

Verificación del Grado de Habilidad del pronóstico del modelo WRF para las principales ciudades del país

en el acumulado diario de las variables de precipitación y temperaturas extremas

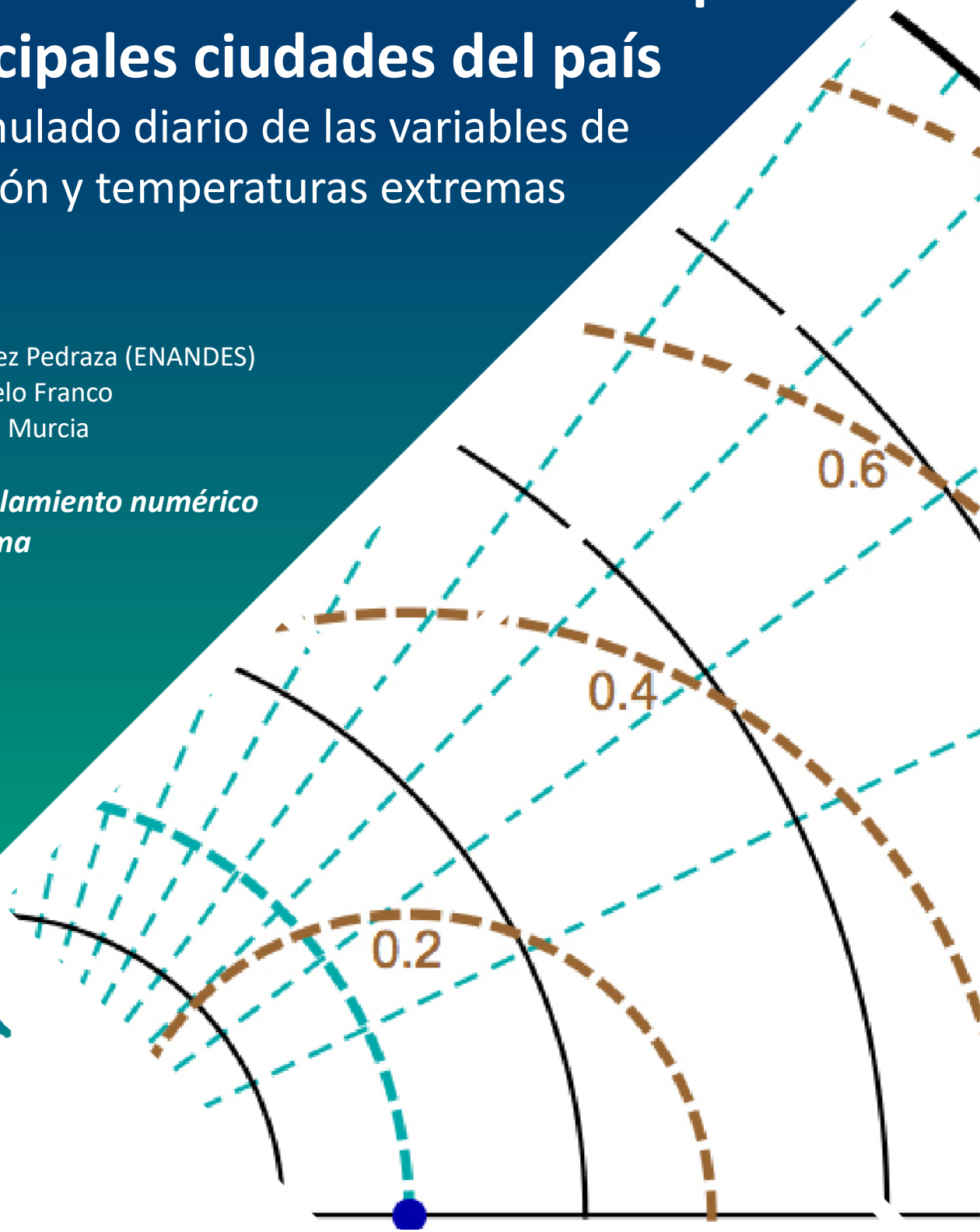
Autores

Alexander Martínez Pedraza (ENANDES)

Jeimmy Yanely Melo Franco

José Franklyn Ruiz Murcia

*Grupo de modelamiento numérico
de tiempo y clima*



Verificación del grado de habilidad de pronóstico del modelo WRF para las principales ciudades del país en el acumulado diario de las variables precipitación y temperaturas extremas

RESUMEN

Esta Nota Técnica analiza el desempeño del pronóstico del modelo WRF implementado en el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) durante el periodo comprendido entre el 1 de marzo de 2020 y 30 de junio de 2023, enfocándose en los agregados diarios de las variables precipitación y temperatura del aire máxima y mínima, para las ubicaciones asociadas a los principales aeropuertos del país. Además de las métricas habituales, se hizo un análisis inferencial sobre la distribución de los errores de pronóstico para establecer la presencia de error sistemático y sus características. Los resultados muestran las variaciones espaciales y temporales de las magnitudes del error, para la precipitación se observó que los errores bajos se asocian con épocas de menos lluvias o con lugares con régimen de precipitación de acumulados bajos, mientras que los valores altos de RMSE se asocian a lugares con acumulados de precipitación altos. En cuanto al error sistemático, se encontró que en el error de pronóstico de precipitación predominan las distribuciones con sesgo positivo y evidencian tendencia a sobreestimar, mientras que los errores de pronóstico para las temperaturas se ajustan mejor a distribuciones Gaussianas y evidencian tendencia a subestimar.

ABSTRACT

The Technical Note analyzes the performance of the WRF model implemented at the Institute of Hydrology, Meteorology, and Environmental Studies (IDEAM) during the period from March 1 2020, to June 30 2023, It focuses on the daily aggregates of precipitation and maximum and minimum air temperatures for locations associated with the country's main airports. In addition to the usual metrics, inferential analysis was conducted on the distribution of forecast errors to establish the presence of systematic error and its characteristics. The results show spatial and temporal variations in error magnitudes. For precipitation, it was observed that low errors are associated with periods of less rainfall or locations with low cumulative precipitation regimes, while high RMSE values are associated with places with high precipitation accumulations. Regarding systematic error, it was found that precipitation forecast errors are predominantly associated with positive bias distributions, indicating a tendency to overestimate, while temperature forecast errors better fit Gaussian distributions and show a tendency to underestimate.

Contenido

1. Introducción	4
2. Datos y Metodología	5
3. Resultados y discusión.....	13
Conclusiones.....	22
Recomendaciones.....	25
Referencias Bibliográficas	26
Anexos.....	28
Anexo 1. Identificación y ubicación geográfica de las estaciones meteorológicas utilizadas en el análisis .	28
Anexo 2. Análisis precipitación	30
<i>Anexo 2.1. Métricas de verificación de pronóstico dicotómico precipitación</i>	<i>30</i>
<i>Anexo 2.2. Métricas de verificación de pronóstico multicategoría precipitación</i>	<i>34</i>
<i>Anexo 2.3. Métricas de verificación de pronóstico de precipitación (variable continua).....</i>	<i>35</i>
<i>Anexo 2.4. Análisis de supuestos de error aleatorio de pronóstico de precipitación</i>	<i>38</i>
Anexo 3. Análisis temperatura máxima	43
<i>Anexo 3.1. Métricas de verificación de pronóstico multicategoría temperatura máxima.....</i>	<i>43</i>
<i>Anexo 3.2. Métricas de verificación de pronóstico de temperatura máxima (variable continua)</i>	<i>44</i>
<i>Anexo 3.3. Análisis de supuestos de error aleatorio de pronóstico de temperatura máxima.....</i>	<i>47</i>
Anexo 4. Análisis temperatura mínima	52
<i>Anexo 4.1. Métricas de verificación de pronóstico multicategoría temperatura mínima.....</i>	<i>52</i>
<i>Anexo 4.2. Métricas de verificación de pronóstico de temperatura mínima (variable continua)</i>	<i>53</i>
<i>Anexo 4.3. Análisis de supuestos de error aleatorio de pronóstico de temperatura mínima</i>	<i>56</i>

1. Introducción

El pronóstico meteorológico es el proceso de estimar los valores futuros de las variables atmosféricas, tales como la temperatura, la presión, la humedad, la nubosidad, la precipitación, etc., que influyen en una región específica (Guerrero-Higuera et al., 2012). Para realizar una predicción meteorológica efectiva en Colombia, se requiere de un modelo que pueda representar las condiciones topográficas complejas, ya que el país está atravesado por tres ramas de la cordillera de los Andes. Uno de los modelos que tiene esta capacidad es el WRF “Weather Research and Forecasting” (Skamarock et al., 2021), que incorpora los últimos avances en el modelado de las variables atmosféricas, y que se considera una herramienta eficaz para simular los diferentes procesos atmosféricos con alta resolución. Sin embargo, el modelo WRF no está exento de errores, y algunos estudios han mostrado que puede sobreestimar o subestimar en algunas regiones los valores de los parámetros meteorológicos. Por estas razones, es indispensable realizar una verificación exhaustiva de las simulaciones del modelo WRF, que permita evaluar su precisión y detectar sus deficiencias, para así poder realizar las correcciones necesarias que mejoren la calidad del pronóstico (García, 2014). La verificación consiste en contrastar los pronósticos de eventos históricos (también llamados pronósticos retrospectivos o hindcasts) con las observaciones correspondientes, permitiendo medir la exactitud (el grado de concordancia entre los pronósticos y las observaciones) y la asociación (la intensidad de la relación entre los pronósticos y las observaciones) del sistema de pronóstico. Para la verificación es conveniente utilizar distintas métricas, ya que una sola no puede capturar todos los aspectos de la calidad del pronóstico (Manubens et al., 2018).

En el caso particular del IDEAM, el modelo WRF se ha implementado desde el año 2007 con fines de predicción del tiempo para Colombia y la ciudad de Bogotá con una resolución espacial de (20kmx20km) y (7kmx7km) respectivamente, siendo una herramienta muy importante para los meteorólogos a la hora de tomar decisiones en la emisión del pronóstico oficial (Arango & Ruiz, 2012). Una primera verificación del modelo fue realizada por Ruiz et al. (2013) donde se utilizaron 27 estaciones del IDEAM de los principales aeropuertos del país.

Con este contexto, el objetivo de esta Nota Técnica es presentar una verificación actualizada del grado de habilidad del pronóstico del modelo WRF implementado en el IDEAM para las principales ciudades del país en acumulado diario de las variables de precipitación y temperaturas extremas. Para esto se complementó la metodología expuesta en la Guía sobre la verificación de pronósticos climáticos estacionales operativos (WMO, 2018) con un análisis inferencial para determinar si los errores cumplen los supuestos de error aleatorio o si por el contrario hay evidencia de error sistemático en los pronósticos.

2. Datos y Metodología

2.1. Cobertura temporal y geográfica

La verificación abarca el periodo comprendido entre el 1 de marzo de 2020 y 30 de junio de 2023. Los análisis consideran la totalidad del territorio colombiano, a través de 47 estaciones meteorológicas asociadas a los principales aeropuertos del país, la distribución espacial se presenta en la Figura 1. La información de identificación y ubicación de cada estación se presenta en el Anexo 1.

2.2. Datos utilizados

El modelo WRF (Ruiz M & Melo F, 2020) ofrece pronósticos para 40 puntos geográficos relacionados con los aeropuertos más importantes del país. Estos pronósticos se pueden consultar en el sitio web: http://bart.ideam.gov.co/wrfideam/new_modelo/WRF00COLOMBIA/meteogramas/COLOMBIA/historicos/. Los datos tienen una resolución temporal de una hora y un horizonte de pronóstico de 7 días. A partir de estos datos fueron calculados los agregados diarios de precipitación, y las temperaturas máximas y mínimas del aire. La verificación se hizo para 6 días de pronóstico, lo que generó 6 bases de datos distintas.

Para la verificación de los pronósticos del modelo WRF se utilizaron los datos de 47 estaciones meteorológicas del IDEAM de los principales aeropuertos del país, cuya ubicación se muestra en la Figura 1. Para el análisis se emplearon agregados diarios de las variables precipitación, temperatura mínima y temperatura máxima.

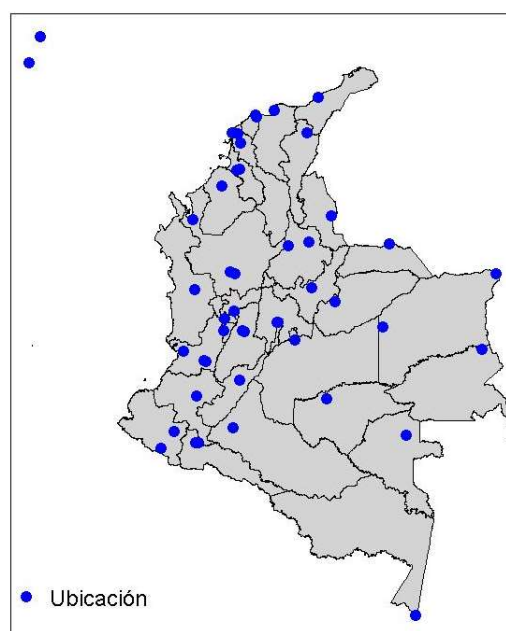


Figura 1. Ubicación de 47 estaciones meteorológicas utilizadas para la verificación de los pronósticos del modelo WRF.

2.3. Metodología

Todo el procesamiento de datos fue desarrollado a través del software libre R (R Core Team, 2022). Para la verificación se siguieron los lineamientos propuestos por el Grupo de trabajo conjunto sobre investigación de verificación de pronósticos (JWGFVR, 2017), asociado al Programa Mundial de Investigaciones Meteorológicas (WWRP, por sus siglas en inglés) de la Organización Meteorológica Mundial (OMM), expuestos formalmente en la Guía sobre la verificación de pronósticos climáticos estacionales operativos (WMO, 2018). Adicionalmente se realizó un análisis inferencial sobre el error de pronóstico para identificar y caracterizar la existencia de error sistemático.

Para poder identificar las diferencias asociadas al ciclo anual, la verificación se realizó para cada uno de los meses del año ya que se evidenció que el análisis sobre las series completas esconde características que solo se observan mediante la desagregación propuesta.

2.3.1. Verificación

En el séptimo Workshop Internacional de Métodos de Verificación realizado en 2017 y llevado a cabo en Berlín (Alemania), se tuvo como objetivo discutir y promover todos los aspectos de la investigación y la práctica de la metodología de verificación, aplicada a los pronósticos y avisos meteorológicos, las predicciones climáticas y sus aplicaciones; resultados que son presentados y descritos en el sitio web de JWGFVR (2017); de los cuales para este trabajo se siguió la propuesta de análisis seleccionando las métricas más usuales y que tienen mejor calificación según las recomendaciones para la verificación expuesta en WMO (2008).

Existen tres familias de métricas de verificación dependiendo del tipo de pronósticos que se analizan:

1. Pronósticos dicotómicos (El evento sucede o no)
2. Pronósticos multicategoría
3. Pronósticos de variables continuas

Métodos para pronósticos dicotómicos

Los pronósticos de precipitación se pueden verificar desde diferentes enfoques, en primera instancia, como un pronóstico dicotómico. En este sólo se considera la ocurrencia o no del evento, sin evaluar los acumulados de precipitación. En este trabajo se considera como evento de precipitación todo aquel que registre acumulados diarios mayores o iguales a 1 mm como se plantea en el trabajo de Venegas (2020).

Para verificar este tipo de pronóstico, se define una tabla de contingencia que muestra la frecuencia de pronósticos y ocurrencias de "sí" y "no". Las cuatro combinaciones de pronósticos (sí o no) y observaciones (sí o no), denominadas distribución conjunta (Como se muestra en la Tabla 1), son:

- Acierto - evento previsto que ocurrirá, y ocurrió
- Error: se pronosticó que el evento no ocurriría, pero ocurrió

- Falsa alarma: se pronosticó que el evento ocurriría, pero no ocurrió.
- Negativo correcto: se pronosticó que el evento no ocurriría y no ocurrió.

El número total de ocurrencias y no ocurrencias observadas y pronosticadas se da en los lados inferior y derecho de la tabla de contingencia y se denomina distribución marginal.

Tabla 1. Resultados posibles para pronósticos categóricos de un evento binario (Jolliffe & Stephenson, 2012)

Evento pronosticado	Evento observado		Total
	Sí	No	
Sí	Aciertos	Falsas alarmas	Pronóstico sí
No	Errores	Negativos correctos	Pronóstico no
Total	Eventos ocurridos	Eventos no ocurridos	Total de pronósticos

Para realizar la verificación de los pronósticos dicotómicos de precipitación se emplearon las siguientes métricas:

- a) Exactitud o proporción correcta

Está definida como:

$$Exactitud = \frac{Aciertos + Negativos correctos}{Total de pronósticos}$$

Responde a la pregunta ¿qué fracción de los pronósticos fueron correctos?, los posibles valores de esta métrica varían entre 0 y 1, donde 1 es la puntuación perfecta cuando todos los eventos son pronosticados de forma correcta.

- b) Bias score

Está definida como:

$$Bias\ score = \frac{Aciertos + Falsas\ alarmas}{Aciertos + Errores}$$

Mide la relación entre la frecuencia de eventos pronosticados y la frecuencia de eventos observados. Indica si el sistema de pronóstico tiene tendencia a subestimar (Bias score < 1) o a sobrepronosticar (Bias score > 1) eventos.

- c) Probabilidad de detección (POD, por sus siglas en inglés) o tasa de aciertos

Está definida como:

$$POD = \frac{Aciertos}{Aciertos + Errores}$$

Responde a la pregunta ¿Qué fracción de los eventos "sí" observados se pronosticaron correctamente?, los posibles valores de esta métrica varían entre 0 y 1. La puntuación perfecta es 1.

d) Tasa de falsas alarmas (FAR, por sus siglas en inglés)

Se define como

$$FAR = \frac{\text{Falsas alarmas}}{\text{Aciertos} + \text{Falsas alarmas}}$$

Responde a la pregunta ¿Qué fracción de los eventos "sí" pronosticados en realidad no ocurrieron (es decir, fueron falsas alarmas) ?, los posibles valores de esta métrica varían entre 0 y 1. La puntuación perfecta es 0.

Método para pronósticos multicategoría

Otro enfoque para verificar los pronósticos es establecer categorías de los posibles valores que puede tomar la variable que se está analizando. En este sentido se define una tabla de contingencia $k \times k$ con $k \geq 3$, como se muestra en la Tabla 2.

Tabla 2. Tabla de contingencia multicategoría

		Categoría observada				Total
		<i>i, j</i>	1	2	...	k
Categoría pronosticada	1	$n(F_1, O_1)$	$n(F_1, O_2)$	...	$n(F_1, O_k)$	$N(F_1)$
	2	$n(F_2, O_1)$	$n(F_2, O_2)$	...	$n(F_2, O_k)$	$N(F_2)$
	...	...	...	...	...	...
	k	$n(F_k, O_1)$	$n(F_k, O_2)$	...	$n(F_k, O_k)$	$N(F_k)$
Total		$N(O_1)$	$N(O_2)$	...	$N(O_k)$	N

Para la precipitación se trabajó con las categorías 1) Muy ligera, 2) Ligera, 3) Moderada, 4) Fuerte y 5) Muy fuerte, establecidas en el trabajo de Barrios (2019), para cada una de las Zonas Pluviométricamente homogéneas de Guzmán et al. (2014). Por otro lado, para las temperaturas, las categorías se estimaron teniendo en cuenta la propuesta de intervalos de normalidad de temperatura con base en percentiles de Martínez et al. (2018).

Para la verificación de los pronósticos multicategoría se empleó la métrica Exactitud que responde a la pregunta ¿qué fracción de los pronósticos estaban en la categoría correcta?, esta se define como:

$$\text{Exactitud multicategoría} = \frac{1}{N} \sum_{i=1}^k n(F_i, O_i)$$

Los posibles valores de esta métrica varían entre 0 y 1 donde la puntuación perfecta es 1 cuando todas las categorías pronosticadas corresponden a la categoría del evento registrado.

Métodos para pronósticos de variables continuas

En este enfoque se verifican los pronósticos midiendo cómo los valores de los pronósticos difieren de los valores de las observaciones. Aquí se nota F_i como el valor de pronóstico en el tiempo i , O_i el valor registrado en el tiempo i y N el total de pronósticos. Las métricas utilizadas se describen a continuación.

a) Error medio (ME, por sus siglas en inglés) o Bias aditivo

Permite realizar una comparación entre los valores observados y los pronosticados, indicando si estos últimos presentan una tendencia a subestimar o sobreestimar las variables meteorológicas simuladas por los modelos (Venegas, 2020). Específicamente estima el promedio del error de pronóstico, un modelo perfecto arrojaría un valor $ME = 0$, sin embargo, es posible obtener una puntuación perfecta para un mal pronóstico si existen errores compensatorios. $ME > 0$ indica tendencia de sobreestimación y $ME < 0$ indica tendencia de subestimación.

$$ME = \frac{1}{N} \sum_{i=1}^N (F_i - O_i)$$

b) Raíz Error cuadrático medio (RMSE, por sus siglas en inglés)

Mide el error "promedio", ponderado según el cuadrado del error. Un modelo perfecto tendría un valor de $RMSE = 0$.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (F_i - O_i)^2}$$

c) Coeficiente de correlación de Spearman

Esta medida no paramétrica mide la relación lineal entre dos variables, "este coeficiente, propuesto por Spearman en 1904, es una versión del coeficiente de correlación de Pearson, pero calculado sobre una sucesión transformada de rangos en las dos variables" (Corzo, 2005) y puede ser expresado de la siguiente forma:

$$r_s = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n d_i^2$$

Donde d_i es la diferencia entre los rangos de las dos variables comparadas. Para la interpretación es suficiente tener en cuenta que los valores de r_s cercanos a uno (1) indican asociación positiva o directa, los valores cercanos a menos uno (-1) indican asociación negativa o inversa y los valores cercanos a cero (0) revelan que no hay ningún tipo de asociación.

2.3.2. Análisis inferencial de los errores de pronóstico

En general en los trabajos sobre verificación de modelos los esfuerzos se centran en cuantificar el error mediante las métricas descritas en la sección anterior, estas permiten evaluar la exactitud en los pronósticos categóricos e identificar qué modelos implican menos error. En este trabajo, se realizó un análisis complementario sobre los errores, tomando como base la Guía para la expresión de la incertidumbre en la medición (JCGM, 2008) donde se plantea que una medición presenta imperfecciones que dan lugar a un error en el resultado de la medición. Tradicionalmente, se considera que un error tiene dos componentes, a saber, un componente aleatorio y un componente sistemático. Aquí la medición corresponde al pronóstico, siguiendo los lineamientos de la guía JCGM (2008) se plantean las siguientes definiciones:

$$error = Valor observado - Valor pronosticado$$

$$error = error aleatorio + error sistemático$$

Donde,

- **error aleatorio** es el componente del error que en una sucesión de pronósticos varía de forma aleatoria.
- **error sistemático** es el componente del error que en una sucesión de pronósticos varía de forma predecible y son constantes en el tiempo. Idealmente se busca que este tipo de error sea nulo o que se pueda reducir lo máximo posible.

En este contexto, dado que los pronósticos analizados en este trabajo son sucesiones temporales, se plantea un enfoque de análisis sobre los errores de pronóstico en una serie temporal, asumiendo que los errores son netamente aleatorios y siguen una distribución de ruido blanco gaussiano, es decir, que los errores se distribuyen normal con media cero, varianza constante ($e_t \sim N(0, \sigma^2)$) y $Cor(e_i, e_j) = 0$ para todo $i \neq j$ (Cowpertwait & Metcalfe, 2009).

Por lo tanto, para verificar los supuestos se utilizan los estadísticos de prueba que se describen a continuación:

Prueba de normalidad de Jarque-Bera

Esta prueba se realiza para verificar si los errores de pronóstico siguen una distribución normal, es decir, se contrasta la hipótesis

H_0 : Los errores de pronóstico se ajustan a una distribución Normal

El estadístico de prueba Jarque-Bera se describe de forma detallada en Jarque & Bera (1980). Bajo H_0 el estadístico se distribuye asintóticamente como una distribución chi cuadrado con dos grados de libertad.

De forma complementaria a la prueba Jarque – Bera se calculó el coeficiente de asimetría de los errores de tal manera que:

- Si se rechaza H_0 y coeficiente de asimetría es mayor que cero da indicios de sesgo positivo en la distribución.
- Si se rechaza H_0 y coeficiente de asimetría es menor que cero da indicios de sesgo negativo en la distribución.

Prueba t-Student

Para verificar si la media de los errores de pronóstico es estadísticamente cero o por el contrario es mayor o menor, asumiendo el supuesto de distribución normal, se contrastan las siguientes pruebas de hipótesis a una y dos vías de la siguiente manera:

- Dos vías: $H_0: \bar{x}_{error} = 0$
- Una vía: $H_0: \bar{x}_{error} < 0$, da indicios de que el pronóstico tiende a sobreestimar (Presencia de error sistemático).
- Una vía: $H_0: \bar{x}_{error} > 0$, da indicios de que el pronóstico tendencia a subestimar (Presencia de error sistemático).

En este caso, el estadístico de prueba es

$$T_c = \frac{\bar{x}_{error}}{s/\sqrt{n}}$$

Donde s es la desviación estándar de los errores y n es el número de pronósticos. Bajo H_0 cierta, T_c tiene una distribución t - Studen con $n-1$ grados de libertad $T_c \sim t_{(n-1)}$.

Prueba de rangos signados de Wilcoxon

Es una prueba no paramétrica, utilizada en lugar de la t de Student cuando los datos no cumplen el requisito de normalidad, como en el caso de la precipitación. En este análisis se utiliza para comprobar si la tendencia central, la mediana, de los errores de pronóstico es cero. Aquí también se contrastaron tres hipótesis de la siguiente manera:

- Dos vías: $H_0: Mediana_{error} = 0$
- Una vía: $H_0: Mediana_{error} < 0$, da indicios de que el pronóstico tiende a sobreestimar (Presencia de error sistemático).

- Una vía: $H_0: Mediana_{error} > 0$, da indicios de que el pronóstico tendencia a subestimar (Presencia de error sistemático).

Para el contraste de la hipótesis se asignan rangos desde 1 hasta n atendiendo únicamente al valor absoluto del error en el tiempo i (e_i) y se suman los rangos correspondientes a los errores positivos y a los errores negativos por separado. Si la hipótesis nula es cierta, es de esperar que los rangos se distribuyan aleatoriamente entre las diferencias positivas y negativas y, por tanto, que ambas sumas de rangos sean aproximadamente iguales. El estadístico de prueba, T , es la menor de las dos sumas de rangos. Cuando $n > 15$ la distribución muestral de T bajo el supuesto de que H_0 es cierta se aproxima a una normal de parámetros:

$$\mu_T = \frac{n(n+1)}{4} \quad \sigma_T^2 = \frac{n(n+1)(2n+1)}{24}$$

El estadístico de prueba es el valor Z :

$$Z = \frac{T - \mu_T}{\sigma_T}$$

Que se distribuye según una normal estándar.

Prueba Breusch-Pagan estudentizada

Esta prueba se usa para verificar el supuesto de homocedasticidad (varianza constante), fue propuesta por Breusch & Pagan (1979) y consiste en ajustar un modelo de regresión lineal con variable respuesta dada por residuales del modelo original al cuadrado e_i^2 y como covariables las variables del modelo original.

Para este trabajo el modelo está dado de la siguiente forma:

$$e_i^2 = \delta_0 + \delta_1 Valor\ pronóstico_i$$

Si se concluye que $\delta_1 = 0$, significa que los errores no son función del pronóstico. Sin embargo, Koenker (1981) propone un estadístico de prueba Breusch-Pagan ajustado ya que los niveles de significancia sugeridos por el estadístico original son correctos sólo bajo condiciones gaussianas en el error.

La hipótesis que se contrasta es $H_0: Varianza\ de\ los\ errores\ es\ constante$, el estadístico en esta prueba está dado por $n \times R^2$ y bajo la hipótesis nula verdadera, el estadístico tiene distribución χ_1^2 .

3. Resultados y discusión

Como se mencionó en la metodología, los análisis se realizaron para las series completas y para las series desagregadas según los meses del año. Sin embargo, se evidenció que para las series completas los resultados no son eficientes para demostrar la calidad de los pronósticos del modelo WRF, ya que la variabilidad asociada al ciclo anual, sobre todo en la precipitación, introduce diferentes comportamientos que difícilmente se pueden caracterizar. Por otro lado, al realizar los análisis por cada mes se pudo identificar de forma más clara fortalezas y debilidades en la habilidad del pronóstico. Por lo tanto, los resultados se presentan con el enfoque de desagregación por meses.

Adicionalmente, como los análisis se realizaron para 47 puntos geográficos con 6 días de pronóstico, se evidenció que las variaciones significativas se dan en la comparación espacial, mientras que la variación entre días no es significativa en términos globales del análisis de este trabajo. En este contexto, los resultados que se describen a continuación corresponden a un horizonte de pronóstico de un día según los meses del año y se presentan gráficamente en los Anexos 2, 3 y 4.

3.1. Precipitación

Los resultados descritos a continuación se soportan gráficamente en el Anexo 2. Análisis precipitación.

Verificación de pronósticos dicotómicos (Lluvia / No lluvia)

Exactitud

Desde el punto de vista de la exactitud, es decir, ¿qué fracción de los pronósticos fueron correctos? Si se tiene en cuenta la climatología que se presenta en IDEAM (2017), se puede observar que el modelo pronostica correctamente el evento de precipitación por encima del 80% en lugares donde el promedio mensual multianual de la precipitación acumulada es menor que 100 mm, especialmente en el periodo diciembre – marzo, en la región Caribe y región Orinoquía, además en julio – septiembre en Neiva. En ese mismo porcentaje de pronóstico, también se observa buen desempeño en los puntos de verificación asociados a Quibdó (abril - enero) y Buenaventura (junio - noviembre).

El modelo muestra una buena precisión, es decir, una alta proporción de pronósticos acertados, en comparación con la climatología reportada por IDEAM (2017). Se destaca el desempeño del modelo en zonas con un promedio mensual multianual de precipitación acumulada menor a 100 mm, especialmente en el Caribe y la Orinoquía entre diciembre y marzo, y en Neiva entre julio y septiembre, donde supera el 80% de acierto. También se obtiene un buen resultado en los puntos de verificación correspondientes a Quibdó (abril - enero) y Buenaventura (junio - noviembre), con el mismo porcentaje de precisión.

En el periodo mayo – julio es cuando el porcentaje de pronóstico correcto predominante está por encima del 60% hasta el 80%. Las regiones mejor representadas por este porcentaje son la región Amazonía e Insular del país.

Los porcentajes de pronóstico correcto menores a 60% predominan en el periodo septiembre – noviembre en la región Andina coincidiendo con la segunda temporada de lluvias y también en el periodo abril- noviembre en el sur de la región Caribe asociada a las cuencas del río Sinú y San Jorge, coincidiendo con su temporada de lluvias. Específicamente Bogotá y Cali tienen porcentaje de pronóstico correcto por debajo del 60% todo el año seguidos de Ipiales y Armenia con 11 y 10 meses respectivamente

Bias score

En la verificación de los eventos pronosticados Sí, frente a los que realmente ocurrieron, en el 68% de los casos analizados se evidencia tendencia a sobreestimar el número de eventos, principalmente en los meses de octubre y diciembre, en la región Andina y Sur de la región Caribe.

El modelo tiende a subestimar el número de eventos de precipitación en el 12.4% de los casos, especialmente en los piedemontes asociados a Villavicencio, Mocoa y Yopal, durante 12, 11 y 9 meses respectivamente. Adicionalmente en Valledupar se evidencia subestimación en 9 meses.

Los puntos de verificación donde el número de eventos Sí fueron bien pronosticados corresponden al 16.1% principalmente en Buenaventura y Florencia durante 9 meses, Arauca y San José del Guaviare durante 8 meses.

Tasa de aciertos

¿Qué proporción de eventos sí, fueron pronosticados correctamente?

La verificación donde los eventos Sí fueron pronosticados correctamente con una proporción mayor al 80% corresponden al 57.4% de los casos, asociados principalmente a la región Andina. En el mes de mayo la tasa de aciertos mayor a ese porcentaje abarca gran parte del país asociado a 34 puntos de verificación.

Las tasas de aciertos menores a 40% corresponde al 9.2% de los casos analizados, se presentan principalmente en Valledupar y Neiva con prevalencia de 10 y 8 meses respectivamente.

Tasa de falsas alarmas

Las tasas de falsa alarma mayores al 60% corresponden al 69% de los casos de verificación, principalmente están asociados a la región Caribe y al periodo de diciembre a mayo.

Las tasas de falsa alarma menores al 40% corresponden al 36.3% de los casos de verificación, principalmente se presentan en los puntos asociados a la región Pacífico y región Amazonía, durante todo el año en Buenaventura, Quibdó y Mocoa. Los meses donde tienen menor incidencia las falsas alarmas son los comprendidos entre mayo y agosto.

Verificación de pronósticos multicategoría

Exactitud

Con base en la categorización de los niveles de precipitación propuestos por Barrios (2019), los pronósticos con mayor exactitud, es decir, aquellos donde la métrica es mayor al 60%, corresponden al 13.8% de los casos de verificación, se dan principalmente en Buenaventura (durante 11 meses), Santa Marta (durante 8 meses), Riohacha y Valledupar (durante 7 meses). Con mayor incidencia en entre los meses de diciembre y febrero.

La exactitud menor al 40%, observada en el 40.2% de los casos, se presenta principalmente en Popayán (durante 12 meses), Bogotá y Tunja (durante 11 meses) y Armenia (durante 10 meses). Esta baja proporción de exactitud se da especialmente en los meses de marzo y abril.

Verificación de pronósticos continuos

Error medio

Los casos con error medio en el intervalo $[-1, 1]$, es decir, con bajo error medio representan el 26.8% del total y se presentan principalmente en Corozal y Riohacha (durante 10 meses), Bucaramanga, Cali, Cúcuta y Sincelejo (durante 8 meses). Los meses donde se registra los valores de error medio cercanos a cero son enero y febrero.

El error medio con valores menores a -1 da indicios de subestimación de la precipitación, estos abarcan el 40% de los casos de verificación y se presentan especialmente en las regiones Orinoquía, Amazonía y Pacífico. En la región Andina en Neiva.

El error medio con valores mayores a 1 da indicios de sobreestimación de la precipitación, estos se registraron en el 32.2% de los casos de verificación y se dan principalmente en la región Andina.

Raíz del Error Cuadrático Medio (RMSE)

Esta es la métrica que representa mejor la bondad de ajuste de un modelo. Los mejores ajustes del modelo WRF con el 44.3% de valores bajos de RMSE tiene mayor incidencia en el periodo diciembre – febrero especialmente en la región Caribe y región Insular. Por otra parte, los puntos donde estos valores bajos tienen mayor prevalencia son Cali y Cúcuta (durante 12 meses), Bucaramanga, Manizales, Riohacha y Santa Marta (durante 10 meses).

El 35% de los valores altos de RMSE se presentan en abril y el periodo octubre – noviembre. Especialmente estos valores se distribuyen principalmente en la región Pacífico, Buenaventura, Popayán, Quibdó (durante 12 meses) y Apartadó (durante 10 meses), Además en los piedemontes asociados a Mocoa (durante 12 meses), Florencia (durante 10 meses) y Villavicencio (durante 9 meses).

Correlación pronóstico Vs. Observaciones

No hay patrones de distribución específicos. Se destaca que en el 48.6% de los casos de verificación predominan correlaciones entre 0.2 y 0.4.

Análisis de supuestos del error aleatorio

Prueba Jarque -Bera

La hipótesis de distribución normal de los errores de pronóstico no se rechaza solo en el 4% de las verificaciones. En los casos donde se rechaza la hipótesis se complementa la prueba con el coeficiente de asimetría, resultando que en el 76% de las verificaciones se evidencia una distribución con sesgo positivo, mientras que en 18% de los casos se evidencia sesgo negativo.

Dentro de los puntos con sesgo negativo se destacan Rionegro y Tunja con 10 meses de prevalencia y Bogotá con 9 meses.

Prueba t-student

Se analiza los resultados del contraste de la hipótesis de media de los errores de pronóstico igual a cero, teniendo presente la premisa que solo en el 4% de los casos se puede asumir normalidad. Con esta aclaración, los resultados de la prueba muestran que en el 47,5% de los casos hay tendencia a subestimar, mientras que en el 40% hay tendencia a sobreestimar y en el 10,9% se puede pensar en que la media es estadísticamente cero.

Prueba de rango signado de Wilcoxon

Dado que el supuesto de normalidad en la distribución no se garantiza, los resultados de esta prueba son más apropiados para verificar si la tendencia central, la mediana, es estadísticamente cero. Solo el 5.7% no rechaza la hipótesis nula, mientras que en el 65.4% hay evidencia de tendencia a sobreestimar y en el 27.8% hay evidencia de tendencia a subestimar.

Las regiones que principalmente tienen tendencia a subestimar son la Orinoquía y la Amazonía, adicionalmente Neiva en la región Andina y Buenaventura en la región Pacífico. Por otra parte, las regiones que tienden a sobreestimar son la Región Andina y la región Caribe.

Prueba Breusch – Pagan

Desde el punto de vista de la hipótesis de varianza constante, en el 69.7% de los casos de verificación no se rechaza la hipótesis nula. En el 27.3% de los casos se rechaza la hipótesis

indicando heterocedasticidad, las ubicaciones con mayor prevalencia son Providencia (9 meses), Rionegro y San Andrés (8 meses). Adicionalmente, el mes que registra mayor heterocedasticidad es abril.

3.2. Temperatura máxima del aire

Los resultados descritos a continuación se soportan gráficamente en el **Anexo 3. Análisis temperatura máxima**.

Verificación de pronósticos multicategoría

Exactitud

Con base en la categorización realizada se evidenció que en ningún caso de verificación la proporción de pronósticos que estaban en la categoría correcta supera al 60%. El 62,6% de los casos de verificación tienen proporción de pronósticos en la categoría correcta menor al 20%. El 28,9% tienen una proporción de exactitud entre el 20% y el 40 %. Estos resultados dan indicios de alto grado de sesgo en los pronósticos de temperatura máxima.

Verificación de pronósticos continuos

Error medio

El error medio con valores cercanos a cero, en el intervalo $[-0.5, 0.5]$ se presenta en 15% de los casos de verificación, principalmente en Arauca (10 meses), Barrancabermeja y Mitú (9 meses), y Quibdó (8 meses). El 74,6% de los casos analizados da indicios de subestimación, con valores de error medio menor a -0.5. Solo el 9.1% de los casos da indicios de sobre estimación, donde Cúcuta y Leticia tienden a la sobreestimación de la temperatura máxima en 10 meses del año.

Raíz del Error Cuadrático Medio (RMSE)

El 35% de los valores más altos de RMSE, es decir, donde hay mayor error de pronóstico están representados principalmente por los puntos de verificación Cali, Ibagué, Medellín, Riohacha, Rionegro, Santa Marta, Tunja, Valledupar y Villavicencio, ya que el alto error de pronóstico prevalece durante todo el año.

Por otro lado, dentro del 46% de los valores más bajos de RMSE se encuentran los puntos Apartadó, Barrancabermeja, Las gaviotas, Mocoa, Neiva, Pereira, Puerto Carreño y Quibdó, donde los valores bajos de error se presentan durante todo el año.

Correlación pronóstico Vs. Observaciones

El 40% de los casos de verificación tienen una correlación mayor a 0.6, principalmente en Riohacha (11 meses), Medellín (10 meses) y Valledupar 9 (meses) y especialmente en los meses de enero y mayo. En el 56% de los casos de verificación la correlación está en el intervalo (0.2, 0.6], Representado por Apartadó (12 meses), Bogotá y Tunja (11 meses).

Análisis de supuestos del error aleatorio

Prueba Jarque -Bera

Sobre la distribución de los errores de pronóstico, la prueba Jarque – Bera da evidencia que en el 49.8% de los casos de verificación los errores se distribuyen normal. En los casos donde se rechaza la hipótesis nula de distribución normal, los análisis se complementan con el cálculo del coeficiente de asimetría, donde se puede afirmar que el 45.1% tienen una distribución con sesgo negativo y el 3.7% tienen una distribución con sesgo positivo.

Los meses donde se presentan más distribuciones con sesgo negativo son octubre y noviembre, en el resto de los meses del año tiene mayor incidencia la distribución normal.

Espacialmente no se observan patrones definidos. Las distribuciones con sesgo negativo se presentan principalmente en los puntos San Andrés (11 meses), Barrancabermeja, Leticia, Villavicencio y Yopal (10 meses), Cali y Montería (9 meses). De otro lado, los puntos donde los errores se ajustan más a una normal son Ipiales, Manizales, Mocoa (11 meses), Barranquilla, Bogotá, Pasto, Riohacha, Santa Marta y Tunja (10 meses).

Prueba t-student

Se analiza los resultados del contraste de la hipótesis que la media de los errores de pronóstico es igual a cero, teniendo presente que solo en el 49.8% se puede hablar de distribución normal en los errores. El estadístico da evidencia que en el 81.5% de los casos de verificación existe tendencia a subestimar, mientras que en el 15.6% la tendencia es a sobreestimar y solo en el 1.6% se podría decir que la media es estadísticamente cero.

Los puntos donde se puede afirmar que el pronóstico de temperatura máxima sobreestima durante todos los meses del año son Cúcuta, Leticia y Puerto Carreño.

Prueba de rango signado de Wilcoxon

Los resultados de esta prueba confirman las características observadas con la prueba anterior, en este caso, se contrasta la hipótesis de que la mediana de los errores de pronóstico es igual a cero. En el 82% de los casos de verificación se tiene evidencia de subestimación, en el 14.8% evidencia de sobre estimación y en el 1.8% evidencia de mediana igual a cero.

Además, se confirma que los puntos donde el pronóstico de temperatura máxima sobreestima durante todos los meses del año son Cúcuta, Leticia y Puerto Carreño.

Prueba Breusch – Pagan

Con esta prueba se verifica la hipótesis nula de varianza constante en los errores (homocedasticidad). Los resultados indican que en el 80% de los casos de verificación hay evidencia de homocedasticidad. Mientras que en el 18.7% se rechaza la hipótesis nula.

Las ubicaciones que más rechazan la hipótesis de homocedasticidad son Medellín, Puerto Carreño, Riohacha (6 meses) y Rionegro (5 meses).

3.3. Temperatura mínima del aire

Los resultados descritos a continuación se soportan gráficamente en el **Anexo 4. Análisis temperatura mínima**.

Verificación de pronósticos multicategoría

Exactitud

Categorizando según los rangos de normalidad de la temperatura basados en percentiles del histórico de las observaciones, se verificó la proporción de pronósticos en la categoría correcta. Se destaca que en el 69.4% de casos de verificación la proporción de exactitud es menor al 20%, en el 23% de los casos la exactitud está entre el 20% y 40%, mientras que solo el 2% de los casos tiene una exactitud entre 40% y 60%. No hay casos que tengan exactitud mayor al 60%. Estos resultados dan indicios de que existe un alto grado de sesgo en los pronósticos de temperatura mínima.

Verificación de pronósticos continuos

Error medio

Se analiza esta métrica para tener evidencias de posible sesgo en los pronósticos. El error medio cercano a cero en el intervalo $[-0.5, 0.5]$ está representado por el 11.7% de los casos de verificación. En el 5% de los casos hay evidencia de sobreestimación principalmente en los puntos San Andrés (9 meses) y Providencia (8 meses). En el 81.5% de los casos hay evidencia de subestimación.

Raíz del Error Cuadrático Medio (RMSE)

Según los resultados de la verificación con esta métrica, se observa que en el 35% de los casos con RMSE más bajo los puntos con mayor prevalencia son Bucaramanga, Mocoa, Pasto, Quibdo (12 meses), las gaviotas, Riohacha (11 meses), Barranquilla y Cúcuta (10 meses). Por otra parte, dentro del 33% de RMSE más altos los puntos donde se mantienen

esos valores todo el año son Bogotá, Cartagena, Ibagué, Medellín, Rionegro, Tunja y Villavicencio.

Correlación pronóstico Vs. Observaciones

Según esta medida de asociación, solo el 2,3% tiene correlaciones mayores a 0.6. En el 66.2% de los casos se tienen correlaciones en el intervalo $[0.2, 0.6]$ que son correlaciones aceptables en las variables meteorológicas. Además, en el 28% de los casos verificados las correlaciones son menores o iguales a 0.2, las ubicaciones con mayor peso en esta categoría de baja correlación son Ibagué y Providencia (8 meses), Bogotá, Cali, Popayán y Rionegro (7 meses)

Análisis de supuestos del error aleatorio

Prueba Jarque -Bera

Según esta prueba hay evidencia para afirmar que en el 79.1% de los casos de verificación los errores de pronóstico siguen una distribución normal. Frente a los casos donde se rechaza la prueba de normalidad se complementa el análisis con el coeficiente de asimetría de donde se obtiene evidencia de que el 11,6% sigue una distribución con sesgo negativo y el 7.7% sigue una distribución con sesgo positivo.

Especialmente no hay patrones definidos. Sin embargo, los puntos que mayor sesgo positivo presentan son Leticia (5 meses) y Quibdó (4 meses), mientras que los puntos con más sesgo negativo son Providencia (5 meses), San Andrés y Villavicencio (4 meses).

Prueba t-student

En el 90% de los casos de verificación se rechaza la hipótesis nula de que el error medio es igual a cero en favor de la alternativa que la media del error es mayor evidenciando tendencia de los pronósticos a subestimar la temperatura mínima. De otro lado, para el 7.7% de los casos hay evidencia de sobreestimación y solo en el 0.8% se puede afirmar que el error medio es estadísticamente cero.

Se resalta que los puntos ubicados en San Andrés y Providencia tienen tendencia a sobreestimar todo el año.

Prueba de rango signado de Wilcoxon

Dada la predominancia de distribución normal en el error de pronóstico de la temperatura mínima, la prueba de rango signado de Wilcoxon para la hipótesis nula de mediana igual a cero confirma las evidencias obtenidas con la prueba t-student. En este caso en el 90% se rechaza la hipótesis nula en favor de tendencia a subestimar, en el 7.5% se rechaza en favor de sobre estimar y solo en el 1% no se rechaza la hipótesis nula.

En providencia la tendencia a sobreestimar se presenta durante 12 meses mientras que en San Andrés durante 11 meses.

Prueba Breusch – Pagan

El contraste de la hipótesis de varianza constante de los errores indica que en el 84.6% de los casos no se rechaza la hipótesis nula garantizando homocedasticidad, mientras que en 13.9% se rechaza en favor de heterocedasticidad.

Las ubicaciones que más rechazan la hipótesis de homocedasticidad son Leticia (5 meses) e Ipiales (4 meses).

Conclusiones

Precipitación

Verificación desde el enfoque dicotómico

- En las regiones con régimen de precipitación bajo o en las temporadas de menos lluvia, se evidencia que los pronósticos tienen mayor exactitud, es decir, que los pronósticos tienen mayor probabilidad de ser correctos. No obstante, en esos escenarios también se tiene alta probabilidad de falsas alarmas, es decir, que los eventos pronosticados de precipitación no ocurran.
- Las temporadas de lluvia están asociadas a los porcentajes de exactitud bajos con alta probabilidad de sobreestimación del número de eventos de precipitación que realmente ocurren.
- En las regiones asociadas a los piedemontes llanero y amazónico, y en Valledupar hay alta probabilidad que el número de eventos de precipitación pronosticados sea menor al número de eventos que sí ocurren.
- Las regiones Pacífico y Amazonía están asociadas con tasas de falsa alarma bajas.
- En la región Andina se evidencian tasas de acierto altas, sin embargo, allí se tiende a sobreestimar la ocurrencia de eventos, principalmente en la segunda temporada de lluvias.

Verificación desde el enfoque multicategoría

- Se verifica que los valores altos de exactitud están más asociados a la temporada de menos lluvias diciembre -febrero.
- En la región Andina la proporción de pronósticos en la categoría correcta es baja indicando la existencia de sesgo.

Verificación desde el enfoque continuo

- Mediante la métrica RMSE se verifica que los errores de pronóstico de precipitación tienen menor magnitud en las temporadas de menos lluvias o en las ubicaciones con régimen de precipitación de acumulados bajos. Al tiempo que los valores altos de RMSE están asociados a las ubicaciones con acumulados de precipitación altos.
- Mediante la verificación con la métrica Error Medio se presume la existencia de error sistemático que se valida de forma más detallada con el análisis inferencial.

Verificación de supuestos de error aleatorio

- Desde el punto de vista de la distribución se observó la predominancia de distribuciones con sesgo positivo, esto puede estar asociado a la naturaleza misma de la variable precipitación que

teóricamente sigue una distribución Gamma. La verificación de esta última afirmación se sale del alcance de este trabajo.

- Dado que no se garantiza normalidad en la distribución, la inferencia más apropiada sobre la tendencia central de los errores es la aplicada con la prueba de Wilcoxon, con esta se confirma la existencia de sesgo en el 93.2% de los casos verificados, predominando la tendencia a sobreestimar los acumulados diarios de precipitación.
- Con respecto al principio de homocedasticidad, este predomina. No obstante, se debe verificar las características de las series donde la varianza no es constante como en el caso de las series asociadas a Providencia, Rionegro y San Andrés, donde la prevalencia de heterocedasticidad se da en la mayoría de los meses del año.

Temperatura del aire máxima

Verificación desde el enfoque multicategoría

- La verificación categorizando los pronósticos y las observaciones de temperatura máxima dan evidencia de existencia de error sistemático.

Verificación desde el enfoque continuo

- El análisis de los valores del error medio indica la existencia de error sistemático en los pronósticos de temperatura máxima ya que en mayor porcentaje se tiende a subestimar los valores observados. Esta evidencia se confirma con las pruebas de hipótesis que se describen en la verificación de supuestos sobre el error.
- Con la métrica RMSE se pudo observar puntos específicos donde el error promedio es bajo y donde es alto. Sin embargo, no se observaron patrones espaciales o temporales asociados a la incidencia de las magnitudes de error.
- De las tres variables analizadas esta es la que mejores correlaciones presenta entre los pronósticos y los valores registrados.

Verificación de supuestos de error aleatorio

- Se evidenció que casi en el 50% de los casos de verificación los errores de pronóstico siguen una distribución normal. Por otro lado, en los casos donde se rechaza la hipótesis de normalidad el coeficiente de asimetría está en favor distribuciones con sesgo negativo.
- Para el análisis sobre la tendencia central, las pruebas t-student y de rango signado de Wilcoxon, la primera sobre la media y la segunda sobre la mediana, coinciden con porcentajes semejantes, en la existencia de error sistemático ya que en menos del 2% de los casos verificados se puede afirmar que la tendencia central es cero. En general se puede afirmar que los pronósticos de temperatura máxima tienden a subestimar los valores reales.
- La verificación de la homocedasticidad en los errores mostró que en menos del 20% se rechaza la hipótesis nula.

Temperatura del aire mínima

Verificación desde el enfoque multicategoría

- La verificación categorizando los pronósticos y las observaciones de temperatura máxima dan evidencia de existencia de error sistemático.

Verificación desde el enfoque continuo

- Confirmando lo observado desde el enfoque multicategoría, los análisis con la métrica Error Medio, dan indicios de un alto grado de sesgo asociado a la tendencia de los pronósticos a subestimar los valores registrados de temperatura mínima. Con las pruebas estadísticas sobre la tendencia central de los errores se confirma lo detectado con esta métrica.
- Con la métrica RMSE no se pueden establecer patrones espaciales y temporales asociados a valores altos o bajos en los errores de pronóstico. Se pudo especificar las ubicaciones donde las magnitudes altas o bajas prevalecen durante todo el año.
- Las correlaciones entre los pronósticos y las observaciones presentan valores aceptables para los análisis de variables meteorológicas.

Verificación de supuestos de error aleatorio

- De las tres variables analizadas, los errores de pronóstico de la temperatura mínima son los que más se ajustan a una distribución normal con casi el 80% de los casos verificados.
- Según las pruebas sobre la tendencia central igual a cero, se evidenció que en el 90% de los casos verificados los pronósticos tienden a subestimar los valores reales.
- La prueba de varianza constate indica que en un alto porcentaje se garantiza la homocedasticidad en los errores de pronóstico.

Recomendaciones

Después de identificar diferentes características en los errores de pronóstico y confirmar la existencia de error sistemático, es natural que el siguiente paso sea buscar estrategias para reducir al máximo ese tipo de error.

Una primera estrategia sería identificar las posibles fuentes de error en la parametrización del modelo WRF, o de otra forma sería pensar en ajustes posteriores al pronóstico dadas las características ya identificadas. En el primer caso implicaría una validación del modelo WRF y en el segundo caso se plantearían ajustes desde un enfoque estadístico.

Desde el enfoque estadístico los ajustes para las temperaturas serían más sencillos, especialmente para la temperatura mínima, ya que la garantía de correlaciones positivas y de los supuestos de normalidad y homocedasticidad en los errores, dan pie a que solo sea necesario hacer ajustes en los niveles de los pronósticos. Por otra parte, la propuesta de ajuste para los pronósticos de la variable precipitación requieren mayor esfuerzo ya que se debe identificar a qué tipo de distribución se ajustan los errores y dependiendo de las propiedades de la distribución, proponer los ajustes correspondientes. Adicionalmente, también es importante identificar los escenarios donde no se garantiza la homocedasticidad ya que implican ajustes específicos para cada caso.

Referencias Bibliográficas

- Arango, C., & Ruiz, J. F. (2012). *Implementación del modelo WRF para la sabana de Bogotá*. http://www.ideam.gov.co/documents/21021/21132/Modelo_WRF_Bogota.pdf/f1d34638-e9f8-4689-b5f4-31957c231c46
- Barrios, J. (2019). *Informe final contrato 360 de 2018, validación preliminar de pronóstico del tiempo en áreas insulares de Colombia*.
- Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, 1287–1294.
- Corzo, J. A. (2005). *Notas de clase. Estadística no paramétrica (Métodos basados en rangos)*. Univ. Nacional de Colombia.
- Cowpertwait, P. S. P., & Metcalfe, A. V. (2009). *Introductory time series with R* (1st ed.). Springer-Verlag New York. <https://doi.org/10.1007/978-0-387-88698-5>
- García, M. J. (2014). *Validación de la capacidad del modelo WRF "Weather Research and Forecasting" para pronosticar lluvia intensa, usando el método orientado a objetos y tablas de contingencia*. Universidad Nacional de Colombia.
- Guerrero-Higueras, A. M., Garcia-Ortega, E., Matellán-Olivera, V., & Sánchez, J. L. (2012). Procesamiento paralelo de los pronósticos meteorológicos del modelo WRF mediante NCL. *Actas de Las XXIII Jornadas de Paralelismo JP2012*, 55–60.
- Guzmán, D., Ruíz, J. F., & Cadena, M. (2014). Regionalización de Colombia según la estacionalidad de la precipitación media mensual, a través de análisis de componentes principales (ACP). *Nota Técnica IDEAM, METEO*, 54.
- IDEAM. (2017). *Atlas climatológico de Colombia*.
- Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6, 255–259.
- JCGM. (2008). *Evaluation of measurement data-Guide to the expression of uncertainty in measurement - JCGM 100:2008*. www.bipm.org
- Jolliffe, I., & Stephenson, D. B. (2012). *Forecast verification: a practitioner's guide in atmospheric science*. John Wiley & Sons.
- JWGFVR. (2017). *7th International Verification Methods Workshop*. <https://www.cawcr.gov.au/projects/verification/>
- Koenker, R. (1981). A note on studentizing a test for heteroscedasticity. *Journal of Econometrics*, 17(1), 107–112.
- Manubens, N., Caron, L. P., Hunter, A., Bellprat, O., Exarchou, E., Fučkar, N. S., Garcia-Serrano, J., Massonnet, F., Ménégos, M., Sicardi, V., Batté, L., Prodhomme, C., Torralba, V., Cortesi, N., Mula-Valls, O., Serradell, K., Guemas, V., & Doblas-Reyes, F. J. (2018). An R package for climate

forecast verification. *Environmental Modelling and Software*, 103, 29–42.
<https://doi.org/10.1016/j.envsoft.2018.01.018>

Martínez, A., Julieta, P., & Serna, J. (2018). *Propuesta de umbrales de normalidad basada en las funciones de distribución de las series de datos y análisis de eventos de extremos para las variables meteorológicas: precipitación, número de días con lluvia y las temperaturas mínima, media y máxima-METEO/001-2018*.

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. <https://www.r-project.org/>

Ruiz, J. F., Arango, C., & Kilpinen, J. (2013). *Verificación del modelo WRF que opera IDEAM*. <http://www.ideam.gov.co/documents/21021/21132/VerificacionWRF.pdf/db010231-8106-4160-a16b-a5aaed198ea8>

Ruiz M, J. F., & Melo F, J. Y. (2020). *Modelación Numérica de Tiempo y Clima IDEAM*. <http://bart.ideam.gov.co/wrfideam/>

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., Wang, W., Powers, J. G., Duda, M. G., Barker, D. M., & Huang, X.-Y. (2021). *A Description of the Advanced Research WRF Model Version 4.3 (No. NCAR/TN-556+STR)*. <https://doi.org/10.5065/1dfh-6p97>

Venegas, T. (2020). *Verificación de los modelos globales GFS y ECMWF, y del modelo regional WRF para las variables precipitación, viento y temperatura en Chile durante los años 2018-2019*.

WMO. (2008). *Recommendations for the Verification and Intercomparison of QPFs and PQPFs from Operational NWP Models WWRP 2009 - 1*.

WMO. (2018). *Guidance on Verification of Operational Seasonal Climate Forecasts WMO-No. 1220*. <http://public.wmo.int/en/>

Anexos

Anexo 1. Identificación y ubicación geográfica de las estaciones meteorológicas utilizadas en el análisis

Código	Nombre	Departamento	Municipio	latitud	longitud
48015050	AEROPUERTO VASQUEZ COBO [48015050]	Amazonas	Leticia	-4,1939	-69,9409
12015070	AEROPUERTO LOS CEDROS [12015070]	Antioquia	Carepa	7,8164	-76,7179
27015330	OLAYA HERRERA - AUT [27015330]	Antioquia	Medellín	6,2247	-75,5882
23085270	AEROPUERTO J.M. CORDOVA - AUT [23085270]	Antioquia	Rionegro (Antioquia)	6,1686	-75,4261
37055010	AEROPUERTO SANTIAGO PÉREZ [37055010]	Arauca	Arauca	7,0694	-70,7381
17015010	AEROPUERTO SESQUICENTENARIO [17015010]	Archipiélago de San Andrés, Providencia y Santa Catalina	San Andrés (Archipiélago de San Andrés, Providencia y Santa Catalina)	12,5878	-81,7011
17025020	AEROPUERTO EL EMBRUJO AUT [17025020]	Archipiélago de San Andrés, Providencia y Santa Catalina	San Andrés Y Providencia (Santa Isabel)	13,3595	-81,3577
29040450	BARRANQUILLA SEDE [29040450]	Atlántico	Barranquilla	10,9976	-74,7967
29045190	AEROPUERTO E. CORTISOZ - AUT [29045190]	Atlántico	Soledad	10,9178	-74,7797
21202280	SEDE IDEAM CALLE 25D KRA [21202280]	Bogotá	Bogotá, D.C	4,6840	-74,1290
21205791	ELDORADO CATAM - AUT [21205791]	Bogotá	Bogotá, D.C	4,7056	-74,1507
29035000	SINCERIN - AUT [29035000]	Bolívar	Arjona	10,1426	-75,2783
14015080	AEROPUERTO RAFAEL NUNEZ [14015080]	Bolívar	Cartagena De Indias	10,4473	-75,5160
14010050	CAÑAVERAL [14010050]	Bolívar	Turbaco	10,4100	-75,3400
24035430	TUNGUAVITA - AUT [24035430]	Boyacá	Paipa	5,7458	-73,1164
26155110	AEROPUERTO LA NUBIA AUT [26155110]	Caldas	Manizales	5,0298	-75,4699
44030060	LARANDIA [44030060]	Caquetá	Florencia (Caquetá)	1,4900	-75,4900
35215020	AEROPUERTO YOPAL - AUT [35215020]	Casanare	Yopal	5,3204	-72,3875
26035030	AEROPUERTO G L VALENCIA [26035030]	Cauca	Popayán	2,4529	-76,6088
28025502	AEROPUERTO ALFONSO LOPEZ - [28025502]	Cesar	Valledupar	10,4362	-73,2477
11045010	AEROPUERTO EL CARANO [11045010]	Chocó	Quibdó	5,6906	-76,6438
13035501	AEROPUERTO LOS GARZONES [13035501]	Córdoba	Montería	8,8258	-75,8251
31095030	PUERTO INIRIDA - AUT [31095030]	Guainía	Inírida	3,8744	-67,9191
31015010	EL TRUENO [31015010]	Guaviare	El Retorno	2,3728	-72,6400
21115020	AEROPUERTO BENITO SALAS [21115020]	Huila	Neiva	2,9488	-75,2931

Código	Nombre	Departamento	Municipio	latitud	longitud
15065180	AEROPUERTO ALM. PADILLA - [15065180]	La Guajira	Riohacha	11,5284	-72,9177
15015050	AEROPUERTO SIMON BOLIVAR [15015050]	Magdalena	Santa Marta	11,1147	-74,2310
35035020	AEROPUERTO VANGUARDIA [35035020]	Meta	Villavicencio	4,1619	-73,6176
52055230	AEROPUERTO SAN LUIS - AUT [52055230]	Nariño	Aldana	0,8571	-77,6778
52045020	AEROPUERTO ANTONIO NARINO [52045020]	Nariño	Chachagüí	1,3941	-77,2909
16015010	AEROPUERTO CAMILO DAZA [16015010]	Norte de Santander	Cúcuta	7,9303	-72,5092
44010110	PUERTO LIMON [44010110]	Putumayo	Mocoa	1,0268	-76,5415
44015010	VILLAGARZON [44015010]	Putumayo	Villagarzón	1,0343	-76,6193
26125061	AEROPUERTO EL EDEN - - AUT [26125061]	Quindío	Armenia (Quindío)	4,4547	-75,7664
26135040	AEROPUERTO MATECANA [26135040]	Risaralda	Pereira	4,8159	-75,7372
23155030	AEROPUERTO YARIGUIES AUT [23155030]	Santander	Barrancabermeja	7,0264	-73,8086
23195502	AEROPUERTO PALONEGRO AUT [23195502]	Santander	Lebrija	7,1215	-73,1845
25025080	AEROPUERTO RAFAEL BARVO [25025080]	Sucre	Corozal	9,3339	-75,2831
25025270	UNISUCRE AUT [25025270]	Sucre	Sampués	9,3164	-75,3875
21245010	PERALES HATO OPIA [21245010]	Tolima	Ibagué	4,4297	-75,2005
21245040	AEROPUERTO PERALES [21245040]	Tolima	Ibagué	4,4241	-75,1394
53115010	AEROPUERTO BUENAVENTUR [53115010]	Valle del Cauca	Buenaventura	3,8201	-76,9924
26075010	PALMIRA ICA [26075010]	Valle del Cauca	Palmira	3,5135	-76,3149
26075040	AEROPUERTO A BONILLA [26075040]	Valle del Cauca	Palmira	3,5367	-76,3866
42075010	MITU [42075010]	Vaupés	Mitú	1,2600	-70,2400
34015010	LAS GAVIOTAS [34015010]	Vichada	Cumaribo	4,5539	-70,9301
38015030	AEROPUERTO PUERTO CARRENO [38015030]	Vichada	Puerto Carreño	6,1824	-67,4912

Anexo 2. Análisis precipitación

Anexo 2.1. Métricas de verificación de pronóstico dicotómico precipitación

Figura 2. Exactitud de pronóstico dicotómico de precipitación por meses

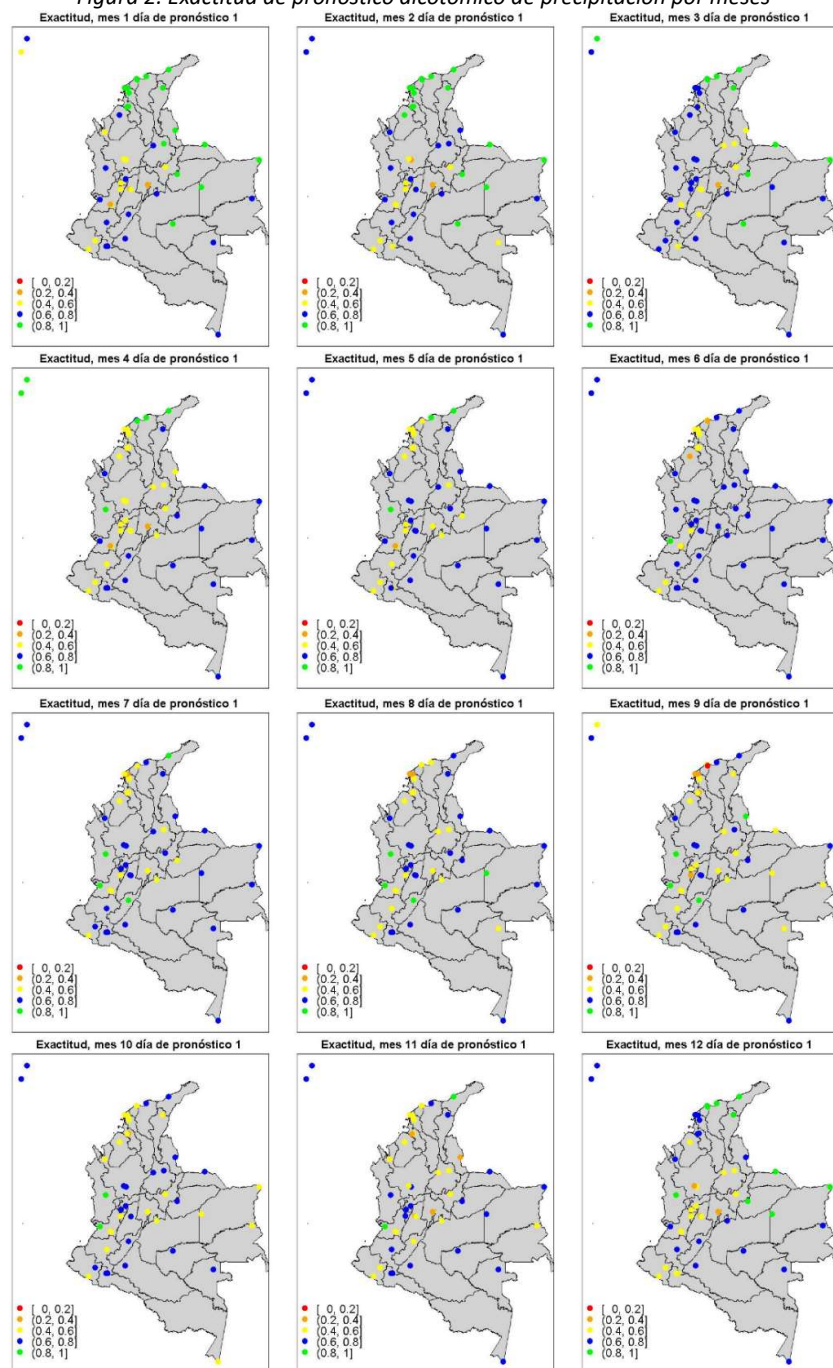


Figura 3. Bias score de pronóstico dicotómico de precipitación por meses.

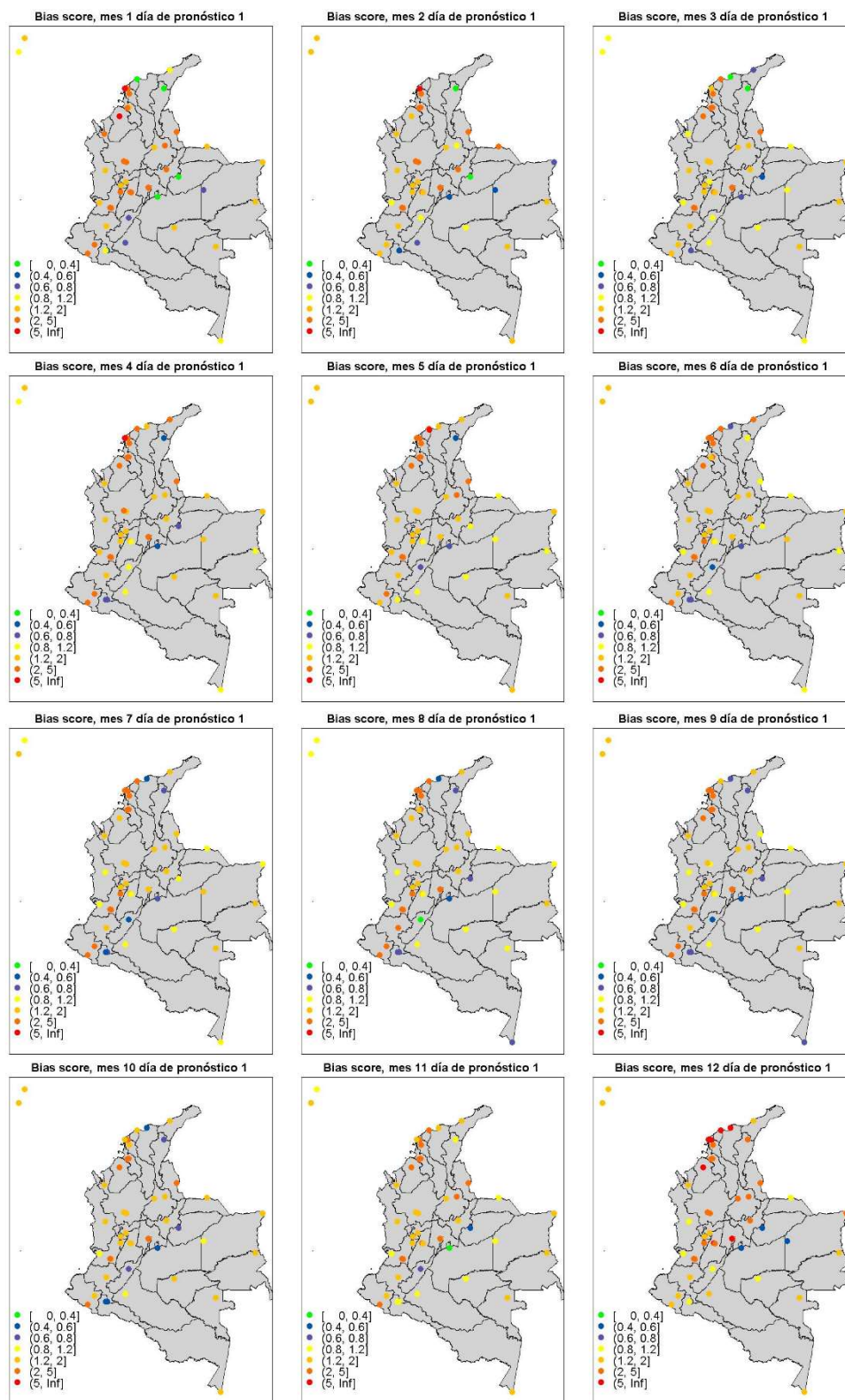


Figura 4. Tasa de aciertos de pronóstico dicotómico de precipitación por meses.

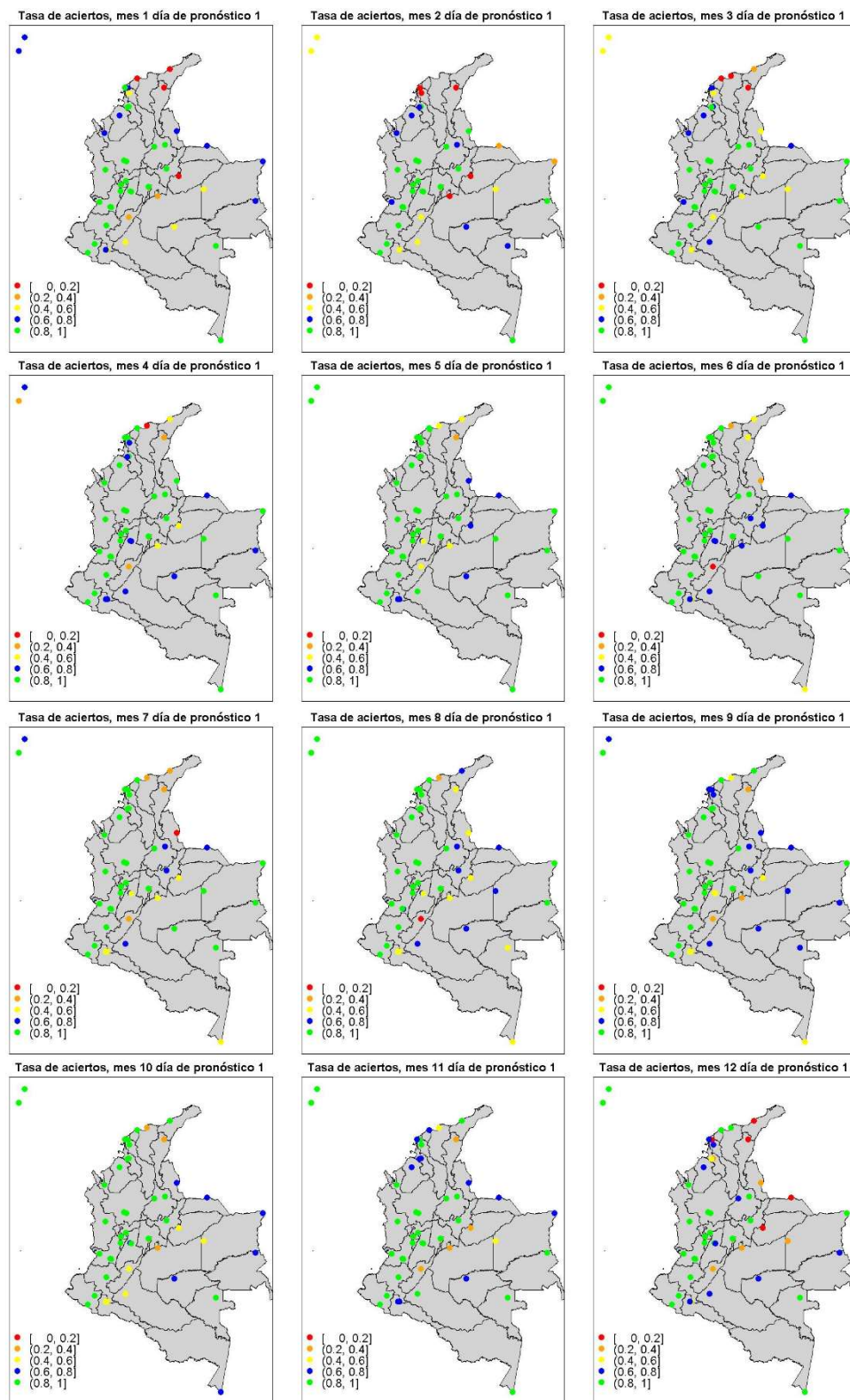
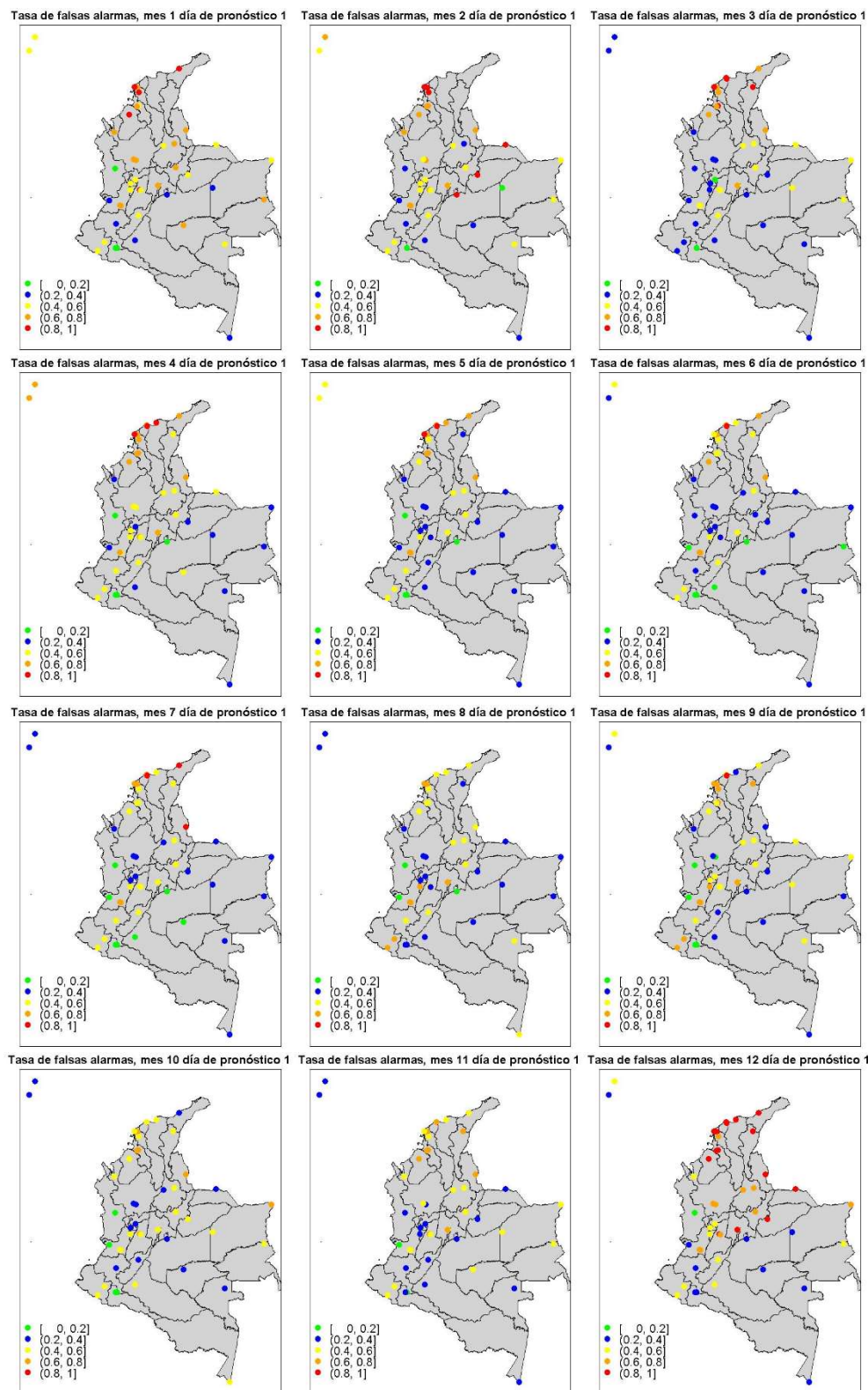
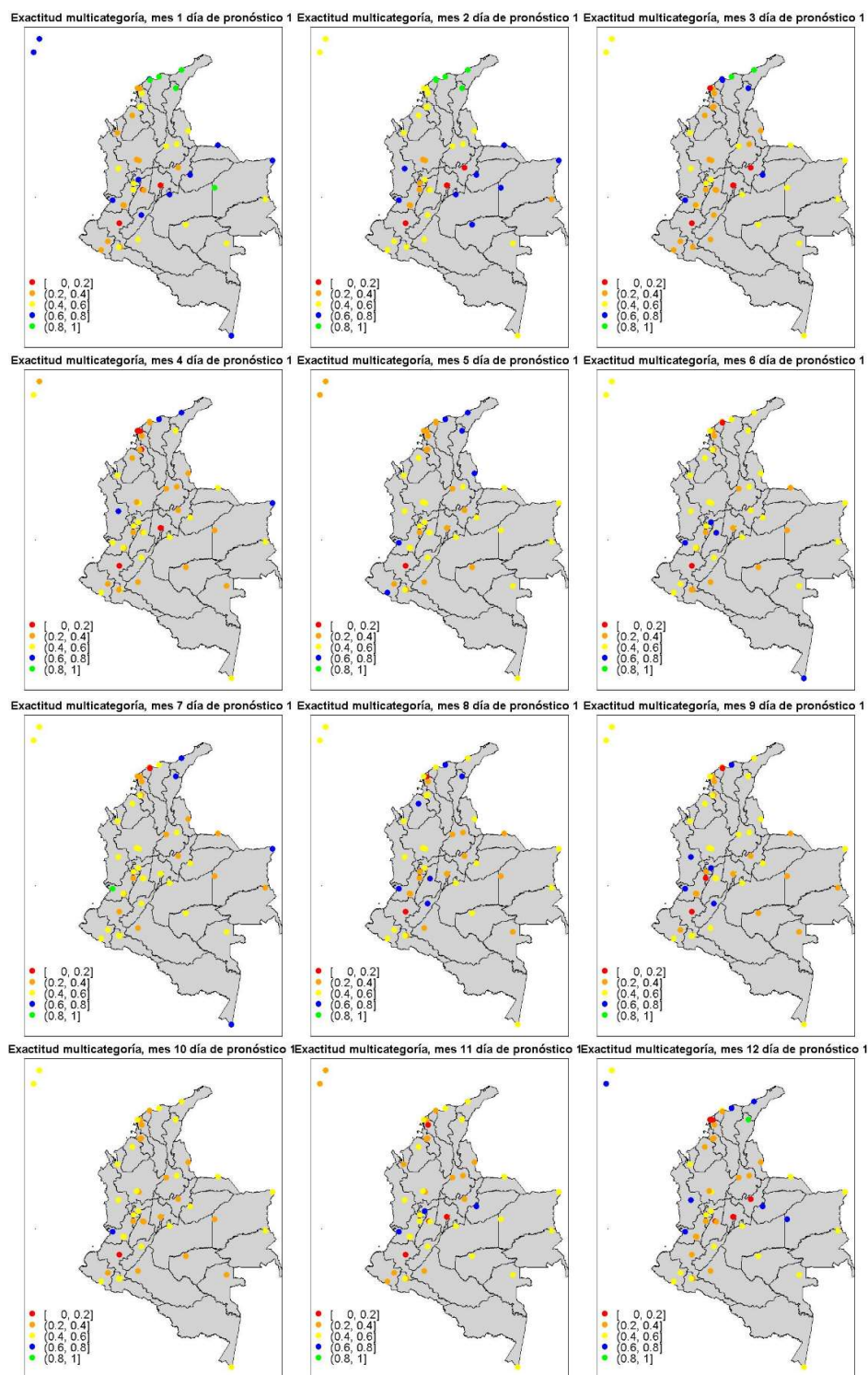


Figura 5. Tasa de falsas alarmas de pronóstico dicotómico de precipitación por meses.



Anexo 2.2. Métricas de verificación de pronóstico multicategoría precipitación

Figura 6. Exactitud de pronóstico multicategoría de precipitación por meses.



Anexo 2.3. Métricas de verificación de pronóstico de precipitación (variable continua)

Figura 7. Error medio (Bias aditivo) de pronóstico de precipitación por meses.

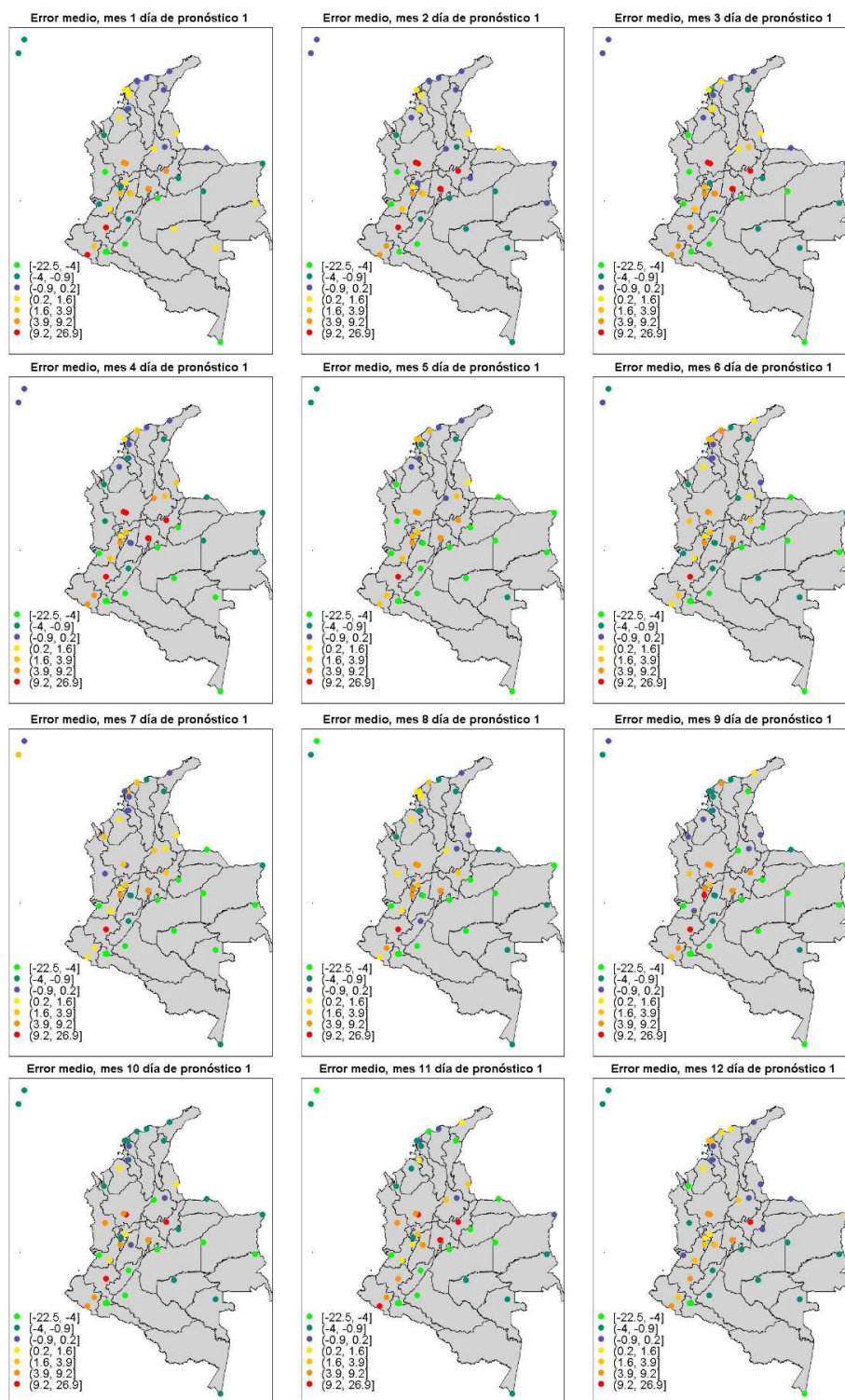


Figura 8. Raíz Error cuadrático medio de pronóstico de precipitación por meses.

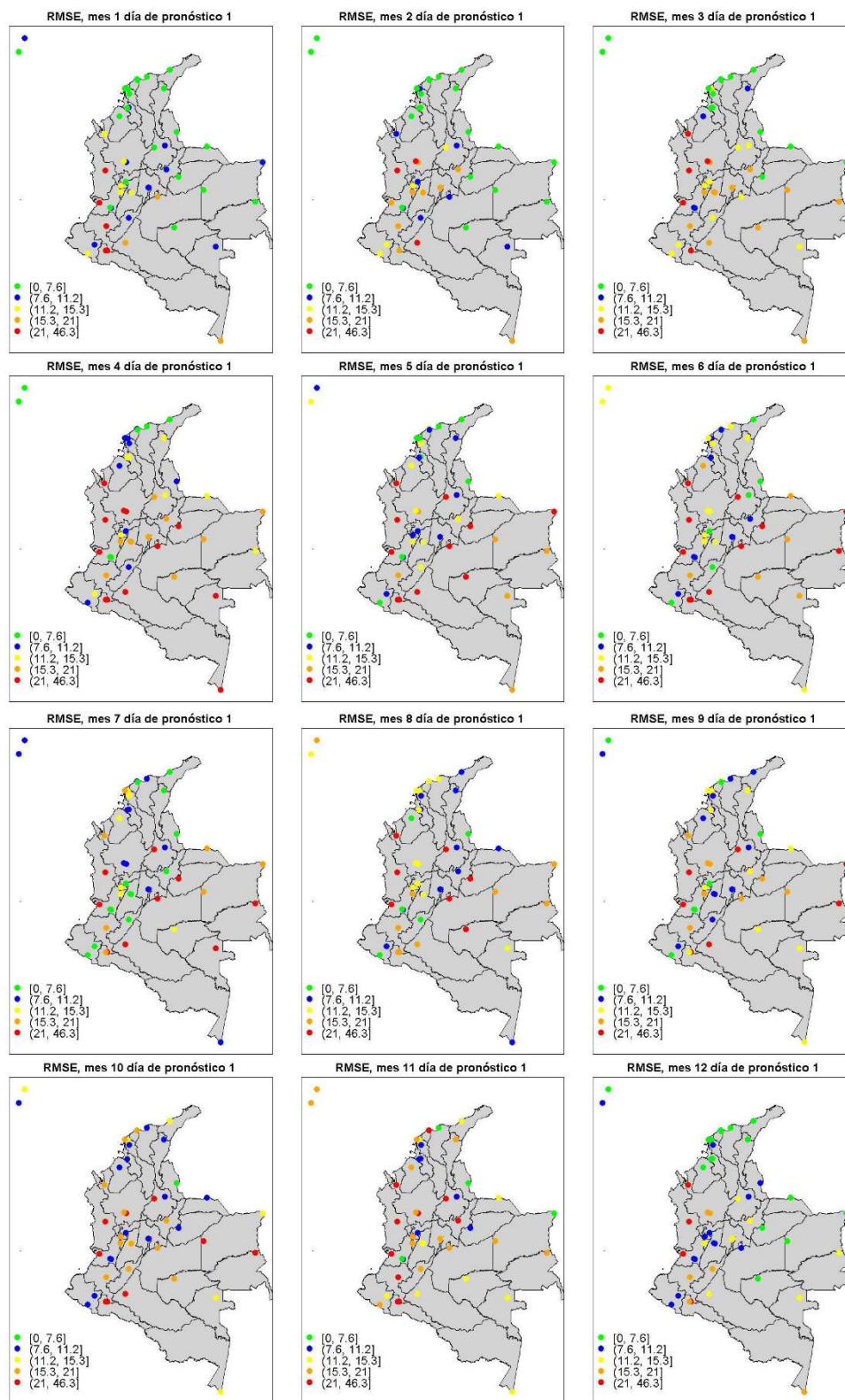
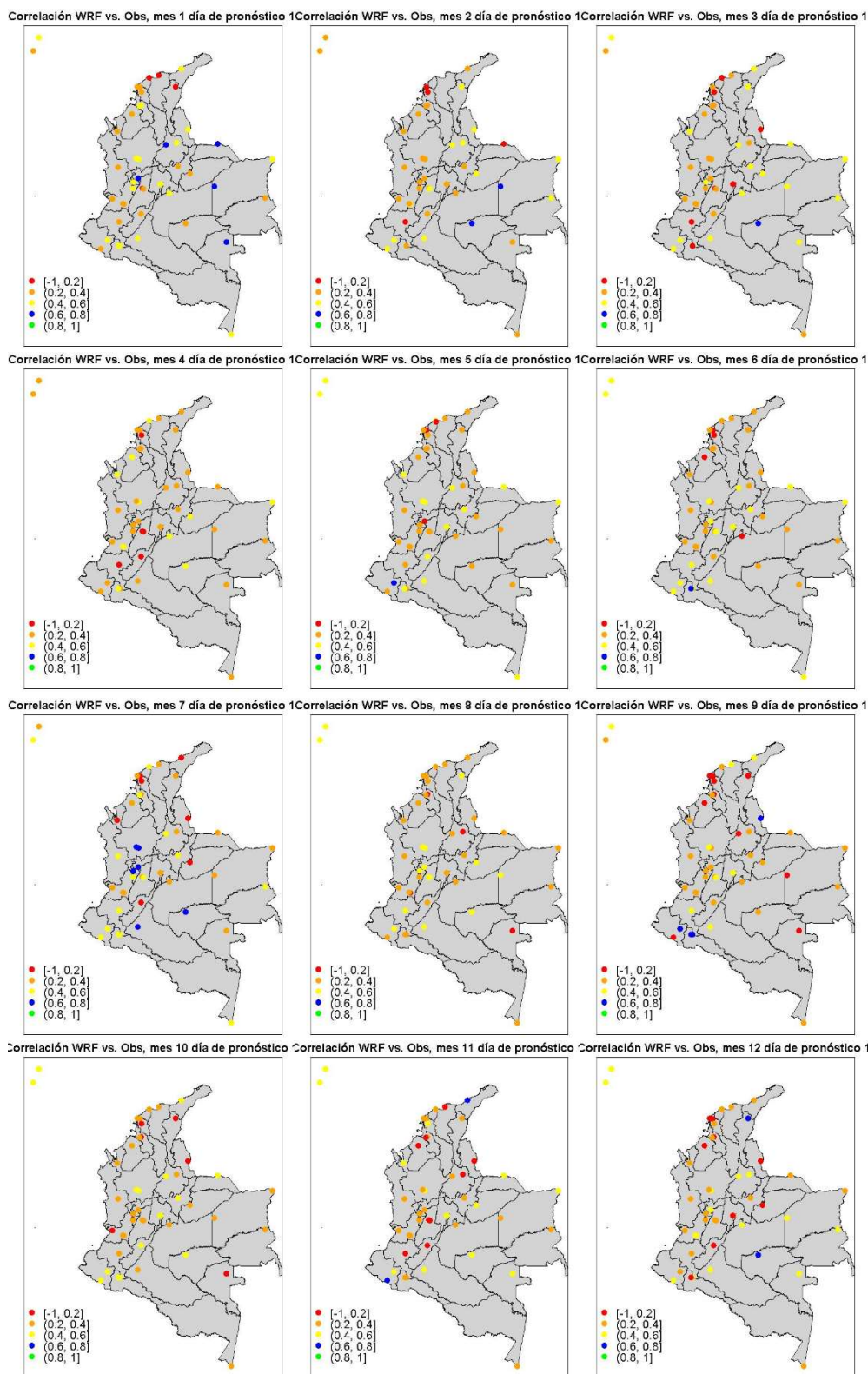


Figura 9. Correlación pronóstico Vs. observaciones de precipitación por meses.



Anexo 2.4. Análisis de supuestos de error aleatorio de pronóstico de precipitación

Figura 10. Prueba Jarque – Bera (H_0 : Distribución normal de los errores de pronóstico de precipitación) por meses.

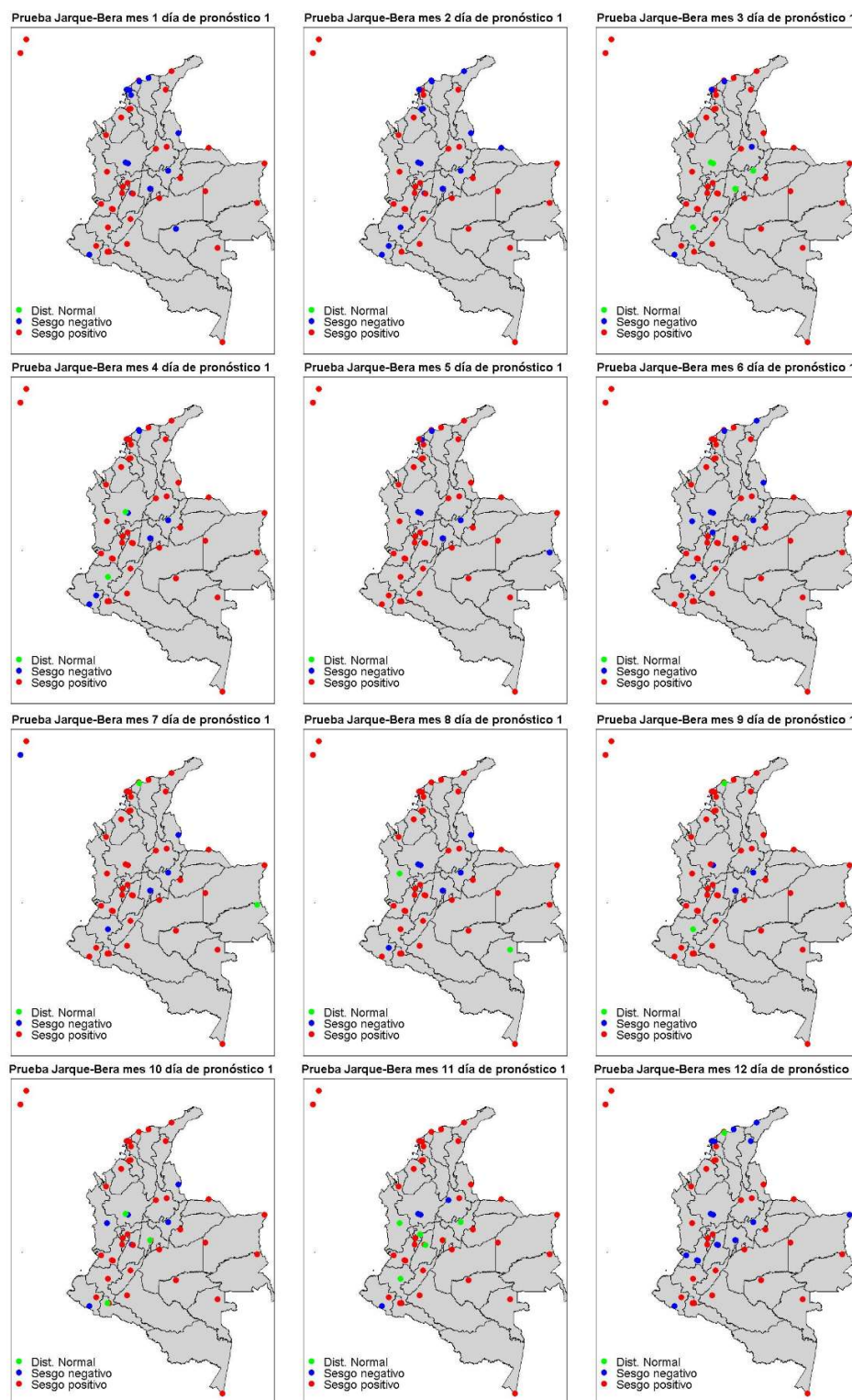


Figura 11. Prueba t-student ($H_0: \bar{X}_{error} = 0$) por meses (precipitación).

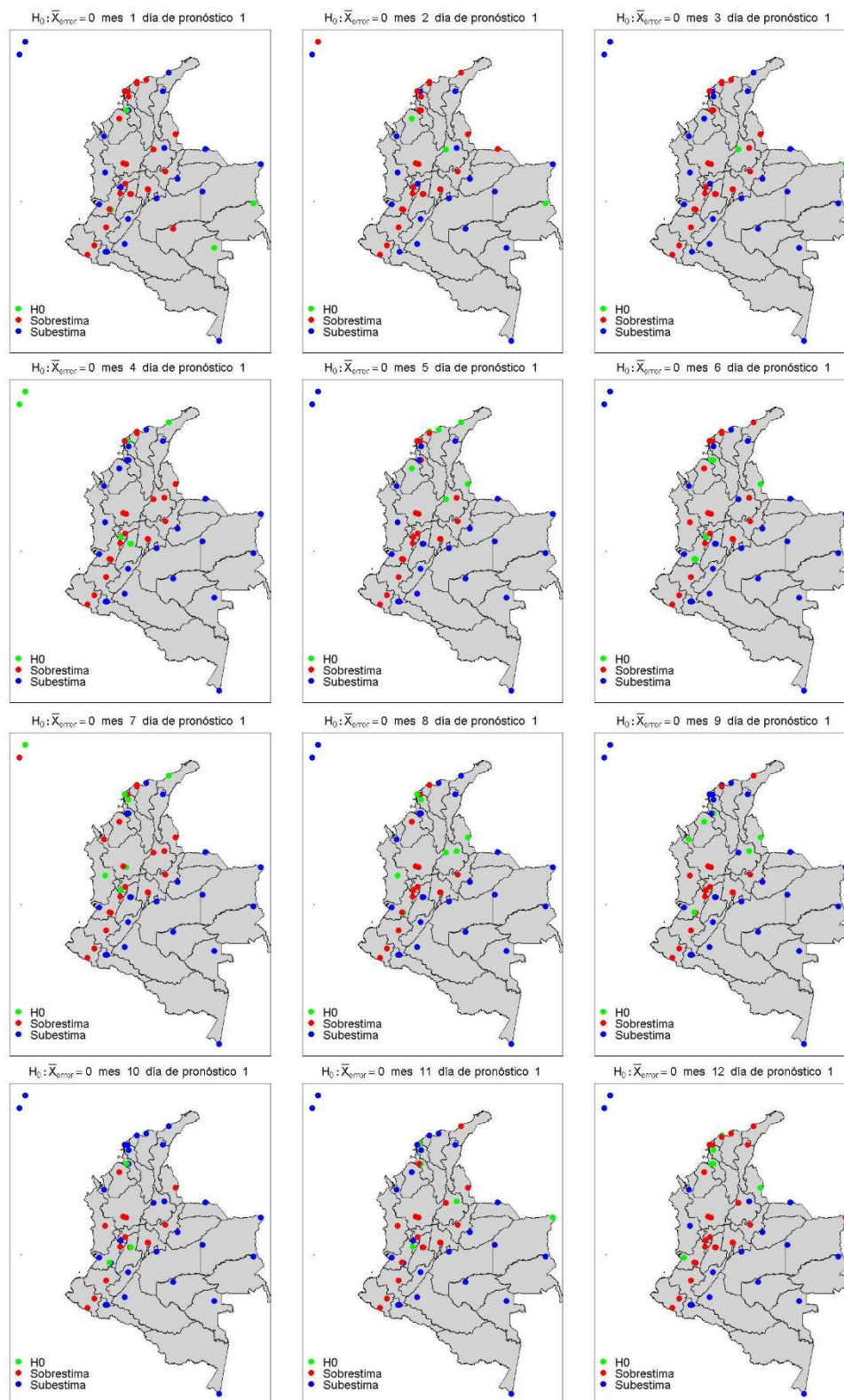


Figura 12. Prueba de rango signado de Wilcoxon ($H_0: \text{Mediana}_{\text{error}} = 0$) por meses (precipitación).

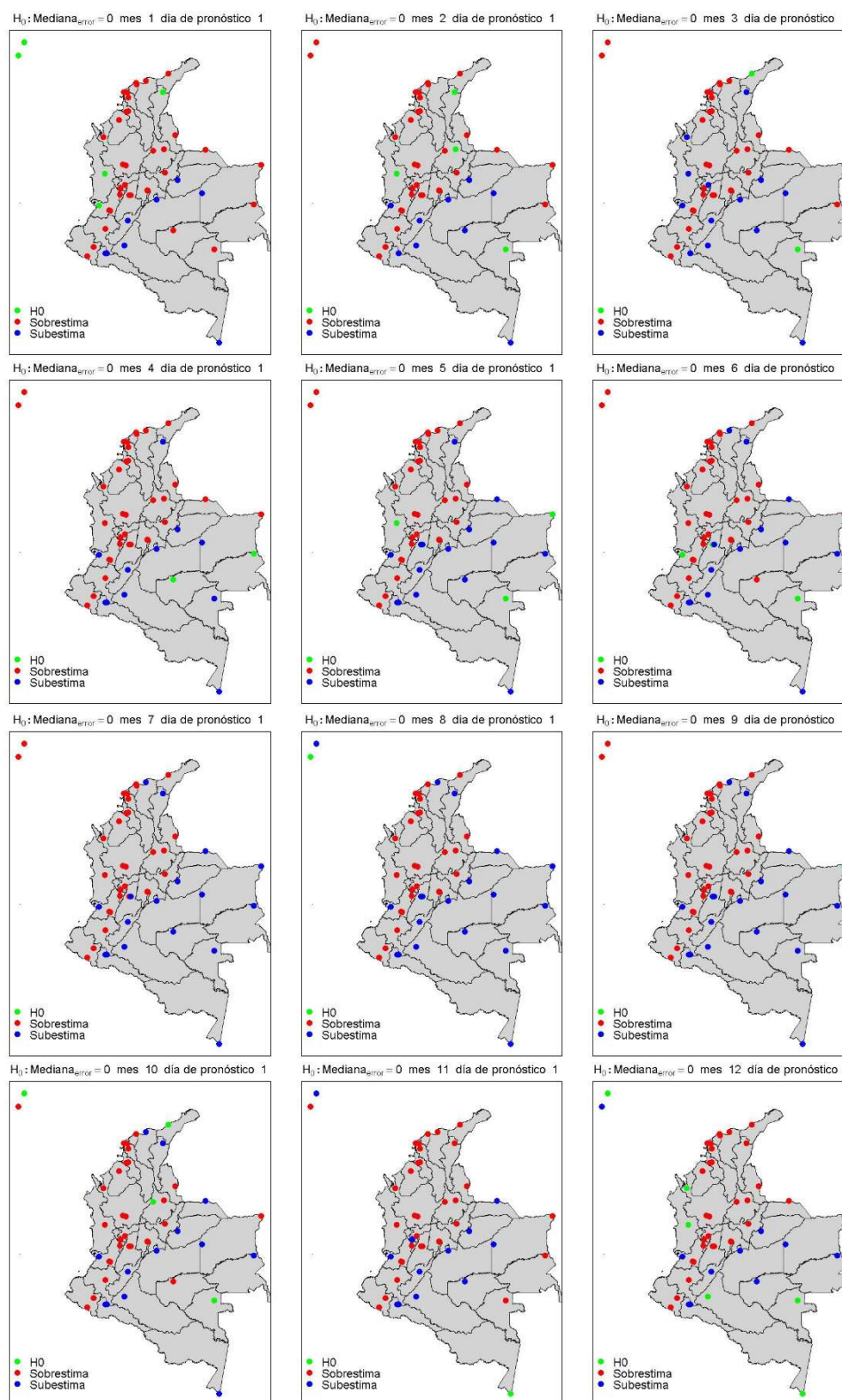


Figura 13. Prueba Breusch – Pagan (H_0 : Varianza constante del error) por meses (precipitación).

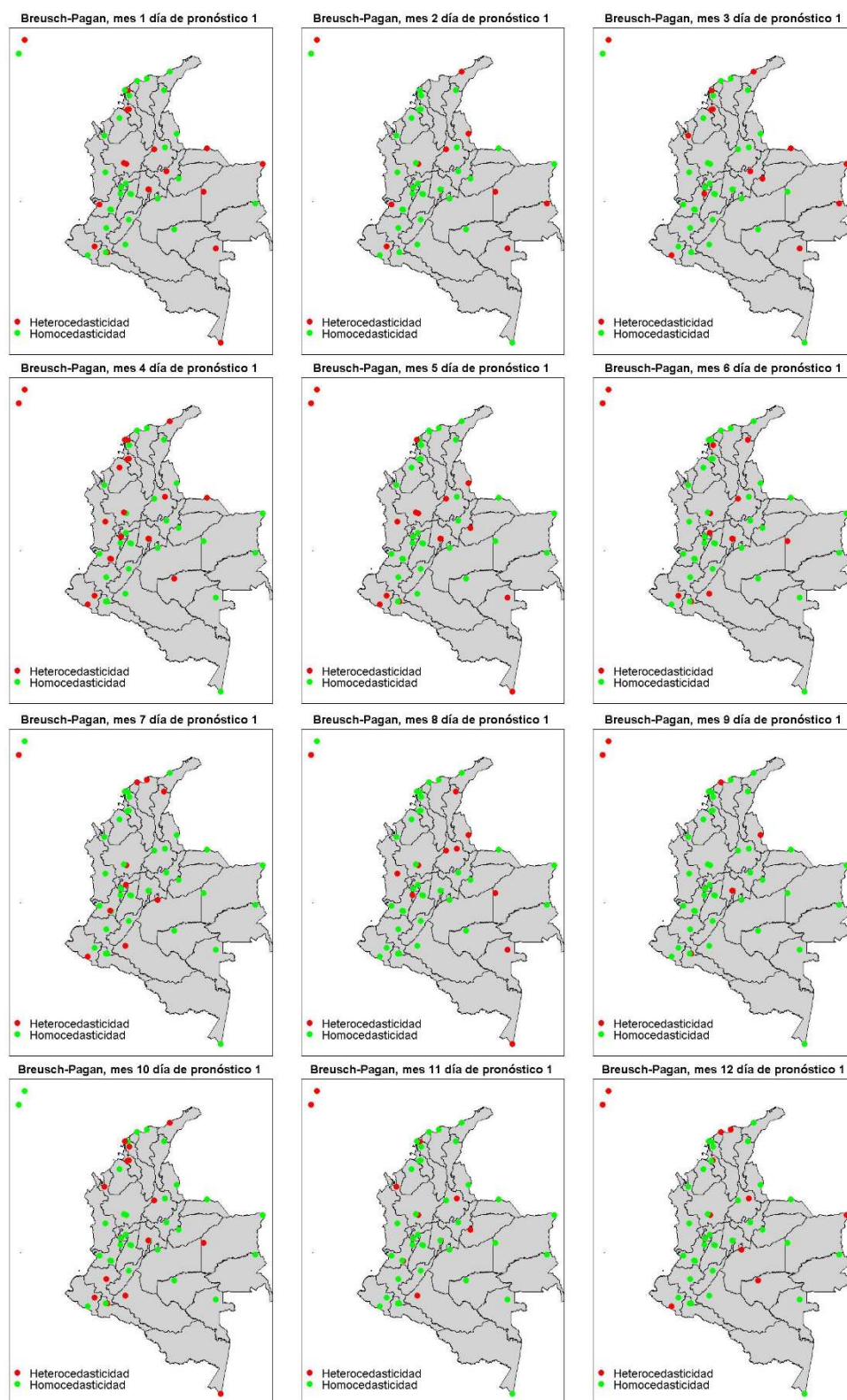
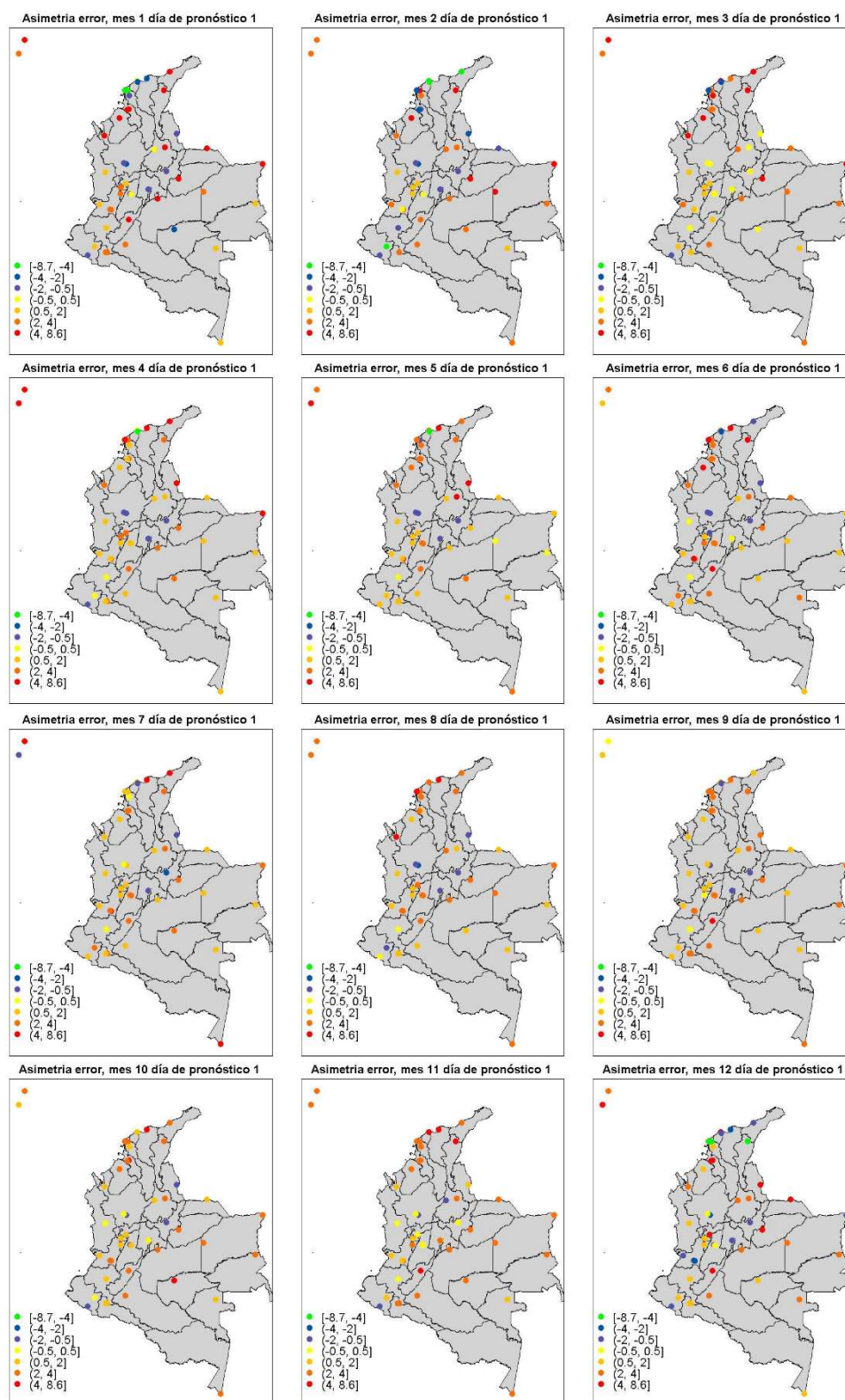


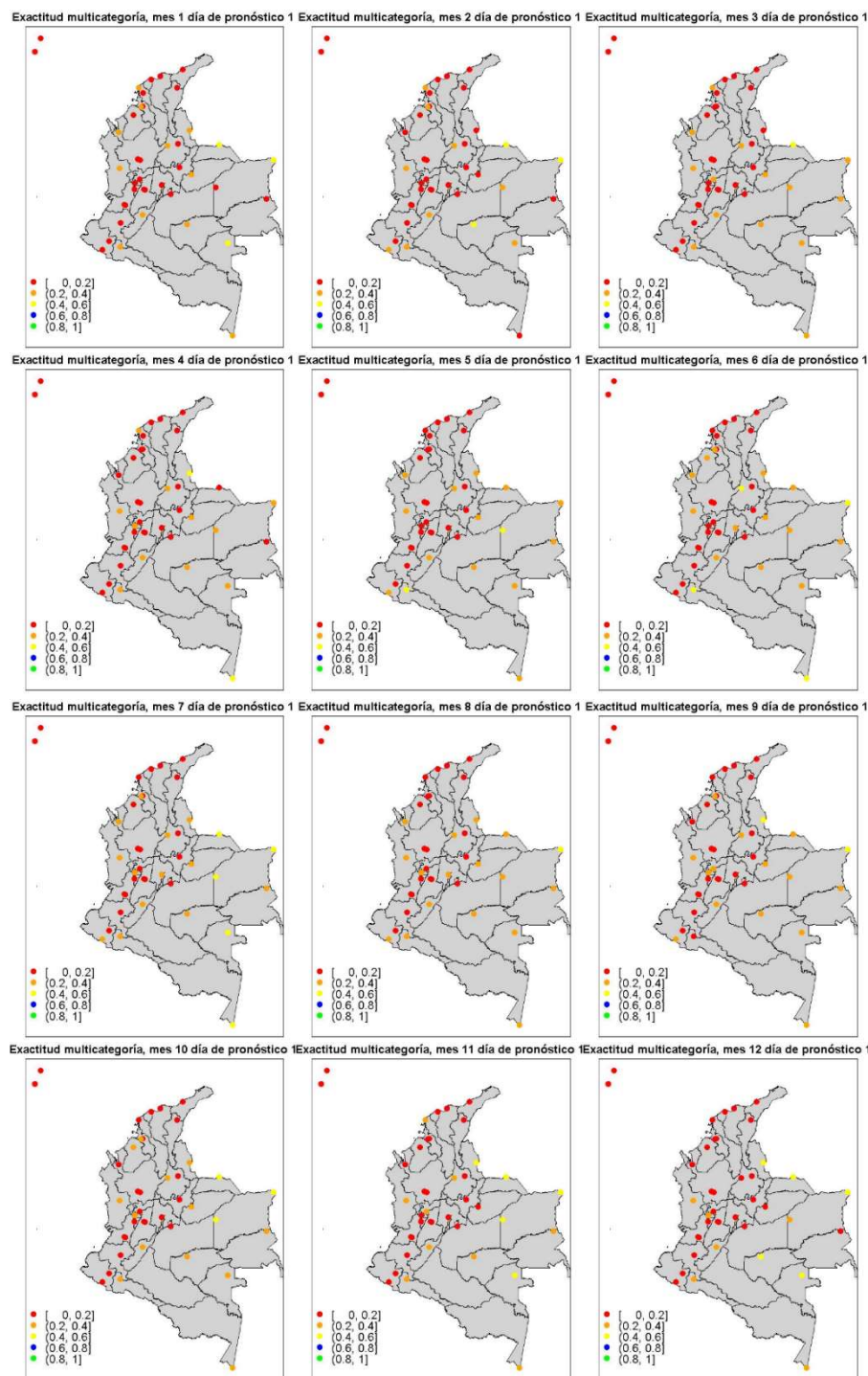
Figura 14. Asimetría del error de pronóstico de precipitación por meses.



Anexo 3. Análisis temperatura máxima

Anexo 3.1. Métricas de verificación de pronóstico multicategoría temperatura máxima

Figura 15. Exactitud de pronóstico multicategoría de temperatura máxima por meses.



Anexo 3.2. Métricas de verificación de pronóstico de temperatura máxima (variable continua)

Figura 16. Error medio (Bias aditivo) de pronóstico de temperatura máxima por meses.

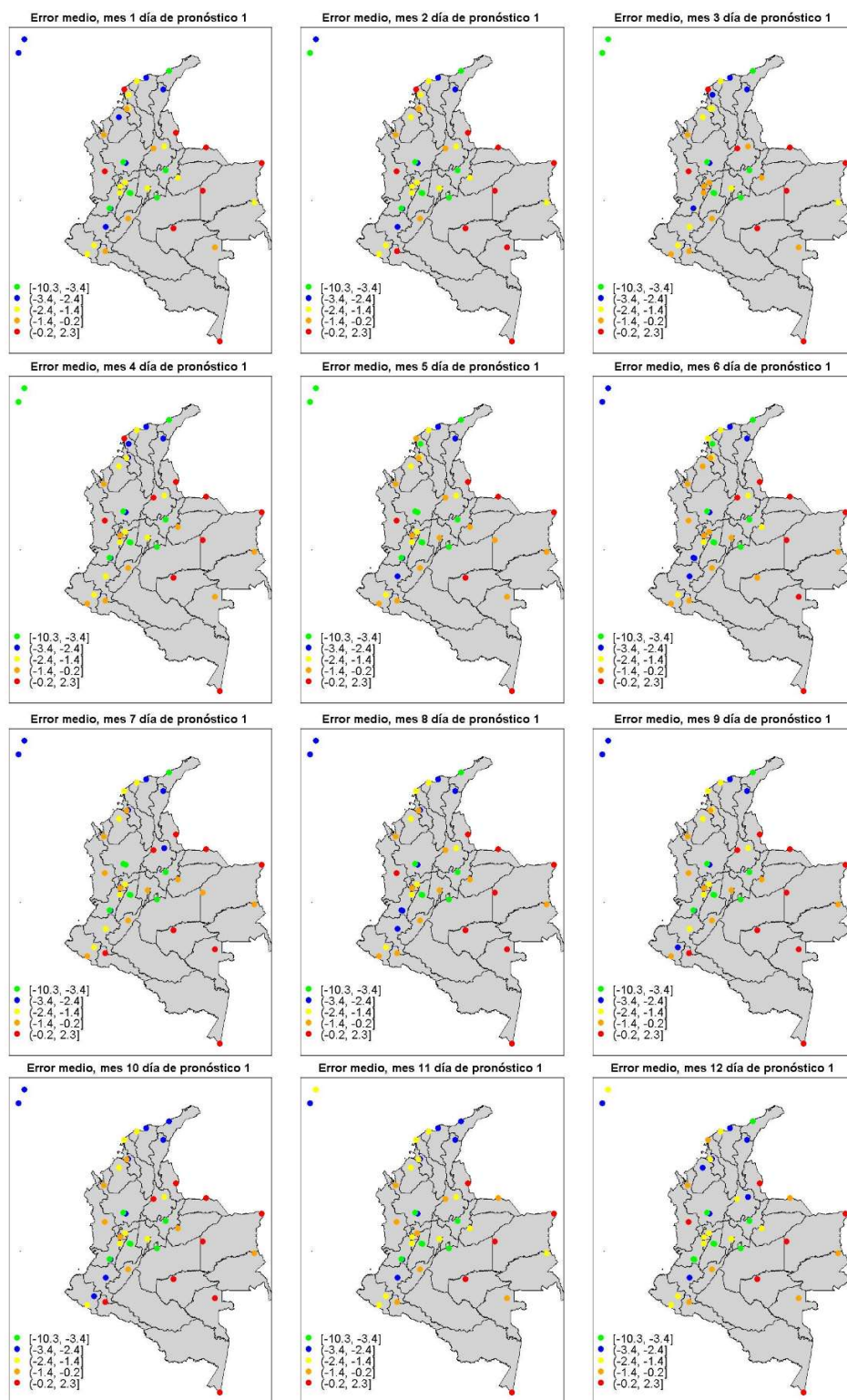


Figura 17. Raíz Error cuadrático medio de pronóstico de temperatura máxima por meses.

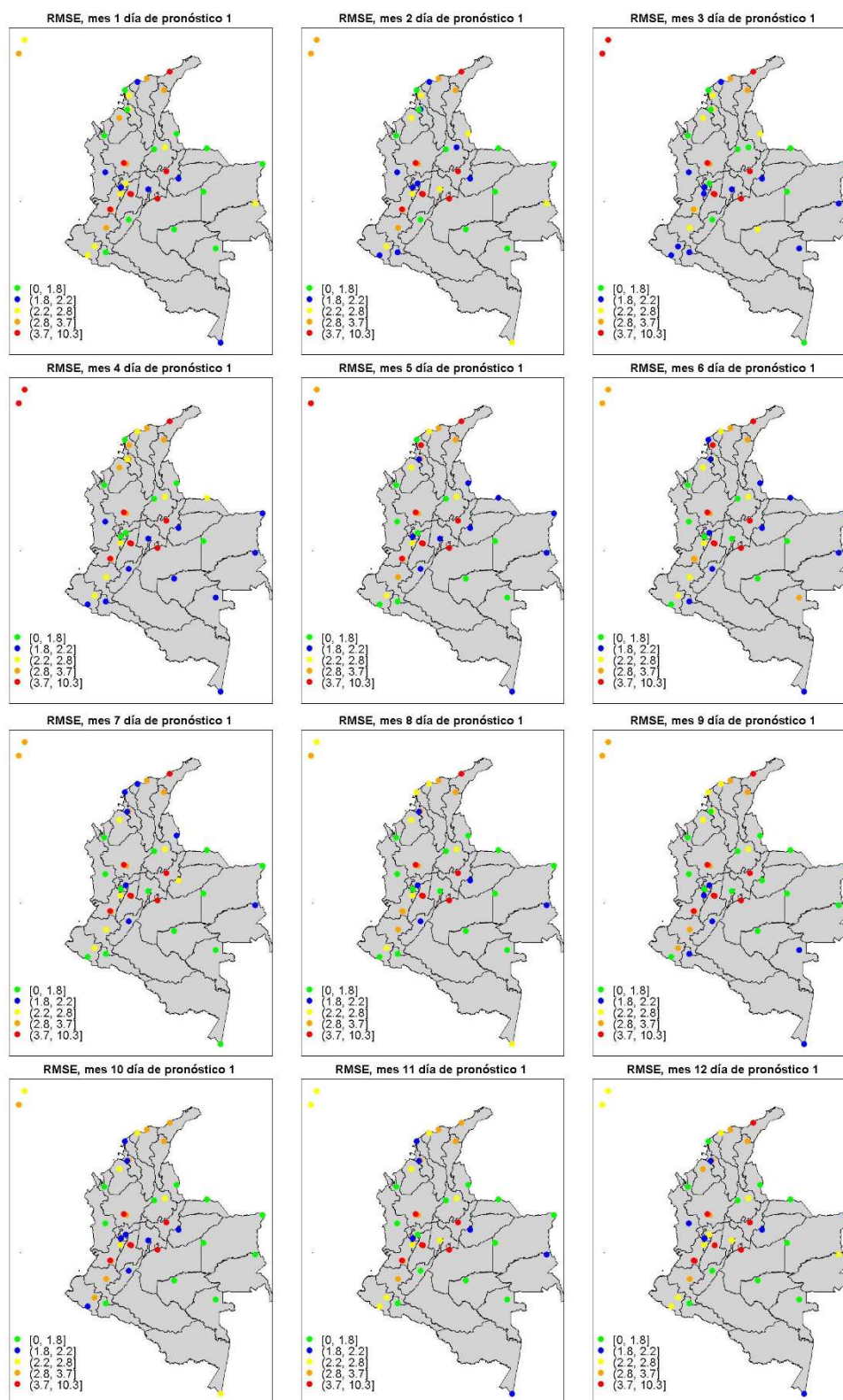
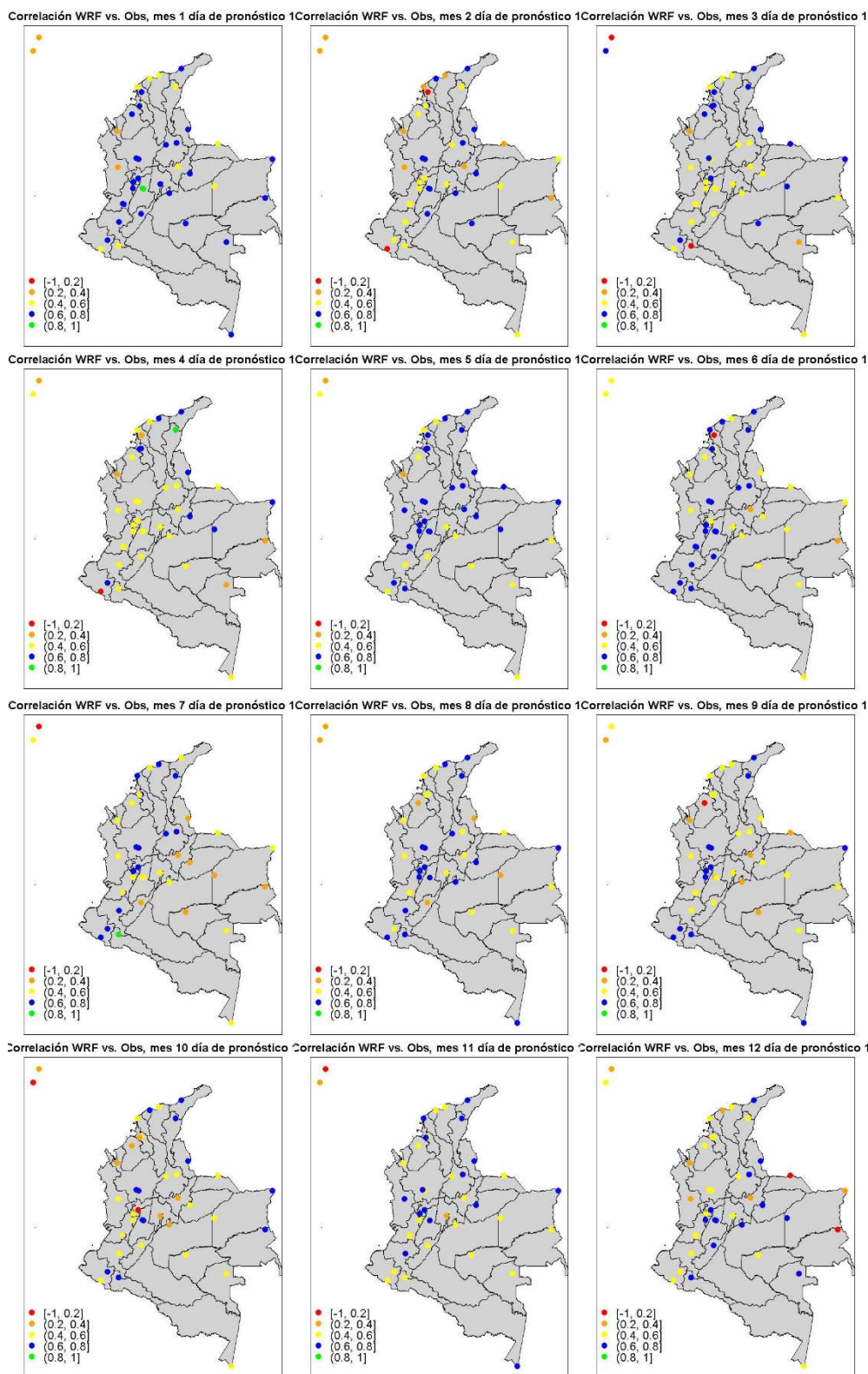


Figura 18. Correlación pronóstico Vs. observaciones de temperatura máxima por meses.



Anexo 3.3. Análisis de supuestos de error aleatorio de pronóstico de temperatura máxima

Figura 19. Prueba Jarque – Bera (H_0 : Distribución normal de los errores de pronóstico de temp. máxima) por meses.

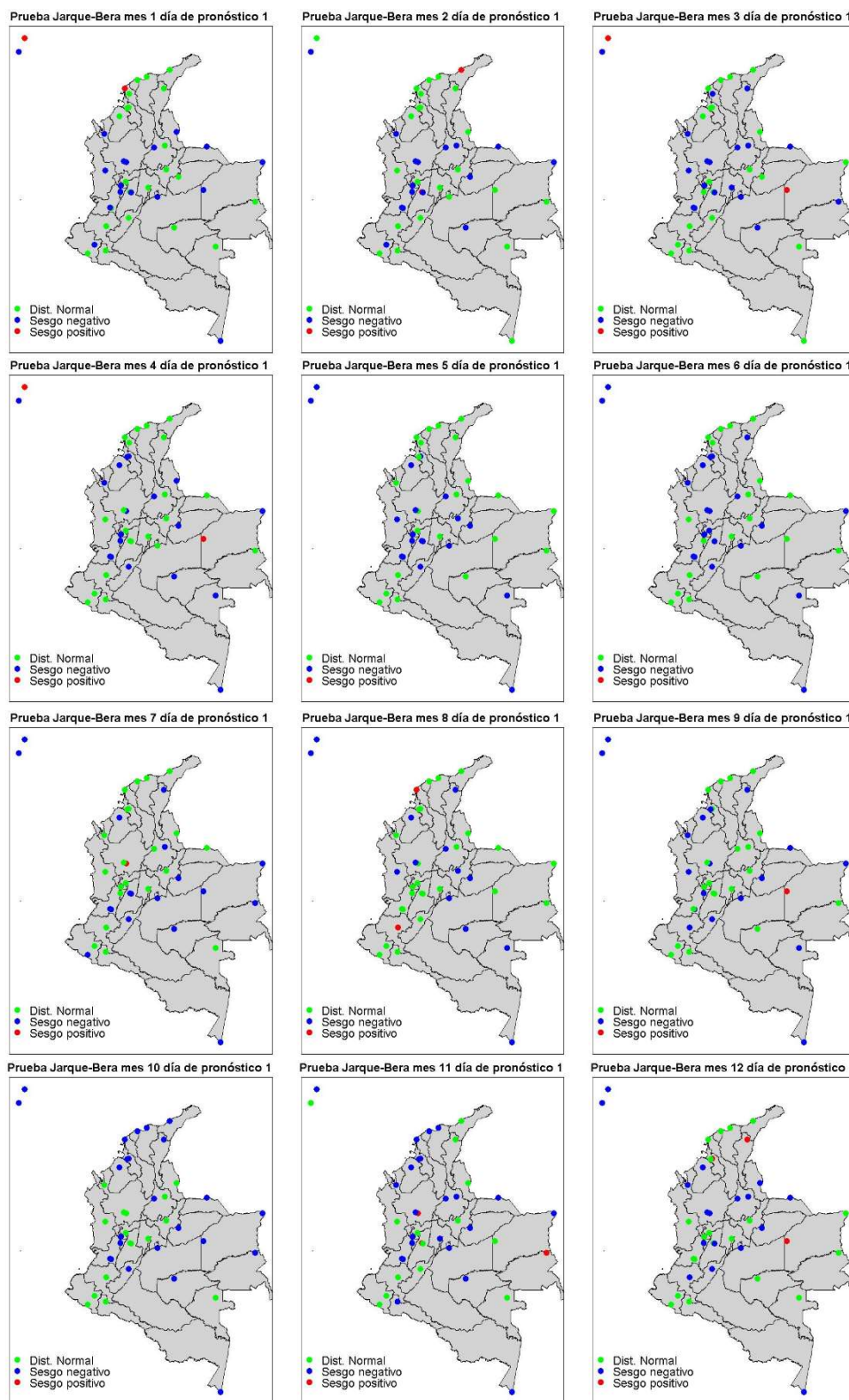


Figura 20. Prueba t-student ($H_0: \bar{X}_{error} = 0$) por meses (temperatura máxima).

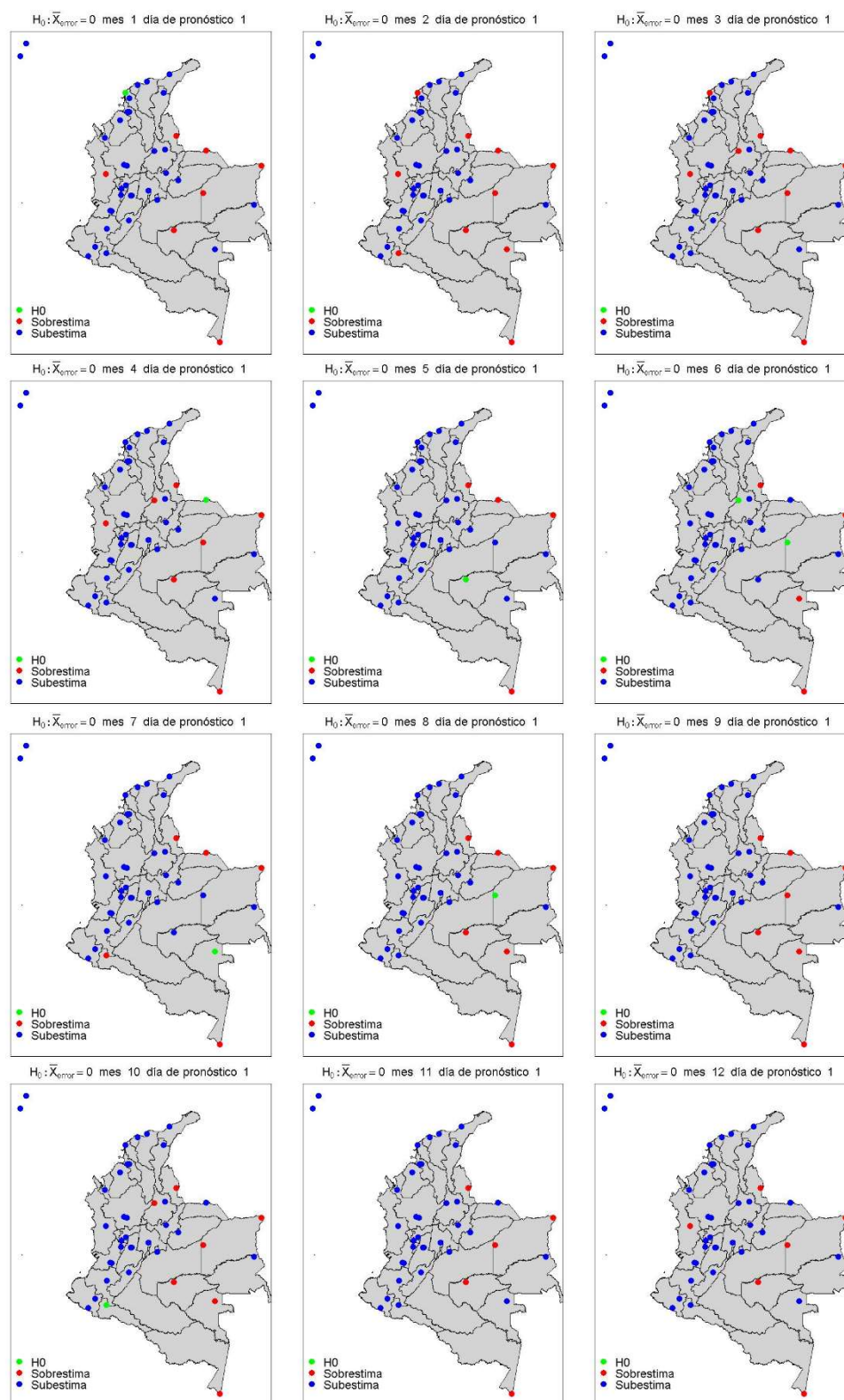


Figura 21. Prueba de rango signado de Wilcoxon ($H_0: \text{Mediana}_{\text{error}} = 0$) por meses (temperatura máxima).

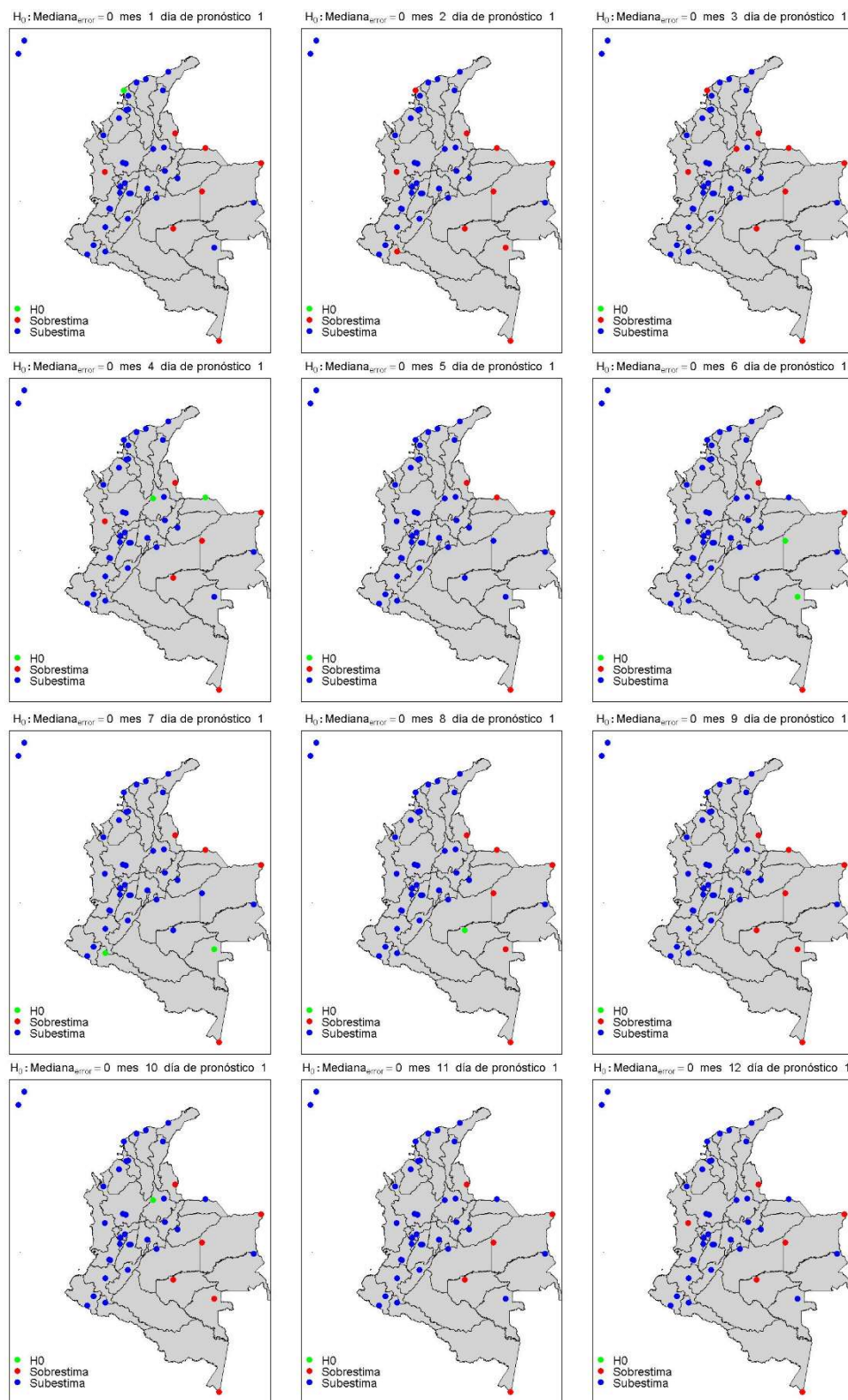


Figura 22. Prueba Breusch – Pagan (H_0 : Varianza constante del error) por meses (temperatura máxima).

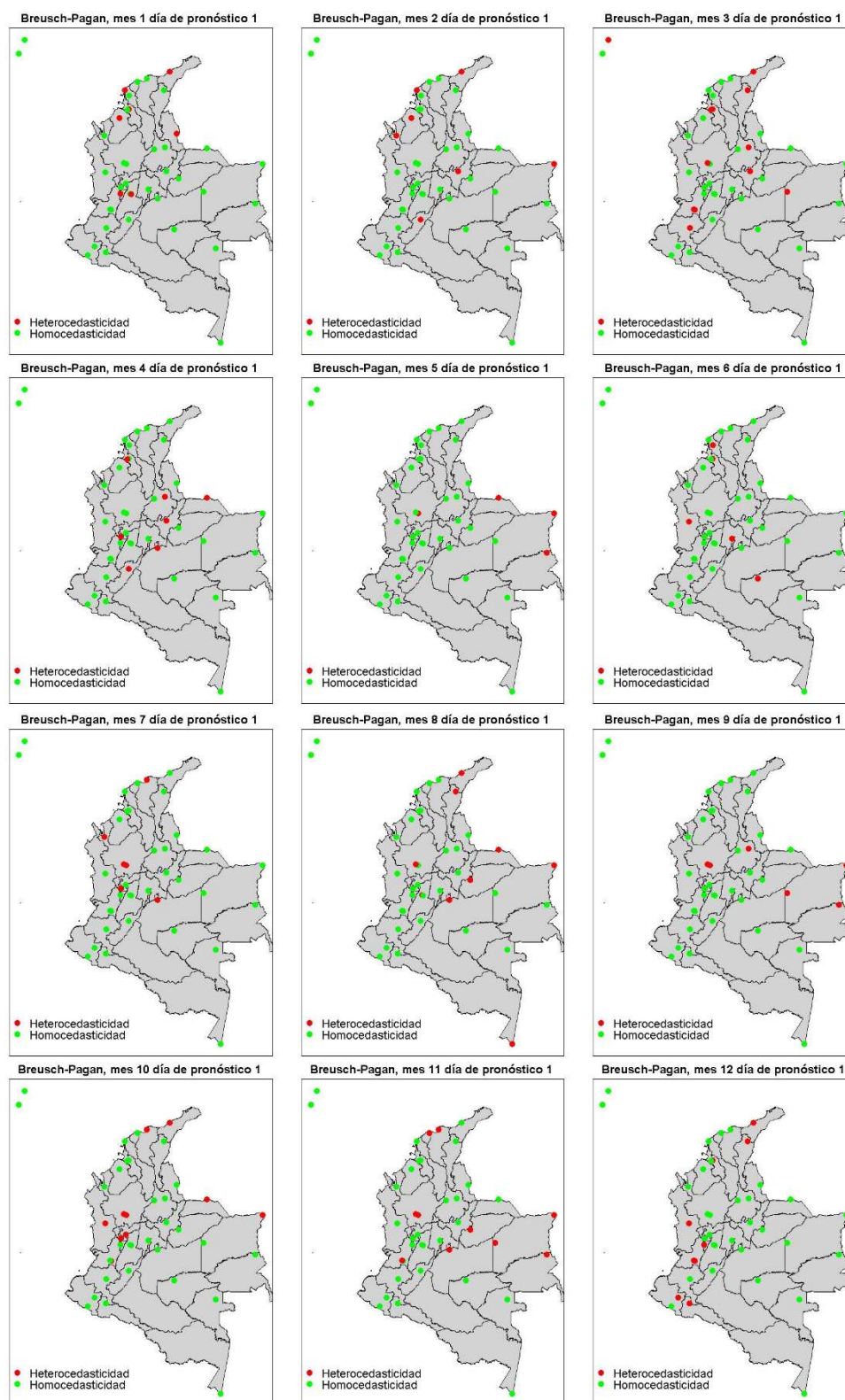
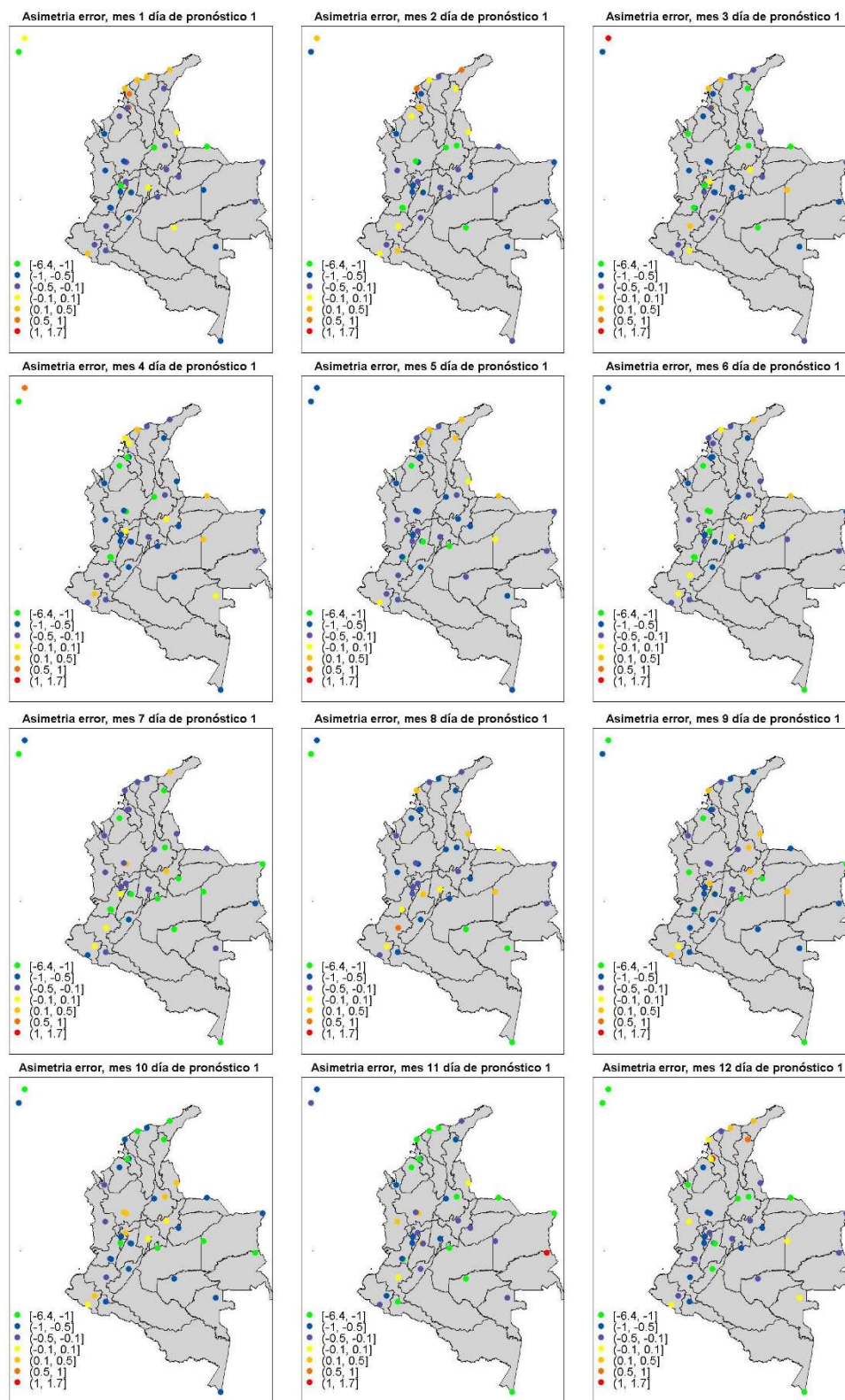


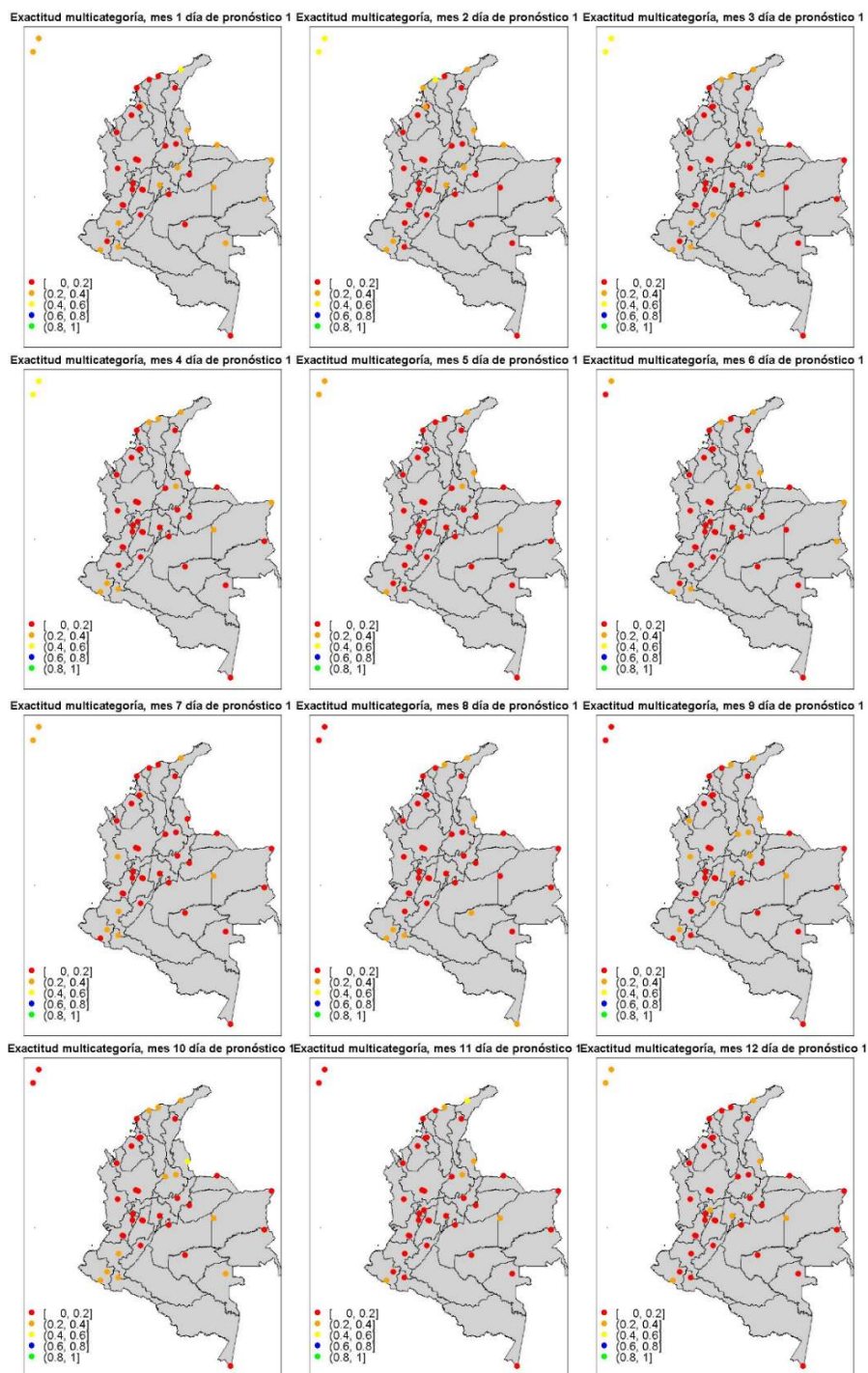
Figura 23. Asimetría del error de pronóstico por meses (temperatura máxima).



Anexo 4. Análisis temperatura mínima

Anexo 4.1. Métricas de verificación de pronóstico multicategoría temperatura mínima

Figura 24. Exactitud de pronóstico multicategoría de temperatura mínima por meses.



Anexo 4.2. Métricas de verificación de pronóstico de temperatura mínima (variable continua)

Figura 25. Error medio (Bias aditivo) de pronóstico de temperatura mínima por meses.

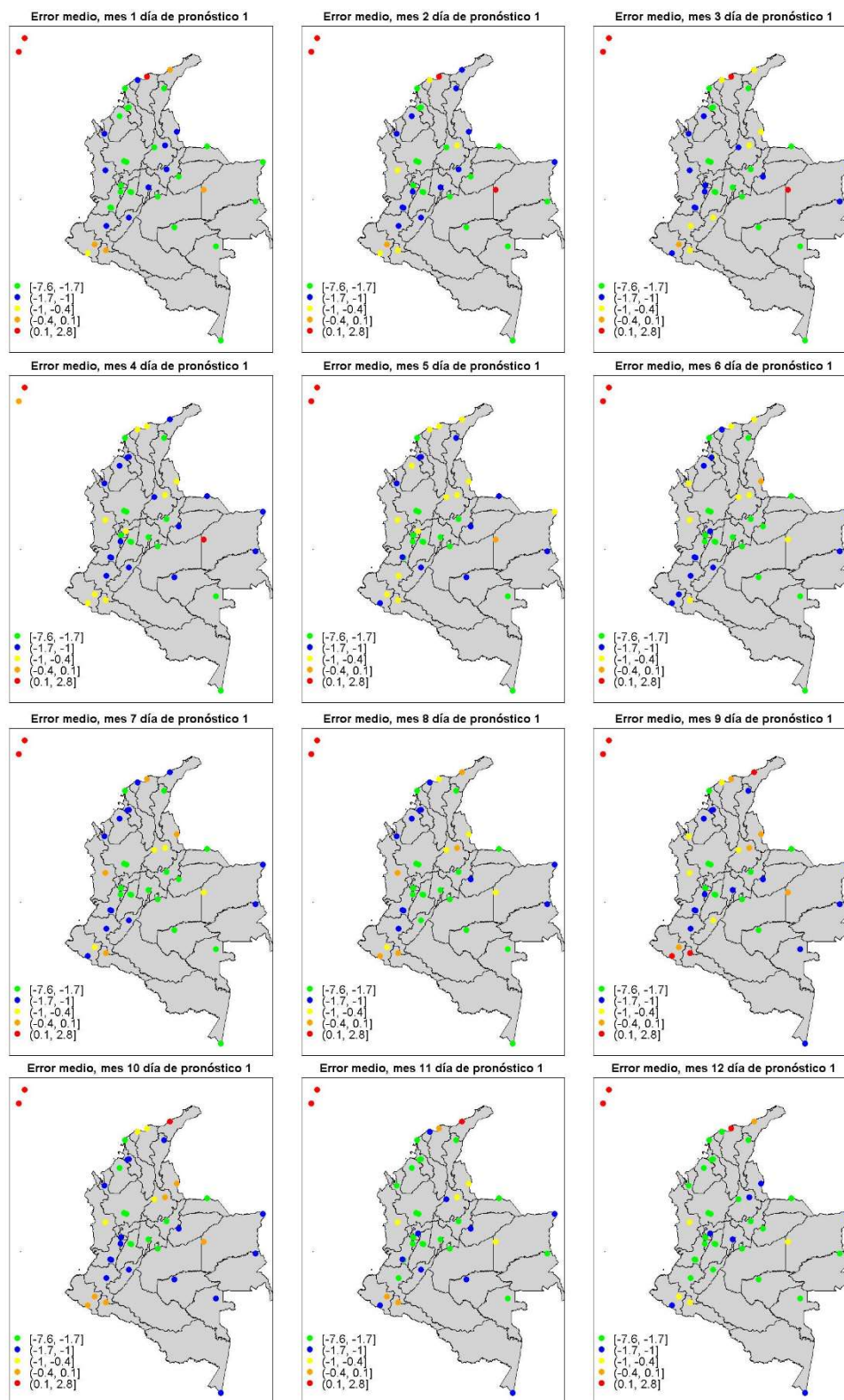


Figura 26. Raíz Error cuadrático medio de pronóstico de temperatura mínima por meses.

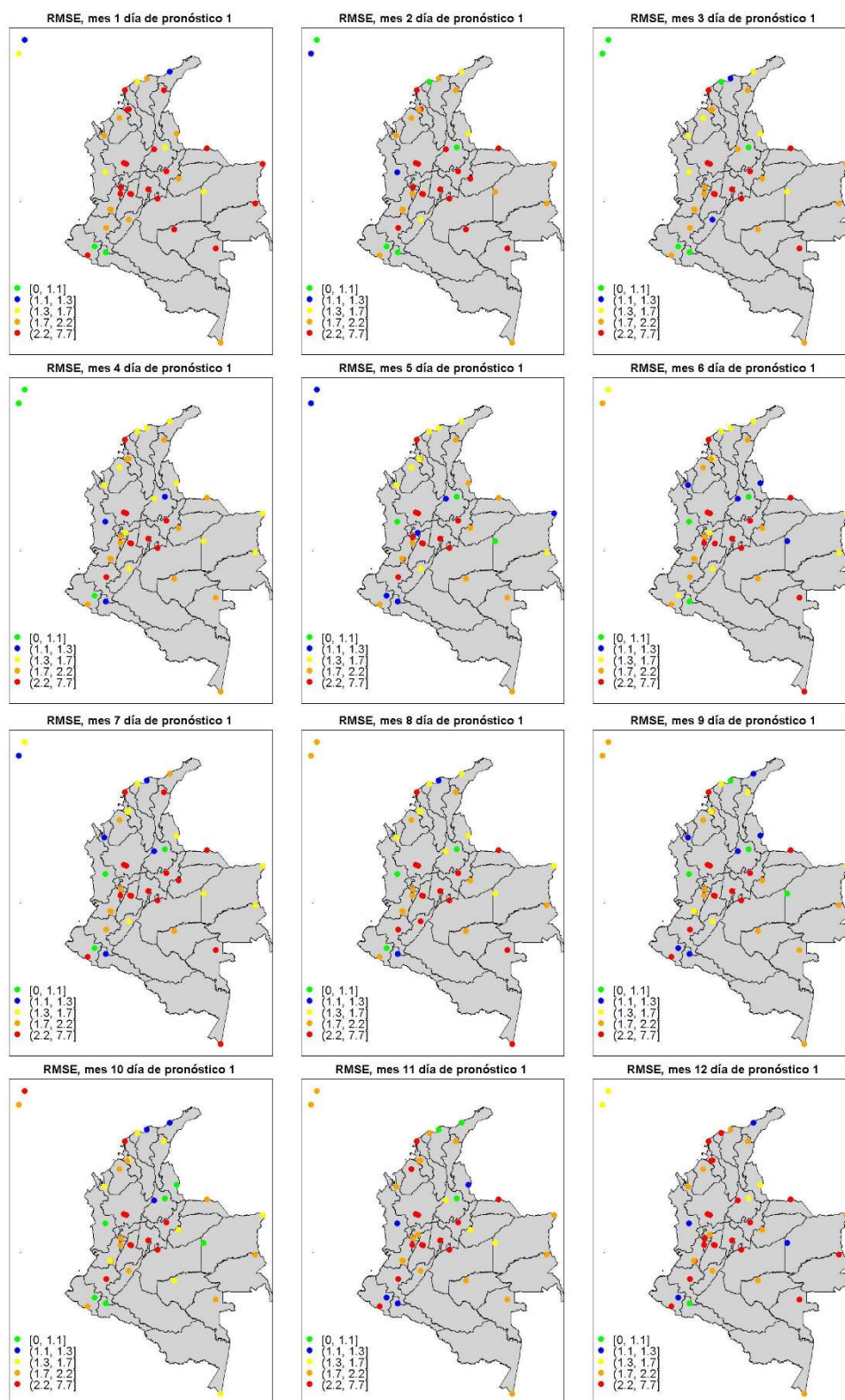
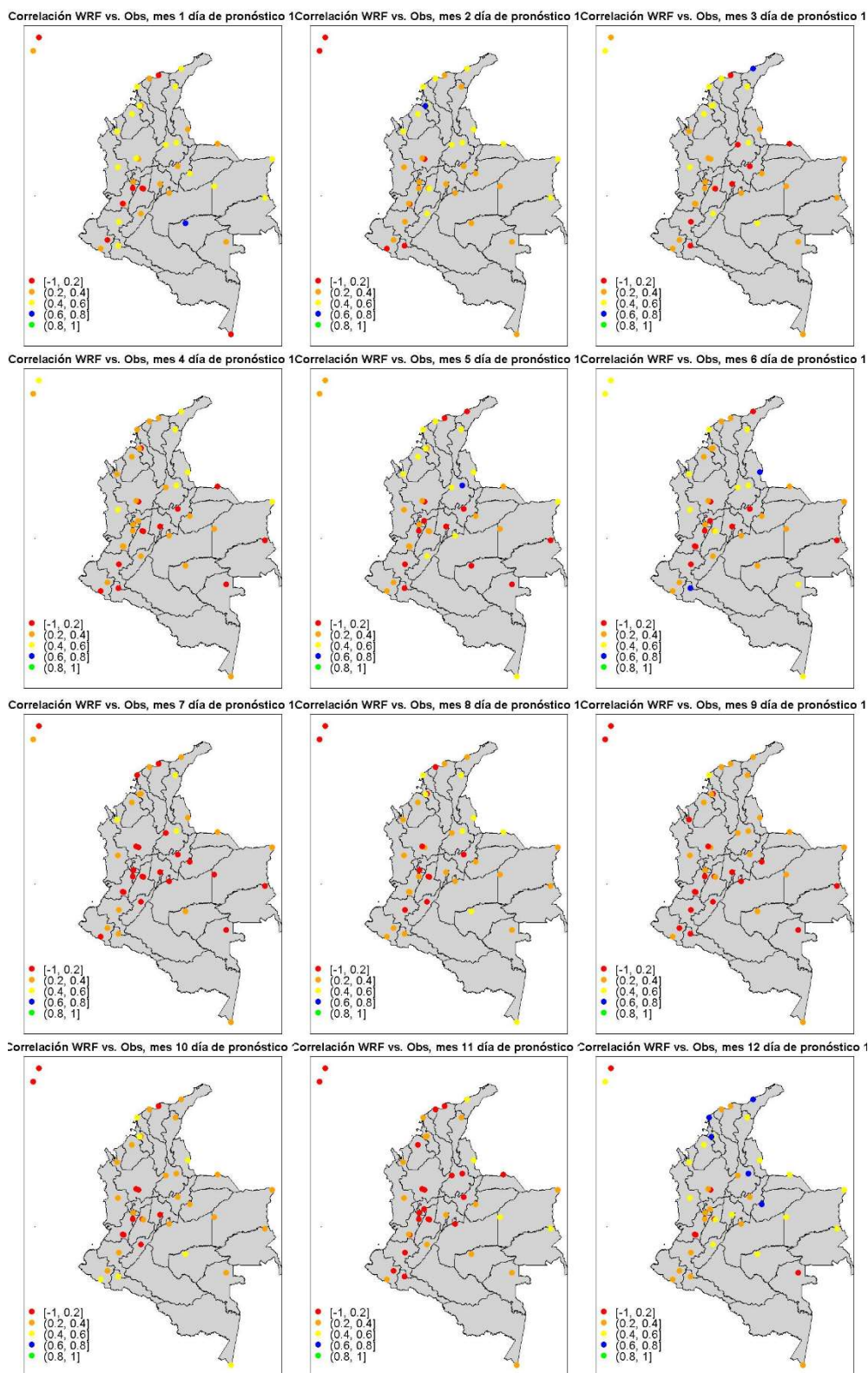


Figura 27. Correlación pronóstico Vs. observaciones de temperatura mínima por meses.



Anexo 4.3. Análisis de supuestos de error aleatorio de pronóstico de temperatura mínima

Figura 28. Prueba Jarque – Bera (H_0 : Distribución normal de los errores de pronóstico de temp. mínima) por meses.

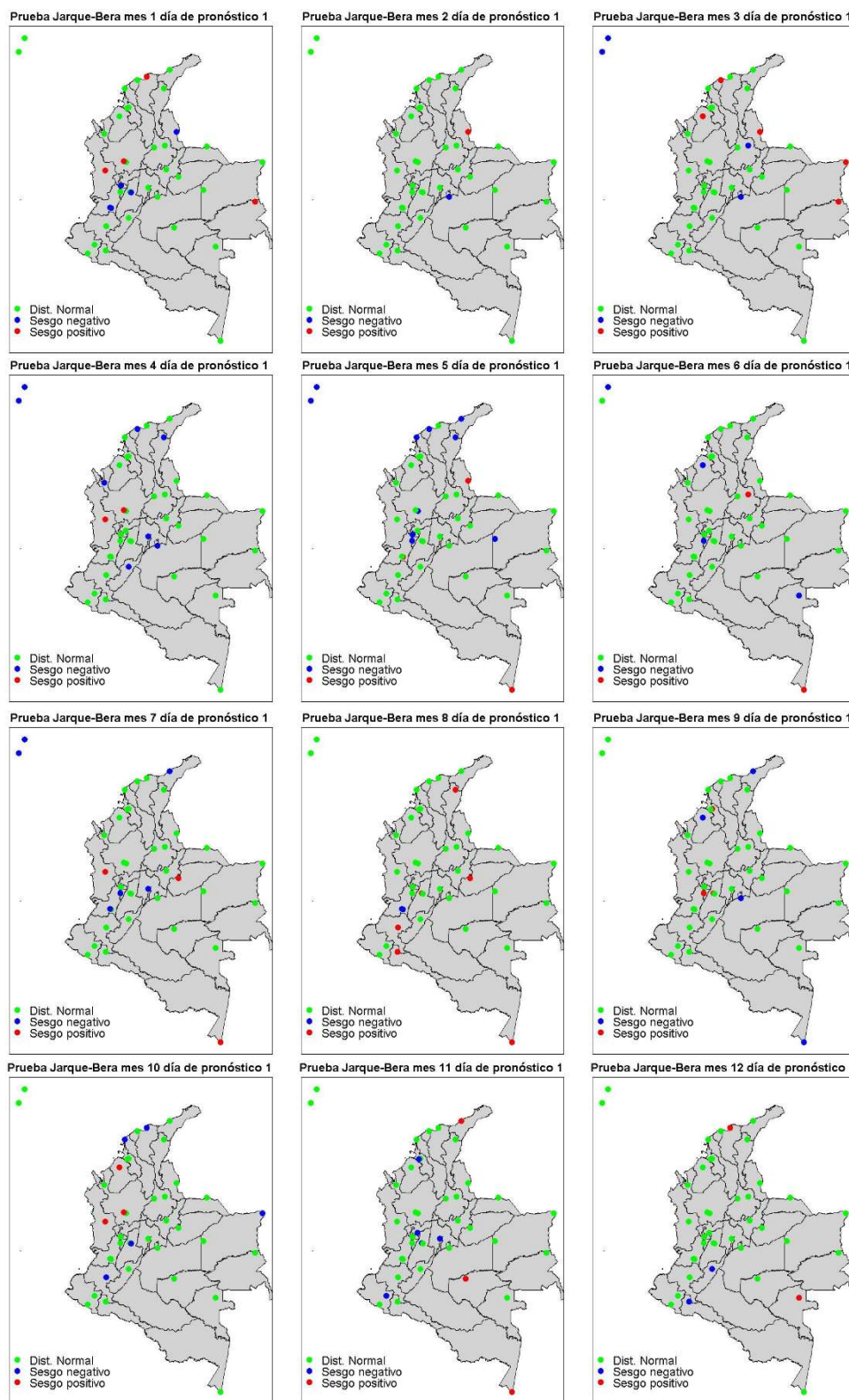


Figura 29. Prueba t-student ($H_0: \bar{X}_{error} = 0$) por meses (temperatura mínima).

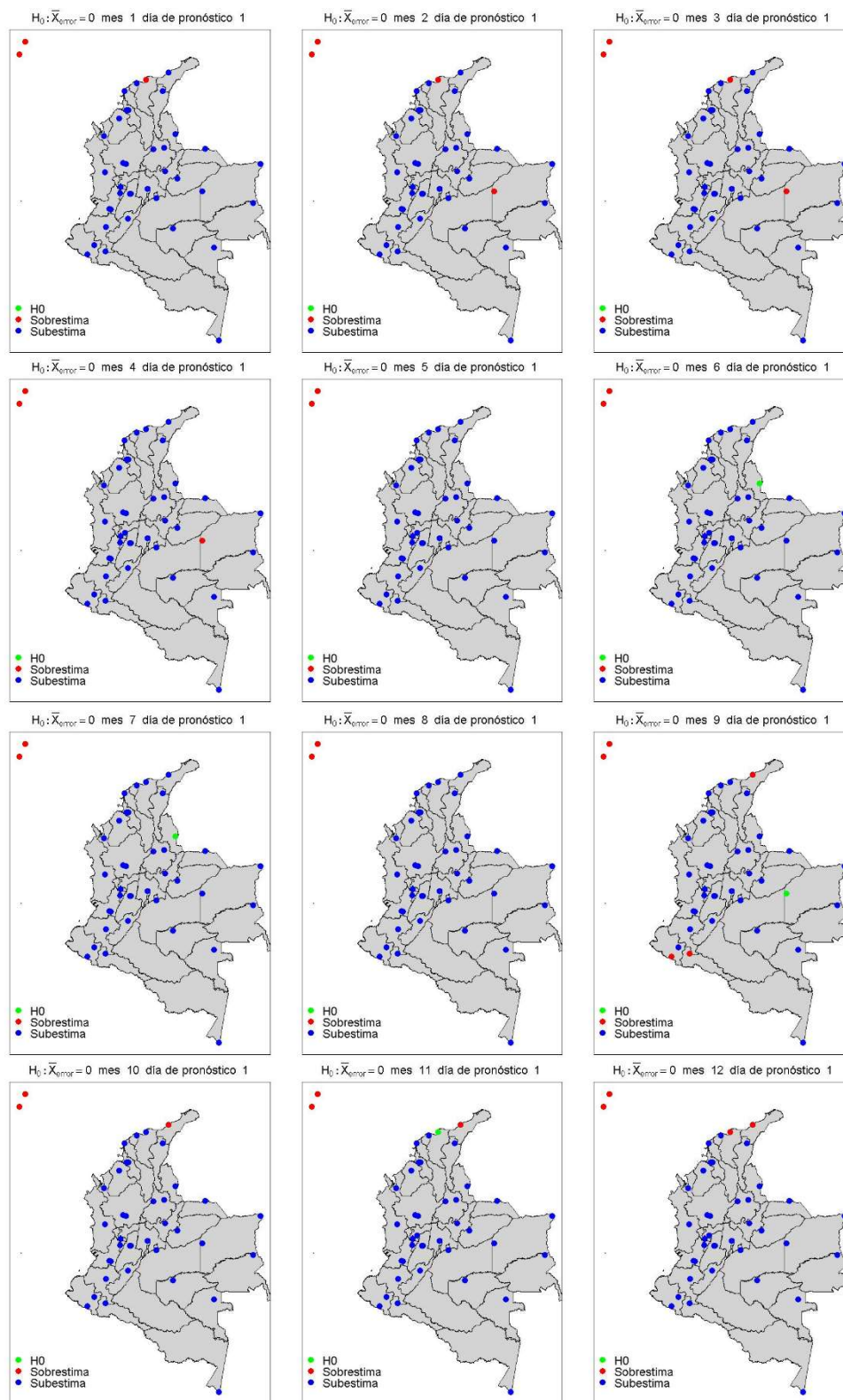


Figura 30. Prueba de rango signado de Wilcoxon ($H_0: \text{Mediana}_{\text{error}} = 0$) por meses (temperatura mínima).

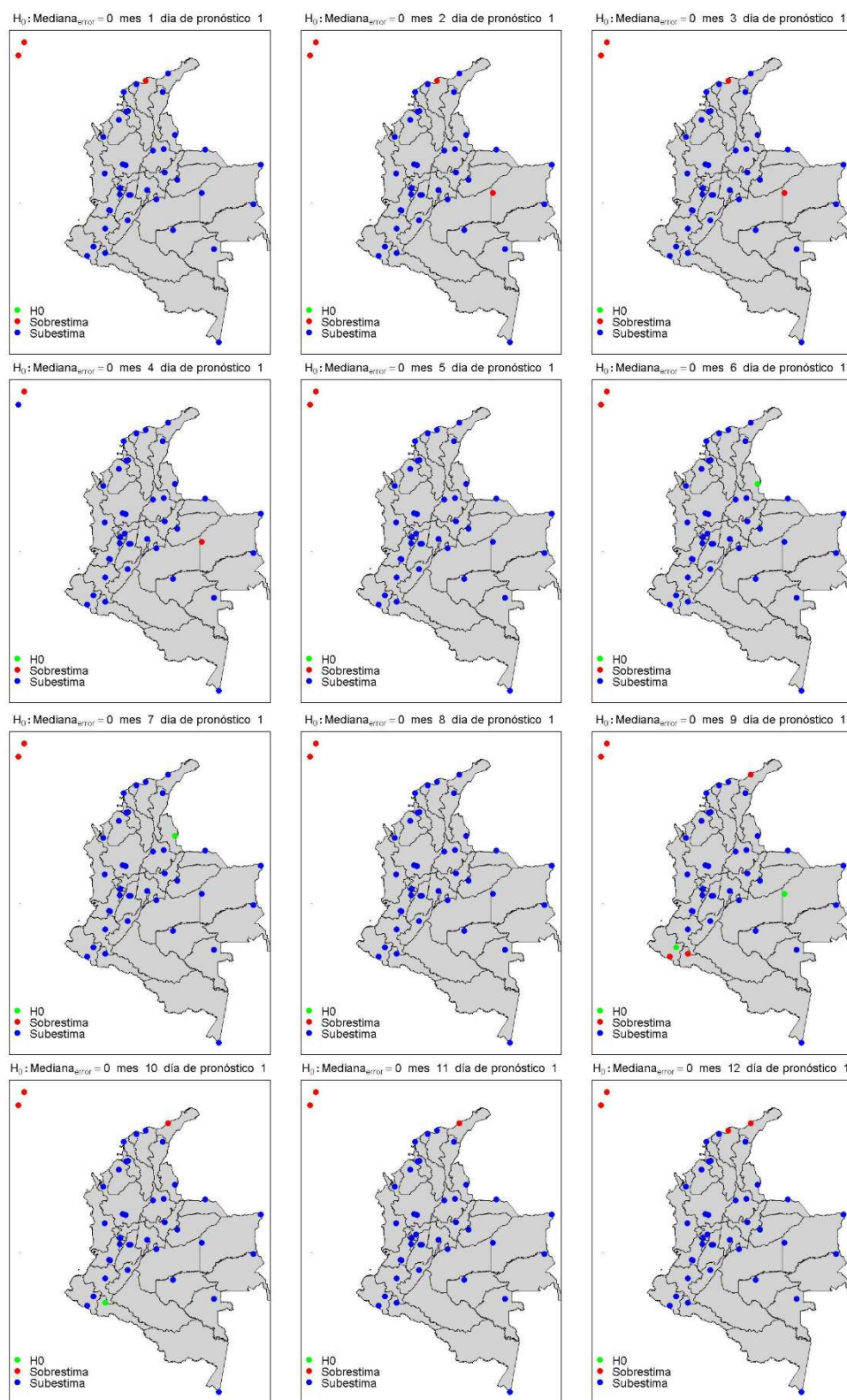


Figura 31. Prueba Breusch – Pagan (H_0 : Varianza constante del error) por meses (temperatura mínima).

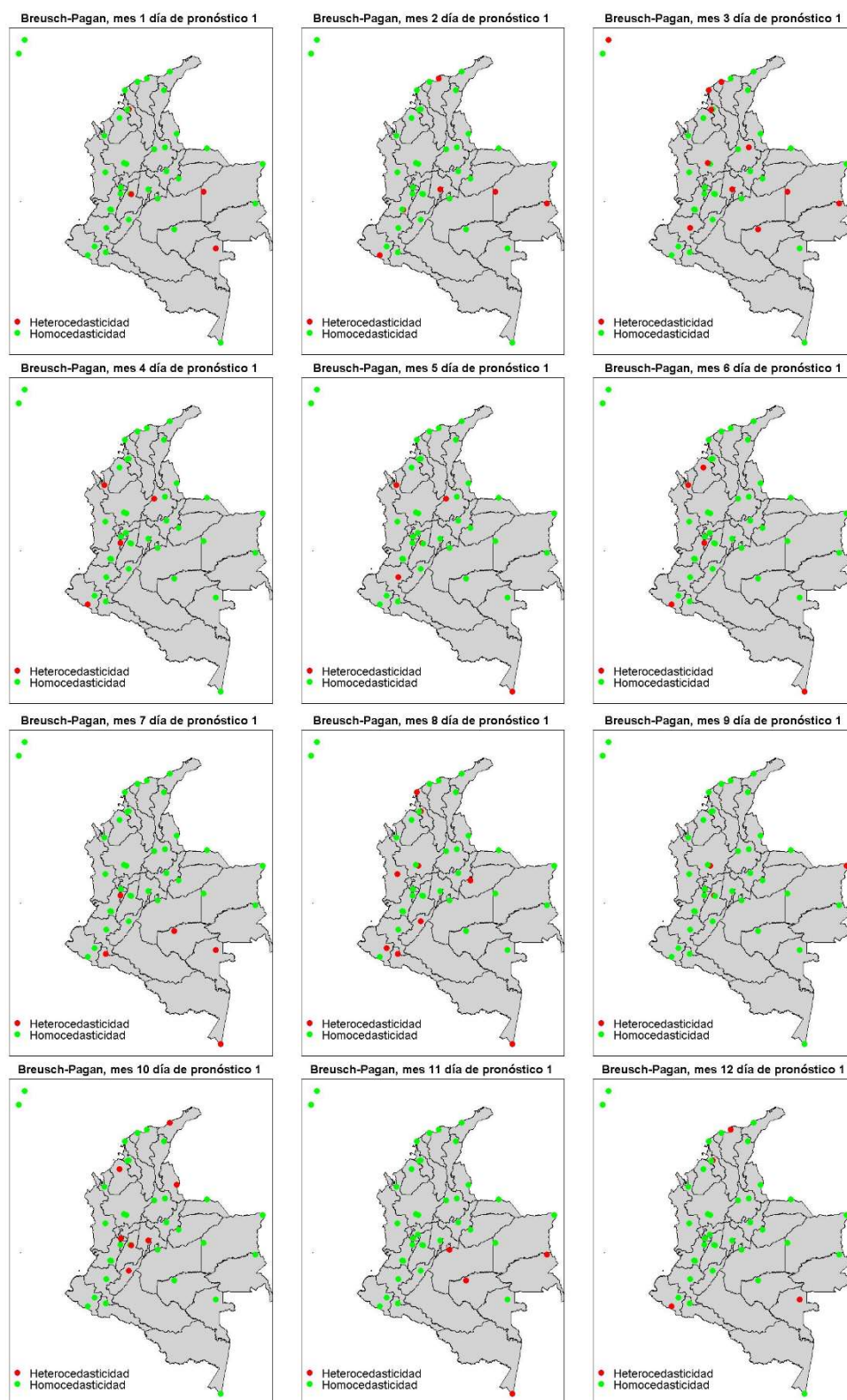


Figura 32. Asimetría del error de pronóstico por meses (temperatura mínima).

