

**ANÁLISIS BAYESIANO DE UN MODELO LINEAL MIXTO PARA  
MEDIR EL IMPACTO DEL CAMBIO CLIMÁTICO EN EL  
RENDIMIENTO DEL FRIJOL PARA EL AÑO 2030 A NIVEL  
MUNDIAL**

Por

Velcy Rosario Palomino Lescano

Tesis sometida en cumplimiento parcial de los requerimientos para el grado de

MAESTRÍA EN CIENCIAS

en

MATEMÁTICAS (ESTADÍSTICA)

UNIVERSIDAD DE PUERTO RICO  
RECINTO UNIVERSITARIO DE MAYAGÜEZ

Julio, 2012

Aprobada por:

---

Edgar Acuña Fernández, Ph.D.  
Miembro, Comité Graduado

---

Fecha

---

Timothy Porch, Ph.D.  
Miembro, Comité Graduado

---

Fecha

---

Dámaris Santana Morant, Ph.D.  
Presidente, Comité Graduado

---

Fecha

---

Eduardo Ortiz, Ph.D.  
Representante de Estudios Graduados

---

Fecha

---

Omar Colón Reyes, Ph.D.  
Director del Departamento

---

Fecha

Resumen de Disertación Presentado a Escuela Graduada  
de la Universidad de Puerto Rico como requisito parcial de los  
Requerimientos para el grado de Maestría en Ciencias

**ANÁLISIS BAYESIANO DE UN MODELO LINEAL MIXTO PARA  
MEDIR EL IMPACTO DEL CAMBIO CLIMÁTICO EN EL  
RENDIMIENTO DEL FRIJOL PARA EL AÑO 2030 A NIVEL  
MUNDIAL**

Por

Velcy Rosario Palomino Lescano

Julio 2012

Consejero: Dámaris Santana Morant, Ph.D.  
Departamento: Ciencias Matemáticas

Desde mitades del siglo XX se vienen observando fenómenos climáticos en nuestro planeta, como por ejemplo el aumento de la temperatura y el deshielo ártico, cambios en gran escala en las cantidades de precipitación y, manifestaciones de cambios extremos en los estados del tiempo. Es de interés el medir el impacto que tienen estos cambios, particularmente en el rendimiento de productos agrícolas. Medir este impacto es un reto en parte por la incertidumbre en la predictibilidad del cambio climático. Sin embargo existen diversos modelos del clima que describen la evolución de una serie de climas variables, como son *atmosphere-ocean general circulation models* (GCMs) que simulan modelos de climas basados en sistemas de ecuaciones diferenciales, bajo diversos supuestos de la composición de la atmósfera y otras influencias como variación de la energía solar. Tebaldi y Sansó, (2008) desarrollaron un modelo Bayesiano que combina información de datos proyectados por modelos GCMs para encontrar la distribución posterior del cambio climático en el futuro. Surge entonces la pregunta de como incorporar la distribución posterior en estudios

del impacto del clima en el rendimiento de cultivos

Encontramos estudios donde se han examinado el cambio climático para soya, maíz, cebada, trigo y arroz desde distintas perspectivas. Lobell y Field (2007) y Tebaldi y Lobell (2008) realizaron sus análisis de manera global, combinado la información a nivel mundial sin considerar el cambio individual por país. Mientras que Lobell y Field (2007) se concentraron en las temperaturas y precipitación observadas, Tebaldi y Lobell (2008) incorporaron la distribución posterior del cambio climático futuro en su análisis, re-estimando los coeficientes de regresión de su modelo utilizando *bootstrap*. Por otra parte Lobell et al. (2011) realizaron su análisis por país pero no incorporan el conocimiento de cambios futuros en el clima, se concentraron en temperatura y precipitación observadas.

En nuestro trabajo examinamos el efecto del cambio climático en el rendimiento del frijol (*Phaseolus vulgaris* L.). Este análisis fue desarrollado incorporando proyecciones del clima futuro por país para los periodos de cultivo del frijol representados en los calendarios de Sacks (2010). Primero modelamos el clima por país usando el método planteado por Tebaldi y Sansó (2008), con el cual encontramos la distribución posterior de los parámetros y la distribución predictiva conjunta de temperatura y precipitación. Luego, estudiamos el efecto del cambio climático de temperatura en el rendimiento del frijol mediante el análisis Bayesiano de un modelo lineal mixto en el cual los países fueron considerados efectos aleatorios. Para incorporar las proyecciones del clima futuro, la distribución predictiva del rendimiento fue calculada usando la distribución posterior de los parámetros correspondientes del modelo lineal mixto con un paso adicional que muestreó las covariables, temperatura y precipitación, de la distribución predictiva de temperatura y precipitación.

De este análisis se obtuvo que la temperatura aumentó con mayor intensidad en los años 1984-2007, con respecto a los años 1961-1984. Para los años 2007-2030 se espera un incremento de temperatura igual o mayor que el período 1984-2007. La distribución posterior de los parámetros de temperatura, y la distribución predictiva conjunta de temperatura y precipitación, pronostican que la temperatura seguirá aumentando en 42 de los 44 países en estudio. Sólo en el caso de Nepal y Perú, se pronostica una disminución en la temperatura. Mientras que para los cinco mayores productores de frijol, se encontró que, cuatro de ellos: Myanmar, Brasil, China e India esperan un aumento en la temperatura, pero la tendencia a aumentar es menos intensa que el período 1984-2007. Sólo para Estados Unidos, se espera, que la tendencia a aumentar sea mayor con respecto al período de 1984-2007. En cuanto al efecto en el rendimiento del frijol se ha observado que este depende del nivel en el que se encuentra la temperatura de un país. Países con temperaturas altas son mas sensitivos a perder en el rendimiento. Mientras que, países con temperaturas muy bajas pueden favorecerse del aumento de la temperatura, hasta cierto punto, a partir del cuál comienzan a perder. En general el aumento de temperatura tiene un impacto negativo en el rendimiento del frijol y con el continuo aumento que se espera al año 2030 las pérdidas serán mayores.

Abstract of Dissertation Presented to the Graduate School  
of the University of Puerto Rico in Partial Fulfillment of the  
Requirements for the Degree of Master of Sciences

**BAYESIAN ANALYSIS OF A LINEAR MIXED MODEL TO  
MEASURE THE IMPACT OF CLIMATE CHANGE IN THE YIELD  
OF COMMON BEAN FOR THE YEAR 2030 WORLDWIDE**

By

Velcy Rosario Palomino Lescano

Julio 2012

Chair: Dámaris Santana Morant, Ph.D.

Major Department: Mathematical Sciences

Since the middle of the 20th century, there have been ongoing and dramatic changes in the earth's climate, such as rising temperatures and loss of sea ice, changes in precipitation, and extreme weather conditions. The impact of these changes, particularly in agriculture, is of critical importance. The change in climate is a challenge, in part, because of the unpredictability of this change. However, there are models such as the atmosphere-ocean general circulation models (GCMs) that simulate future climate using differential equations, under various assumptions, based on the composition of the atmosphere and other factors like solar energy variation. Tebaldi and Sansó (2008) developed a Bayesian model that combined the projections of different GCM models to obtain a posterior distribution for future climate change. Thus, the question arises on how to incorporate this posterior distribution into a statistical analysis of the impact of climate change in the yield of cultivars.

There are studies that have examined the impact of climate change on soybean, maize, barley, wheat and rice under different climate scenarios. Lobell and Field

(2007) and Tebaldi and Lobell (2008) performed a global analysis, combining the information for the world and not considering change in individual countries. Whereas Lobell and Field (2007) used only observed weather data on temperature and precipitation, Tebaldi and Lobell (2008) incorporated the posterior distribution of future climate change in their analysis by re-estimating the regression coefficients of their model using bootstrap. Lobell et al. (2011) performed their analysis by country, but using only observed climate data and not incorporating projections of future climate.

We studied the impact of climate change in the yield of common bean (*Phaseolus vulgaris L.*). The analysis was performed by country incorporating projections of future climate for the periods of cultivation of beans presented in the calendars of Sacks (2010). We begin by modeling future climate by country using the model of Tebaldi and Sansó (2008) by finding the posterior distribution of the corresponding parameters and estimating the joint predictive distribution of temperature and precipitation. Then, we studied the impact of temperature and precipitation on the yield of bean using a Bayesian analysis of a linear mixed model in which countries were considered random effects. To incorporate the projections for future climate, the predictive distribution of yield was calculated using the posterior distribution of the corresponding parameters of the linear mixed model with an additional step that sampled the model covariates, temperature and precipitation, from the joint predictive distribution of temperature and precipitation.

From the analysis on climate, we concluded that temperature is increasing and that the increase was more intense in the years 1984-2007 compared to 1961-1984. For the years 2007-2030, the increase is expected to be equal to or greater than 1984-2007. From the joint posterior distribution of temperature and precipitation,

as well as from their joint predictive distribution, we conclude that temperature will continue to increase in 42 of the 44 countries in this study. Only for Nepal and Peru, is the temperature expected to decrease. Of the five major producers of beans, four of them (Myanmar, Brazil, China and India), have increases in temperature estimated to be less intense in 2007-2030 than it was in 1984-2007. But for the United States, the increase is estimated to be greater. From the analysis of yield and climate, we observed that the impact of climate on yield depends on the level of temperature that the country has to begin with. Countries with higher temperatures are more sensitive and have greater reduction in yield with temperature increases, whereas yields in countries with lower temperatures could be favored with increases in temperatures; this would occur only to a certain point, at which these countries too will experience reductions in yield. In general, the increase in temperature has a negative impact in the yield of beans and since temperatures are expected to keep increasing, losses in yields in 2030 could be much higher.

Copyright © 2012

por

Velcy Rosario Palomino Lescano

A mis padres Lourdes y Marcelino por su infinito amor y confianza.

A mi esposo John por su amor, paciencia y comprensión.

A mis hermanos por el apoyo y confianza que me brindan.

## AGRADECIMIENTOS

A Dios por ser mi fortaleza en momentos difíciles.

A la Dra. Dámaris Santana por sus importantes aportes, disposición y apoyo incondicional para el desarrollo del presente trabajo.

Al Dr. Timothy Porch por habernos presentado el problema que motiva este trabajo, por sus sugerencias en el desarrollo y conclusión del mismo.

Al Dr. Edgar Acuña por brindarnos las facilidades para el uso de su laboratorio y por sus aportes para la versión final del trabajo.

Al Dr. George Casella por su sugerencia de usar la distribución predictiva del rendimiento como medio para aprender sobre el rendimiento y para incorporar la distribución de las proyecciones del clima.

Al Dr. Raúl Macchiavelli por su contribución en el entendimiento de los modelos lineales mixtos.

Al Departamento de Ciencias Matemáticas del Recinto Universitario de Mayagüez, por el apoyo brindado para cursar mis estudios de Maestría.

A Ysela, Walter, Héctor, Roberto, Charlie y Fabián, y a todos mis amigos que confían en mí.

## Índice general

	<u>página</u>
RESUMEN EN ESPAÑOL . . . . .	II
ABSTRACT ENGLISH . . . . .	V
AGRADECIMIENTOS . . . . .	X
Índice de cuadros . . . . .	XIII
Índice de figuras . . . . .	XIV
1. INTRODUCCIÓN . . . . .	1
1.1. Justificación . . . . .	1
1.1.1. El frijol . . . . .	4
1.2. Publicaciones previas . . . . .	7
1.3. Objetivos . . . . .	10
2. REVISIÓN DE LITERATURA . . . . .	12
2.1. Análisis Bayesiano . . . . .	12
2.1.1. Modelo Bayesiano Jerárquico . . . . .	14
2.2. Métodos de Cadenas de Markov Monte Carlo (MCMC) . . . . .	16
2.2.1. Gibbs Sampling . . . . .	17
2.2.2. Metropolis-Hastings . . . . .	18
2.3. Modelos Longitudinales . . . . .	18
2.3.1. Modelo Lineal Mixto . . . . .	21
3. MODELO DEL CAMBIO DE TEMPERATURA Y PRECIPITACIÓN . . . . .	25
3.1. Descripción de los Datos Climáticos . . . . .	25
3.2. Tendencias en el Clima Observado . . . . .	27
3.3. Modelando la Distribución Conjunta de Temperatura y Precipitación Incorporando Modelos de Proyección (GCM) . . . . .	33
3.3.1. Determinación de las Distribuciones a Priori y a Posteriori . . . . .	35
3.3.2. Gibbs Sampler Para la Distribución Posterior . . . . .	37
3.3.3. Distribución Predictiva . . . . .	39
3.4. Tendencias en el Clima Proyectado . . . . .	40
3.4.1. Proyecciones del Clima Usando la Distribución Posterior . . . . .	41
3.4.2. Comparación de las Tendencias Observadas con las Tendencias Futuras . . . . .	47
3.4.3. Proyecciones del Clima Usando la Distribución Predictiva . . . . .	51

4.	MODELO DEL RENDIMIENTO . . . . .	56
4.1.	Descripción de los Datos de Rendimiento . . . . .	56
4.2.	Modelo del Rendimiento en Respuesta a la Variación del Clima . . . . .	57
4.2.1.	Modelo Lineal Mixto Para Datos Observados . . . . .	61
4.3.	Rendimiento Incorporando la Distribución Predictiva de Temperatura y Precipitación . . . . .	68
4.4.	Resultados del Modelo 1 . . . . .	69
4.5.	Resultados del Modelo 2 . . . . .	75
5.	CONCLUSIONES Y TRABAJOS FUTUROS . . . . .	82
	APÉNDICES . . . . .	85
A.	FIGURAS . . . . .	86
A.1.	Distribución predictiva de temperatura y precipitación para el período actual y futuro . . . . .	86
A.2.	Series de Temperatura y precipitación para los modelos proyectados, observados, tendencias centrales y distribución predictiva . . . . .	96
B.	TABLAS . . . . .	106
C.	DISTRIBUCIONES CONDICIONALES Y FORMULAS . . . . .	108
C.1.	Cálculo de las distribuciones condicionales de los parámetros . . . . .	108
C.2.	Aplicación de la frmula de Schur . . . . .	116

<u>Tabla</u>	Índice de cuadros	<u>página</u>
B-1. Países productores de frijol . . . . .		107

## Índice de figuras

<u>Figura</u>		<u>página</u>
3-1.	Correspondencia entre datos de temperatura y la temporada de crecimiento para Ecuador . . . . .	26
3-2.	Series del promedio anual de precipitación 1961-2007 . . . . .	28
3-3.	Series del promedio anual de precipitación 1961-2007 (Continuación de la Figura 3-2 . . . . .	28
3-4.	Series del promedio anual de temperatura 1961-2007 . . . . .	30
3-5.	Series del promedio anual de temperatura 1961-2007 (Continuación Figura 3-4) . . . . .	30
3-6.	Tendencia de temperatura y precipitación 1961-1984 vs 1984-2007 . . . . .	31
3-7.	Cambio de temperatura y precipitación 1961-1984 vs 1984-2007 . . . . .	33
3-8.	Intervalo de credibilidad del 95 % de la tendencia actual y futura de temperatura en la temporada del crecimiento del frijol . . . . .	42
3-9.	Intervalo de credibilidad del 95 % de la tendencia actual y futura de precipitación en la temporada del crecimiento del frijol . . . . .	43
3-10.	Intervalo de credibilidad del 95 % de la tendencia futura de temperatura y precipitación . . . . .	44
3-11.	Intervalo de credibilidad del 95 % de los coeficientes de correlación . . . . .	45
3-12.	Intervalo de credibilidad del 95 % de los coeficientes de correlación (Continuación de la Figura 3-11) . . . . .	46
3-13.	Intervalo de credibilidad del 95 % de los coeficientes de correlación (Continuación de la Figura 3-12) . . . . .	46
3-14.	Intervalo de credibilidad del 95 % de los coeficientes de correlación (Continuación de la Figura 3-13) . . . . .	47
3-15.	Comparación de las tendencias por períodos . . . . .	48
3-16.	Gráfico del cambio de temperatura de 1961-1984 vs 1984-2007 y 1984-2030 vs 2007-2030 . . . . .	49

3-17.	Gráfico del cambio de precipitación de de 1961-1984 vs 1984-2007 sin Nepal . . . . .	50
3-18.	Gráfico del cambio de precipitación de de 1961-1984 vs 1984-2007 . . . . .	51
3-19.	Distribución conjunta para Togo, Uganda, USA y Venezuela . . . . .	53
3-20.	Series de tiempo de temperatura y precipitación para Togo, Uganda, USA y Venezuela . . . . .	54
4-1.	Series del promedio anual de rendimiento de frijol de 1961-2007 . . . . .	57
4-2.	Series del promedio anual de rendimiento (hg/ha) - temperatura (°C) de frijol de 1961-2007 . . . . .	58
4-3.	Series del promedio anual de rendimiento - temperatura eliminando la magnitud de temperatura . . . . .	59
4-4.	Series del promedio anual de rendimiento - precipitación de frijol de 1961-2007 . . . . .	60
4-5.	Series del promedio anual de rendimiento - precipitación eliminando la magnitud de precipitación . . . . .	60
4-6.	Distribución de los coeficientes de temperatura ( <i>d</i> ) y precipitación ( <i>e</i> ) en el modelo del rendimiento . . . . .	71
4-7.	Intervalo de credibilidad del 95 % para el cambio en el rendimiento medio debido cambio en temperatura, estimado por el modelo I . . . . .	73
4-8.	Intervalo de credibilidad del 95 % para el cambio en el rendimiento medio debido al incremento en precipitación estimado por el modelo I . . . . .	74
4-9.	Intervalo de credibilidad del 95 % para el cambio en el rendimiento medio debido al cambio en el clima estimado por el modelo I . . . . .	75
4-10.	Intervalo de credibilidad del 95 % para el cambio en el rendimiento medio debido al cambio en temperatura, estimado por el modelo II . . . . .	77
4-11.	Intervalo de credibilidad del 95 % para el cambio en el rendimiento medio debido al cambio en precipitación, estimado por el modelo II . . . . .	78
4-12.	Intervalo de credibilidad del 95 % para el cambio en el rendimiento medio debido al cambio climático, estimado por el modelo II . . . . .	79
4-13.	Efecto del cambio de temperatura en climas extremos, cuando $\Delta = 1, 2, 3, 4^{\circ}C$ . . . . .	80

4-14. Efecto del cambio de precipitación en climas extremos cuando $\Delta =$ 1, 2, 3, 4 mm. . . . .	81
A-1. Distribución conjunta para Albania, Argentina, Australia y Bangladesh	86
A-2. Distribución conjunta para Benin, Brasil, Bulgaria y Cameroon . . .	87
A-3. Distribución conjunta para Canada, Chile, China y Colombia . . .	88
A-4. Distribución conjunta para Costa Rica, Ecuador, Egipto y El Salvador	89
A-5. Distribución conjunta para Francia, Grecia, Guatemala y Honduras	90
A-6. Distribución conjunta para India, Iraq, Japon y Korea del Sur . . .	91
A-7. Distribución conjunta para Madagascar, Mexico, Myanmar y Nepal	92
A-8. Distribución conjunta para Holanda, Nicaragua, Pakistan y Panama	93
A-9. Distribución conjunta para Perú, Filipinas, República Dominicana y Romania . . . . .	94
A-10. Distribución conjunta para Africa del Sur, Sri Lanka, Syria y Tailandia	95
A-11. Series de tiempo de temperatura y precipitación para Albania, Ar- gentina, Australia y Bangladesh . . . . .	96
A-12. Series de tiempo de temperatura y precipitación para Benin, Brasil, Bulgaria y Camerún . . . . .	97
A-13. Series de tiempo de temperatura y precipitación para Canada, Chile, China, Colombia . . . . .	98
A-14. Series de tiempo de temperatura y precipitación para Costa Rica, Ecuador, Egipto y El Salvador . . . . .	99
A-15. Series de tiempo de temperatura y precipitación para Francia, Gre- cia, Guatemala y Honduras . . . . .	100
A-16. Series de tiempo de temperatura y precipitación para India, Iraq, Japón y Korea del Sur . . . . .	101
A-17. Series de tiempo de temperatura y precipitación para Madagascar, Mexico, Myanmar y Nepal . . . . .	102
A-18. Series de tiempo de temperatura y precipitación para Holanda, Ni- caragua, Pakistan y Panama . . . . .	103
A-19. Series de tiempo de temperatura y precipitación para Perú, Filipinas, República Dominicana y Romania . . . . .	104

A-20. Series de tiempo de temperatura y precipitación para Africa del Sur,  
Sri Lanka, Syria y Tailandia . . . . . 105

# Capítulo 1

## INTRODUCCIÓN

### 1.1. Justificación

Desde mitades del siglo XX se vienen observando fenómenos climáticos en nuestro planeta, como por ejemplo el aumento de la temperatura y el deshielo ártico, cambios en gran escala en las cantidades de precipitación y, manifestaciones de cambios extremos en los estados del tiempo. Estudios basados en los últimos records observados indican que la temperatura promedio de la tierra se ha incrementando al rededor de  $0.7\text{ }^{\circ}\text{C}$  desde 1906 al 2005 (IPCC, 2007), mostrando mayor calentamiento en los últimos 50 años. El más reciente estudio (Rowlands et al., 2012) estima cambios de  $1.4\text{-}3.0\text{ }^{\circ}\text{C}$  para el 2050. Muchos de estos cambios se atribuyen a la actividad humana, dado que el hombre está realizando actividades que provocan cambios climáticos en el planeta como: la emisión de dióxido de carbono, óxido nitroso, gas metano y otros contaminates que combinan el uso de tierras como la deforestación. Se proyecta que la concentración de gases seguirá en aumento (Lobell y Burke, 2009).

IPCC (2007) estima que un cambio de  $2^{\circ}\text{C}$  tendría un efecto catastrófico en la agricultura. No obstante la disponibilidad de alimentos abarca cuestiones globales y regionales, en general a nivel mundial la cantidad de alimentos que se produce es más que suficiente. Sin embargo hay discrepancias con respecto a la disponibilidad de alimentos a nivel regional, es así que otros estudios indican que la necesidad de alimentos para el 2050 será mayor, ya que el cambio climático puede contrarrestar

el incremento que puede haber en el rendimiento (producción total de un cultivo cosechado, por área de terreno utilizada), debido a otros factores como: tecnología, fertilización, de manera que existen regiones dónde el incremento del rendimiento puede hasta triplicarse y otras en dónde el aumento es más dificultoso como el caso de Africa y Asia (Lobell y Burke, 2009), donde se considera que el efecto del clima sería mayor y mientras la población sigue en aumento la disponibilidad de alimentos para estas regiones puede ser afectada sustancialmente.

Durante el desarrollo de los cultivos, la temperatura puede afectar muchos aspectos del crecimiento, la velocidad con la que absorben agua y nutrientes, la expansión de las hojas, la calidad del cultivo y el rendimiento (Wild, 1992). Sin embargo, cambios en la temperatura pueden tener efectos positivos o negativos sobre el rendimiento de los cultivos, la diferencia depende de la ubicación y magnitud de dichos cambios (Reilly et al., 1999). Por ejemplo, algunos cultivos pueden resultar beneficiados al existir menores heladas, no obstante las altas temperaturas durante la fase de reproducción de muchos cultivos (como caupí, frijol, maíz, arroz, trigo y sorgo) producen una reducción en el rendimiento por el aborto de órganos reproductivos, como botones, flores y vainas (Porch, 2001; Konsens and Kigel, 1992). También, las altas temperatura pueden facilitar el aumento severo de insectos e incrementar los efectos de enfermedades en las plantas, agrabando el riesgo de la pérdida de las plantaciones. Así mismo, cambios en la precipitación afectan directamente la humedad del suelo y por lo tanto, la producción de alimentos, en este sentido el rendimiento de cultivos agrícolas está fuertemente relacionado con la temperatura y la precipitación.

Medir el impacto del cambio del clima en el rendimiento es un problema de debate en el mundo (Reilly y Schimmelpfenning, 2008). En parte por la falta de

certeza en modelos de predictibilidad del cambio climático, y porque el entendimiento del efecto de los cambios climáticos pasados en la agricultura aún no es muy clara. Entender cómo los cambios climáticos afectaron el rendimiento en la agricultura hasta la fecha y conocer el comportamiento del cambio del clima ayudarían a predecir el impacto de estos cambios en el rendimiento futuro. Existen diversos modelos de clima, desde modelos simples que aproximan sólo la trayectoria global de la temperatura hasta modelos complejos (Clausser et al., 2002), que son los llamados *atmosphere-ocean general circulation models* (GCMs).

Los GCMs describen la evolución de una serie de climas variables. Estos son modelos de simulación basados en sistemas de ecuaciones diferenciales, bajo diversos supuestos de la composición de la atmósfera y otras influencias como variación de la energía solar. Estos modelos han logrado explicar las variaciones en los climas pasados, y se utilizan para simular las variaciones futuras en el clima, suponiendo diversas alteraciones de la atmósfera, principalmente por la emisión de gases de efecto invernadero, y otras sustancias producidas por la actividad humana, como los aerosoles de sulfato, que se sabe que influyen en el clima. A las distintas composiciones de la atmósfera, se les conoce como escenarios climáticos, tres de ellos por ejemplo: SRESA2, SRESA1B y SRESA1A. A las simulaciones futuras, se les conoce como proyecciones, que son una herramienta importante, en la evaluación de la influencia de la actividad humana en el clima. Una reconocida fuente de referencia para los modelos climáticos son, los informes de Evaluación del Grupo Intergubernamental de Expertos sobre el cambio climático, el más reciente es el IPCC, 2007. Sin embargo, diferentes modelos varían sustancialmente en sus proyecciones (Smith et al., 2008) y no hay manera de clasificarlos como mejores o peores. Por lo que surge la pregunta de cómo estos modelos pueden ser combinados para tener una distribución de probabilidad del cambio climático en el futuro. Un enfoque es considerar

los modelos como iguales y trabajar con promedios, el otro surge de que no todos los modelos son igualmente confiables y que algunos deben ser ponderados más que otros (Giorgi y Means, 2002 ; Greene et al., 2006 ; Furrer et al. 2007 ). Siguiendo el segundo enfoque, Tebaldi y Sansó (2008) estiman la distribución conjunta, mediante un análisis Bayesiano que combina los diferentes modelos (GCMs) de clima. Tebaldi y Lobell (2008) combinan el análisis de Tebaldi y Sansó con un modelo de regresión y usando *bootstrap* miden el impacto del cambio de temperatura promedio global en el rendimiento global de tres cultivos: maíz, trigo y cebada.

En nuestro trabajo estudiamos el impacto de los cambios climáticos en el frijol. Entendiendo que el cambio climático varía según regiones geográficas, y que el efecto de este cambio en el rendimiento depende de las condiciones climáticas de dichas regiones, en este estudio utilizamos el modelo Bayesiano de Tebaldi y Sansó (2008) para derivar la distribución conjunta de temperatura y precipitación en los países productores de frijol, en vez de hacer un análisis global. Evaluamos los efectos de las temperaturas y precipitación proyectadas por nuestro análisis en el del rendimiento del frijol para los años 2007-2030, usando análisis Bayesiano de un modelo lineal mixto.

### **1.1.1. El frijol**

Debido al alto porcentaje de proteínas en las leguminosas, las semillas de leguminosas han constituido el complemento más utilizado para aumentar el contenido de proteínas de las raciones de alimentos y son un alimento básico mundial especialmente en países pobres (Reyes y Padilla, 2008). Dentro de las leguminosas, el frijol es una fuente importante de proteínas e hidratos, llegando a comprender alrededor del 20-25 % de su peso en cuanto a proteínas, por lo que trasciende como fuente de alimento y sustituto de otros nutrientes. Según la Food and Agriculture Organization

(FAO statistical database, 2006, <http://faostat.fao.org>), la producción de frijol en el mundo se concentra en 129 países de los cinco continentes. Entre 1961-2007 se produjo en promedio casi 15 millones de toneladas al año. Entre los países productores de frijol destacan; India con 19.0 %, Brasil con 16.6 %, Myanmar con 13.6 % (país que muestra mayor tasa de crecimiento en los últimos años), China con 7 %, Estados Unidos con 5.6 % y México con 4.8 % ( en el caso de India y Myanmar se debe tener en cuenta que FAO combina datos de diferentes leguminosas). Estas naciones contribuyen el 66 % del total producido (FAO, 2006). Así mismo se cree que el mayor consumo de frijol se manifiesta principalmente en naciones en vías de desarrollo, dándose mayores niveles de aceptación y uso de este producto en América Latina, Asia y Africa (ITESM, 2004). Destacándose que de los trece países de mayor consumo de la leguminosa en el mundo, ocho de ellos se encuentran en América Latina: Nicaragua, Brasil, México, Paraguay, Belice, Costa Rica, Guatemala y Honduras, lo que confirma la relación entre los niveles de consumo y los ingresos percapita de países menos y más desarrollados y a quienes el cambio climático podría estar afectando directamente. Además el consumo de frijoles tiene amplias posibilidades de incremento dada la difusión que se está realizando para producir su ingesta, con la finalidad de prevenir los riesgos de enfermedades. Se ha determinado que el frijol no sólo suministra proteínas y carbohidratos también tiene cantidades importantes de vitaminas y minerales. Salinas et al. (2005) destaca la presencia de las antocianinas indispensables en la prevención de enfermedades en especial aquellas asociadas al cáncer de colon o la obstrucción de arterias coronarias. El frijol está también relacionado a la prevención de la arteresclerosis y las inflamaciones intestinales, gracias a los altos contenidos de fibra, ácido fólico y proteína de origen vegetal (Ortiz, 1998).

Por otro lado alimentos con altos índices de glicémico causan una subida de peso más rápida, además de que estimulan a las personas a comer más pronto de

lo normal. El exceso de grasa corporal conduce a elevadas concentraciones de glucosa (hiperglucemia) e insulina (hiperinsulinemia) y estudios recientes sugieren que elevadas concentraciones de estas contribuyen al desarrollo de ciertos cánceres. Los frijoles son alimentos con baja concentración de glicémico y son buenos reguladores de glucosa e insulina (Bennik, 2003).

El ciclo biológico del frijol puede variar entre 80 y 180 días, dicho lapso se encuentra determinado sobre todo por el genotipo de la variedad, hábito de crecimiento, clima, suelo, radiación solar y fotoperiodo (Ortiz, 1998). El frijol comprende dos fases: la fase vegetativa y la fase reproductiva. La fase vegetativa se inicia cuando se le brindan a la semilla las condiciones para iniciar la germinación, y termina cuando aparecen los primeros botones florales o los primeros racimos. En esta fase se desarrolla la estructura vegetativa necesaria para iniciar la actividad reproductiva de la planta. La fase reproductiva, por su parte, está comprendida entre la aparición de los primeros botones florales o racimos y la madurez de cosecha. El frijol es una planta que se cultiva en zonas tropicales y regiones templadas así que esto permite ubicarle en las especies termófilas, dado que no soportan bajas temperaturas (Debouck e Hidalgo, 1985). El rango de temperatura promedio del frijol varía entre 15 y 27°C, sin embargo, las temperaturas óptimas dependen de la fase de cultivo en la que se encuentra (Arias et al., 2007). Durante la germinación la temperatura óptima está entre 20 y 30°C, temperaturas por debajo de 10°C causan daño, además temperaturas inferiores tardan la maduración, mientras que temperaturas altas causan aceleración, las temperaturas extremas (5°C o 40°C) pueden ser soportadas por períodos cortos, pero por tiempos prolongados causan daños irreversibles (White, 1985). Sin embargo la variación que presenta el frijol en los niveles de producción entre un año y otro se corresponde con la presencia de lluvias, ya que una proporción significativa se obtiene bajo condiciones de temporal (ITESM, 2004). Durante

la etapa de floración son convenientes 110 -180 mm, las épocas más críticas por la necesidad de agua son 15 días antes de la floración y 18-22 días antes de la maduración de las primeras vainas. Los 15 días previos a la cosecha, deberían ser secos (Benacchio, 1982).

Es así que los diversos estudios demuestran la enorme importancia del estudio del frijol en razón de su alto valor nutricional y las cualidades saludables de quienes lo consumen, se trata de un producto con amplia relevancia social en estratos de bajos ingresos y de trascendencia económica para quienes lo cultivan. Sin embargo, las altas temperatura durante el período de reproducción, causan pérdidas significativas en el rendimiento (Hall, 1992). En adición, el frijol es más sensible a cambios en temperatura mínima, así que el cambio climático asimétrico (la temperatura mínima está subiendo más rápido que la máxima) puede causar más daño aún en este cultivo.

En Puerto Rico, Ramírez et al. (2011), Wasonga et al. (2010), Porch et al. (2010, 2009) y Urrea y Porch (2010) han estudiado los efectos de altas temperaturas y la sequía en la producción de frijoles y trabajan en el desarrollo de algunas variedades tolerantes a cambios climatológicos.

## **1.2. Publicaciones previas**

Tebaldi et al. (2005) mediante un enfoque Bayesiano determinaron la función de distribución de probabilidad del cambio de temperatura en una escala regional, combinando 9 modelos GCMs bajo dos escenarios climáticos de SRES para 22 regiones. Estas regiones fueron construidas sobre estándares sub continentales de  $10^7 Km^2$  propuestas por Giorgi y Mearns (2002). La distribución posterior del cambio de temperatura incorpora el criterio de sesgo en los pesos implícitamente asignados a cada modelo y la correlación entre los datos observados y los datos proyectados. La

función de distribución se determinó asumiendo independencia entre temperatura y precipitación. La forma de la función de distribución de probabilidad del cambio de temperatura futura varió ampliamente, mostrando además modas múltiples y colas largas. Años más tarde, Tebaldi et al. (2008) propone 2 versiones de este modelo: un análisis univariado en el que cada región es analizada separadamente, y un análisis multivariado que combina las 22 regiones. Cada uno de estos modelos fueron validados mediante el uso de validación cruzada. Por otro lado Tebaldi y Sansó (2008) mediante un modelo Bayesiano jerárquico determinaron la distribución conjunta para las proyecciones de los cambios de las tendencias en temperatura y precipitación, en términos de las tendencias actuales y las tendencias futuras. El resultado son funciones bivariantes ya que se considera que la temperatura y precipitación están correlacionadas.

En nuestro trabajo consideramos como regiones aquellos países productores de frijol, y las variables precipitación y temperatura como dependientes, por lo que realizaremos un análisis por país incorporando correlación entre temperatura y precipitación, siguiendo el modelo estimado por Tebaldi y Sansó (2008).

Como mencionamos antes los cambios climáticos afectan el rendimiento de los cultivos. Lobell y Field (2007) estimaron las tendencias del cambio del rendimiento global debido a la variación anual del clima para los cultivos de soya, maíz y cebada. La temperatura y precipitación fueron promediadas sobre las áreas de cultivo y según las temporadas de crecimiento para cada cultivo. Como el rendimiento, la temperatura y precipitación varían en el tiempo eliminaron la tendencia, utilizando las primeras diferencias. Finalmente aplicaron un modelo de regresión a estas diferencias, estimando así el efecto de la variación de temperatura y precipitación en el cambio del rendimiento. Los resultados demostraron efectos negativos al incremento

de temperatura.

Por su parte Tebaldi y Lobell (2008) estudiaron el impacto en el rendimiento global del cambio de temperatura y precipitación promedio, en las regiones donde se cultivan maíz, trigo y cebada incorporando proyecciones para el clima de GCMs. Usando el método empleado en Tebaldi y Sansó (2008) encontraron la distribución posterior conjunta del cambio de las tendencias en temperatura promedio y precipitación para un escenario dado (SRESA2). Luego muestrearon de la distribución posterior conjunta de las diferencias en el clima y reemplazaron estas diferencias en las variables predictoras del modelo estadístico de cultivo propuesto por Lobell y Field (2007) lo que genera una muestra de valores para el rendimiento. Para incorporar la incertidumbre entre variaciones del clima y la variación de los cultivos, utilizaron el método de *bootstrap* en los valores del rendimiento, y reestimaron los coeficientes de regresión.

Posteriormente Lobell et al. (2011) cambian de un análisis global a uno por país, estudiando el impacto de los cambios climáticos en la producción de maíz, soya, arroz y trigo, para países con más de 10000 hectáreas de cultivo para el período comprendido desde 1980 hasta el 2008. El rendimiento anual para cada cultivo lo obtuvieron de la página web de *Food and Agriculture Organization of the United Nations*(FAO <http://faostat.fao.org/>). Luego extrajeron la temperatura y precipitación mensual de la base de datos de la Universidad de Delaware (<http://climate.geog.udel.edu/climate>). Para cada cultivo: arroz, trigo, maíz y soya, modelaron el rendimiento en función de la temperatura y precipitación mediante un análisis de datos longitudinales con interceptos para cada país. En este análisis muestran que la producción comienza a disminuir en cuanto un país está más caliente. En particular un aumento de la temperatura en 1 °C tiende a bajar el rendimiento en hasta 10 %, excepto en

países con gran latitud. Por otro lado, el arroz resulta ser beneficiado por el calentamiento y la precipitación incrementa su producción hasta cierto punto, el incremento de la lluvia resulta ser perjudicial a partir de un valor. En general el maíz y el trigo percibieron un impacto negativo, el impacto neto en arroz y soya no fue significativo con ganancia en algunos países y pérdida en otros.

Notamos entonces que se ha examinado el efecto en el cambio climático para soya, maíz, cebada, trigo y arroz. Lobell y Field (2007) y Tebaldi y Lobell (2008) realizaron sus análisis de manera global, no por país. Mientras que Lobell y Field (2007) se concentraron en las temperaturas y precipitación observadas, Tebaldi y Lobell (2008) incorporaron en su análisis la distribución conjunta para cambios futuros en temperatura y precipitación y estimaron su impacto re-estimando los coeficientes de regresión de su modelo utilizando *bootstrap*. Por otra parte Lobell et al. (2011) realizan su análisis por país pero no incorporan el conocimiento de cambios futuros en el clima, se concentraron en temperaturas y precipitación observadas.

En este trabajo examinaremos el efecto del cambio climático para el frijol. Incorporamos el conocimiento que ganamos sobre las temperaturas y precipitaciones futuras por país y calculamos la distribución predictiva del rendimiento usando análisis Bayesiano de un modelo lineal mixto.

### 1.3. Objetivos

- Crear una base de datos por país con rendimiento para frijol, temperatura y precipitación correspondientes a su época de cultivo para los años 1961-2007.
- Medir el impacto de los cambios climáticos en el rendimiento del frijol desde 1961-2007 para los países en el mundo.

- Encontrar la distribución conjunta del cambio de temperatura y precipitación para el 2030 por país.
- Predecir la distribución del impacto del cambio climático en el rendimiento del frijol para el 2030.
- Estimar el impacto medio del cambio de temperatura y precipitación en el rendimiento del frijol para el 2030.

## Capítulo 2

# REVISIÓN DE LITERATURA

### 2.1. Análisis Bayesiano

En el análisis bayesiano, además de especificar el modelo de los datos observados  $x = (x_1, x_2, \dots, x_n)$ , dado un vector de parámetros desconocidos  $\theta \in \Theta$ ,  $f(x | \theta)$  se considera que  $\theta$  es una variable aleatoria que tiene una distribución a priori  $\pi(\theta)$ . La inferencia concerniente a  $\theta$  está basada en su distribución a posteriori, dada por,

$$\pi(\theta | x) = \frac{f(x | \theta) \pi(\theta)}{\int f(x | \theta) \pi(\theta) d\theta} \quad (2.1)$$

donde  $\int f(x | \theta) \pi(\theta)$  es la distribución marginal de  $x$ , por lo que la ecuación(2.1) puede ser escrita en forma proporcional como:

$$\pi(\theta | x) \propto f(x | \theta) \pi(\theta) \quad (2.2)$$

La ecuación (2.1) es conocida como teorema de Bayes. Las inferencias en estadística Bayesiana se basan en  $\pi(\theta | x)$  en lugar de  $f(x | \theta)$ .

#### Distribución a Priori $\pi(\theta)$

La distribución a priori cumple un papel importante en el análisis bayesiano ya que mide el grado de conocimiento inicial que se tiene de los parámetros en estudio antes de recolectar el conjunto de datos. En la práctica es raro que este grado de conocimiento inicial guie la determinación de la distribución a priori que se considere correcta, pues es posible que varias distribuciones a priori sean compatibles con la información que se tiene. Así será posible plantear tantas distribuciones iniciales

como estados iniciales de conocimiento existan y los diferentes resultados obtenidos en la distribución a posteriori bajo cada uno de los enfoques, adquirirán una importancia en relación con la convicción que tenga el investigador sobre cada estado inicial. El uso de una u otra distribución a priori tendrá influencia en la inferencia (Robert, 2007). Cuando poco o nada es conocido sobre los parámetros de la distribución a priori su selección entonces se puede hacer de manera mecánica. Una opción es usar distribuciones a priori no informativas. Ejemplo de estas son la distribución Uniforme que asigna la misma probabilidad a cada valor de los parámetros y la distribución a priori de Jeffreys que se basa en la matriz de información de Fisher, por lo que dependerá de la forma de  $f(x|\theta)$  (Robert, 2007). La otra opción es usar distribuciones a priori conjugadas, donde se asume que la distribución a priori para  $\theta$  pertenece a una familia paramétrica de distribuciones  $\pi(\theta)$ . La cual es conjugada para la función  $f(\theta|x)$ , es decir que resulta en una distribución a posterior  $\pi(\theta|x)$  que pertenece a la misma familia de distribuciones de la distribución a priori. Los modelos Bayesianos jerárquicos (Sección 2.1.1) presentan una posible disminución del efecto en la inferencia de la distribución a priori moviendo su selección a niveles más altos en la jerarquía (Robert, 2007).

### **Distribución Posterior $\pi(\theta | x)$**

Uno de los objetivos de la inferencia Bayesiana es encontrar la distribución posterior de los parámetros. Hace algunos años, la aplicación de los métodos Bayesianos a problemas prácticos no era tan extendida como ahora, pues encontrar la distribución posterior o funciones de ésta, como por ejemplo, el valor esperado posterior, requería de métodos numéricos complicados o resultados asintóticos para resolver problemas de integración asociados a esos cálculos. Actualmente, el uso de estas herramientas de aproximación no son tan necesarias para llevar a cabo un análisis de datos Bayesiano, pero siguen siendo útiles pues proveen un marco de referencia

conceptual para la interpretación. En la Sección 2.2 se discutirá la obtención de la distribución posterior a través de métodos de simulación. Estos métodos también aplicarán a la distribución predictiva posterior.

### Distribución Predictiva Posterior

Una de las ventajas de los métodos bayesianos es que todas las cantidades desconocidas son tratadas como variables aleatorias con distribución de probabilidad. Así, el problema de predecir una nueva observación se puede pensar como el de hacer inferencia sobre una cantidad observable pero desconocida, por lo que predecir una nueva observación puede pensarse como realizar una inferencia predictiva. Si  $x$  es desconocida pero observable:

$$\begin{aligned} f(x) &= \int f(x, \theta) d\theta \\ &= \int \pi(\theta) f(x | \theta) d\theta \end{aligned} \tag{2.3}$$

Esta es la distribución marginal de  $x$  o distribución predictiva a priori. Después de que  $x$  ha sido observada, se puede predecir una cantidad desconocida  $\tilde{x}$  observable. La distribución predictiva para una nueva  $\tilde{x}$  es:

$$\begin{aligned} f(\tilde{x} | x) &= \int f(\tilde{x}, \theta | x) d\theta \\ &= \int f(\tilde{x} | \theta, x) \pi(\theta | x) d\theta \\ &= \int f(\tilde{x} | \theta) \pi(\theta | x) d\theta \end{aligned} \tag{2.4}$$

Aquí se supone que  $\tilde{x}$  y  $x$  son independientes.

#### 2.1.1. Modelo Bayesiano Jerárquico

Estos son modelos de múltiples parámetros que están relacionados o conectados por la estructura del problema que se está estudiando. Como su nombre lo indica, se supone un modelo de probabilidad conjunta para los parámetros formando una

jerarquía que refleja la dependencia entre ellos. En particular, la distribución a priori  $\pi(\theta)$ , se puede escribir a través de una jerarquía condicional de las previas.

Un modelo jerárquico tiene la siguiente estructura:

*Nivel I* (Observaciones)

$$\begin{aligned} f(x | \theta) &= f(x_1, \dots, x_J | \theta_1, \dots, \theta_J) \\ &= \prod_{j=1}^J f(x_j | \theta_j) \end{aligned} \quad (2.5)$$

*Nivel II* (Parámetros)

$$\begin{aligned} \pi(\theta | \phi) &= \pi(\theta_1, \dots, \theta_J | \phi) \\ &= \prod_{j=1}^J \pi(\theta_j | \phi) \end{aligned} \quad (2.6)$$

*Nivel III* (Hiperparámetros)

$$\pi(\phi) \quad (2.7)$$

### Tratamiento Bayesiano Para Modelos Jerárquicos

Se considera la distribución a priori conjunta  $\pi(\phi, \theta) = \pi(\theta) \pi(\theta | \phi)$ . La distribución conjunta viene dada por

$$\begin{aligned} \pi(\phi, \theta | x) &\propto \pi(\phi, \theta) \pi(x | \phi, \theta) \\ &= \pi(\phi, \theta) \pi(x | \theta) \end{aligned} \quad (2.8)$$

Ya que la distribución de los datos sólo depende de  $\theta$  dado que los hiperparámetros  $\phi$  afectan a  $x$  a través de  $\theta$ . Es usual considerar a  $\phi$  como un parámetro de molestia, por lo que típicamente se le asigna una distribución a priori no informativa.

Para encontrar las distribuciones marginales y condicionales se puede proceder de la siguiente forma:

- Escribir  $\pi(\theta, \phi | x)$  en forma no-normalizada. Esto implica calcular:

$$\pi(\theta, \phi | x) \propto \pi(\phi) \pi(\theta | \phi) \pi(x | \theta) \quad (2.9)$$

- Determinar analíticamente  $\pi(\theta | \phi, x)$  dados los hiperparámetros  $\phi$ .
- Determinar  $\phi$  marginalizando la distribución posteriori. Esto implica hallar la integral

$$\pi(\phi | x) = \int \pi(\theta, \phi | x) d\theta \quad (2.10)$$

Para algunos modelos se puede usar:

$$\pi(\phi | x) = \frac{\pi(\theta, \phi | x)}{\pi(\theta | \phi, x)} \quad (2.11)$$

- Simular el vector de hiperparámetros  $\phi$  de la marginal  $\pi(\phi | x)$ .
- Simular  $\theta$  a partir de  $\pi(\theta | \phi, x)$ .
- Dependiendo del problema puede ser necesario simular un nuevo valor  $\tilde{x}$  dado  $\theta$ .

Sin embargo, no siempre se pueden encontrar estas distribuciones condicionales, en forma analítica, por lo que una alternativa es simular de ellas utilizando métodos de Cadenas de Markov Monte Carlo, particularmente, utilizando el método de Gibbs sampling que se explicará en la siguiente sección.

## 2.2. Métodos de Cadenas de Markov Monte Carlo (MCMC)

Los métodos MCMC tuvieron su origen en la estadística física, cuando fueron utilizados para estudiar las propiedades de sistemas de partículas interactuando. Su uso en la inferencia Bayesiana, data de los años 80.

Estos métodos han sido altamente exitosos por su capacidad de encontrar aproximaciones de modelos complicados y de alta dimensionalidad. El objetivo de los métodos de Cadenas de Markov Monte Carlo es encontrar una cadena de Markov

en el espacio de los parámetros, cuya distribución de equilibrio o estacionaria de la cadena coincide con la distribución posterior.

### 2.2.1. Gibbs Sampling

Este algoritmo, también llamado de muestreo condicional alternante, es sumamente útil cuando el espacio de parámetros es alternante multidimensional. Se supone que  $\theta$  tiene  $d$  componentes, es decir  $\theta = (\theta_1, \dots, \theta_d)$ . En cada iteración  $r$ , del algoritmo se escoje un ordenamiento de los  $d$  subvectores, y cada  $\theta_j^r$  es muestreado de la distribución condicional dados todos los demás componentes,

$$\pi(\theta_j | \theta_{-j}^{r-1}, x) \quad (2.12)$$

donde  $\theta_{-j}^{r-1}$  representa todos los componentes de  $\theta$  excepto por  $\theta_j$  en sus valores actuales,

$$\theta_{-j}^{r-1} = (\theta_1^r, \dots, \theta_{j-1}^r, \theta_{j+1}^{r-1}, \dots, \theta_d^{r-1}) \quad (2.13)$$

En la práctica es necesario establecer un valor inicial  $\theta_0 = (\theta_1^0, \dots, \theta_d^0)$ . En cada iteración del algoritmo se muestrea de las siguientes distribuciones condicionales.

Muestrear de:

$$\begin{aligned} \theta_1^r &\sim \pi(\theta_1 | \theta_2^{r-1}, \dots, \theta_d^{r-1}) \\ \theta_2^r &\sim \pi(\theta_2 | \theta_1^r, \theta_3^{r-1}, \dots, \theta_d^{r-1}) \\ \theta_3^r &\sim \pi(\theta_3 | \theta_1^r, \theta_2^r, \theta_4^{r-1}, \dots, \theta_d^{r-1}) \\ &\vdots \\ \theta_d^r &\sim \pi(\theta_d | \theta_1^r, \theta_2^r, \dots, \theta_{d-1}^r) \end{aligned}$$

Si estas distribuciones condicionales, resultan tener una forma cerrada (conocida), entonces fácilmente se puede muestrear de ellas, de lo contrario se requiere de otros métodos. Una alternativa es el método de Metropolis Hastings que será descrito a

continuación.

### 2.2.2. Metropolis-Hastings

Supongamos que queremos simular la distribución posterior de  $\phi(\theta|x)$  o en general de una distribución  $g(\theta)$ . El algoritmo de Metropolis-Hastings comienza con un valor inicial  $\theta^0$ , y especifica una regla de aceptación o rechazo para el  $t$ -ésimo valor de la secuencia  $(\theta^t)$  teniendo en cuenta el  $(t - 1)$ -ésimo valor en la secuencia  $(\theta^{t-1})$ . Para esto es necesario tener una distribución  $p(\theta)$  de la cuál sea fácil simular. Específicamente, este algoritmo puede describirse como sigue.

- Simular un valor candidato  $\theta^*$ , para la densidad  $p(\theta^*|\theta^{t-1})$
- Calcular la razón

$$R = \frac{g(\theta^*) p(\theta^{t-1} | \theta^*)}{g(\theta^{t-1}) p(\theta^* | \theta^{t-1})}$$

- Calcular la probabilidad de aceptación  $p = \min\{R, 1\}$
- Definir  $\theta^t = \theta^*$  con probabilidad  $p$ , de otro modo  $\theta^t = \theta^{t-1}$

Bajo algunas condiciones de regularidad en  $p(\theta^*|\theta^{t-1})$ , la secuencia  $\theta^1, \theta^2, \dots$  convergerá a la distribución  $g(\theta)$ .

### 2.3. Modelos Longitudinales

Los modelos longitudinales combinan observaciones de corte transversal y de series de tiempo. En ese sentido, incorporan más información que un modelo de sólo series de tiempo o sólo de corte transversal (regresión). Por series de tiempo se entiende un conjunto de datos que recoge observaciones de un fenómeno a lo largo del tiempo. Dichos conjuntos de datos están ordenados y la información relevante respecto al fenómeno estudiado es la que proporciona su evolución en el tiempo. Un conjunto transversal de datos contiene observaciones sobre múltiples fenómenos en

un momento determinado, en este caso, el orden de las observaciones es irrelevante (Daniels y Hogan, 2008). El estudio de datos longitudinales tiene por objetivo capturar la heterogeneidad no observable, ya sea entre individuos así como también en el tiempo, dado que esta heterogeneidad no se puede detectar ni con estudios de series temporales ni tampoco con los de corte transversal (Mayorga y Muños, 2000). Por lo tanto un modelo de datos longitudinales puede escribirse como el resultado de la evolución conjunta de la variable respuesta y de sus covariables.

Sea,  $Y_i = (y_{i1}, y_{i2}, \dots, y_{i\tau_0})^T$  el vector de la variable respuesta medidos en el tiempo  $j = 1, 2, \dots, \tau_0$  y  $x_{ij} = (x_{ij1}, \dots, x_{ijp})$  el vector de  $p$  covariables del individuo  $i$  al tiempo  $j$ . La matriz completa de las covariables para el individuo  $i$  es, entonces,

$$x_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ij} \\ \vdots \\ x_{i\tau_0} \end{bmatrix}_{\tau_0 \times p}$$

En regresión estamos interesados en caracterizar los efectos de las covariables  $x_i$  sobre una variable dependiente  $Y_i$ . Es decir se desea realizar inferencia acerca de la distribución conjunta de  $Y_i$  condicionado sobre  $x_i$ . En modelos de regresión con datos longitudinales la función de verosimilitud requiere la especificación de la distribución conjunta  $f(Y_i|x_i, \theta)$ . Donde  $\theta$  es un vector de parámetros, que podría incluir coeficientes de regresión, componentes de varianza y parámetros de correlación serial. La inferencia sobre la distribución conjunta de respuestas, según el objetivo propuesto puede especificarse marginal (directa) o condicionalmente.

Un enfoque marginal de la media caracteriza  $E(Y_i|x_i)$  sin recurrir a estructuras latentes. Estos modelos incorporan la correlación entre individuos modelando las correlaciones o las covarianzas, en vez de permitir efectos aleatorios. Los parámetros de la covarianza, y/o correlación, no son de interés sino que son tratados como un estorbo (nuisance), que se controla pero no se estima. Un modelo marginal, usualmente llamado modelo de efectos fijos, se utiliza cuando se pretenden realizar inferencias poblacionales, es decir, si lo que se quiere es explicar la relación entre la variable dependiente y las variables explicativas con independencia de la variabilidad intra-individual (Sáez, 2011). El principal problema de este modelo, es encontrar una parametrización para la matriz de covarianzas o correlaciones, y, a menudo en el estudio del análisis longitudinal, observaciones del mismo individuo (grupo) tienden a estar correlacionadas, por lo que frecuentemente se requiere el uso de un modelo condicional.

Un modelo de especificación condicional, también conocido como modelo de efectos mixtos, considera que la media está dada condicionalmente sobre respuestas previas o efectos aleatorios, que reflejan aspectos de la distribución conjunta de la variable respuesta (Daniels, 2007). Con este modelo se pretende realizar inferencias individuales y permite capturar la variabilidad dentro del sujeto y entre sujetos (individuos). Para ello, modela simultáneamente la media de la variable dependiente (la variabilidad entre sujetos) y la estructura de covarianzas o correlaciones (la variabilidad dentro de los sujetos) (Sáez, 2011). Si estamos interesados en estimar la relación entre sujetos, un modelo condicional provee información de ello mediante la varianza de los efectos aleatorios, mientras que en un modelo de efectos fijos, se tendría que usar un parámetro por sujeto. de modo que un modelo de especificación condicional reduce la cantidad de parámetros a ser estimados.

### 2.3.1. Modelo Lineal Mixto

En general se considera que los efectos de un modelo son especificados por el experimentador, sin embargo no siempre pueden ser especificados, si no que también pueden ser el resultado de una selección al azar. Con la inclusión de efectos aleatorios, los modelos lineales generalizados se pueden utilizar para modelar datos longitudinales y agrupados.

El modelo más común de efectos mixtos para datos longitudinales especifica la distribución conjunta  $f(y, \mu|x, \theta)$  como

$$f(y|\mu, x, \theta_1)f(\mu|x, \theta_2)$$

Donde  $\theta_1$  captura los efectos condicionales de  $X$  sobre  $Y$  y  $\theta_2$  captura las características de la distribución de efectos aleatorios  $\mu$ . La distribución no condicional sobre los efectos aleatorios se obtiene integrando sobre  $\mu$ , es decir,

$$f(y|x, \theta_1, \theta_2) = \int f(y|x, \mu, \theta_1)f(\mu|x, \theta_2)d\mu$$

El modelo lineal mixto es comúnmente usado para modelar la variable respuesta de datos longitudinales, además que permite tener gran flexibilidad en el modelamiento e interpretabilidad de los parámetros.

Este modelo se puede describir en dos niveles. En el primer nivel, sea  $Y_i$  el vector de respuestas medido en los tiempos  $\{1, 2, \dots, \tau_0\}$ , condicionado normalmente sobre los efectos aleatorios  $\mu_i$ .

$$Y_i|x_i, \mu_i \sim N(E(Y_i|x_i, \mu_i), \Sigma_i(\phi))$$

$\Sigma_i(\phi)$  es parametrizada por el vector  $\phi$  de parámetros no redundantes, esta condicionado sobre los efectos aleatorios y captura la variabilidad dentro de los sujetos.

Incorporando los efectos de las covariables se obtiene que

$$E(Y_i|x_i, \mu_i) = x_i\varphi + w_i\mu_i$$

donde  $w_i$  es la matriz de diseño de los efectos aleatorios. Aquí  $\theta_1 = (\varphi, \phi)$

Esta representación de la media condicional motiva el término de modelo de efectos mixtos, porque considera la cuantificación de los efectos de los coeficientes a nivel poblacional ( $\varphi$ ) e individual ( $\mu_i$ ).

En el segundo nivel al vector de efectos aleatorios se le asigna una distribución que puede depender de las covariables, la distribución normal es una elección común.

$$\mu_i | x_i \sim N(0, \Omega(\eta))$$

donde  $\Omega(\eta)$  es la matriz de covarianzas de los efectos aleatorios y es parametrizada por  $\eta$ . Aquí  $\theta_2 = \eta$

Los estimados de los parámetros se pueden encontrar mediante estimadores de Máxima Verosimilitud o utilizan el método Bayesiano.

### Estimadores de Máxima Verosimilitud (MLE)

El método de Máxima Verosimilitud es una técnica popular para estimar parámetros (Casella, 2001). Para una muestra independiente  $Y_1, \dots, Y_n$  con función de densidad  $f(y|\theta)$ , donde  $\theta$  es un vector de parámetros. La función de verosimilitud está definida como:

$$\begin{aligned} L(Y | \theta) &= f(y | \theta) \propto f(y_1, \dots, y_n | \theta) \\ &= \prod_{i=1}^n f(y_i | \theta) \end{aligned} \tag{2.14}$$

De modo que los estimadores de máxima verosimilitud se encuentran maximizando la ecuación 2.14.

En el modelo de efectos aleatorios.

Sea  $Y_i = (y_{i1}, y_{i2}, \dots, y_{i\tau_0})^T$  la respuesta observada para el individuo  $i$ , donde  $i =$

$1, \dots, n$ . Asumiendo normalidad e independencia entre sujetos, la función de verosimilitud es,

$$\begin{aligned} L(\theta | Y) &\propto \prod_{i=1}^n f(y | x, \theta_1) f(\mu | x, \theta_2) \\ &= \prod_{i=1}^n |\Sigma_i|^{-1/2} e^{\left\{-\frac{1}{2}(y_i - x_i\beta - w_i\mu_i)^T (\Sigma_i^b)^{-1} (y_i - x_i\beta - w_i\mu_i)\right\}} f(u_i | \Omega) \end{aligned} \quad (2.15)$$

Los estimadores para los parámetros, se obtienen maximizando la ecuación (2.15), esto significa encontrar las derivadas parciales para cada parámetro e igualar a cero, cuyas soluciones podrían ser complicadas, por lo que se usan métodos numéricos para encontrar el MLE.

### Enfoque Bayesiano

En este enfoque además de los efectos aleatorios, los parámetros  $\varphi$ ,  $\Omega$  y  $\Sigma_i$  son también aleatorios. La inferencia Bayesiana mediante la distribución a priori, les asigna una distribución a los parámetros mencionados para luego ser actualizada en la distribución posterior, usando la información de los datos que está cuantificada en  $f(y|\mu_i, \varphi, \Omega, \Sigma_i)$  (ver ecuación 2.2).

En algunos casos la distribución posterior tiene una forma conocida y en otros no, por lo que para muestrear de la distribución posterior se requiere el uso de algoritmos como Gibbs Sampling y Metropolis Hastings discutidos en las secciones 2.2.1 y 2.2.2 respectivamente.

Para un modelo lineal mixto  $Y_i \sim N(x_i\varphi + w_i\mu_i, \Sigma_i)$ , con  $\mu_i \sim N(0, \Omega)$  y  $\Sigma_i = \sigma^2 I$ , donde  $I$  es la matriz identidad, usualmente se asume distribuciones a priori condicionalmente conjugadas (Daniels y Hogan, 2008), como:

$$\varphi \sim N(\varphi_0, \tau I)$$

$$\frac{1}{\sigma^2} \sim G(h, g)$$

$$\Omega^{-1} \sim \text{Wishart}(v, v/A)$$

donde,  $v$  es el número de efectos aleatorios y  $A$  es una matriz de dimensión  $v \times v$  cuyos elementos de la diagonal representan la variabilidad de los estimadores de los parámetros obtenidos mediante un modelo de regresión. Bajo estas condiciones, para muestrear de la distribución posterior, sólo es necesario el uso del algoritmo Gibbs Sampling, ya que la distribución posterior condicional para cada uno de los parámetros tiene una forma conocida.

## Capítulo 3

# MODELO DEL CAMBIO DE TEMPERATURA Y PRECIPITACIÓN

### 3.1. Descripción de los Datos Climáticos

Los datos consisten de temperatura y precipitación observados y proyectados por los modelos GCMs, promediados sobre la temporada de crecimiento de frijol por país y por años, desde 1961 hasta el 2007, para los datos observados y desde 1961 al 2030 para los modelos proyectados. Los datos observados se obtuvieron del conjunto de datos de *University of Delaware Terrestrial Air Temperature and Precipitation* (<http://climate.geog.udel.edu/climate>). Estos están disponibles para cuadrículas de cada 0.5 grados mensuales y por año. Como nuestro estudio se enfoca en el clima durante la temporada de crecimiento del frijol, y cada país realiza el cultivo de esta legumbre en diferentes temporadas, fue necesario realizar una correspondencia entre el clima y la temporada de crecimiento (Lobell et al., 2011). Teniendo en cuenta que en una región el clima puede ser muy variable, se consideró la cantidad de área en la que se cultiva, para describir mejor la temperatura y precipitación promedio durante la temporada de crecimiento para cada país.

Los datos de la temporada de crecimiento y el área dedicada al cultivo se obtuvieron de los calendarios de Sacks (2010). Estos calendarios basados en los datos del año 2010, proveen información del inicio y final de período y el área de cultivo por cada cuadrícula de 5 min para 19 cultivos, entre ellos pulses, categoría dentro del cual se encuentra el frijol. Para hacer corresponder la “temporada de crecimiento”

a un punto (coordinada según longitud y latitud) promediamos las fechas de inicio y final de todas las cuadrículas de 5 min que caen dentro de una distancia de .25 grados de esta coordenada (definiendo una nueva cuadrícula de 0.5 grados, ver Figura 3-1). De manera análoga, se encontró el área total dedicada al cultivo para la misma coordenada, estableciéndose así una “temporada de crecimiento” y un “área dedicada al cultivo” para la “nueva cuadrícula” (cuadrícula en negro en la Figura 3-1). La temperatura promedio para esta cuadrícula se calculó dentro de la “temporada de crecimiento”. Este análisis se hizo para todas las cuadrículas en cada país, por lo tanto cada cuadrícula de .5 grados en un país tiene una temperatura y un área de cultivo. Para determinar la temperatura promedio para el cultivo del frijol en cada país para un año en particular, se calculó un promedio ponderado de las temperaturas de todas las cuadrículas que están dentro del país, siendo los pesos de cada temperatura proporcional al área de cultivo (Lobell et al., 2011). Para obtener la cantidad de precipitación promedio en el país, se hizo el mismo análisis.

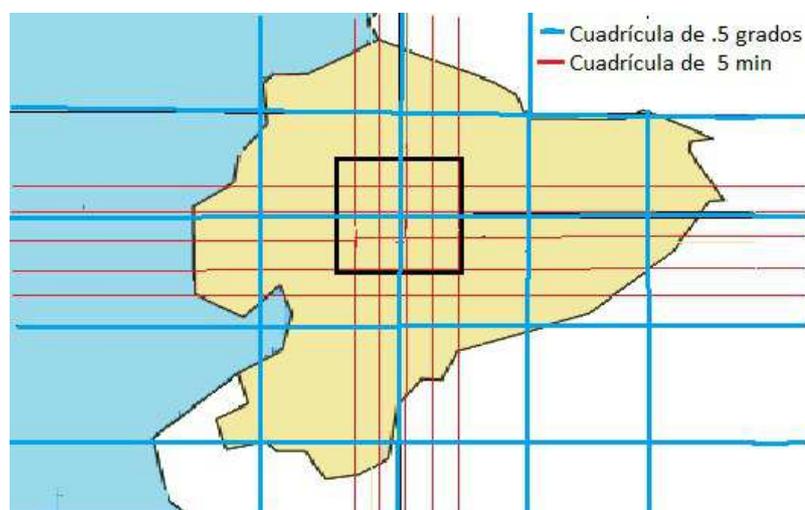


Figura 3-1: Correspondencia entre datos de temperatura y la temporada de crecimiento para Ecuador

Las temperaturas y precipitaciones son diferentes para cada país, porque promediamos sobre diferentes áreas de cultivo y diferentes temporadas de crecimiento. Como nuestro interés es el efecto del clima en el rendimiento, se seleccionó la data

del clima para sólo 44 países Tabla B-1 que cumplieron con ciertos requisitos de rendimiento (Sección 4.1).

Los datos anuales de temperatura y precipitación actual y futura proyectados bajo distintos escenarios climáticos se encuentran disponibles en <http://www.pcmdi.llnl.gov>. El número de modelos GCM para cada escenario varía entre 15 a 18 modelos. En nuestro estudio utilizamos sólo un escenario climático: SRESA1B (Special Report on Emission Scenarios), este escenario es considerado el más completo, porque considera muchos factores para estimar los modelos. De este escenario climático tomamos 15 GCMs. Los datos están dados por cuadrículas de 5 grados por lo que para nuestro análisis promediamos el clima sobre las regiones de cultivo de frijol y la temporada de crecimiento para cada país, siguiendo el procedimiento que acabamos de describir para los datos observados.

### **3.2. Tendencias en el Clima Observado**

Las Figuras 3-2 y 3-3 muestran las series anuales para precipitación y las Figuras 3-4 y 3-5 muestran las series anuales de temperatura para los 44 países. En el caso de precipitación se encuentran tendencias positivas y negativas, resaltando más las tendencias de países cuya precipitación ha disminuido en el tiempo, como es el caso de Myanmar, Romania, Nepal y algunos países centroamericanos como: Guatemala, Nicaragua y El Salvador

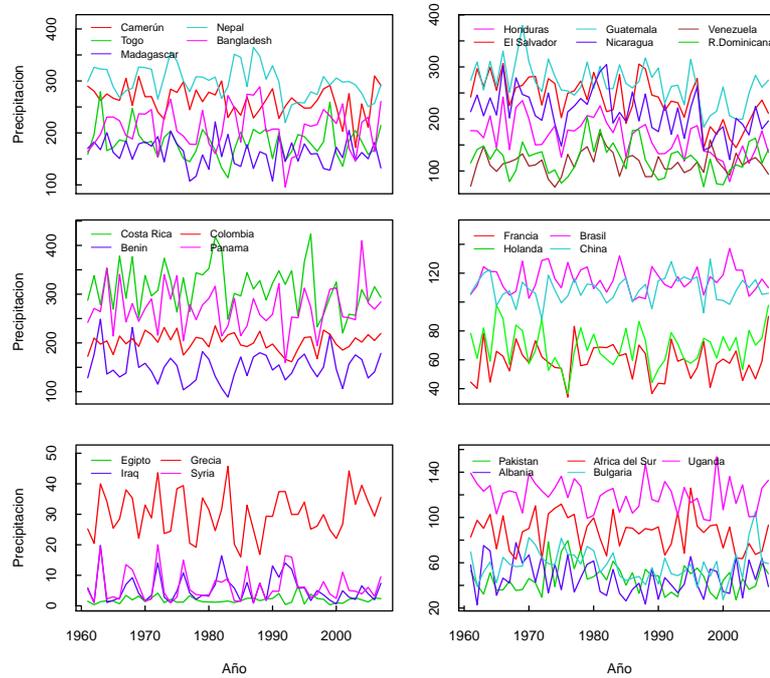


Figura 3-2: Series del promedio anual de precipitación ( $mm/mes$ ) en la temporada de crecimiento del frijol de 1961-2007 para 27 de los 44 países bajo estudio.

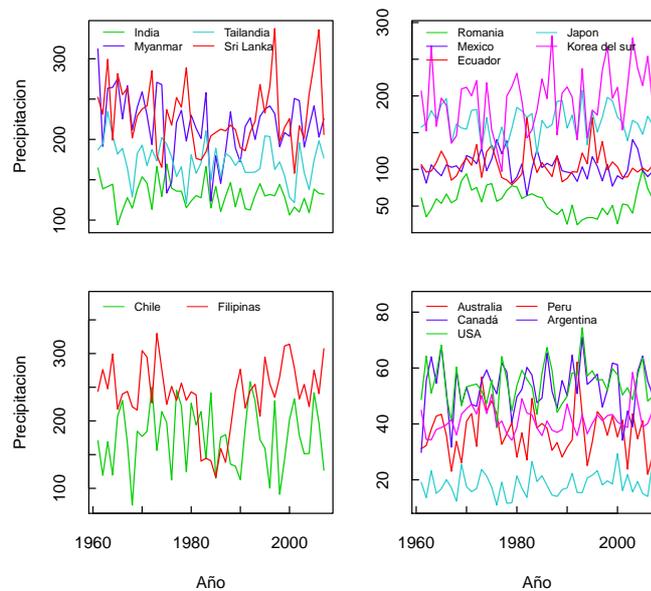


Figura 3-3: (Continuación de la Figura 3-2) series del promedio anual de precipitación ( $mm/mes$ ) en la temporada de crecimiento del frijol de 1961-2007 para 17 países restantes.

Para temperatura, algunos países revelaron tendencias positivas, negativas y otros tendencias no lineales. En la mayoría de los países la temperatura fue aumentando cada año, de manera particular se puede ver esta característica para los países centroamericanos (primera fila y segunda columna de la Figura 3-4). Particularmente, las series de temperatura revelan tendencias positivas desde 1980 para casi todos los países. (Lobell et al., 2011) muestran que a partir de los años 80 el aumento de temperatura fue mayor con respecto a los 20 años anteriores para la temporada de crecimiento de algunos cultivos como soya, trigo y maíz. Tomando esta referencia realizamos una comparación similar para la temporada del frijol entre los períodos 1961-1984 y 1984-2007, consideramos 1984 por ser la mitad del período. Para comparar la magnitud de las tendencias en estos 2 períodos, se estimó la tendencia usando los datos para cada período y para cada país usando un modelo de regresión para temperatura (precipitación) en función del tiempo, y para cuantificar esta tendencia, se hizo la diferencia entre los valores de temperatura (precipitación) que predijo el modelo para los años 1984 y 1961. De manera similar se encontró las tendencias para el período 1984-2007. Finalmente estas tendencias fueron normalizadas por la desviación estándar histórica (desviación estándar de temperatura y precipitación en el período de 1961 a 2000 (Lobell et al., 2011). Interpretamos que si la tendencia normalizada es 1, significa que la temperatura (precipitación) al final del período fue  $1\sigma$  más grande que al inicio del período.

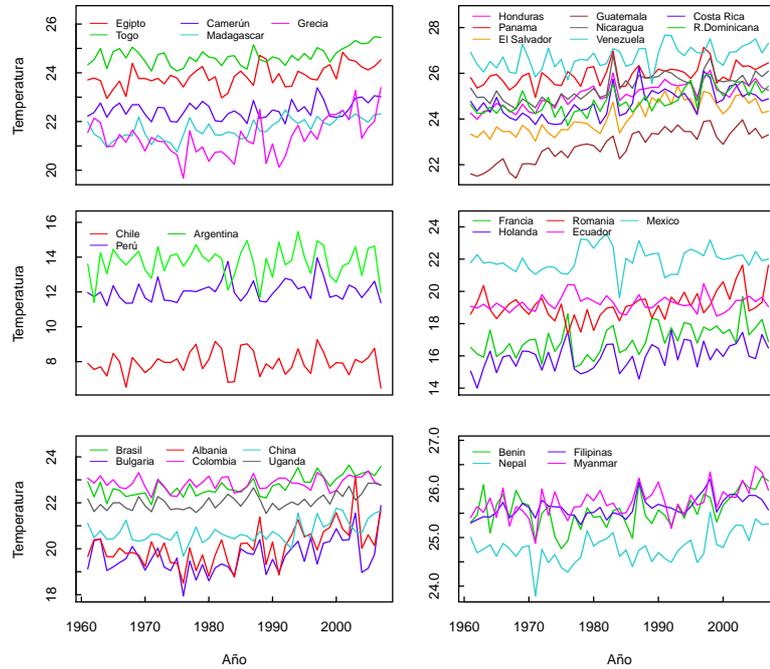


Figura 3-4: Series del promedio anual de temperatura ( $^{\circ}C$ ) en la temporada de crecimiento del frijol de 1961-2007 para 31 de los 44 países bajo estudio.

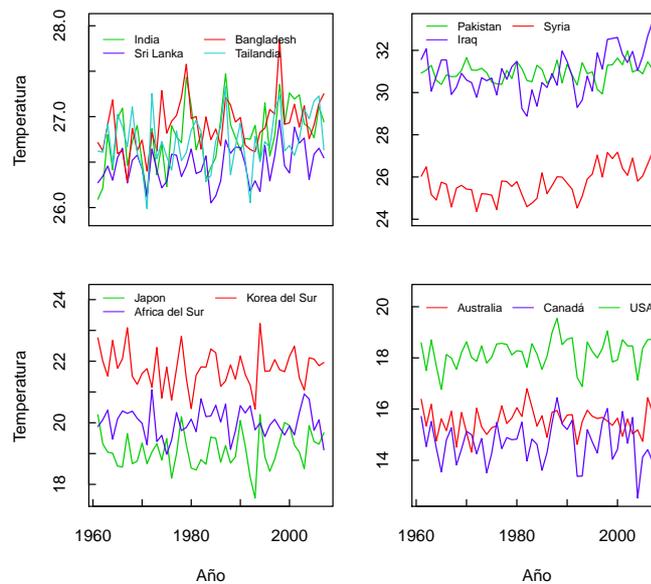


Figura 3-5: (Continuación Figura 3-4) series del promedio anual de temperatura ( $^{\circ}C$ ) en la temporada de crecimiento del frijol de 1961-2007 para 13 países restantes.

De la Figura 3-6 (izquierda) se puede concluir que en general hubo mayor aumento de temperatura en el período de 1984-2007 respecto a las 2 décadas anteriores, sin embargo, en precipitación (figura de la derecha), el comportamiento general en las tendencias de estos dos periodos fue muy similar. El hecho de decir que hubo un aumento de temperatura en el período 1984-2007 respecto de 1961-1984, no significa que todos los países se calentaron, similarmente si hablamos de precipitación en forma global decir que las tendencias en los dos períodos fueron similares no quiere decir que no haya países donde la cantidad de precipitación para la temporada de cultivo de frijol haya aumentado o disminuido.

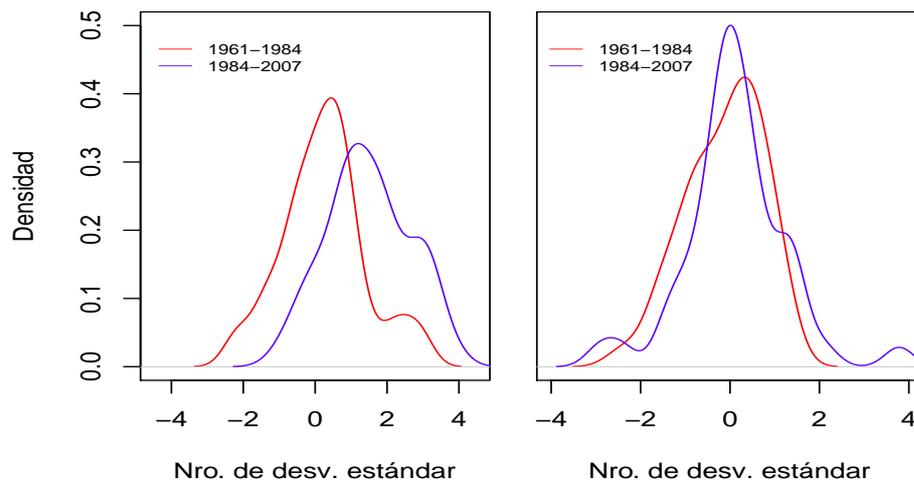


Figura 3-6: Tendencia de temperatura y precipitación para la temporada de crecimiento de 1961-1984 vs 1984-2007 (número de desviaciones estándar respecto a la data histórica). Gráfica de la izquierda para temperatura y derecha para precipitación.

La Figura 3-7 muestra una comparación de las tendencias en el clima entre los dos periodos (1961-1984 vs 1984-2007) para cada país (a la izquierda temperatura y a la derecha precipitación). Puntos por encima de la recta  $y = x$  representan países donde el aumento o disminución de temperatura (precipitación) en 1984-2007 fue mayor con respecto de los 24 años anteriores, puntos bajo la recta representa

países donde el aumento o disminución de temperatura (precipitación) fue menor en el período 1984-2007 respecto al período anterior. Note que 30 de los 44 países aumentan su temperatura con respecto a los 24 años anteriores.

En cuanto a la temperatura, se observa que tanto en el período de 1961-1984 y 1984-2007 hubo un aumento de temperatura en casi todos los países (puntos a la derecha del eje  $y$  y puntos sobre el eje  $x$  respectivamente). Además, 18 de 44 países que se enfriaron en el período 1961-1984 pasaron a calentarse, en el período 1984-2007 (puntos del cuadrante II). Por otro lado, 3 países que mostraron un calentamiento en el periodo 1961-1984, pasaron a enfriarse en el período 1984-2007 (puntos en el cuadrante IV). En adición hay 2 países que muestran disminución de temperatura en ambos períodos (puntos bajo la línea roja en el cuadrante III). Notamos que en particular, de los 5 primeros productores de frijol en el mundo, 4 de ellos muestran mayor aumento de temperatura en los últimos 24 años (Bazil, China, Myanmar e India). Esto quiere decir que se calientan más que en el período 1961-1984 y sólo Estados Unidos muestra un aumento en temperatura en el período 1984-2007 que fue menor que en el periodo anterior.

En cuanto a precipitación, la cantidad de países con aumentos o disminución en precipitación en el segundo período en comparación al primero son iguales, lo que explica porque no hubo un desplazamiento de las curvas en la Figura 3-6. Sin embargo, algo que se puede resaltar en las tendencias de precipitación es que 12 países que mostraban aumento en la cantidad de precipitación en el período 1961-1984 pasaron a disminuir en el segundo período (puntos del cuadrante IV). De los primeros productores resalta Myanmar que muestra aumento de precipitación en el período 1984-2007, mientras que Estados Unidos, China e India muestran disminución en precipitación. Brazil, por su parte, muestra aumento en precipitación

en ambos períodos, pero el aumento es menor en el período 1984-2007 respecto del primero.

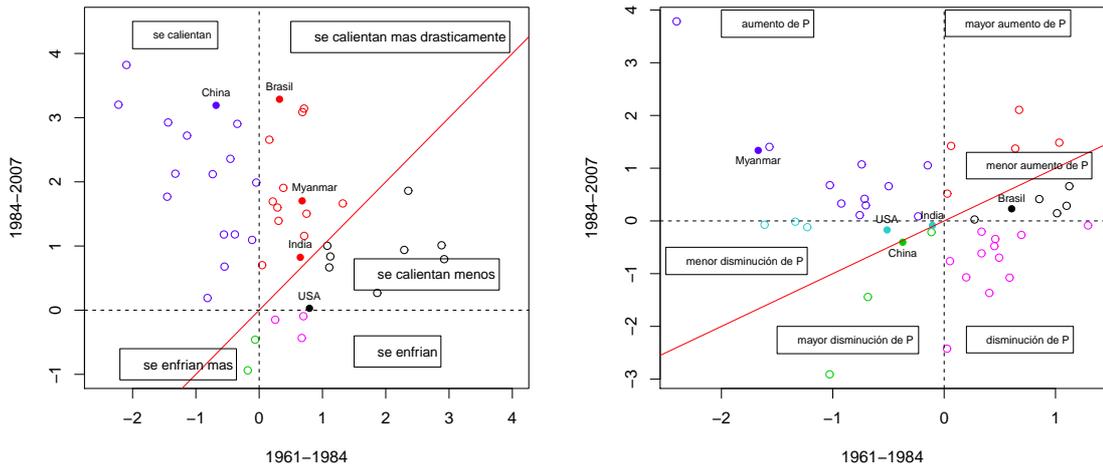


Figura 3-7: Gráfico del cambio de temperatura (izquierda) y precipitación  $P$  (derecha) de 1961-1984 vs 1984-2007 identificando los 5 primeros productores.

### 3.3. Modelando la Distribución Conjunta de Temperatura y Precipitación Incorporando Modelos de Proyección (GCM)

En esta sección nos enfocamos en el modelo desarrollado por Tebaldi y Sansó, (2008), quienes estiman la distribución conjunta del cambio de temperatura y precipitación. Utilizando un modelo Bayesiano jerárquico estimaron promedios por década de temperatura y precipitación actual y futura en una región determinada, basados en datos observados y datos simulados usando GCMs. A diferencia de ellos en nuestro trabajo consideramos promedios anuales de temperatura y precipitación para cada uno de los 44 países durante la temporada de cultivo del frijol.

Mencionamos en la Sección 3.1 que el conjunto de datos observados comprende el periodo de 1961 al 2007. Denotamos a  $\tau_0$  como el número de años en este período así,  $\tau_0 = 47$ . Los datos proyectados para los 15 GCMs bajo el escenario SRESA1B son desde 1961 al 2030 y denotamos con  $\tau^*$  como el número de años durante este

período por lo que  $\tau^* = 70$ . Utilizaremos el sub índice  $j$  para referirnos al  $j$ -ésimo modelo GCM, de modo que  $j = 1, 2, \dots, M$ , dónde  $M = 15$  ya que estamos usando 15 GCM. El tiempo en el período lo denotaremos por  $t$  por lo que para los datos observados  $t = 1, 2, 3, \dots, \tau_0 = 47$  y para los modelos GCM  $t = 1, 2, \dots, \tau^* = 70$ .

Sea  $O_t = (O_t^T, O_t^P)$  el vector bidimensional de temperatura y logaritmo de precipitación observados en el tiempo  $t = 1, \dots, 47$  y  $X_{jt} = (X_{jt}^T, X_{jt}^P)$  el vector de temperatura y logaritmo de precipitación proyectadas por el  $j$ -ésimo GCM para  $t = 1, \dots, \tau_0$ . Note que los subíndices T y P se utilizan para denotar temperatura y precipitación, respectivamente. El modelo de Tebaldi y Sansó, (2008) considera los siguientes supuestos:

1. El vector de valores observados tiene distribución normal bivalente con correlación entre temperatura y precipitación.
2. El verdadero proceso es una función lineal por partes para temperatura y precipitación. Se supone un “elbow” en el año 2007, considerando que las tendencias futuras pueden cambiar con respecto a las tendencias observadas.
3. Las proyecciones  $X_{jt}$  son sesgadas con respecto al verdadero valor observado de temperatura y precipitación, se asume un sesgo aditivo y ruido normal bivariado.

Entonces definen el modelo estadístico de la siguiente manera:

Para la temperatura observada

$$O_t^T \sim N[\mu_t, \eta^T]', \text{ para } t = 1, \dots, \tau_0$$

y la precipitación

$$O_t^P \sim N[\mu_t^P + \beta_{xo}(O_t^T - \mu_t^T), \eta^P], \text{ para } t = 1, 2, \dots, \tau_0,$$

donde

$$X_{jt}^T \sim N[\mu_t^T + d_j^T, \varepsilon_j^T], \text{ para } j = 1, \dots, M \text{ y } t = 1, \dots, \tau^*$$

y

$$X_{jt}^P \sim N[\mu_t^P + \beta_{xj}(X_{jt}^T - \mu_t^T - d_j^T) + d_j^P, \varepsilon_j^P], \text{ para } j = 1, \dots, M \text{ y } t = 1, \dots, \tau^*$$

donde

$$\beta_{xo} \sim N[\beta_0, \lambda_o]$$

note que  $N[v, R]$  denota la distribución normal con media  $v$  y varianza  $R^{-1}$ .

Estos modelos tienen una distribución normal bivariada para  $O_t$  y  $X_{jt}$ , y las correlaciones  $\beta_{x0}, \beta_{x1}, \dots, \beta_{xM}$  se utilizan para modelar la correlación entre temperatura y precipitación. En los modelos de precipitación y temperatura asumen que todos los parámetros son aleatorios a excepción de  $\beta_{x0}, \eta^T$  y  $\eta^P$  que son valores fijos calculados a partir de los datos observados.

Además asumen que la evolución verdadera del clima se modela por  $\mu'_t = (\mu_t^T, \mu_t^P)$ , donde cada componente del vector  $\mu'_t$  consta de una tendencia lineal por partes definida por:

$$\begin{pmatrix} \mu_t^T \\ \mu_t^P \end{pmatrix} \equiv \begin{pmatrix} \alpha^T + \beta^T t + \gamma^T (t - \tau_0) I_{(t > \tau_0)} \\ \alpha^P + \beta^P t + \gamma^P (t - \tau_0) I_{(t > \tau_0)} \end{pmatrix} \quad (3.1)$$

En la ecuación (3.1) el hecho de considerar  $\gamma^T$  significa que la tendencia durante el período 2007-2030 puede cambiar de pendiente, debido a que las tendencias del clima no son constantes. En general se espera una tendencia mayor durante este período que las tendencias observadas.

### 3.3.1. Determinación de las Distribuciones a Priori y a Posteriori

La distribución a priori para los parámetros del modelo formulado por Tebaldi y Sansó, (2008) se especifican mediante un modelo Bayesiano jerárquico:  $\beta_{xj} \sim N[\beta_0, \lambda_B]$ ,  $d_j^T \sim N[a^T, \lambda_D^T]$ ,  $d_j^P \sim N[a^P, \lambda_D^P]$  para  $j = 1, \dots, M$ ,  $\xi_j^T \sim$

$G[a_{\xi^T}, b_{\xi^T}]$  y  $\xi_j^P \sim G[a_{\xi^P}, b_{\xi^P}]$ .  $\lambda_0$  es un valor fijo que se calcula de los datos observados. Para el resto de los parámetros se asignan distribuciones a priori no informativas:  $\beta_0, a^T, a^P \sim U[-\infty, \infty]$  y  $\lambda_B, \lambda_D^T, \lambda_D^P, a_{\xi^T}, b_{\xi^T}, a_{\xi^P}, b_{\xi^P} \sim G[g, h]$ , donde  $g = h = 0.01$ . De forma similar para los parámetros de la ecuación (3.1) asumimos que  $\alpha^T, \beta^T, \gamma^T, \alpha^P, \beta^P, \gamma^P \sim U[-\infty, \infty]$ , donde  $G[a, b]$  denota la distribución Gamma con función de densidad proporcional a  $x^{a-1}e^{-bx}$ .

Los coeficientes de correlación  $\beta_{x_0}$  y  $\beta_{x_j}$  tienen una media común  $\beta_0$ , y los sesgos  $d_j^T$  y  $d_j^P$  tiene medias comunes  $a^T$  y  $a^P$  respectivamente. Estas modelan errores sistemáticos de las variable simuladas por cada GCM. Todos los sesgos de los GCM tanto para temperatura como para precipitación son realizaciones de una distribución normal con medias posiblemente diferentes de 0, cuando el conjunto de las trayectorias del modelo no están distribuidas simétricamente alrededor del verdadero valor de temperatura y/o precipitación. También consideran que cada modelo tiene su propia precisión (inversa de la varianza) en la simulación de las verdaderas series de tiempo de temperatura y precipitación, asumen distribuciones a priori comunes para  $\xi_j^T$  y  $\xi_j^P, \forall j$ , cuyos parámetros son estimados a partir de los datos. Bajo estas consideraciones. Sea,

$$\theta = (\beta_{x_j}, \beta_0, \lambda_B, d_j^T, a^T, d_j^P, a^P, \lambda_D^T, \lambda_D^P, \varepsilon_j^T, a_{\varepsilon^T}, b_{\varepsilon^T}, \varepsilon_j^P, a_{\varepsilon^P}, b_{\varepsilon^P}, \alpha^T, \beta^T, \gamma^T, \alpha^P, \beta^P, \gamma^P, \lambda_0) \quad (3.2)$$

La función de densidad conjunta de las distribuciones a priori es:

$$f(\theta) = \left[ \prod_{j=1}^M f(\beta_{x_j} | \beta_0, \lambda_B) f(d_j^T | a^T, \lambda_D^T) f(d_j^P | a^P, \lambda_D^P) f(\varepsilon_j^T | a_{\varepsilon^T}, b_{\varepsilon^T}) f(\varepsilon_j^P | a_{\varepsilon^P}, b_{\varepsilon^P}) \right] \\ \times f(\beta_{x_0} | \beta_0, \lambda_0) f(\beta_0) f(\lambda_B) f(a^T) f(\lambda_D^T) f(a^P) f(\lambda_D^P) f(a_{\varepsilon^T}) f(b_{\varepsilon^T}) f(a_{\varepsilon^P}) \\ f(b_{\varepsilon^P}) f(\alpha^T) f(\beta^T) f(\gamma^T) f(\alpha^P) f(\beta^P) f(\gamma^P)$$

y la función de densidad de lo observado

$$f(O_t^T, O_t^P, X_{jt}^T, X_{jt}^P | \theta) = \left[ \prod_{t=1}^{\tau_0} f(O_t^T | \mu_t^T, \eta^T) f(O_t^P | \mu_t^T, \eta^P, \eta^T, \beta_{xo}, \mu_t^P) \right] \prod_{j=1}^M \left[ \prod_{t=1}^{\tau^*} f(X_{jt}^T | \mu_t^T, d_j^T, \varepsilon_j^T) f(X_{jt}^P | \mu_t^T, \mu_t^P, d_j^T, d_j^P, \varepsilon_j^P, \beta_{xj}) \right] \quad (3.3)$$

Como se mencionó en la Sección 2.2 el objetivo del método Bayesiano es actualizar la distribución de los parámetros, encontrando la distribución posterior de los mismos dado lo observado. En este caso, por la ecuación 2.2 la función de distribución posterior está dada por:

$$\begin{aligned} \pi(\theta | O_t^T, O_t^P, X_{jt}^T, X_{jt}^P) &\propto f(O_t^T, O_t^P, X_{jt}^T, X_{jt}^P | \theta) f(\theta) \\ &= \left[ \prod_{t=1}^{\tau_0} f(O_t^T | \mu_t^T, \eta^T) f(O_t^P | \mu_t^T, \eta^P, \eta^T, \beta_{xo}, \mu_t^P) \right] \\ &\quad \prod_{j=1}^M \left[ \prod_{t=1}^{\tau^*} f(X_{jt}^T | \mu_t^T, d_j^T, \varepsilon_j^T) f(X_{jt}^P | \mu_t^T, \mu_t^P, d_j^T, d_j^P, \varepsilon_j^P, \beta_{xj}) \right] \\ &\quad \times \left[ \prod_{j=1}^M f(\beta_{xj} | \beta_0, \lambda_B) f(d_j^T | a^T, \lambda_D^T) f(d_j^P | a^P, \lambda_D^P) f(\varepsilon_j^T | a_{\varepsilon^T}, b_{\varepsilon^T}) f(\varepsilon_j^P | a_{\varepsilon^P}, b_{\varepsilon^P}) \right] \\ &\quad \times f(\beta_{xo} | \beta_0, \lambda_o) f(\beta_0) f(\lambda_B) f(a^T) f(\lambda_D^T) f(a^P) f(\lambda_D^P) f(a_{\varepsilon^T}) f(b_{\varepsilon^T}) f(a_{\varepsilon^P}) f(b_{\varepsilon^P}) \\ &\quad f(\alpha^T) f(\beta^T) f(\gamma^P) f(\alpha^P) f(\beta^P) f(\gamma^P) \end{aligned} \quad (3.4)$$

### 3.3.2. Gibbs Sampler Para la Distribución Posterior

Para hacer inferencia debemos muestrear de la distribución posterior. Para esto usamos el método de Gibbs Sampler descrito en la Sección 2.2.1. Para usar este método necesitamos las distribuciones condicionales. Usaremos la notación  $\lambda_B | \theta_{\lambda_B}$ , por ejemplo, para denotar la distribución condicional de  $\lambda_B$  dado todos los parámetros de la ecuación 3.2 menos el parámetro  $\lambda_B$ . A continuación las distribuciones condicionales para el modelo Tebaldi y Sansó, (2008). Para detalles de su derivación ver el Apéndice C.1.

$$\lambda_B | \theta_{\lambda_B} \sim G \left( \frac{M}{2} + g - 1, \frac{\sum_j (\beta_{x_j} - \beta_0)^2}{2} + h \right)$$

$$\lambda_D^T | \theta_{\lambda_D^T} \sim G \left( \frac{M}{2} + g - 1, \frac{\sum_j (d_j^T - a^T)^2}{2} + h \right)$$

$$\lambda_D^P | \theta_{\lambda_D^P} \sim G \left( \frac{M}{2} + g - 1, \frac{\sum_j (d_j^P - a^P)^2}{2} + h \right)$$

$$\varepsilon_j^T | \theta_{\varepsilon_j^T} \sim G \left( \frac{\tau^*}{2} + a_{\varepsilon^T} - 1, \sum_{\tau^*} \frac{(X_{jt}^{T'} - d_j)^2}{2} + b_{\varepsilon^T} \right) \text{ donde } X_{jt}^{T'} = X_{jt}^T - \mu_t^T$$

$$\varepsilon_j^P | \theta_{\varepsilon_j^P} \sim G \left( \frac{\tau^*}{2} + a_{\varepsilon^P} - 1, \sum_{\tau^*} \frac{(X_{jt}^{P'} - d_j^P - \beta_{x_j} (X_{jt}^{T'} - d_j^T))^2}{2} + b_{\varepsilon^P} \right)$$

$$\beta_0 | \theta_{\beta_0} \sim N \left( \frac{\lambda_B \sum \beta_{x_j} + \lambda_0 \beta_{x_0}}{\lambda_B M + \lambda_0}, \frac{1}{\lambda_B M + \lambda_0} \right)$$

$$\beta_{x_j} | \theta_{\beta_{x_j}} \sim N \left( \frac{\varepsilon_j^P \sum_{t^*} (X_{jt}^{P'} - d_j^P) (X_{jt}^{T'} - d_j^T) + \lambda_B \beta_0}{\varepsilon_j^P \sum_{t^*} (X_{jt}^{T'} - d_j^T)^2 + \lambda_B}, \sum_{t^*} \varepsilon_j^P (X_{jt}^{T'} - d_j^T)^2 + \lambda_B \right)$$

$$a^P | \theta_{a^P} \sim N \left( \frac{B}{A}, A \right) \text{ donde } A = M \lambda_D^P, B = \lambda_D^P \frac{\sum d_j^P}{M}$$

$$a^T | \theta_{a^T} \sim N \left( \frac{\lambda_D^T \sum d_j^P}{\lambda_D^T M}, M \lambda_D^P \right)$$

La distribuciones condicionales para

$$d_j^P | \theta_{d_j^P}, d_j^T | \theta_{d_j^T}, \gamma^P | \theta_{\gamma^P}, \gamma^T | \theta_{\gamma^T}, \beta^P | \theta_{\beta^P}, \beta^T | \theta_{\beta^T}, \alpha^P | \theta_{\alpha^P}, \alpha^T | \theta_{\alpha^T}$$

todas son normales y están definidas en el Apéndice C.

Muestrear de estas distribuciones es fácil dado que son distribuciones conocidas. Sin embargo, para generar valores de los parámetros  $a_{\varepsilon^T}$ ,  $b_{\varepsilon^T}$ ,  $a_{\varepsilon^P}$  y  $b_{\varepsilon^P}$  se debe utilizar el algoritmo Metropolis Hastings descrito en la Sección 2.1. Usando el método

propuesto por Smith et al., (2007) la generación de estos parámetros se hace en forma conjunta para  $(a_{\varepsilon^T}, b_{\varepsilon^T})$  y  $(a_{\varepsilon^P}, b_{\varepsilon^P})$  (para detalles de la derivación de las condicionales ver el Apéndice C.1) usando el siguiente algoritmo:

1. Genera 2 uniformes independientes  $U_1, U_2$  sobre el intervalo (0,1) y calcular:

$$a'_{\varepsilon^T} = a_{\varepsilon^T} e^{\delta(U_1-1/2)} \text{ y } b'_{\varepsilon^T} = b_{\varepsilon^T} e^{\delta(U_2-1/2)}.$$

Los valores de  $\delta$  son arbitrario en este caso  $\delta = 1$ .

2. Calcular:

$$\begin{aligned} l_1 &= M a_{\varepsilon^T} \log(b_{\varepsilon^T}) - M \log\{\Gamma(\varepsilon^T)\} + (a_{\varepsilon^T} - 1) \sum \log(\varepsilon_j^T) - b_{\varepsilon^T} \sum \varepsilon_j^T + \\ &\quad g * \log(a_{\varepsilon^T} b_{\varepsilon^T}) - h(a_{\varepsilon^T} + b_{\varepsilon^T}) \\ l_2 &= M a'_{\varepsilon^T} \log(b'_{\varepsilon^T}) - M \log\{\Gamma(a'_{\varepsilon^T})\} + (a'_{\varepsilon^T} - 1) \sum \log(\varepsilon^T) - b'_{\varepsilon^T} \sum \varepsilon_j^T + \\ &\quad g * \log(a'_{\varepsilon^T} b'_{\varepsilon^T}) - h(a'_{\varepsilon^T} + b'_{\varepsilon^T}) \end{aligned}$$

Este último cálculo se realiza sobre el logaritmo de la verosimilitud para ambos  $(a_\lambda, b_\lambda)$  y  $(a'_\lambda, b'_\lambda)$

3. Si  $l_2 < l_1$  entonces aceptamos el nuevo  $(a'_\lambda, b'_\lambda)$ , de lo contrario nos quedamos con los valores actuales.

De manera similar se genera para  $(a_{\varepsilon^P}, b_{\varepsilon^P})$ . Se corrieron 150,000 iteraciones del Gibbs Sampler (con el paso de Metropolis-Hastings). Desechamos las primeras 130,000 iteraciones.

### 3.3.3. Distribución Predictiva

Además de estimar la distribución posterior del parámetro  $\theta$  usando el modelo de Tebaldi y Sansó, (2008) nuestro interés es hacer proyecciones del clima del año 2030 para medir su impacto en el rendimiento del frijol (Capítulo 4). Para esto presentamos una distribución predictiva para temperatura y precipitación que denotamos por  $X_{*t} = (X_{*t}^T, X_{*t}^P)$ .

Dada la distribución posterior de  $\theta$ , primero muestreamos de la distribución predictiva para cada modelo GCM. Esto lo hacemos muestreando

$$\theta_j = (\beta_{xj}, d_j^T, d_j^P, \varepsilon_j^T, \varepsilon_j^P, \alpha^T, \beta^T, \gamma^T, \alpha^P, \beta^P)$$

para  $j = 1, 2, \dots, 15$  de  $\pi(\theta|O_t^T, O_t^P, X_{jt}^T, X_{jt}^P)$  para cada  $t = 1, \dots, 47$ . Luego muestreamos de

$$X_{jt}^T \sim N [\mu_t^T + d_j^T, \varepsilon_j^T]$$

y

$$X_{jt}^P \sim N [\mu_t^P + \beta_{xj}(X_{jt}^T - \mu_t^T - d_j^T) + d_j^P, \varepsilon_j^P].$$

$X_{*t} = (X_{*t}^T, X_{*t}^P)$  para  $t = 1, \dots, 47$  es el promedio sobre  $j$  de  $X_{jt}^T$  y  $X_{jt}^P$ , respectivamente, es decir  $\left( \frac{\sum_{j=1}^{15} X_{jt}^T}{15}, \frac{\sum_{j=1}^{15} X_{jt}^P}{15} \right)$

### 3.4. Tendencias en el Clima Proyectado

El objetivo es observar cómo el clima va cambiando a través del tiempo durante la temporada de cultivo del frijol, lo haremos mostrando dos tipos de proyecciones: una usando la distribución posterior de las tendencias centrales de temperatura y precipitación denotada por  $\mu_t = (\mu_t^T, \mu_t^P)$  y la otra usando la distribución predictiva posterior de  $X_{*t} = (X_{*t}^T, X_{*t}^P)$  que presentamos en la Sección 3.3.

### 3.4.1. Proyecciones del Clima Usando la Distribución Posterior

En el caso de la distribución posterior conjunta de las tendencias centrales del clima se asume que el comportamiento en temperatura y precipitación puede ser modelado con una función lineal por partes (ver ecuación 3.1), por lo que nuestros parámetros de interés son las tendencias observadas ( $\beta^T$  y  $\beta^P$ ) y las tendencias futuras ( $\gamma^T$  y  $\gamma^P$ ) para temperatura y precipitación respectivamente. Examinaremos el efecto de estos parámetros usando intervalos de credibilidad que se estiman a partir de la distribución posterior.

La Figura 3-8 muestra los intervalos de credibilidad del 95 % para las tendencias anuales en temperatura, por país. La gráfica superior muestra los intervalos de credibilidad para  $\beta^T$  que representa el cambio anual observado. De esta gráfica se puede ver que todos los intervalos excepto uno están sobre la recta  $x = 0$  de manera que indican que el incremento actual en temperatura estimado por nuestro modelo hasta el año 2007 es positivo y significativo para estos países. Sólo para Perú hay tendencia de disminución en la temperatura. La gráfica inferior muestra el incremento (decremento) en las tendencias de temperatura a partir del 2007 respecto al incremento en los años anteriores. En varios países se observa que el aumento de temperatura a partir de este año será mayor dado que los intervalos son positivos, como son el caso de Australia, Romania, Africa del Sur y Francia. Sin embargo, se puede observar 2 países (Nepal y Venezuela) en dónde se espera que las tendencias bajen respecto de las tendencias observadas.

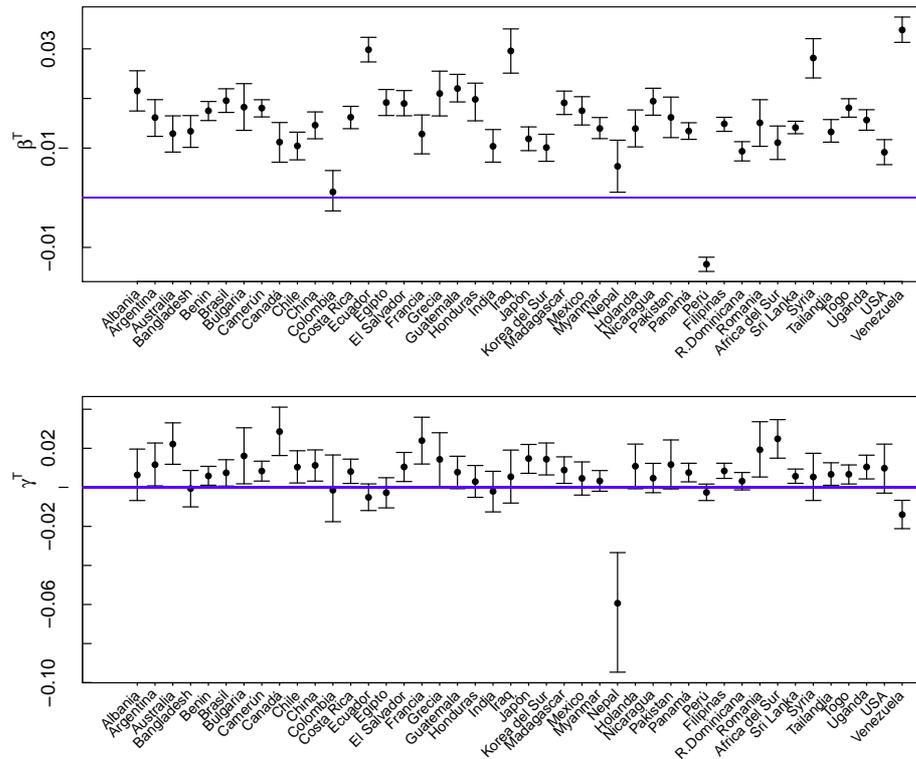


Figura 3–8: Intervalo de credibilidad del 95 % de la tendencia actual ( $\beta^T$ ) y futura ( $\gamma^T$ ) de temperatura en la temporada del crecimiento del frijol para todos los países incorporando GCMs.

Para precipitación, en la gráfica superior de la Figura 3–9, se observa que en algunos países como Albania, Egipto y Grecia las tendencias actuales son significativamente diferente de cero y son negativas, lo que indica que hay tendencia a que la precipitación anual disminuya. Sólo para el caso de Costa Rica e India se observa tendencias positivas significativas. La gráfica inferior muestra que a partir del 2007 la disminución en las tendencias es mayor respecto a las tendencias observadas, para Australia, El Salvador y Francia. Para Nepal, se espera que la tendencia anual en la cantidad de precipitación aumente.

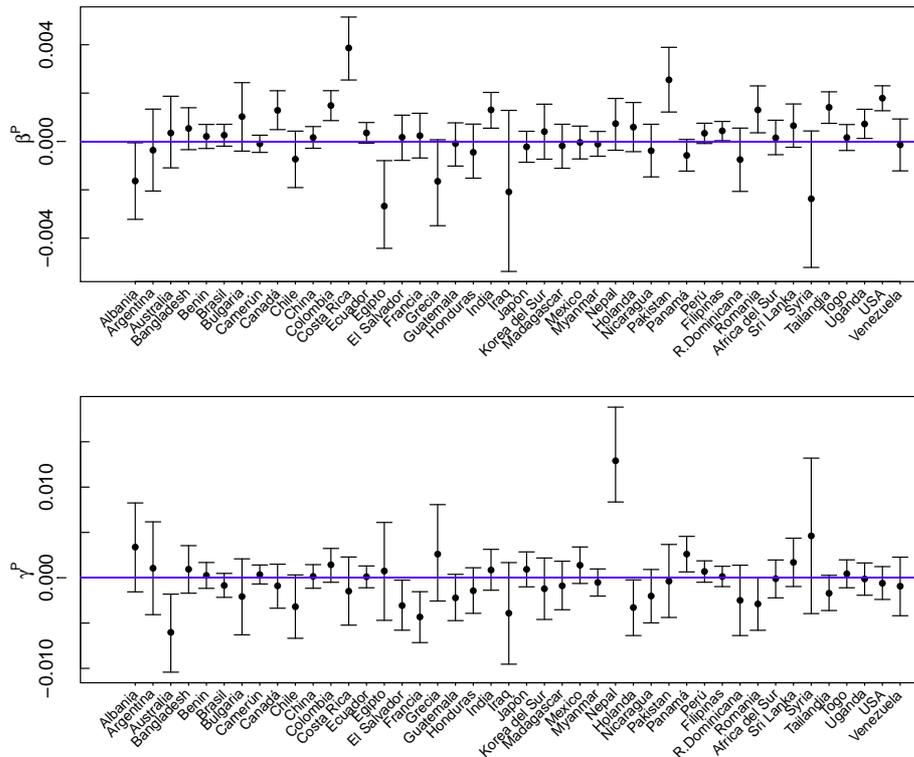


Figura 3-9: Intervalo de credibilidad del 95 % de la tendencia actual ( $\beta^P$ ) y futura ( $\gamma^P$ ) de precipitación en la temporada del crecimiento del frijol para todos los países incorporando GCMs

Para conocer cual sería la tendencia en el clima en el período 2007-2030 encontramos  $\beta^T + \gamma^T$  y  $\beta^P + \gamma^P$ , que toman en cuenta tendencias observadas más el incremento (decremento) a partir del 2007. La gráfica superior de la Figura 3-10 muestra que para este período las tendencias en temperatura serán positivas y significativamente diferentes de cero para 41 de los 44 países. En el caso de Nepal se esperan tendencias negativas para este período a diferencia de las tendencias observadas. Perú sigue mostrando tendencias negativas aunque ya no en la misma intensidad que las tendencias observadas. En cuanto a precipitación (gráfica inferior de la Figura 3-10) las tendencias para el período 2007-2030 son negativas y significativas para Australia, El Salvador y Holanda. Sólo en el caso de Nepal, hay posibilidad de observar tendencias positivas significativas. Los demás países no muestran tendencias significativas.

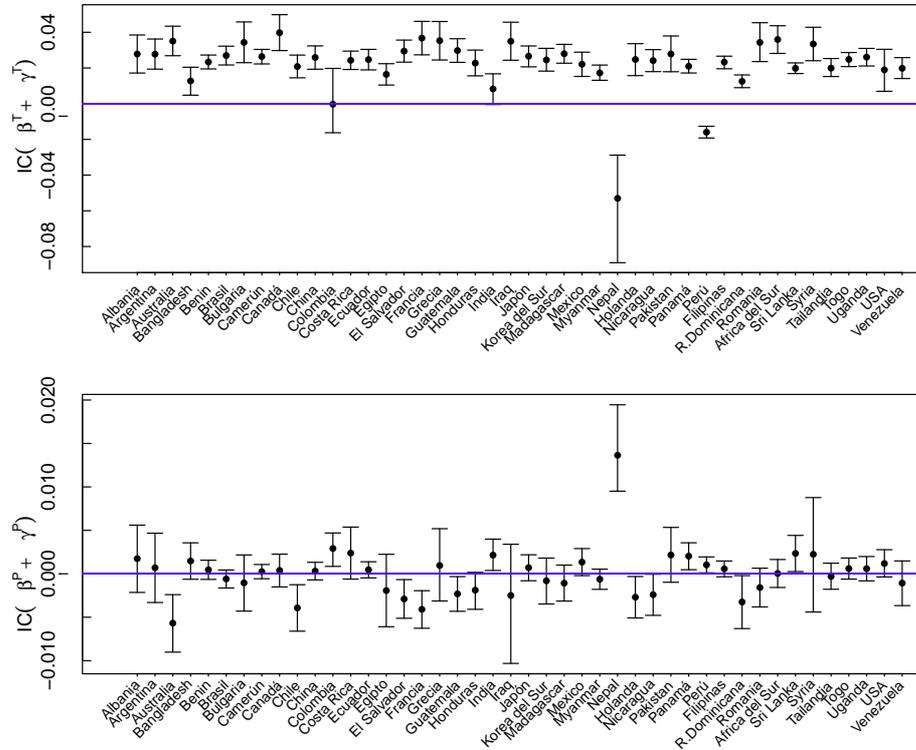


Figura 3–10: Intervalo de credibilidad del 95 % de la tendencia futura  $(\beta^T + \gamma^T)$  y  $(\beta^P + \gamma^P)$  de temperatura y precipitación, respectivamente, para la temporada de crecimiento del frijol para todos los países incorporando GCMs.

Además de las tendencias del clima, otros parámetros de interés son la correlación entre temperatura y precipitación. Para estudiarlos usaremos la distribución posterior para los coeficientes de correlación  $(\beta_{xj}, j = 1, 2, 3, \dots, 15)$  de cada modelo y para  $\beta_0$  que representa la correlación media.

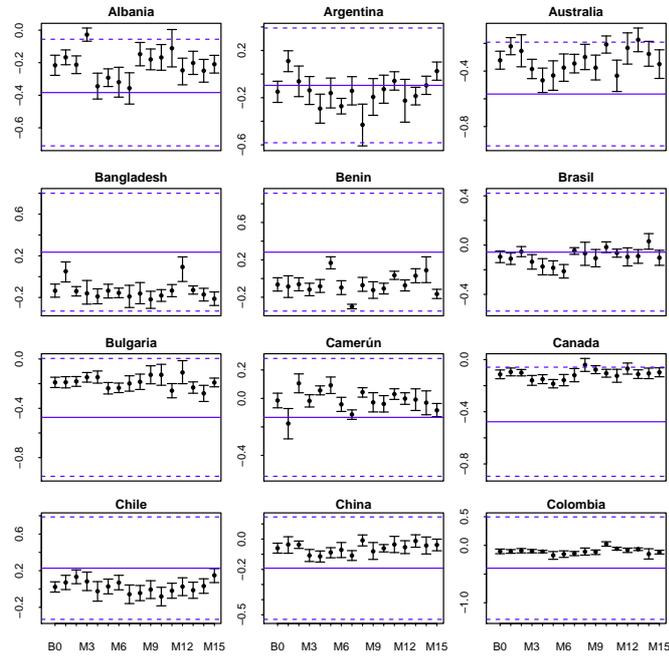


Figura 3–11: Intervalo de credibilidad del 95 % de los coeficientes de correlación ( $\beta_0$ ) y futura ( $\beta_{x_j}$ ), identificando la correlación observada ( $\beta_{x_0}$ )

Las Figura 3–11 a 3–14 muestran los intervalos de credibilidad para la correlación entre temperatura y precipitación en cada país. El primer intervalo de la izquierda de cada recuadro representa el parámetro  $\beta_0$  y los siguientes en cada cuadro son los  $\beta_{x_j}$ ,  $j = 1, 2, \dots, 15$ . El valor observado  $\beta_{x_0}$  viene siendo la línea horizontal y las líneas entrecortadas marcan una distancia de 2 desviaciones estándar de  $\beta_{x_0}$ . Intervalos que quedan dentro de las líneas entrecortadas muestran que la estimación de la correlación correspondiente por el modelo concuerda con el observado. Lo que se observa en general para todos los países y modelos GCMs, excepto para África del Sur y Estados Unidos. En algunos países es evidente la diferencia en la estimación para los distintos GCMs. Tebaldi y Sansó, (2008) sugieren incluir una distribución para el parámetro  $\beta_{x_0}$ .

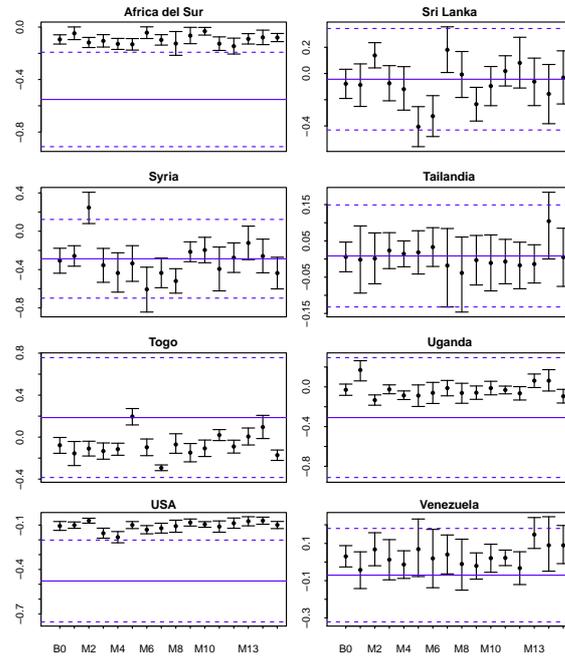


Figura 3–12: (Continuación de la Figura 3–11) intervalo de credibilidad del 95% de los coeficientes de correlación ( $\beta_0$ ) y futura ( $\beta_{xj}$ ), identificando la correlación oservada ( $\beta_{x0}$ ).

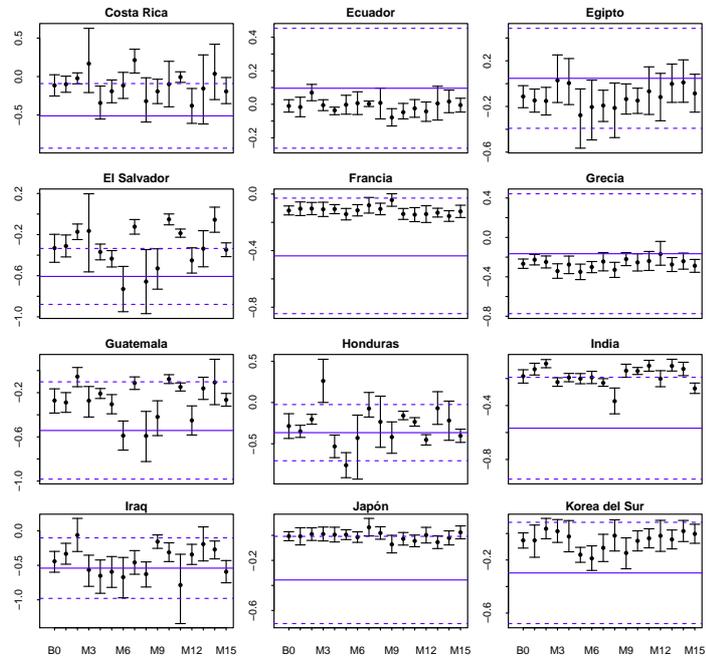


Figura 3–13: (Continuación de la Figura 3–11) intervalo de credibilidad del 95% de los coeficientes de correlación ( $\beta_0$ ) y futura ( $\beta_{xj}$ ), identificando la correlación oservada ( $\beta_{x0}$ ).

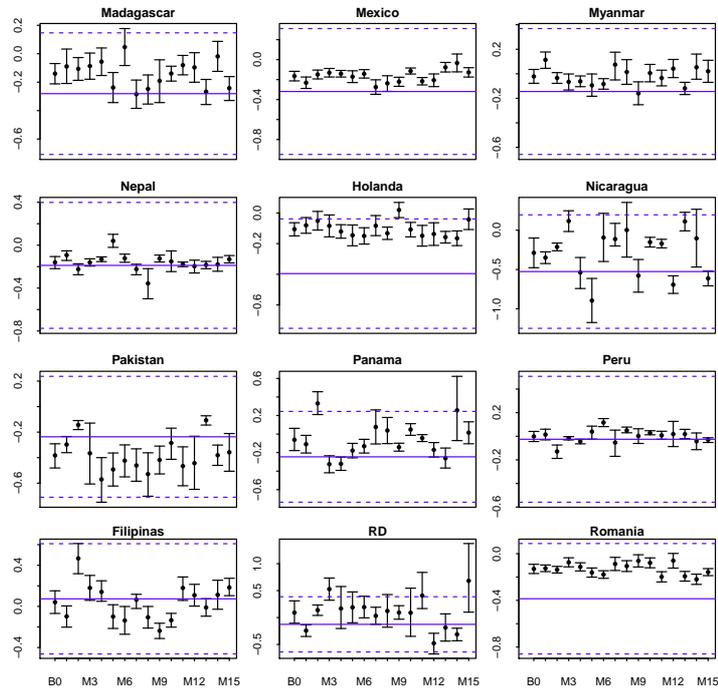


Figura 3–14: (Continuación de la Figura 3–11) intervalo de credibilidad del 95% de los coeficientes de correlación ( $\beta_0$ ) y futura ( $\beta_{xj}$ ), identificando la correlación observada ( $\beta_{x0}$ ).

Este modelo ofrece muchas otras distribuciones marginales posteriores de los parámetros que pueden ser de interés, de forma aislada o conjuntamente. Por ejemplo, podrían ser de interés modelos específicos, o en general, los sesgos ( $d_{jt}$ ), y precisiones ( $\varepsilon_{jt}$ ) tanto para temperatura como para precipitación.

### 3.4.2. Comparación de las Tendencias Observadas con las Tendencias Futuras

En la Sección 3.2 comparamos las tendencias de dos períodos: 1961-1984 y 1984-2007. Se observó que en el período 1984-2007, treinta de los 44 países aumentaron su temperatura con respecto a los 24 años anteriores y que la precipitación no mostró tendencias diferentes en los dos períodos. Para cuantificar este cambio se dividió entre la desviación estándar histórica (1961-2000). Ahora que tenemos la distribución posterior de  $\gamma^T$  y  $\gamma^P$  que representan las tendencias futuras del clima,

podemos comparar las tendencias del clima futuro (2007-2030) respecto a los dos períodos anteriores. Para esto, muestreamos para cada país la distribución posterior de  $\mu_t^T$  y  $\mu_t^P$ , para los años 2007 y 2030 y calculamos la diferencia promedio entre estos dos años dividida entre su desviación estándar histórica para obtener una tendencia normalizada comparable a la calculada en la Sección 3.2

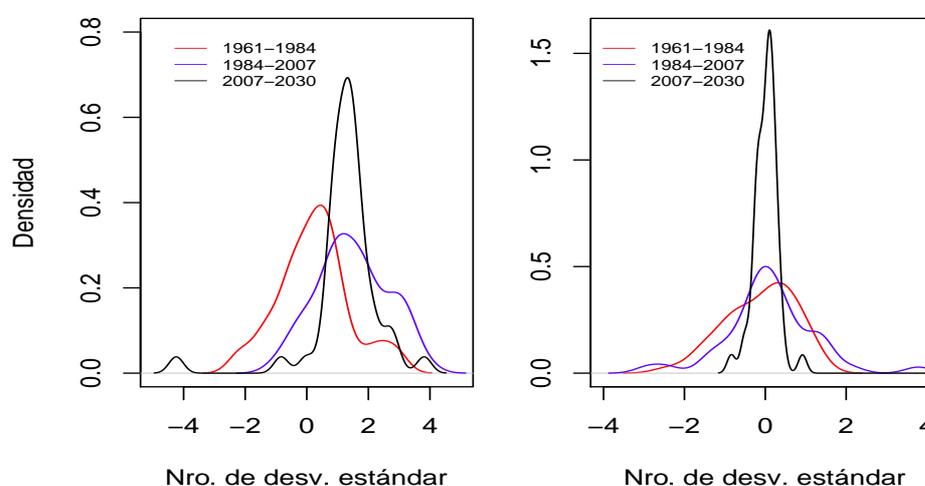


Figura 3–15: Comparación de las tendencias por períodos. Figura de la izquierda para temperatura y derecha para precipitación.

La Figura 3–15 muestra las densidades para los tres períodos. La gráfica de la izquierda para temperatura y la gráfica de la derecha para precipitación. Se puede observar de manera general que las tendencias futuras de temperatura son similares a las tendencias del período 1984-2007, y que las tendencias para los últimos dos períodos tienden a ser mayores comparadas con la de los años 1961-1984. Para precipitación se observa tendencias similares en los tres períodos.

Si nos concentramos en las tendencias del período 2007-2030 se observa que hay mayor probabilidad de que la magnitud de las tendencias en temperatura estén entre 0 y 4, lo que significa que las tendencias en este período podrían ser hasta 4 desviaciones estándar mayores que el año 2007. Las tendencias son positivas en la

mayoría de los países estudiados lo que también mostró la Figura 3-10, sin embargo se puede ver que existe una posibilidad de observar una tendencia negativa de  $-4\sigma$  respecto al año 2007, pero esto ocurre en el caso particular de Nepal que cómo se observó en la Figura 3-10 muestra una tendencia drásticamente negativa.

En el caso de precipitación (gráfica de la derecha de la Figura 3-15) se puede observar que el rango de valores para las tendencias están distribuidas alrededor de cero, esto significa que el número de países donde hay probabilidad de aumento en la cantidad de precipitación es similar al número de países donde hay probabilidad de disminución.

La Figura 3-16 muestra el cambio en las tendencias de temperatura por país. El gráfico de la izquierda muestra la comparación de las tendencias en el período 1961-1984 y 1984-2007, que vimos en la Sección 3.2 y el gráfico de la derecha muestra la comparación de las tendencias observadas en temperatura en el período 1984-2007 y las proyectadas para el período 2007-2030.

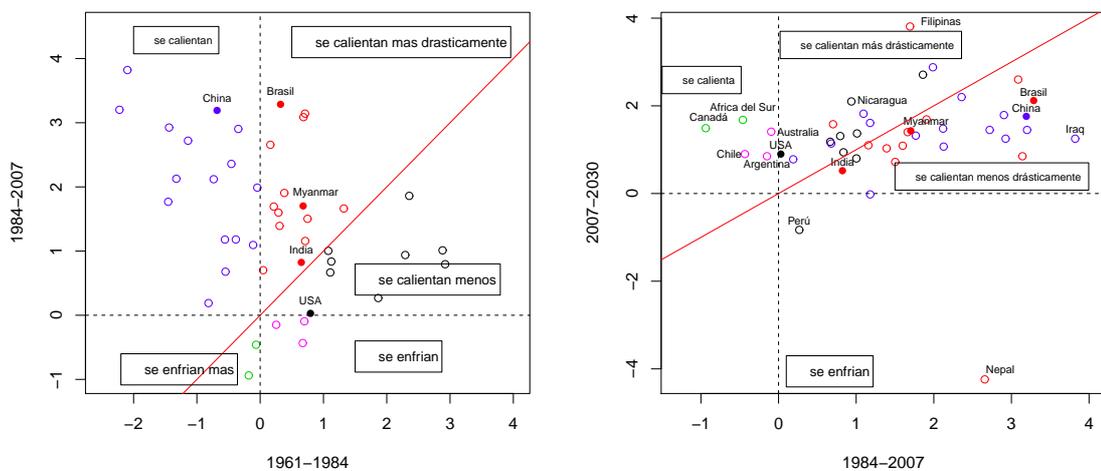


Figura 3-16: Gráfico del cambio de temperatura de 1961-1984 vs 1984-2007 (izquierda) y 1984-2030 vs 2007-2030 (derecha) identificando los 5 primeros productores.

De la gráfica podemos decir que países que se estaban enfriando (Canadá, Africa del Sur, Argentina, Australia y Chile) en el período 1984-2007 pasarían a calentarse para los años 2007-2030. También, países que experimentaban mayor aumento de temperatura, seguirán mostrando esta característica pero muchos de ellos en menor intensidad. De manera similar países que pasaron a calentarse en el período 1984-2007 con respecto al 1961-1984, en la gráfica de la izquierda seguirán calentándose pero con menor intensidad. En general las tendencias para los años 2007-2030 son positivas, lo que significa que seguirán calentándose en este período, con excepción de Perú y Nepal. Para los 5 primeros productores de frijol en el mundo (puntos sólidos) se observa un aumento en temperatura pero las tendencias son menores comparadas a los 20 años anteriores para el caso de Myanmar, Brasil, China e India. Sólo en el caso de Estados Unidos el aumento de temperatura sería mayor que el aumento en los 20 años pasados.

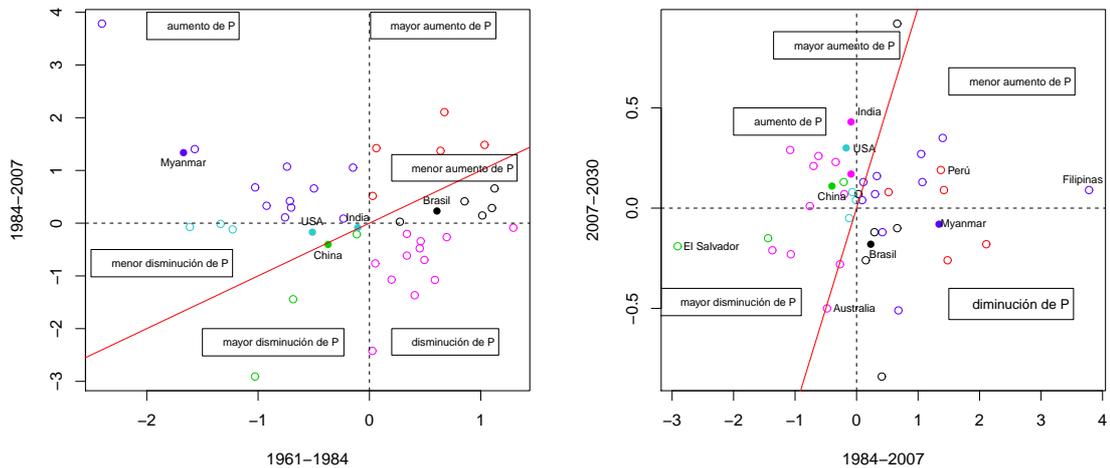


Figura 3-17: Gráfico del cambio de precipitación de de 1961-1984 vs 1984-2007 identificando los 5 primeros productores, sin incluir Nepal

Para precipitación (Figura 3-17) el patrón que observamos antes continua para el periodo 2007-2030; de la cantidad de países con mayor aumento en precipitación es similar a la cantidad de países con menor aumento. Pero resalta el hecho de que la

intensidad de los cambios son menos drásticos, con excepción de Nepal (ver Figura 3–18). Esto se puede observar tanto en la variabilidad de la densidad para 2007-2030 en la Figura 3–15 como en la escala del eje  $y$  en las gráficas de la Figura 3–17. Note que en particular, Filipinas que en el período 1984-2007 tenía un aumento de casi  $4\sigma$  al final del período para el año 2030 el aumento en precipitación es muy pequeño. Por otro lado, en la Figura 3–18, se observa que Nepal sigue mostrando un comportamiento atípico pues antes presentó una baja en la cantidad de precipitación y experimentaría un aumento de casi  $5\sigma$  para el 2030. Esto es contrario a lo que vimos antes con temperatura, lo que muestra una correlación negativa entre temperatura y precipitación.

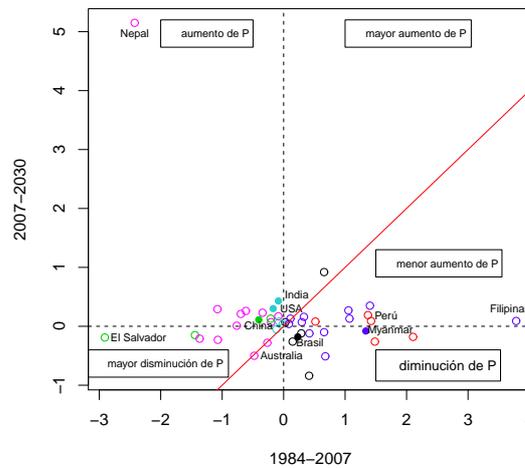


Figura 3–18: Gráfico del cambio de precipitación de de 1961-1984 vs 1984-2007

### 3.4.3. Proyecciones del Clima Usando la Distribución Predictiva

En la Sección 3.3.3 presentamos la distribución predictiva como una alternativa de un nuevo modelo de proyección. Con esta  $X_{*t}$  derivamos 3 distribuciones: la primera que representa el clima actual y es el promedio de las distribuciones predictivas en los años 1997-2007, la segunda que representa el clima futuro y es calculada a partir del promedio de las distribuciones predictivas de los años 2020-2030, y la tercera que representa la distribución predictiva del cambio del clima entre estos dos

periodos.

En las Figuras 3-19 y A-1 hasta A-10, se presentan estas distribuciones para cada país, usando azul para el clima actual y negro para el clima futuro. En la primera columna se observan los contornos de la distribución predictiva posterior conjunta para precipitación y temperatura actual y futura. En la segunda columna presentamos la distribución de la diferencia (futuro-actual) y en la tercera y cuarta columna se muestran las densidades marginales de temperatura y precipitación, respectivamente, para ambos períodos. En la primera columna notamos que para todos los países hay un desplazamiento hacia la derecha de los contornos lo que indica que habría un incremento en la temperatura. En el caso de Estados Unidos (Figura 3-19) el desplazamiento es menos intenso.

En cuanto a precipitación, sólo en Estados Unidos se observa un ligero desplazamiento hacia arriba, que estaría indicando un aumento en la cantidad de precipitación. Otro aspecto importante que se observa en los contornos es la correlación entre temperatura y precipitación, es que en el caso de Estados Unidos y Togo (Figura 3-19) la correlación es negativa, esto se deduce por la forma inclinada del contorno. No obstante, en Venezuela, se observa correlación positiva entre temperatura y precipitación. Para Uganda parece no haber correlación. Esto también se puede verificar en la Figura 3-14 donde el primer intervalo que muestra la correlación media contiene al cero.

De la segunda y tercera columna de estas figuras podemos observar claramente un aumento en temperatura dado que los valores de las diferencias de temperatura son positivos (futuro-actual). El desplazamiento en las densidades marginales es

claro. Para precipitación, sólo para Estados Unidos y Uganda (Figura 3–19) se observa un desplazamiento hacia la derecha, lo que ya habíamos mencionado con los contornos, pero este desplazamiento no es significativo puesto que en el contorno de las diferencias incluye el cero en el eje  $y$ .

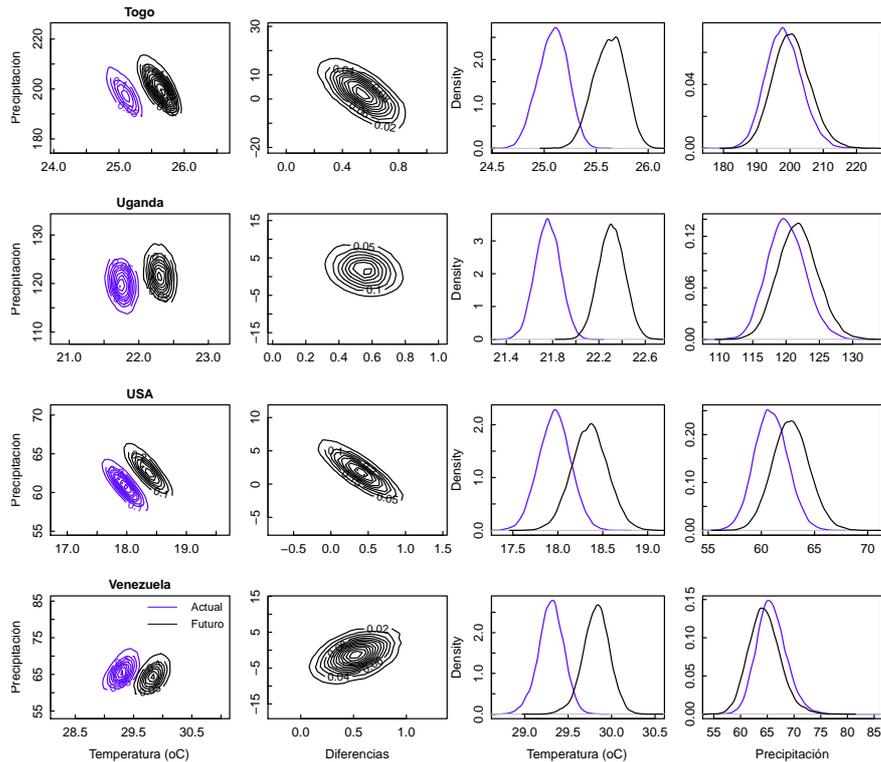


Figura 3–19: Contours de la distribución predictiva conjunta de temperatura y precipitación comparando el período actual (1997-2007) y futuro (2020-2030) por países y densidades marginales para temperatura y precipitación comparando estos 2 períodos, el período actual es azul y futuro negro

De manera similar se pueden realizar la interpretaciones para los demás países con las figuras del Apéndice A.1. De ellas, podemos resaltar a Francia (Figura A–5) que muestra aumento en temperatura y disminución en precipitación, además, estos cambios son significativos puesto que no contienen al cero, esto se observa en la gráfica del contorno de las diferencias. Otros países a resaltar son: Chile que muestra una ligera disminución en precipitación (Figura A–3) y Nepal que presenta un

aumento en precipitación (Figura A-8), sin embargo en ambos casos el contorno de las diferencias contiene al cero. En general se espera un aumento en la temperatura para el período 2020-2030, y la precipitación no muestra cambios significativos.

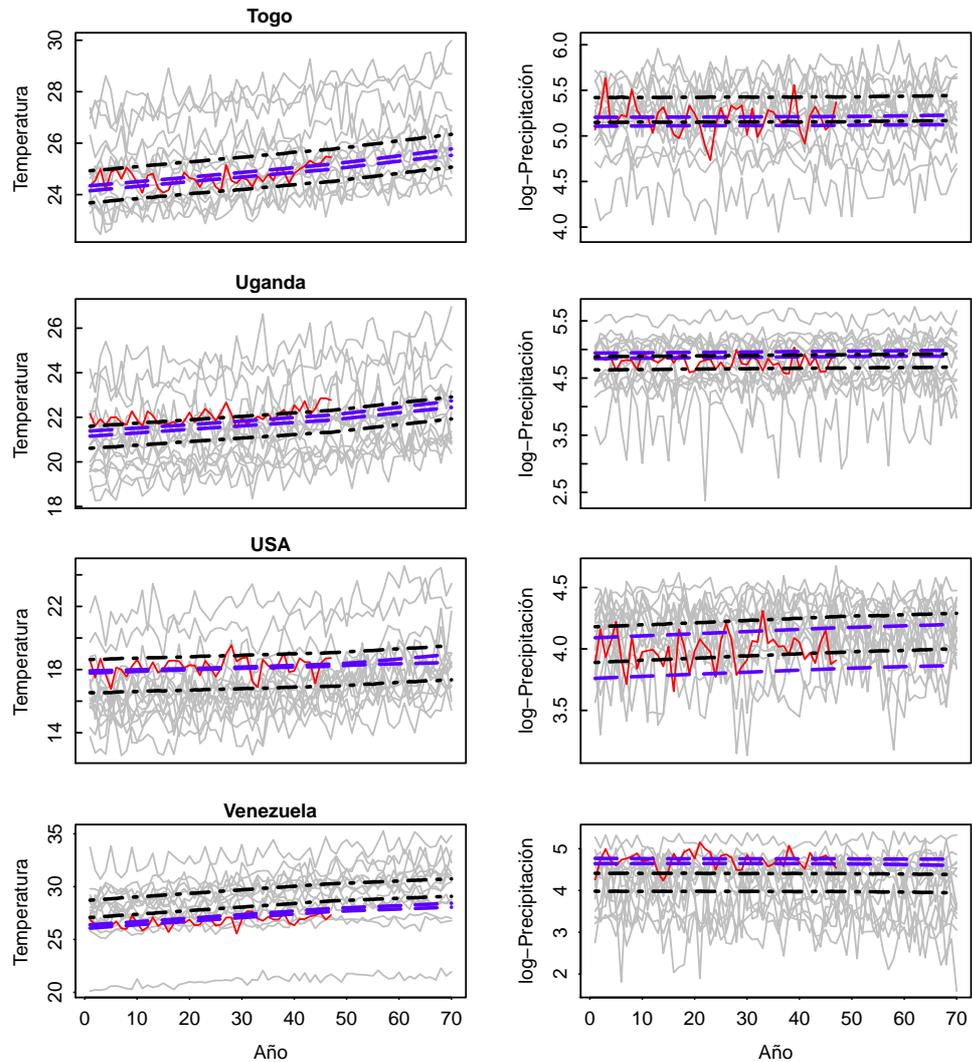


Figura 3-20: Series de tiempo de temperatura y log-precipitación, líneas plomas son las series proyectadas por los 15 GCMs, líneas rojas series observadas, líneas azules representan los intervalos de credibilidad del 95 % para para el modelo del clima ( $\mu_t$ ) y líneas negras representan el IC del 95 % para la distribución predictiva.

Para las Figuras 3-20 y A-11 hasta A-20 las series de color gris son las series proyectadas para los 15 modelos GCM. Las líneas de color azul representa el intervalo de confianza de las tendencias centrales ( $\mu_t$ ), y las líneas de color negro

representan el intervalo de confianza de la distribución predictiva, y las líneas de color rojo representan a las series observadas.

Se observa que los intervalos de credibilidad para  $\mu_t$  para los países mostrados en esta gráfica (Togo, Uganda, Estados Unidos y Venezuela), contienen o pasan por los valores observados pero esto no sucede en general, por ejemplo, para Argentina (Figura A-11 del Apéndice A.2) las temperaturas observadas están fuera del intervalo de credibilidad para las tendencias centrales, esto se da porque, este intervalo depende de los modelos de proyección GCMs, y  $\mu_t$  se entiende como la tendencia central que podría seguir el clima. En el caso del intervalo de credibilidad de la distribución predictiva, se observa que es más ancho que el de la distribución posterior de  $\mu_t$ , lo cual tiene sentido porque esperamos que el rango de valores para la distribución predictiva sea más amplio, y para tres de estos países (Togo, Uganda y Estados Unidos) este intervalo contiene a los valores observados, sin embargo esto no necesariamente debe suceder siempre, así observamos el caso de Venezuela en el que se observa lo contrario, esto es porque la distribución predictiva se entiende como un nuevo modelo GCM que describe un posible clima diferente o igual al observado, basado en los modelos de proyección.

## Capítulo 4

# MODELO DEL RENDIMIENTO

### 4.1. Descripción de los Datos de Rendimiento

Datos de promedios anuales de rendimiento (producción por hectárea) de frijol obtenidos de la página de *Food and Agricultural Organization of the United Nations* (FAO, <http://faostat.fao.org>) para los años 1961-2007 fueron combinados con temperatura y precipitación (temperatura y precipitación fueron agregados de acuerdo a lo descrito en el Capítulo 3), de manera que contamos con datos de rendimiento, temperatura y precipitación para cada año y para cada país, el rendimiento está medido en *hg/ha*, la temperatura en °C y la precipitación en *mm/mes*. Se consideró sólo hasta el año 2007 debido a que no se cuentan con datos de clima para los años restantes.

Para medir propiamente el efecto del clima en el rendimiento tratamos de controlar aquellos factores que podrían sesgar los resultados (Lobell et al., 2011), por lo que se consideraron países que tienen más de 1000 hectáreas dedicadas al cultivo de frijol (información que también fue obtenida de FAO), y países que no muestran el mismo rendimiento repetido para dos o tres años consecutivos. Los 44 países que consideramos en nuestro estudio están en la Tabla B-1.

## 4.2. Modelo del Rendimiento en Respuesta a la Variación del Clima

El rendimiento del frijol está sujeto a factores climáticos (temperatura y precipitación). En el Capítulo 3 vimos que el clima está cambiando y es de interés medir el impacto de este cambio sobre el rendimiento de frijol. Sin embargo, el rendimiento de cultivos no sólo varía en función de factores climáticos, si no que también responde a la variación en el tiempo, debido al mejoramiento y el acceso a nuevas tecnologías.

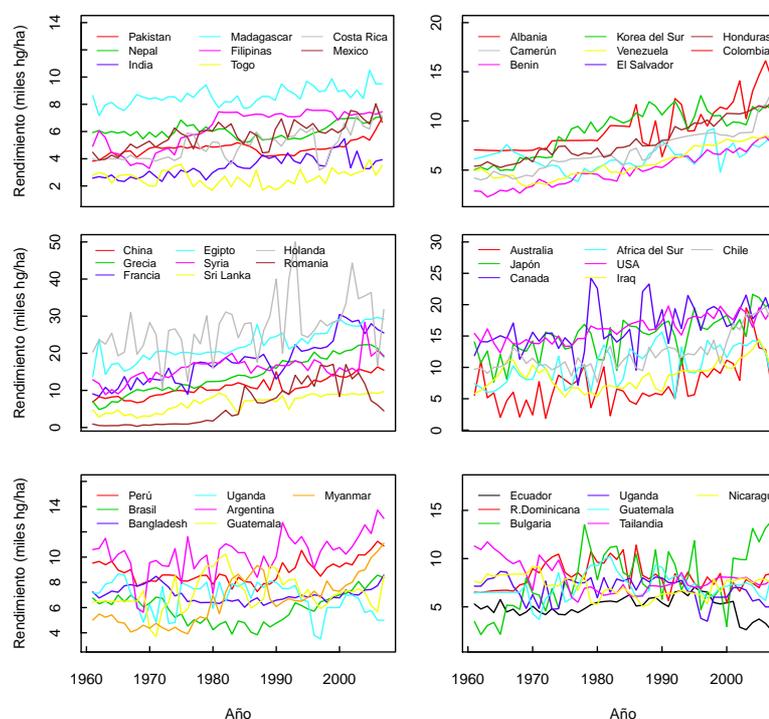


Figura 4-1: Series del promedio anual de rendimiento de frijol de 1961-2007

La Figura 4-1 muestra cómo el rendimiento va cambiando a través del tiempo. En ella se puede observar que el rendimiento anual de frijol para la mayoría de los países está en aumento, en algunos, el aumento es más rápido que en otros. También se observa en las gráficas de la última fila que el cambio del rendimiento no necesariamente es lineal, si no que más bien parece haber una relación cuadrática entre rendimiento y tiempo, aspecto que será considerado en nuestro análisis.

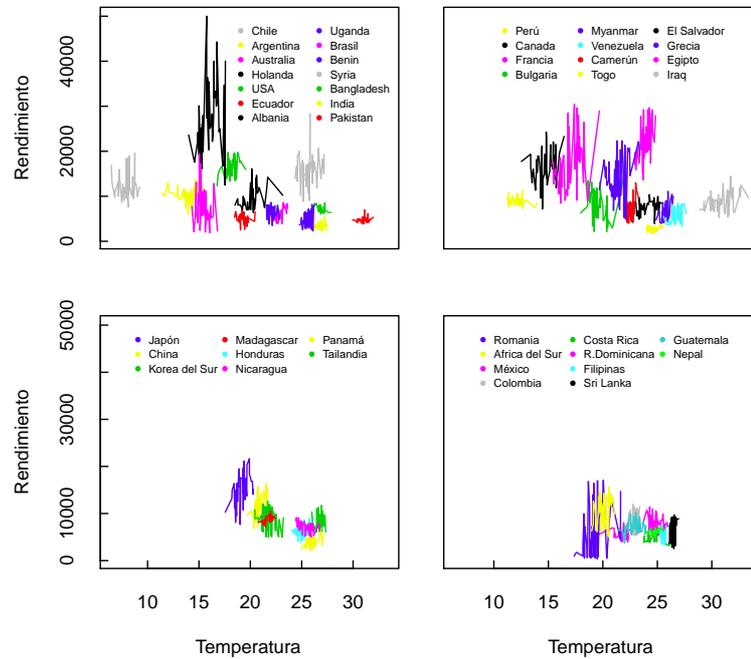


Figura 4-2: Series del promedio anual de rendimiento (hg/ha) - temperatura ( $^{\circ}$ C) de frijol de 1961-2007

La Figura 4-2 muestra que el rendimiento en función de temperatura varía por países. Si observamos los datos sin considerar la variable país, podríamos decir que el rendimiento tiende a bajar, cuando la temperatura aumenta, pero dentro de cada país no necesariamente sucede esto. Para ver esto más claramente en la Figura 4-3 vemos las mismas series de rendimiento de la Figura 4-2, pero eliminando la magnitud de la temperatura. Se observa que la relación entre rendimiento y temperatura no es la misma para todos los países.

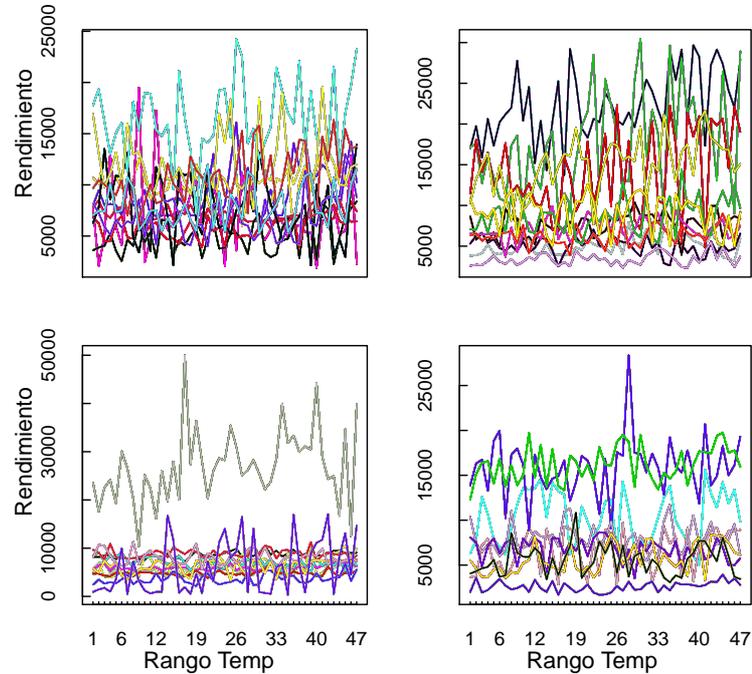


Figura 4-3: Series del promedio anual de rendimiento (hg/ha) - temperatura de frijol de 1961-2007 eliminando la magnitud de temperatura.

En cuanto a la precipitación en la Figura 4-4, sucede algo similar a lo que ocurre en temperatura. El rendimiento tiende a bajar a medida que la precipitación aumenta, pero dentro de cada país no ocurre necesariamente esto. En algunos casos, el rendimiento no varía respecto a la precipitación (Figura superior izquierda, líneas amarillas). La Figura 4-5 muestra el rendimiento en función de la precipitación. Similarmente a la temperatura se observa que el rendimiento en función de precipitación varía por país.

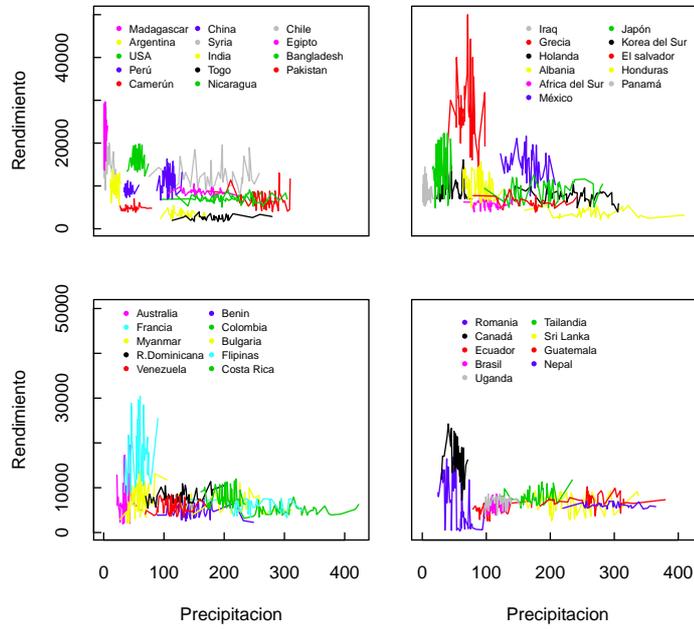


Figura 4-4: Series del promedio anual de rendimiento (hg/ha) - precipitación (mm/mes) de frijol de 1961-2007

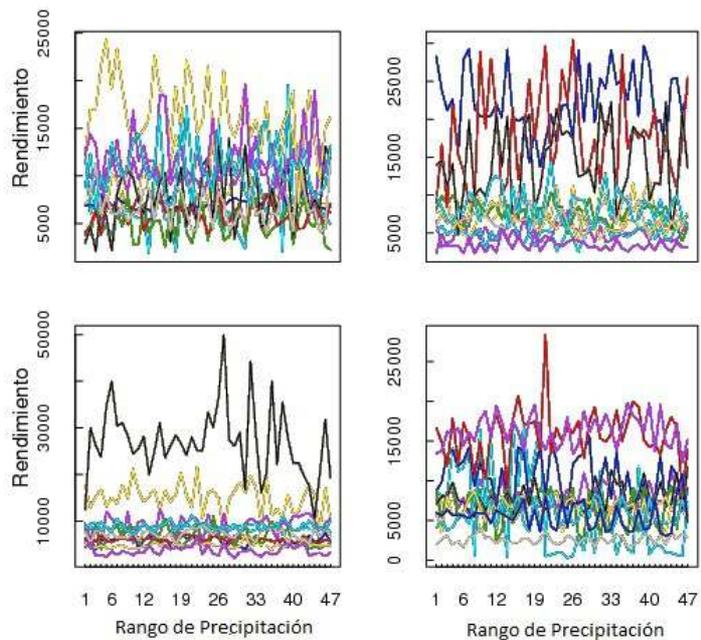


Figura 4-5: Series del promedio anual de rendimiento (hg/ha)- precipitación de frijol de 1961-2007 eliminando la magnitud de precipitación.

Del estudio del clima (Capítulo 3) concluimos que, la temperatura está y muy probablemente seguirá en aumento, y que la precipitación muestra cambios variables por país. En la Figura 4-1 se ve un claro aumento en el rendimiento, pero el hecho de que dos variables aumenten en el tiempo no implica una relación causal, es decir, un aumento en temperatura y un aumento en el rendimiento no necesariamente implica que el calentamiento sea beneficioso; dado que existen otros factores que pueden impulsar el rendimiento, como por ejemplo el uso de tecnología, y el uso de fertilizantes. De aquí que, además de incorporar en nuestro modelo tendencias en el tiempo y las tendencias en el tiempo al cuadrado, también consideraremos que tanto éstas como el intercepto pueden producir efectos específicos para cada país (no comunes a todos), ya que respuestas al clima pueden depender de otros factores y condiciones (variables omitidas o no observables), que varían por país.

En nuestro estudio utilizaremos el modelo lineal mixto discutido en la sección 2.3 debido a que existe correlación dentro de los países (individuos), así como también consideramos que existe variabilidad entre países debido a factores no observables, además de que el modelo de efectos mixtos, a diferencia de un modelo de efectos fijos, nos permite reducir el número de parámetros a estimar. Usaremos un modelo logarítmico del rendimiento para tener una distribución log-normal y porque la varianza del rendimiento se mantiene similar en términos relativos (Lobell et al., 2011), pero no en términos absolutos. Utilizar el logaritmo del rendimiento nos permite tener una variable que se ajusta más a una distribución normal.

#### 4.2.1. Modelo Lineal Mixto Para Datos Observados

Consideremos el modelo, (al que llamaremos modelo 1)

$$\log(y_{ij}) = (a + \mu_{i1}) + (b + \mu_{i2})t_{ij} + (c + \mu_{i3})t_{ij}^2 + (d)T_{ij} + (e)\log(P_{ij}) + \varepsilon_{ij} \quad (4.1)$$

donde se asume que  $\epsilon_{ij} \sim iidN(0, \sigma^2)$ ,  $\mu_i \sim N(0, \Omega)$ , y

$$\Omega = \begin{bmatrix} \sigma_{\mu_{i1}}^2 & \sigma_{\mu_{i1}\mu_{i2}} & \sigma_{\mu_{i1}\mu_{i3}} \\ \sigma_{\mu_{i2}\mu_{i1}} & \sigma_{\mu_{i2}}^2 & \sigma_{\mu_{i2}\mu_{i3}} \\ \sigma_{\mu_{i3}\mu_{i1}} & \sigma_{\mu_{i3}\mu_{i2}} & \sigma_{\mu_{i3}}^2 \end{bmatrix}$$

es la matriz de covarianzas de los efectos aleatorios.

$t_{ij}$ : Es el tiempo.

$t_{ij}^2$ : Es el cuadrado de el tiempo.

$T_{ij}$ : Representa la temperatura.

$\log(P_{ij})$ : Es el logaritmo de la precipitación.

$i = 1, \dots, n$  ( $n$  número de países, en nuestro caso  $n = 44$ ) y

$j = 1, \dots, \tau$  ( $\tau$  número de observaciones por país, que son los años en el período 1961-2007), entonces  $\tau = 47$

Si denotamos  $Y_i$  como el vector de observaciones para el país  $i$ , tenemos

$$Y_i = x_i\varphi + w_i\mu_i + \epsilon_i \quad (4.2)$$

donde

$$x_i = \begin{bmatrix} 1 & t_1 & t_1^2 & T_1 & \log(P_1) \\ 1 & t_2 & t_2^2 & T_2 & \log(P_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{47} & t_{47}^2 & T_{47} & \log(P_{47}) \end{bmatrix}_{47 \times 5}, \quad w_i = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_{47} & t_{47}^2 \end{bmatrix}_{47 \times 3},$$

$$\varphi = \begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix} \quad \text{y} \quad \mu_i = \begin{bmatrix} \mu_{i1} \\ \mu_{i2} \\ \mu_{i3} \end{bmatrix}$$

Como se mencionó en la Sección 2.1 en el contexto Bayesiano nos interesa muestrear de la distribución a posteriori de los parámetros dado el conjunto de datos que está dada por

$$\begin{aligned}
& \pi(\varphi, \Omega, \sigma^2 \mid y_1, \dots, y_n) \\
&= \prod_{i=1}^n f(\varphi, \Omega, \sigma^2 \mid y_i) \\
&= \prod_{i=1}^n \int f(\varphi, \Omega, \sigma^2, \mu_i \mid y_i) d\mu_i \\
&= \int \cdots \int \left[ \prod_{i=1}^n f(\varphi, \Omega, \sigma^2, \mu_i \mid y_i) \right] d\mu_1^1, \dots, \mu_i^n \\
&\propto \int \cdots \int \left[ \prod_{i=1}^n f(y_i \mid \varphi, \Omega, \sigma^2, \mu_i) f(\varphi, \Omega, \sigma^2, \mu_i) \right] d\mu_1^1, \dots, \mu_i^n \\
&= \int \cdots \int \left[ \prod_{i=1}^n f(y_i \mid \varphi, \Omega, \sigma^2, \mu_i) f(\varphi \mid \sigma^2) f(\sigma^2) f(\mu_i \mid \Omega) f(\Omega) \right] d\mu_1^1, \dots, \mu_i^n
\end{aligned} \tag{4.3}$$

donde  $f(y_i \mid \varphi, \Omega, \sigma^2, \mu_i)$  es la función de distribución de  $y_i$  y  $f(\varphi \mid \sigma^2)$ ,  $f(\sigma^2)$ ,  $f(\mu_i \mid \Omega)$  y  $f(\Omega)$  son funciones a priori para los parámetros  $\varphi$ ,  $\sigma^2$ ,  $\mu_i (i = 1, \dots, n)$  y  $\Omega$  respectivamente.

Si consideramos las distribuciones a priori conjugadas (Daniels y Hogan, 2008)

$$\varphi \sim N(0, \tau^2 I)$$

$$\frac{1}{\sigma^2} \sim G(h, g)$$

$$\Omega^{-1} \sim Wishart(p, A)$$

$$\mu_i \sim N(0, \Omega)$$

donde  $\tau^2$ ,  $h$ ,  $g$  y  $p$  son hiperparámetros, entonces

$$\begin{aligned} \pi(\varphi, \Omega, \sigma^2 \mid y_1, \dots, y_n) \propto & \int \cdots \int \left\{ \prod_{i=1}^n \left[ \left( \frac{1}{\sigma^2} \right)^{n/2} e^{-\frac{1}{2\sigma^2} (y_i - x_i \varphi - w_i \mu_i)' (y_i - x_i \varphi - w_i \mu_i)} \right. \right. \\ & e^{-\frac{1}{2\tau^2} \varphi' \varphi} e^{-\frac{1}{\sigma^2 g} \left( \frac{1}{\sigma^2} \right)^{h-1} \mid \Omega_r^{-1} \mid^{-\frac{1}{2}} e^{-\frac{1}{2} \mu_i' \Omega^{-1} \mu_i} \\ & \left. \left. \mid \Omega_r^{-1} \mid^{\frac{r-p-1}{2}} e^{-\tau(A^{-1} \Omega^{-1})} \right] \right\} d\mu_1, \dots, d\mu_n \end{aligned} \quad (4.4)$$

Para obtener muestras de esta distribución podemos usar el algoritmo de Gibbs que describimos en la Sección 2.2.1. Para implementar este algoritmo necesitamos las distribuciones condicionales de cada uno de los parámetros. En este caso tenemos dos opciones, la primera es calcular las integrales en (4.3) y luego hallar las distribuciones condicionales. Por ejemplo, para un  $i$  particular, la distribución marginal de  $y_i$  está dada por

$$\begin{aligned} f(y_i \mid \sigma^2, \Omega, \varphi) & \propto \int e^{-\frac{1}{2\sigma^2} (y_i - x_i \varphi - w_i \mu_i)' (y_i - x_i \varphi - w_i \mu_i)} e^{-\frac{1}{2} \mu_i' \Omega^{-1} \mu_i} d\mu_i \\ & = e^{-\frac{1}{2} \left( (y_i - x_i \varphi)' \left( \frac{1}{\sigma^2} \right)^{-1} (y_i - x_i \varphi) - \left( (y_i - x_i \varphi)' \left( \frac{1}{\sigma^2} \right)^{-1} w_i \right) \mu_i \right)} \\ & * e^{\frac{1}{2} \left( \mu_i' \left( (w_i)' \left( \frac{1}{\sigma^2} \right)' (y_i - x_i \varphi) \right) + \mu_i' \left( w_i' \left( \frac{1}{\sigma^2} \right)' w_i + \Omega^{-1} \right) \mu_i \right)} \\ & = e^{-\frac{1}{2} \left( (y_i - x_i \varphi)' \left( \frac{I}{\sigma^2} - \frac{w_i' V^{-1} w_i}{\sigma^2} \right) (y_i - x_i \varphi) \right)} \\ & = e^{-\frac{1}{2} \left( (y_i - x_i \varphi)' \left( w_i \Omega w_i' + I \sigma^2 \right)^{-1} (y_i - x_i \varphi) \right)} \end{aligned} \quad (4.5)$$

donde  $V = \frac{w_i' w_i}{\sigma^2} + \Omega^{-1}$ . La última línea de la ecuación (4.5) se obtiene por la regla de Schur (Ver en el Apéndice C.2) El producto en la ecuación (4.3) ahora incluirá términos de  $\Omega$  y  $\sigma^2$  dentro de un determinante lo que complica el cálculo de las distribuciones condicionales para  $\Omega$  y  $\sigma^2$ .

La otra opción es considerar los  $\mu_i$  ( $i = 1, \dots, n$ ) como parte del Gibbs Sampler de modo que las distribuciones condicionales esten condicionados en los  $\mu_i$ . A partir de la ecuación (4.4) derivamos las ecuaciones condicionales. Para esto consideramos

sólo los términos que incluyen el parámetro correspondiente de interés, y tenemos que:

La distribución condicional de  $\frac{1}{\sigma^2} \mid \varphi, \sigma^2, \Omega, \mu_i, y_i, x_i$  ( $i = 1, \dots, n$ ) está dada por

$$\begin{aligned} f\left(\frac{1}{\sigma^2} \mid \varphi, \sigma^2, \Omega, \mu_i, y_i, x_i\right) &\propto \prod_{i=1}^n \left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(y_i - x_i\varphi + w_i\mu_i)'(y_i - x_i\varphi + w_i\mu_i)} e^{-\frac{1}{\sigma^2 g}} \left(\frac{1}{\sigma^2}\right)^{h-1} \\ &= \left(\frac{1}{\sigma^2}\right)^{\frac{n\tau_0}{2} + h - 1} e^{-\frac{1}{\sigma^2} \left\{ \frac{\sum (y_i - x_i\varphi + w_i\mu_i)'(y_i - x_i\varphi + w_i\mu_i)}{2} + \frac{1}{g} \right\}} \end{aligned} \quad (4.6)$$

por lo que la distribución de  $\frac{1}{\sigma^2} \mid \varphi, \sigma^2, \Omega, \mu_i, y_i, x_i$  ( $i = 1, \dots, n$ )

$$\text{es } G\left(\frac{N\tau_0}{2} + h, \left[ \frac{\sum (y_i - x_i\varphi + w_i\mu_i)'(y_i - x_i\varphi + w_i\mu_i)}{2} + \frac{1}{g} \right]^{-1}\right)$$

La distribución condicional de  $\Omega_r^{-1} \mid \sigma^2, \varphi, \mu_i, y_i, x_i$  ( $i = 1, \dots, n$ ) está dado por

$$\begin{aligned} f(\Omega_r^{-1} \mid \sigma^2, \varphi, \mu_i, y_i, x_i) &\propto \prod_{i=1}^n \left| \Omega_r^{-1} \right|^{-\frac{1}{2}} e^{\left\{ -\frac{1}{2} \mu_i' \Omega_r^{-1} \mu_i \right\}} \left| \Omega_r^{-1} \right|^{\frac{p-r-1}{2}} e^{-\tau(A^{-1}\Omega_r^{-1})} \\ &= \left| \Omega_r^{-1} \right|^{\frac{n+p-r-1}{2}} e^{\sum_{i=1}^n -\tau\left(-\frac{1}{2} \mu_i' \Omega_r^{-1} \mu_i\right)} e^{-\tau(A^{-1}\Omega_r^{-1})} \\ &= \left| \Omega_r^{-1} \right|^{\frac{n+p-r-1}{2}} e^{\sum_{i=1}^n -\tau\left(-\frac{1}{2} \mu_i \mu_i' \Omega_r^{-1}\right)'} e^{-\tau(A^{-1}\Omega_r^{-1})} \\ &= \left| \Omega_r^{-1} \right|^{\frac{n+p-r-1}{2}} e^{-\tau\left(\left[ \sum_{i=1}^n \frac{\mu_i \mu_i'}{2} \right] \Omega_r^{-1}\right)'} e^{-\tau(A^{-1}\Omega_r^{-1})} \\ &= \left| \Omega_r^{-1} \right|^{\frac{n+p-r-1}{2}} e^{-\tau\left(\left\{ \left[ \sum_{i=1}^n \frac{\mu_i \mu_i'}{2} \right] + A^{-1} \right\} \Omega_r^{-1}\right)} \end{aligned} \quad (4.7)$$

por lo que la distribución de  $\Omega_r^{-1} \mid \sigma^2, \varphi, \mu_i, y_i, x_i$  ( $i = 1, \dots, n$ )

$$\text{es } Wishart\left(n + p, A^{-1} + \sum_{i=1}^n \frac{\mu_i \mu_i'}{2}\right)$$

La distribución de  $\mu_i \mid \varphi, \sigma^2, \Omega, y_i x_i$  ( $i = 1, \dots, n$ ) está dado por

$$\begin{aligned}
f(\mu_i \mid \varphi, \sigma^2, \Omega, y_i x_i) &\propto e^{-\frac{1}{\sigma^2}(y_i - x_i \varphi - w_i \mu_i)'(y_i - x_i \varphi - w_i \mu_i)} e^{\left\{-\frac{1}{2} \mu_i' \Omega_r^{-1} \mu_i\right\}} \\
&\propto e^{-\frac{1}{2} \left\{ \mu_i' \left( \frac{w_i' w_i}{\sigma^2} + \Omega_r^{-1} \right) \mu_i - 2 \mu_i' w_i' \frac{(y_i - x_i \varphi)}{\sigma^2} \right\}} \\
&= e^{-\frac{1}{2} \left\{ \mu_i' Z \mu_i - 2 \mu_i' w_i' \frac{(y_i - x_i \varphi)}{\sigma^2} \pm \frac{(y_i - x_i \varphi)' w_i Z^{-1} w_i' (y_i - x_i \varphi)}{\sigma^4} \right\}} \\
&\propto e^{-\frac{1}{2} \left\{ \left( \mu_i - Z^{-1} w_i' \frac{(y_i - x_i \varphi)}{\sigma^2} \right)' Z \left( \mu_i - Z^{-1} w_i' \frac{(y_i - x_i \varphi)}{\sigma^2} \right) \right\}}
\end{aligned} \tag{4.8}$$

donde  $Z = \frac{w_i' w_i}{\sigma^2} + \Omega^{-1}$ , por lo que la distribución de  $\mu_i \mid \varphi, \sigma^2, \Omega, y_i x_i$  ( $i = 1, \dots, n$ ) es  $N\left(Z^{-1} w_i' \frac{(y_i - x_i \varphi)}{\sigma^2}, Z^{-1}\right)$

La distribución de  $\varphi \mid \sigma^2, \Omega, \mu_i x_i, y_i$  ( $i = 1, \dots, n$ ) está dada por

$$\begin{aligned}
f(\varphi \mid \sigma^2, \Omega, \mu_i x_i, y_i) &= \prod_{i=1}^n e^{-\frac{1}{2\sigma^2}(y_i - x_i \varphi - w_i \mu_i)'(y_i - x_i \varphi - w_i \mu_i)} e^{-\frac{1}{2\tau^2} \varphi' \varphi} \\
&\propto e^{-\frac{1}{2} \varphi' \left( \left[ \sum_{i=1}^n \frac{x_i x_i'}{\sigma^2} + \frac{I_k}{\tau^2} \right] \right) \varphi - 2 \varphi' \sum_{i=1}^n \frac{x_i' (y_i - w_i \mu_i)}{\sigma^2}} \\
&= e^{-\frac{1}{2} \sum_{i=1}^n \left\{ \varphi' S \varphi - 2 \varphi' \sum_{i=1}^n \frac{x_i' (y_i - w_i \mu_i)}{\sigma^2} + \frac{1}{\sigma} \left( \sum_{i=1}^n (y_i - w_i \mu_i)' x_i \right) S^{-1} \left( \sum_{i=1}^n x_i' (y_i - w_i \mu_i) \right) \right\}} \\
&\propto e^{-\frac{1}{2} \left\{ \left( \varphi - S^{-1} \sum_{i=1}^n \frac{x_i' (y_i - w_i \mu_i)}{\sigma^2} \right)' S \left( \varphi - S^{-1} \sum_{i=1}^n \frac{x_i' (y_i - w_i \mu_i)}{\sigma^2} \right) \right\}}
\end{aligned} \tag{4.9}$$

donde  $S = \left( \left[ \sum_{i=1}^n \frac{x_i x_i'}{\sigma^2} \right] + \frac{I_k}{\tau^2} \right)$ , por lo que la distribución de  $\varphi \mid \sigma^2, \Omega, \mu_i x_i, y_i$  ( $i = 1, \dots, n$ ) es  $N\left(S^{-1} \sum_{i=1}^n \frac{x_i' (y_i - w_i \mu_i)}{\sigma^2}, S^{-1}\right)$

Tomando los hiperparámetros como  $\tau^2 = 10^6$ ,  $h=g=0.01$ ,  $p = r$  y

$$A = \begin{bmatrix} \text{var}(\hat{a}_i) & 0 & 0 \\ 0 & \text{var}(\hat{b}_i) & 0 \\ 0 & 0 & \text{var}(\hat{c}_i) \end{bmatrix} \tag{4.10}$$

La matriz diagonal de las varianzas de los estimadores de los parámetros de un modelo de regresión (Daniels y Hogan, 2008) de la forma

$$\log(Y_i) = a_i + b_i t_i + c_i t_i^2 + d_i T_i + e * \log(P_i) \text{ para cada } i = 1, 2, \dots, n$$

Usando estas distribuciones condicionales el algoritmo Gibbs se implementa como sigue:

- Inicializar  $\Omega^{-1(0)} \sim Wishart(r, A)$ ,  $\mu_i^{(0)} \sim N(0, \Omega^{(0)})$
- Para un cierto número de iteraciones  $l$
- Generar  $(\Omega^{-1})^{(l)}$  de una distribución  $Wishart\left(n + p, A^{-1} + \sum_{i=1}^n \frac{\mu_i^{(l-1)} \mu_i'^{(l-1)}}{2}\right)$  para  $i = 1, \dots, n$
- Generar  $(\frac{1}{\sigma^2})^{(l)}$  de una distribución

$$G\left(\frac{N\tau}{2} + h, \left[\frac{\sum (y_i - x_i \varphi^{(l-1)} + w_i \mu_i^{(l-1)})' (y_i - x_i \varphi^{(l-1)} + w_i \mu_i^{(l-1)})}{2} + \frac{1}{g}\right]^{-1}\right)$$

- Generar  $\mu_i^{(l)}$  de una distribución  $N\left(Z^{-1(l)} w_i' \frac{(y_i - x_i \varphi^{(l-1)})}{\sigma^{2(l)}}, Z^{-1(l)}\right)$
- Generar  $\varphi^{(l)}$  de una distribución  $N\left(S^{-1(l)} \sum_{i=1}^n \frac{x_i' (y_i - w_i \mu_i^{(l)})}{\sigma^{2(l)}}, S^{-1(l)}\right)$

Como vimos en la Sección 2.2.1, esto define una cadena de Markov que converge a una distribución estacionaria a partir de cierto número de iteraciones por lo que desechamos las primeras 1,700,000 simulaciones y nos quedamos con las últimas 30,000.

Para hacer inferencia para los  $y_i$  debemos muestrear los  $y_i$ s usando la distribución predictiva. La distribución predictiva para un número nuevo  $y_i$  es la mezcla

$$\int f(y_i | \sigma^2, \Omega, \varphi) f(\varphi, \Omega, \sigma^2 | y_1, \dots, y_n) d_{\varphi, \Omega, \sigma^2}$$

No es necesario calcular la integral en forma explícita para simular una nueva  $y_i$ , sólo debemos obtener  $\varphi, \Omega, \sigma^2$  de  $f(\varphi, \Omega, \sigma^2 | y_1, \dots, y_n)$ , y luego  $y_i$  de  $f(y_i | \sigma^2, \Omega, \varphi)$  para  $i = 1, \dots, n$ . Por la ecuación (4.5)  $(y_i | \sigma^2, \Omega, \varphi)$  tiene distribución

$N(x_i\varphi, \sigma^2I + w_i\Omega w_i)$ , así que con la cadena de Markov generada por el Gibbs Sampler podemos muestrear triplas  $(\sigma_{(1)}^2, \Omega_{(1)}, \varphi_{(1)}), \dots, (\sigma_{(l)}^2, \Omega_{(l)}, \varphi_{(l)})$  y generar  $y_{i(1)}, \dots, y_{i(l)}$  para  $i = 1, \dots, n$

Los  $y_i$ , generados de esta forma no son independientes. Dado que queremos una muestra independiente de la distribución predictiva para hacer inferencias, muestremos sistemáticamente de los  $y_{i(1)}, \dots, y_{i(l)}$ . Para obtener el número de salto se consideró un plot de autocorrelación en el que se observa que aproximadamente existe autocorrelación hasta un orden 10, entonces se tomó un número aleatorio entre 1 y 10, para elegir la posición de arranque, y las muestras se tomaran a partir de ésta con un salto de 10 hasta conseguir el tamaño deseado.

Presentamos los resultados del modelo 1 en la Sección 4.4 para los datos observados de temperatura y precipitación para los años 1961-2007. Pero vale la pena preguntar como podríamos incorporar lo aprendido sobre temperatura y precipitación en el Capítulo 3 en este modelo.

### 4.3. Rendimiento Incorporando la Distribución Predictiva de Temperatura y Precipitación

En esta sección utilizamos la distribución predictiva del clima encontrado en el Capítulo 3, para actualizar los parámetros del modelo de rendimiento hasta el año 2030. Para ello consideremos el siguiente proceso.

Se muestrea un valor de  $\varphi_1, \mu_1, \sigma_1$  y  $\Omega_1$  de  $f(\varphi, \mu_i, \sigma, \Omega|X_i)$  y un  $X_{*t}$  de  $f(X_{*t})$ , donde,  $f(X_{*t}) = f(X_{*t}^T, X_{*t}^P)$  es la función de distribución predictiva de temperatura y precipitación que se encontró en el Capítulo 3. Con este valor se genera un vector de  $Y_{i*t}$  para  $t = 47, 48, \dots, 70$ , considerando que  $Y_{i*t} \sim N(X_{*i}\varphi_1, \sigma_1^2I + w_{*i}\Omega_1 w_{*i}')$ , para  $i = 1, \dots, 44$ . De este modo tenemos información para el rendimiento y el clima

desde 1961 hasta el 2030. Con esta información tenemos un nuevo conjunto de datos en el que  $t = 1, 2, \dots, 70$ , y con éste se realiza una iteración del Gibbs Sampler definido en la Sección 4.2.1. Este proceso se repite, de modo que cada iteración del Gibbs Sampler cuenta con un escenario diferente de clima seleccionando al azar.

Por el hecho de que ahora el clima es aleatorio, la convergencia probablemente será más lenta. Por lo que fueron necesarias 7,000,000 de los cuales conservamos las últimas 50,000.

#### 4.4. Resultados del Modelo 1

Aplicando el Gibbs Sampler descrito en la Sección 4.2 obtuvimos muestras de la distribución posterior de los parámetros del modelo 1 usando los datos de temperatura y precipitación observados (1961-2007) y también incorporando la distribución predictiva de temperatura y precipitación encontrada para cada país en el Capítulo 3 (2007-2030). Nos concentramos en la interpretación de los parámetros  $d$  y  $e$  que son los coeficientes de temperatura y precipitación respectivamente.

De la Figura 4-6 se observa que el rango de valores para la distribución del coeficiente de temperatura  $d$  asumen valores negativos, tanto para lo observado (línea sólida), como incorporando la distribución predictiva (línea entrecortada)  $X_{*t}$  de temperatura y precipitación. El intervalo de credibilidad del 95% de probabilidad en ambos casos (datos observados y modelo predictivo) es negativo, de modo que se puede concluir que el incremento de temperatura tiene un efecto significativo negativo en el rendimiento del frijol, y que el efecto negativo del incremento de temperatura para el futuro puede ser mayor a lo que se observa, puesto que la distribución del coeficiente de temperatura para el futuro, alcanza a considerar valores más a la izquierda que la de los valores observados. En cuanto al parámetro del logaritmo de la precipitación  $e$ , se observa que el intervalo de credibilidad contiene

al cero, por lo que concluimos que la precipitación no tiene un efecto significativo en el rendimiento. Sin embargo, la mayoría de los valores de este parámetro son menores a 0, lo que significa mayor probabilidad de que el aumento de precipitación genere un efecto negativo en el rendimiento.

Como la variable respuesta en nuestro modelo es el logaritmo del rendimiento es necesario transformarla para medir el efecto del coeficiente de temperatura. Consideremos que el modelo es de la forma  $\log(Y) = A + d * T$ , donde A es una expresión que contiene los términos que no involucran temperatura. Si denotamos a  $Y_1$  como un rendimiento inicial y  $Y_2$  como el rendimiento luego de un incremento  $\Delta$  en temperatura, entonces tenemos que  $Y_1 = \exp(A + dT)$  y  $Y_2 = \exp(A + d(T + \Delta))$  de modo que  $Y_2 = Y_1 * \exp(d * \Delta)$ , esto significa que si la temperatura se incrementa en  $\Delta$  unidades, entonces el rendimiento inicial se multiplica por  $\exp(d * \Delta)$ . El razonamiento para la precipitación es similar. Note que esto expone una limitación del modelo 1, pues implica que la reducción es la misma en cualquier país. Esto no es realista porque en países que de por sí son calientes el aumento de temperatura debe ser más perjudicial que en países menos calientes. Inclusive países fríos pueden experimentar un aumento en el rendimiento. Esto lo tomaremos en cuenta en el modelo 2 que presentamos en la Sección 4.5. De todos modos continuamos con la interpretación del modelo 1.

Nuestro intervalo de credibilidad del 95 % para  $d$  contiene sólo valores negativos lo que significa que, el rendimiento se reduce por  $\exp(d)$ , si tenemos un aumento de  $1^\circ C$  en temperatura. Para nuestra distribución  $\bar{d}=0.028$  por lo que el rendimiento se reduce en promedio en  $1/\exp(0.028)=1/1.028$ , es decir un 3%. Podemos decir que con un 95 % de probabilidad, la reducción del rendimiento está entre 1 % y 4 %

cuando la temperatura aumenta por 1°C.

Cuando se incorpora el clima futuro (la Distribución del clima determinado en el Capítulo 3) el efecto de temperatura será mayor. En promedio, el rendimiento futuro del frijol se reduciría por  $1/\exp(0.038)$  que equivale a un 4%. Un intervalo de credibilidad del 95% muestra que se esperaría una reducción del rendimiento entre 1% y 7% para cada país, si en promedio el aumento es de 1°C.

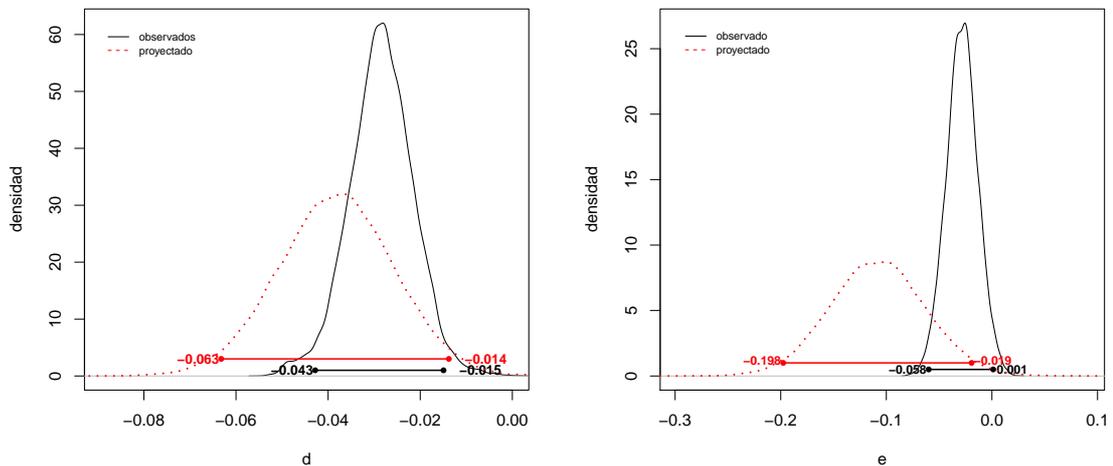


Figura 4-6: Distribución de los coeficientes de temperatura y precipitación en el modelo del rendimiento. *Observado* significa que la estimación se obtuvo a partir del modelo de los datos observados; *Proyectado* que el estimado se obtuvo incorporando las distribuciones predictivas de temperatura y precipitación.

Como vimos en la Figura 4-1 el rendimiento del frijol varía por país. El modelo 1 asume el mismo efecto en términos porcentuales para todos los países, pero varía en términos de la magnitud del rendimiento del país, así que el efecto neto en cada país es diferente. Para saber cuánto se perdió en cada país en el período de 1961-2007 con respecto al rendimiento medio del país utilizaremos el cambio promedio que hubo en temperatura durante este período y calculamos intervalos de credibilidad para  $\bar{y}_i * \exp(\bar{d} * \bar{\Delta}_i)$ , donde  $\bar{y}_i$  es el rendimiento promedio para el

país en el período de 1961-2007,  $\bar{d}$  es el promedio de la distribución posterior de  $d$  para el modelo que usa las temperaturas observadas y  $\overline{\Delta}_i$  es el cambio promedio en temperaturas que se observó en el período de 1961-2007. La Figura 4-7 muestra estos intervalos de credibilidad para cada país (intervalos de color negro). Se puede observar que la mayoría de los países pierden debido a la temperatura, a excepción de Canadá y Korea del Sur que muestran una ganancia pequeña. Esto se da porque experimentaron disminución de temperatura. Los países que más perdieron en este período fueron Francia y Holanda, que alcanzaron una pérdida de casi 1500 hg/ha (150 kg/ha), y la pérdida promedio por país fue aproximadamente 193 hg/ha.

En la Figura 4-7 también se presentan los intervalos (líneas entrecortadas) de credibilidad correspondientes cuando se incorpora la distribución predictiva del clima.  $\bar{y}_i$  se toma como el promedio del rendimiento que predice el modelo para los años 2007-2030,  $\bar{d}$  se toma como el promedio de la distribución posterior de  $d$  y  $\overline{\Delta}_i$  como el promedio del cambio en temperatura que estima el modelo. Estos intervalos muestran que la mayoría de los países experimentarían pérdida en el rendimiento. En algunos casos la magnitud en la pérdida será menor que lo observado en el período 1961-2007, como son el caso de Francia, El Salvador, Egipto y Holanda. En la mayoría de los casos la pérdida es mayor. Sólo en los casos de Nepal y Perú el incremento en el rendimiento es positivo, esto se debe a que de la distribución predictiva de temperatura para el futuro mostró una disminución de temperatura.

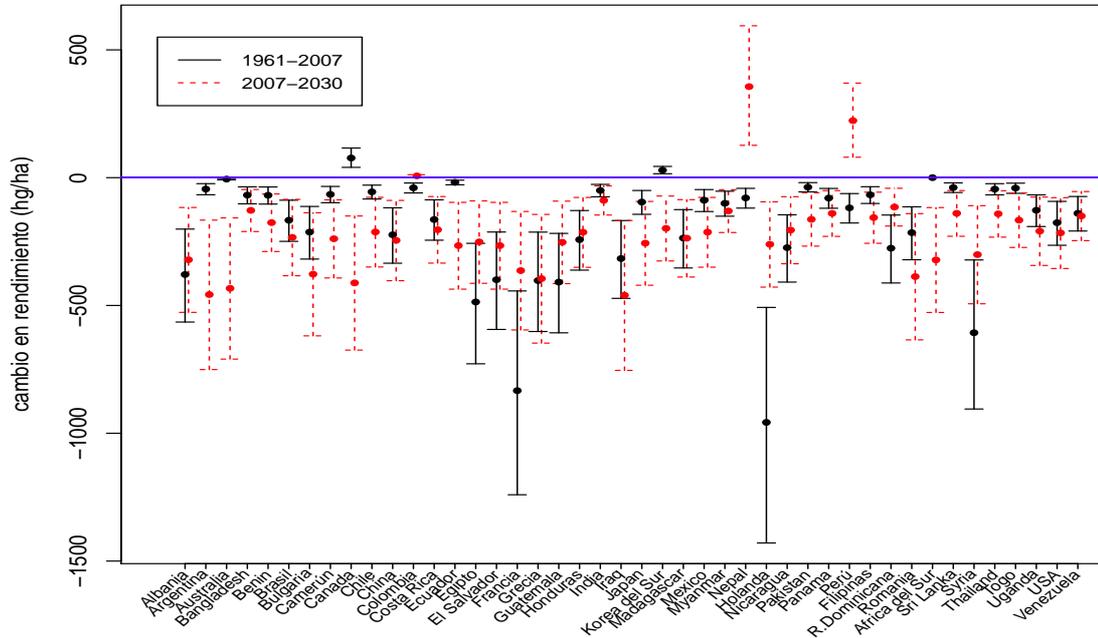


Figura 4-7: Intervalo de credibilidad del 95 % para el cambio en el rendimiento medio en el período 1961-2007 (líneas sólidas) y el período 2007-2030 (líneas entrecortadas), debido al cambio en temperatura.

La Figura 4-8 muestra el efecto de la precipitación en el rendimiento por país. Se observa que el efecto de la precipitación por país es muy variable y no es significativo para muchos países. Esto se refleja en el hecho de que muchos intervalos se encuentran próximos a la recta  $x = 0$ . Se puede ver que en algunos países la precipitación puede aumentar el rendimiento, como por ejemplo El Salvador y Honduras que alcanzan valores próximos a 200 hg/ha de incremento. Sin embargo, esta cantidad es pequeña comparada con la pérdida que se experimenta con el efecto de la temperatura.

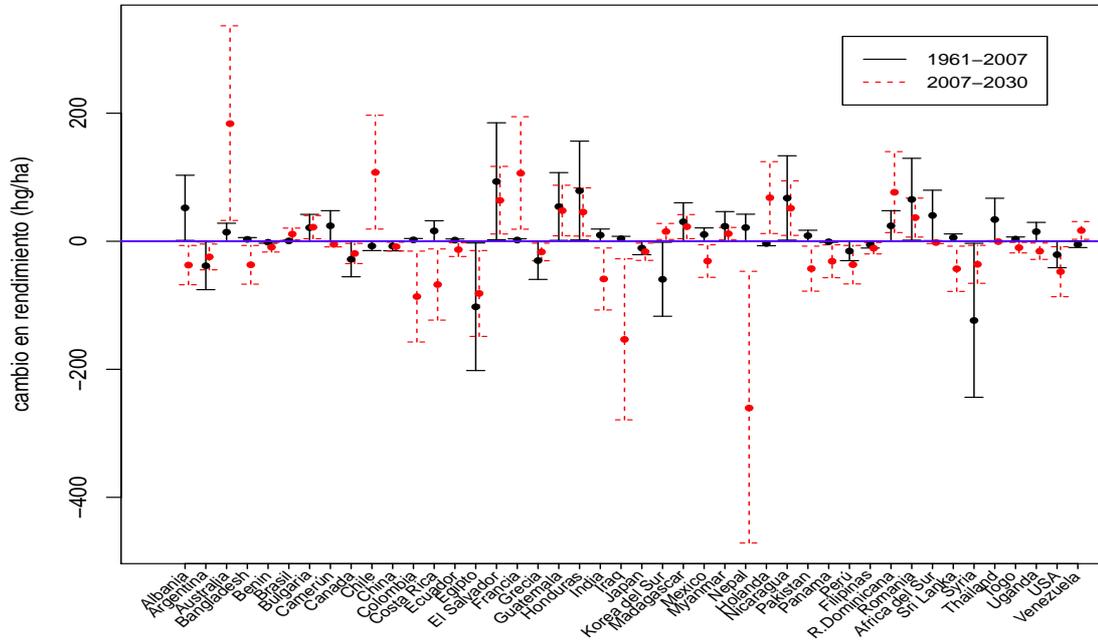


Figura 4-8: Intervalo de credibilidad del 95 % para el cambio en el rendimiento medio en el período 1961-2007 y el período 2007-2030 debido al cambio en precipitación.

Dado que nuestra intención es medir el efecto del clima (temperatura y precipitación conjunta), la Figura 4-9 muestra el efecto que provoca la variación del clima. En este caso los intervalos de predicción se calculan para  $\bar{y}_i * \exp(\Delta_T * d + \Delta_P * e)$ , donde  $\Delta_T$  es el cambio en temperatura y  $\Delta_P$  es el cambio en precipitación. Esta gráfica es muy similar a la que obtuvimos cuando consideramos sólo la temperatura lo que confirma el hecho de que la precipitación no es significativa.

La pérdida promedio por país es 202 hg/ha, esto significa que aunque el rendimiento aumentó al final del período, se pudo haber obtenido, por país 202 hg más de frijol por hectárea, si no hubiera efecto del clima.

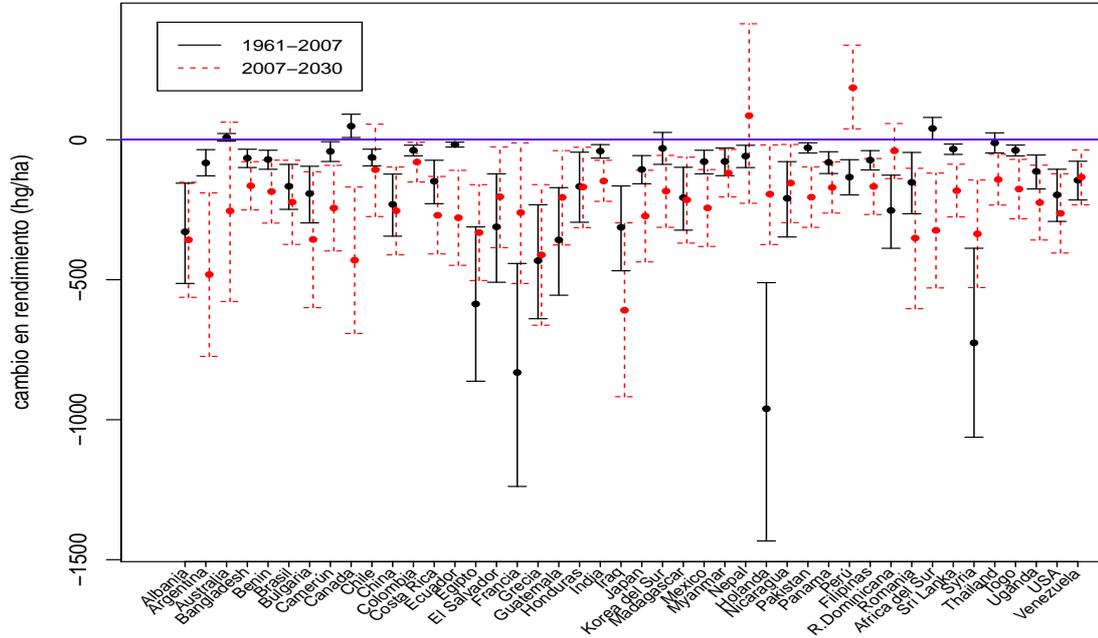


Figura 4-9: Intervalo de credibilidad del 95 % para el cambio en el rendimiento medio en el período 1961-2007 y el período 2007-2030 debido al cambio en el clima.

#### 4.5. Resultados del Modelo 2

Como mencionamos anteriormente el modelo 1, tiene la limitación de considerar el mismo efecto de temperatura y precipitación en el rendimiento para todos los países, pero sabemos que el efecto de temperatura depende del nivel de la temperatura en el que se encuentra un país (similarmente para precipitación). Por esta razón decidimos analizar el modelo incorporando términos cuadráticos en temperatura y precipitación. Este es nuestro modelo 2 y nos permitirá saber que sucede en países con climas extremos, por ejemplo, climas muy fríos o muy calientes.

El modelo 2 es de la forma

$$\log(y_{ij}) = (a + \mu_{i1}) + (b + \mu_{i2})t_{ij} + (c + \mu_{i3})t_{ij}^2 + d * T_{ij} + e * T_{ij}^2 + f * \log(P_{ij}) + g * \log(P_{ij})^2 + \varepsilon_{ij} \quad (4.11)$$

El análisis es el mismo al del modelo 1, lo único que cambia es la matriz de diseño  $X$ , que ahora tendrá dimensión  $47 \times 7$  y el vector de parámetros  $\varphi$  que ahora tendrá dimensión  $7 \times 1$ . Como en el modelo 1 consideremos el efecto de la variable temperatura. Ahora nuestro modelo es de la forma  $\log(Y) = A + dT + eT^2$ , donde  $A$  es una expresión que incluye el resto de los términos que no incluyen la variable temperatura. Si denotamos a  $Y_1$  como un rendimiento inicial y  $Y_2$  como el rendimiento luego de un incremento  $\Delta$  en temperatura, entonces tenemos que  $Y_1 = \exp(A + dT + eT^2)$  y  $Y_2 = \exp(A + d(T + \Delta) + e(T + \Delta)^2)$  de modo que  $Y_2 = Y_1 * \exp(\Delta * (d + 2T * e) + e * \Delta^2)$ , esto significa que si la temperatura se incrementa en  $\Delta$  unidades, entonces el rendimiento inicial se multiplica por  $\exp(\Delta * (d + 2T * e) + e * \Delta^2)$ . Ahora el efecto no sólo depende del cambio ( $\Delta$ ) si no también de la temperatura. El razonamiento para la precipitación es similar.

La Figura 4-10 muestra el efecto de temperatura en el rendimiento de frijol en el período 1961-2007 y el período 2007-2030 estimado con el modelo 2. Como para el modelo 1, estos intervalos fueron calculados con los datos observados e incorporando la distribución predictiva para el clima. Se observa que el efecto del incremento de temperatura fue negativo en el rendimiento del frijol y se espera que este efecto sea mayor para la mayoría de los países en el período 2007-2030, con excepción de Netherlands, Nepal y Syria. Nepal muestra incremento porque la proyección de temperatura fue que disminuiría. Para el período 1961-2007, se muestra que hubo una pérdida en la mayoría de países. El país que más pierde es Syria que muestra una pérdida cerca de 500hg/ha. En cuanto a la pérdida por país se obtiene que en promedio cada país pierde 243 hg/ha para el período 1961-2007.

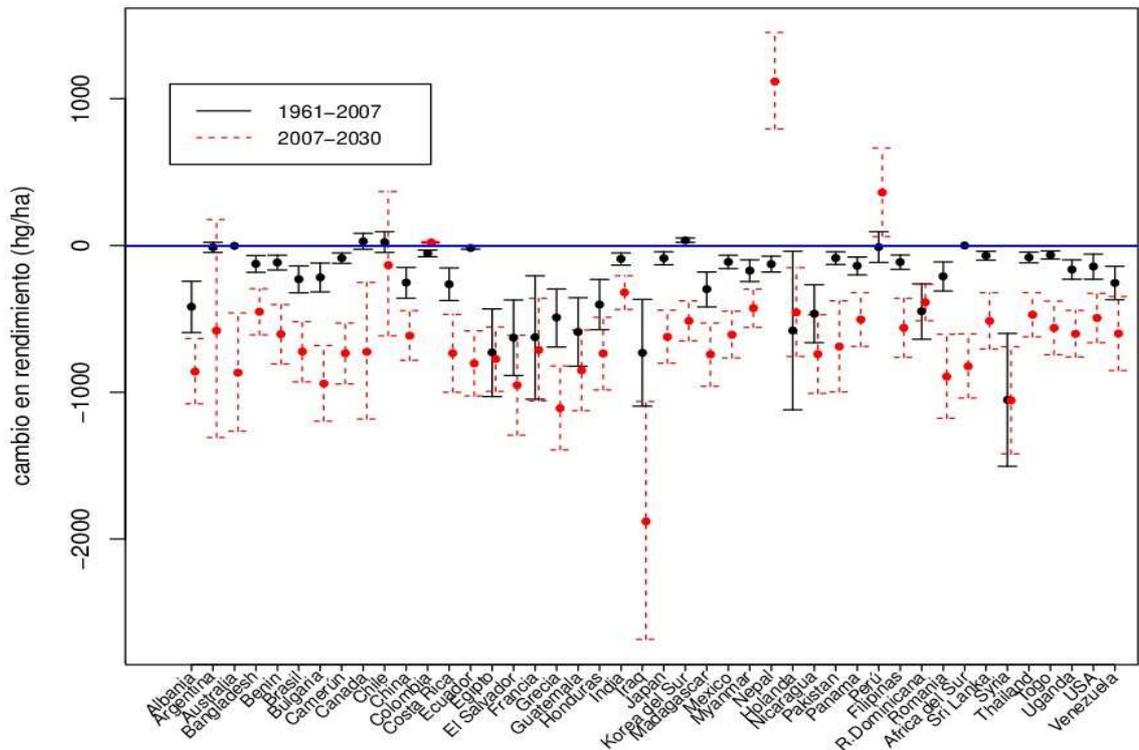


Figura 4–10: Intervalo de credibilidad del 95 % para el cambio en el rendimiento medio en el período 1961-2007 y el período 2007-2030 debido al cambio en temperatura, incorporando términos cuadráticos.

La Figura 4–11 muestra el efecto del cambio de la precipitación para los períodos 1961-2007 y 2007-2030. En el período 1961-2007 se observa que la mayoría de los intervalos para el cambio en el rendimiento contiene al cero por lo que el efecto de la precipitación no es significativo excepto para Argentina y Francia que muestran pérdidas significativas. Sólo en el caso de Albania, el intervalo de credibilidad para el período 1961-2007 muestra que el rendimiento se incrementó. En cuanto al período 2007-2030, casi todos los intervalos contienen al cero a excepción de Argentina, lo que significa que la precipitación no es significativo para el rendimiento, y el ancho del intervalo para muchos países es pequeño.

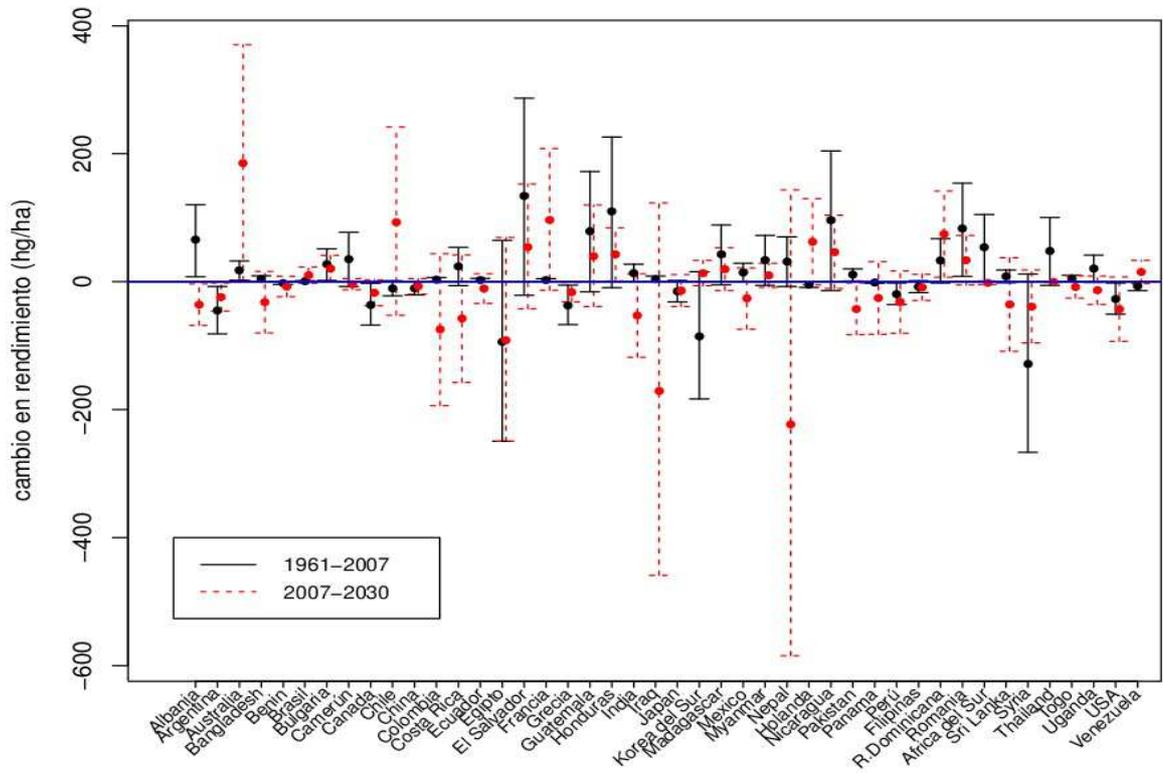


Figura 4-11: Intervalo de credibilidad del 95 % para el cambio en el rendimiento medio en el período 1961-2007 y el período 2007-2030 debido al cambio en precipitación, incorporando términos cuadráticos.

En la figura 4-12 se muestra que el efecto del cambio del clima (temperatura y precipitación) fue negativo en el rendimiento de frijol en el período 1961-2007 y que la mayor pérdida fue de casi 1500 hg/ha. Además la pérdida promedio por país en el rendimiento se estima que fue de 233 hg/ha (23 Kg/ha), lo que significa que la temperatura contribuye de manera significativa en las pérdidas del rendimiento, lo que no sucede con precipitación. Además para el período 2007-2030 se observa que los efectos del clima serán más negativos. Esto muestra que si las temperaturas siguen incrementándose, el efecto negativo en el rendimiento podría ser mayor.

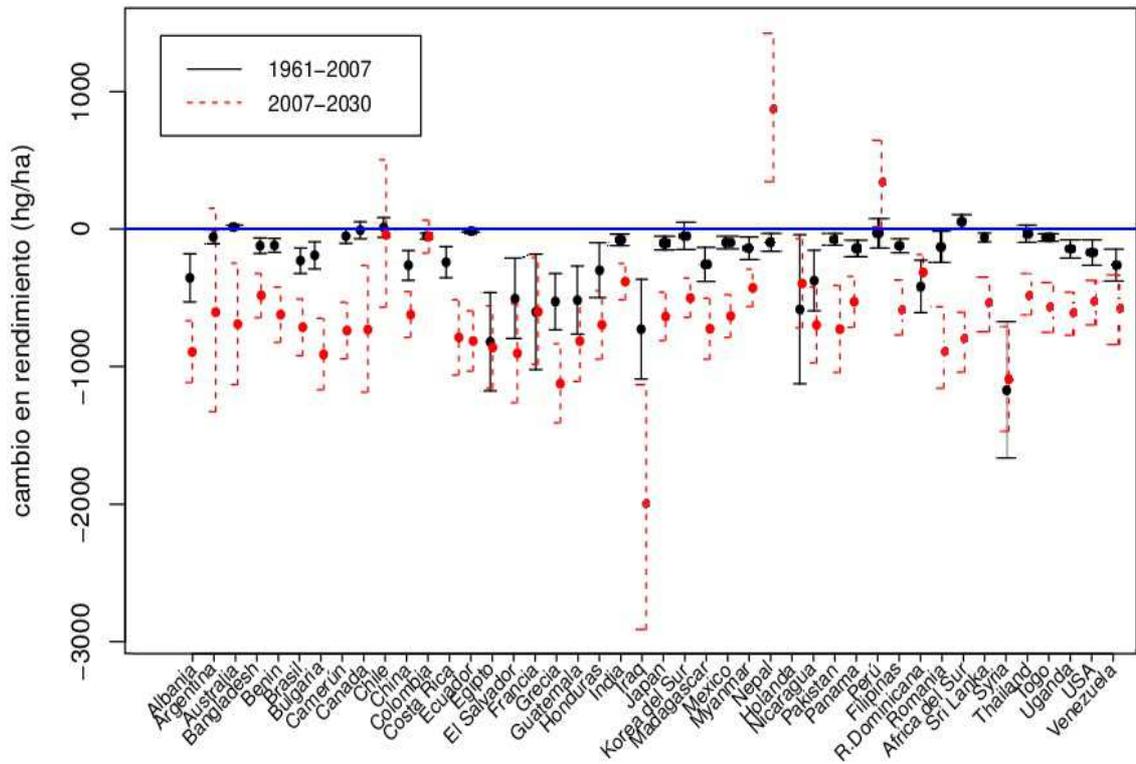


Figura 4-12: Intervalo de credibilidad del 95 % para el cambio en el rendimiento medio en el período 1961-2007 y el período 2007-2030 debido al incremento en el clima, incorporando términos cuadráticos.

Anteriormente mencionamos que este modelo, nos permite conocer que pasa en climas extremos. La Figura 4-13, nos permite comparar el efecto en el rendimiento causado por un incremento fijo  $\Delta$  para todos los países. Consideremos  $\Delta = 1, 2, 3, 4$  °C. Se observa que a mayor incremento de temperatura, países que son más calientes experimentan mayor pérdida en el rendimiento del frijol, esto significa que son más sensibles al incremento de temperatura, como es el caso de Iraq y Pakistan. Por otro lado, Chile que es un país comparativamente frío gana con el aumento en temperatura. El rendimiento en Chile tiene una tendencia a aumentar con el incremento en temperatura. Sin embargo, cuando el  $\Delta$  es aproximadamente  $2.5^{\circ}\text{C}$ , todos los países bajan su rendimiento, por lo que se podría decir que a partir de cierto

nivel de temperatura, el incremento en temperatura tiene un efecto negativo en el rendimiento.

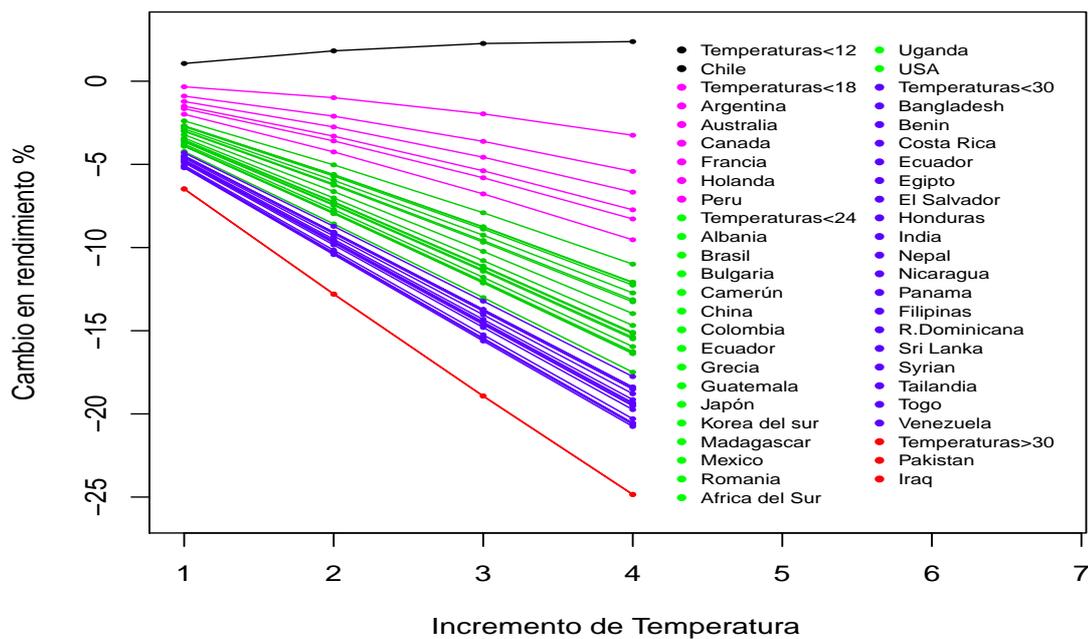


Figura 4-13: Efecto del cambio de temperatura en climas extremos, cuando  $\Delta = 1, 2, 3, 4^{\circ}C$ .

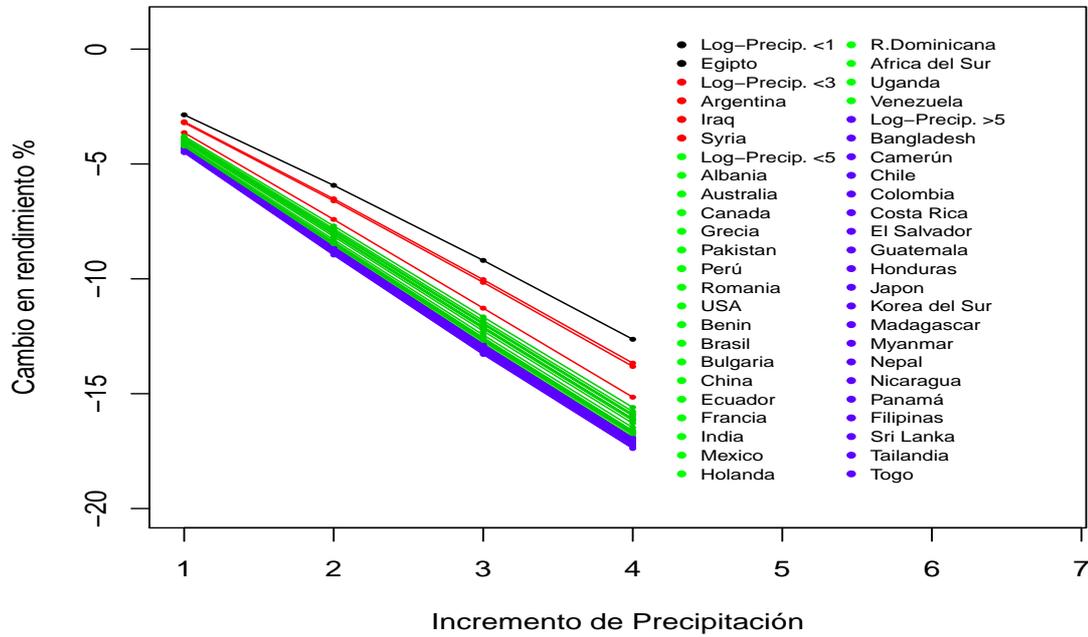


Figura 4-14: Efecto del cambio de precipitación en climas extremos cuando  $\Delta = 1, 2, 3, 4$  mm.

En cuanto a precipitación se observa que la razón con la que el rendimiento baja es parecido para todos los países. Debemos considerar que un incremento de  $\Delta = 1, 2, 3, 4$  mm. en el logaritmo de la precipitación equivale a multiplicar la precipitación por  $exp(\Delta)$ , que es un incremento mayor con respecto a lo que la temperatura varía. Esto corrobora también el hecho de que la precipitación no es significativa en el modelo, y que para observar efectos en el rendimiento el cambio en precipitación debe ser bien grande.

## Capítulo 5

# CONCLUSIONES Y TRABAJOS FUTUROS

En este trabajo estudiamos los cambios en temperatura y precipitación para la temporada del cultivo del frijol y el impacto de estos cambios en el rendimiento del mismo. Concluimos que:

- La temperatura aumentó con mayor intensidad en los años 1984-2007, con respecto a los años 1961-1984. Para los años 2007-2030 se espera un incremento de temperatura igual o mayor que el período 1984-2007. La distribución posterior de los parámetros de temperatura, y la distribución predictiva conjunta de temperatura y precipitación, pronostican que la temperatura seguirá aumentando en 42 de los 44 países. Sólo en el caso de Nepal y Perú, se pronostica una disminución en la temperatura.
- Para la precipitación, no se observa una tendencia general para todos los países, y los cambios que se registran no son significativos. Sin embargo resalta el caso de Nepal, que esperaríamos un aumento en precipitación.
- Para los cinco mayores productores de frijol, se encontró que, cuatro de ellos: Myanmar, Brasil, China e India esperaríamos un aumento en la temperatura, pero la tendencia a aumentar es menos intensa que el período 1984-2007. Sólo para Estados Unidos, se espera, que la tendencia a aumentar sea mayor con respecto al período de 1984 - 2007.
- El efecto del incremento de las temperaturas en el rendimiento del frijol depende del nivel en el que se encuentra la temperatura de un país. El modelo con efectos

cuadráticos confirma que países con temperaturas altas son más sensitivos a perder en el rendimiento, mientras que, países con temperaturas muy bajas pueden favorecerse del aumento de la temperatura, hasta cierto punto, a partir del cuál comienzan a perder.

- Se ha observado un impacto negativo del aumento de temperatura en el rendimiento del frijol y con el continuado aumento que se espera al año 2030 las pérdidas serán mayores. El esfuerzo de los fitomejoradores para producir variedades de frijol resistentes a altas temperaturas es muy importante.
- El análisis Bayesiano de un modelo lineal mixto para el rendimiento permitió incorporar las proyecciones que obtuvimos para el clima y medir el efecto en el rendimiento del período 2007-2030.

Como trabajo futuro se propone:

- Incorporar variables adicionales en el modelo para rendimiento, como el número de fitomejoradores y la inversión en tecnología para el mejoramiento agrícola. Así como también considerar la concentración de  $CO_2$  en la atmósfera.
- Realizar el análisis considerando temperatura y precipitación máximas y mínimas, puesto que el frijol es sensible a climas extremos.
- Incorporar la información de un calendario de las fechas y áreas de cultivo variables por lo menos cada 5 años para determinar el área y la temporada de crecimiento, de modo que se tome en cuenta el movimiento de las áreas de producción y el cambio de la temporada de cultivo. Los agricultores como medidas de adaptabilidad al cambio climático tienden a variar su áreas y temporadas de cultivo. Entendemos que esto no está disponible actualmente. El calendario de Sacks (2010) sólo provee las fechas de cultivo para el año 2010 lo que es una limitación para nuestro estudio.
- Realizar el estudio del efecto del cambio climático en el rendimiento por cuadrículas de .5 grados.

- Estudiar el efecto del cambio climático en el cultivo de maíz, y compararlo con el estudio del frijol para los años 2050 y 2100. Muchos agricultores cultivan estos dos productos en forma conjunta, porque el frijol mejora la absorción de nitrógeno por parte de la tierra, lo cual es necesario para el desarrollo del maíz.
- Estudiar el impacto de la precipitación en el rendimiento del frijol, de modo que los datos sean considerados en intervalos de tiempo cortos (diarios) e intervalos de tiempo prolongados (10 años), para capturar el posible efecto de precipitación. El frijol es sensible a cambios en precipitación y sin embargo la mayoría de los estudios reportan que la precipitación es no significativa.
- Mejorar el modelo del clima permitiendo que el código que marca el cambio en las tendencias centrales del clima varíe en forma aleatoria.
- Estudiar el impacto del aumento de temperatura en otras áreas como en la Ingeniería Eléctrica y el uso de paneles solares y en Ingeniería Agrícola y el estudio de suelos y flujo de agua.

# APÉNDICES

# Apéndice A

## FIGURAS

### A.1. Distribución predictiva de temperatura y precipitación para el período actual y futuro

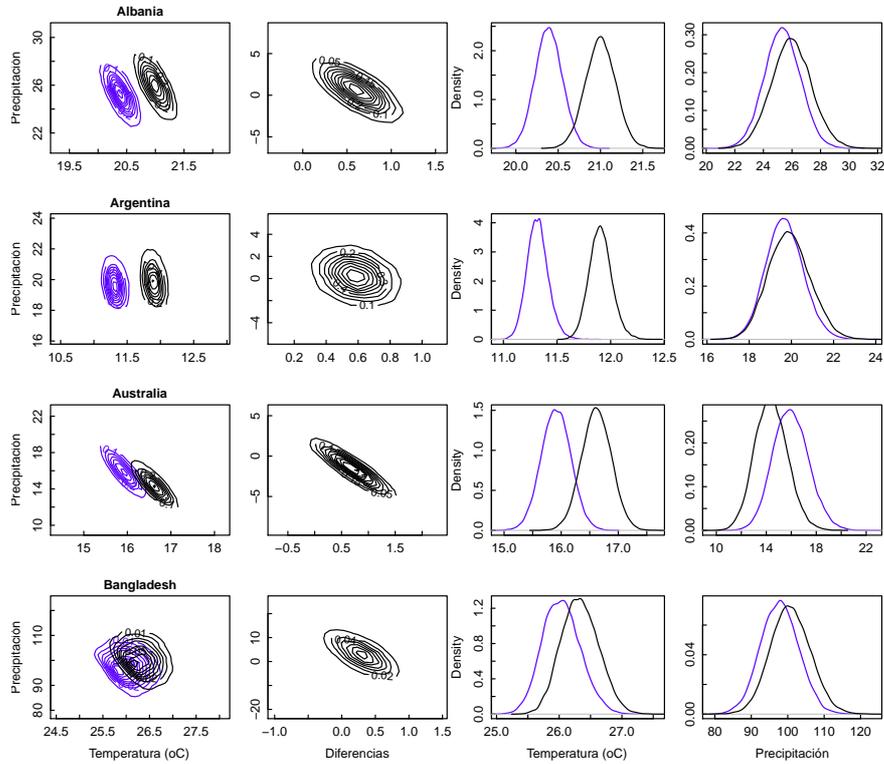


Figura A-1: En la primera columna contours de la distribución predictiva conjunta de temperatura y precipitación comparando el período actual (1997-2007) y futuro (2020-2030). En la segunda columna la distribución de la diferencia (futuro-actual). En la tercera y cuarta columna las distribuciones marginales de temperatura y precipitación para (Albania, Argentina, Australia, Bangladesh).

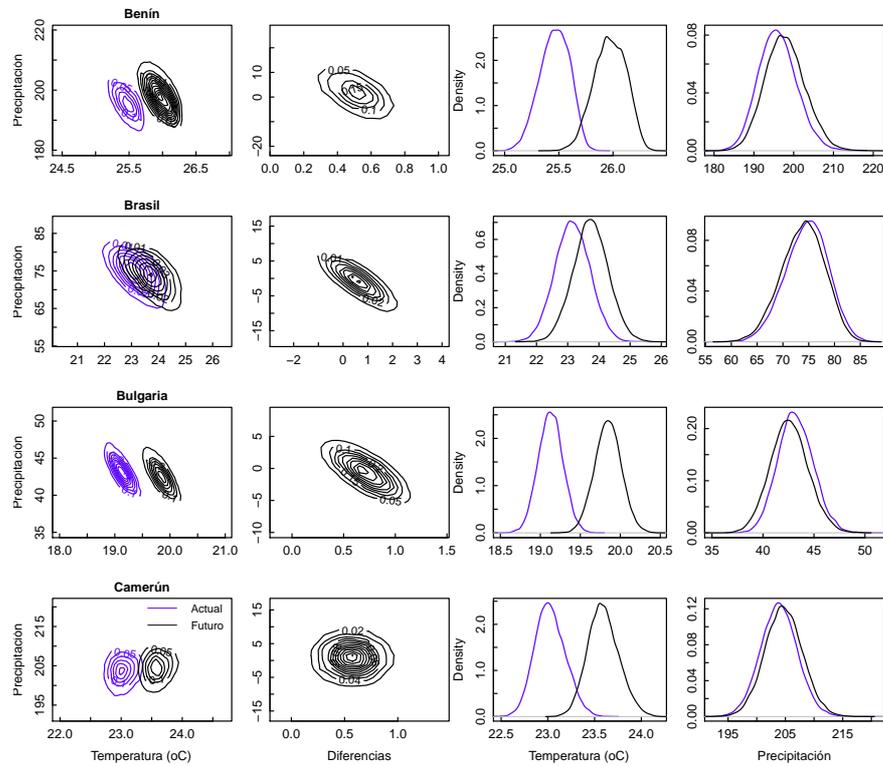


Figura A-2: En la primera columna contours de la distribución predictiva conjunta de temperatura y precipitación comparando el período actual (1997-2007) y futuro (2020-2030). En la segunda columna la distribución de la diferencia (futuro-actual). En la tercera y cuarta columna las distribuciones marginales de temperatura y precipitación para (Benin, Brasil, Bulgaria, Camerún).

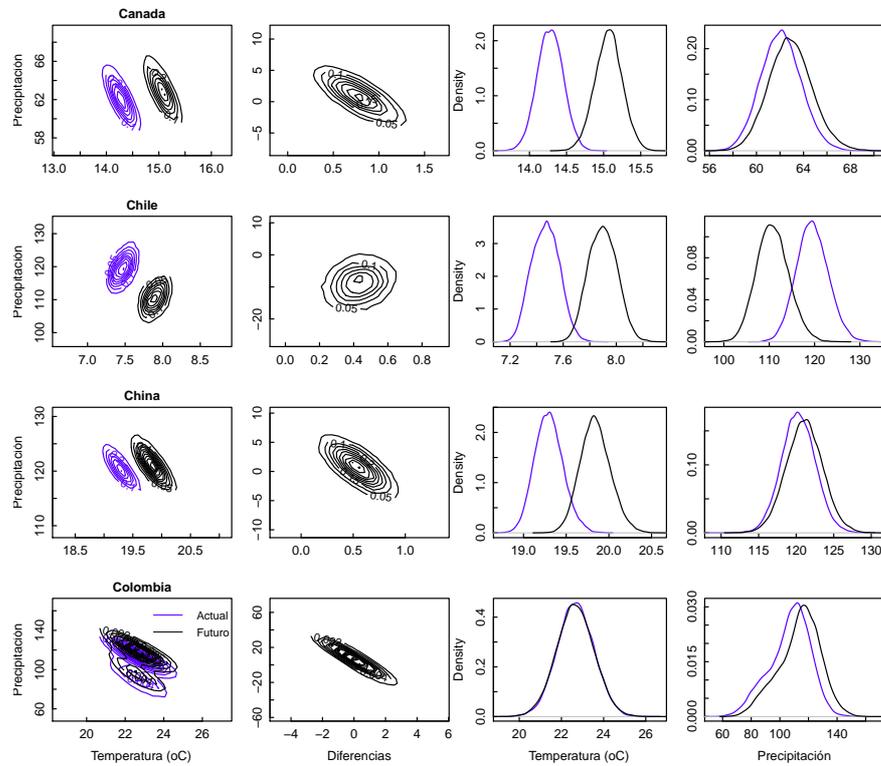


Figura A-3: En la primera columna contours de la distribución predictiva conjunta de temperatura y precipitación comparando el período actual (1997-2007) y futuro (2020-2030). En la segunda columna la distribución de la diferencia (futuro-actual). En la tercera y cuarta columna las distribuciones marginales de temperatura y precipitación para (Canada, Chile, China, Colombia ).

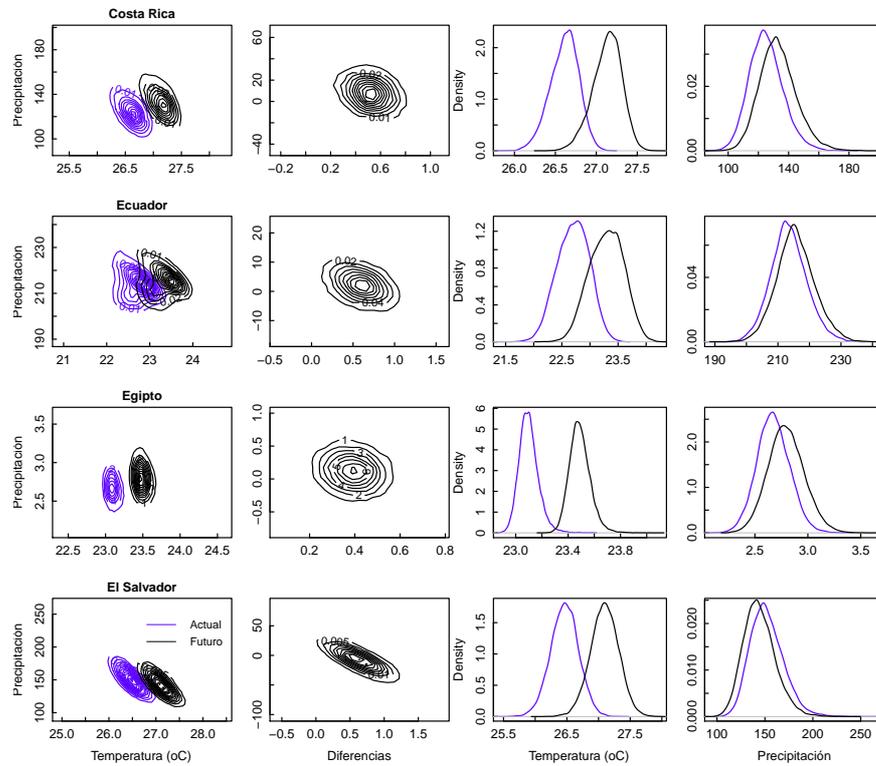


Figura A-4: En la primera columna contours de la distribución predictiva conjunta de temperatura y precipitación comparando el período actual (1997-2007) y futuro (2020-2030). En la segunda columna la distribución de la diferencia (futuro-actual). En la tercera y cuarta columna las distribuciones marginales de temperatura y precipitación para (Costa Rica, Ecuador, Egipto, El Salvador).

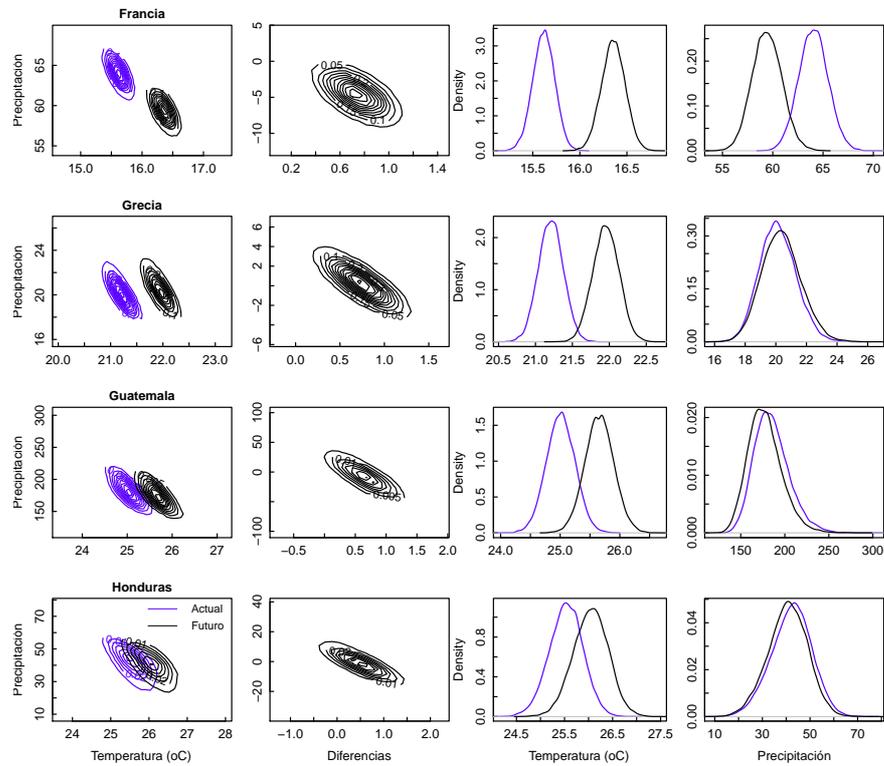


Figura A-5: En la primera columna contours de la distribución predictiva conjunta de temperatura y precipitación comparando el período actual (1997-2007) y futuro (2020-2030). En la segunda columna la distribución de la diferencia (futuro-actual). En la tercera y cuarta columna las distribuciones marginales de temperatura y precipitación para (Francia, Grecia, Guatemala, Honduras).

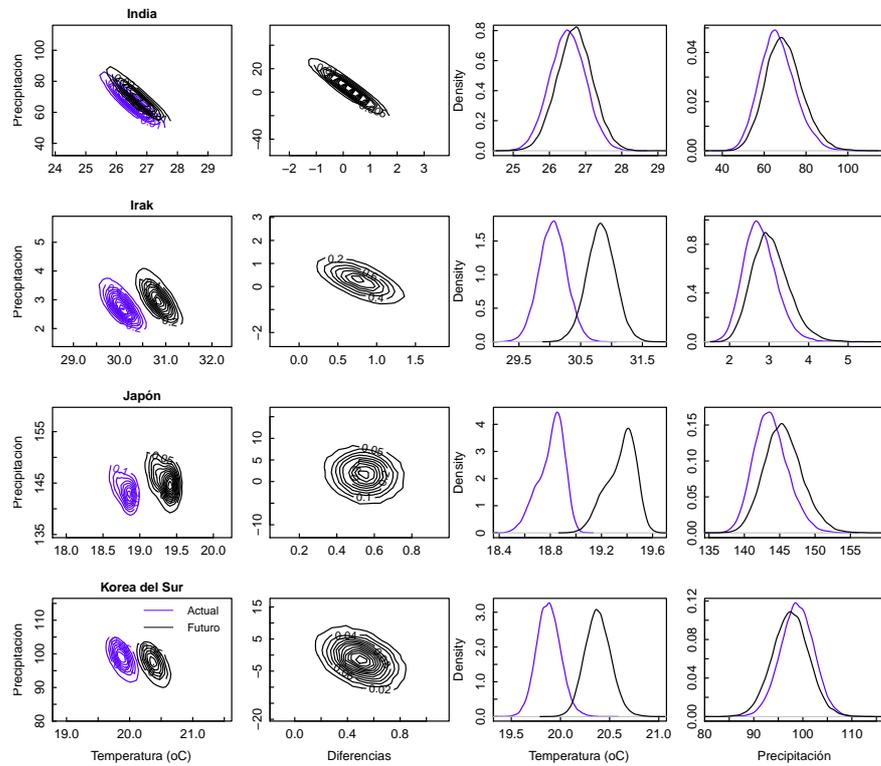


Figura A-6: En la primera columna contours de la distribución predictiva conjunta de temperatura y precipitación comparando el período actual (1997-2007) y futuro (2020-2030). En la segunda columna la distribución de la diferencia (futuro-actual). En la tercera y cuarta columna las distribuciones marginales de temperatura y precipitación para ( India, Iraq, Japon, Korea del Sur).

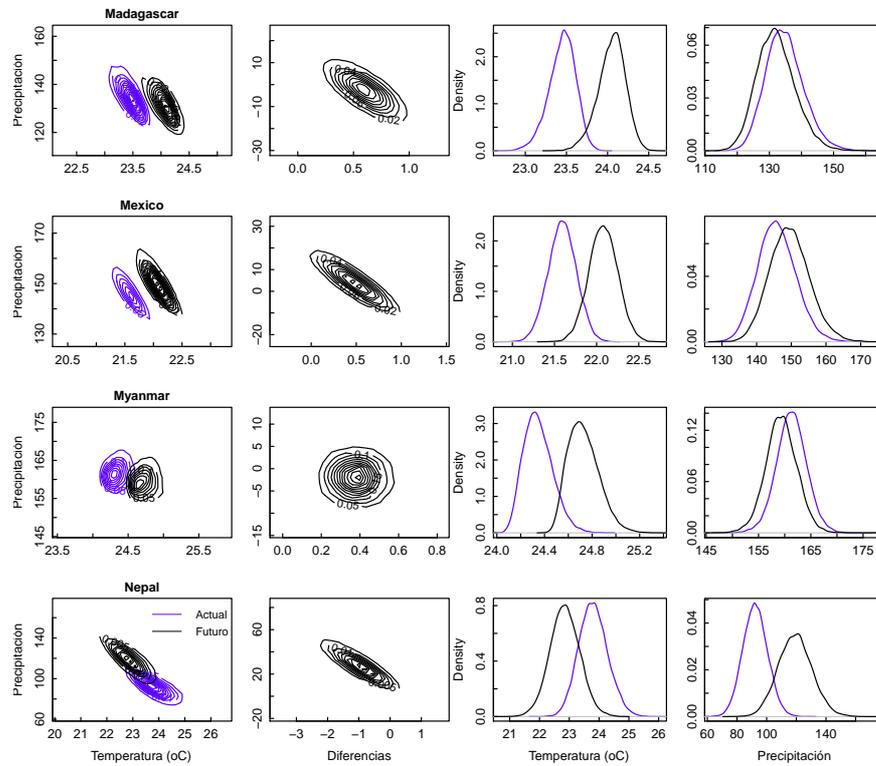


Figura A-7: En la primera columna contours de la distribución predictiva conjunta de temperatura y precipitación comparando el período actual (1997-2007) y futuro (2020-2030). En la segunda columna la distribución de la diferencia (futuro-actual). En la tercera y cuarta columna las distribuciones marginales de temperatura y precipitación para (Madagascar, Mexico, Myanmar, Nepal).

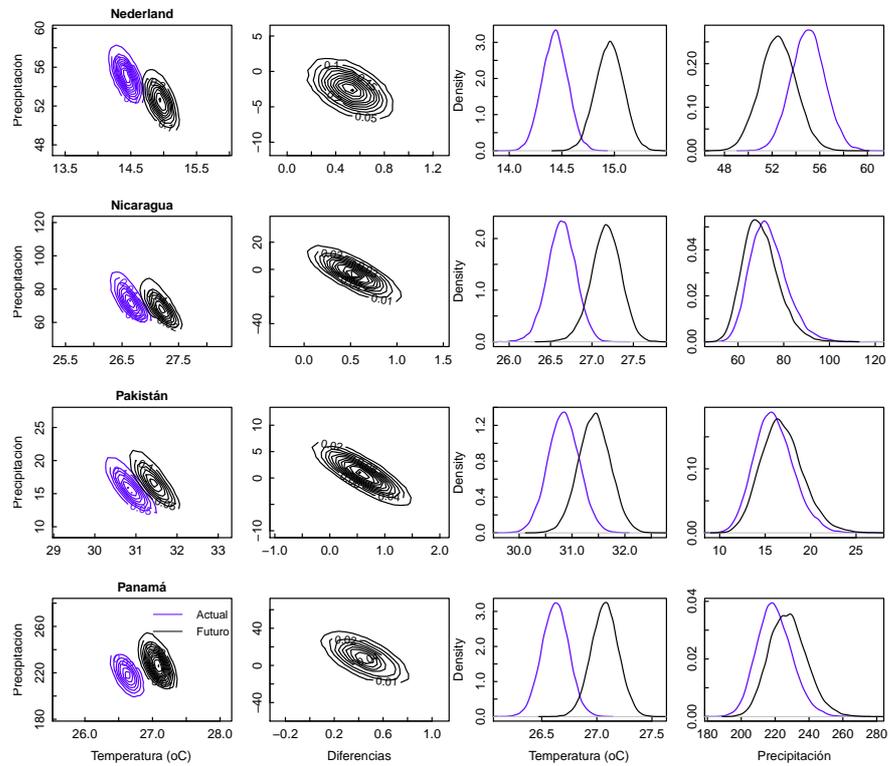


Figura A-8: En la primera columna contours de la distribución predictiva conjunta de temperatura y precipitación comparando el período actual (1997-2007) y futuro (2020-2030). En la segunda columna la distribución de la diferencia (futuro-actual). En la tercera y cuarta columna las distribuciones marginales de temperatura y precipitación para (Holanda, Nicaragua, Pakistan, Panama).

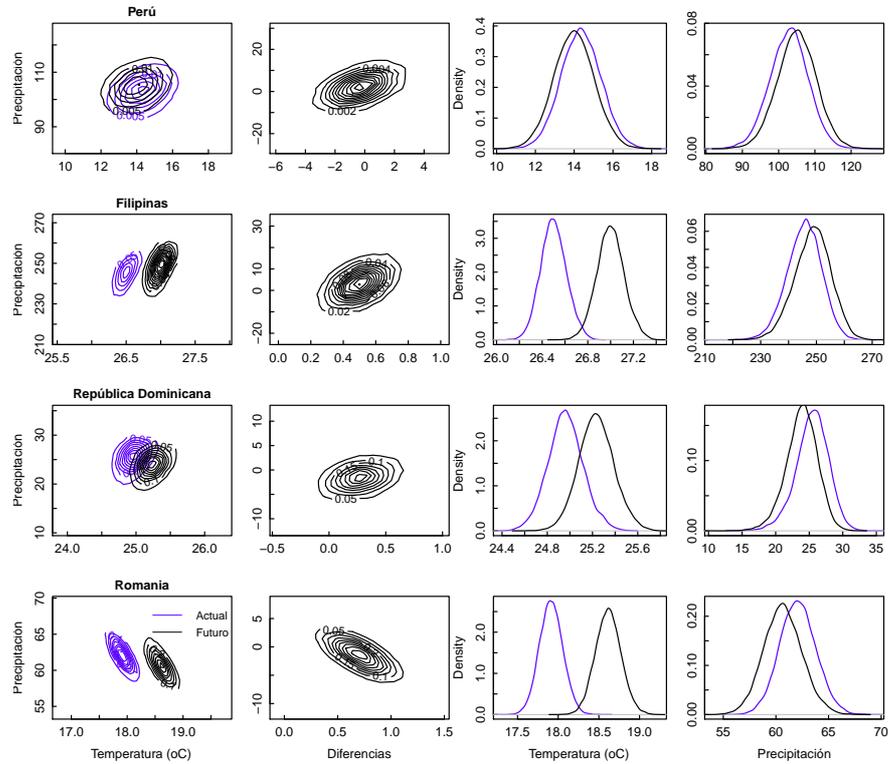


Figura A-9: En la primera columna contours de la distribución predictiva conjunta de temperatura y precipitación comparando el período actual (1997-2007) y futuro (2020-2030). En la segunda columna la distribución de la diferencia (futuro-actual). En la tercera y cuarta columna las distribuciones marginales de temperatura y precipitación para (Perú, Filipinas, República Dominicana y Romania).

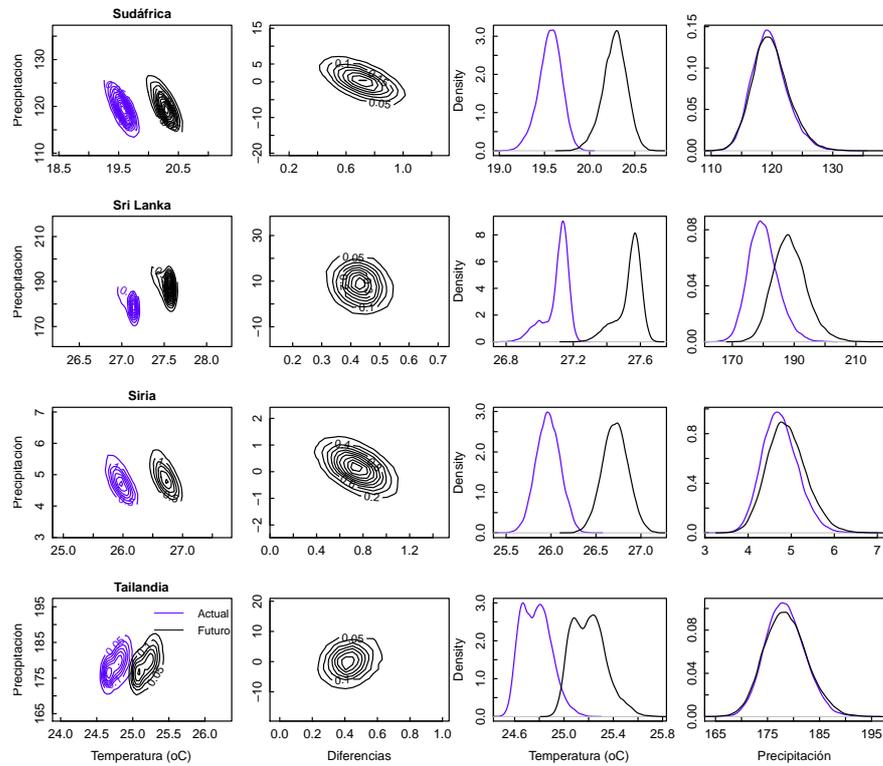


Figura A-10: En la primera columna contours de la distribución predictiva conjunta de temperatura y precipitación comparando el período actual (1997-2007) y futuro (2020-2030). En la segunda columna la distribución de la diferencia (futuro-actual). En la tercera y cuarta columna las distribuciones marginales de temperatura y precipitación para (Africa del Sur, Sri Lanka, Siria, Tailandia).

## A.2. Series de Temperatura y precipitación para los modelos proyectados, observados, tendencias centrales y distribución predictiva

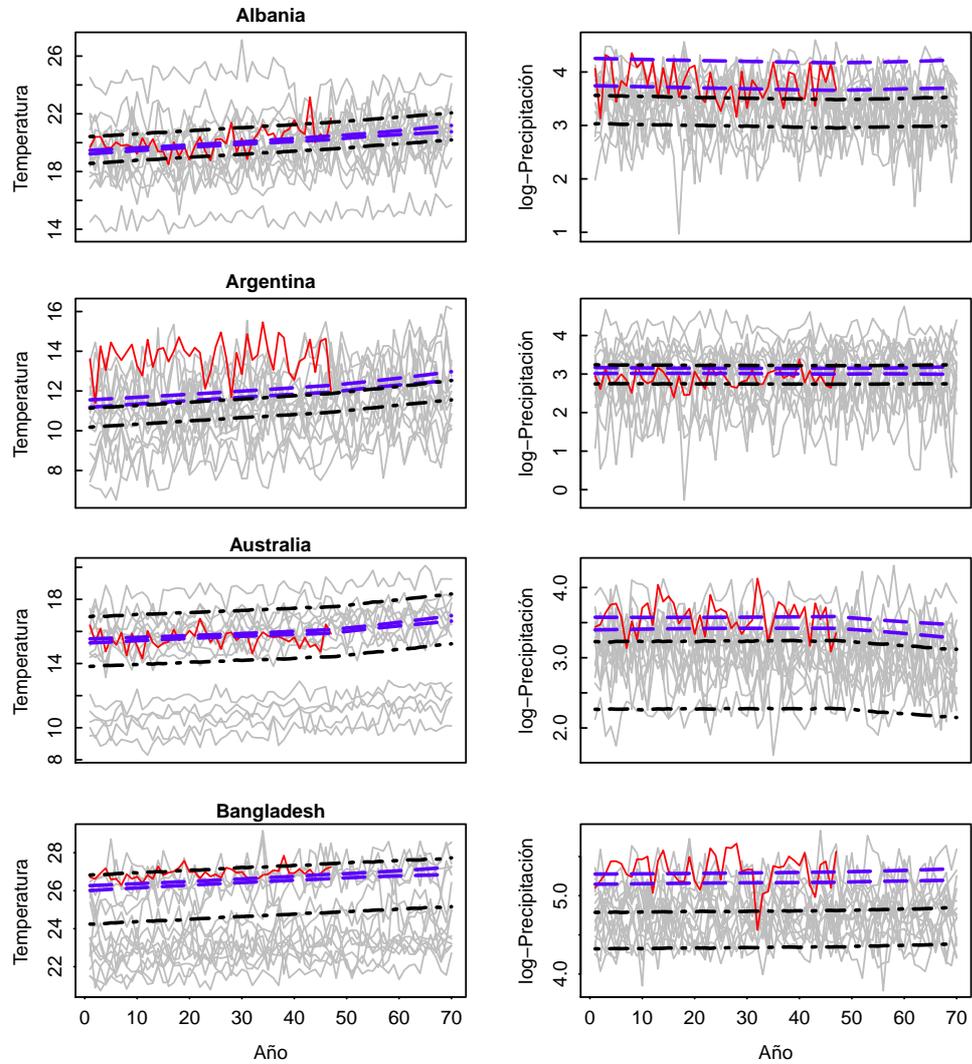


Figura A–11: Series de tiempo de temperatura y log-precipitación, líneas plomas son las series proyectadas por los 15 GCMs, líneas rojas series observadas, líneas azules representan los intervalos de credibilidad del 95 % para para el modelo del clima ( $\mu_t$ ) y líneas negras representan el IC del 95 % para la distribución predictiva para (Albania, Argentina, Australia, Bangladesh).

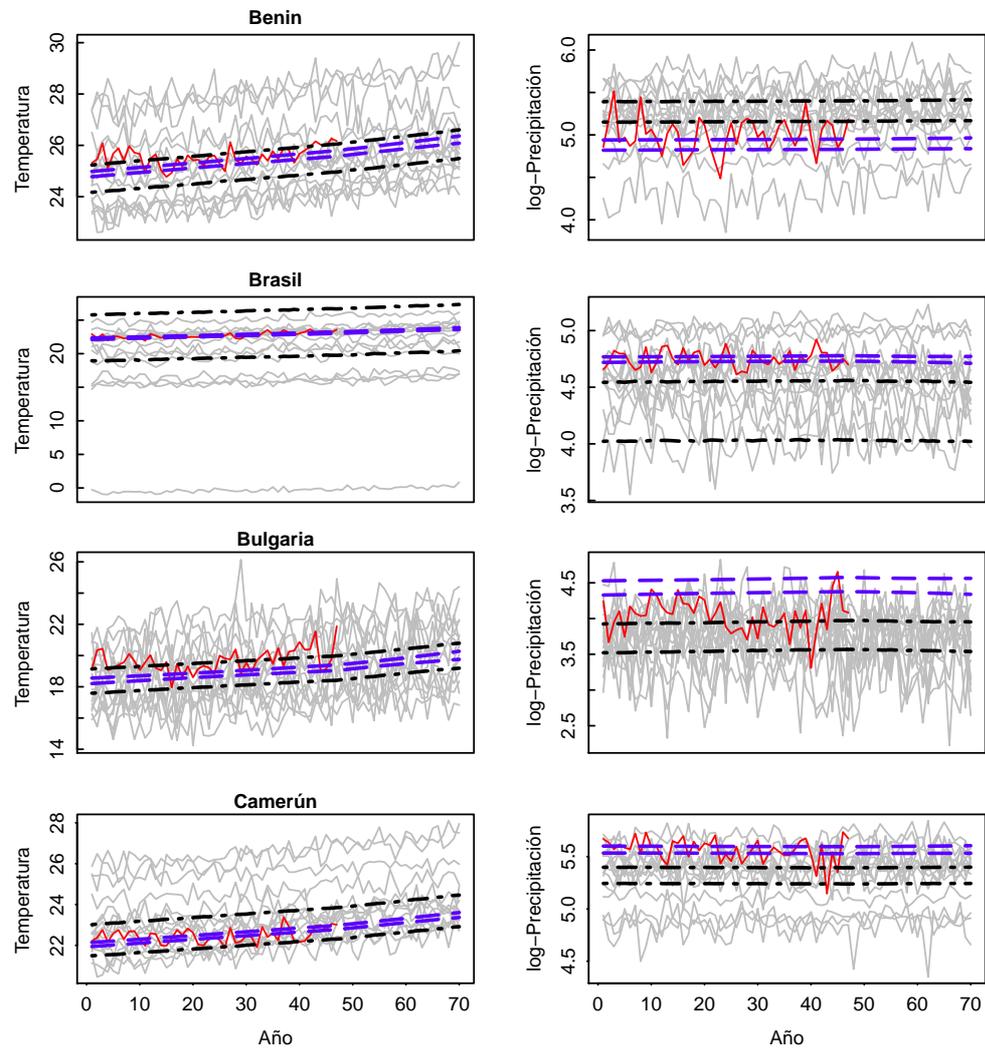


Figura A-12: Series de tiempo de temperatura y log-precipitación, líneas plomas son las series proyectadas por los 15 GCMs, líneas rojas series observadas, líneas azules representan los intervalos de credibilidad del 95% para para el modelo del clima ( $\mu_t$ ) y líneas negras representan el IC del 95% para la distribución predictiva para (Benin, Brasil, Bulgaria, Camerún).

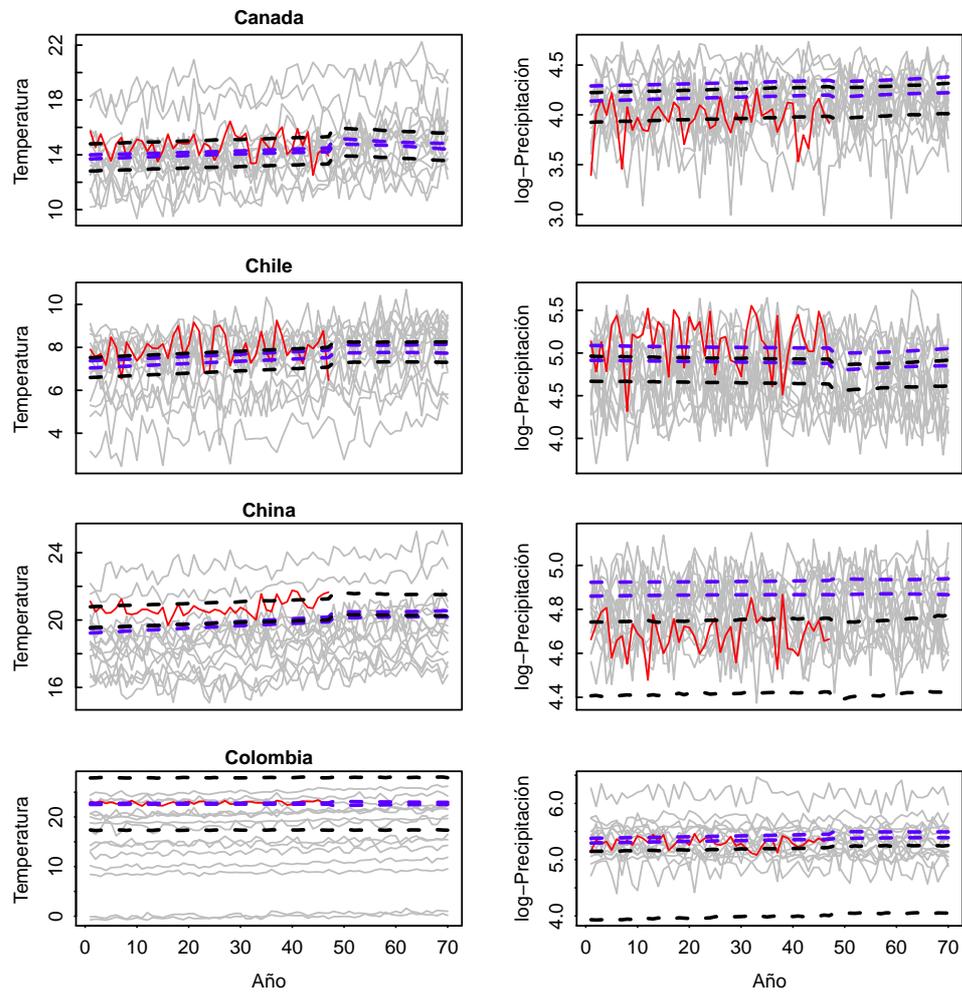


Figura A-13: Series de tiempo de temperatura y log-precipitación, líneas plomas son las series proyectadas por los 15 GCMs, líneas rojas Series observadas, líneas azules representan los intervalos de credibilidad del 95 % para para el modelo del clima ( $\mu_t$ ) y líneas negras representan el IC del 95 % para la distribución predictiva para (Canada, Chile, China, Colombia).

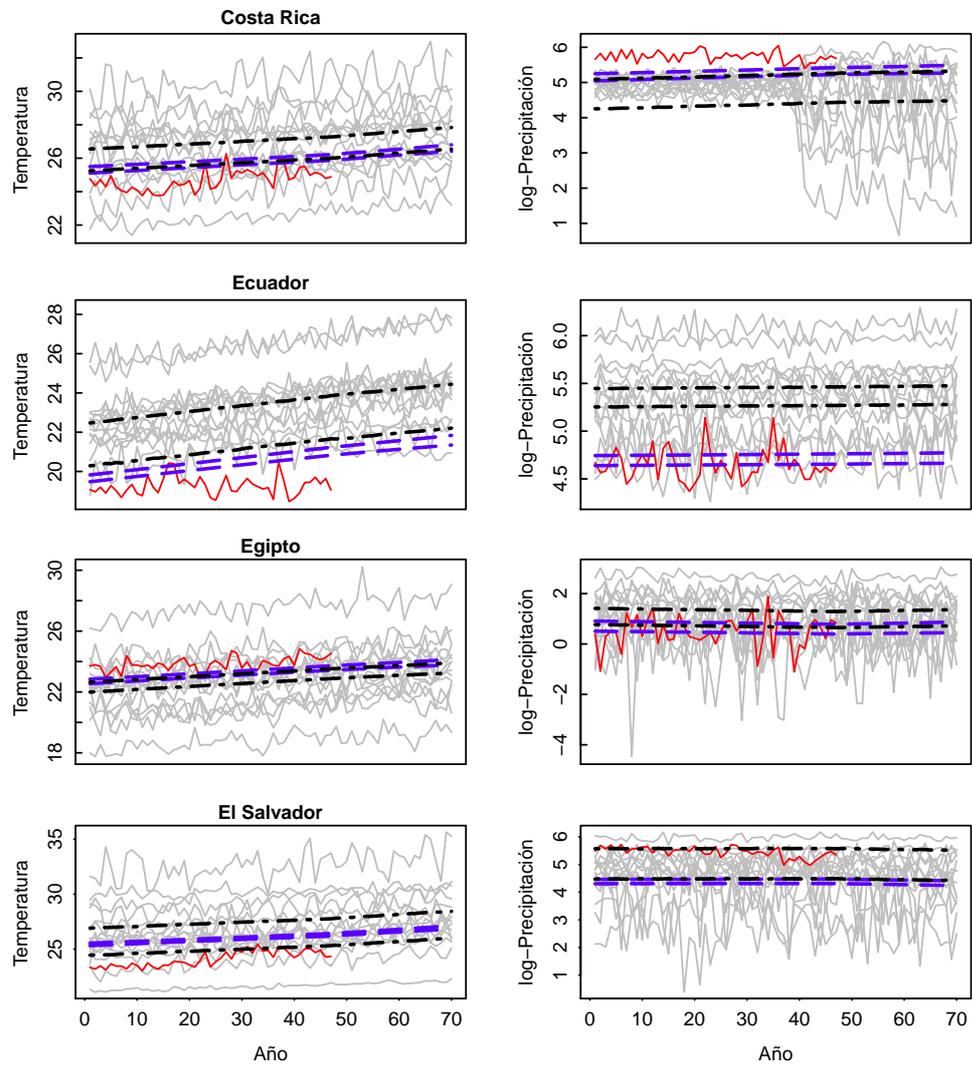


Figura A-14: Series de tiempo de temperatura y log-precipitación, líneas plomas son las series proyectadas por los 15 GCMs, líneas rojas series observadas, líneas azules representan los intervalos de credibilidad del 95% para el modelo del clima ( $\mu_t$ ) y líneas negras representan el IC del 95% para la distribución predictiva (Costa Rica, Ecuador, Egipto y El Salvador).

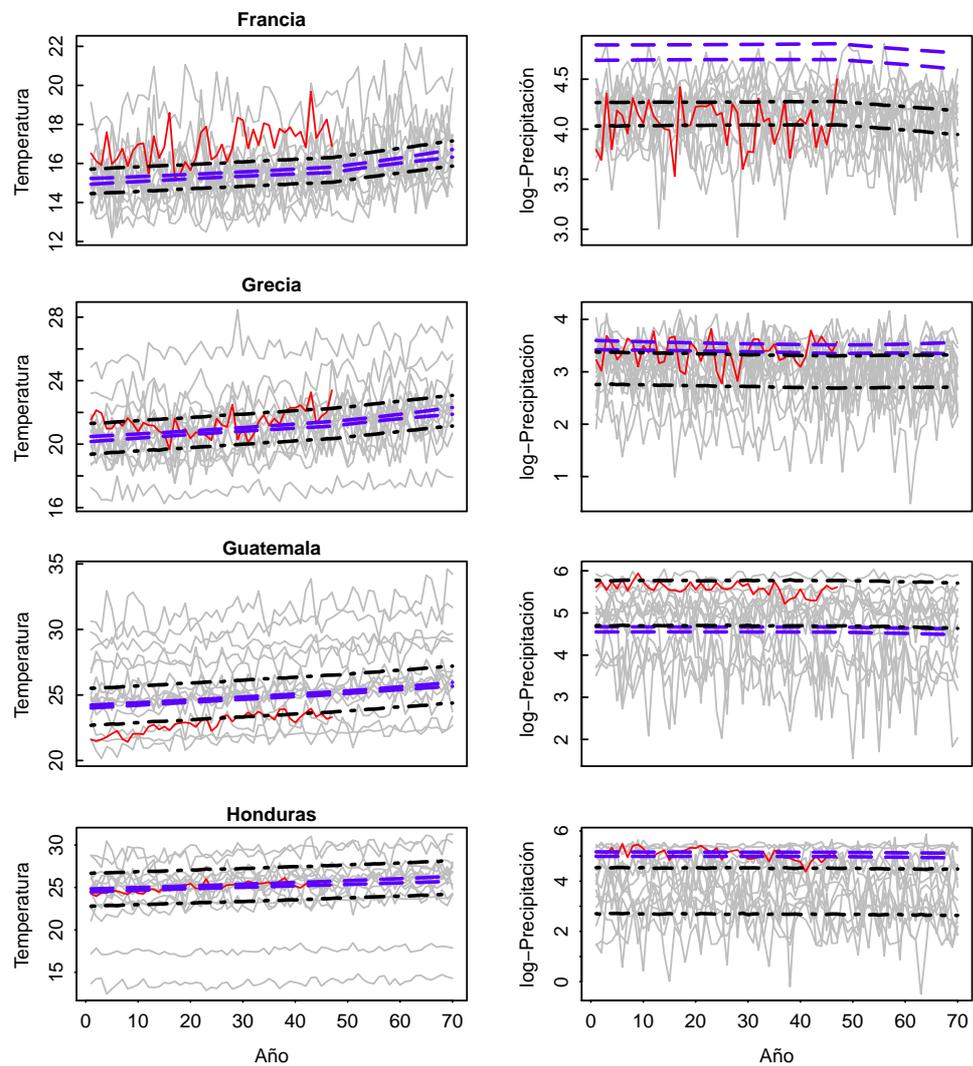


Figura A-15: Series de tiempo de temperatura y log-precipitación, líneas plomas son las series proyectadas por los 15 GCMs, líneas rojas series observadas, líneas azules representan los intervalos de credibilidad del 95 % para el modelo del clima ( $\mu_t$ ) y líneas negras representan el IC del 95 % para la distribución predictiva (Francia, Grecia, Guatemala, Honduras).

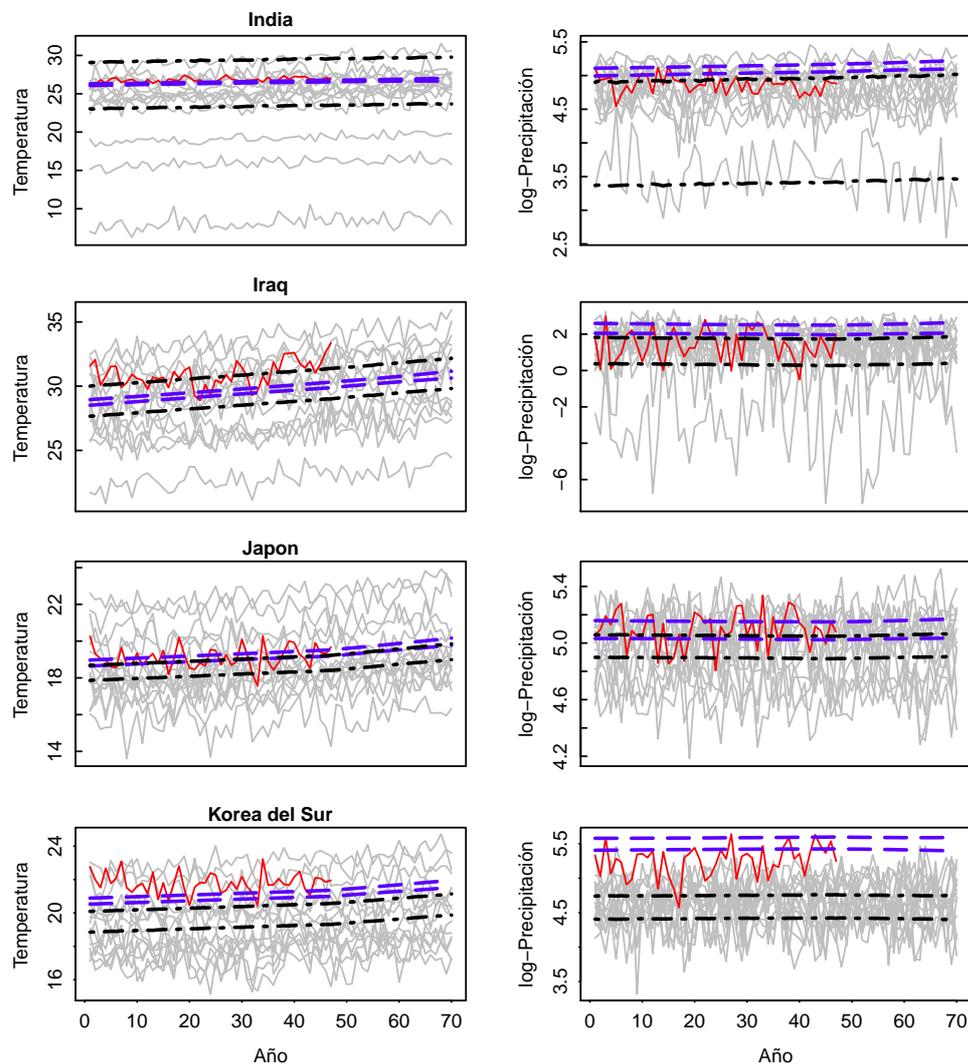


Figura A-16: Series de tiempo de temperatura y log-precipitación, líneas plomas son las series proyectadas por los 15 GCMs, líneas rojas series observadas, líneas azules representan los intervalos de credibilidad del 95 % para para el modelo del clima ( $\mu_t$ ) y líneas negras representan el IC del 95 % para la distribución predictiva (India, Iraq, Japón, Korea del Sur).

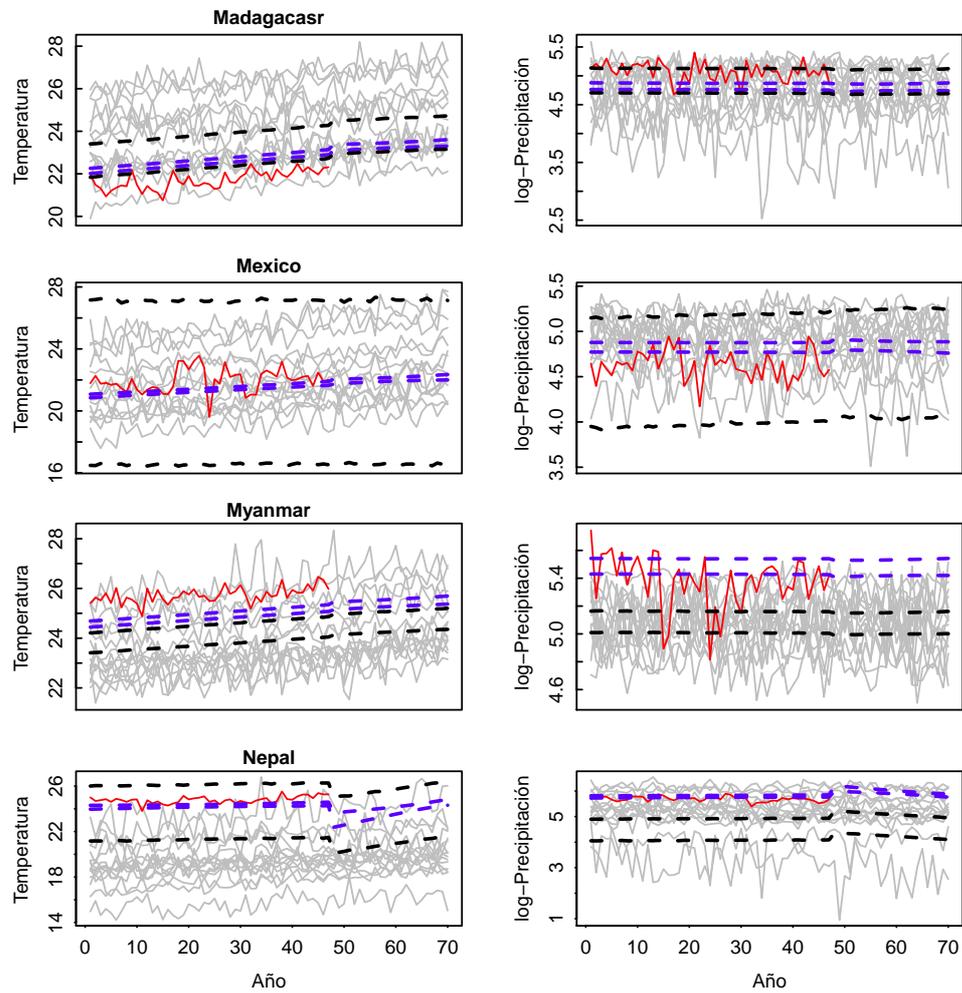


Figura A-17: Series de tiempo de temperatura y log-precipitación, líneas plomas son las series proyectadas por los 15 GCMs, líneas rojas series observadas, líneas azules representan los intervalos de credibilidad del 95 % para para el modelo del clima ( $\mu_t$ ) y líneas negras representan el IC del 95 % para la distribución predictiva (Madagascar, Mexico, Myanmar, Nepal).

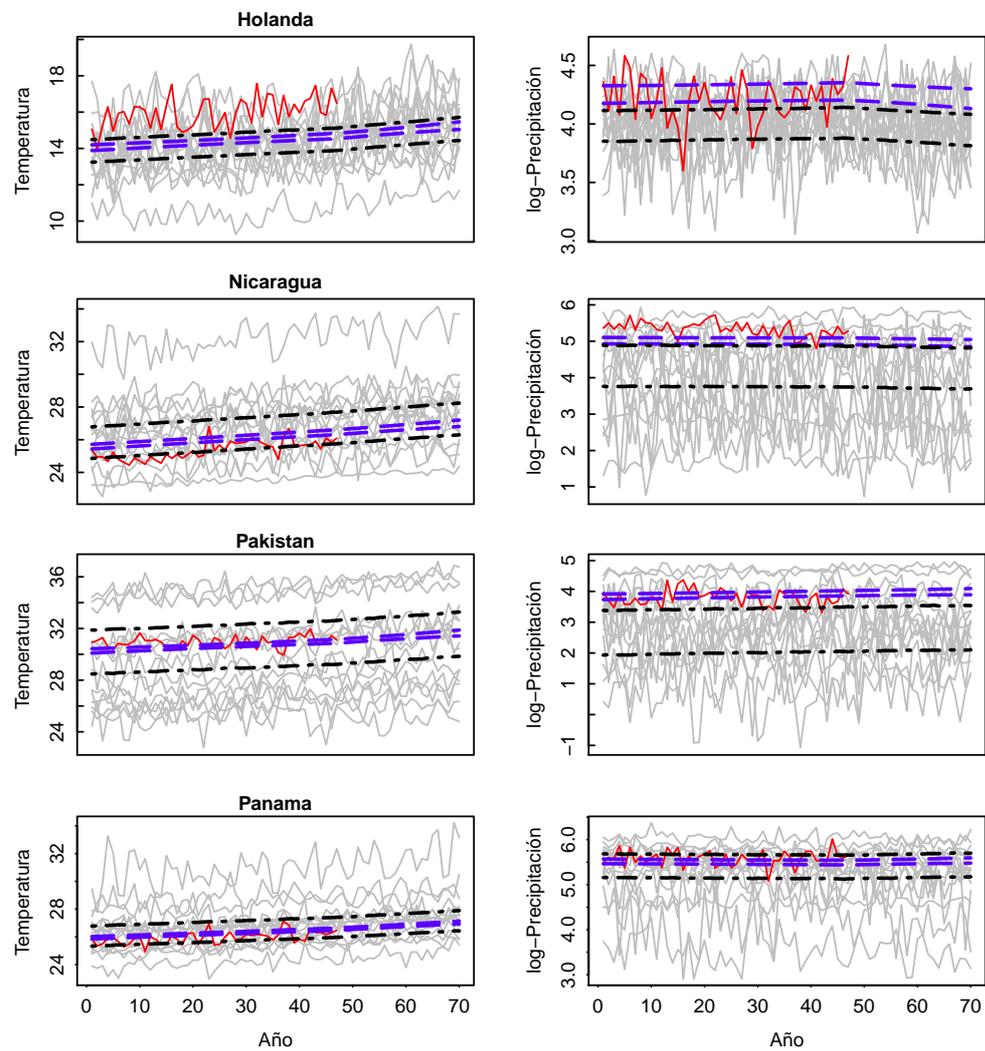


Figura A-18: Series de tiempo de temperatura y log-precipitación, líneas plomas son las series proyectadas por los 15 GCMs, líneas rojas series observadas, líneas azules representan los intervalos de credibilidad del 95% para para el modelo del clima ( $\mu_t$ ) y líneas negras representan el IC del 95% para la distribución predictiva (Holanda, Nicaragua, Pakistan, Panama).

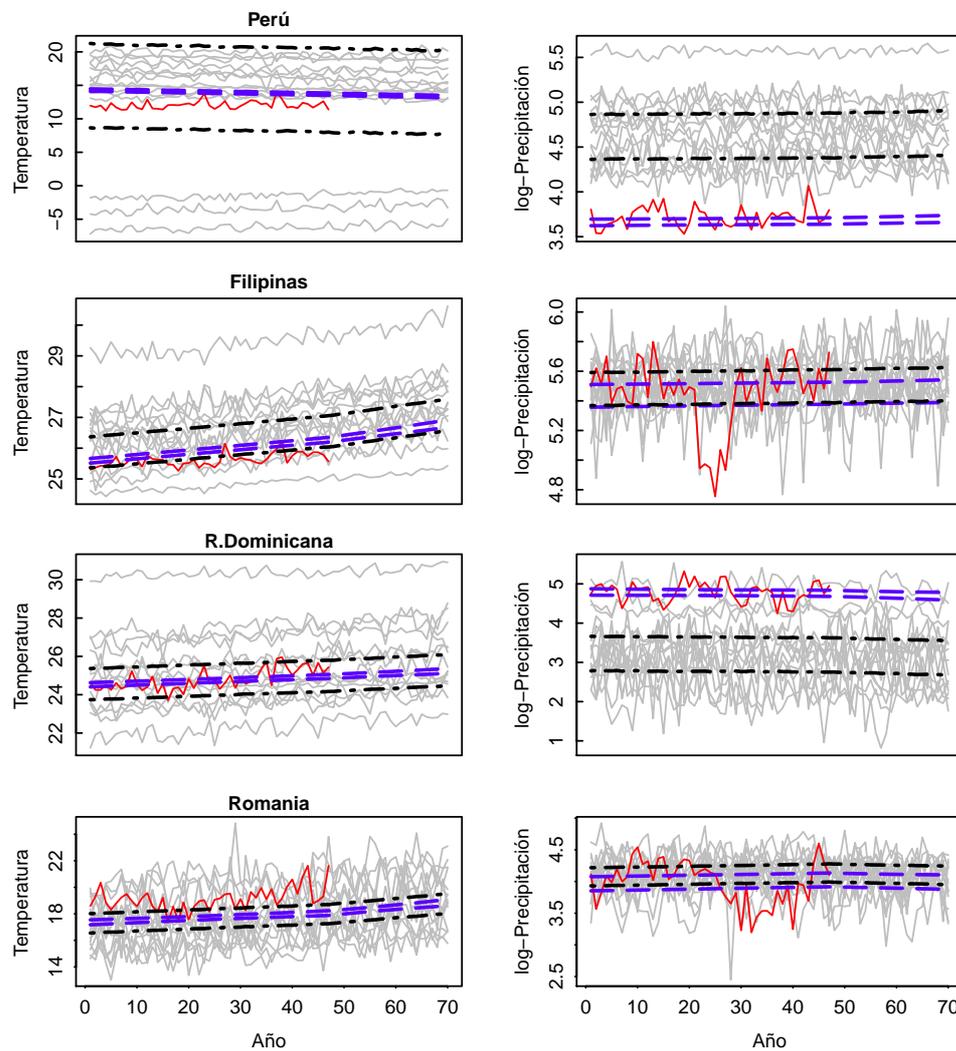


Figura A-19: Series de tiempo de temperatura y log-precipitación, líneas plomas son las series proyectadas por los 15 GCMs, líneas rojas series observadas, líneas azules representan los intervalos de credibilidad del 95 % para para el modelo del clima ( $\mu_t$ ) y líneas negras representan el IC del 95 % para la distribución predictiva (Perú, Filipinas, República Dominicana y Romania).

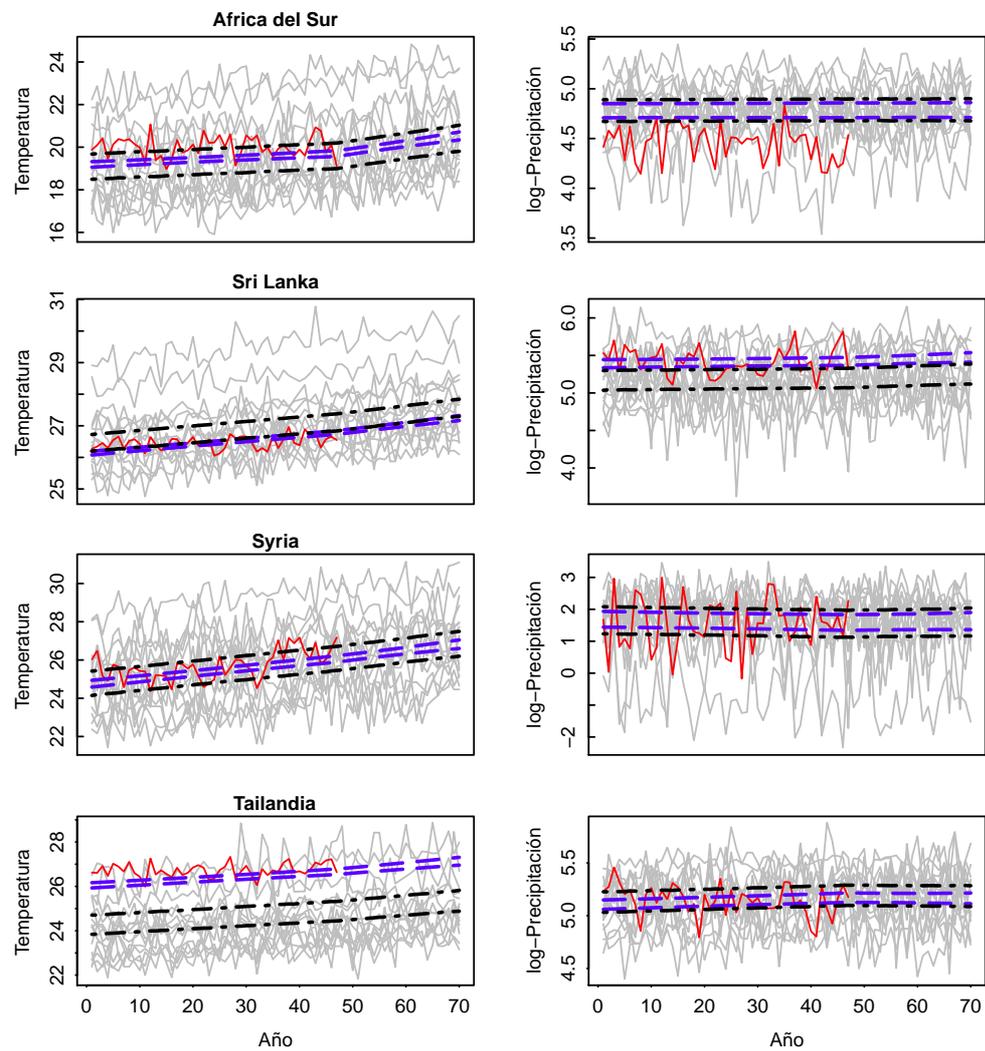


Figura A-20: Series de tiempo de temperatura y log-precipitación, líneas plomas son las series proyectadas por los 15 GCMs, líneas rojas series observadas, líneas azules representan los intervalos de credibilidad del 95 % para para el modelo del clima ( $\mu_t$ ) y líneas negras representan el IC del 95 % para la distribución predictiva (Africa del Sur, Sri Lanka, Syria, Tailandia).

## Apéndice B

### TABLAS

<b>País</b>	<b>País</b>
Albania	Japón
Argentina	Korea del sur
Australia	Madagascar
Bangladesh	Mexico
Benin	Myanmar
Brasil	Nepal
Bulgaria	Holanda
Camerún	Nicaragua
Canada	Pakistan
Chile	Panamá
China	Perú
Colombia	Filipinas
Costa Rica	R. Dominicana
Ecuador	Romania
Egipto	África del Sur
El Salvador	Sri Lanka
Francia	Syria
Grecia	Tailandia
Guatemala	Togo
Honduras	Uganda
India	USA
Iraq	Venezuela

Tabla B-1: Países incluidos en el presente estudio; productores de frijol que tienen más de 1000 hectáreas dedicadas al cultivo y no muestran rendimiento repetido para dos o 3 años consecutivos en data de FAO, 2006.

## Apéndice C

### DISTRIBUCIONES CONDICIONALES Y FORMULAS

#### C.1. Cálculo de las distribuciones condicionales de los parámetros

Para encontrar las distribuciones condicionales a partir de la ecuación 3.3 consideramos sólo los términos que incluyen al parámetro correspondiente, por tanto tenemos:

1.- Distribución condicional para  $\lambda_B | \theta_{\lambda_B}$

$$\lambda_B | \theta_{\lambda_B} \propto \left[ \prod_{j=1}^M f(\beta_{x_j} | \beta_0, \lambda_0) \right] f(\lambda_0) = (\lambda_B)^{\frac{M}{2} + g - 1} e^{\left\{ -\lambda_B \left[ \frac{\sum_j (\beta_{x_j} - \beta_0)^2}{2} + h \right] \right\}} \frac{\lambda^g}{\Gamma(g) (\sqrt{2\pi})^M} \quad (C.1)$$

$$\therefore \lambda_B | \theta_{\lambda_B} \sim G \left( \frac{M}{2} + g - 1, \frac{\sum_j (\beta_{x_j} - \beta_0)^2}{2} + h \right)$$

2.- Distribución condicional para  $\lambda_D^T | \theta_{\lambda_D^T}$

$$\lambda_D^T | \theta_{\lambda_D^T} \propto \left[ \prod_{j=1}^M f(d_j^T | a^T, \lambda_D^T) \right] f(\lambda_D^T) = \frac{(\lambda_D^T)^{\frac{M}{2} + g - 1}}{(\sqrt{2\pi})^M} e^{\left\{ -\lambda_D^T \left[ \frac{\sum_j (d_j^T - a^T)^2}{2} + h \right] \right\}} \frac{\lambda^g}{\Gamma(g)} \quad (C.2)$$

$$\therefore \lambda_D^T | \theta_{\lambda_D^T} \sim G \left( \frac{M}{2} + g - 1, \frac{\sum_j (d_j^T - a^T)^2}{2} + h \right)$$

3.- Distribución condicional para  $\lambda_D^p | \theta_{\lambda_D^p}$

$$\lambda_D^p | \theta_{\lambda_D^p} \propto \left[ \prod_{j=1}^M f(d_j^p | a^p, \lambda_D^p) \right] f(\lambda_D^p) = (\lambda_D^p)^{\frac{M}{2} + g - 1} e^{\left\{ -\lambda_D^p \left[ \frac{\sum_j (d_j^p - a^p)^2}{2} + h \right] \right\}} \frac{h^g}{\Gamma(g) (\sqrt{2\pi})^M} \quad (\text{C.3})$$

$$\therefore \lambda_D^p | \theta_{\lambda_D^p} \sim G \left( \frac{M}{2} + g - 1, \frac{\sum_j (d_j^p - a^p)^2}{2} + h \right)$$

4.- Distribución condicional para  $\varepsilon_j^T | \theta_{\varepsilon_j^T}$ , para un solo  $j$

$$\begin{aligned} \varepsilon_j^T | \theta_{\varepsilon_j^T} &\propto \prod_{t \leq \tau^*} (f(X_{jt}^T | \mu_t^T + d_t^T, \varepsilon_j^T)) f(\varepsilon_j^T | a_{\varepsilon^T}, b_{\varepsilon^T}) \\ &= (\varepsilon_j^T)^{\frac{\tau^*}{2} + a_{\varepsilon^T} - 1} e^{\left\{ -\varepsilon_j^T \left( \sum_{t \leq \tau^*} \frac{(X_{jt}^{T'} - d_j)^2}{2} + b_{\varepsilon^T} \right) \frac{(b_{\varepsilon^T})^{a_{\varepsilon^T}}}{\Gamma(a_{\varepsilon^T}) (\sqrt{2\pi})^{\tau^*}} \right\}} \end{aligned} \quad (\text{C.4})$$

$$\therefore \varepsilon_j^T | \theta_{\varepsilon_j^T} \sim G \left( \frac{\tau^*}{2} + a_{\varepsilon^T} - 1, \sum_{t \leq \tau^*} \frac{(X_{jt}^{T'} - d_j)^2}{2} + b_{\varepsilon^T} \right) \text{ donde } X_{jt}^{T'} = X_{jt}^T - \mu_t^T$$

5.- Distribución condicional para  $\varepsilon_j^p | \theta_{\varepsilon_j^p}$ , para un solo  $j$  y  $t = 1, \dots, \tau^*$

$$\begin{aligned} \varepsilon_j^p | \theta_{\varepsilon_j^p} &\propto \left[ \prod_{t \leq \tau^*} f(X_{jt}^p | \mu_t^p, \beta_{X_j}, \mu_t^T, d_t^T, d_t^p, X_{jt}^T, \varepsilon_j^p) \right] f(\varepsilon_j^p | a_{\varepsilon^p}, b_{\varepsilon^p}) \\ &= (\varepsilon_j^p)^{\frac{\tau^*}{2} + a_{\varepsilon^p} - 1} e^{\left\{ -\varepsilon_j^p \left[ \sum_{t \leq \tau^*} \frac{\left( X_{jt}^{p'} - d_j^p - \beta_{X_j} (X_{jt}^{T'} - d_j^T) \right)^2}{2} + b_{\varepsilon^p} \right] \right\}} \end{aligned} \quad (\text{C.5})$$

$$\therefore \varepsilon_j^p | \theta_{\varepsilon_j^p} \sim G \left( \frac{\tau^*}{2} + a_{\varepsilon^p} - 1, \sum_{t \leq \tau^*} \frac{\left( X_{jt}^{p'} - d_j^p - \beta_{X_j} (X_{jt}^{T'} - d_j^T) \right)^2}{2} + b_{\varepsilon^p} \right)$$

6.- Distribución condicional para  $\beta_0|\theta_{\beta_0}$

$$\begin{aligned}\beta_0|\theta_{\beta_0} &\propto \left[ \prod_{j=1}^M f(\beta_{x_j} | \beta_0, \lambda_B) \right] f(\beta_{x_0} | \beta_0, \lambda_0) \\ &= (\lambda_B)^{\frac{M}{2}} (\lambda_0)^{\frac{1}{2}} e^{\left\{ -\frac{(\lambda_B M - \lambda_0)}{2} \left( \beta_0 - \frac{\lambda_B \sum \beta_{x_j} + \lambda_0 \beta_{x_0}}{\lambda_B M + \lambda_0} \right)^2 \right\}} \\ &\quad * e^{\left\{ -\lambda_B \sum \beta_{x_j} + \lambda_0 \beta_{x_0}^2 + \left( \frac{\lambda_B \sum \beta_{x_j} + \lambda_0 \beta_{x_0}}{2} \right)^2 \right\}}\end{aligned}\quad (C.6)$$

$$\therefore \beta_0|\theta_{\beta_0} \sim N\left(\frac{\lambda_B \sum \beta_{x_j} + \lambda_0 \beta_{x_0}}{\lambda_B M + \lambda_0}, \frac{1}{\lambda_B M + \lambda_0}\right)$$

7.- Distribución condicional para  $\beta_{x_j}|\theta_{\beta_{x_j}}$ , para un solo  $j$

$$\begin{aligned}\beta_{x_j}|\theta_{\beta_{x_j}} &\propto \prod_{t \leq \tau^*} f(X_{jt}^p | \mu_t^p, \beta_{x_j}, \mu_t^T, d_j^T, d_j^p, \varepsilon_j^p, X_{jt}^T) f(\beta_{x_j} | \beta_0, \lambda_B) \\ &= e^{\left\{ -\frac{1}{2} \left( \sum_{t \leq \tau^*} \varepsilon_j^p (X_{jt}^{T'} - d_j^T)^2 + \lambda_B \right) \left( \beta_{x_j} - \frac{\varepsilon_j^p \sum_{t \leq \tau^*} (X_{jt}^{p'} - d_j^p) (X_{jt}^{T'} - d_j^T) + \lambda_B \beta_0}{\varepsilon_j^p \sum_{t \leq \tau^*} (X_{jt}^{T'} - d_j^T)^2 + \lambda_B} \right)^2 \right\}}\end{aligned}\quad (C.7)$$

$$\therefore \beta_{x_j}|\theta_{\beta_{x_j}} \sim N\left(\frac{\varepsilon_j^p \sum_{t \leq \tau^*} (X_{jt}^{p'} - d_j^p) (X_{jt}^{T'} - d_j^T) + \lambda_B \beta_0}{\varepsilon_j^p \sum_{t \leq \tau^*} (X_{jt}^{T'} - d_j^T)^2 + \lambda_B}, \frac{1}{\varepsilon_j^p \sum_{t \leq \tau^*} (X_{jt}^{T'} - d_j^T)^2 + \lambda_B}\right),$$

consideremos

$$B = \varepsilon_j^p \sum_{t \leq \tau^*} (X_{jt}^{p'} - d_j^p) (X_{jt}^{T'} - d_j^T) + \lambda_B \beta_0$$

y

$$A = \varepsilon_j^p \sum_{t \leq \tau^*} (X_{jt}^{T'} - d_j^T)^2 + \lambda_B$$

8.- Distribución condicional para  $a^p|\theta_{a^p}$

$$\begin{aligned}a^p|\theta_{a^p} &\propto \prod_{j=1}^M f(d_j^p | a^p, \lambda_D^p) \\ &= \frac{(\lambda_D^p)^M}{(\sqrt{2\pi})^M} e^{\left\{ -\frac{M\lambda_D^p}{2} \left( a^p - \frac{\sum d_j^p}{M} \right)^2 + \frac{M\lambda_D^p}{2} \left( \left( \frac{\sum d_j^p}{M} \right)^2 - \frac{\sum (d_j^p)^2}{M} \right) \right\}}\end{aligned}\quad (C.8)$$

$$\therefore a^p|\theta_{a^p} \sim N\left(\frac{B}{A}, A\right), \text{ donde } A = M\lambda_D^p, B = \lambda_D^p \frac{\sum d_j^p}{M}$$

9.- Distribución condicional para  $a^T | \theta_{a^T}$

De forma similar que en la parte 8 se puede encontrar para  $a^T$

$$\therefore a^T | \theta_{a^T} \sim N \left( \frac{\lambda_D^T \sum_j d_j^p}{\lambda_D^T M}, M \lambda_D^P \right)$$

10.- Distribución condicional para  $d_j^p | \theta_{d_j^p}$ , para un solo  $j$

$$\begin{aligned} d_j^p | \theta_{d_j^p} &\propto \prod_{t \leq \tau^*} [f(X_{jt}^p | \mu^p, \beta_{x_j}, \mu_t^T, d_j^T, d_j^p, \varepsilon_j, X_{jt}^T)] f(d_j^p | a^p, \lambda_D^P) \\ &= e \left\{ -\frac{1}{2} (\tau^* \varepsilon_j^P + \lambda_D^P) \left[ d_j^p - \frac{\varepsilon_j^p \sum_{t \leq \tau^*} (X_{jt}^{p'} - \beta_{x_j} (X_{jt}^{T'} - d_j^T) + a^p \lambda_D^P)}{\tau^* \varepsilon_j^p + \lambda_D^P} \right]^2 \right\} \end{aligned} \quad (C.9)$$

Si consideramos

$$A = \tau^* \varepsilon_j^p + \lambda_D^P$$

y

$$B = \varepsilon_j^p \sum_{t \leq \tau^*} \left( X_{jt}^{p'} - \beta_{x_j} (X_{jt}^{T'} - d_j^T) + a^p \lambda_D^P \right)$$

$$\therefore d_j^p | \theta_{d_j^p} \sim N \left( \frac{B}{A}, \frac{1}{A} \right)$$

11.- Distribución condicional para  $d_j^T | \theta_{d_j^T}$ , para un solo  $j$

$$\begin{aligned} d_j^T | \theta_{d_j^T} &\propto \left\{ \prod_{t \leq \tau^*} f(X_{jt}^T | \mu_t^T, d_j^T, \varepsilon_j^T) f(X_{jt}^p | \mu_t^p, \beta_{x_j}, \mu_t^T, d_j^T, d_j^p, \varepsilon_j^p, X_{jt}^T) \right\} f(d_j^T | a_D^T) \\ &= e \left\{ -\frac{(\tau^* \varepsilon_j^T + \tau^* \varepsilon_j^p \beta_{x_j}^2 + \lambda_D^T)}{2} \left( d_j^T - \frac{\varepsilon_j^T \sum_{t \leq \tau^*} X_{jt}^{T'} - \varepsilon_j^p \sum_{t \leq \tau^*} (X_{jt}^{p'} - \beta_{x_j} X_{jt}^{T'} - d_j^p) + \lambda_D^T a^T}{\tau^* \varepsilon_j^T + \tau^* \varepsilon_j^p \beta_{x_j}^2 + \lambda_D^T} \right)^2 \right\} \end{aligned} \quad (C.10)$$

Si consideramos

$$A = \tau^* \varepsilon_j^T + \tau^* \varepsilon_j^p \beta_{x_j}^2 + \lambda_D^T$$

$$B = \varepsilon_j^T \sum_{t \leq \tau^*} X_{jt}^{T'} - \varepsilon_j^p \sum_{t \leq \tau^*} \left( X_{jt}^{p'} - \beta_{x_j} X_{jt}^{T'} - d_j^p \right) + \lambda_D^T a^T$$

$$\therefore d_j^T | \theta_{d_j^T} \sim N \left( \frac{B}{A}, \frac{1}{A} \right)$$

12.- Distribución condicional para  $\gamma^p | \theta_{\gamma^p}$

$$\begin{aligned} \gamma^p | \theta_{\gamma^p} &\propto \prod_t [f(O_t^p | \mu_t^p, n^p)] \prod_j f(X_{jt}^P | \mu_t^p, \beta_{x_j}, X_{jt}^T, \mu_t^T, d_j^T, d_j^p, \varepsilon_j^p) \\ &= \prod_j f(X_{jt}^P | \mu_t^p, \beta_{x_j}, X_{jt}^T, \mu_t^T, d_j^T, d_j^p, \varepsilon_j^p) \\ &= e \left\{ -\frac{\sum_{t \geq \tau_0} (t - \tau_0)^2 \sum_j \varepsilon_j^p}{2} \left( \gamma^p - \frac{\sum_{t \geq \tau_0} (t - \tau_0) \sum_j \varepsilon_j \left( X_{jt}^p - \alpha^p - \beta^p t - \beta_{x_j} \left( X_{jt}^{T'} - d_j^T \right) - d_j^p \right)}{\sum_{t \geq \tau_0} (t - \tau_0)^2 \sum_j \varepsilon_j^p} \right) \right\} \end{aligned} \quad (C.11)$$

Si consideramos

$$B = \sum_{t \geq \tau_0} (t - \tau_0) \sum_j \varepsilon_j \left( X_{jt}^p - \alpha^p - \beta^p t - \beta_{x_j} \left( X_{jt}^{T'} - d_j^T \right) - d_j^p \right)$$

y

$$A = \sum_{t \geq \tau_0} (t - \tau_0)^2 \sum_j \varepsilon_j^p$$

$$\therefore \gamma^p | \theta_{\gamma^p} \sim N \left( \frac{B}{A}, \frac{1}{A} \right)$$

13.- Distribución condicional para  $\gamma^T | \theta_{\gamma^T}$  solo para  $t \geq \tau_0$

$$\begin{aligned} \gamma^T | \theta_{\gamma^T} &\propto \prod_t \left[ \prod_j f(X_{jt} | \mu_t^T, d_j^T, \varepsilon_j^T) f(X_{jt}^p | \mu_t^p, \beta_{x_j}, \mu_t^T, d_j^T, d_j^p, \varepsilon_j^p, X_{jt}^T) \right] \\ &= e \left\{ -\frac{A}{2} \left[ \gamma^p - \frac{B}{A} \right] \right\} \end{aligned} \quad (C.12)$$

donde

$$A = \sum_{t \geq \tau_0} (t - \tau_0)^2 \left( \sum_j \varepsilon_j^T + \sum_j \varepsilon_j^p \beta_{x_j}^2 \right)$$

y

$$\begin{aligned}
B &= \sum_{t \geq \tau_0} t \sum_j \varepsilon_j^T (X_{jt}^T - d_j^T - \alpha^T - \beta^T t) \\
&\quad - \sum_{t \geq \tau_0} t \sum_j \varepsilon_j^p \beta_{x_j} \left( X_{jt}^{p'} - \beta_{x_j} (X_t^T - \alpha^T - \beta^T t - d_j^T) - d_j^p \right) \\
&\quad - \sum_{t \geq \tau_0} \tau_0 \sum_j \varepsilon_j^T (X_{jt}^T - d_j^T - \alpha^T - \beta^T t) \\
&\quad + \sum_{t \geq \tau_0} \tau_0 \sum_j \varepsilon_j^p \beta_{x_j} \left( X_{jt}^{p'} - \beta_{x_j} (X_t^T - \alpha^T - \beta^T t - d_j^T) - d_j^p \right)
\end{aligned}$$

$$\therefore \gamma^T | \theta_{\gamma^T} \sim N \left( \frac{B}{A}, \frac{1}{A} \right)$$

14.- Distribución condicional para  $\beta^p | \theta_{\beta^p}$ 

$$\begin{aligned}
\beta^p | \theta_{\beta^p} &\propto \prod_{t \leq \tau^0} [f(O_t^p | \mu_t^p, n^p)] \prod_{t \leq \tau^*} \prod_j f(X_{jt}^p | \mu_t^p, \beta_{x_j}, \mu_t^T, d_j^p, d_j^T, \varepsilon_j^p, X_{jt}^T) \\
&= e^{\left\{ -\frac{A}{2} [(\beta^p)^2 - \frac{B}{A}] \right\}}
\end{aligned} \tag{C.13}$$

donde

$$A = n^p \sum_{t \leq \tau^0} t^2 + \sum_{t \leq \tau^0} t^2 \varepsilon_j^p$$

y

$$\begin{aligned}
B &= \sum_{t \leq \tau^0} n^p t (O_t^p - \beta_{x_0} (O_t^T - \mu_t^T)) - \alpha^T \\
&\quad + \sum_{t \leq \tau^*} t \sum_j \varepsilon_j^p \left( X_{jt}^p - \beta_{x_j} (X_{jt}^{T'} - d_t^T) - d_j^p - \alpha^p - \gamma^p (t - \tau_0)_I \right)
\end{aligned}$$

$$\therefore \beta^p | \theta_{\beta^p} \sim N \left( \frac{B}{A}, \frac{1}{A} \right)$$

15.- Distribución condicional para  $\beta^T | \theta_{\beta^T}$ 

$$\begin{aligned}
\beta^T | \theta_{\beta^T} &\propto \prod_{t \leq \tau^0} f(O_t^T | \mu_t^T, n^T) f(O_t^p | \mu_t^p, n^p, O_t) \prod_{\tau^*} \prod_{j=1}^M f(X_{jt}^T | \mu_t^T, d_t^T, \varepsilon_j^T) \\
&\quad \cdot f(X_{jt}^p | \mu_t^p, \beta_{x_j}, \mu_t^T, d_j^T, d_j^p, \varepsilon_j^p, X_{jt}^T) = e^{\left\{ -\frac{1}{2} [(\beta^T)^2 A - 2\beta^T B] \right\}}
\end{aligned} \tag{C.14}$$

donde

$$A = n^T \sum_{t \leq \tau^0} t^2 + n^p \beta_{x_0}^2 \sum_{t \leq \tau^0} t^2 + \sum_{\tau^*} t^2 \sum_j \varepsilon_j^T + \sum_{\tau^*} t^2 \sum_j \varepsilon_j^p \beta_{x_j}^2$$

y

$$\begin{aligned} B &= n^T \sum_{t \leq \tau^0} t (O_t^T - \alpha^T) - n^p \beta_{x_0} \sum_{t \leq \tau^0} \left[ O_t^{p'} - \beta_{x_0} (O_t^T - \alpha^T) \right] \\ &\quad + \sum_{t \leq \tau^*} t \sum_j \varepsilon_j^T \left[ X_{jt}^{T'} - \alpha^T - \gamma^T (t - t_0)_I - d_j^T \right] \\ &\quad - \sum_{t \leq \tau^*} t \sum_j \varepsilon_j^p \beta_{x_j} (X_{jt}^p - \beta_{x_j} [X_{jt}^T - \alpha^T - \gamma^T (t - t_0)_I - d_j^T] - d_j^p) \end{aligned}$$

$$\therefore \beta^T | \theta_{\beta^T} \sim N \left( \frac{B}{A}, A \right)$$

16.- Distribución condicional para  $\alpha^p | \theta_{\alpha^p}$

$$\begin{aligned} \alpha^p | \theta_{\alpha^p} &\propto \prod_{\tau^0} [f(O_t^p | \mu_t^p, n^p)] \prod_{\tau^*} \prod_j f(X_{jt}^p | \mu_t^p, \beta_{x_j}, \mu_t^T, d_j^T, d_j^p, \varepsilon_j^p, X_{jt}^T) \\ &= e^{\left\{ -\frac{A}{2} (\alpha^p - \frac{B}{A})^2 \right\}} \end{aligned} \quad (\text{C.15})$$

donde

$$A = n^p \tau^0 + \tau^* \sum_j \varepsilon_j^p$$

y

$$\begin{aligned} B &= n^p \sum_{t \leq \tau_0} \left( O_t^p - \beta_{x_0} O_t^{T'} - \beta^0 t \right) \\ &\quad + \sum_{t \leq \tau^*} \sum_j \varepsilon_j^p \left( X_{jt}^p - \beta_{x_j} \left( X_{jt}^{T'} - d_j^T \right) - d_j^p - \beta^p t - \gamma^p (t - t_0)_I \right) \end{aligned}$$

$$\therefore \alpha^p | \theta_{\alpha^p} \sim N \left( \frac{B}{A}, \frac{1}{A} \right)$$

17.- Distribución condicional para  $\alpha^T | \theta_{\alpha^T}$

$$\begin{aligned} \alpha^T | \theta_{\alpha^T} &\propto \left[ \prod_{t \leq \tau^0} f(O_t^T | \mu_t^T, n^T) f(O_t^P | \mu_t^P, n^P, O_t^T) \right] \left[ \prod_{t \leq \tau^*} \prod_{j=1}^M f(X_{jt}^T | \mu_t^T, d_j^T, \varepsilon_j^T) \right] \\ &= e^{\left\{ -\frac{A}{2} (\alpha^T - \frac{B}{A})^2 \right\}} \end{aligned} \quad (\text{C.16})$$

donde

$$A = \tau_0 n^T + \tau^0 n^P \beta_{x_0}^2 + \tau^* \sum_j \varepsilon_j^T + \tau^* \sum_j \varepsilon_j^P \beta_{x_j}^2$$

y

$$\begin{aligned} B &= n^T \sum_{t \leq \tau^0} (O_t^T - \beta^T t) - n^P \beta_{x_0} \sum_{t \leq \tau^0} [O_t^P - \beta_{x_0} (O_t^T - \beta^T t)] \\ &\quad + \sum_{t \leq \tau^*} \sum_j \varepsilon_j^T [X_{jt}^T - \beta^T t - \gamma^t (t - \tau_0)_I - d_j^T] \\ &\quad - \sum_{t \leq \tau^*} \sum_j \varepsilon_j^P \beta_{x_j} [X_{jt}^P - \beta_{x_j} (X_{jt}^T - \beta^T t - \gamma^T (t - \tau_0)_I - d_j^T) - d_j^P] \end{aligned}$$

$$\therefore \alpha^T | \theta_{\alpha^T} \sim N\left(\frac{B}{A}, \frac{1}{A}\right)$$

18.- Distribución conjunta condicional para  $(a_{\varepsilon^T}, b_{\varepsilon^T})$

$$\begin{aligned} \prod_j^M f(\varepsilon_j^T | a_{\varepsilon^T}, b_{\varepsilon^T}) f(a_{\varepsilon^T}) f(b_{\varepsilon^T}) &= \left[ \prod_{j=1}^M \frac{(b_{\varepsilon^T})^{a_{\varepsilon^T}}}{\Gamma(a_{\varepsilon^T})} (\varepsilon_j^T)^{a_{\varepsilon^T}-1} e^{-b_{\varepsilon^T} \varepsilon_j^T} \right] \\ &\quad \frac{h^g}{\Gamma(h)} (a_{\varepsilon^T})^{g-1} e^{-h a_{\varepsilon^T}} \frac{h^g}{\Gamma(h)} (b_{\varepsilon^T})^{g-1} e^{-h b_{\varepsilon^T}} \end{aligned}$$

Eliminando términos que no contienen a  $a_{\varepsilon^T}$  y  $b_{\varepsilon^T}$ , se obtiene:

$$\frac{(b_{\varepsilon^T})^{M a_{\varepsilon^T}}}{\Gamma(a_{\varepsilon^T})^M} \left( \prod_{j=1}^M \varepsilon_j^T \right)^{a_{\varepsilon^T}-1} (a_{\varepsilon^T} b_{\varepsilon^T})^{g-1} e^{-h(a_{\varepsilon^T} + b_{\varepsilon^T}) - b_{\varepsilon^T} \sum_{j=1}^M \varepsilon_j^T} \quad (\text{C.17})$$

19.- Distribución conjunta condicional para  $(a_{\varepsilon^P}, b_{\varepsilon^P})$

$$\prod_j^M f(\varepsilon_j^P | a_{\varepsilon^P}, b_{\varepsilon^P}) f(a_{\varepsilon^P}) f(b_{\varepsilon^P}) = \left[ \prod_{j=1}^M \frac{(b_{\varepsilon^P})^{a_{\varepsilon^P}}}{\Gamma(a_{\varepsilon^P})} (\varepsilon_j^P)^{a_{\varepsilon^P}-1} e^{-b_{\varepsilon^P} \varepsilon_j^P} \right] \frac{h^g}{\Gamma(h)} (a_{\varepsilon^P})^{g-1} e^{-h a_{\varepsilon^P}} \frac{h^g}{\Gamma(h)} (b_{\varepsilon^P})^{g-1} e^{-h b_{\varepsilon^P}}$$

Eliminando términos que no contienen a  $a_{\varepsilon^P}$  y  $b_{\varepsilon^P}$ , se obtiene:

$$\frac{(b_{\varepsilon^P})^{M a_{\varepsilon^P}}}{\Gamma(a_{\varepsilon^P})^M} \left( \prod_{j=1}^M \varepsilon_j^P \right)^{a_{\varepsilon^P}-1} (a_{\varepsilon^P} b_{\varepsilon^P})^{g-1} e^{-h(a_{\varepsilon^P} + b_{\varepsilon^P}) - b_{\varepsilon^P} \sum_{j=1}^M \varepsilon_j^P} \quad (\text{C.18})$$

## C.2. Aplicación de la frmula de Schur

Formula de Schur:

$$(ZDZ' + I\sigma^2)^{-1} = \frac{I}{\sigma^2} - \frac{1}{\sigma^2} Z(\sigma^2 D^{-1} + Z'Z)^{-1} Z' \quad (\text{C.19})$$

Para nuestro caso.

$$\begin{aligned} \frac{I}{\sigma^2} - \frac{w_i}{\sigma^2} N^{-1} \frac{w_i'}{\sigma^2} &= \frac{I}{\sigma^2} - \frac{1}{\sigma^2} w_i \left( \frac{N^{-1}}{\sigma^2} \right) w_i' \\ &= \frac{I}{\sigma^2} - \frac{1}{\sigma^2} w_i \left( w_i' \frac{1}{\sigma^2} w_i + \Omega^{-1} \right)^{-1} \frac{w_i}{\sigma^2} \\ &= \frac{I}{\sigma^2} - \frac{1}{\sigma^2} w_i \left( w_i' w_i + \sigma^2 \Omega^{-1} \right)^{-1} w_i \\ &= (w_i \Omega w_i' + I \sigma^2)^{-1} \end{aligned}$$

La ultima desigualdad se obtuvo aplicando la regla de Schur (ecuación C.19)

## Bibliografía

- [1] J. Arias Restrepo, T. Rengifo Martínez, and M. Jaramillo Carmona. *Manual Técnico. Buenas Prácticas Agrícolas en la Producción del Frijol*. Gerencia Seguridad Alimentaria y Nutricional MANA (FAO), 2007.
- [2] S. Benacchio. Algunas exigencias agroecológicas en 58 especies de cultivo con potencial de producción en el trópico americano. *Fonaiap-CPNIA*, page 202, 1982.
- [3] M. Bennik. Eat beans for good health. *Food Science and Human Nutrition*. Michigan State University, <http://www.michiganbean.org/research.html>.
- [4] G. Casella and R. Berger. *Statistical Inference*. Thomson Learning, Second edition, 2001.
- [5] M. Claussen, L. Mysak, and A. Weaver. Earth system models of intermediate complexity: Closing the gap in the spectrum of climate system models. *Climate Dynamics*, 18(7):579–586, 2002.
- [6] M. Daniels and J. Hogan. *Missing Data en Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Chapman & Hall, 2007.
- [7] D. Debouck and R. Hidalgo. Morfología de la planta de frijol común. *Frijol: Investigación y producción*, pages 7–41, 1985.
- [8] R. Furrer, S. Sain, D. Nychka, and G. Meehl. Multivariate bayesian analysis of atmosphere-ocean general circulation models. *Environmental ecological Statistic*, 14:249–266, 2007.
- [9] F. Giorgi and L. Mearns. Calculation of average uncertainty range and reliability of regional climate changes from AOGCM simulations via the

- “Reliability ensemble averaging” (REA) method. *Journal of Climate*, 15:1141–1158, 2002.
- [10] A. Greene, L. Goddard, and U. Lall. Probabilistic multimodel regional temperature change projections. *Journal of Climate*, 19:4326–4343, 2006.
- [11] A. Hall. Breeding for heat tolerance. *Plant Breeding Reviews. American Society of Horticultural Science, Crop Society of America, Society of American Foresters and National Council of Commercial Plant Breeders*, pages 131–168, 1992.
- [12] IPCC. Grupo intergubernamental de expertos sobre el cambio climático. 2007.
- [13] ITESM. Diagnóstico nacional y estatal sobre problemática y perspectivas de la producción de frijol en los estados de Sinaloa y Zacatecas. *FAO SAGARPA*, 81, 2004.
- [14] D. Lobell and M. Burke. *Climate Changes and Food Security: Adapting Agriculture to a Warmer World*. Springer Verlag, 2009.
- [15] D. Lobell and C. Field. Global scale climate-crop yield relationships and the impacts of recent warming. *Environmental Research Letters*, 2007.
- [16] D. Lobell, W. Schlenker, and J. Costa-Roberts. Climate trends and global crop production. *Science express*, 16-May-2011.
- [17] M. Mayorga and E. Muños. Las técnicas de datos de panel: Una guía para su uso e interpretación. *Banco Central de Costa Rica*, Sep-2000.
- [18] V. Ortiz. El frijol en el estado de zacatecas. *Gobierno del Estado de Zacatecas*, page 181, 1998.
- [19] J. Reilly, A. Richard, and B. Hurd. A review of impacts to U.S. agricultural resources. *Documento preparado para el Pew Center on Global Climate Change*, 1999.

- [20] J. Reilly and D. Schimmelfenning. Irreversibilidad, uncertainty, and learnig: Portraits of adaptation to long term climate change. *Climatic Change*, 45:253–278, 2000.
- [21] R. Reyes, B. Padilla, V. Pérez, and J. López. Historia, naturaleza y cualidades alimentarias del frijol. *Revista Investigación Científica*, 4, 2008.
- [22] C. Robert. *The Bayesian Choice*. Springer, Second edition, 2007.
- [23] D. Rowlands. Broad range of 2050 warming from an observationally constrained large climate model ensemble. *Nature Geoscience*, 2012.
- [24] W. Sacks, D. Deryng, J. Foley J., and N. Ramankutty. Global ecology and biogeography. *in press*, 2010.
- [25] C. Salinas, H. Rojas, M. Sosa, and H. Péres. Composición de antocianinas en variedades de frijol negro (*Phaseolus vulgaris* L.) cultivadas en México. *Agrociencia*, 39 (4):385–394, 2005.
- [26] M. Sáenz. El problema de las medidas repetidas. análisis longitudinal en epidemiología. *Grupo de Investigación en Estadística, Economía Aplicada y Salud (GRECS). Departamento de Economía*, Jun-2011.
- [27] R. Smith, C. Tebaldi, D. Nycha, and L. Mears. Bayes modeling and Uncertainty in ensembles of climates models. *Journal of the American Statistical Association*, 2008.
- [28] C. Tebaldi and D. Lobell. Towards probabilistic projections of climate impacts on global crop yields. *Geophysical Research Letters*, 35, 2008.
- [29] C. Tebaldi and B. Sansó. Joint projections of temperature and precipitation change from multiple climate models: A hierarchical bayesian approach. *Royal Statistical Society*, 172:83–106, 2008.
- [30] C. Tebaldi, R. Smith, D. Nychka, and L. Mearns. Quantifying uncertainty in projections of regional climate change: A bayesian approach to the analysis of multimodel ensembles. *Journal of Climate*, 18:1524–1540, 2005.

- [31] J. White. Conceptos básicos de fisiología del frijol. *CIAT*, pages 43–60, 1985.
- [32] A. Wild. Condiciones del suelo y desarrollo de las plantas según russell. *Mundi Prensas*, 1992.

**ANÁLISIS BAYESIANO DE UN MODELO LINEAL MIXTO PARA  
MEDIR EL IMPACTO DEL CAMBIO CLIMÁTICO EN EL  
RENDIMIENTO DEL FRIJOL PARA EL AÑO 2030 A NIVEL  
MUNDIAL**

Velcy Rosario Palomino Lescano

[velcy.palomino@upr.edu](mailto:velcy.palomino@upr.edu)

Departamento de Ciencias Matemáticas

Consejero: Dámaris Santana Morant, Ph.D.

Grado: Maestría en Ciencias

Fecha de Graduación: Julio 2012