA-Interim. For each CT, results are sorted by the latitude of the stations (decreasing from left to right). CT-specific spatial average values are shown with points on the right hand side of each panel.

completeness, results are also shown for a deterministic GLM considering as predictors the standardized anomalies at the 4 nearest grid boxes (left column). Results are very similar in both cases, obtaining the highest rs in the northern part of the archipelago and along the eastern coastline, with a gradual decrease towards the south.

6.2 Season-Specific versus Yearly Data for Calibration

Since seasons may change in the future —e.g., more summer-like days are expected in Europe (Ruostenoja and Räisänen, 2013),— calibrating and applying the SDMs separately for each season (e.g. only JJA training data is considered for JJA predictions) could have uncontrollable effects in the downscaled results from GCM scenario runs (Im-

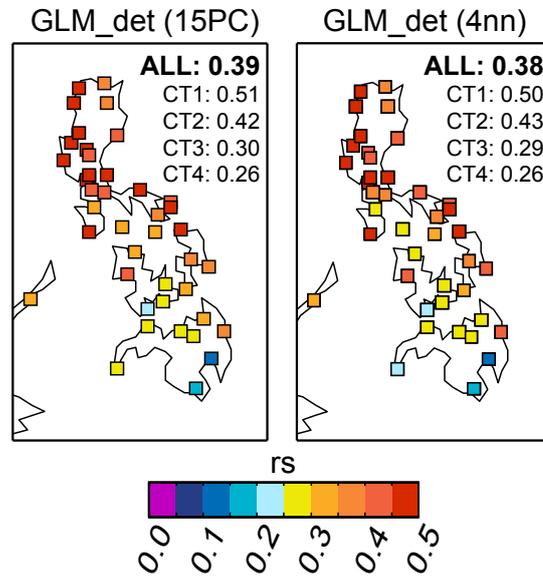


Figure 6.3: Cross-validated results —as measured by the Spearman correlation coefficient (r_s) between observed and downscaled daily precipitation time-series for the period 1981-2005— for the P4 predictor combination (see Table 6.1) from ERA-Interim. The numbers indicate the spatial mean value for all the stations (‘all’) and for those stations pertaining to a specific CT (‘CT1-4’).

bert and Benestad, 2005). Thus, the entire available *yearly* predictor dataset is typically used in SD of climate change projections (see, e.g., Gutiérrez et al., 2013), being the use of *season-specific* data not recommended. However, the latter approach is suitable for SD of seasonal forecasts. Therefore, in this case, it is important to assess the performance of these two alternative procedures, validating to this aim those aspects relevant for seasonal forecasting. In particular, interannual ACC is the key validation metric to take into account since it summarizes the ability of the SDM to preserve the temporal structure of the observed interannual anomalies. Figure 6.4 shows the interannual ACC obtained when considering season-specific and yearly data (solid and dashed lines, respectively) for calibrating a deterministic GLM using as predictors (left) the 15 leading PCs and (right) standardized anomalies at the 4 nearest grid boxes (the same cross-validation framework used in the previous section for the period 1981-2005 was considered here). Following from the results of that section, the P4 predictor combination was considered for this analysis. For all 3-month seasons along the year, each line (see colors in legend) shows the average interannual ACC for the corresponding CT. In general, season-specific data yields better correlations for all CTs, and especially in the central part of the year (JJA), for which most of the rainfall occur over the area of interest.

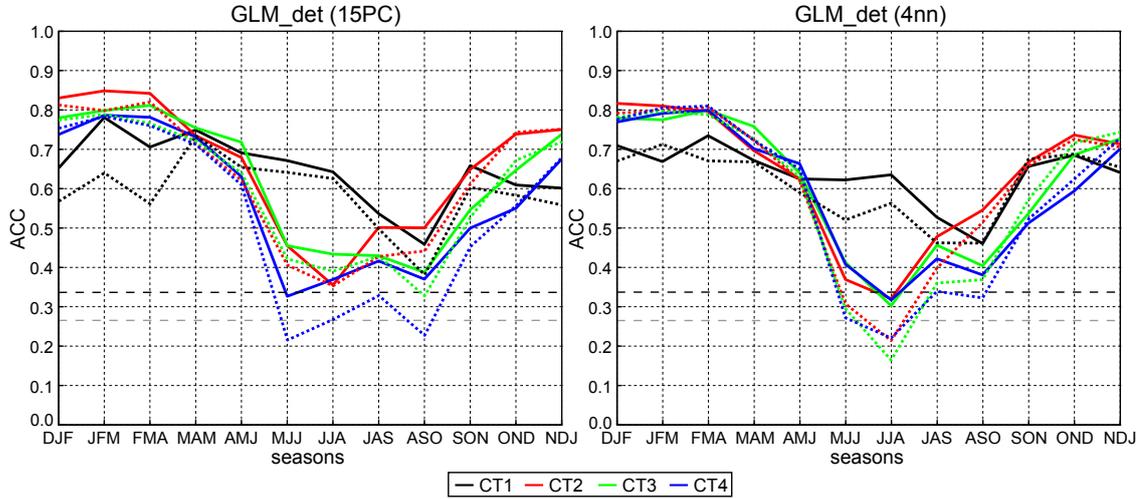


Figure 6.4: Cross-validated results —as measured by the ACC between observed and downscaled interannual accumulated precipitation time-series for the period 1981-2005—when considering season-specific and yearly ERA-Interim predictor data (solid and dashed lines, respectively) for calibrating a deterministic GLM using (left) the 15 leading PCs and (right) the standardized anomalies at the 4 nearest grid boxes. For all 3-month seasons along the year (x -axis), each line shows the average interannual ACC for the corresponding CT (see colors in legend). Values above the gray (black) horizontal dashed line are statistically significant — $\alpha = 0.1$ ($\alpha = 0.05$).—

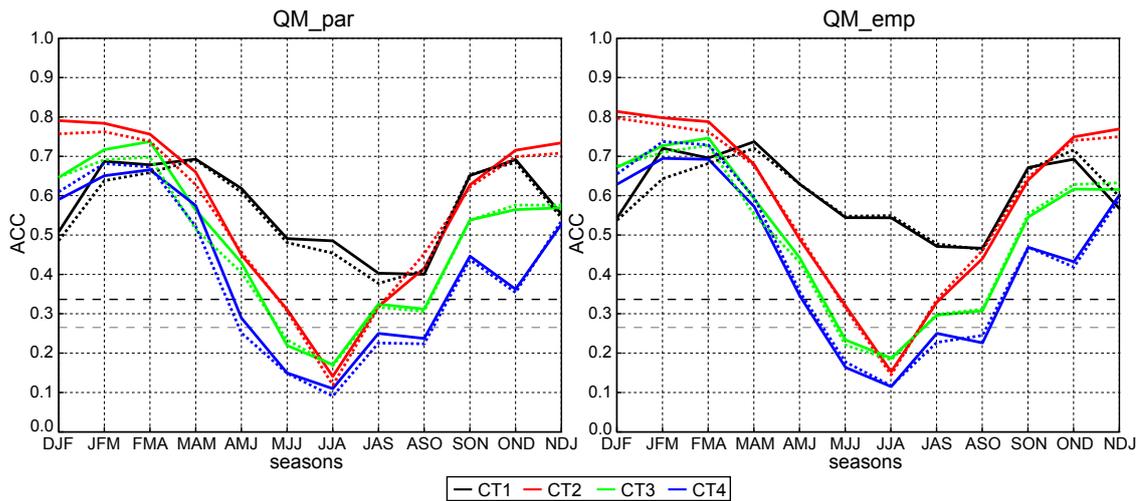


Figure 6.5: As Figure 6.4 but for the two MOS-BC methods used in this Thesis, considering as only predictor ERA-Interim precipitation.

For completeness, Figure 6.5 is the equivalent to Figure 6.4 but for the two MOS-BC methods used in this Thesis, considering as only predictor ERA-Interim precipitation at the nearest grid box. Differently to the case of PP, MOS-BC methods seem to be not sensitive to the choice of data (season-specific or yearly) considered for calibration, providing very similar results in both cases.

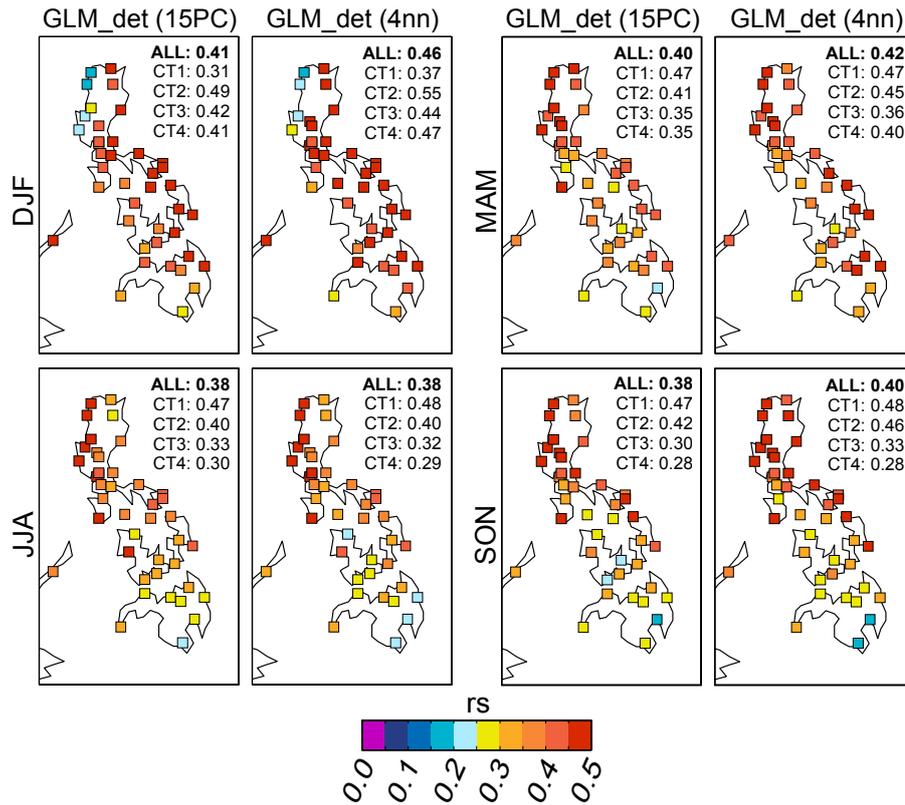


Figure 6.6: Cross-validated results —as measured by the Spearman correlation coefficient (r_s) between observed and downscaled daily precipitation time-series for the period 1981-2005— obtained for each individual season when using season-specific data for calibrating a deterministic GLM considering (left column) the 15 leading PCs and (right column) the standardized anomalies at the 4 nearest grid boxes. The P4 predictor combination (see Table 6.1) from ERA-Interim is considered. The numbers indicate the spatial mean value for all the stations (‘all’) and for those stations pertaining to a specific CT (‘CT1-4’).

Hence, based on the results from Figures 6.4 and 6.5, season-specific data is considered for the calibration of all the SDMs (both PP and MOS-BC) later applied in this Thesis. For an example complementing Figure 6.3, Figure 6.6 shows the r_s between observed and downscaled daily precipitation time-series obtained for each individual season when using season-specific data for calibrating a deterministic GLM considering (left column) the 15

leading PCs and (right column) the standardized anomalies at the 4 nearest grid boxes. As above, the P4 predictor combination was considered and a 5-fold cross-validation for the period 1981-2005 was applied. It is seen that both configurations of the SDM yield very similar results, with rs values which are in general similar to those from Figure 6.3. However, as compared to the latter figure, changes in the spatial patterns appear for different seasons. For instance, whilst the north-south gradient of predictive potential is roughly kept for the rest of the year, it shifts in DJF, when the worst results are found for the stations in the northwestern part of the archipelago.

6.3 Harmonization of GCM Predictor Data

As shown in Section 3.4.1-2, GCMs exhibit large biases for precipitation. For a more detailed analysis over our area of interest —the domain considered for SD (Figure 6.1),— Figure 6.7 displays, for four of the five ENSEMBLES models, the bias —ERA-Interim was taken as reference— for one-month lead precipitation for DJF, MAM, JJA and SON, as well as for each of the individual months forming each season (in rows). Thus, within each particular season, predictions for each target month would correspond to different lead-times. For instance, for DJF, predictions for December/January/February would correspond to 1/2/3 months lead-time forecasts.

Beyond the expected differences between models, varying spatial patterns are found in some cases among the different months forming the same season. To properly take this model feature into account, different bias correction strategies based on moving-windows have been applied in the literature. For instance, for the i^{th} day of the year ($i = \{1, \dots, 365\}$), the corresponding bias might be corrected based upon the data subset encompassed by a moving window of a certain width centred around that day. Although this approach could be equally used for both climate change projections and seasonal forecasts, it has been only applied in the former case to-date (Table 6.2). Note that, given the short length of each particular season, very narrow windows would be needed in seasonal forecasting —for instance, for a one-month moving window, the first and the last 15 days of the season would not be used,— which would make the process prone to over-fitting and computationally inefficient. For this reason, the MOS-BC methods considered in this Thesis (Section 5.4.2) are directly applied on raw GCM precipitation and are expected to correct it, at a distributional level, based on local observations.

Differently to the MOS-BC methods, as explained in Chapter 5, PP methods need to be calibrated with reanalysis data before being applied to a GCM (see Section 5.1.1), so they are sensitive to biases in the GCM predictors (as compared to reanalysis). Noteworthy, the large biases shown for precipitation in Figure 6.7 suggest that the associated large-

Reference	Window width
Räsänen and Rätty (2013)	1, 2, 3 months
Thiemeßl et al. (2011); Wilcke et al. (2013); Maurer and Pierce (2014)	31 days
Thiemeßl et al. (2011)	61 days
Boé et al. (2007); Maraun (2013)	seasonal (\simeq 90 days)

Table 6.2: Typical window sizes used in previous works (all of them have been undertaken in climate change conditions). This list is not comprehensive, but provides just a few illustrative examples.

scale variables might also suffer from important model errors. Figures 6.8 to 6.11 are the equivalent ones to Figure 6.7 but for the predictor variables listed in Table 6.1. Note that U300 is replaced here by U200, the most similar variable available for the ENSEMBLES models.

As in Figure 6.7, varying spatial patterns are found for different months within the same season, particularly important for some variables and models. Therefore, before entering any of the PP SDMs considered in this Thesis (Section 5.4.1), every large-scale GCM predictor is properly harmonized (i.e., made compatible with the counterpart reanalysis variable used for calibration). To this, we first remove the model bias (the reanalysis used for calibration is taken as reference) separately for each month rather than for the entire season. By doing this, we force the GCM to follow the mean and variability of the reanalysis to some extent. Note that, although more sophisticated strategies as the aforementioned moving windows applied for precipitation in climate change studies could be also considered to this aim, this first-class approach adapts better to the case of seasonal forecasts and is more efficient from a computational point of view. Afterwards, GCM predictor data is standardized (grid box by grid box) by subtracting the mean and dividing by the standard deviation of the reanalysis used for calibration. On the one hand, these standardized fields are used as input data for those SDMs considering anomalies at nearby grid boxes. Note that standardization brings the first and second order moments of the reanalysis and GCM into agreement and thereby provides a better approximation for the assumption of ‘perfect’ GCM performance than using untransformed data. Moreover, it prevents from the negative effects related to the distinct ranges of the different predictor variables considered. On the other hand, for the SDMs considering PCs as predictors, the PCs of the GCM are obtained by projecting the standardized GCM fields onto the Empirical Orthogonal Functions (EOFs) of the reanalysis used for calibration.

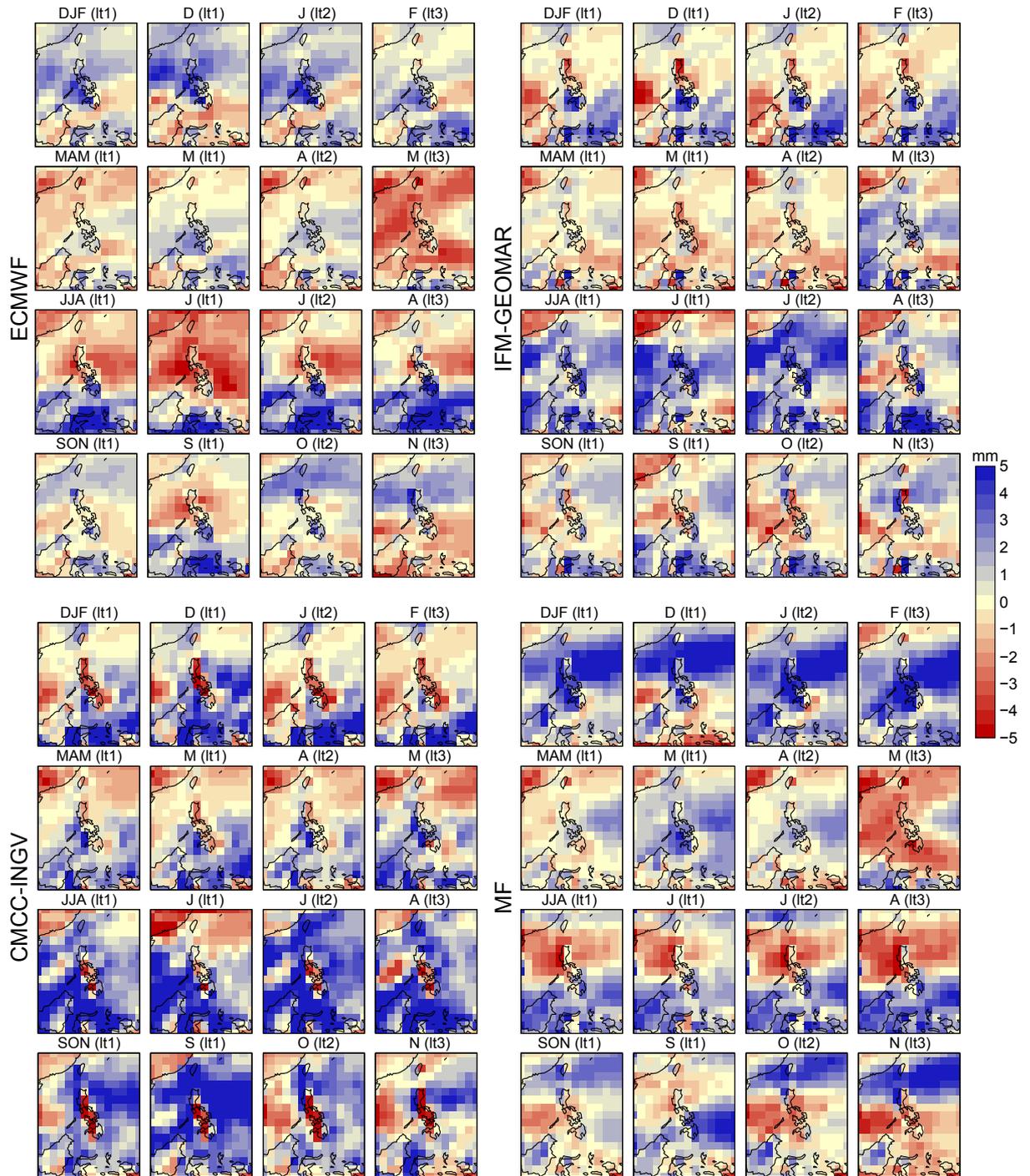


Figure 6.7: For each of the ENSEMBLES models, panels show the bias —with respect to ERA-Interim— for precipitation for DJF, MAM, JJA and SON (one-month lead predictions are considered) and for each of the individual months forming the corresponding season (in rows).

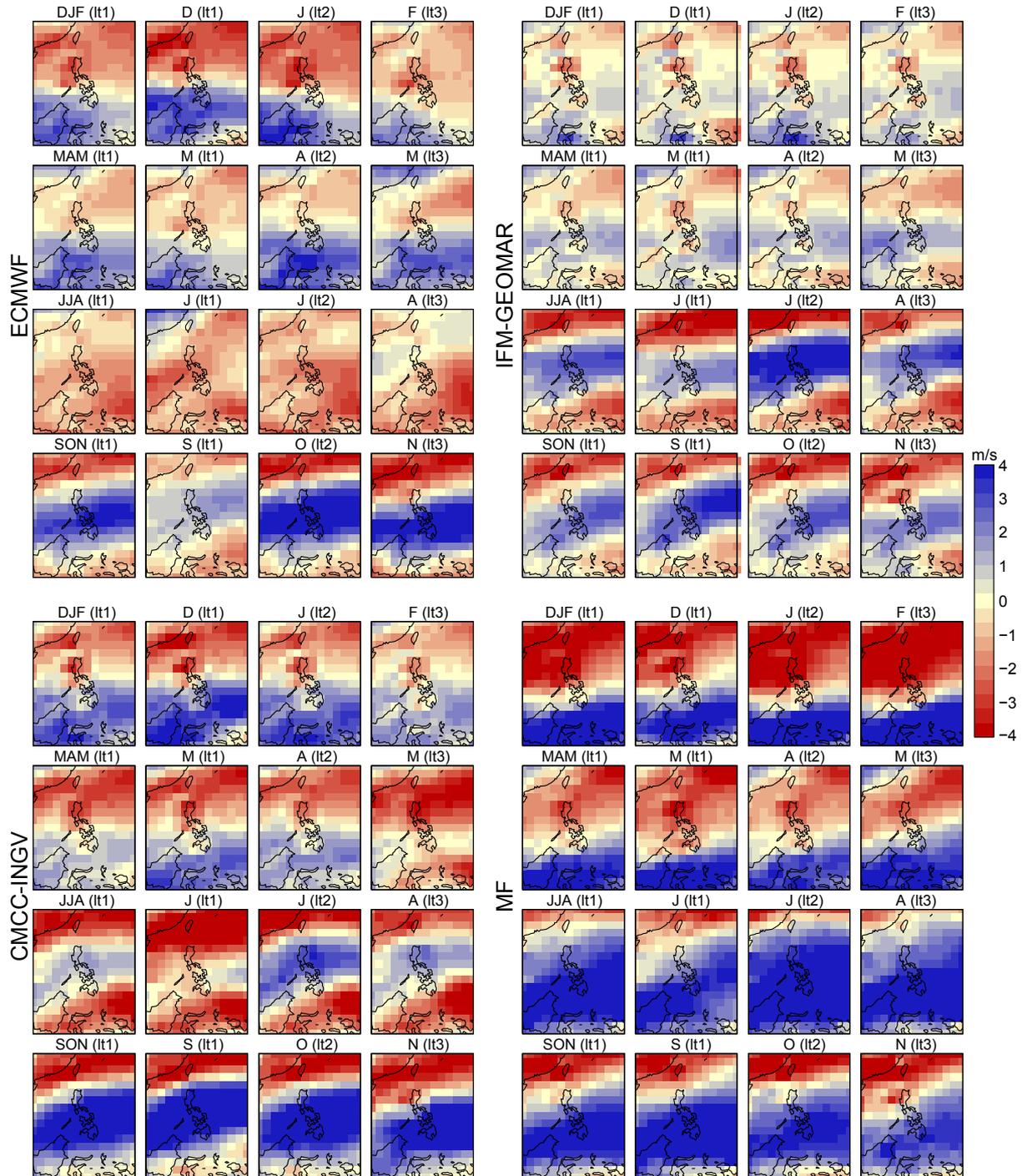


Figure 6.8: As Figure 6.7 but for U50.

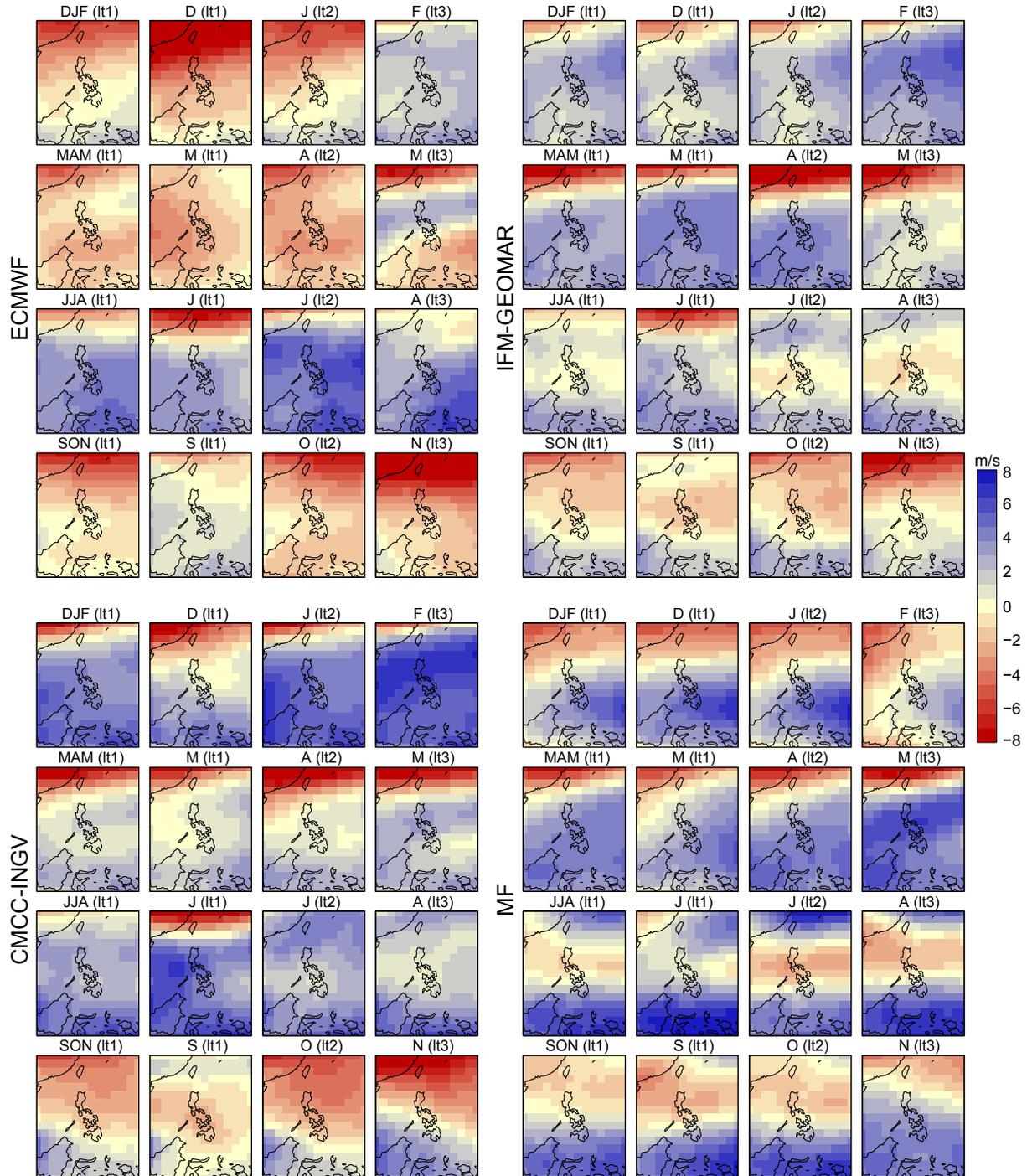


Figure 6.9: As Figure 6.7 but for U200.

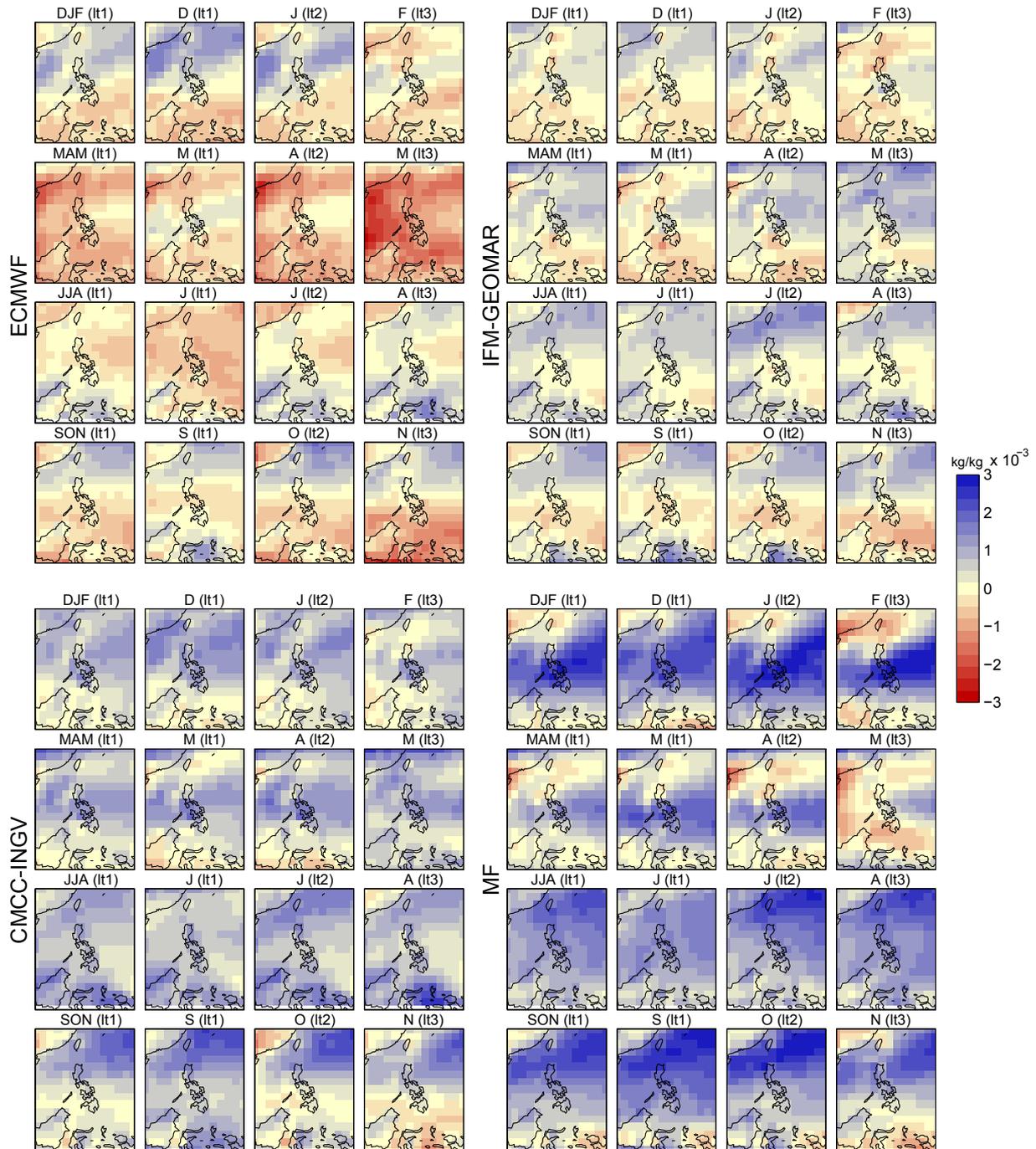


Figure 6.10: As Figure 6.7 but for Q850.

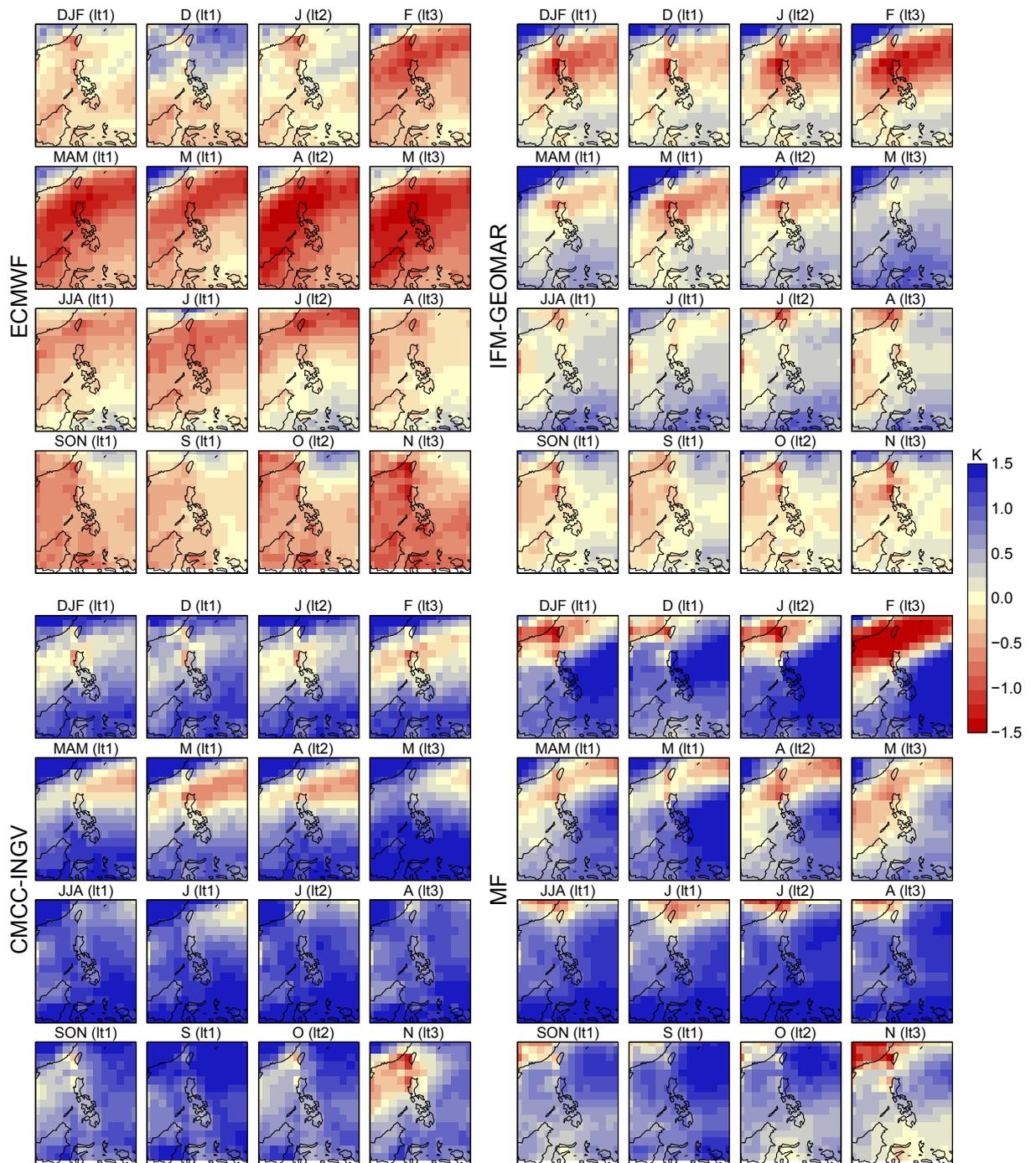


Figure 6.11: As Figure 6.7 but for T850.

Part III

Main Results

CHAPTER 7

Validation of Global Seasonal Precipitation Forecasts

Nowadays, one of the limiting factors that still hinder the practical use of seasonal forecasts (see, e.g., Goddard et al., 2010) is that predictability at this particular time-scale strongly varies with the target variable, region and season (see, e.g., Halpert and Ropelewski, 1992; van Oldenborgh, 2004; Barnston et al., 2010; Doblas-Reyes et al., 2010). Hence, in order to properly communicate the uncertainties related to global seasonal precipitation predictions, it is needed to develop a comprehensive assessment of the performance of the different forecasting models worldwide. Nevertheless, the majority of verification studies for this variable have been conducted over limited areas of the world and for concrete seasons (see, e.g., Batté and Déqué, 2011; Lim et al., 2011; Kim et al., 2012a; Landman and Beraki, 2012). A few studies have also been conducted worldwide (van Oldenborgh et al., 2005; Wang et al., 2009; Barnston et al., 2010; Doblas-Reyes et al., 2010) using different validation scores; however, the limited hindcast period available in the latter works does not ensure a robust statistical validation. For instance, Doblas-Reyes et al. (2010) analyzed the ENSEMBLES multimodel seasonal dataset computing averaged scores over six large-scale regions of the world for the period 1991-2005.

This chapter aims to fill the lack of a user-oriented validation of global seasonal precipitation forecasts for a long enough period (1961-2000) which allows for robustly identifying those regions of the world with significant seasonal skill. To this, we consider the ENSEMBLES multimodel seasonal hindcast (see Section 3.4.1) and apply a simple tercile-based probabilistic validation scheme, obtaining worldwide maps of ROC Skill Score (ROCSS), a score which is recommended by the SVS-LRF for the verification of probabilistic seasonal forecasts (see Section 3.5). Additionally, since ENSO is known to be the major driving factor for seasonal predictability (see Chapter 3), we also analyze its role on the seasonal skill.

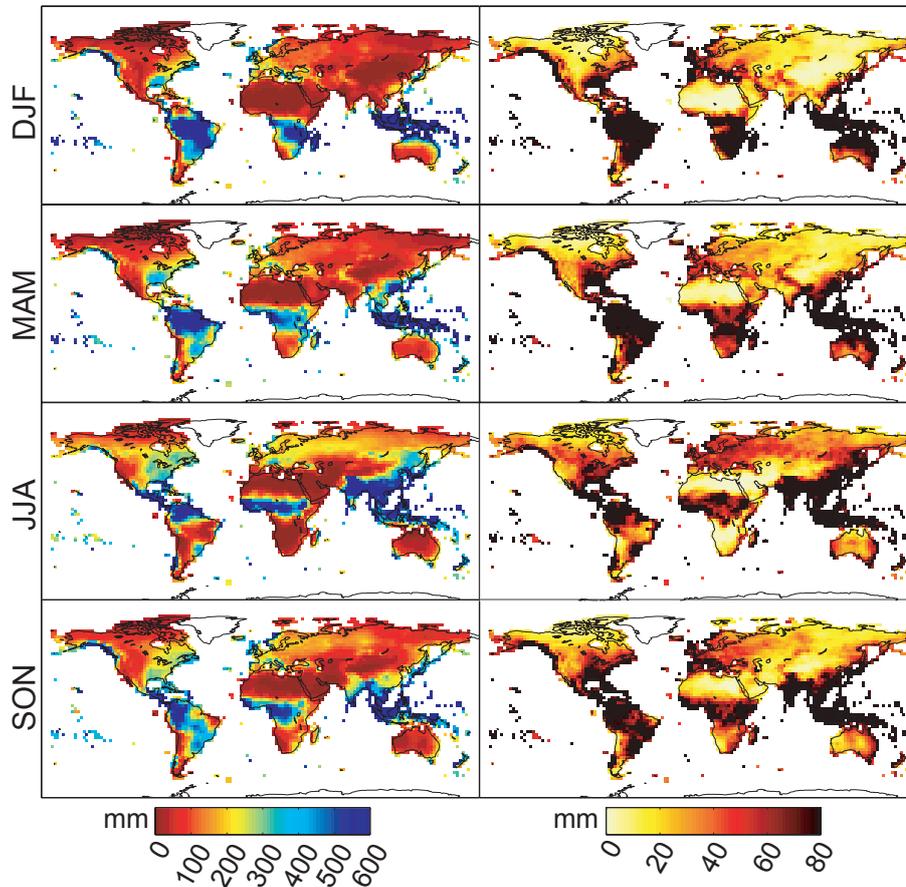


Figure 7.1: Mean climatology and interannual standard deviation (left and right column, respectively) of seasonal accumulated precipitation from VASClmO v1.1 for the four standard seasons (in rows) within the period 1961-2000.

7.1 Data Used

VASClmO v1.1 (Beck et al., 2005) was considered as reference observations for validation. This gauge-based product provides monthly precipitation totals on a 2.5° resolution grid for the global land areas (except the Antarctica) for the period 1951-2000. Figure 7.1 shows the mean seasonal totals and the corresponding interannual standard deviation for this dataset for the period of study: 1961-2000.

In order to test the sensitivity to the reference data in the validation process, all calculations were also done for an alternative precipitation dataset, the Global Precipitation Climatology Centre full data reanalysis version 6 (GPCC v6) (Becker et al., 2013). The results obtained in both cases were very similar. Thus, only VASClmO v1.1 is considered hereafter.

Predictions were obtained from the most comprehensive and longest-to-date multi-model seasonal hindcast, provided by the EU project ENSEMBLES (see Section 3.4.1). For each of the five contributing models, seven-month long runs were issued four times a year within the period 1960-2005, starting the first of February, May, August and November (see Weisheimer et al., 2009, for more details about the experiment). Therefore, the seasons considered for validation were the standard boreal winter (DJF), spring (MAM), summer (JJA) and autumn (SON), since this allows to analyze one- and four-month lead predictions —e.g., the initializations of August and May can be used to forecast SON.— Note that although alternative three-month seasons could be more informative in particular regions of the world, there would be a single lead-time available for them, thus limiting the study. The validation period considered was 1961-2000, common to VASCLimO v1.1 observations and the ENSEMBLES models. All the models were bi-linearly interpolated to the grid of the observations —similar results were obtained using the nearest grid box interpolation technique (not shown).—

7.2 Validation Methodology

The validation methodology used is a tercile-based probabilistic approach previously applied in other studies (see, e.g., Frías et al., 2010; Vellinga et al., 2013). For each particular grid box and each particular model, member and season, the forty-year interannual series of predicted seasonal precipitation were classified into three categories (dry, normal and wet), according to their respective climatological terciles within the period 1961-2000. Then, a probabilistic forecast was computed year by year by considering the number of members falling within each category, out of a total of $n = 9$ members. The terciles were defined independently for each model, considering the interannual series of its nine members (a total of $40 \times 9 = 360$ values) —terciles were not computed at a member-level since no significant overlap among the dry and wet terciles of the nine members was found applying a Student's t-test.— In the case of the multimodel (MM), $n = 45$ members were used to compute the probabilistic forecasts, thus assuming equal weights for all the models. The terciles for the MM were computed independently for each model. Note that working with precipitation categories instead of with raw values implicitly entails a bias correction grid box by grid box, so the different region- and season-dependent biases shown by the ENSEMBLES models in Section 3.4.1-2 do not affect the results obtained here.

Rather than using deterministic scores (e.g., van Oldenborgh et al., 2005; Batté and Déqué, 2011; Lim et al., 2011; Li et al., 2012; Singh et al., 2012), the forecast performance is assessed in terms of the probabilistic ROCSS. The statistical significance of this score was obtained by bootstrapping (Mason and Graham, 2002) with 1000 samples, i.e., by

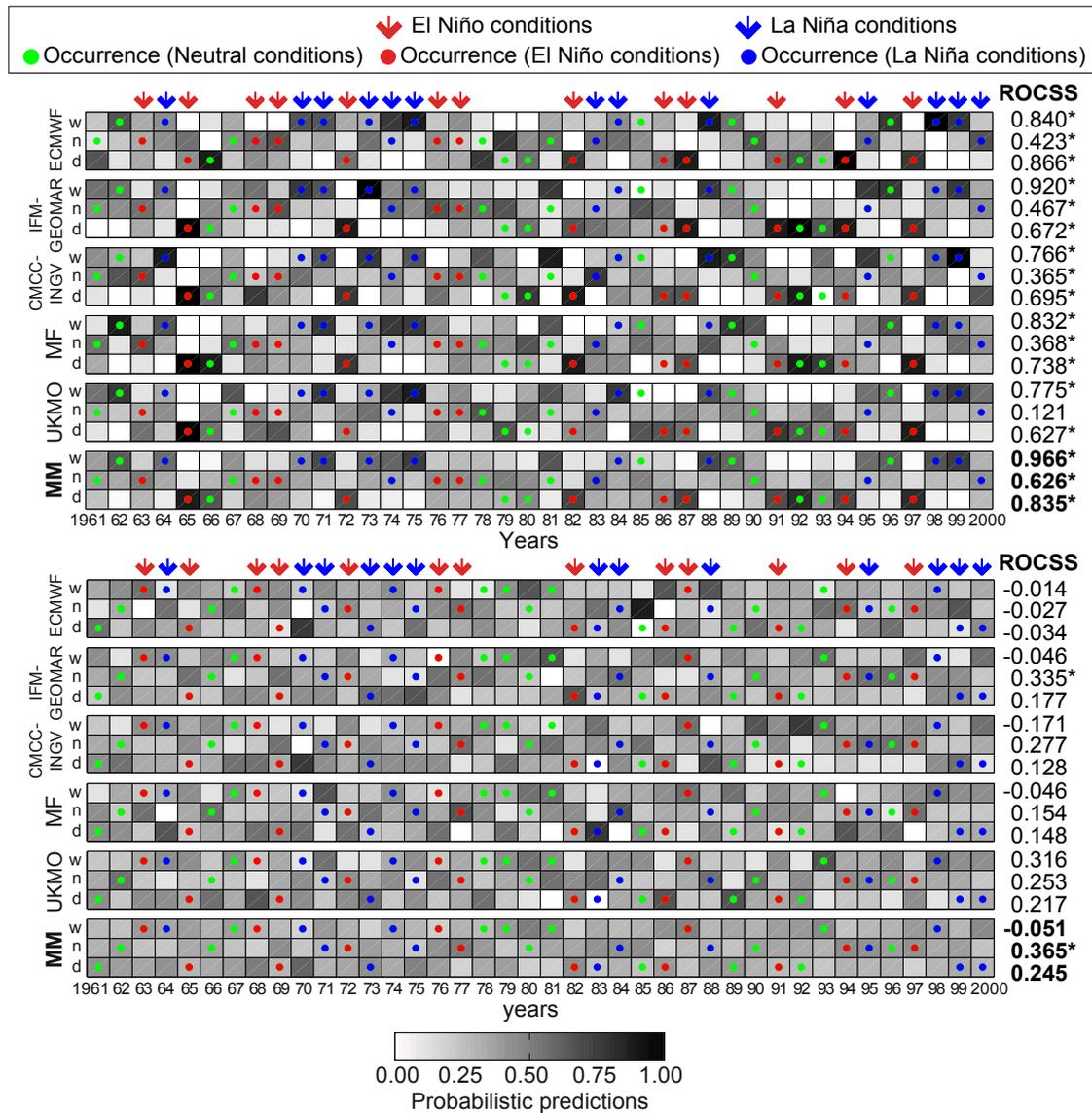


Figure 7.2: One-month lead probabilistic predictions from the five ENSEMBLES models and the MM (in bold) for SON in an illustrative grid box in (top panel) the Malay archipelago -11.25° S, 151.25° E— and (bottom panel) Europe -48.75° N, 16.25° E.— For each tercile — d , n and w stand for dry, normal and wet, respectively,— probabilities are displayed in a white (0)-to-black (1) scale. Red/blue/green points mark the observed tercile in El Niño/La Niña/neutral years. Numbers on the right show the ROCSS for each model and each tercile. Asterisks indicate significant values ($\alpha = 0.05$).

generating 1000 time series of probabilistic forecasts by randomly resampling the original 1961-2000 sequence. As an illustrative example of the validation scheme followed, Figure 7.2 shows the 1961-2000 interannual time-series of probabilistic predictions from the five models and the MM as well as the binary occurrence/non occurrence for the three terciles in two particular grid boxes—one in the Malay archipelago (top panel) and the other in Europe (bottom panel)— at one month lead-time for SON. Although varying from year to year and from model to model, predictions exhibit a higher resolution (probabilities far from 1/3) in the former point. Furthermore, in this case resolution increases in general in El Niño and La Niña conditions (marked with red and blue arrows, respectively), suggesting the existence of a predictability signal linked to ENSO in this region of the world for this season. Numbers on the right correspond to the ROCSS for the different models and terciles. High skill (over 0.7 in most of the cases) is found for the dry and wet terciles for the point in the Malay archipelago. On the contrary, almost no skill (i.e., non significant ROCSS) is found for the point in Europe.

7.3 Overall Skill (Worldwide)

The above described methodology was applied worldwide, grid box by grid box, in order to compute the ROCSS (and its corresponding significance) for the five models and the MM within the period 1961-2000, thus obtaining a measure of overall skill. As a summary of the results obtained, Figure 7.3(a-d) shows the percentage of grid boxes with significant ($\alpha = 0.05$) skill in the tropical—region between 23.5° N and 23.5° S latitudes—and the extratropical land areas, for both one- and four-month lead predictions. Although predictability varies with region, season, model and lead-time, several general conclusions can be obtained. First, the skill concentrates in the extreme (wet and dry) terciles, whereas almost no skill is obtained for normal conditions (note that the percentage of significant grid boxes is around 5% in this case, which can be explained by chance according to the significance level considered). This lack of skill for the near normal category is in agreement with previous studies (see, e.g., Van Den Dool and Toth, 1991). Second, predictability is mainly located in the tropics (with 20 to 40% of total land areas showing significant skill) rather than in the extratropics (only 10%), which is also in agreement with previous studies (see, e.g., van Oldenborgh et al., 2005). Furthermore, SON (MAM) is in general the most (least) skillful season. Third, all models yield similar results for a concrete region, season and lead-time, with the MM outperforming any of them in all cases, which is also in agreement with previous studies (see, e.g., Doblas-Reyes et al., 2009; Bundel et al., 2011; Ma et al., 2012). Finally, the spatial coverage of the skillful areas decays at four months lead-time (particularly in JJA), although not sharply. This general low decrease in skill

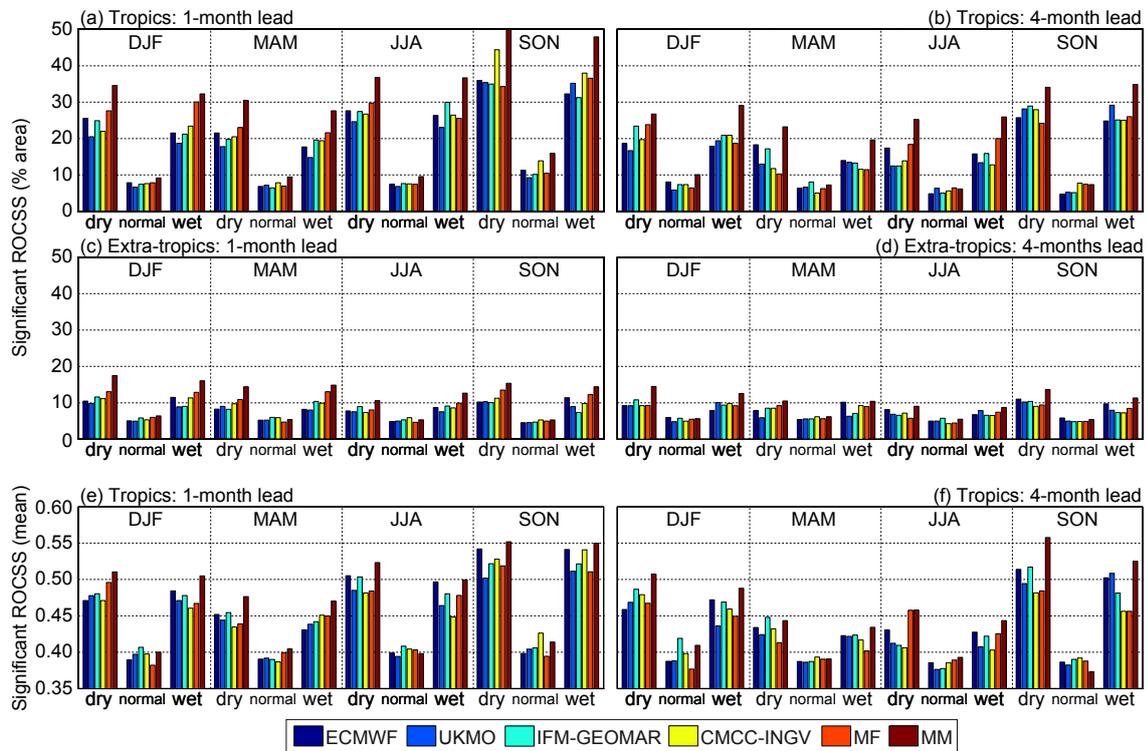


Figure 7.3: Percentage of land areas with significant ($\alpha = 0.05$) ROCSS for one- and four-month lead predictions (left and right column, respectively) from the five ENSEMBLES models and the MM (see colors in legend) in (a,b) the tropics and (c,d) the extratropics. (e,f) Mean value of the significant ROCSS in the tropics at one and four months lead-time.

with lead-time was also found by Barnston (1994), who attributed it to a persistent ENSO signal. For a full interpretation of the previous results, Figure 7.3(a-b) should be analyzed jointly with Figure 7.3(e-f), which displays the mean value of the significant ROCSS in the tropics at one and four months lead-time. Note that there is a clear correspondence between Figure 7.3(e-f) and Figure 7.3(a-b), so all the previous comments apply.

In order to further analyze the above results in the different regions of the world, global spatial maps of ROCSS were obtained for the five models and the MM. For conciseness, and given its better performance, only results for the MM are shown. Figures 7.4 and 7.5 show the significant skill for the dry (left column) and wet (right column) terciles at one and four months lead-time, respectively, by seasons (in rows). It can be seen that significant skill is mainly located over the tropics. Moreover, there is clear symmetry for dry and wet terciles. In addition, although both the signal and spatial coverage of the skillful areas slightly reduce at four months lead-time (as compared to the one-month lead case) the skill patterns are broadly preserved, particularly in DJF and SON.

By seasons, the main skillful regions at one month lead-time in DJF are the Gulf of California, northern South America, central and southern Africa, western Australia and the Pacific islands of Oceania in Melanesia, Micronesia and Polynesia. Except in Africa, where the predictability signal weakens, most of this skill remains at four months lead-time. In MAM, skill at one month lead-time is located over parts of western U.S.A., northeastern Brazil, southern Africa, parts of the Arabic peninsula, Indochina and the Malay archipelago. Most of this predictability vanishes at four months lead-time over Africa and the Arabic peninsula. In JJA, central America, northern Brazil, the Gulf of Guinea, the Malay archipelago, eastern Australia and the Pacific islands of Oceania are the main skillful regions at one month lead-time. However, most of this skill is only maintained in the Malay archipelago and the Pacific islands of Oceania at four months lead-time. Finally, one-month lead skill in SON is located over northern South America, a belt in central Africa (especially in the Somali peninsula), parts of Middle East, the Malay archipelago, Australia and the Pacific islands of Oceania. Moreover, this skill remains almost unaltered at four months lead-time for all the aforementioned regions except the Somali peninsula, thus indicating a persisting predictability signal.

In the light of the previous results, northern South America and the Malay archipelago seem to be the most skillful regions of the world for seasonal forecasting of precipitation. Note that seasonal predictability in these regions has been analyzed in previous studies (Haylock and McBride, 2001; Aldrian et al., 2007), considering also its derived socio-economic impacts (Kirono and Tapper, 1999).

7.4 Overall Skill (in the Philippines)

Both MOS-BC and PP methods are limited by the quality of the driving GCM in simulating the needed predictors; precipitation in the former case and large-scale circulation variables in the latter. Therefore, when assessing the potential to downscale precipitation it is important to first evaluate the skill of the GCMs considered over the region of interest (Maraun et al., 2010). For the case of the Philippines, a simple deterministic validation in terms of bias and interannual ACC (Figures 4.4 and 4.5, respectively) indicated that the ENSEMBLES models poorly represent local precipitation —particularly for certain key seasons and regions,— which might be related to the known errors of GCMs in simulating important large-scale phenomena like ENSO (Latif et al., 2001; Leloup et al., 2007), monsoonal circulation and tropical and extratropical cyclones (Meehl et al., 2007). In this section, we extend these results by applying the same tercile-based probabilistic validation conducted worldwide (previous section), but locally for the 42 PAGASA stations of Figure 4.1*b* over the period 1981-2005, gaining thus regional detail. Figure 7.6 shows

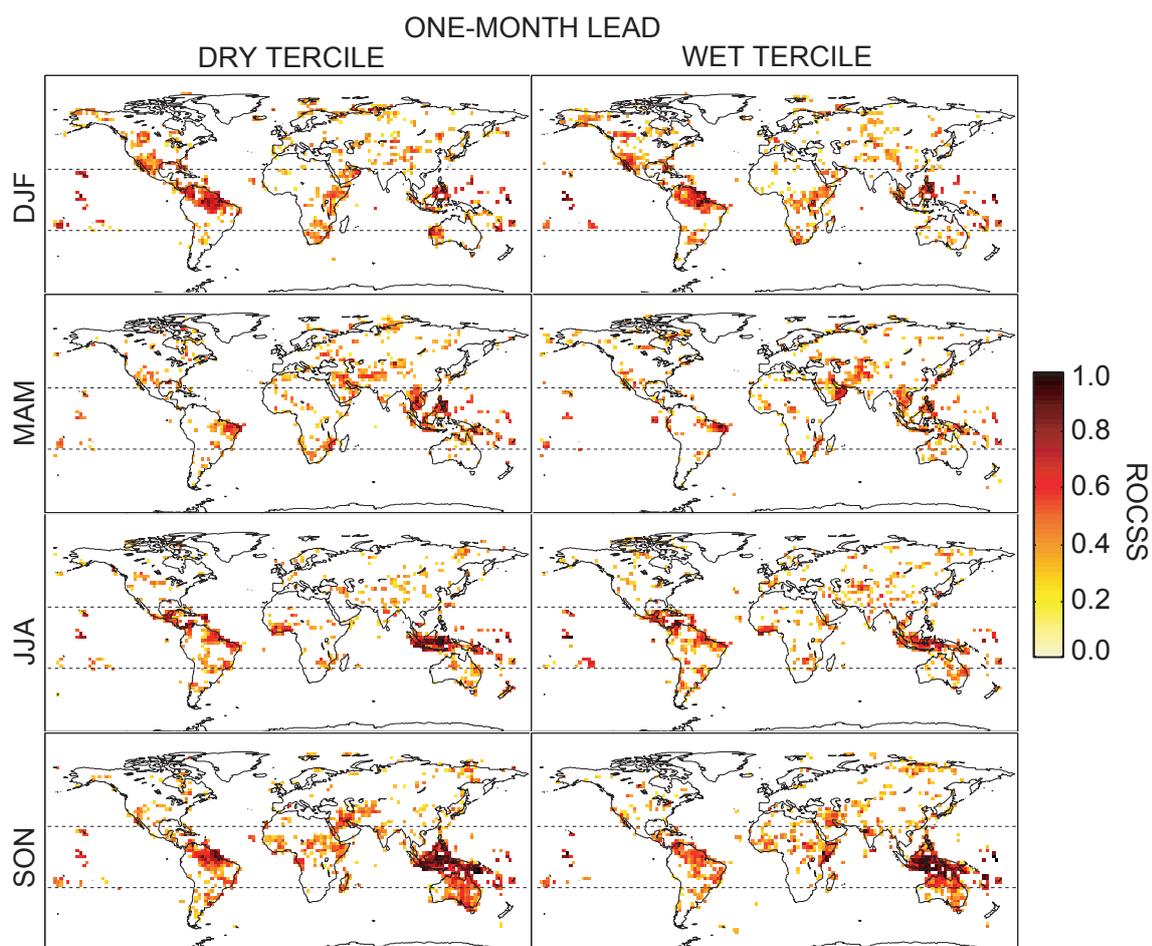


Figure 7.4: MM skill for the dry and wet tercile (left and right column, respectively) at one month lead-time for the period 1961-2000, by seasons (in rows). Only significant ($\alpha = 0.05$) ROCSS are shown. Dashed lines indicate the tropics/extratropics division.

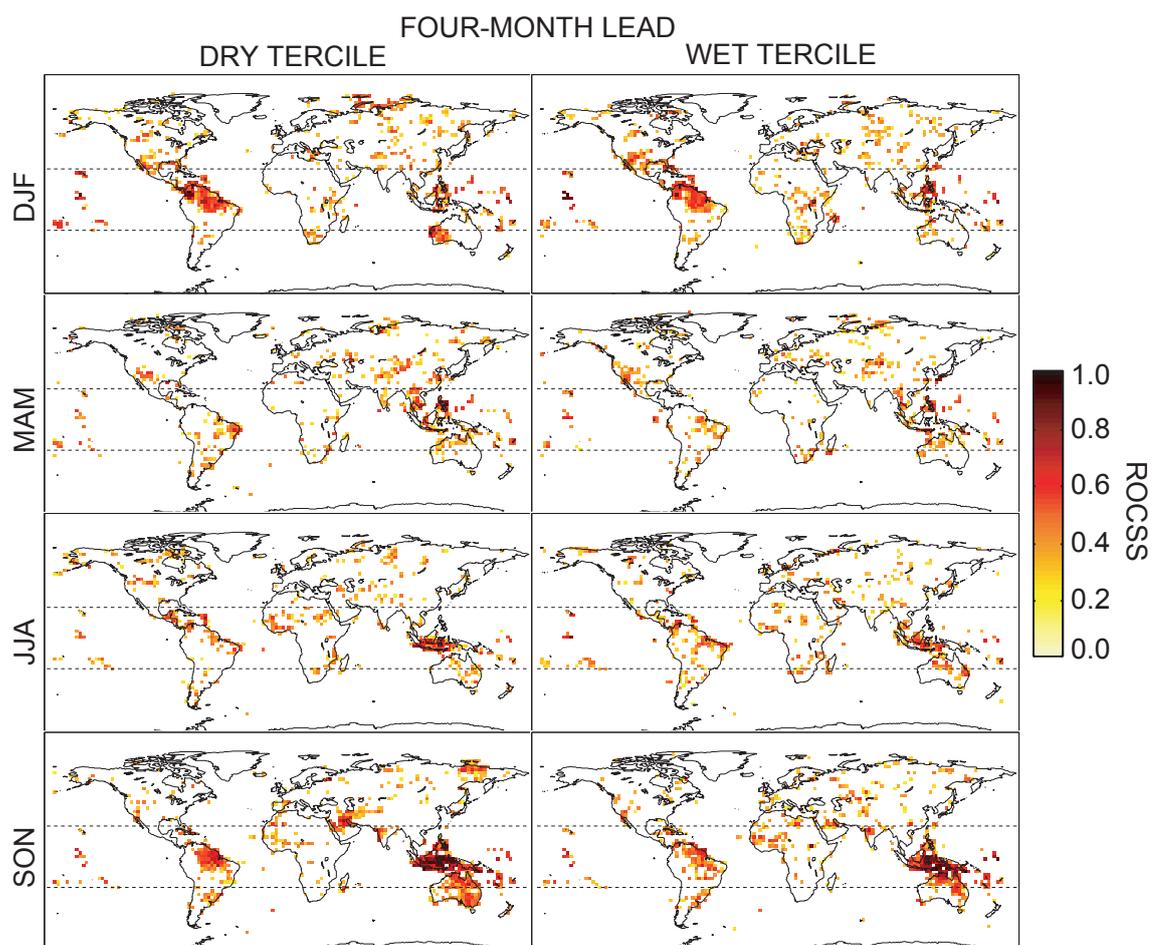


Figure 7.5: As Figure 7.4, but for the four-month lead predictions.

the results obtained for four of the ENSEMBLES models. Only one month lead-time predictions are shown (results for the four-month lead case are very similar). Significant ($\alpha = 0.05$) ROCSS are indicated with a black dot (as in Figures 7.4 and 7.5, significance was computed upon bootstrapping with 1000 samples). As found for the interannual ACC in Figure 4.5, although results vary among models, seasons and also from region to region, acceptable ROCSS are found throughout the year (especially in DJF and MAM) except for JJA, which points out the potential of the selected area for SD. Remarkably, SDMs might be particularly beneficial to overcome the limited skill found in JJA, especially for the CT1 region, where most of the annual rainfall is received due to the presence of the southwest monsoon. This possibility will be analyzed in detail in Chapter 9 by applying different PP and MOS-BC methods.

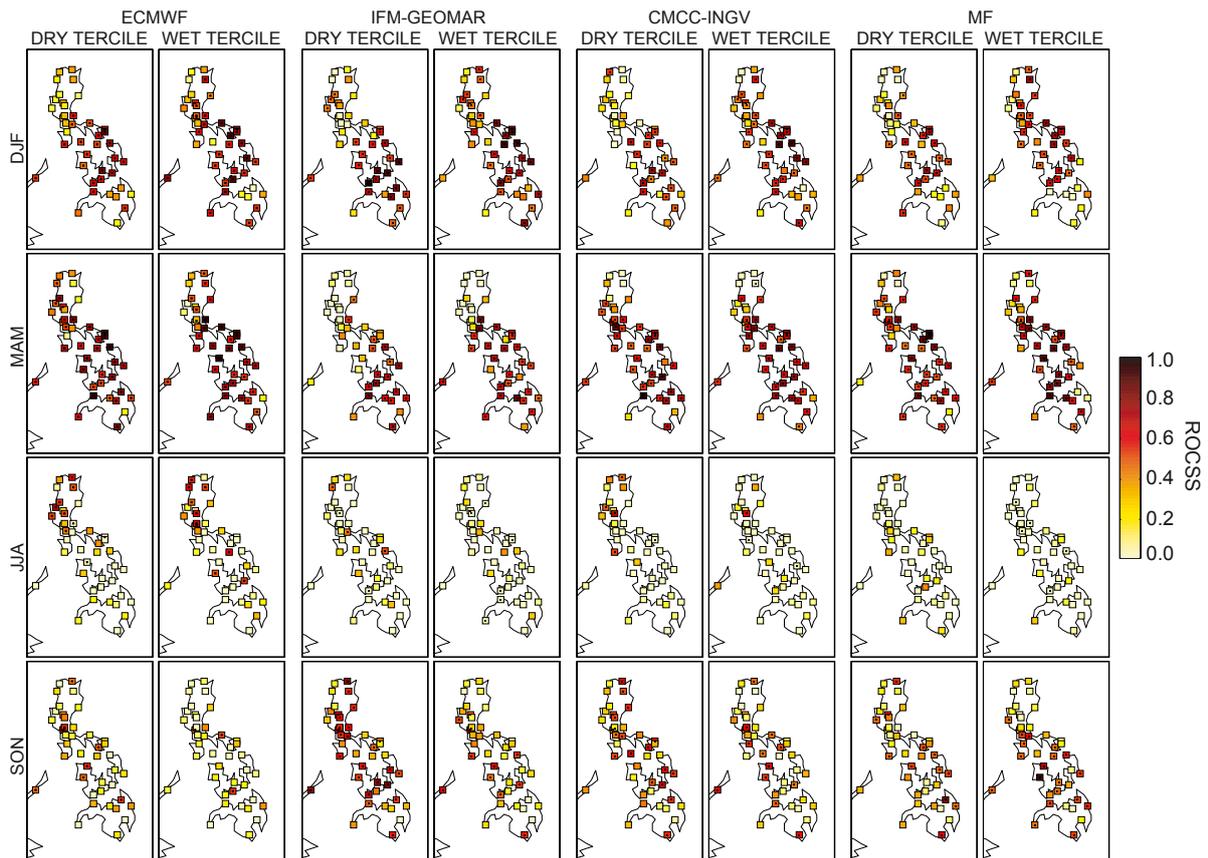


Figure 7.6: ROCSS maps for four of the ENSEMBLES models (dry and wet terciles) for the 42 PAGASA stations of Figure 4.1b at one month lead-time for the period 1981-2005, by seasons (in rows). Significant ($\alpha = 0.05$) values are indicated with a black dot.

7.5 ENSO-Driven Skill

Despite the important achievements reached in seasonal forecasting in the last ten years, significant levels of skill for precipitation are only generally found over regions connected with ENSO (see, e.g., Coelho et al., 2006; Barnston et al., 2010; Arribas et al., 2011; Lim et al., 2011; Kim et al., 2012a,b; Landman and Beraki, 2012), which is known to be the dominant mode of seasonal variability (see, e.g., Goddard and Dilley, 2005; Doblas-Reyes et al., 2010). Therefore, in this section we analyze both the direct (through the SST anomalies in El Niño 3.4 region) and indirect (through its associated atmospheric teleconnections) influence of ENSO on the skill obtained for precipitation in Sections 7.3 and 7.4.

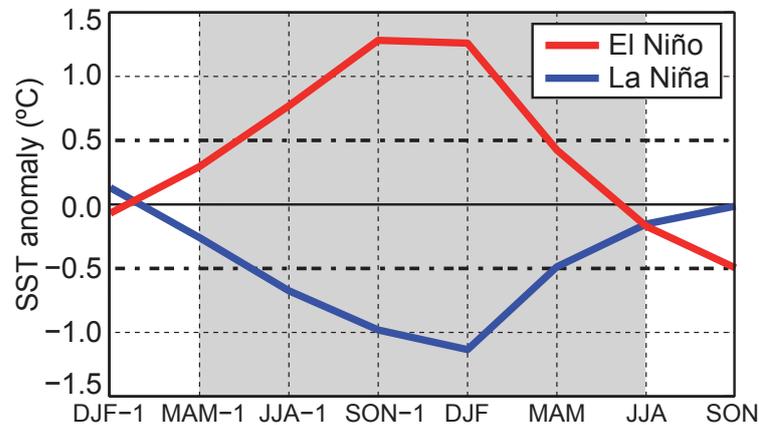


Figure 7.7: Mean observed SST anomaly in El Niño 3.4 region for the (red) El Niño and (blue) La Niña events considered. Observations come from the ERSST v3b dataset (Smith et al., 2008).

7.5.1 Sea Surface Temperature in El Niño 3.4 Region

Although alternative indices for the definition of warm (El Niño) and cool (La Niña) ENSO events have been proposed, the Oceanic Niño Index (ONI) —based on the SST anomalies in El Niño 3.4 region (5°N-5°S, 120°W-170°W)— has become the de-facto standard used by the National Oceanic and Atmospheric Administration of the United States (NOAA: <http://www.noaa.gov>). According to this index, an El Niño (La Niña) event is defined when the SST anomalies in five consecutive overlapping 3-month seasons remain equal or above (equal or below) the 0.5°C (−0.5°C) threshold. Here, we adopted this definition and analyzed the 18-month period spanning from the spring of the onset year to the summer of the decay year —as indicated by the shading in Figure 7.7.— Note that this period is centered in SON and DJF, when the SST anomalies reach their maximum (minimum)

value. The following (decay) years were considered for El Niño (La Niña) events: 1964, 1966, 1969, 1970, 1973, 1977, 1978, 1983, 1987, 1988, 1992, 1995 and 1998 (1965, 1971, 1972, 1974, 1975, 1976, 1984, 1985, 1989, 1996, 1999, 2000 and 2001). In the following, we use the notation MAM-1 and JJA-1 (MAM and JJA) to refer to the seasons of onset (decay) year of the event.

In order to assess the performance of the different models to predict ENSO, we computed the correlation between the observed (ERSST v3b) and the simulated SST in El Niño 3.4 region during the above El Niño and La Niña episodes. Figure 7.8 shows the results for (left) one- and (right) four-month lead predictions. The poorest skill is obtained for summer (both JJA-1 and JJA), when correlations decrease substantially from one to four months lead-time. Note that this is in agreement with the spring predictability barrier for ENSO documented in previous studies (see, e.g., Zheng and Zhu, 2010; Tippett et al., 2011; Yan and Yu, 2012; Duan and Wei, 2013).

However, to properly disentangle the role of ENSO in the overall skill it is needed to analyze not only the models' ability to forecast the phenomenon itself (as characterized by the SST in El Niño 3.4 region), but also the remote effect of its associated atmospheric teleconnections —carrying the predictability signal to the different regions of the world (time-lagged in some cases).—

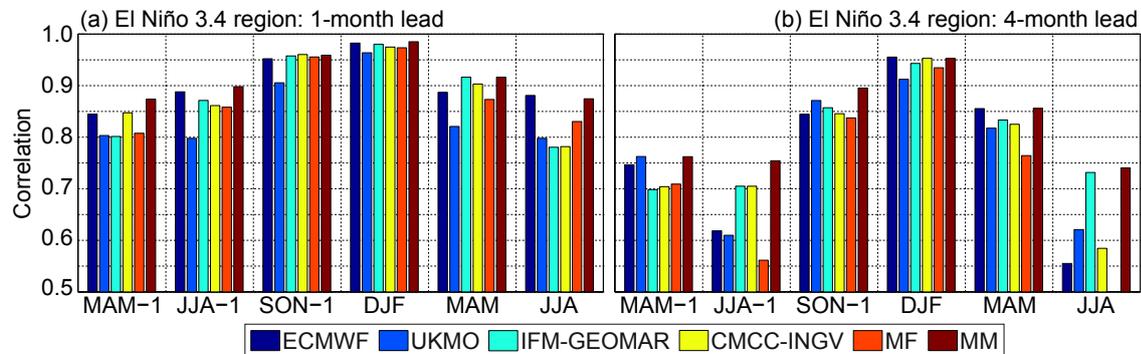


Figure 7.8: Correlation between the observed (ERSST v3b) and simulated SST in El Niño 3.4 region during El Niño and La Niña episodes for the five ENSEMBLES models and the MM (see colors in legend), at (a) one and (b) four months lead-time.

7.5.2 ENSO Teleconnections (Worldwide)

ENSO teleconnections with precipitation were calculated, following a tercile-based approach, in terms of the frequency of occurrence of each category (dry, normal and wet) in El Niño and La Niña events considered —as compared to the expected climatological frequency 1/3.— A chi-square test for equality of proportions was applied to detect

those frequencies significantly higher (lower) than $1/3$, which were considered as significant positive (negative) ENSO teleconnections.

Figure 7.9 shows the percentage of areas exhibiting significant ($\alpha = 0.05$) El Niño (red) and La Niña (blue) teleconnections in (a) the tropics and (b) the extratropics, by seasons. Tropical teleconnections are stronger than extratropical ones, with different influence of El Niño and La Niña for different seasons; particularly in MAM, when La Niña has a greater effect than El Niño. Thus, in the following we restrict the analysis to the tropics.

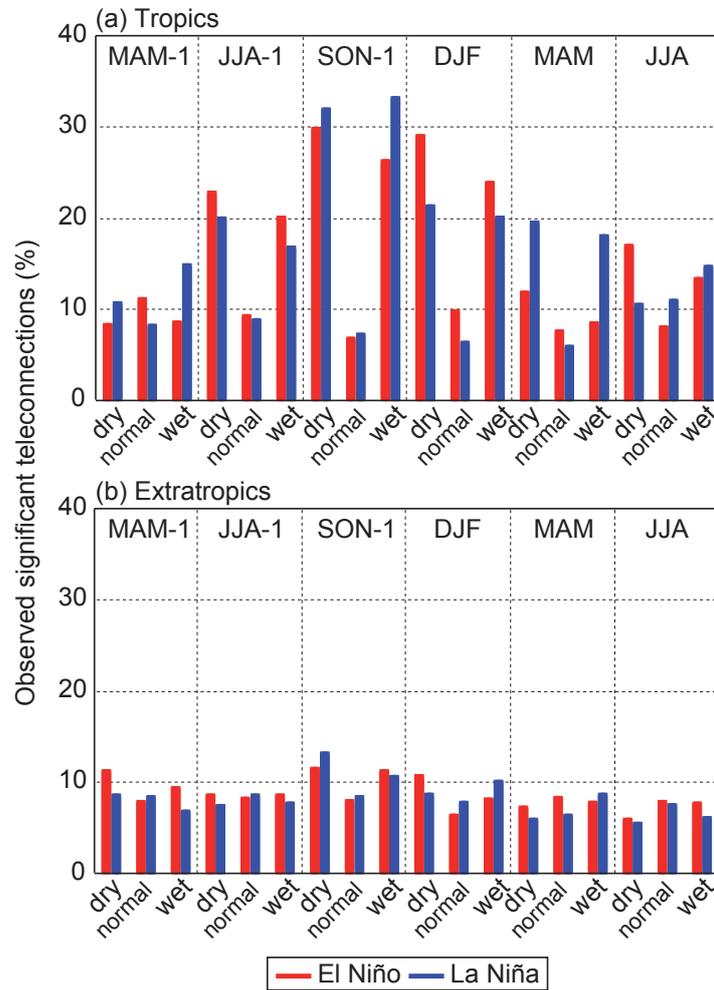


Figure 7.9: Percentage of areas showing significant (red) El Niño and (blue) La Niña teleconnections with precipitation in (a) the tropics and (b) the extratropics, by seasons.

Figure 7.10 shows the maps of ENSO teleconnections in this region for the different seasons (in rows). Red/blue colors indicate high/low frequency of occurrence of the corresponding dry (left) and wet (right) terciles during El Niño (top panel) and La Niña

(bottom panel) events. Notice that El Niño and La Niña —as opposite phases of the same underlying phenomenon— tend to yield similar patterns but with opposite signals, although there are some exceptions —e.g., the Malay archipelago and northern Australia are teleconnected in MAM with La Niña but not with El Niño.— Overall, the results are in agreement with previous studies (see, e.g., Ropelewski and Halpert, 1987; van Oldenborgh et al., 2000; Kayano et al., 2009; Shaman and Tziperman, 2011; Zhang et al., 2012; Yadav et al., 2013; Zhang et al., 2013).

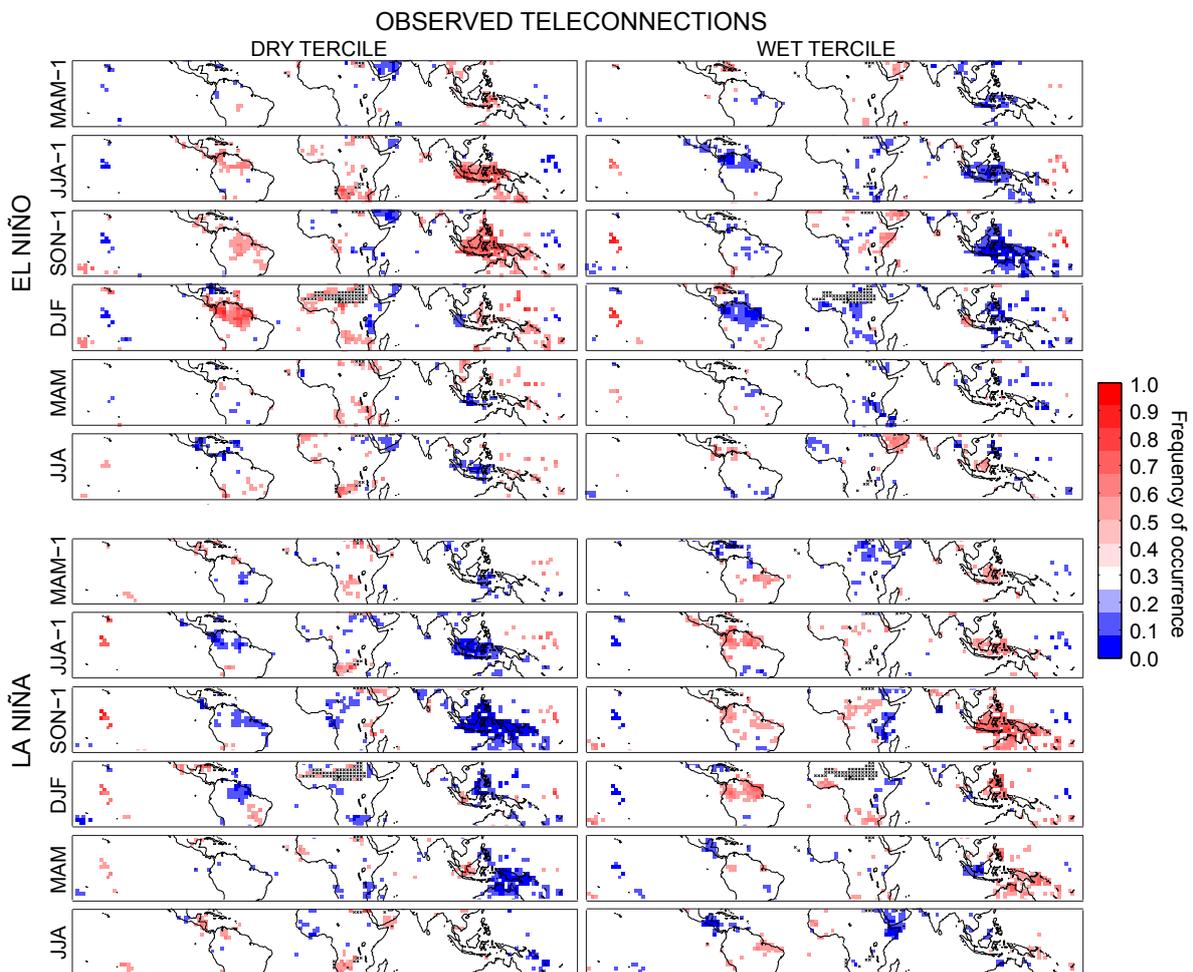


Figure 7.10: Relative frequency of occurrence for the dry and wet tercile (left and right column, respectively) in the tropics in (top panel) El Niño and (bottom panel) La Niña periods, for the different seasons (in rows). Red (blue) colors correspond to values above (below) $1/3$ —the expected climatological frequency.— Only significant ($\alpha = 0.05$) teleconnections, according to a chi-square test, are displayed. Black crosses indicate grid boxes where precipitation categories can not be properly defined (series with less than three different values).

Comparison of Figure 7.10 with Figures 7.4 and 7.5 reveals that, in general, the skillful zones are significantly teleconnected with ENSO —e.g. northern South America in DJF and SON and the Malay archipelago in JJA and SON.— This suggests that the seasonality and the spatial distribution of the overall skill could be explained by this phenomenon through its associated teleconnections. Therefore, it is important to assess the ability of the models to properly reproduce El Niño and La Niña precipitation teleconnections. Figure 7.11 shows the spatial correlation between the observed and simulated El Niño and La Niña teleconnections patterns —as given by the maps of terciles frequencies,— considering one- (left) and four-month (right) lead predictions from the five single models and the MM (see colors in legend), over their corresponding skillful regions within the tropics. Note that, although a similar analysis has been done in Yang and Delsole (2012), who compared observed and simulated ENSO teleconnections using a field regression analysis, they did not assess the predictive skill of the teleconnections found, which is the aim here. As can be seen, the agreement is good for both lead-times, with correlations over 0.8 in some seasons. On the one hand, the observed patterns —for both El Niño and La Niña— are best reproduced in SON-1 and DJF, the central seasons of the phenomenon. On the other hand, they are worst reproduced in MAM-1 in all cases.

Notice that in spite of the aforementioned spring predictability barrier for ENSO (Figure 7.8), the models exhibit a relative good performance in reproducing the existing —both El Niño and La Niña— teleconnections in summer (particularly in JJA-1), which could explain the overall skill found for precipitation in this season (Figure 7.3).

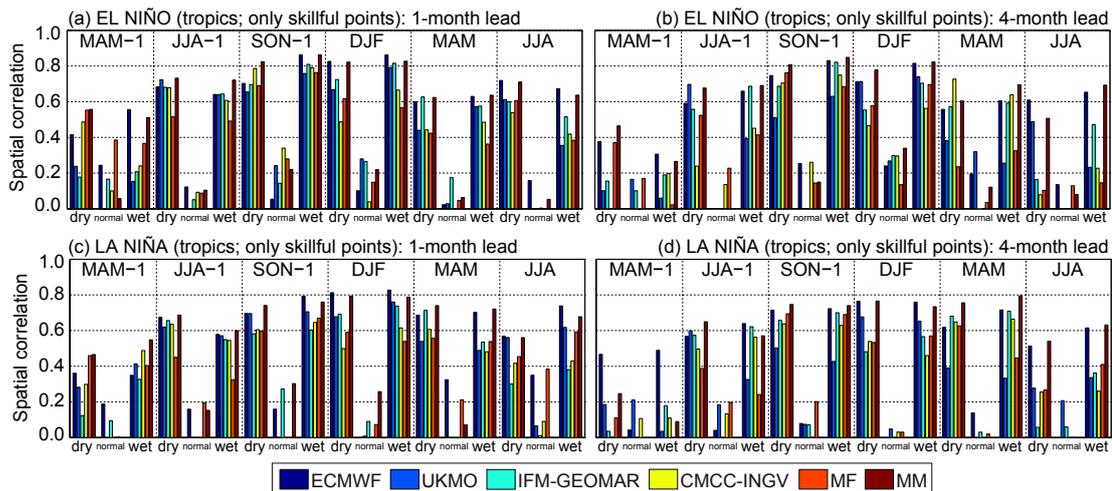


Figure 7.11: Spatial correlation between the observed and predicted (*a-b*) El Niño and (*c-d*) La Niña teleconnection patterns for the tropical regions showing significant skill at one and four months lead-time (left and right column, respectively), for the five ENSEMBLES models and the MM (see colors in legend).

7.5.3 ENSO Teleconnections (in the Philippines)

The methodology introduced in the previous section was also applied for a regional study of ENSO teleconnections in the Philippines, considering observed precipitation at the 42 PAGASA stations for the period 1981-2005. In this case, the El Niño (La Niña) years considered were: 1983, 1987, 1988, 1992, 1995, 1998, 2003 and 2005 (1984, 1985, 1989, 1996, 1999, 2000 and 2001). In agreement with previous results from Lyon et al. (2006), Figure 7.12 highlights the strong influence of ENSO on the climate of this country (Koide et al., 2012), bringing in general below (above) normal rainfall during El Niño (La Niña), except for JJA, when the signal reverses (being the latter clearer for El Niño than for La Niña episodes). Note that the weakest teleconnections are found for summer (both JJA-1 and JJA), the season for which the models exhibit the poorest skill (Figure 7.6).

To better understand the results obtained for this region, Figure 7.13 shows the observed interannual variability of spatial average precipitation totals for each CT (in colors), by seasons. The red (blue) vertical lines mark the above indicated El Niño (La Niña) years. Note that the low year-to-year observed variability found in JJA (with the exception of CT1) is associated with the aforementioned weaker effect of ENSO in this season —ENSO amplifies the local observed interannual seasonal climate variations in teleconnected regions.—

7.5.4 Contribution of ENSO to the Overall Skill

All the previous results suggest the idea that the seasonality and the spatial distribution of the overall skill found in Sections 7.3 and 7.4 may be mainly driven by the indirect effect ENSO through its associated atmospheric teleconnections and therefore limited by the ability of the different models to reproduce the observed El Niño and La Niña teleconnections with precipitation. Thus, to further assess the role of ENSO on the global skill for each particular model (and the MM) we computed the spatial correlation between the corresponding ROCSS maps for 1) the full period 1961-2000 —shown in Figures 7.4 and 7.5 for the MM— and 2) the twenty-six El Niño and La Niña events considered. Figure 7.14 shows the results obtained for one- (left) and four-month (right) lead predictions in the tropics for JJA-1, SON-1, DJF, and MAM, the seasons with the strongest teleconnections (Figure 7.9).

Correlations are very high (over 0.85) in most of the seasons at both one- and four-month lead, which confirms that the overall (1961-2000) skill attained in this region may be mainly explained by the contribution of El Niño and La Niña years.

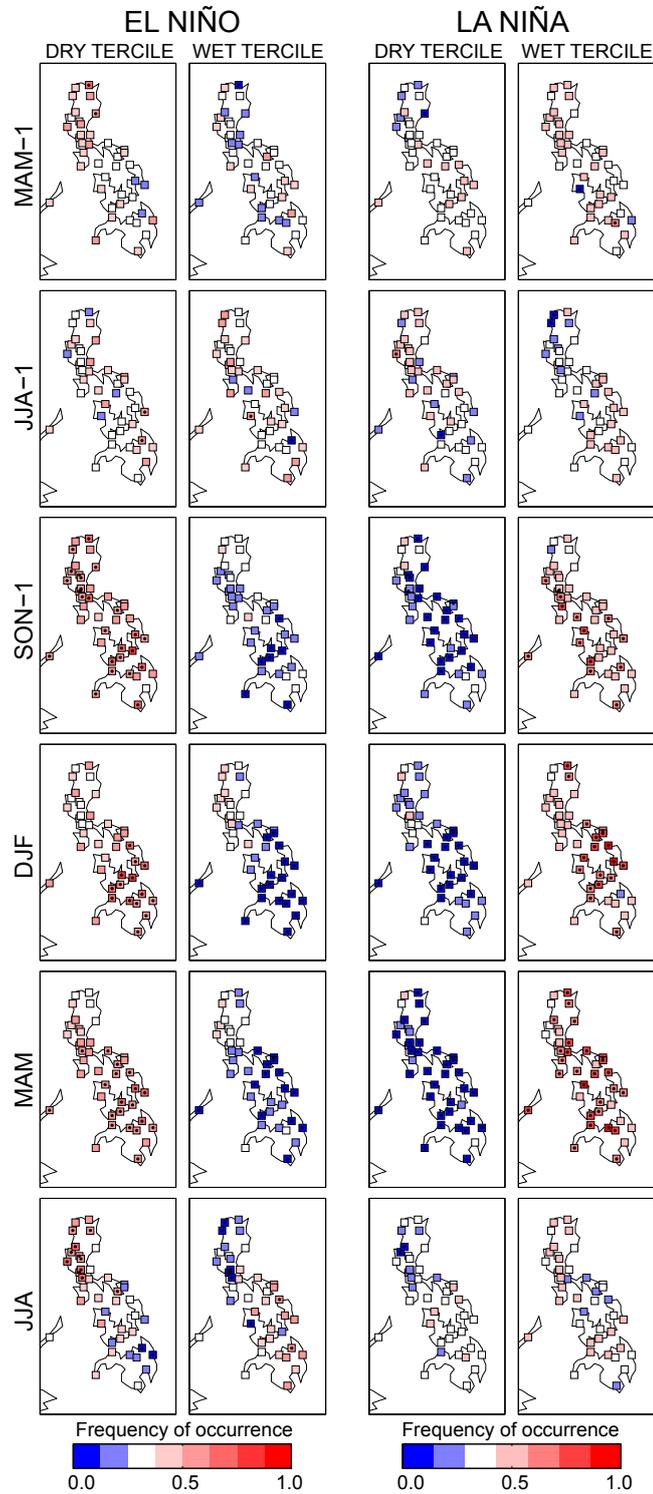


Figure 7.12: As Figure 7.10 but for observed precipitation at the 42 PAGASA stations for the period 1981-2005. Significant ($\alpha = 0.05$) teleconnections, according to a chi-square test, are marked with a black dot.

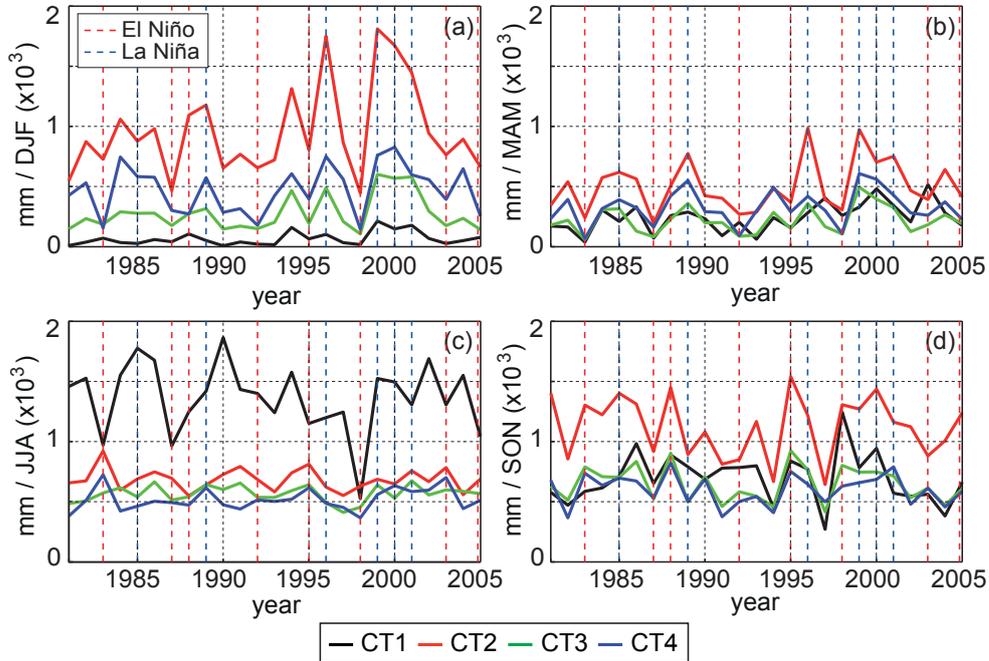


Figure 7.13: Interannual variability of spatial average precipitation totals for each CT (see colors in the legend) for the period 1981–2005, by seasons. The red (blue) vertical lines mark the El Niño (La Niña) years.

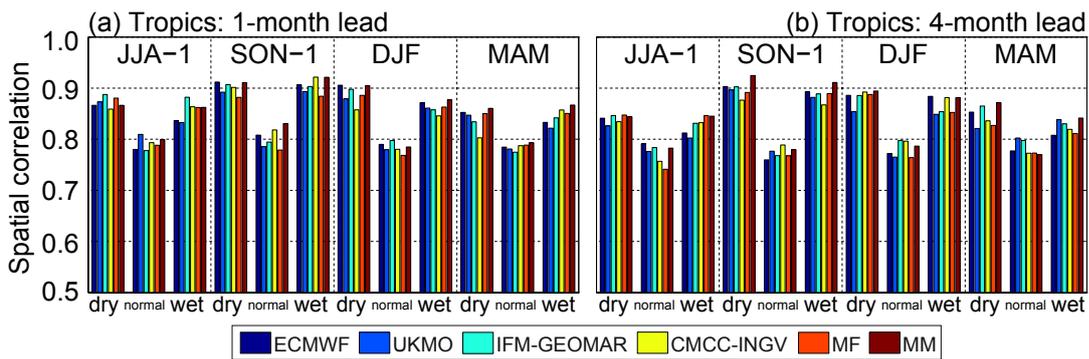


Figure 7.14: Spatial correlation between the ROCSS maps for the forty-year period (1961–2000) and those for the twenty-six El Niño and La Niña events at (a) one- and (b) four-month lead for the JJA-1, SON-1, DJF and MAM seasons (see Figure 7.7 for further details on the definition of seasons). Results are shown for each model and the MM (see colors in legend).

CHAPTER 8

The Effect of Reanalysis Uncertainty on Perfect Prog

In previous chapters it was shown that the skill of global seasonal forecasts to reproduce local precipitation in the Philippines was limited for certain key seasons and regions (see Figures 4.4, 4.5 and 7.6). SD might be suitable to overcome these shortcomings. However, as mentioned in Chapter 4, SDMs have been developed and applied almost exclusively for extratropical regions since manifold problems still hinder their successful application in the tropics (Hewitson et al., 2014). In particular, the success of PP methods in the extratropics relies on the fact that a large fraction of local-scale climate variability can be described by atmospheric phenomena operating on spatial scales of the order of thousands of kilometres, typically having a lifetime of several days. At this scale, reanalyses are known to be skillful, in the sense that their spatio-temporal resolution captures the relevant processes such as extratropical cyclones and the associated fronts (Groth and MacCracken, 1991; Widmann et al., 2003). At lower-latitudes, however, the atmospheric drivers of local climate variability operate on much finer spatial and temporal scales and are generally poorly captured by reanalyses. Moreover, observational coverage is generally sparse in the tropics, leading to considerable differences between distinct reanalyses (Trenberth et al., 2001; Sterl, 2004; Brands et al., 2012, 2013) and errors with respect to observational records (Manzanas et al., 2014a), which in turn, can complicate the detection of a proper predictor-predictand relationship. As a result from these considerations, reanalysis choice could be a source of uncertainty (or error) for the application of the PP methods considered in this Thesis (note that MOS-BC methods do not use reanalysis data and, therefore, are not affected by this issue). Thus, focussing on the Philippines, this chapter assesses the impact of the choice of reanalysis on the downscaled results, both in ‘perfect’ (i.e., in cross-validation mode using reanalysis predictor data for both

the calibration and the prediction phase) and ‘non-perfect’ conditions. In the latter case, the statistical models calibrated with different reanalyses are applied to GCM predictors and the differences in the downscaled results are quantified, characterizing the sensitivity of the method considered to the choice of reanalysis used for calibration. Two different time-horizons are considered: seasonal forecasts (for the period 1981-2005) and climate change projections (up to the end of the 21st century). Note that, although climate change is not the main concern of this Thesis, the results found allow to better understand the influence of the choice of reanalysis on SD, which is shown to be of minor importance for seasonal forecasting, as compared to the case of climate change.

8.1 *Data and SD Method Used*

For this chapter, the predictor variables listed in Table 6.1 were obtained from two distinct reanalyses —ERA-Interim¹ (Dee et al., 2011) and JRA-25² (Onogi et al., 2007),— four out of the five ENSEMBLES models (see Section 3.4.1) and the Max Planck Institute (MPI) ECHAM5³ model (Giorgetta et al., 2006). For the latter, control and A1B scenario data from the 3rd transient run developed within the ENSEMBLES project were retrieved. To keep consistency between the time-steps available for both reanalyses and the different GCMs considered, daily instantaneous values at 00 UTC were chosen in all cases. Due to the distinct native resolutions, predictor data from all sources were regridded onto the regular 2° grid shown in Figure 6.1 (the domain considered for SD in this Thesis) by means of bi-linear interpolation. For each of the 42 PAGASA gauges (red points in Figure 6.1), the standardized anomalies at the 4 nearest grid boxes were considered as predictor data for a deterministic GLM (see Section 5.4.1-2). For the case of the ENSEMBLES models and the MPI-ECHAM5, predictor data was properly harmonized as explained in Section 6.3 before entering the SDM. Finally, for the MPI-ECHAM5, scenario data was standardized by removing the mean of the control period from the mean of the corresponding scenario period and dividing by the standard deviation of the control period.

8.2 *Reanalysis Uncertainty (in the Philippines)*

Figure 8.1 shows a comparison between ERA-Interim (taken as reference) and JRA-25 reanalyses —see <http://reanalyses.org/atmosphere/comparison-table> for details of both reanalyses— for the four daily predictor variables in Table 6.1 over the CORDEX-East Asian domain (http://wcrp-cordex.ipsl.jussieu.fr/images/pdf/cordex_regions).

¹<http://www.ecmwf.int/en/research/climate-reanalysis/era-interim>

²http://jra.kishou.go.jp/JRA-25/index_en.html

³<http://cera-www.dkrz.de/WDCC/ui>

pdf) for the period 1981-2005. For regional detail, the top panel of Figure 8.2 is limited to the domain considered for SD (Figure 6.1).

The left column shows the mean difference (bias) between both reanalyses —expressed in % of ERA-Interim’s standard deviation.— The middle column displays the ratio of variances (RV) —defined as σ_J^2/σ_E^2 ,— where σ_J (σ_E) is the variance of JRA-25 (ERA-Interim), respectively. In the right column, the Pearson correlation coefficient (r) between the two reanalyses is depicted.

There are appreciable differences (systematically lower for U850 and U300 than for Q850 and T850) between both reanalyses for the three validation measures considered, indicating that the PP assumption (reanalysis data reflecting ‘real’ large-scale atmospheric conditions) does not hold for the area under study. In particular, uncertainty is high over the Tibetan Plateau —probably due to different orography representations— and generally increases towards the Equator, affecting the Philippines. Nevertheless, with respect to their application for SD, recall that the reanalysis time-series are standardized to have zero mean and unit variance. Consequently, differences in the mean and variance between the two reanalyses (left and middle columns) do not affect the downscaled results, whereas differences in the third and fourth order moments —i.e., skewness and kurtosis, (see, e.g., Brands et al., 2011)— and in day-to-day variations (right column) remain and are expected to affect them.

Additionally, in the bottom panel of Figure 8.2, the zonally averaged r between the predictor time-series from JRA-25 and ERA-Interim is displayed for the specific case of the Philippines. The grid box coordinates are mapped on the left hand side and r as a function of latitude is displayed on the right hand side. Noticeably, U850 exhibits values around 0.95 at all latitudes, which indicates that both reanalyses are in nearly perfect agreement for this variable. However, a north-south gradient is found for the remaining variables. In particular, correlations for T850 and Q850 drop from 0.95 to 0.70 and from 0.75 to 0.50, respectively, which probably reflect the increasing influence of sub-grid processes —subject to reanalysis/model-dependent parametrization schemes— towards the Equator.

8.3 Sensitivity to Reanalysis Choice in Cross-Validation Mode

To assess the effect of the reanalysis uncertainty found in the previous section, the SDM considered is separately calibrated with ERA-Interim and JRA-25 data in order to downscale daily precipitation for the 42 PAGASA stations for the period 1981-2005. As in previous chapters, the 5-fold cross-validation introduced in Section 5.2 is used. Similarly as in Figure 6.2, Figure 8.3 displays the Spearman correlation coefficient (r_s) between daily observed and downscaled precipitation time-series over the period 1981-2005 for four

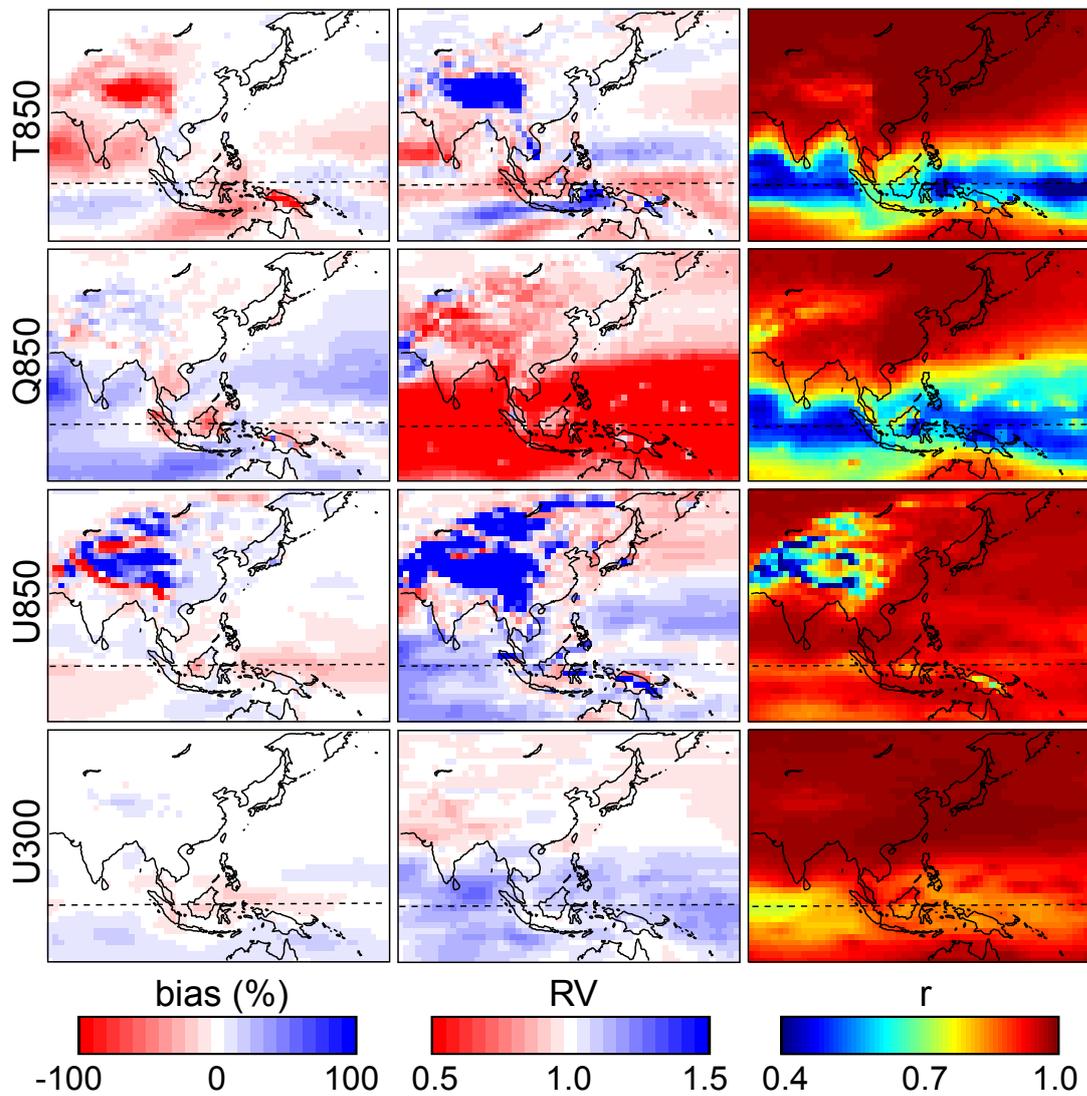


Figure 8.1: Comparison between ERA-Interim (taken as reference) and JRA-25 for the four daily predictor variables in Table 6.1 (in rows) over the CORDEX-East Asian domain for the period 1981-2005. The metrics considered (in columns) are explained in the text. The dashed lines in the maps indicate the Equator.

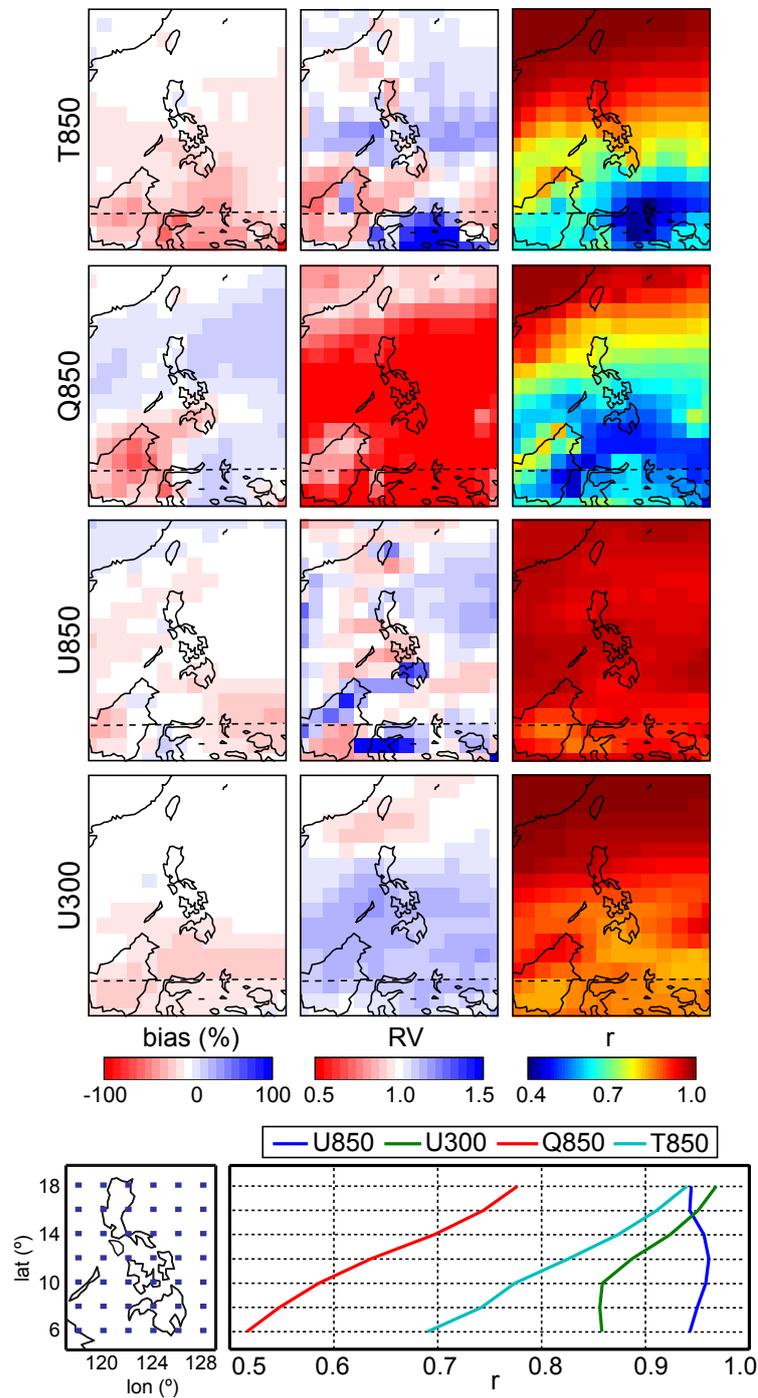


Figure 8.2: Top: As Figure 8.1 but limited to the geographical domain considered for SD (Figure 6.1). Bottom right: Pearson correlation coefficient between the two reanalysis daily time-series, as a function of latitude (displayed are zonal averages) for the Philippines. Different colors indicate different predictor variables. Bottom left: Grid box coordinates used for computing the zonal averages.

illustrative predictor combinations—in particular, the simplest (most complete) P1 (P4) plus Q850 and T850 alone—but for the case of considering predictor data from both ERA-Interim and JRA-25 (solid and dashed lines, respectively). In each panel, the results for a specific CT (see Section 4.1) are shown. Along the x -axis, stations are sorted by decreasing latitude (from left to right). On the right hand side, the CT-averaged r_s are indicated. Points (asterisks) correspond to ERA-Interim (JRA-25) predictor data.

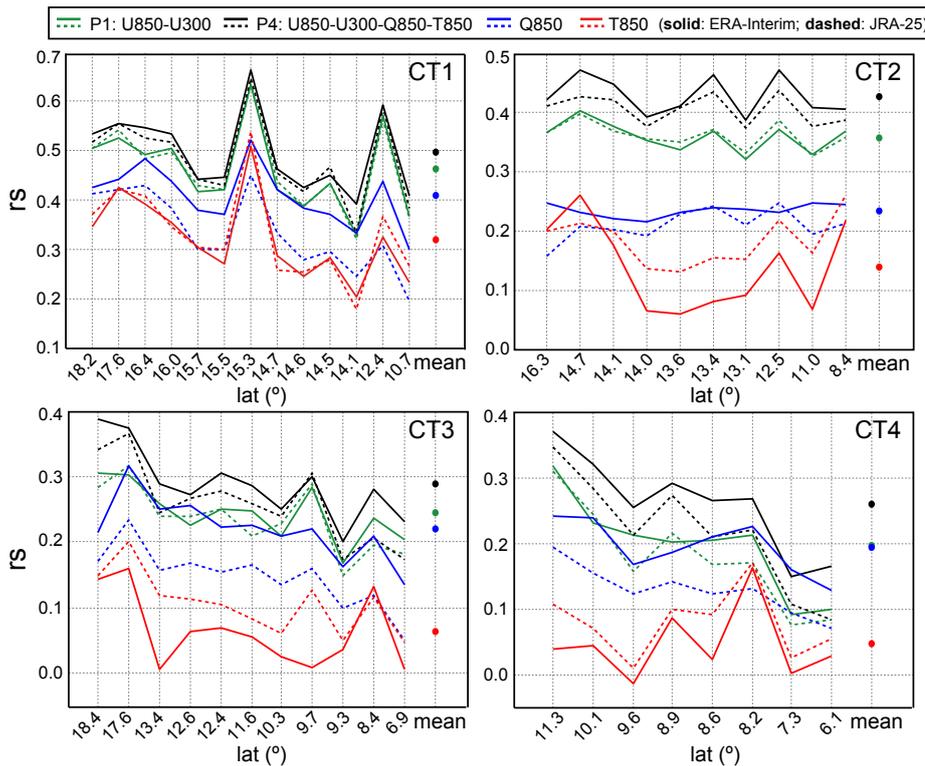


Figure 8.3: Cross-validated results for each CT, as measured by the Spearman correlation coefficient between observed and downscaled daily precipitation amount (period: 1981–2005). Different colors correspond to different predictor combinations (see the legend) and solid (dashed) lines refer to the results obtained from using ERA-Interim (JRA-25) predictor data. For each CT, results are sorted by the latitude of the stations (decreasing from left to right). CT-specific spatial average values are shown on the right hand side of each panel; points (asterisks) correspond to ERA-Interim (JRA-25).

For the sole use of circulation variables (P1), the downscaled results are generally not sensitive to reanalysis choice, except for the stations situated in the south (CT4). This is in agreement with the small differences found between ERA-Interim and JRA-25 for U850 and the slight north-south uncertainty gradient detected for U300 (see Figure 8.2). However, for Q850 and T850, appreciable reanalysis-induced differences are found. In particular, Q850 from ERA-Interim yields better results than Q850 from JRA-25, whereas

the opposite is the case for T850 (with the exception of CT1). This indicates that the ‘real’ statistical relationship between Q850 (T850) and local-scale precipitation is more accurately captured by ERA-Interim (JRA-25). Moreover, when considering the best predictor combination (P4), results are systematically better for ERA-Interim than for JRA-25. Notably, the southward loss of predictive potential occurring in all CTs except CT2 is in agreement with the southward increase of reanalysis uncertainty found (compare Figure 8.3 with Figure 8.2).

For a geographical overview of these results, Figure 8.4 shows the mean point-wise cross-validated rs when considering ERA-Interim (left column) and JRA-25 (middle column) predictor data, for the P1-P4 predictor combinations (in rows). The corresponding differences —JRA-25 minus ERA-Interim— are displayed in the right column, so positive (negative) values indicate that JRA-25 (ERA-Interim) is more appropriate for SD. Due to the lower predictive potential described above, results for the single predictor variables (Q850 and T850) are not included in this figure.

For circulation predictors only, both reanalyses perform similarly (first row). However, as shown in the second (third) row, if Q850 (T850) is added to circulation, better results are obtained for ERA-Interim (JRA-25). Notably, for the case of including T850, the advantage of JRA-25 over ERA-Interim is most obvious along the eastern coastline. When considering the full predictor combination (P4), ERA-Interim systematically outperforms JRA-25 at all stations.

To further assess the increase in predictive potential from adding temperature and moisture information to circulation, Figure 8.5 shows the difference in rs — $\delta(rs)$ — obtained when adding Q850 and T850 separately (P2 and P3, respectively) and in combination (P4) to the basic circulation variables (P1). Results for calibrating with ERA-Interim and JRA-25 are given in the left and middle column, respectively. Additionally, the corresponding performance differences —JRA-25 minus ERA-Interim— are shown in the right column. Positive (negative) values indicate a larger increment for JRA-25 (ERA-Interim).

In congruence with Figures 8.2 and 8.3, the performance improvement attained when adding Q850 (T850) is larger for ERA-Interim (JRA-25) than for JRA-25 (ERA-Interim). Moreover, when including Q850 + T850 to the basic circulation-based predictor, the improvement is larger for ERA-Interim than for JRA-25. These results prove that reanalysis uncertainty can slightly affect the results from SD in cross-validation mode. In particular, for the case of the Philippines, up to 0.1 correlation points can be missed for particular locations depending on the choice of reanalysis if climate change signal bearing variables such as temperature and/or humidity are included in the predictor field.

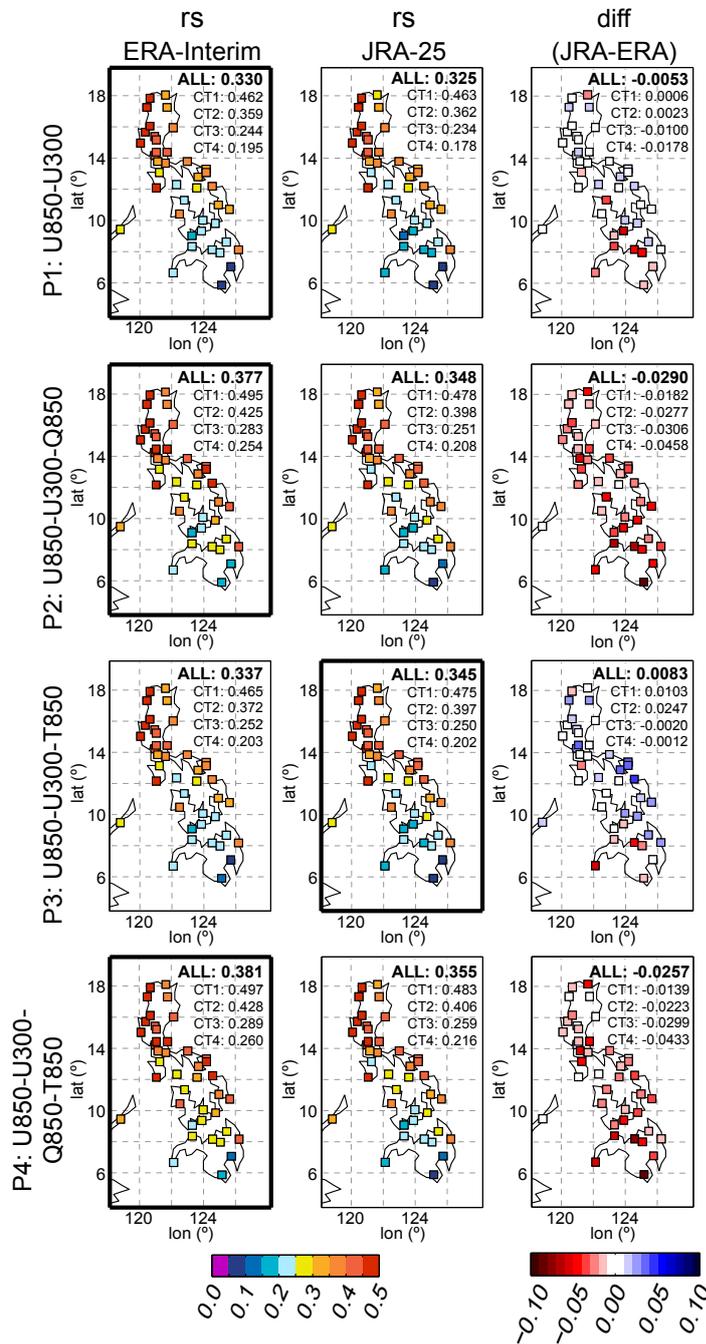


Figure 8.4: Spearman correlation coefficient between observed and downscaled daily precipitation amount for different predictor combinations (in rows), when considering predictor data from ERA-Interim (left column) and JRA-25 (middle column). Performance differences (JRA-25 minus ERA-Interim) are shown in the right column. For each specific predictor combination, black borders indicate the best performing reanalysis. The numbers in each panel indicate the spatial mean value for all the stations ('all') and for those stations pertaining to a specific CT ('CT1-4').

8.4 Sensitivity to Reanalysis Choice in Seasonal Forecasts

In order to assess the extent to which reanalysis uncertainty affects the results from SD of seasonal forecasts, the deterministic GLM used in the previous section was independently calibrated with ERA-Interim and JRA-25 predictor data using the period 1981-2005 and the regression coefficients learnt from both reanalyses were subsequently applied to predictor data from four of the ENSEMBLES models. As discussed in Section 6.2, this calibration was done separately for DJF, MAM, JJA and SON, considering season-specific data. Based on the results from Sections 6.1 and 8.3, where P4 (Table 6.1) was shown to provide the best performance in cross-validation mode, only the combination U850-U200-Q850-T850—a minor variation of P4 in which U300 is replaced by U200, the most similar variable available for the ENSEMBLES models—was considered. SD is carried out for each ensemble member individually, thus obtaining nine downscaled time-series for each model. Afterwards, these nine series are averaged to get a unique downscaled time-series per model.

Figure 8.6 shows the Spearman correlation coefficient between the daily downscaled predictions obtained when applying the regression coefficients calibrated with ERA-Interim and JRA-25 reanalyses to one-month lead predictor data (similar conclusions were obtained for four-month lead predictors; not shown) from the four ENSEMBLES models (in columns) for the period 1981-2005, by seasons (in rows). In all cases, the numbers in each map indicate the spatial mean Spearman correlation for all the stations ('all') and for those stations pertaining to a specific CT ('CT1-4').

In agreement with the north-south gradient of reanalysis uncertainty shown in Section 8.2 for Q850 and T850, smooth differences appear towards the south. However, the agreement between the downscaled results obtained when using ERA-Interim and JRA-25 predictor data for calibration is very high for all models, seasons and CTs (correlations around 0.9 in most of cases), which suggests that the issue of reanalysis uncertainty is not of special relevance in seasonal forecasting. Therefore, and given that ERA-Interim yields overall better results than JRA-25 in cross-validation mode (Section 8.3), only the former is considered for the calibration of the different PP methods applied in this Thesis.

For completeness, Figure 8.7 is the equivalent to Figure 8.6 but for a deterministic GLM considering the 15 leading PCs as predictors instead of standardized anomalies at nearby grid boxes. Results are overall very similar to those from Figure 8.6. However, the use of PCs is slightly less sensitive to reanalysis uncertainty than nearby grid boxes, especially for a few particular points in which the local character of standardized anomalies lead to larger differences.

8.5 Sensitivity to Reanalysis Choice in Climate Change Projections

Differently to the case of seasonal forecasting, we show in this section that local-scale climate projections obtained by means of PP methods are highly sensitive to the choice of reanalysis used for calibration. To this aim, the regression coefficients obtained from separately calibrating the deterministic GLM used in the previous sections (which considers as predictors the standardized anomalies at the 4 nearest grid boxes) with either ERA-Interim or JRA-25 are applied to predictor data from the MPI-ECHAM5, both for the reference period 1981-2000 (using control run data) and for three different future periods (2011-2040, 2041-2070 and 2071-2100), using scenario run data (A1B, run 3). The underlying assumption of this procedure is that the predictor-predictand relationships obtained in present climate conditions remain stationary in time (Vrac et al., 2007). To calculate the projected climate change signals we apply the ‘delta-method’, i.e., the reference/control period’s mean is subtracted from the mean of the particular target scenario period (Räisänen, 2007). Deltas are shown as relative (%) deviations from the mean in the reference period (0% = no deviation). For the sake of comparison with the results obtained in cross-validation mode (Section 8.3), we consider here all the predictor combinations listed in Table 6.1 (P1 to P4).

Figure 8.8 shows three panels, one for each of the future periods considered. In each panel, the deltas projected by applying the coefficients learnt from ERA-Interim (JRA-25) are shown in the left (middle) column, while the corresponding differences (JRA-25’s delta minus ERA-Interim’s delta) are given in the right column —each row corresponds to a particular predictor combination.— The numbers in each map indicate the spatial mean value for all the stations (‘all’) and for those stations pertaining to a specific CT (‘CT1-4’).

Independently from the reanalysis used for calibration, negligible deltas are found for any future period if precipitation is downscaled from circulation variables alone (first row). However, if Q850 and/or T850 are added to circulation (second, third and forth rows), increasing precipitation is projected across the entire country (with deltas increasing as a function of forecast-time). Moreover, the magnitude of the projected deltas seems to be related to the cross-validated results of Section 8.3. In particular, larger deltas are obtained when the ‘better’ performing reanalysis is considered for calibration, i.e., ERA-Interim (JRA-25) in case Q850 (T850) is added to circulation. This suggest that regional climate change projections obtained by means of PP methods might be not accurate if the reanalysis used for calibration is not able to properly capture the relationship between the predictors considered and the local target predictand.

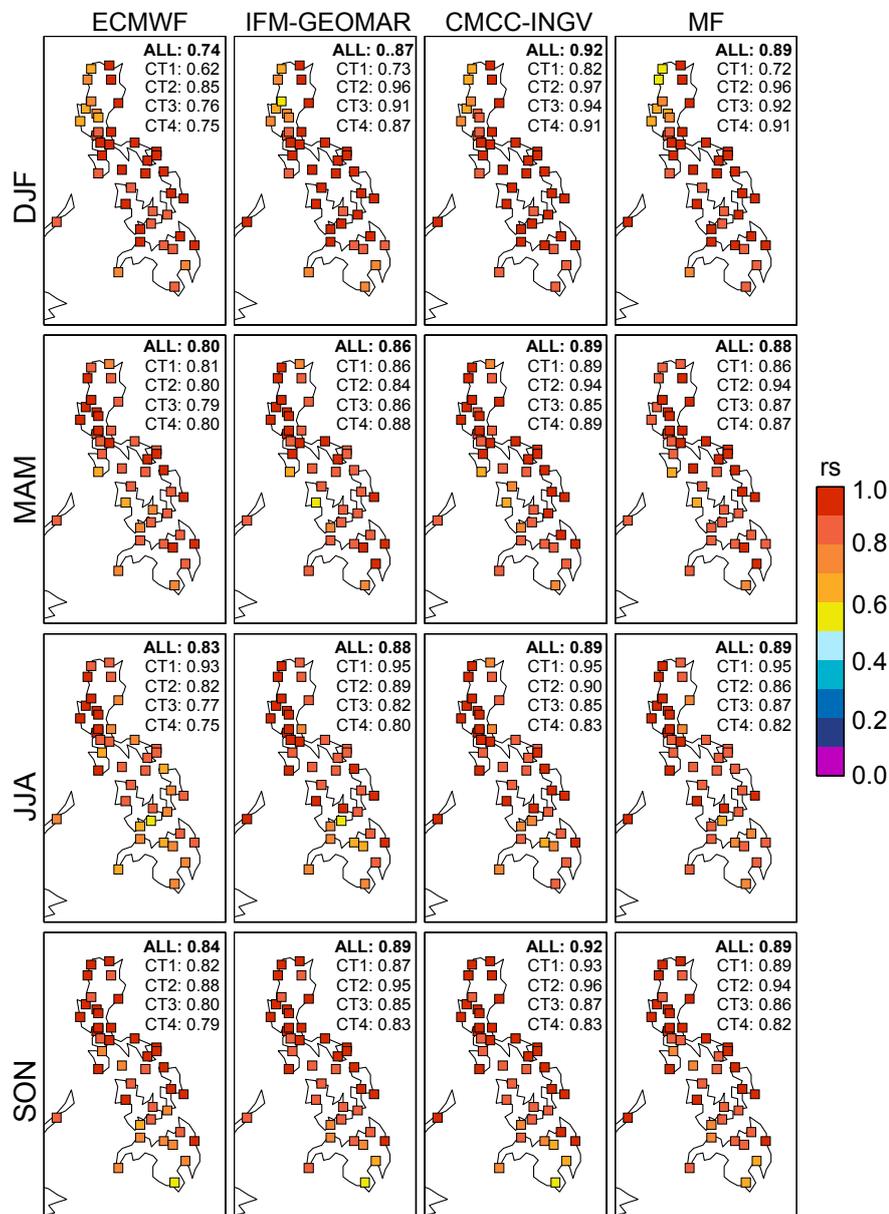


Figure 8.6: Spearman correlation between the daily downscaled predictions obtained when applying the regression coefficients calibrated with ERA-Interim and JRA-25 reanalyses to one-month lead predictor data from the four ENSEMBLES models (in columns) for the period 1981-2005, by seasons (in rows). A deterministic GLM using the standardized In all cases, the numbers in each map indicate the spatial mean value for all the stations ('all') and for those stations pertaining to a specific CT ('CT1-4').

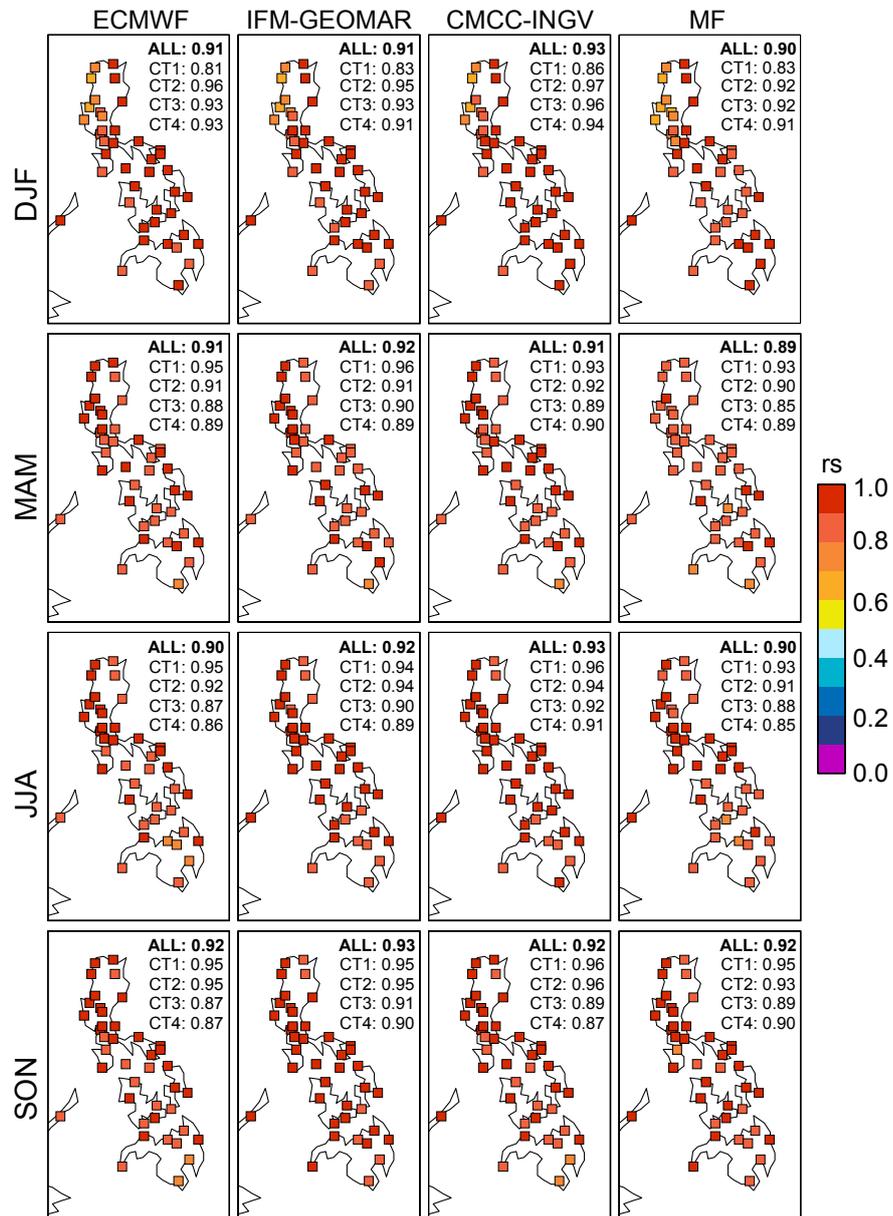


Figure 8.7: As Figure 8.6 but for a deterministic GLM considering the 15 leading PCs as predictors.

Importantly, the reanalysis-induced differences detected in the downscaled results in present climate conditions (see Figures 8.3 and 8.4) are considerably amplified, leading to ‘delta-change’ estimates which strongly differ depending on the reanalysis used for calibration. In particular, the downscaled results are especially sensitive to reanalysis choice when Q850 is included among the predictors. For instance, for the P2 combination, the reanalysis-induced delta differences reach 35% (45%) for the entire country (CT1) for the end of the century (2071-2100).

Finally, Figure 8.9 shows the mean value —as simulated by the MPI-ECHAM5 (A1B scenario, run 3)— for each of the predictor variables in Table 6.1 over the domain considered for SD (Figure 6.1) for the three future periods considered: 2011-2040, 2041-2070 and 2071-2100. This figure suggests that the reanalysis-induced differences in the downscaled projections are proportional to the climate change signal imposed by the GCM considered (compare Figure 8.8 to Figure 8.9). For instance, the negligible deltas found if circulation variables alone are considered as predictors would be in agreement with the time evolution of U850 and U300, which is virtually constant throughout the whole 21st century, indicating that the large-scale circulation —as simulated by the MPI-ECHAM5— over the target region is not sensitive to climate change.

In the light of the previous results, the choice of reanalysis used for calibration in SD is shown to be an important uncertainty source for the case of local-scale climate change projections —for which it should be treated with equal care as other, well-known, uncertainty sources such as the choice of GCM or downscaling method (Dibike and Coulibaly, 2005; Chen et al., 2012),— whereas it is not of special relevance in the context of seasonal forecasting —an issue which, to the author’s knowledge, had not been assessed before this Thesis.— Albeit these conclusions have been deduced for a specific region (the Philippines), they are likely to hold valid for the entire tropics since previous studies point out that reanalysis uncertainty is a general problem at low-latitudes —especially for climate change signal bearing predictor variables on the daily time-scale (Brands et al., 2012, 2013).—

Additionally, apart from being relevant for SD under the PP approach, reanalysis uncertainty is expected to be also relevant for recent MOS downscaling schemes operating on a daily time-scale, in which GCMs have been nudged to reanalysis data in order to force them to follow the ‘real’ large-scale variability (Eden et al., 2012). Here, it has been shown that the ‘real’ large-scale atmospheric variability in the tropics is likely to be misrepresented by reanalyses and, consequently, also by the aforementioned nudged GCMs. Likewise, since RCMs can be nested into either different reanalyses or bias corrected (using reanalysis data) GCMs, reanalysis uncertainty is also likely to affect the results from dynamical downscaling (Park et al., 2013).

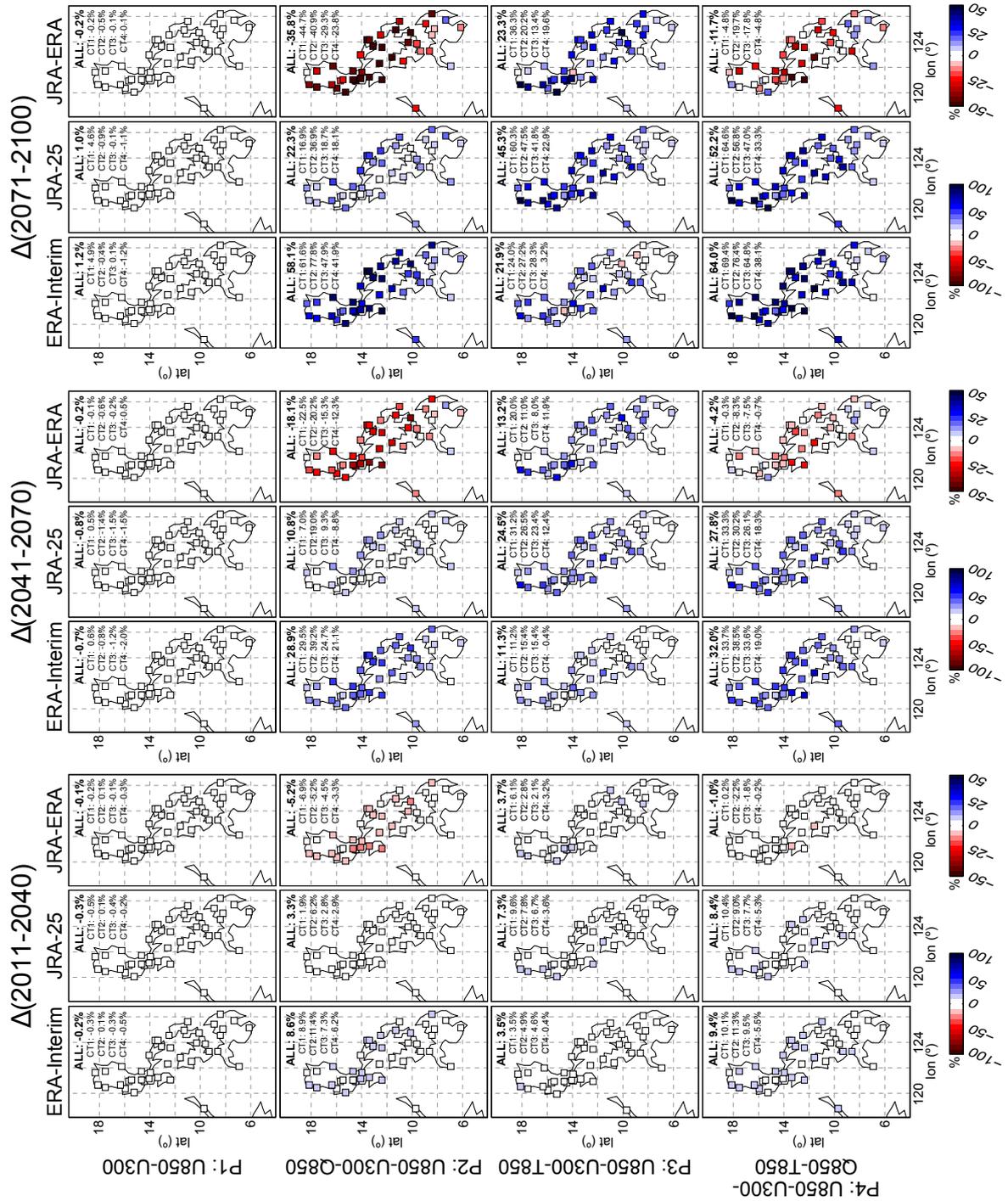


Figure 8.8: Precipitation deltas for three future periods: 2011–2040, 2041–2070 and 2071–2100 (all with respect to the control period 1981–2000). For each of these periods, the left (middle) column shows the deltas obtained from applying the regression coefficients learnt from ERA-Interim (JRA-25) to predictor data from the MPI-ECHAM5, whereas the right column displays the difference between the JRA-25’s delta and ERA-Interim’s delta. The numbers within each map indicate the spatial mean value for all the stations (‘all’) and for those stations pertaining to a specific CT (‘CT1-4’).

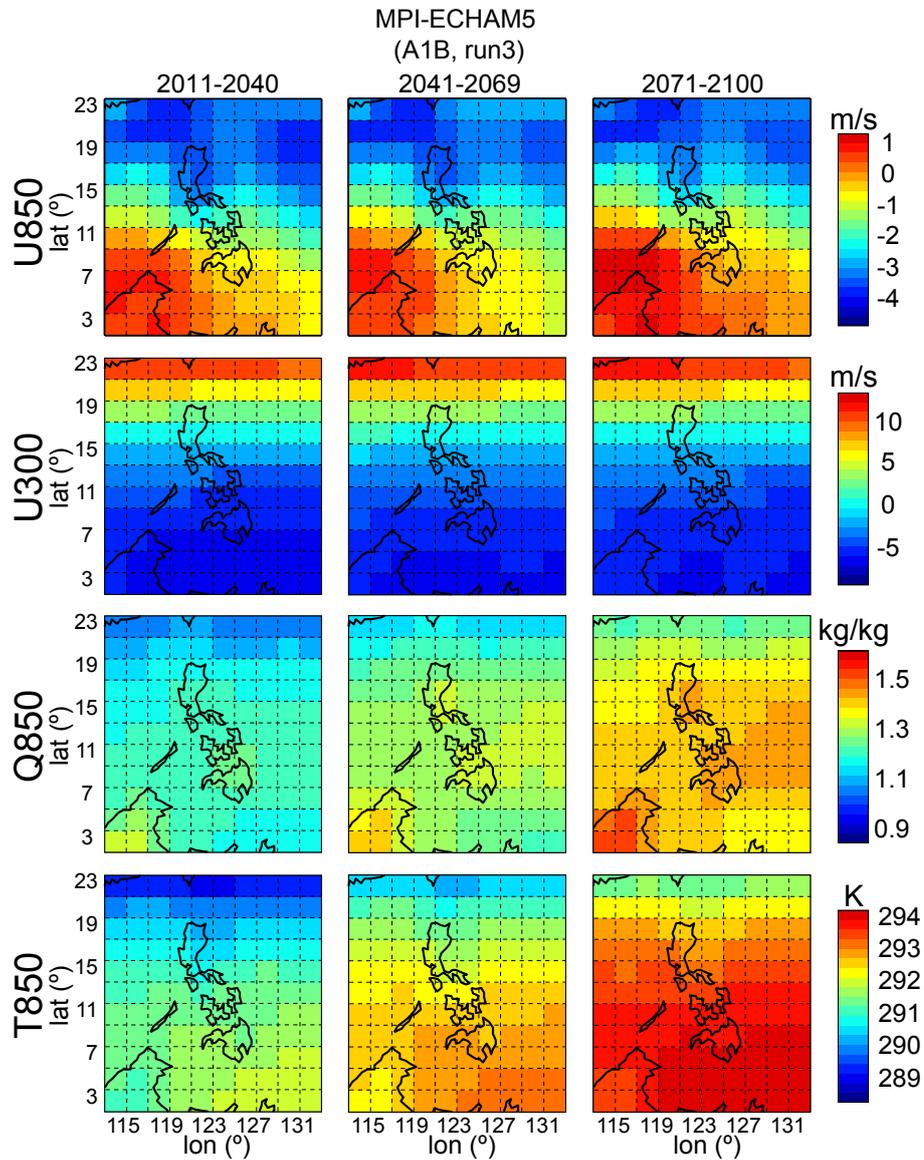


Figure 8.9: Mean value —as simulated by the MPI-ECHAM5 (A1B scenario, run 3)— for each of the predictor variables listed in Table 6.1 (in rows), for the three future periods considered (2011-2040, 2041-2070 and 2071-2100, in columns).

CHAPTER 9

SD Methods for Seasonal Forecasting: Advantages and Limitations

As explained in previous chapters, the coarse spatial resolution of seasonal GCMs and their systematic biases (see Section 3.4.1-2) hamper their direct application in local impact studies. To overcome this, SD is nowadays routinely used. Nevertheless, whereas the wide range of existing approaches and techniques (see Chapter 5) have been extensively used and critically assessed in climate change applications, their advantages and limitations for seasonal forecasting are not well understood yet (Feddersen and Andersen, 2005). Moreover, despite most of the skill for seasonal precipitation forecasts concentrates in the tropics (see Chapter 7), these techniques have been developed and applied almost exclusively for extratropical regions (Hewitson and Crane, 1996; Wilby and Wigley, 1997; Hanssen-Bauer et al., 2005; Fowler et al., 2007; Maraun et al., 2010; Gutiérrez et al., 2013). In particular, studies under the PP approach are rare to-date for low-latitudes because manifold problems still hinder its successful application in these regions (Hewitson et al., 2014). As a consequence, the MOS approach, and in particular MOS-BC methods — which do not require a predictor screening process and do not need reanalysis data for calibration (see Section 5.1.2)— have emerged as an alternative for SD purposes, even though several limiting problems have been identified (see, e.g., Ehret et al., 2012). Yet, a key problem in the context of seasonal forecasting is whether SD can serve to improve the skill of the raw model precipitation forecasts beyond reducing their systematic biases (see Section 3.4.1-2).

In this chapter, we analyze this matter for the Philippines, a challenging region affected by different large-scale phenomena where the use of SDMs may be especially relevant since

GCMs are not expected to capture the fine-scale forcings of mountains and land-sea contrasts which determine local rainfall (Robertson et al., 2012). To this, we apply two MOS-BC and two PP methods —representative of the different families introduced in Chapter 5— to four of the ENSEMBLES models (those listed in Table 6.1, with the exception of the UKMO model), comparing the downscaled results against the corresponding raw model precipitation at the 42 PAGASA stations for the period 1981-2005, focusing on accuracy and reliability aspects.

Notice that all the SDMs applied in this Thesis work on a daily basis. The reader interested in the application of seasonal MOS techniques for the Philippines is referred to the existing literature (Kang et al., 2007; Robertson et al., 2012).

9.1 *SD Methods Used*

As representative of the MOS-BC approach, we considered the two distributional q - q methods introduced in Section 5.4.2 (*QM_{par}* and *QM_{emp}*). These methods were calibrated and applied following the same 5-fold cross-validation introduced in Section 5.2, using as only predictor one-month lead precipitation from the ENSEMBLES models at the nearest grid box for each of the 42 gauges.

In addition, based on the results of Section 5.4, the *GLM_{det}* and the *AN_{det}* methods (belonging to the *transfer functions* and *weather typing* families, respectively) were considered as representative of the PP approach. For the *AN_{det}* method, only the closest analog was considered. Following from the screening process carried out in Section 8.3, a minor variation of P4 in which U300 is replaced by U200 (denoted as P4* hereafter), defined over the domain shown in Figure 6.1, was considered as predictor data for both methods. In order to minimize the effect of reanalysis uncertainty, the 30 leading PCs were considered (see Section 8.4). Building on the results from Sections 6.2 and 8.3, season-specific data from the ERA-Interim reanalysis were used for the calibration phase. For the prediction phase, predictor data from the ENSEMBLES models (one-month lead predictions were considered) were properly harmonized as explained in Section 6.3. For coherence, 1981-2005 was considered for both the calibration and the prediction phases.

The two MOS-BC and the two PP methods were separately applied to each of the nine available ensemble members for each model. In addition to the results obtained for each of the four individual models, a multimodel (MM) time-series was constructed by considering the 36 (4 models x 9 members) downscaled predictions, thus giving equal weights to all models and members. Note that, for the computation of probabilistic MM forecasts, precipitation categories (e.g. terciles) were computed independently for each model.

9.2 Results for Reanalysis Predictors

Solid (dashed) lines in Figure 9.1 (a composition of Figures 6.4 and 6.5) show the cross-validated results —as measured by the ACC between observed and downscaled interannual precipitation totals for the period 1981-2005— obtained when applying the *GLM_det* (*QM_emp*) method using ‘perfect’ (reanalysis) predictors under the 5-fold cross-validation framework introduced in Section 5.2. Note that the P4* combination defined over the domain shown in Figure 6.1 (precipitation at the nearest model grid box) from ERA-Interim was considered as predictor data for the *GLM_det* (*QM_emp*) method.

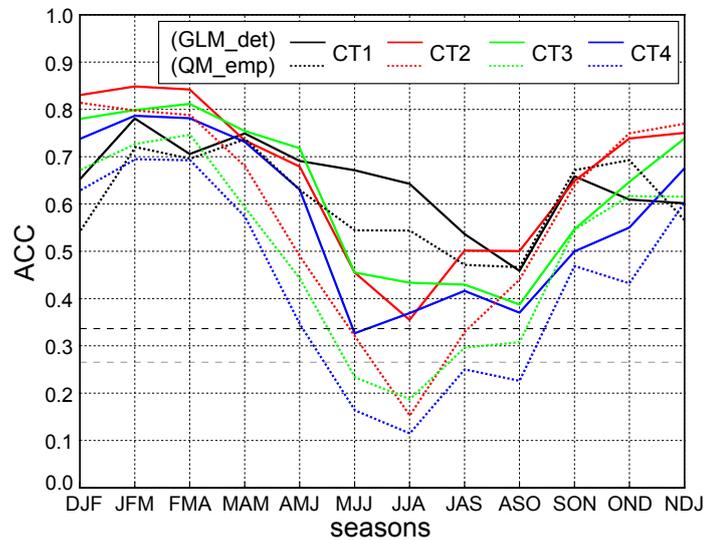


Figure 9.1: Composition of Figures 6.4 and 6.5: cross-validated results —as measured by the ACC between observed and downscaled interannual accumulated precipitation time-series for the period 1981-2005— when applying the *GLM_det* (*QM_emp*) method using ‘perfect’ predictors, i.e., considering large-scale variables (precipitation) from ERA-Interim as predictors, as represented by the solid (dashed) lines. For all 3-month seasons along the year (x -axis), each line shows the average correlation for the corresponding CT (see colors in the legend). Significant — $\alpha = 0.1$ ($\alpha = 0.05$)— values are those above the gray (black) dashed lines. A Student’s t -distribution with $N - 2$ degrees of freedom ($N =$ number of years) was considered to compute this significance.

Recall that all the SDMs applied in this Thesis work on a daily basis (the predictor-predictand statistical links are established based on daily data) so all of them have an intrinsic error to reproduce the daily observed time-series. This figure shows how this error is propagated to the relevant time-scale for seasonal forecasting, thus representing the upper-limit of local interannual seasonal predictability that can be obtained from SD in this region (i.e., the extent to which the SDMs preserve the observed interannual seasonal variability). The *GLM_det* method provides better results than the *QM_emp* for all CTs

and seasons, especially for JJA (note that the correlations obtained from the QM_{emp} in this season are not significant).

To give more insight on the results found, Figure 9.2 shows the observed interannual variability of spatial average precipitation totals for each CT (in colors) for the period 1981-2005, by seasons. The poor local predictability obtained from SD for JJA (Figure 9.1) may be related to the low year-to-year observed variability in this season (with the exception of CT1), since the negative impact of the day-to-day error from the SDMs is in turn amplified. At the same time, as explained in Section 7.5.3, this low year-to-year observed variability is associated with the weaker effect of ENSO in this region for this season (note that the ENSO phenomenon amplifies the local interannual seasonal climate variations in teleconnected regions).

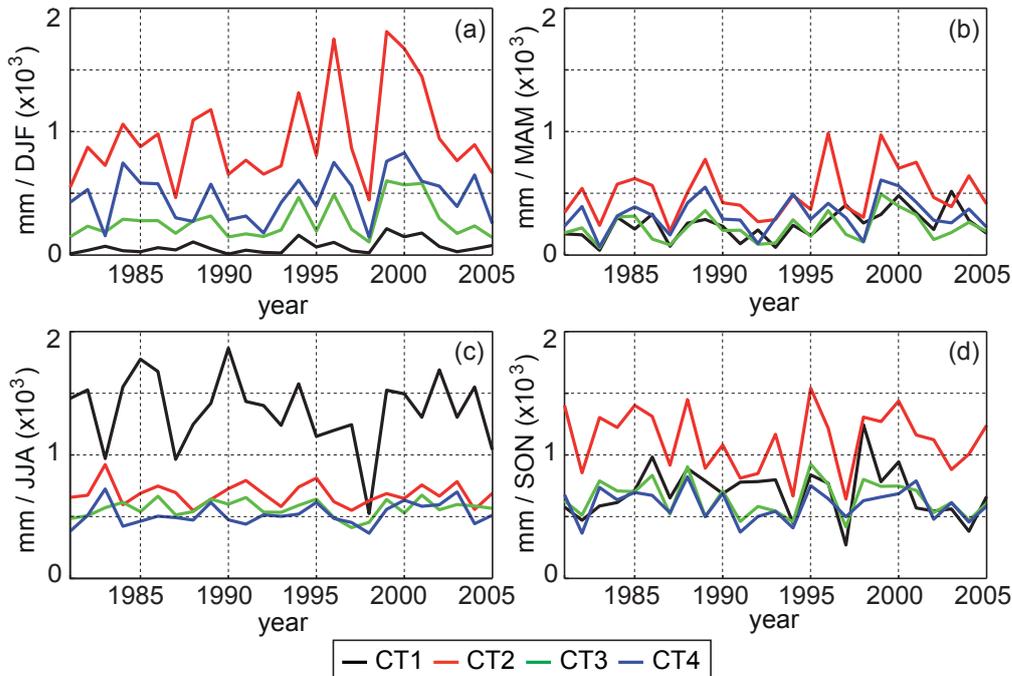


Figure 9.2: Interannual variability of spatial average precipitation totals for each CT (see colors in the legend) for the period 1981-2005, by seasons.

Although the pattern found in Figure 9.1 may change when substituting the ‘perfect’ (reanalysis) by ‘non-perfect’ (GCM) predictors—in particular, the local predictability from SD is expected to diminish,— these results suggest that PP methods might provide better results than simpler and more pragmatic MOS-BC ones, especially for JJA. The potential usefulness of PP methods for this season might be explained by situations in which the large-scale predictor variables are better simulated by the model than the target precipitation. We will further explore this hypothesis in the following sections.

9.3 Results for GCM Predictors

In order to validate the skill of the raw seasonal precipitation forecasts from the ENSEMBLES models and the (possible) added value of the corresponding downscaled results, we considered the following metrics: ACC —as a simple measure of accuracy for the deterministic ensemble mean forecasts— and a measure of reliability based on the different categories introduced in Section 3.5.1 for tercile-based (T1: dry, T2: normal, T3: wet) probabilistic forecasts. Note that both ACC and reliability were calculated upon the interannual anomalies of seasonal precipitation totals for the period 1981-2005, which were derived from the daily time-series. Note also that, as opposite to mean squared skill scores, the two validation metrics considered are not sensitive to biases in the mean and the amplitude of the predictions. Therefore, we assess the relevant (temporal) aspects which can provide added value for seasonal forecasting and refer the reader to some comprehensive validation experiments for further information on the performance of the different SDMs from the point of view of biases and marginal statistics (Maraun et al., 2015).

9.3.1 Accuracy

For the different seasons (in rows) and CTs (in columns), panels in Figure 9.3 show the interannual ACC values obtained for each of the ENSEMBLES models (see the colors in the legend). Boxplots display the results along the 42 stations for the Direct Model Output (DMO; indicated by a light gray shadow) at the nearest model grid box and the *QM-par*, *QM-emp*, *GLM-det* and *AN-det* methods (denoted hereafter as BC1, BC2, PP1 and PP2, respectively).

Overall, results vary mainly among seasons, but also among CTs, models and downscaling methods. For the latter, results are more sensitive to the approach considered (MOS-BC or PP) than to the particular SDM used within each approach. As already shown in Figure 7.6, the highest scores for the DMO are obtained for DJF and MAM whereas the worst results are found for JJA, with no significant correlations for any model except for the ECMWF in the CT1 region. In general, neither MOS-BC nor PP methods yield relevant accuracy improvements (with respect to the DMO) for DJF and MAM —the most skillful seasons (see Figures 4.5 and 7.6),— which suggests that the added value that can be obtained by means of SD is limited in those cases when the models properly simulate precipitation. However, whereas MOS-BC methods do not clearly improve (or even worsen) the accuracy of the DMO in JJA and SON, PP methods provide in general better (worse) results than the DMO in the former (latter) season. In particular, notice that PP methods yield large accuracy improvements in JJA for the stations pertaining to CT1 for all models (with the exception of the ECMWF), which exhibit nearly-zero ACC values in this season.

Besides this general good (bad) performance in JJA (SON), there are a few interesting cases for DJF and MAM in which PP methods can add value to the DMO.

First, it is found that PP methods can improve raw precipitation from the relatively bad performing models—those exhibiting small ACC values, as compared to the rest of models—in these seasons. This is the case for the MF model in DJF (CT4) and the IFM-GEOMAR model in MAM (CT1), indicated by a black dashed border in Figure 9.3. Second, there are also cases in which PP methods can still improve the accuracy of raw model precipitation despite its good performance, as occurs for the MF model in DJF (CT1), the ECMWF model in MAM (CT1) and the IFM-GEOMAR model in MAM (CT2), marked with a blue dashed border in Figure 9.3.

Third, PP methods can add important local value for particular outlier stations (i.e., stations where the accuracy of raw model precipitation drops, as compared with the rest of locations). This occurs for the MF model in DJF (CT3) and the CMCC-INGV model in MAM (CT2 and CT3), indicated by a green dashed border in Figure 9.3. Notice that, as opposite to the DMO and the MOS-BC methods—which depend on model precipitation at the nearest grid box and can be affected by local features such as wrong orographical gradients, land-sea interfaces, etc.—PP methods rely on large-scale predictors to infer local precipitation and allow thus to properly reproduce the observed interannual variability in these cases.

As mentioned, the results obtained are more sensitive to the approach considered (MOS-BC or PP) than to the particular SDM used within each approach. Therefore, in order to gain insight on the advantages and limitations of the two approaches whilst keeping the analysis simple, we considered the ensemble mean of the two SDMs applied for each approach. In particular, the downscaled predictions from BC1 and BC2 (PP1 and PP2) were averaged into a single time-series—denoted as BCens (PPens)—which was compared against the DMO. For each season (in rows) and model (in columns), panels in Figure 9.4 show the percentage of stations exhibiting significant ($\alpha = 0.1$) ACC values for the DMO and the BCens (yellow and blue bars, respectively) in the different CTs. Moreover, within each yellow (blue) bar, a thinner black bar indicates the percentage of stations in which the ACC obtained for the DMO (BCens) outperforms in a 25% the ACC obtained for the BCens (DMO). Figure 9.5 is the equivalent to Figure 9.4 but for the PPens instead of the BCens (note that yellow bars are the same in both figures). These two figures provide a good summary of the results presented in Figure 9.3.

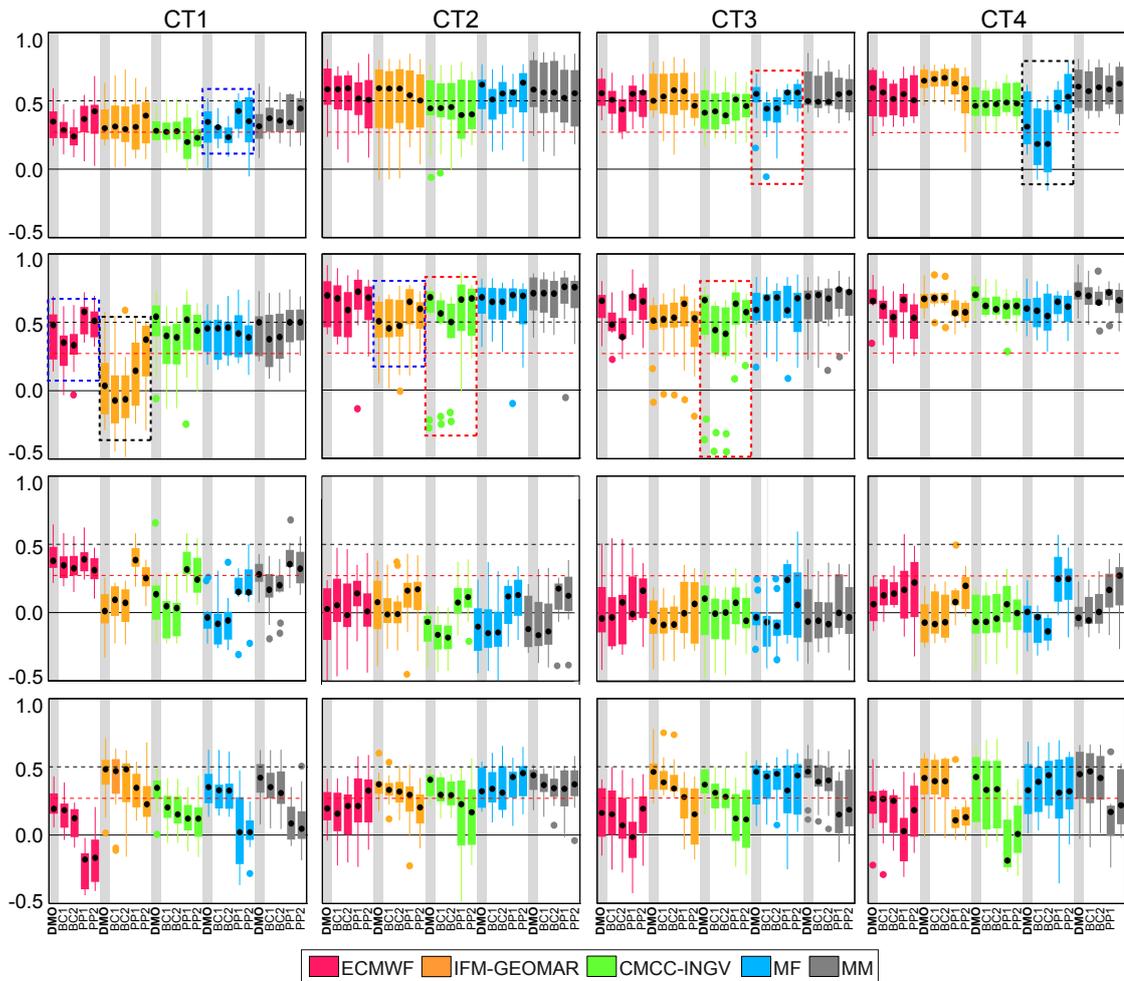


Figure 9.3: Interannual ACC obtained for each season (in rows) and CT (in columns). In each panel, results from each model are shown in different colors (see the legend). From left to right, boxplots display the correlations obtained along the 42 PAGASA stations for the DMO (indicated by a light gray shadow) and the BC1: QM_{par} , BC2: QM_{emp} , PP1: GLM_{det} and PP2: AN_{det} methods. Significant ($\alpha = 0.1$) values are those above the red dashed lines. A Student's t-distribution with $N - 2$ degrees of freedom ($N =$ number of years) was considered to compute this significance.

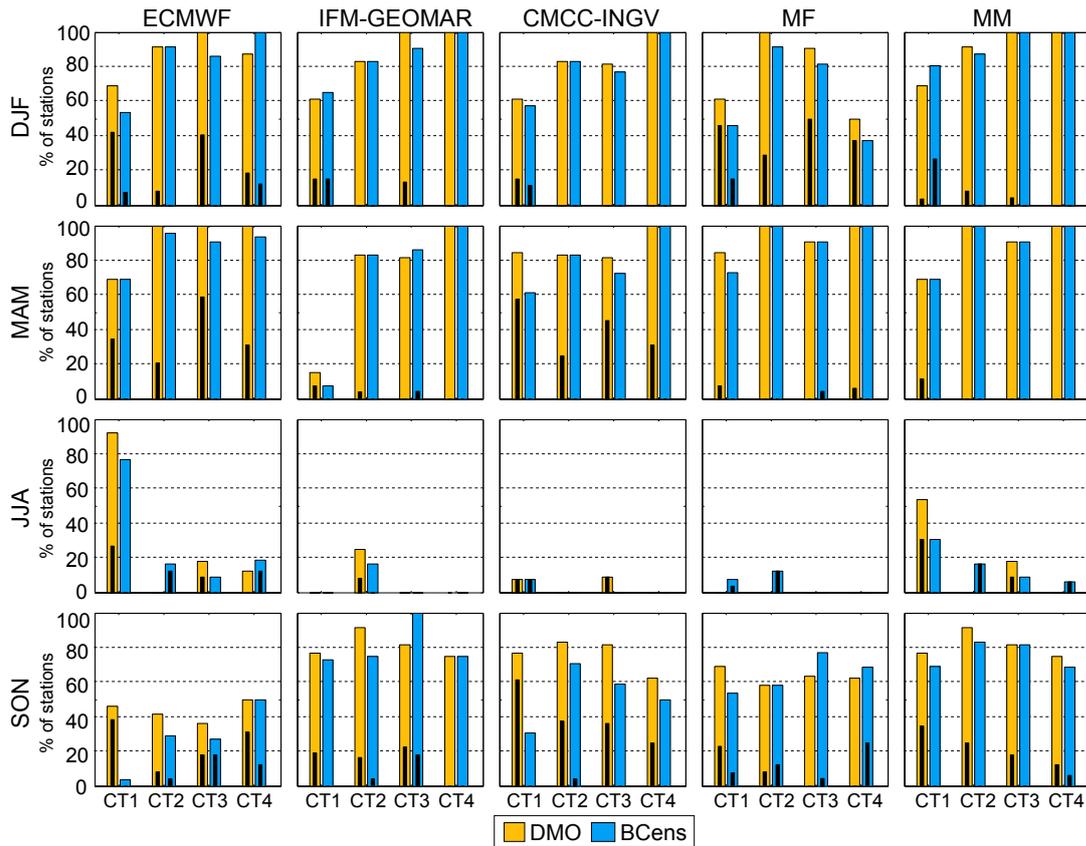


Figure 9.4: For each season (in rows) and model (in columns), panels show the percentage of stations exhibiting significant ($\alpha = 0.1$) interannual ACC values for the DMO and the BCens (yellow and blue bars, respectively) in the different CTs. Within each yellow (blue) bar, a thinner black bar indicates the percentage of stations in which the correlation obtained for the DMO (BCens) outperforms in a 25% the value obtained for the BCens (DMO).

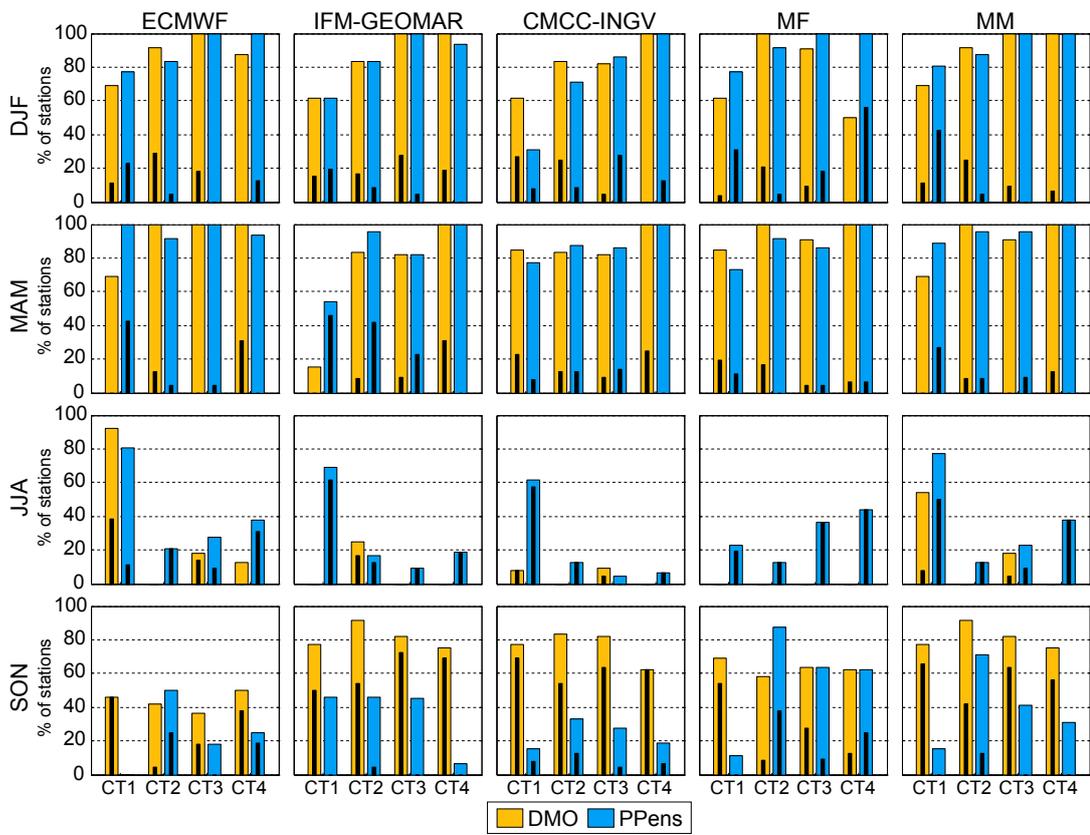


Figure 9.5: As Figure 9.4 but for PPens instead of BCens.

Figure 9.4 shows that the DMO outperforms in general the BCens. Moreover, results for the latter are limited by the quality of the model: note that the best (worst) accuracy for the BCens is found for DJF and MAM (JJA), the seasons with the (highest) lowest precipitation skill (Figures 4.5 and 7.6). Note also that the gain in accuracy found for the BCens in some cases is limited to a few stations and is counteracted by the loss found in others. As a result, no robust signal of added value is obtained for the MOS-BC approach (see the corresponding boxplots in Figure 9.3).

Nonetheless, Figure 9.5 evidences that the PPens can either improve or spoil the accuracy of the DMO, depending on the case. In particular, the above mentioned cases in which PP methods can add value to the DMO in DJF and MAM are reflected in this figure. Also, the accuracy improvements found in CT1 for JJA for all models (with the exception of the ECMWF) and the worsening obtained in SON are clear from this figure. To further assess the results for this interesting case (which will be later analyzed with care) at a local level, the top (bottom) row of Figure 9.6 shows the interannual ACC differences between the DMO —taken as reference— and BC1/BC2/BCens (columns 1/2/3) and PP1/PP2/PPens (columns 4/5/6) for the MM at the 13 stations pertaining to CT1, for JJA (SON). Notice from the large differences obtained at some locations that the choice of an appropriate SD approach can have important implications for local applications.

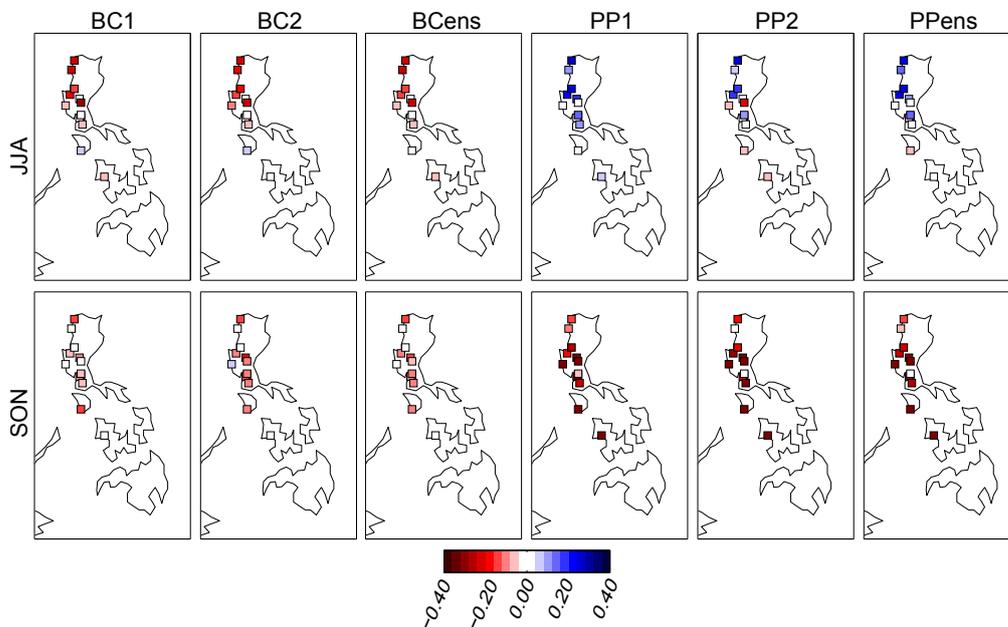


Figure 9.6: Interannual ACC differences between the DMO —taken as reference— and BC1/BC2/BCens (columns 1/2/3) and PP1/PP2/PPens (columns 4/5/6) for the MM in the 13 stations pertaining to the CT1, for JJA and SON (top and bottom row).

9.3.2 Reliability

The methodology proposed by Weisheimer and Palmer (2014) is used here to assess the reliability of the tercile-based probabilistic forecasts obtained from the DMO and the different SDMs applied. As explained in Section 3.5.1, this methodology allows to differentiate between five reliability categories —*perfect* (green), *still useful* (blue), *marginally useful* (yellow), *not useful* (orange) and *dangerous* (red)— based on the relative position of the weighted reliability line with respect to the perfect reliability (diagonal), no-skill and no-resolution lines, as well as on the uncertainty range around it (as obtained from bootstrapping). Recall from Section 3.5.1 that a 50% of the total range is used here — instead of the 75% used in Weisheimer and Palmer (2014)— since it is more suitable for the ensemble size of the ENSEMBLES models considered for this Thesis. Additionally, within the *marginally useful* (yellow) category, we differentiate those cases in which the reliability line falls within the skill region but the uncertainty range around it is not fully contained (they are identified with the dark yellow color here). These two slight adaptations were found to provide further insight for regional analysis such as the one undertaken here —the original classification was developed for the 21 global land-areas defined in Giorgi and Francisco (2000).— For illustrative purposes, Figure 9.7 shows examples of this classification for some raw model precipitation forecasts and the corresponding MOS-BC and PP downscaled values.

Figure 9.8 shows the reliability categories (in colors) obtained from applying the above described methodology for the different models (in columns), and the different seasons and CTs (in rows). Each block shows the results for the DMO, the two MOS-BC (BC1 and BC2) and the two PP (PP1 and PP2) methods considered, for the three terciles. Overall, this figure is in good agreement with the results found for accuracy in the previous section, with the best reliability obtained in DJF and MAM and the worst in JJA. Moreover, the results for the two MOS-BC methods are very similar to those for the DMO, with slight differences due to spurious changes of category (as illustrated in the top row of Figure 9.7). However, the two PP methods exhibit major reliability differences with respect to the DMO, especially for JJA and SON. In particular, both PP1 and PP2 improve the results of the DMO in the former season, especially for the CT1, where *marginally useful* categories are obtained instead of *not useful* and *dangerous* ones. Nevertheless, the opposite situation is found for SON. Additionally, in agreement with the conclusions from Chapter 7, the best results are obtained for the extreme terciles (in contrast with the normal one) and for the MM, which generally outperforms any single model.

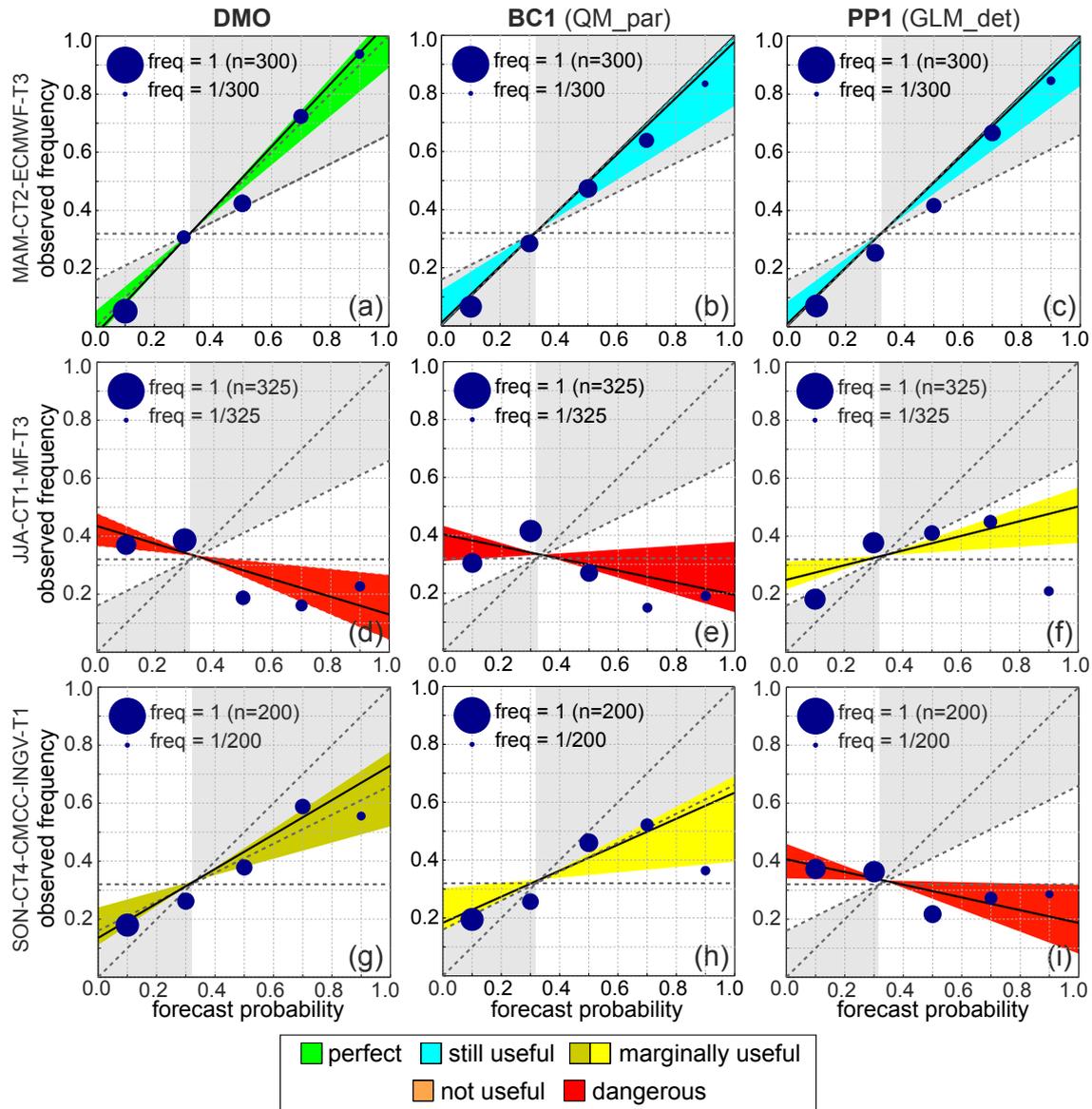


Figure 9.7: Reliability diagrams for the DMO, the BC1 and the PP1 method (in columns), for three different illustrative examples of seasonal forecasts in MAM, JJA and SON (in rows), for different CTs and models (see the labels on the left-hand side). The gray area defines the region contributing positively to the BSS. Colors correspond to the six categories used, which are based on the original scale proposed by Weisheimer and Palmer (2014) (see the text for details). Note that the diagrams are calculated considering the joined series of predictions of the different stations falling within each CT. Numbers in the upper left corner shows the sample size used in each case (the effective sample size would be lower due to the spatial dependence between the stations).

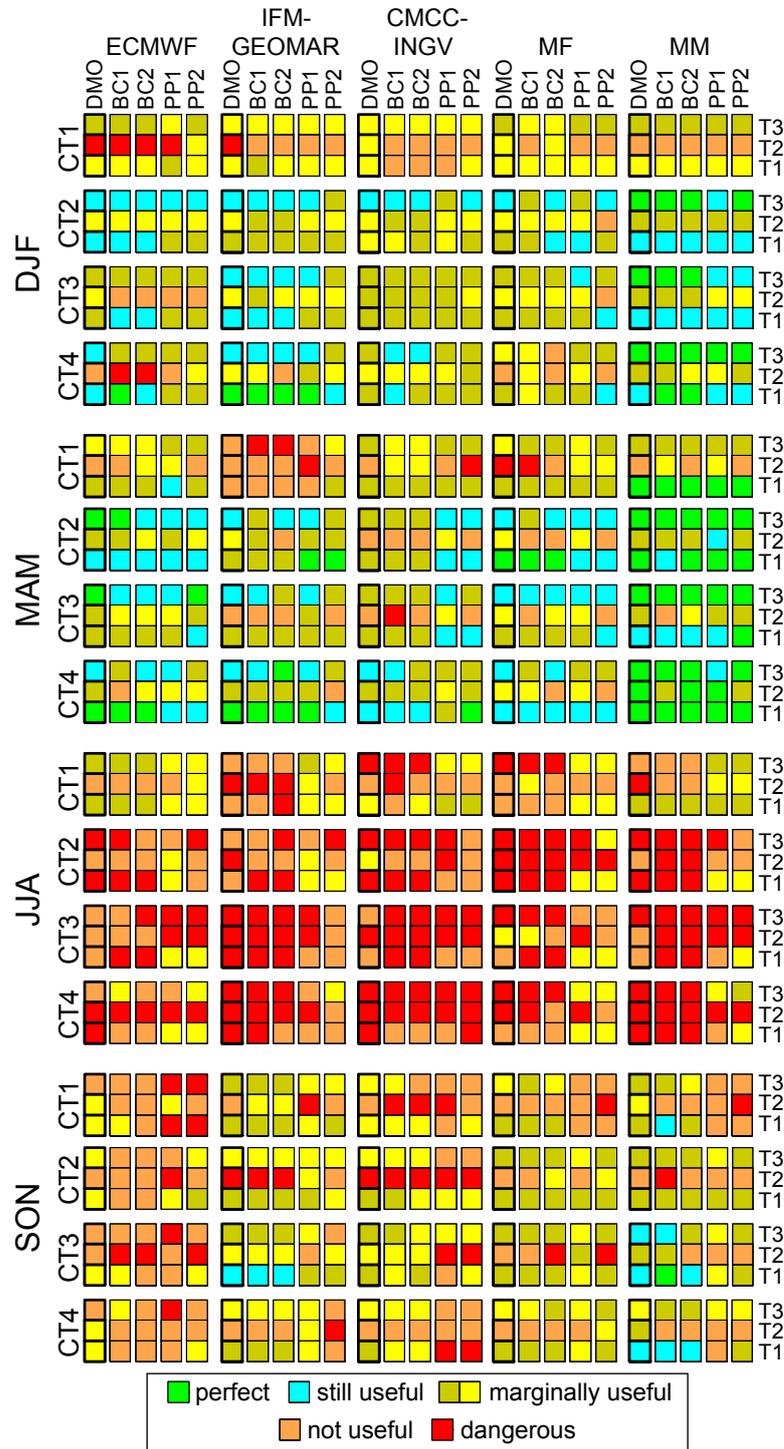


Figure 9.8: Reliability categories obtained for the different ENSEMBLES models (in columns) along the different seasons and CTs (in rows). Each block shows the results obtained for the DMO, the two MOS-BC (BC1 and BC2) and the two PP (PP1 and PP2) methods considered, for the three terciles. Colors correspond to the six categories used, which are based on the original scale proposed by Weisheimer and Palmer (2014) (see the text for details).

In order to summarize the results from Figure 9.8 and to better quantify the added value of the different approaches for SD, Figure 9.9 shows in stacked bar charts the percentage of reliability categories obtained from the DMO and the different downscaling methods for the different seasons (panels *a-d*) and CTs (panels *e-h*). Note that, for each approach (MOS-BC and PP), a joined analysis containing the two methods considered (BC1–BC2 and PP1–PP2) is performed. For clarity, the results from the MM and from the normal tercile are excluded from this analysis.

This figure shows that MOS-BC methods do not provide clear added value (or even worsen the DMO), neither for a particular season nor for a particular CT. However, PP methods yield substantial added value for JJA, leading to *marginally useful* categories in over 50% of the cases, as compared to less than 10% for the DMO (and for MOS-BC methods). Note that the opposite situation is found for the PP methods in SON, with *not useful* or *dangerous* categories obtained in nearly 50% of the cases (as compared with 10% for the DMO). Regarding the different CTs, PP methods exhibit a reduced fraction of useful results in CT4.

Remarkably, the good alignment between the results found for reliability in this section and those found for accuracy in the previous one points out the suitability and usefulness of the methodology proposed by Weisheimer and Palmer (2014) —slightly modified here— for regional studies (beyond its use for global areas).

9.4 The Added Value of PP Methods

As explained in Chapter 5, PP methods rely on large-scale predictors to infer local precipitation. As such, the above presented cases leading to a gain (loss) of skill for the PP approach could be explained by situations where large-scale variables, defined over a synoptic domain, are better (worse) predicted by the model than precipitation, which is more affected by particular local features. In order to check this premise, we focus on the 13 stations pertaining to the CT1, where PP methods were shown to improve (deteriorate) in general the skill of the DMO in JJA (SON). Figure 9.10 displays the ACC values between observed interannual precipitation at the 13 stations and the corresponding ERA-Interim and ENSEMBLES models outputs —the nearest grid box is considered— for precipitation (PR) and the different predictors used (U850, U200, Q850, T850) for the period 1981–2005.

The gain of skill found in JJA for all models except the ECWMF (Figures 9.3 and 9.8) is in agreement with the results found in the top panel, which shows significant ACC values (similar to the benchmark provided by ERA-Interim) for precipitation for this model. However, when analyzing the different predictors used by the PP methods

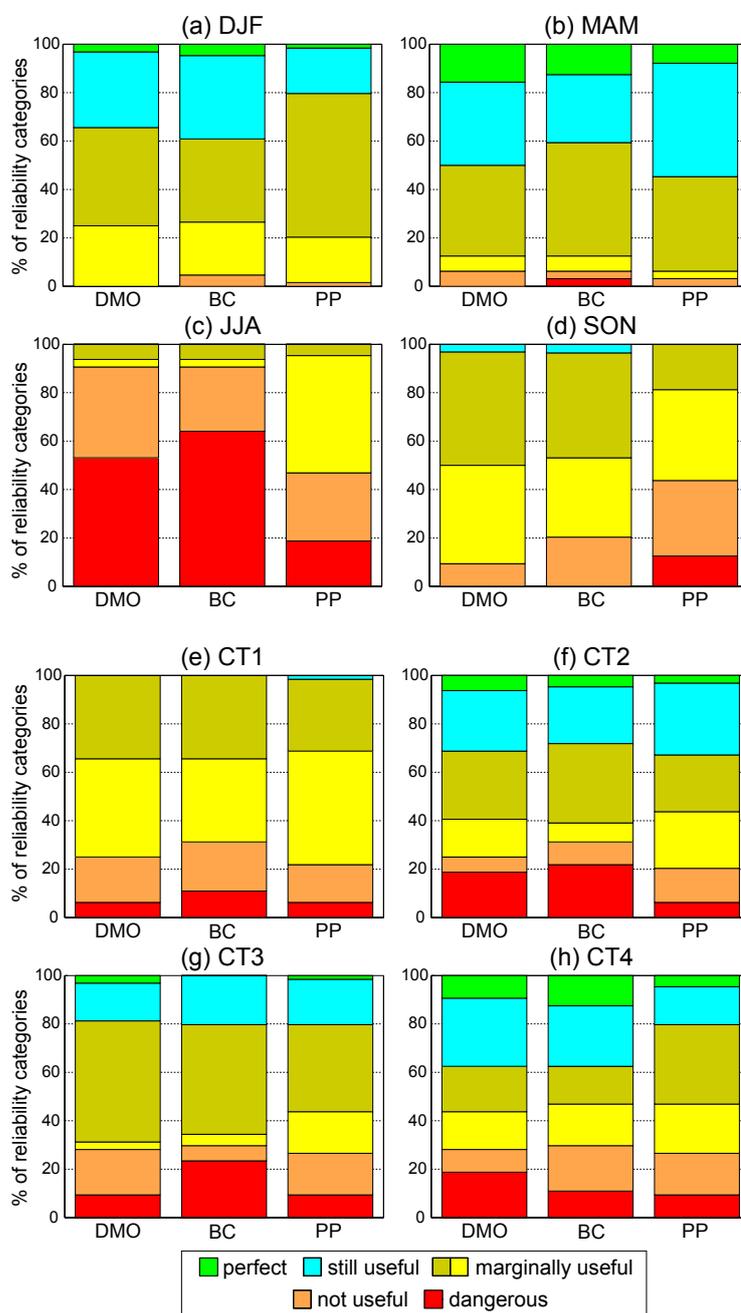


Figure 9.9: Stacked bar charts with the percentage of reliability categories (in colors) for the DMO and the MOS-BC and PP approaches (within each approach, the two methods considered are jointly analyzed) for different (a-d) seasons and (e-h) CTs. For clarity, results from the MM and from the normal tercile are excluded from this analysis.

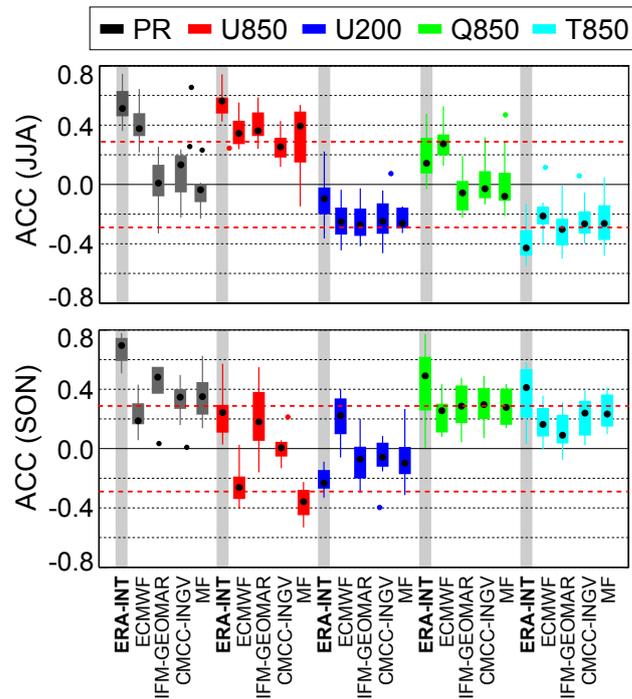


Figure 9.10: ACC values between observed interannual precipitation at the 13 stations pertaining to CT1 and the corresponding ERA-Interim and ENSEMBLES models outputs—the nearest grid box is considered—for precipitation (PR) and the different predictors used (U850, U200, Q850, T850) for (top) JJA and (bottom) SON over the period 1981–2005. Significant $-\alpha = 0.1$ —positive (negative) values are those above (below) the upper (lower) red dashed line.

—in particular U850 and T850, the most correlated ones with observed precipitation, as given by ERA-Interim—the results are similar for all models and are in agreement with ERA-Interim, indicating a good model performance for these variables. This suggests that PP methods might be able to exploit the models’ ability for reproducing the large-scale predictor variables to indirectly obtain more skillful local precipitation forecasts (as compared to the DMO).

The opposite situation is found for SON (bottom panel). In this case, the ACC values found for precipitation are significant (although smaller than the benchmark provided by ERA-Interim) in most cases. However, the results for the large-scale predictors are not significant in general. Moreover, opposite correlations with observations (as compared to ERA-Interim) are found in some cases. The joint effect of these errors might lead to wrong downscaled predictions—as occurs for the ECMWF model—exhibiting negative ACC values with observed interannual precipitation (see Figure 9.3) and *dangerous* reliabilities (extreme terciles in Figure 9.8).

Finally, the negligible added value found for the MOS-BC methods might be consequence of their lack of a physical basis (Haerter et al., 2011), which often spoils the merits of the different GCMs by altering their spatio-temporal field consistency (Ehret et al., 2012). Notice that the MOS-BC are limited since the temporal structure is still determined by the dynamics represented in the model grid box and do not describe local phenomena (Hempel et al., 2013).

Part IV

Concluding Remarks

CHAPTER 10

Conclusions, Achievements and Future Work

10.1 *Main Conclusions*

This section aims to summarize what has been done in order to achieve the three main objectives of the Thesis (which were identified in Section 2.2) as well as to briefly expose the most important conclusions which have been obtained in relation to them.

1. Regarding the first objective, the skill of global seasonal precipitation forecasts was assessed worldwide (grid box by grid box) for the forty-year period 1961-2000 in Chapter 7. To this, we considered the ENSEMBLES models (see Section 3.4.1) and applied a tercile-based probabilistic approach in terms of the ROC Skill Score (ROCSS). Although predictability varies with the region, season and lead-time considered, results indicate that 1) significant skill is mainly located in the tropics —20 to 40% of the total land areas,— 2) overall, SON (MAM) is the most (least) skillful season, and 3) the skill weakens (with respect to the one-month lead case) at four-month lead —especially in JJA,— although the ROCSS spatial patterns are broadly preserved —particularly in northern South America and the Malay archipelago.— Additionally, since ENSO is known to be the major driver of seasonal predictability (see, e.g., Goddard and Dilley, 2005; Doblas-Reyes et al., 2010), the role that this phenomenon plays on the skill was also analyzed. Results from a conditioned —restricted to El Niño and La Niña events— validation (also in terms of the ROCSS) and a study of El Niño and La Niña teleconnections with precipitation (see also Sections 3.2 and 3.4.1-1) suggested that the seasonality and spatial distribution of the skill found may be not only determined by the direct effect of ENSO —and therefore by the skill of the different models to predict the SST in El Niño 3.4 region,— but

rather by its indirect effect through its associated El Niño and La Niña teleconnections —and consequently limited by the models’ ability to simulate the observed teleconnection patterns.—

Moreover, following the same methodologies used for the worldwide study, a regional probabilistic validation and a regional study of ENSO teleconnections were carried out over the Philippines, considering in this case observed precipitation at the 42 PAGASA stations (see Section 4.1) over the period 1981-2005. In agreement with the results obtained for the deterministic validation presented in Section 4.2, the lowest skill in this region is found for JJA, the season exhibiting the weakest ENSO teleconnections.

2. With respect to the second objective, Chapter 5 provides a comprehensive description of the different approaches and techniques available for SD, as well as a detailed analysis of the SD methods which are considered throughout the Thesis. In order to optimize the configuration chosen for each technique, as well as for intercomparison purposes, the different techniques were applied in ‘perfect’ conditions (i.e., considering predictors from reanalysis for both the calibration and the prediction phase) over four illustrative PAGASA stations —representative of the four different climate types present in the Philippines (see Section 4.1),— under a cross-validation framework for the period 1981-2005 (see the chapter for details). These were the most relevant conclusions obtained:

- As representative of the PP approach (see the chapter for details), we considered the analog technique and Generalized Linear Models (GLMs). For the former, a deterministic version considering only the closest analog and a stochastic one in which the prediction is given by random selection among the observations corresponding to the 15 closest analogs were considered. Both SDMs perform alike in terms of distributional similarity. In particular, both predict less rainy days than observed (which in turn leads to a little dry bias) and underestimate the variance to a small extent. However, the deterministic version yields slightly better correlations than the stochastic does, and is therefore the only used in this Thesis.

Regarding the GLMs, we considered a deterministic configuration in which predictions are obtained from the estimated expected values and a stochastic one in which a simulation procedure is introduced in both *GLMo* and *GLMa* (see Section 5.4.1-2 for details). The deterministic version is not able to predict light precipitation amounts. Furthermore, it greatly underestimates the observed variance —with most of the predicted rain values in a small range,—

which leads to a wet bias (over a 10%) and a very low distributional similarity. Differently, bias is centred around zero and the predicted variance matches better the observed one in the stochastic version —simulating allows to predict the full range of observed precipitation values,— greatly improving the distributional similarity. Nevertheless, correlations strongly decay in the stochastic version as an effect of the simulation introduced, which implies a reduction of the predictive capacity of the method. Based on their better performance in terms of distributional consistency, stochastic GLMs are needed for SD of climate change projections. However, in seasonal forecasting it is key to assess the accuracy of the predictions and, hence, it is important to keep the deterministic signal isolated from the stochastic one. Therefore, the GLMs used in this Thesis are deterministic, i.e., predictions are based on the expected values. As compared to the analog methods, the correlations exhibited by the deterministic GLM are clearly higher.

- With regard to the MOS-BC approach (see the chapter for details), two common distributional methods (one parametric and one empirical) were considered. Both methods allow to improve the bad distributional similarity exhibited by ERA-Interim precipitation (being the empirical one better). However, neither the parametric nor the empirical version improves the correlations attained by the reanalysis; on the contrary, they deteriorate them. In addition, both perform similarly in reproducing the occurrence event and also in terms of correlations, yielding better results than analogs do and similar ones to those obtained for the deterministic GLM (with the exception of interannual correlation, which is better in the latter). In terms of distributions, the bias and the predicted variance (especially the latter) are better in the empirical method. Whilst the clear overestimation of the variance that occurs for the parametric method could indicate an intrinsic limitation to simulate extreme values (this method assumes that both simulated and observed precipitation fit to a gamma distribution, which might not be true), the good performance of the empirical one might be related to certain over-fitting.

Also in relation to the second objective, Chapter 6 assesses, focusing on the Philippines, a number of aspects which are relevant for the SD of seasonal forecasts. As derived from the conclusions drawn below, it establishes the methodological framework under which the different SDMs considered in this Thesis are applied:

- By applying a PP method in ‘perfect’ conditions for the 42 PAGASA stations (under a cross-validation framework for the period 1981-2005), we found that the combination of circulation (zonal wind velocity at different vertical levels) and thermodynamic (humidity and temperature) variables over a domain covering the country (Figure 6.1) yielded the highest predictive potential, being the latter lower for the case of using circulation variables alone. Therefore, the predictor combination P4 (see Table 6.1) is considered for all the PP methods applied in this Thesis.
- Under a cross-validation framework for the period 1981-2005, we applied in ‘perfect’ conditions two different configurations of a PP method based on GLMs and two MOS-BC methods which considered for calibration both season-specific and yearly predictor data over the 42 PAGASA stations (see the chapter for details). Whereas MOS-BC methods were shown to be not sensitive to the choice of data considered for calibration, season-specific data yielded better results in the case of the PP method (especially for the central part of the year). Therefore, season-specific data is used for the calibration of all the SD methods (both PP and MOS-BC) considered in this Thesis.
- We computed the biases of the ENSEMBLES models —ERA-Interim was used as reference— for the large-scale predictors variables included in P4 (see Table 6.1) over the geographical domain considered for SD (Figure 6.1). Beyond the expected differences among models, varying spatial patterns of bias were found in many cases for different target months within the same season. Thus, in order to be compatible with the reanalysis used for calibration, model predictor data must be properly preprocessed (or *harmonized*) before entering PP methods. To this, in this Thesis we first remove the model bias (the reanalysis used for calibration is taken as reference) separately for each month rather than for the entire season. By doing this, we force the model to follow the mean and variability of the reanalysis to some extent.

To fully accomplish the second objective of the Thesis, Chapter 8 assesses, also for the Philippines, the impact of the choice of reanalysis used for the calibration of the PP methods on the downscaled results —an issue which had never been addressed before,— both in ‘perfect’ and ‘non-perfect’ (i.e., using GCM predictor data for the prediction phase) conditions (see the chapter for details). The most important conclusions obtained are given next:

- We compared ERA-Interim and JRA-25 (two state-of-the-art reanalyses of reference) over a large region containing the Philippines and found that whereas reanalysis uncertainty (differences between both reanalyses) was trifling for circulation variables (zonal winds), it was important for thermodynamic ones (humidity and temperature). Thus, the PP assumption (reanalysis data reflecting ‘real’ large-scale atmospheric conditions) does not hold for the region considered.
- First, we applied an illustrative PP method (in cross-validation mode for the period 1981-2005) considering separately ERA-Interim and JRA-25 reanalyses in order to obtain the regression coefficients relating the local-scale precipitation at the 42 PAGASA stations to the large-scale predictors considered (selected based on the results from Chapter 6). The downscaled results were found to be sensitive to the reanalysis used for calibration if humidity and/or temperature—the variables showing the largest reanalysis uncertainty—were included in the predictor field. However, with local (spatial average) daily correlation differences of 0.1 (0.03) at the utmost, this sensitivity was relatively small. Afterwards, the coefficients calibrated with ERA-Interim and JRA-25 were separately applied to predictor data from the MPI-ECHAM5 (both control and scenario runs), as well as to four ENSEMBLES models in order to generate local-scale climate change projections (up to the end of the century) and seasonal forecasts (for the period 1981-2005), respectively. In the former case, the reanalysis-induced differences detected in present climate conditions were considerably amplified when climate signal bearing variables, i.e., humidity and temperature—which are indispensable for capturing the ‘correct’ climate change signal (Goodess and Palutikof, 1998; Wilby et al., 1998)—were included in the predictor field. In particular, the projected deltas for the end of the century (2071–2100 minus 1981–2000) were found to differ by up to a 35% (on average for the whole country) between the two reanalyses considered. In contrast, in the latter case, very similar local-scale seasonal predictions were obtained for most of the models, locations and seasons, independently from the reanalysis used for calibration. Therefore, the choice of reanalysis used for calibration of the PP methods was shown to be an important uncertainty source for climate change studies—for which it should be treated with equal care as other, well-known, uncertainty sources such as the choice of GCM or SDM (Dibike and Coulibaly, 2005; Chen et al., 2012),—whereas it is not of special relevance for the SD of seasonal forecasts.

3. Finally, with regard to the third objective, Chapter 9 assesses, also for the Philippines, whether SD can serve to improve the skill of the raw model global precipitation forecasts, beyond reducing their systematic biases (see Section 3.4.1-2). To this, and building on the lessons learnt from Chapters 6 and 8, we applied two MOS-BC and two PP methods to four of the ENSEMBLES models and analyzed their relative advantages and limitations by comparing the downscaled results against the corresponding raw model outputs at the 42 PAGASA stations for the period 1981-2005. In particular, we focused on accuracy and reliability aspects (for deterministic and tercile-based probabilistic predictions, respectively). The main conclusions found were:

- Overall, results vary mainly among seasons, but also among regions, models and SD methods. For the latter, they are more sensitive to the approach considered (MOS-BC or PP) than to the particular method used within each approach. In terms of accuracy, neither MOS-BC nor PP methods yield relevant improvements (with respect to raw model precipitation) for DJF and MAM, suggesting that the added value that can be obtained by means of SD is limited for those cases in which the models properly simulate precipitation. However, whereas MOS-BC methods do not clearly improve (or even worsen) the direct model output in JJA and SON, PP methods provide in general better (worse) accuracy than raw model precipitation does in the former (latter) season. In particular, PP methods yield large accuracy improvements in JJA over the northwestern part of the country for all models (with the exception of the ECMWF). Regarding reliability, the results obtained are in general very similar to those found for accuracy. In particular, whereas substantial added value is found for PP methods in JJA —when they lead to *marginally useful* categories over 50% of the cases (as compared to less than 10% for the direct model output and the MOS-BC methods),— the opposite situation is found in SON —when they lead to *not useful* or *dangerous* categories in nearly 50% of the cases (as compared to 10% for the direct model output).— The general agreement between the results found for accuracy and reliability points out the suitability of the classification proposed by Weisheimer and Palmer (2014) —which is slightly modified in this Thesis (see Section 3.5.1 for details)— for regional studies such as the one undertaken here.
- Those cases in which PP methods lead to a gain (loss) of skill —as measured by both accuracy and reliability— can be explained by situations where large-scale variables, which are defined over a synoptic domain, are better (worse)

predicted by the model than precipitation, which is more affected by particular local features. This suggests that PP methods might be able to exploit the models' ability for reproducing the large-scale predictor variables to indirectly obtain more skillful local precipitation forecasts (as compared to the direct model global outputs).

10.2 Related Publications

Part of the main results of this Thesis (Part III) have led to publications in international journals of relevance in the field of the atmospheric sciences:

- Chapter 7 is based on the results from “Manzanas, R., M. D. Frías, A. S. Cofiño, and J. M. Gutiérrez, 2014: Validation of 40 year multimodel seasonal precipitation forecasts: The role of ENSO on the global skill. *Journal of Geophysical Research: Atmospheres*, **119** (4), 1708–1719, doi:10.1002/2013JD020680.”
- Chapter 8 is based on the results from “Manzanas, R., S. Brands, D. San–Martín, A. Lucero, C. Limbo, and J. M. Gutiérrez, 2015: Statistical downscaling in the tropics can be sensitive to reanalysis choice: A case study for precipitation in the Philippines. *Journal of climate*, **28** (10), 4171–4184, doi:10.1175/JCLI-D-14-00331.1”
- Chapter 9 is based on the results from “Manzanas, R., J. M. Gutiérrez, and A. Lucero, 2016: Can statistical downscaling and bias correction methods improve the accuracy and reliability of seasonal forecasts? Submitted to *Climate Dynamics*.”

Additionally, in parallel to the development of this Thesis, other publications whose results have not been included here have also emerged:

- Alongside part of the methodological knowledge gained, some of the SD methods used in this Thesis have been applied over Spain in “San–Martín, D., R. Manzanas, S. Brands, S. Herrera, and J. M. Gutiérrez, 2016: Reassessing model uncertainty for regional projections of precipitation with an ensemble of statistical downscaling methods. Under review in *Journal of Climate*.”
- Manzanas, R., 2016: Can statistical downscaling improve the skill of global seasonal forecasts in Senegal? Under review in *Theoretical and Applied Climatology*.
- Manzanas, R., L. K. Amekudzi, K. Preko, S. Herrera, and J. M. Gutiérrez, 2014: Precipitation variability and trends in Ghana: An intercomparison of observational and reanalysis products. *Climatic Change*, **124** (4), 805–819, doi:10.1007/s10584-014-1100-9.

- Gutiérrez, J. M., D. San-Martín, S. Brands, R. Manzanas, and S. Herrera, 2013: Reassessing statistical downscaling techniques for their robust application under climate change conditions. *Journal of Climate*, **26** (1), 171–188, doi:10.1175/JCLI-D-11-00687.1.
- Brands, S., R. Manzanas, J. M. Gutiérrez, and J. Cohen, 2012: Seasonal predictability of wintertime precipitation in Europe using the Snow Advance Index. *Journal of Climate*, **25** (12), 4023–4028, doi:10.1175/JCLI-D-12-00083.1.

10.3 Developed Software: *MeteoLab*

Most of the calculations of this Thesis have been performed using *MeteoLab*, a *Matlab*[®] toolbox for statistical downscaling developed by the Santander Meteorology Group which can be freely downloaded from <http://meteo.unican.es/trac/MLToolbox>. Whereas the analog and the bias correction techniques (see Chapter 5) were already implemented in this toolbox, the methods based on GLMs have been developed in this Thesis and are available in the current version of *MeteoLab*. The code needed for reproducing all the SD methods used in this Thesis with *MeteoLab* is given below (the nomenclature introduced in Chapter 5 is used). The reader is referred to <http://meteo.unican.es/trac/MLToolbox/wiki/Downscaling> for details on further configuration possibilities.

```
method.type = 'ANALOGES';           % Method
method.properties.NumberOfPCs = 30; % Number of PCs
method.properties.AnalogueNumber = 1; % Number of analogs
method.properties.InferenceMethod = 'mean'; % Inference method
```

MeteoLab code for defining the *AN_det* SDM.

```
method.type = 'ANALOGES';           % Method
method.properties.NumberOfPCs = 30; % Number of PCs
method.properties.AnalogueNumber = 15; % Number of analogs
method.properties.InferenceMethod = 'rand'; % Inference method
```

MeteoLab code for defining the *AN_sto* SDM.

```
method.type = 'GLM';                 % Method
method.properties.ThresholdPrecip = 0.1; % Threshold for wet days (mm)
method.properties.NumberOfPCs = 15; % Number of PCs
method.properties.SimOccurrence = 'false'; % Not simulate occurrence
method.properties.SimAmount = 'false'; % Not simulate amount
```

MeteoLab code for defining the *GLM_det* SDM.

```

method.type = 'GLM'; % Method
method.properties.ThresholdPrecip = 0.1; % Threshold for wet days (mm)
method.properties.NumberOfPCs = 15; % Number of PCs
method.properties.SimOccurrence = 'true'; % Simulate occurrence
method.properties.SimAmount = 'true'; % Simulate amount

```

MeteoLab code for defining the *GLM_sto* SDM.

```

method.type = 'QGM'; % Method
method.properties.Variable = 'pr'; % Variable
method.properties.threshold = 0.1; % Threshold for wet days (mm)
method.properties.FreqCorrection = 'true'; % Frequency correction

```

MeteoLab code for defining the *QM_par* SDM.

```

method.type = 'EQM'; % Method
method.properties.Variable = 'pr'; % Variable
method.properties.threshold = 0.1; % Threshold for wet days (mm)
method.properties.extrapolation = 'constant'; % Type of extrapolation
method.properties.quantiles = 1:99; % Corrected percentiles
method.properties.FreqCorrection = 'true'; % Frequency correction

```

MeteoLab code for defining the *QM_emp* SDM.

10.4 Future Work

On the one hand, some of the results obtained during the realization of this Thesis have opened the door for the development of new works which constitute the natural continuation of some of the analysis presented here:

- In Section 3.4.1-2 we have characterized the drift of seasonal forecasts, considering precipitation from the ENSEMBLES models. In order to assess both the possible reduction of this drift in recent global models, but also its sensitivity to the ensemble size, the analysis was extended to the System 4 (the new version of the ECMWF model included in ENSEMBLES, the System 3). Preliminary results indicate that the drift is still important, with high values along the tropics and over the oceans. However, small ensembles (around 5 members) seem to be enough for its accurate characterization. Due to the lack of time, this work could not be finalized during the course of the Thesis and will be one of the first tasks to be addressed after its end, focusing on the implications that may arise for the bias correction of seasonal forecasts.

- As explained in Section 3.5.1, Weisheimer and Palmer (2014) assessed the reliability of global seasonal precipitation forecasts from the ECMWF System 4 (51 members) for the 21 only-land regions defined in Giorgi and Francisco (2000) and the results were presented in a scale ranging from 1 (dangerous) to 5 (perfect). Nevertheless, our preliminary investigation indicates that this classification is sensitive to a number of factors. For instance, the results from the latter may be altered if the confidence interval considered for the slope of the weighted reliability line is not suitable for the available ensemble size. Also, the results may substantially change depending on the region considered for spatial aggregation. Therefore, we will carefully analyze all these factors in collaboration with the authors of the original methodology.

On the other hand, in the framework of international initiatives and collaborations which have emerged during the Thesis, we also contemplate some works and research lines which will be developed in the coming future:

- Part of the methodological knowledge and some of the downscaling techniques developed during this Thesis have been applied to different regions in SPECS and EUPORIAS projects. In particular, an experiment in which dynamical and statistical (considering different approaches) downscaling are compared has been carried out over Brazil. In EUPORIAS, the added value of downscaled seasonal predictions has been assessed for Ethiopia (in the context of a climate service for early warning of droughts) and over two pilot regions in Europe (England and Italy). All these works will be completed during the coming months (for instance, in the case of Ethiopia we have found that the uncertainty in the observations may have an impact on the downscaled forecasts) and the most important results will be gathered in a number of publications which, given the heterogeneity of the regions considered (in terms of climatic variability, seasonal predictability, influence of ENSO, etc.), will provide a reference for future SD studies in the context of seasonal forecasting.
- Following from a research stay at the National Service for Meteorology and Hydrology of Perú (SENAMHI: <http://www.senamhi.gob.pe>), a high-quality observational dataset covering the whole country is available. Given the complex climatic characteristics of Perú (most of the climates of the world are present in the country), and the important influence of ENSO over the region, we consider a future research line in which, in collaboration with SENAMHI and taking advantage of the experience and methodological knowledge gained during this Thesis, the potential for the application of the different techniques for SD of seasonal forecasts in the country will be explored.

-
- Finally, the downscaling techniques developed in this Thesis have contributed to the *experiment 1a* of the VALUE COST action (<http://www.value-cost.eu>), which aims to carry out a systematic intercomparison of the different approaches and techniques for SD in the context of climate change. As a result of this initiative, a database with the predictions from more than 40 downscaling methods (the largest ensemble to-date) has been produced. Additionally, the Santander Meteorology Group has developed a portal which allows for defining indices and validation metrics which can be applied to the whole ensemble of available methods. Therefore, another future research line will be focused on including new verification metrics of interest for seasonal forecasting (such as the interannual correlation) in order to assess the potential of the different methods for this new context of application.

Bibliography

- Abaurrea, J. and J. Asín, 2005: Forecasting local daily precipitation patterns in a climate change scenario. *Climate Research*, **28 (3)**, 183–197, doi:10.3354/cr028183.
- Adler, R. F., et al., 2003: The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-present). *Journal of Hydrometeorology*, **4 (6)**, 1147–1167, doi:10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2.
- Aldrian, E., L. Dümenil Gates, and F. H. Widodo, 2007: Seasonal variability of Indonesian rainfall in ECHAM4 simulations and in the reanalyses: The role of ENSO. *Theoretical and Applied Climatology*, **87 (1-4)**, 41–59, doi:10.1007/s00704-006-0218-8.
- Amengual, A., V. Homar, R. Romero, S. Alonso, and C. Ramis, 2012: A statistical adjustment of regional climate model outputs to local scales: Application to Platja de Palma, Spain. *Journal of Climate*, **25 (3)**, 939–957, doi:10.1175/JCLI-D-10-05024.1.
- Arribas, A., et al., 2011: The GloSea4 ensemble prediction system for seasonal forecasting. *Monthly Weather Review*, **139 (6)**, 1891–1910, doi:10.1175/2010MWR3615.1.
- Balmaseda, M., 2012: Initialization techniques in seasonal forecasts. *Seminar on seasonal prediction: Science and applications*, ECMWF.
- Balmaseda, M. and D. L. T. Anderson, 2009: Impact of initialization strategies and observations on seasonal forecast skill. *Geophysical Research Letters*, **36 (1)**, n/a–n/a, doi:10.1029/2008GL035561.
- Balmaseda, M., D. L. T. Anderson, and A. Vidard, 2007: Impact of Argo on analyses of the global ocean. *Geophysical Research Letters*, **34 (16)**, n/a–n/a, doi:10.1029/2007GL030452.

- Barnston, A. G., 1994: Linear statistical short-term climate predictive skill in the Northern Hemisphere. *Journal of Climate*, **7** (10), 1513–1564, doi:10.1175/1520-0442(1994)007<1513:LSSTCP>2.0.CO;2.
- Barnston, A. G., S. Li, S. J. Mason, D. G. DeWitt, L. Goddard, and X. Gong, 2010: Verification of the first 11 years of IRI’s seasonal climate forecasts. *Journal of Applied Meteorology and Climatology*, **49** (3), 493–520, doi:10.1175/2009JAMC2325.1.
- Batté, L. and M. Déqué, 2011: Seasonal predictions of precipitation over Africa using coupled ocean-atmosphere general circulation models: Skill of the ENSEMBLES project multimodel ensemble forecasts. *Tellus A*, **63** (2), 283–299, doi:10.1111/j.1600-0870.2010.00493.x.
- Beck, C., J. Grieser, and B. Rudolf, 2005: A new monthly precipitation climatology for the global land areas for the period 1951 to 2000. Tech. rep., Klimastatusbericht, 181-190 pp.
- Becker, A., P. Finger, A. Meyer-Christoffer, B. Rudolf, K. Schamm, U. Schneider, and M. Ziese, 2013: A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901-present. *Earth System Science Data*, **5** (1), 71–99, doi:10.5194/essd-5-71-2013.
- Beckmann, B. R. and T. A. Buishand, 2002: Statistical downscaling relationships for precipitation in the Netherlands and North Germany. *International Journal of Climatology*, **22** (1), 15–32, doi:10.1002/joc.718.
- Beersma, J. J. and T. A. Buishand, 2003: Multi-site simulation of daily precipitation and temperature conditional on the atmospheric circulation. *Climate Research*, **25** (2), 121–133, doi:10.3354/cr025121.
- Benestad, R. E., 2010: Downscaling precipitation extremes: Correction of analog models through PDF predictions. *Theoretical and Applied Climatology*, **100** (1), 1–21, doi:10.1007/s00704-009-0158-1.
- Benson, C., 1997: The economic impact of natural disasters in the Philippines. Tech. rep., Overseas Development Institute (ODI), Working Paper 99.
- Berner, J., K. R. Smith, S. Y. Ha, J. P. Hacker, and C. Snyder, 2014: Increasing the skill of probabilistic forecasts: Model-error representations versus calibration and debiasing. *Monthly Weather Review* (submitted).

- Boé, J., L. Terray, F. Habets, and E. Martin, 2007: Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. *International Journal of Climatology*, **27** (12), 1643–1655, doi:10.1002/joc.1602.
- Brands, S., J. M. Gutiérrez, S. Herrera, and A. S. Cofiño, 2012: On the use of re-analysis data for downscaling. *Journal of Climate*, **25** (7), 2517–2526, doi:10.1175/JCLI-D-11-00251.1.
- Brands, S., S. Herrera, J. Fernández, and J. M. Gutiérrez, 2013: How well do CMIP5 Earth System Models simulate present climate conditions in Europe and Africa? *Climate Dynamics*, **41** (3-4), 803–817, doi:10.1007/s00382-013-1742-8.
- Brands, S., S. Herrera, D. San-Martín, and J. M. Gutiérrez, 2011: Validation of the ENSEMBLES global climate models over southwestern Europe using probability density functions, from a downscaling perspective. *Climate Research*, **48** (2-3), 145–161, doi:10.3354/cr00995.
- Brands, S., R. Manzananas, J. M. Gutiérrez, and J. Cohen, 2012: Seasonal predictability of wintertime precipitation in Europe using the Snow Advance Index. *Journal of Climate*, **25** (12), 4023–4028, doi:10.1175/JCLI-D-12-00083.1.
- Brandsma, T. and T. A. Buishand, 1997: Statistical linkage of daily precipitation in Switzerland to atmospheric circulation and temperature. *Journal of Hydrology*, **198** (1-4), 98–123, doi:10.1016/S0022-1694(96)03326-4.
- Branković, Č. and T. N. Palmer, 2000: Seasonal skill and predictability of ECMWF PROVOST ensembles. *Quarterly Journal of the Royal Meteorological Society*, **126** (567), 2035–2067, doi:10.1002/qj.49712656704.
- Bundel, A. Y., V. N. Kryzhov, Y. M. Min, V. M. Khan, R. M. Vilfand, and V. A. Tishchenko, 2011: Assessment of probability multimodel seasonal forecast based on the APCC model data. *Russian Meteorology and Hydrology*, **36** (3), 145–154, doi:10.3103/S1068373911030010.
- Bürger, G. and Y. Chen, 2005: Regression-based downscaling of spatial variability for hydrologic applications. *Journal of Hydrology*, **311** (1-4), 299–317, doi:10.1016/j.jhydrol.2005.01.025.
- CAIICP, 2010: *Assessment of intraseasonal to interannual climate prediction and predictability*. National Academies Press.

- Carrassi, A., R. J. T. Weber, V. Guemas, F. J. Doblas-Reyes, M. Asif, and D. Volpi, 2014: Full-field and anomaly initialization using a low-order climate model: A comparison and proposals for advanced formulations. *Nonlinear Processes in Geophysics*, **21** (2), 521–537, doi:10.5194/npg-21-521-2014.
- Casanueva, A., S. Herrera, J. Fernández, M. D. Frías, and J. M. Gutiérrez, 2013: Evaluation and projection of daily temperature percentiles from statistical and dynamical downscaling methods. *Natural Hazards and Earth System Sciences*, **13**, 2089–2099, doi:10.5194/nhess-13-2089-2013.
- Chandler, R. E. and H. S. Wheater, 2002: Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland. *Water Resources Research*, **38** (10), 10–1–10–11, doi:10.1029/2001WR000906.
- Charles, A., B. Timbal, E. Fernández, and H. Hendon, 2013: Analog downscaling of seasonal rainfall forecasts in the Murray Darling basin. *Monthly Weather Review*, **141** (3), 1099–1117, doi:10.1175/MWR-D-12-00098.1.
- Charles, S. P., C. B. Bryson, P. H. Whetton, and J. P. Hughes, 1999: Validation of downscaling models for changed climate conditions: Case study of southwestern Australia. *Climate Research*, **12** (1), 1–14, doi:10.3354/cr012001.
- Chen, H., C. Y. Xu, and S. Guo, 2012: Comparison and evaluation of multiple GCMs, statistical downscaling and hydrological models in the study of climate change impacts on runoff. *Journal of Hydrology*, **434**, 36–45, doi:10.1016/j.jhydrol.2012.02.040.
- Christensen, J. H., F. Boberg, O. B. Christensen, and P. Lucas-Picher, 2008: On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophysical Research Letters*, **35** (20), n/a–n/a, doi:10.1029/2008GL035694.
- Chu, J. L., H. Kang, C. Y. Tam, C. K. Park, and C. T. Chen, 2008: Seasonal forecast for local precipitation over northern Taiwan using statistical downscaling. *Journal of Geophysical Research: Atmospheres*, **113** (D12), n/a–n/a, doi:10.1029/2007JD009424.
- Chu, J. L. and P. S. Yu, 2010: A study of the impact of climate change on local precipitation using statistical downscaling. *Journal of Geophysical Research: Atmospheres*, **115** (D10), n/a–n/a, doi:10.1029/2009JD012357.
- Coelho, C. A. S. and S. M. S. Costa, 2010: Challenges for integrating seasonal climate forecasts in user applications. *Current Opinion in Environmental Sustainability*, **2** (5–6), 317–325, doi:10.1016/j.cosust.2010.09.002.

- Coelho, C. A. S., D. B. Stephenson, F. J. Doblas-Reyes, and M. Balmaseda, 2006: The skill of empirical and combined/calibrated coupled multi-model South American seasonal predictions during ENSO. *Advances in Geosciences*, **6**, 51–55, doi:10.5194/adgeo-6-51-2006.
- Cohen, J. and J. Jones, 2011: A new index for more accurate winter predictions. *Geophysical Research Letters*, **38** (21), n/a–n/a, doi:10.1029/2011GL049626.
- Coronas, J., 1920: *The climate and weather of the Philippines, 1903-1918*. PAGASA.
- Cubasch, U., H. von Storch, J. Waszkewitz, and E. Zorita, 1996: Estimates of climate change in Southern Europe derived from dynamical climate model output. *Climate Research*, **7** (2), 129–149, doi:10.3354/cr007129.
- de Castro, T. N., F. Souza, R. S. T. Pontes, L. L. N. dos Reis, and S. Daher, 2013: Neo-fuzzy neuron model for seasonal rainfall forecast: A case study of Ceara's eight homogenous regions. *Journal of Intelligent and Fuzzy Systems*, **25** (2), 389–394, doi:10.3233/IFS-2012-0645.
- Dee, D. P., et al., 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137** (656), 553–597, doi:10.1002/qj.828.
- Déqué, M., 2007: Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values. *Global and Planetary Change*, **57** (1-2), 16–26, doi:10.1016/j.gloplacha.2006.11.030.
- Dibike, Y. B. and P. Coulibaly, 2005: Hydrologic impact of climate change in the Saguenay watershed: Comparison of downscaling methods and hydrologic models. *Journal of Hydrology*, **307** (1-4), 145–163, doi:10.1016/j.hydrol.2004.10.012.
- Díez, E., C. Primo, J. A. García-Moya, J. M. Gutiérrez, and B. Orfila, 2005: Statistical and dynamical downscaling of precipitation over Spain from DEMETER seasonal forecasts. *Tellus A*, **57** (3), 409–423, doi:10.1111/j.1600-0870.2005.00130.x.
- Doblas-Reyes, F. J., C. A. S. Coelho, and D. B. Stephenson, 2008: How much does simplification of probability forecasts reduce forecast quality? *Meteorological Applications*, **15** (1), 155–162, doi:10.1002/met.50.
- Doblas-Reyes, F. J., M. Déqué, and J. P. Pielieuvre, 2000: Multi-model spread and probabilistic seasonal forecasts in PROVOST. *Quarterly Journal of the Royal Meteorological Society*, **126** (567), 2069–2087, doi:10.1002/qj.49712656705.

- Doblas-Reyes, F. J., J. García-Serrano, F. Lienert, A. P. Biescas, and L. R. L. Rodrigues, 2013: Seasonal climate predictability and forecasting: Status and prospects. *Wiley Interdisciplinary Reviews: Climate Change*, **4** (4), 245–268, doi:10.1002/wcc.217.
- Doblas-Reyes, F. J., A. Weisheimer, T. N. Palmer, J. M. Murphy, and D. Smith, 2010: Forecast quality assessment of the ENSEMBLES seasonal-to-decadal Stream 2 hindcasts. Tech. Rep. 621, ECMWF.
- Doblas-Reyes, F. J., et al., 2009: Addressing model uncertainty in seasonal and annual dynamical ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, **135** (643), 1538–1559, doi:10.1002/qj.464.
- Duan, W. and C. Wei, 2013: The ‘spring predictability barrier’ for ENSO predictions and its possible mechanism: Results from a fully coupled model. *International Journal of Climatology*, **33** (5), 1280–1292, doi:10.1002/joc.3513.
- Durman, C. F., J. M. Gregory, D. C. Hassell, R. G. Jones, and J. M. Murphy, 2001: A comparison of extreme European daily precipitation simulated by a global and a regional climate model for present and future climates. *Quarterly Journal of the Royal Meteorological Society*, **127** (573), 1005–1015, doi:10.1002/qj.49712757316.
- Eden, J. M., G. J. van Oldenborgh, E. Hawkins, and E. B. Suckling, 2015: A global empirical system for probabilistic seasonal climate prediction. *Geoscientific Model Development*, **8** (12), 3947–3973, doi:10.5194/gmd-8-3947-2015.
- Eden, J. M., M. Widmann, D. Grawe, and S. Rast, 2012: Skill, correction, and downscaling of GCM-simulated precipitation. *Journal of Climate*, **25** (11), 3970–3984, doi:10.1175/JCLI-D-11-00254.1.
- Ehret, U., E. Zehe, V. Wulfmeyer, K. Warrach-Sagi, and J. Liebert, 2012: HESS Opinions “Should we apply bias correction to global and regional climate model data?”. *Hydrology and Earth System Sciences*, **16** (9), 3391–3404, doi:10.5194/hess-16-3391-2012.
- Fealy, R. and J. Sweeney, 2007: Statistical downscaling of precipitation for a selection of sites in Ireland employing a generalised linear modelling approach. *International Journal of Climatology*, **27** (15), 2083–2094, doi:10.1002/joc.1506.
- Fedderson, H. and U. Andersen, 2005: A method for statistical downscaling of seasonal ensemble predictions. *Tellus A*, **57** (3), 398–408, doi:10.1111/j.1600-0870.2005.00102.x.
- Fedderson, H., A. Navarra, and M. N. Ward, 1999: Reduction of model systematic error by statistical correction for dynamical seasonal predictions. *Journal of Climate*, **12** (7), 1974–1989, doi:10.1175/1520-0442(1999)012<1974:ROMSEB>2.0.CO;2.

- Ferro, C. A. T., D. S. Richardson, and A. P. Weigel, 2008: On the effect of ensemble size on the discrete and continuous ranked probability scores. *Meteorological Applications*, **15** (1), 19–24, doi:10.1002/met.45.
- Flores, J. F. and V. F. Balagot, 1969: *World Survey of Climatology*, Climates of Northern and Eastern Asia, Vol. 8, chap. Climate of the Philippines, 159–213. Arakawa.
- Fowler, H. J., S. Blenkinsop, and C. Tebaldi, 2007: Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology*, **27** (12), 1547–1578, doi:10.1002/joc.1556.
- Frías, M. D., S. Herrera, A. S. Cofiño, and J. M. Gutiérrez, 2010: Assessing the skill of precipitation and temperature seasonal forecasts in Spain: Windows of opportunity related to ENSO events. *Journal of Climate*, **23** (2), 209–220, doi:10.1175/2009JCLI2824.1.
- Giorgetta, M. A., G. P. Brasseur, E. Roeckner, and J. Marotzke, 2006: Preface to special section on climate models at the Max Planck Institute for Meteorology. *Journal of Climate*, **19** (16), 3769–3770, doi:10.1175/JCLI9023.1.
- Giorgi, F. and R. Francisco, 2000: Uncertainties in regional climate change prediction: A regional analysis of ensemble simulations with the HADCM2 coupled AOGCM. *Climate Dynamics*, **16** (2-3), 169–182, doi:10.1007/PL00013733.
- Giorgi, F. and L. O. Mearns, 1999: Introduction to special section: Regional climate modeling revisited. *Journal of Geophysical Research: Atmospheres*, **104** (D6), 6335–6352, doi:10.1029/98JD02072.
- Giorgi, F., et al., 2001: *Climate change 2001: The scientific basis*. Cambridge University Press, 583–638.
- Glahn, H. R. and D. A. Lowry, 1972: The use of Model Output Statistics (MOS) in objective weather forecasting. *Journal of Applied Meteorology*, **11** (8), 1203–1211, doi:10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2.
- Goddard, L., A. G. Barnston, and S. J. Mason, 2003: Evaluation of the IRI’s “net assessment” seasonal climate forecasts: 1997–2001. *Bulletin of the American Meteorological Society*, **84** (12), 1761–1781, doi:10.1175/BAMS-84-12-1761.
- Goddard, L. and M. Dilley, 2005: El Niño: Catastrophe or opportunity. *Journal of Climate*, **18** (5), 651–665, doi:10.1175/JCLI-3277.1.

- Goddard, L., et al., 2010: Providing seasonal-to-interannual climate information for risk management and decision-making. *Procedia Environmental Sciences*, **1**, 81–101, doi:10.1016/j.proenv.2010.09.007.
- Goodess, C. M. and J. P. Palutikof, 1998: Development of daily rainfall scenarios for southeast Spain using a circulation-type approach to downscaling. *International Journal of Climatology*, **18** (10), 1051–1083, doi:10.1002/(SICI)1097-0088(199808)18:10<1051::AID-JOC304>3.0.CO;2-1.
- Grotch, S. L. and M. C. MacCracken, 1991: The use of general circulation models to predict regional climatic change. *Journal of Climate*, **4** (3), 286–303, doi:10.1175/1520-0442(1991)004<0286:TUOGCM>2.0.CO;2.
- Gutiérrez, J. M., R. Cano, A. S. Cofiño, and C. Sordo, 2005: Analysis and downscaling multi-model seasonal forecasts in Peru using self-organizing maps. *Tellus A*, **57** (3), 435–447, doi:10.1111/j.1600-0870.2005.00128.x.
- Gutiérrez, J. M., A. S. Cofiño, R. Cano, and M. A. Rodríguez, 2004: Clustering methods for statistical downscaling in short-range weather forecasts. *Monthly Weather Review*, **132** (9), 2169–2183, doi:10.1175/1520-0493(2004)132<2169:CMFSDI>2.0.CO;2.
- Gutiérrez, J. M., D. San-Martín, S. Brands, R. Manzanas, and S. Herrera, 2013: Reassessing statistical downscaling techniques for their robust application under climate change conditions. *Journal of Climate*, **26** (1), 171–188, doi:10.1175/JCLI-D-11-00687.1.
- Gutjahr, O. and G. Heinemann, 2013: Comparing precipitation bias correction methods for high-resolution regional climate simulations using COSMO-CLM. *Theoretical and Applied Climatology*, **114** (3), 511–529, doi:10.1007/s00704-013-0834-z.
- Haddad, Z. S., 2004: Global variability of precipitation according to the Tropical Rainfall Measuring Mission. *Journal of Geophysical Research: Atmospheres*, **109** (D17), doi:10.1029/2004JD004607.
- Haerter, J. O., S. Hagemann, C. Moseley, and C. Piani, 2011: Climate model bias correction and the role of timescales. *Hydrology and Earth System Sciences*, **15** (3), 1065–1079, doi:10.5194/hess-15-1065-2011.
- Hagemann, S., C. Chen, J. O. Haerter, J. Heinke, D. Gerten, and C. Piani, 2011: Impact of a statistical bias correction on the projected hydrological changes obtained from three GCMs and two hydrology models. *Journal of Hydrometeorology*, **12** (4), 556–578, doi:10.1175/2011JHM1336.1.

- Halpert, M. S. and C. F. Ropelewski, 1992: Surface temperature patterns associated with the Southern Oscillation. *Journal of Climate*, **5 (6)**, 577–593, doi:10.1175/1520-0442(1992)005<0577:STPAWT>2.0.CO;2.
- Hansen, J. W., A. Challinor, A. V. M. Inés, T. Wheeler, and V. Moron, 2006: Translating climate forecasts into agricultural terms: Advances and challenges. *Climate Research*, **33 (1)**, 27–41, doi:10.3354/cr033027.
- Hanssen-Bauer, I., C. Achberger, R. E. Benestad, D. Chen, and E. J. Forland, 2005: Statistical downscaling of climate scenarios over Scandinavia. *Climate Research*, **29 (3)**, 255–268, doi:10.3354/cr029255.
- Hastenrath, S., 1995: Recent advances in tropical climate prediction. *Journal of Climate*, **8 (6)**, 1519–1532, doi:10.1175/1520-0442(1995)008<1519:RAITCP>2.0.CO;2.
- Haylock, M. and J. McBride, 2001: Spatial coherence and predictability of Indonesian wet season rainfall. *Journal of Climate*, **14 (18)**, 3882–3887, doi:10.1175/1520-0442(2001)014<3882:SCAPOI>2.0.CO;2.
- Haylock, M. R., G. C. Cawley, C. Harpham, R. L. Wilby, and C. M. Goodess, 2006: Downscaling heavy precipitation over the United Kingdom: A comparison of dynamical and statistical methods and their future scenarios. *International Journal of Climatology*, **26 (10)**, 1397–1415, doi:10.1002/joc.1318.
- Hempel, S., K. Frieler, L. Warszawski, J. Schewe, and F. Piontek, 2013: A trend-preserving bias correction: The ISI-MIP approach. *Earth Systems Dynamics*, **4 (1)**, 49–92, doi:10.5194/esdd-4-49-2013.
- Hertig, E. and J. Jacobeit, 2008: Assessments of Mediterranean precipitation changes for the 21st century using statistical downscaling techniques. *International Journal of Climatology*, **28 (8)**, 1025–1045, doi:10.1002/joc.1597.
- Hertig, E., S. Seubert, A. Paxian, G. Vogt, H. Paeth, and J. Jacobeit, 2013: Changes of total versus extreme precipitation and dry periods until the end of the twenty-first century: Statistical assessments for the Mediterranean area. *Theoretical and Applied Climatology*, **111 (1-2)**, 1–20, doi:10.1007/s00704-012-0639-5.
- Hewitson, B. C. and R. G. Crane, 1996: Climate downscaling: Techniques and application. *Climate Research*, **7 (2)**, 85–95, doi:10.3354/cr007085.
- Hewitson, B. C. and R. G. Crane, 2006: Consensus between GCM climate change projections with empirical downscaling: Precipitation downscaling over South Africa. *International Journal of Climatology*, **26 (10)**, 1315–1337, doi:10.1002/joc.1314.

- Hewitson, B. C., J. Daron, R. G. Crane, M. F. Zermoglio, and C. Jack, 2014: Interrogating empirical-statistical downscaling. *Climatic Change*, **122** (4), 539–554, doi:10.1007/s10584-013-1021-z.
- Hilario, F., R. de Guzman, D. Ortega, P. Hayman, and B. Alexander, 2009: El Niño–Southern Oscillation in the Philippines: Impacts, forecasts, and risk managements. *Philippine Journal of Development*, **36** (1), 9–34.
- Hirschi, M., C. Spirig, A. P. Weigel, P. Calanca, J. Samietz, and M. W. Rotach, 2012: Monthly weather forecasts in a pest forecasting context: Downscaling, recalibration, and skill improvement. *Journal of Applied Meteorology and Climatology*, **51** (9), 1633–1638, doi:10.1175/JAMC-D-12-082.1.
- Holland, M., D. Bailey, and S. Vavrus, 2011: Inherent sea ice predictability in the rapidly changing Arctic environment of the Community Climate System Model, version 3. *Climate Dynamics*, **36** (7-8), 1239–1253, doi:10.1007/s00382-010-0792-4.
- Huth, R., 1999: Statistical downscaling in central Europe: Evaluation of methods and potential predictors. *Climate Research*, **13** (2), 91–101, doi:10.3354/cr013091.
- Huth, R., C. Beck, and O. E. Tveito, 2010: Classifications of atmospheric circulation patterns: Theory and applications (Preface). *Physics and Chemistry of the Earth*, **35** (9-12), 307–308, doi:10.1016/j.pce.2010.06.005.
- Imbert, A. and R. Benestad, 2005: An improvement of analog model strategy for more reliable local climate change scenarios. *Theoretical and Applied Climatology*, **82** (3-4), 245–255, doi:10.1007/s00704-005-0133-4.
- Israel, D. C. and R. M. Briones, 2012: Impacts of natural disasters on agriculture, food security, and natural resources and environment in the Philippines. Tech. rep., Philippine Institute for Development Studies.
- Jolliffe, I. T. and D. B. Stephenson, 2003: *Forecast verification: A practitioner's guide in atmospheric sciences*. John Wiley & Sons.
- Kang, H., K. H. An, C. K. Park, A. L. S. Solís, and K. Stitthichivapak, 2007: Multi-model output statistical downscaling prediction of precipitation in the Philippines and Thailand. *Geophysical Research Letters*, **34** (15), n/a–n/a, doi:10.1029/2007GL030730.
- Kang, I. S., J. Y. Lee, and C. K. Park, 2004: Potential predictability of summer mean precipitation in a dynamical seasonal prediction system with systematic error correction. *Journal of Climate*, **17** (4), 834–844.

- Kay, J. K., H. M. Kim, Y. Y. Park, and J. Son, 2013: Effect of doubling the ensemble size on the performance of ensemble prediction in the warm season using MOGREPS implemented at the KMA. *Advances in Atmospheric Sciences*, **30** (5), 1287–1302, doi:10.1007/s00376-012-2083-y.
- Kayano, M. T., C. P. de Oliveira, and R. V. Andreoli, 2009: Interannual relations between South American rainfall and tropical sea surface temperature anomalies before and after 1976. *International Journal of Climatology*, **29** (10), 1439–1448, doi:10.1002/joc.1824.
- Kharin, V. V. and F. W. Zwiers, 2003: On the ROC score of probability forecasts. *Journal of Climate*, **16** (24), 4145–4150, doi:10.1175/1520-0442(2003)016<4145:OTRSOP>2.0.CO;2.
- Kim, H. M., P. J. Webster, and J. A. Curry, 2012a: Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere winter. *Climate Dynamics*, **39** (12), 2957–2973, doi:10.1007/s00382-012-1364-6.
- Kim, H. M., P. J. Webster, J. A. Curry, and V. E. Toma, 2012b: Asian summer monsoon prediction in ECMWF System 4 and NCEP CFSv2 retrospective seasonal forecasts. *Climate Dynamics*, **39** (12), 2975–2991, doi:10.1007/s00382-012-1470-5.
- Kintanar, R. L., 1984: Climate of the Philippines. Tech. rep., PAGASA, 38 pp.
- Kirono, D. G. C. and N. J. Tapper, 1999: ENSO rainfall variability and impacts on crop production in Indonesia. *Physical Geography*, **20** (6), 508–519, doi:10.1080/02723646.1999.10642693.
- Kirtman, B. and A. Pirani, 2009: The state of the art of seasonal Prediction: Outcomes and recommendations from the first World Climate Research Program workshop on seasonal prediction. *Bulletin of the American Meteorological Society*, **90** (4), 455–458, doi:10.1175/2008BAMS2707.1.
- Kirtman, B., et al., 2014: The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, **95** (4), 585–601, doi:10.1175/BAMS-D-12-00050.1.
- Klein, W. H., B. M. Lewis, and I. Enger, 1959: Objective prediction of five-day mean temperatures during winter. *Journal of Meteorology*, **16** (6), 672–682, doi:10.1175/1520-0469(1959)016<0672:OPOFDM>2.0.CO;2.

- Koide, N., A. W. Robertson, A. V. M. Inés, J. H. Qian, D. G. DeWitt, and A. Lucero, 2012: Prediction of rice production in the Philippines using seasonal climate forecasts. *Journal of Applied Meteorology and Climatology*, **52** (3), 552–569, doi:10.1175/JAMC-D-11-0254.1.
- Koster, R. D., et al., 2010: Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment. *Geophysical Research Letters*, **37** (2), n/a–n/a, doi:10.1029/2009GL041677.
- Kumar, A., 2009: Finite samples and uncertainty estimates for skill measures for seasonal prediction. *Monthly Weather Review*, **137** (8), 2622–2631, doi:10.1175/2009MWR2814.1.
- Lachenbruch, P. A. and M. R. Mickey, 1968: Estimation of error rates in discriminant analysis. *Technometrics*, **10** (1), 1–11, doi:10.2307/1266219.
- Landman, W. A. and A. Beraki, 2012: Multi-model forecast skill for mid-summer rainfall over southern Africa. *International Journal of Climatology*, **32** (2), 303–314, doi:10.1002/joc.2273.
- Landman, W. A., M. J. Kgatuke, M. Mbedzi, A. Beraki, A. Bartman, and A. du Piesanie, 2009: Performance comparison of some dynamical and empirical downscaling methods for South Africa from a seasonal climate modelling perspective. *International Journal of Climatology*, **29** (11), 1535–1549, doi:10.1002/joc.1766.
- Landman, W. A. and W. J. Tennant, 2000: Statistical downscaling of monthly forecasts. *International Journal of Climatology*, **20** (13), 1521–1532, doi:10.1002/1097-0088(20001115)20:13<1521::AID-JOC558>3.0.CO;2-N.
- Lansigan, F. P., 2005: Coping with climate variability and change in rice production systems in the Philippines. *Rice is Life: Scientific Perspectives for the 21st Century*, International Rice Research Institute (IRRI).
- Laprise, R., 2008: Regional climate modelling. *Journal of Computational Physics*, **227** (7), 3641–666, doi:10.1016/j.jcp.2006.10.024.
- Latif, M., et al., 2001: ENSIP: The El Niño simulation intercomparison project. *Climate Dynamics*, **18** (3-4), 255–276, doi:10.1007/s003820100174.
- Lazar, A., A. Vintzileos, F. J. Doblas-Reyes, P. Rogel, and P. Delécluse, 2005: Seasonal forecast of tropical climate with coupled ocean–atmosphere general circulation models: On the respective role of the atmosphere and the ocean components in the drift of the

- surface temperature error. *Tellus A*, **57** (3), 387–397, doi:10.1111/j.1600-0870.2005.00135.x.
- Leloup, J., M. Lengaigne, and J. P. Boulanger, 2007: Twentieth century ENSO characteristics in the IPCC database. *Climate Dynamics*, **30** (2), 277–291, doi:10.1007/s00382-007-0284-3.
- Lemos, M. C., C. J. Kirchhoff, and V. Ramprasad, 2012: Narrowing the climate information usability gap. *Nature Climate Change*, **2** (11), 789–794, doi:10.1038/nclimate1614.
- Li, C., R. Lu, and B. Dong, 2012: Predictability of the western North Pacific summer climate demonstrated by the coupled models of ENSEMBLES. *Climate Dynamics*, **39** (1-2), 329–346, doi:10.1007/s00382-011-1274-z.
- Li, H., J. Sheffield, and E. F. Wood, 2010: Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *Journal of Geophysical Research: Atmospheres*, **115** (D10), n/a–n/a, doi:10.1029/2009JD012882.
- Lim, E. P., H. H. Hendon, D. L. T. Anderson, A. Charles, and O. Alves, 2011: Dynamical, statistical-dynamical, and multimodel ensemble forecasts of Australian spring season rainfall. *Monthly Weather Review*, **139** (3), 958–975, doi:10.1175/2010MWR3399.1.
- Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogues. *Journal of the Atmospheric Sciences*, **26** (4), 636–646, doi:10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2.
- Luo, Q., L. Wen, J. L. McGregor, and B. Timbal, 2013: A comparison of downscaling techniques in the projection of local climate change and wheat yields. *Climatic Change*, **120** (1-2), 249–261, doi:10.1007/s10584-013-0802-8.
- Lyon, B., H. Cristi, E. R. Verceles, F. D. Hilario, and R. Abastillas, 2006: Seasonal reversal of the ENSO rainfall signal in the Philippines. *Geophysical Research Letters*, **33** (24), n/a–n/a, doi:10.1029/2006GL028182.
- Ma, S., X. Rodó, and F. J. Doblas-Reyes, 2012: Evaluation of the DEMETER performance for seasonal hindcasts of the Indian summer monsoon rainfall. *International Journal of Climatology*, **32** (11), 1717–1729, doi:10.1002/joc.2389.
- Magnusson, L., M. Balmaseda, S. Corti, F. Molteni, and T. Stockdale, 2013a: Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. *Climate Dynamics*, **41** (9-10), 2393–2409, doi:10.1007/s00382-012-1599-2.

- Magnusson, L., M. Balmaseda, and F. Molteni, 2013b: On the dependence of ENSO simulation on the coupled model mean state. *Climate Dynamics*, **41** (5-6), 1509–1525, doi:10.1007/s00382-012-1574-y.
- Manzanas, R., L. K. Amekudzi, K. Preko, S. Herrera, and J. M. Gutiérrez, 2014a: Precipitation variability and trends in Ghana: An intercomparison of observational and reanalysis products. *Climatic Change*, **124** (4), 805–819, doi:10.1007/s10584-014-1100-9.
- Manzanas, R., S. Brands, D. San-Martín, A. Lucero, C. Limbo, and J. M. Gutiérrez, 2015: Statistical downscaling in the tropics can be sensitive to reanalysis choice: A case study for precipitation in the Philippines. *Journal of Climate*, **28** (10), 4171–4184, doi:10.1175/JCLI-D-14-00331.1.
- Manzanas, R., M. D. Frías, A. S. Cofiño, and J. M. Gutiérrez, 2014b: Validation of 40 year multimodel seasonal precipitation forecasts: The role of ENSO on the global skill. *Journal of Geophysical Research: Atmospheres*, **119** (4), 1708–1719, doi:10.1002/2013JD020680.
- Maraun, D., 2012: Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums. *Geophysical Research Letters*, **39** (6), n/a–n/a, doi:10.1029/2012GL051210.
- Maraun, D., 2013: Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *Journal of Climate*, **26** (6), 2137–2143, doi:10.1175/JCLI-D-12-00821.1.
- Maraun, D., et al., 2010: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, **48** (3), n/a–n/a, doi:10.1029/2009RG000314.
- Maraun, D., et al., 2015: VALUE: A framework to validate downscaling approaches for climate change studies. *Earth's Future*, **3** (1), 1–14, doi:10.1002/2014EF000259, 2014EF000259.
- Markatou, M., H. Tian, S. Biswas, and G. Hripcsak, 2005: Analysis of variance of cross-validation estimators of the generalization error. *Journal of Machine Learning Research*, **6**, 1127–1168, doi:10.7916/D86D5R2X.
- Marzban, C., S. Sandgathe, and E. Kalnay, 2006: MOS, Perfect-Prog and reanalysis. *Monthly Weather Review*, **134** (2), 657–663, doi:10.1175/MWR3088.1.
- Mason, S. J. and L. Goddard, 2001: Probabilistic Precipitation Anomalies Associated with ENSO. *Bulletin of the American Meteorological Society*, **82** (4), 619–638, doi:10.1175/1520-0477(2001)082<0619:PPAAWE>2.3.CO;2.

- Mason, S. J. and N. E. Graham, 2002: Areas beneath the Relative Operating Characteristics (ROC) and Relative Operating Levels (ROL) curves: Statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society*, **128** (584), 2145–2166, doi:10.1256/003590002320603584.
- Matulla, C., X. Zhang, X. Wang, J. Wang, E. Zorita, S. Wagner, and H. von Storch, 2008: Influence of similarity measures on the performance of the analog method for downscaling daily precipitation. *Climate Dynamics*, **30** (2-3), 133–144, doi:10.1007/s00382-007-0277-2.
- Maurer, E. P. and D. W. Pierce, 2014: Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean. *Hydrology and Earth System Sciences*, **18** (3), 915–925, doi:10.5194/hess-18-915-2014.
- Maxino, C. C., B. J. McAvaney, A. J. Pitman, and S. E. Perkins, 2008: Ranking the AR4 climate models over the Murray Darling basin using simulated maximum temperature, minimum temperature and precipitation. *International Journal of Climatology*, **28** (8), 1097–1112, doi:10.1002/joc.1612.
- Meehl, G., et al., 2007: *Climate change 2007: The physical science basis*, chap. Global Climate Projections. Cambridge University Press.
- Mo, R. and D. M. Straus, 2002: Statistical-dynamical seasonal prediction based on Principal Component regression of GCM ensemble integrations. *Monthly Weather Review*, **130** (9), 2167–2187, doi:10.1175/1520-0493(2002)130<2167:SDSPBO>2.0.CO;2.
- Molteni, F., et al., 2011: The new ECMWF seasonal forecast system (System 4). Tech. rep., ECMWF.
- Moron, V., A. Lucero, F. Hilario, B. Lyon, A. W. Robertson, and D. DeWitt, 2009: Spatio-temporal variability and predictability of summer monsoon onset over the Philippines. *Climate Dynamics*, **33** (7-8), 1159–1177, doi:10.1007/s00382-008-0520-5.
- Murphy, J., 1999: An evaluation of statistical and dynamical techniques for downscaling local climate. *Journal of Climate*, **12** (8), 2256–2284, doi:10.1175/1520-0442(1999)012<2256:AEOSAD>2.0.CO;2.
- Müller, W. A., C. Appenzeller, and C. Schär, 2005: Probabilistic seasonal prediction of the winter North Atlantic Oscillation and its impact on near surface temperature. *Climate Dynamics*, **24** (2-3), 213–226, doi:10.1007/s00382-004-0492-z.

- Nelder, J. A. and R. W. M. Wedderburn, 1972: Generalized linear models. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, **135** (3), 370–384, doi:10.2307/2344614.
- Nicholls, N., 1984: The stability of empirical long-range forecast techniques: A case study. *Journal of Climate and Applied Meteorology*, **23** (1), 143–147, doi:10.1175/1520-0450(1984)023<0143:TSOELR>2.0.CO;2.
- Onogi, K., et al., 2007: The JRA-25 reanalysis. *Journal of the Meteorological Society of Japan*, **85** (3), 369–432, doi:10.2151/jmsj.85.369.
- Paeth, H. and M. Diederich, 2011: Postprocessing of simulated precipitation for impact research in West Africa. Part II: A weather generator for daily data. *Climate Dynamics*, **36** (7-8), 1337–1348, doi:10.1007/s00382-010-0840-0.
- Palmer, T. N., et al., 2004: Development of a European Multimodel Ensemble system for seasonal to inTERannual prediction (DEMETER). *Bulletin of the American Meteorological Society*, **85** (6), 853–872, doi:10.1175/BAMS-85-6-853.
- Panofsky, H. A. and G. W. Brier, 1968: *Some applications of statistics to meteorology*. Penn State University (College of Earth and Mineral Sciences).
- Park, J. H., S. G. Oh, and M. S. Suh, 2013: Impacts of boundary conditions on the precipitation simulation of RegCM4 in the CORDEX East Asia domain. *Journal of Geophysical Research: Atmospheres*, **118** (4), 1652–1667, doi:10.1002/jgrd.50159.
- Paul, S., C. M. Liu, J. M. Chen, and S. H. Lin, 2008: Development of a statistical downscaling model for projecting monthly rainfall over East Asia from a general circulation model output. *Journal of Geophysical Research: Atmospheres*, **113** (D15), n/a–n/a, doi:10.1029/2007JD009472.
- Pavan, V., S. Marchesi, A. Morgillo, C. Cacciamani, and F. J. Doblas-Reyes, 2005: Downscaling of DEMETER winter seasonal hindcasts over Northern Italy. *Tellus A*, **57** (3), 424–434, doi:10.1111/j.1600-0870.2005.00111.x.
- Perkins, S. E., A. J. Pitman, N. J. Holbrook, and J. McAneney, 2007: Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. *Journal of Climate*, **20** (17), 4356–4376, doi:10.1175/JCLI4253.1.
- Piani, C., J. O. Haerter, and E. Coppola, 2010: Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoretical and Applied Climatology*, **99** (1-2), 187–192, doi:10.1007/s00704-009-0134-9.

- Pozo-Vázquez, D., S. R. Gámiz-Fortis, J. Tovar-Pescador, M. J. Esteban-Parra, and Y. Castro-Díez, 2005: El Niño-Southern Oscillation events and associated European winter precipitation anomalies. *International Journal of Climatology*, **25** (1), 17–31, doi:10.1002/joc.1097.
- Preisendorfer, R., 1988: *Principal component analysis in meteorology and oceanography*. 1st ed., Elsevier.
- Quintana-Seguí, P., A. Ribes, E. Martín, F. Habets, and J. Boé, 2010: Comparison of three downscaling methods in simulating the impact of climate change on the hydrology of Mediterranean basins. *Journal of Hydrology*, **383** (1-2), 111–124, doi:10.1016/j.jhydrol.2009.09.050.
- Räisänen, J., 2007: How reliable are climate models? *Tellus A*, **59** (1), 2–29, doi:10.1111/j.1600-0870.2006.00211.x.
- Räisänen, J. and O. Räty, 2013: Projections of daily mean temperature variability in the future: Cross-validation tests with ENSEMBLES regional climate simulations. *Climate Dynamics*, **41** (5-6), 1553–1568, doi:10.1007/s00382-012-1515-9.
- Ribalaygua, J., L. Torres, J. Pórtoles, R. Monjo, E. Gaitán, and M. Pino, 2013: Description and validation of a two-step analogue/regression downscaling method. *Theoretical and Applied Climatology*, **114** (1-2), 253–269, doi:10.1007/s00704-013-0836-x.
- Richardson, D. S., 2001: Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Quarterly Journal of the Royal Meteorological Society*, **127** (577), 2473–2489, doi:10.1002/qj.49712757715.
- Robertson, A. W., S. Kirshner, and P. Smyth, 2004: Downscaling of daily rainfall occurrence over Northeast Brazil using a Hidden Markov Model. *Journal of Climate*, **17** (22), 4407–4424, doi:10.1175/JCLI-3216.1.
- Robertson, A. W., J. H. Qian, M. K. Tippett, V. Moron, and A. Lucero, 2012: Downscaling of seasonal rainfall over the Philippines: Dynamical versus statistical approaches. *Monthly Weather Review*, **140** (4), 1204–1218, doi:10.1175/MWR-D-11-00177.1.
- Ropelewski, C. F. and M. S. Halpert, 1987: Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Monthly Weather Review*, **115** (8), 1606–1626, doi:10.1175/1520-0493.
- Rowell, D. P., 1998: Assessing potential seasonal predictability with an ensemble of multidecadal GCM simulations. *Journal of Climate*, **11** (2), 109–120, doi:10.1175/1520-0442(1998)011<0109:APSPWA>2.0.CO;2.

- Rowell, D. P., C. K. Folland, K. Maskell, and M. N. Ward, 1995: Variability of summer rainfall over tropical North Africa (1906–92): Observations and modelling. *Quarterly Journal of the Royal Meteorological Society*, **121** (523), 669–704, doi:10.1002/qj.49712152311.
- Ruosteenoja, K. and P. Räisänen, 2013: Seasonal changes in solar radiation and relative humidity in Europe in response to global warming. *Journal of Climate*, **26** (8), 2467–2481, doi:10.1175/JCLI-D-12-00007.1.
- San-Martín, D., R. Manzananas, S. Brands, S. Herrera, and J. M. Gutiérrez, 2016: Re-assessing model uncertainty for regional projections of precipitation with an ensemble of statistical downscaling methods. *Journal of Climate* (submitted).
- Sauter, T. and V. Venema, 2011: Natural three-dimensional predictor domains for statistical precipitation downscaling. *Journal of Climate*, **24** (23), 6132–6145, doi:10.1175/2011JCLI4155.1.
- Scaife, A. A., et al., 2014: Skillful long-range prediction of European and North American winters. *Geophysical Research Letters*, **41** (7), 2514–2519, doi:10.1002/2014GL059637.
- Schmidli, J., C. M. Goodess, C. Frei, M. R. Haylock, Y. Hundecha, J. Ribalaygua, and T. Schmith, 2007: Statistical and dynamical downscaling of precipitation: An evaluation and comparison of scenarios for the European Alps. *Journal of Geophysical Research: Atmospheres*, **112** (D4), n/a–n/a, doi:10.1029/2005JD007026.
- Schoof, J. and S. Pryor, 2001: Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks. *International Journal of Climatology*, **21** (7), 773–790, doi:10.1002/joc.655.
- Shaman, J. and E. Tziperman, 2011: An atmospheric teleconnection linking ENSO and Southwestern European precipitation. *Journal of Climate*, **24** (1), 124–139, doi:10.1175/2010JCLI3590.1.
- Shao, Q. and M. Li, 2013: An improved statistical analogue downscaling procedure for seasonal precipitation forecast. *Stochastic Environmental Research and Risk Assessment*, **27** (4), 819–830, doi:10.1007/s00477-012-0610-0.
- Sharma, D., A. Das Gupta, and M. S. Babel, 2007: Spatial disaggregation of bias-corrected GCM precipitation for improved hydrologic simulation: Ping River basin, Thailand. *Hydrology and Earth System Sciences*, **11** (4), 1373–1390, doi:10.5194/hess-11-1373-2007.

- Shi, W., N. Schaller, D. MacLeod, T. N. Palmer, and A. Weisheimer, 2015: Impact of hindcast length on estimates of seasonal climate predictability. *Geophysical Research Letters*, **42** (5), 1554–1559, doi:10.1002/2014GL062829.
- Shukla, J., et al., 2000: Dynamical seasonal prediction. *Bulletin of the American Meteorological Society*, **81** (11), 2593–2606, doi:10.1175/1520-0477(2000)081<2593:DSP>2.3.CO;2.
- Silva, G. A. M. and D. Mendes, 2013: Comparison results for the CFSv2 hindcasts and statistical downscaling over the northeast of Brazil. *Advances in Geosciences*, **35**, 79–88, doi:10.5194/adgeo-35-79-2013.
- Singh, A., N. Acharya, U. C. Mohanty, A. W. Robertson, and G. Mishra, 2012: On the predictability of Indian summer monsoon rainfall in general circulation model at different lead time. *Dynamics of Atmospheres and Oceans*, **58**, 108–127, doi:10.1016/j.dynatmoce.2012.09.004.
- Sinha, P., U. C. Mohanty, S. C. Kar, S. K. Dash, A. W. Robertson, and M. K. Tippett, 2013: Seasonal prediction of the Indian summer monsoon rainfall using canonical correlation analysis of the NCMRWF global model products. *International Journal of Climatology*, **33** (7), 1601–1614, doi:10.1002/joc.3536.
- Smith, D. M., R. Eade, and H. Pohlmann, 2013: A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. *Climate Dynamics*, **41** (11–12), 3325–3338, doi:10.1007/s00382-013-1683-2.
- Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore, 2008: Improvements to NOAA’s historical merged land–ocean surface temperature analysis (1880–2006). *Journal of Climate*, **21** (10), 2283–2296, doi:10.1175/2007JCLI2100.1.
- Sohn, S. J., C. Y. Tam, and J. B. Ahn, 2013: Development of a multimodel-based seasonal prediction system for extreme droughts and floods: Case study for South Korea. *International Journal of Climatology*, **33** (4), 793–805, doi:10.1002/joc.3464.
- Sokol, Z., 2003: MOS-based precipitation forecasts for river basins. *Weather and Forecasting*, **18** (5), 769–781, doi:10.1175/1520-0434(2003)018<0769:MPFFRB>2.0.CO;2.
- Sterl, A., 2004: On the (in)homogeneity of reanalysis products. *Journal of Climate*, **17** (19), 3866–3873, doi:10.1175/1520-0442(2004)017<3866:OTIORP>2.0.CO;2.
- Stockdale, T., et al., 2010: Understanding and predicting seasonal-to-interannual climate variability: The producer perspective. *Procedia Environmental Sciences*, **1**, 55–80, doi:10.1016/j.proenv.2010.09.006.

- Stockdale, T. N., D. L. T. Anderson, J. O. S. Alves, and M. Balmaseda, 1998: Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. *Nature*, **392** (6674), 370–373, doi:10.1038/32861.
- Stone, M., 1974: Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **36** (2), 111–147, doi:10.2307/2984809.
- Straus, D., J. Shukla, D. Paolino, S. Schubert, M. Suárez, P. Pegion, and A. Kumar, 2003: Predictability of the seasonal mean atmospheric circulation during autumn, winter, and spring. *Journal of Climate*, **16** (22), 3629–3649, doi:10.1175/1520-0442(2003)016<3629:POTSMA>2.0.CO;2.
- Sun, J. and H. Chen, 2012: A statistical downscaling scheme to improve global precipitation forecasting. *Meteorology and Atmospheric Physics*, **117** (3-4), 87–102, doi:10.1007/s00703-012-0195-7.
- Teng, J., N. J. Potter, F. H. S. Chiew, L. Zhang, J. Vaze, and J. P. Evans, 2014: How does bias correction of RCM precipitation affect modelled runoff? *Hydrology and Earth System Sciences Discussions*, **11** (9), 10 683–10 724, doi:10.5194/hessd-11-10683-2014.
- Thiemeßl, M. J., A. Gobiet, and G. Heinrich, 2012: Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change*, **112** (2), 449–468, doi:10.1007/s10584-011-0224-4.
- Thiemeßl, M. J., A. Gobiet, and A. Leuprecht, 2011: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *International Journal of Climatology*, **31**, 1530–1544, doi:10.1002/joc.2168.
- Thiaw, W. M., A. G. Barnston, and V. Kumar, 1999: Predictions of African rainfall on the seasonal timescale. *Journal of Geophysical Research: Atmospheres*, **104** (D24), 31 589–31 597, doi:10.1029/1999JD900906.
- Timbal, B., A. Dufour, and B. McAvaney, 2003: An estimate of future climate change for western France using a statistical downscaling technique. *Climate Dynamics*, **20** (7-8), 807–823, doi:10.1007/s00382-002-0298-9.
- Tippett, M. K., A. G. Barnston, and S. Li, 2011: Performance of recent multimodel ENSO forecasts. *Journal of Applied Meteorology and Climatology*, **51** (3), 637–654, doi:10.1175/JAMC-D-11-093.1.

- Trenberth, K. E., D. P. Stepaniak, J. W. Hurrell, and M. Fiorino, 2001: Quality of reanalyses in the Tropics. *Journal of Climate*, **14** (7), 1499–1510, doi:10.1175/1520-0442(2001)014<1499:QORITT>2.0.CO;2.
- Trigo, R. M. and J. P. Palutikof, 2001: Precipitation scenarios over Iberia: A comparison between direct GCM output and different downscaling techniques. *Journal of Climate*, **14** (23), 4422–4446, doi:10.1175/1520-0442(2001)014<4422:PSOIIAC>2.0.CO;2.
- Tung, Y. L., C. Y. Tam, S. J. Sohn, and J. L. Chu, 2013: Improving the seasonal forecast for summertime South China rainfall using statistical downscaling. *Journal of Geophysical Research: Atmospheres*, **118** (11), 5147–5159, doi:10.1002/jgrd.50367.
- Turco, M., P. Quintana-Seguí, M. C. Llasat, S. Herrera, and J. M. Gutiérrez, 2011: Testing MOS precipitation downscaling for ENSEMBLES regional climate models over Spain. *Journal of Geophysical Research: Atmospheres*, **116** (D18), n/a–n/a, doi:10.1029/2011JD016166.
- Uppala, S. M., et al., 2005: The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, **131** (612), 2961–3012, doi:10.1256/qj.04.176.
- Van Den Dool, H. M. and Z. Toth, 1991: Why do forecasts for "near normal" often fail? *Weather and Forecasting*, **6** (1), 76–85, doi:10.1175/1520-0434(1991)006<0076:WDFNO>2.0.CO;2.
- van Oldenborgh, G. J., 2004: Assessing the skill of seasonal forecasts. Tech. rep., KNMI Research Biennial Reports, 72-77 pp.
- van Oldenborgh, G. J., M. Balmaseda, L. Ferranti, T. N. Stockdale, and D. L. T. Anderson, 2005: Evaluation of atmospheric fields from the ECMWF seasonal forecasts over a 15 year period. *Journal of Climate*, **18** (16), 3250–3269, doi:10.1175/JCLI3421.1.
- van Oldenborgh, G. J., G. Burgers, and A. K. Tank, 2000: On the El Niño teleconnection to spring precipitation in Europe. *International Journal of Climatology*, **20** (5), 565–574, doi:10.1002/(SICI)1097-0088(200004)20:5<565::AID-JOC488>3.0.CO;2-5.
- Vannitsem, S., 2008: Dynamical properties of MOS forecasts: Analysis of the ECMWF operational forecasting system. *Weather and Forecasting*, **23** (5), 1032–1043, doi:10.1175/2008WAF2222126.1.
- Vannitsem, S. and C. Nicolis, 2008: Dynamical properties of model output statistics forecasts. *Monthly Weather Review*, **136** (2), 405–419, doi:10.1175/2007MWR2104.1.

- Vellinga, M., A. Arribas, and R. Graham, 2013: Seasonal forecasts for regional onset of the West African monsoon. *Climate Dynamics*, **40** (11-12), 3047–3070, doi:10.1007/s00382-012-1520-z.
- von Storch, H., E. Zorita, and U. Cubasch, 1993: Downscaling of global climate change estimates to regional scales: An application to Iberian rainfall in wintertime. *Journal of Climate*, **6** (6), 1161–1171, doi:10.1175/1520-0442(1993)006<1161:DOGCCE>2.0.CO;2.
- Vrac, M., M. L. Stein, K. Hayhoe, and X. Z. Liang, 2007: A general method for validating statistical downscaling methods under future climate change. *Geophysical Research Letters*, **34** (18), doi:10.1029/2007GL030295.
- Wang, B., 2002: Rainy season of the Asian-Pacific summer monsoon. *Journal of Climate*, **15** (4), 386–398, doi:10.1175/1520-0442(2002)015<0386:RSOTAP>2.0.CO;2.
- Wang, B., et al., 2009: Advance and prospectus of seasonal prediction: Assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004). *Climate Dynamics*, **33** (1), 93–117, doi:10.1007/s00382-008-0460-0.
- Ward, M. N. and A. Navarra, 1997: Pattern analysis of SST-forced variability in ensemble GCM simulations: Examples over Europe and the tropical Pacific. *Journal of climate*, **10** (9), 2210–2220, doi:10.1175/1520-0442(1997)010<2210:PAOSFV>2.0.CO;2.
- Weisheimer, A. and T. N. Palmer, 2014: On the reliability of seasonal climate forecasts. *Journal of the Royal Society Interface*, **11** (96), doi:10.1098/rsif.2013.1162.
- Weisheimer, A., et al., 2009: ENSEMBLES: A new multi-model ensemble for seasonal-to-annual prediction. Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs. *Geophysical Research Letters*, **36** (21), n/a–n/a, doi:10.1029/2009GL040896.
- Widmann, M., C. S. Bretherton, and E. P. Salathe, 2003: Statistical precipitation downscaling over the Northwestern United States using numerically simulated precipitation as a predictor. *Journal of Climate*, **16** (5), 799–816, doi:10.1175/1520-0442(2003)016<0799:SPDOTN>2.0.CO;2.
- Wilby, R. L., S. Charles, E. Zorita, B. Timbal, P. Whetton, and L. Mearns, 2004: Guidelines for use of climate scenarios developed from statistical downscaling methods. Tech. rep., IPCC-TGCIA.
- Wilby, R. L., H. Hassan, and K. Hanaki, 1998: Statistical downscaling of hydrometeorological variables using general circulation model output. *Journal of Hydrology*, **205** (1-2), 1–19, doi:10.1016/S0022-1694(97)00130-3.

- Wilby, R. L. and T. M. L. Wigley, 1997: Downscaling general circulation model output: A review of methods and limitations. *Progress in Physical Geography*, **21** (4), 530–548, doi:10.1177/030913339702100403.
- Wilcke, R., T. Mendlik, and A. Gobiet, 2013: Multi-variable error correction of regional climate models. *Climatic Change*, **120** (4), 871–887, doi:10.1007/s10584-013-0845-x.
- Winkler, J. A., et al., 2011: Climate scenario development and applications for local/regional climate change impact assessments: An overview for the non-climate scientist. *Geography Compass*, **5** (6), 275–300, doi:10.1111/j.1749-8198.2011.00425.x.
- WMO, 1992: *Manual on the Global Data-processing and Forecasting System*. Secretariat of the World Meteorological Organization (WMO).
- Wu, W., et al., 2012: Statistical downscaling of climate forecast system seasonal predictions for the Southeastern Mediterranean. *Atmospheric Research*, **118**, 346–356, doi:10.1016/j.atmosres.2012.07.019.
- Yadav, R. K., D. A. Ramu, and A. P. Dimri, 2013: On the relationship between ENSO patterns and winter precipitation over North and Central India. *Global and Planetary Change*, **107**, 50–58, doi:10.1016/j.gloplacha.2013.04.006.
- Yan, L. and Y. Yu, 2012: The spring prediction barrier in ENSO hindcast experiments using the FGOALS-g model. *Chinese Journal of Oceanology and Limnology*, **30** (6), 1093–1104, doi:10.1007/s00343-012-1271-z.
- Yang, X. and T. Delsole, 2012: Systematic comparison of ENSO teleconnection patterns between models and observations. *Journal of Climate*, **25** (2), 425–446, doi:10.1175/JCLI-D-11-00175.1.
- Ying, L. and F. Ke, 2012: Improve the prediction of summer precipitation in the South-eastern China by a hybrid statistical downscaling model. *Meteorology and Atmospheric Physics*, **117** (3-4), 121–134, doi:10.1007/s00703-012-0201-0.
- Zhang, Q., J. Li, V. P. Singh, C. Y. Xu, and J. Deng, 2013: Influence of ENSO on precipitation in the East River basin, South China. *Journal of Geophysical Research: Atmospheres*, **118** (5), 2207–2219, doi:10.1002/jgrd.50279.
- Zhang, Y., Y. Qian, V. Dulière, E. P. Salathé Jr., and L. R. Leung, 2012: ENSO anomalies over the Western United States: Present and future patterns in regional climate simulations. *Climatic Change*, **110** (1-2), 315–346, doi:10.1007/s10584-011-0088-7.

- Zheng, F. and J. Zhu, 2010: Spring predictability barrier of ENSO events from the perspective of an ensemble prediction system. *Global and Planetary Change*, **72 (3)**, 108–117, doi:10.1016/j.gloplacha.2010.01.021.
- Zorita, E., J. P. Hughes, D. P. Lettemaier, and H. von Storch, 1995: Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation. *Journal of Climate*, **8 (5)**, 1023–1042, doi:10.1175/1520-0442(1995)008<1023:SCORCP>2.0.CO;2.
- Zorita, E. and H. von Storch, 1999: The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. *Journal of Climate*, **12 (8)**, 2474–2489, doi:10.1175/1520-0442(1999)012<2474:TAMAAS>2.0.CO;2.



Instituto de Física de Cantabria

Santander Meteorology Group

<http://www.meteo.unican.es>

Santander (Spain)

June 2016