

Nota Técnica sobre la habilidad predictiva que tiene los modelos da la NOAA para predecir la lluvia en Colombia.

2024

Diana Marcela Guzmán Lugo, Jeimmy Yanely Melo Franco, Jose Franklyn Ruiz Murcia.

*Instituto de Hidrología, Meteorología y Estudios Ambientales – IDEAM.
Subdirección de Meteorología.*



Tabla de contenido

Introducción.....	5
1.Modelos y Herramientas para la Predicción Climática del IDEAM.....	7
1.1 Modelos globales del North American Multi-Model ensemble (NMME).....	7
1.2 Herramienta Climate Predictability Tool (CPT)	8
2. Metodología	9
2.1 Recopilación de Datos.....	9
2.2 Calculo de las métricas estadísticas	9
2.3 . Descripción Detallada de las Métricas Estadísticas.....	10
2.4 Interpretación de Resultados.....	13
2.4.1 Correlaciones (pearson, spearman, cc)	13
2.4.2 Errores y Desviaciones (rmse, MAE)	13
2.4.3 Estadísticos de Varianza y Desviación Estándar (var_ratio, p_var, SD):....	13
2.4.4 Habilidad de Pronóstico (hit, hss)	14
2.4.5 Evaluación de Consistencia y Discriminación en las Predicciones Climáticas: Mean Bias, 2AFC, LEPS y Gerrity	14
2.4.6 Evaluadores de Discriminación 2 (roc_bel, roc_above):.....	15
3. Análisis de Resultados	16
3.1 Análisis Integrado de Correlaciones Canónica, Pearson y Spearman	16
3.1.1 Análisis por Correlación Canónica (CC)	16
3.1.2 Análisis de Correlaciones Pearson y Spearman.....	16
3.1.3 Comparación General para las correlaciones	16
3.2 Análisis de errores absolutos medios (MAE) y errores cuadráticos medios (RMSE) por modelo, mes y lead time.....	17
3.2.1 Error absoluto medio(MAE).....	17
3.2.2 Error cuadrático medio (RMSE).....	17
3.2.3 Tendencias generales para MAE y RMSE	17
3.3 Análisis de los Estadísticos de Varianza y Desviación Estándar por Modelo y Mes.....	23
3.3.1 Relación de Varianza (var_ratio):	23
3.3.2 Proporción de Varianza Explicada (p_var):.....	23
3.3.3 Desviación Estándar (SD)	23

3.3.4 Tendencias generales para ratio y proporción de la varianza y desviación estándar.	24
3.4 Análisis sobre Tasa de Aciertos (Hit Rate) y Habilidad Predictiva (HSS) por Mes y Lead Time	26
3.4.1 Hit Rate.	26
3.4.2 HSS.....	26
3.4.3 Tendencias generales para Hit Rate y HSS.....	26
3.5 Evaluación de Consistencia y Discriminación en las Predicciones Climáticas: Análisis de Mean Bias, 2AFC, LEPS y Gerrity.....	29
3.5.1 Mean Bias: Identificación del Sesgo Promedio en las Predicciones.....	29
3.5.2 Análisis del 2AFC: Capacidad del Modelo para Diferenciar Eventos Binarios	30
3.5.3 Análisis de la Precisión de las Predicciones en el Espacio de Probabilidad (LEPS).	33
3.5.4 Análisis del Estadístico Habilidad de Clasificación en Eventos Ordenados (Gerrity Score) por Mes y Modelo	35
3.6 Análisis Resumido de los ROC (ROC Above y ROC Below) por Mes y Lead Time	37
Conclusiones.....	40
Recomendaciones.....	40
Referencias	41

Lista de figuras

Figura 1. Comparación de la Correlación Canónica (CC) por Modelo) y Mes (arriba, y Lead Time (abajo)	18
Figura 2. Correlaciones Pearson y Spearman por Modelo y Lead Time – Primer Semestre (Enero-Junio)	
Figura 3. Correlaciones Pearson y Spearman por Modelo y Lead Time – Segundo Semestre (Julio-Diciembre)	19
Figura 4. Desempeño de Modelos en MAE y RMSE durante el Primer Semestre (Enero-Junio).....	21
Figura 5. Desempeño de Modelos en MAE y RMSE durante el Segundo Semestre (Julio-Diciembre).	22
Figura 6. Mapa de Calor del Ratio de Varianza (var_ratio) por Modelo y Mes	24
Figura 7. Mapa de Calor del Porcentaje de Varianza Explicada (p_var) por Modelo y Mes	25

Figura 8 Mapa de Calor de Desviación Estándar (SD) por Modelo y Mes	25
Figura 9. Comparativo de Habilidad de Acierto (HSS) y Tasa de Aciertos (Hit Rate) por Modelo, Mes y Lead Time (Enero - Junio).....	27
Figura 10. Comparativo de Habilidad de Acierto (HSS) y Tasa de Aciertos (Hit Rate) por Modelo, Mes y Lead Time (Julio - Diciembre).....	28
Figura 11. Comparativo del Sesgo por Mes y Lead Time.....	31
Figura 12. Comparativo del estadístico 2AFC por Mes y Lead Time	32
Figura 13. Comparativo del estadístico LEPS por Mes y Lead Time	34
Figura 14. Comparativo del estadístico Gerrity Score por Mes y Lead Time.....	36
Figura 15. Comparativo de ROC Below (ROCBEL) y ROC Above (ROCABO) por Mes y Lead Time (Enero - Junio)	38
Figura 16. Comparativo de ROC Below (ROCBEL) y ROC Above (ROCABO) por Mes y Lead Time (Julio - Diciembre)	39

Introducción

La validación estadística es un componente fundamental en la predicción climática, ya que permite evaluar la precisión y fiabilidad de los modelos predictivos. Estas predicciones son esenciales en la toma de decisiones en sectores como la agricultura, la gestión de recursos hídricos, la energía, y la reducción del riesgo de desastres. Sin embargo, la eficacia de estas predicciones depende de su precisión y fiabilidad, lo que hace que la verificación estadística sea una práctica indispensable.

La verificación estadística implica comparar las predicciones del modelo con los datos observados y calcular una serie de métricas que revelan las fortalezas y debilidades de los modelos. Según la guía de la Organización Meteorológica Mundial (OMM), la verificación de las predicciones climáticas operacionales es importante para mantener la confianza de los usuarios en las predicciones y para garantizar que se utilicen eficazmente en la toma de decisiones (OMM, 2023). Además, la verificación no solo ayuda a cuantificar la calidad de las predicciones, sino que también es esencial para el desarrollo continuo de los modelos y la mejora de su precisión (Manrique-Suñén, Gonzalez-Reviriego, Torralba, Cortesi, & Doblas-Reyes, 2020).

Este documento presenta un análisis sobre la capacidad predictiva de los modelos climáticos utilizados por el **Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM)** de Colombia. La verificación estadística evaluará cómo los modelos predicen la lluvia en el territorio colombiano, así como también identificar fortalezas y debilidades en los modelos: Esta verificación incluyó medidas como correlación de Pearson y Spearman, el RMSE, y el MAE. Además, se analizaron métricas de discriminación y habilidad predictiva, como el HSS, Hit Rate y 2AFC, esenciales para evaluar la capacidad de los modelos de predecir eventos extremos y comportamientos no lineales.

Los resultados permitieron identificar los **mejores modelos** en función de estas métricas. Los modelos **CFSv2** y **GFDL** se destacaron en los **primeros meses del año (enero-marzo)** y hacia el **final del año (noviembre-diciembre)**, con correlaciones entre **0.3 y 0.6**. Estos modelos demostraron **estabilidad en lead times cortos (1-3 meses)**, aunque en **horizontes largos (6-9 meses)** se observó una **disminución en su rendimiento**, un comportamiento esperado estadísticamente debido al aumento de la incertidumbre en predicciones a largo plazo. Por otro lado, los modelos **NMME** y **CMC1** presentaron mayor dispersión en meses como **abril y octubre**, afectando la precisión de sus predicciones, lo que evidencia la necesidad de ajustes en sus algoritmos para estos periodos de mayor variabilidad.

En resumen, los resultados de esta verificación guiarán los **ajustes y mejoras** en los modelos climáticos que alimentan la **Climate Predictability Tool (CPT)**. Esto permitirá mejorar la confiabilidad y utilidad de las predicciones para los usuarios finales y una mejor gestión de los recursos, así como una respuesta más efectiva ante **eventos climáticos extremos**.

1. Modelos y Herramientas para la Predicción Climática del IDEAM

1.1 Modelos globales del North American Multi-Model ensemble (NMME)

El Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) de Colombia utiliza un conjunto de modelos climáticos globales proporcionados por el North American Multi-Model Ensemble (NMME) de la NOAA (National Oceanic and Atmospheric Administration) para realizar predicciones climáticas. El NMME es una iniciativa colaborativa internacional que integra varios modelos climáticos de diversas instituciones para mejorar la precisión y confiabilidad de las predicciones. Los modelos utilizados incluyen:

- **NCAR CCSM4 (National Center for Atmospheric Research Community Climate System Model 4):** Desarrollado por el Centro Nacional para la Investigación Atmosférica, este modelo integra componentes atmosféricos, oceánicos, de hielo y de tierra para simular el sistema climático global.
- **NASA (National Aeronautics and Space Administration):** Modelo desarrollado por la NASA, que contribuye al NMME con simulaciones que incorporan datos satelitales para mejorar las predicciones climáticas.
- **GFDL (Geophysical Fluid Dynamics Laboratory):** Modelo del Laboratorio de Dinámica de Fluidos Geofísicos, centrado en la simulación de las interacciones entre la atmósfera, los océanos y el hielo, lo que permite una representación precisa de los fenómenos climáticos.
- **CMC1 y CMC2 (Canadian Meteorological Centre Models 1 y 2):** Modelos desarrollados por el Centro Meteorológico Canadiense, que contribuyen con dos versiones diferentes para captar diversas incertidumbres en las predicciones climáticas.
- **CFSv2 (Climate Forecast System Version 2):** Desarrollado por el Centro de Predicción Climática de la NOAA, este modelo es ampliamente utilizado para predicciones climáticas a escala mensual y estacional, y es una de las piezas fundamentales del sistema NMME.

Para procesar y analizar los datos de estos modelos, el IDEAM utiliza la CPT, que, no solo facilita la generación de predicciones, sino que también ofrece capacidades robustas para la verificación y validación de los modelos predictivos.

Los datos que se usan de estos modelos provienen de dos fuentes principales: Los datos de **Hindcast**, que permiten evaluar la precisión de los modelos en la predicción de eventos climáticos pasados, y los datos de **Forecast**, utilizados para las predicciones operativas de los próximos seis meses. Es importante señalar que, debido a las políticas de manejo de datos de la NOAA, la información anterior a 2021 no está disponible públicamente. Aunque esto puede limitar algunos análisis de largo plazo, las predicciones recientes siguen proporcionando información importante para llevar a cabo la verificación estadística de los modelos mencionados.

Estos modelos generan predicciones con un horizonte de seis meses, ofreciendo diversos escenarios que capturan las incertidumbres inherentes a las predicciones climáticas. Las ventajas de utilizar el NMME incluyen su capacidad para proporcionar predicciones globales y regionales, la actualización continua de datos, la compatibilidad con herramientas de procesamiento como la CPT, y su historial de desempeño confiable a nivel internacional.

1.2 Herramienta Climate Predictability Tool (CPT)

Para el procesamiento y validación de los datos obtenidos de los modelos NMME, el IDEAM utiliza la CPT, desarrollada por el IRI. Este proceso incluye la evaluación de múltiples estadísticos, como la correlación de Pearson, Spearman, RMSE, entre otros que se explican en mayor detalle en el numeral 2 (Metodología), para cada uno de los modelos y sus diferentes horizontes de predicción (lead times).

Así, La CPT no solo permite la generación de predicciones, sino que también ofrece herramientas para la verificación y validación de los modelos predictivos. Una de las grandes ventajas de la CPT es su accesibilidad y usabilidad, combinada con su capacidad para procesar grandes volúmenes de datos que la convierten en una herramienta ideal para las necesidades del IDEAM, permitiendo una mejora continua en la precisión y confiabilidad de las predicciones climáticas del país a través de la verificación estadística.

2. Metodología

En esta sección, se describe el proceso general utilizado para la verificación estadística de las predicciones climáticas generadas por el IDEAM. Este proceso permite evaluar la precisión, confiabilidad y utilidad de las predicciones, permitiendo ajustes continuos y mejoras en los modelos utilizados.

2.1 Recopilación de Datos

El primer paso en el proceso de verificación estadística consiste en la recopilación de datos de observación y predicción. Para ello, se utiliza el conjunto de datos CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data), que proporciona información de precipitación mensual para el período 1981-2024. CHIRPS combina datos de estaciones meteorológicas con estimaciones de satélite para todo el dominio de Colombia.

Por otro lado, las predicciones se obtienen de los modelos climáticos globales del North American Multi-Model Ensemble (NMME), descritos en detalle en la sección 1.1. Estos modelos, que incluyen el NCAR CCSM4, NASA, GFDL, CMC1, CMC2 y CFSv2, proporcionan un conjunto diverso de predicciones climáticas. La comparación entre los datos observados de CHIRPS y las predicciones de los modelos NMME forma la base del proceso de verificación

2.2 Calculo de las métricas estadísticas

El cálculo de los estadísticos para la verificación de la predicción de precipitación se realizó mediante un proceso automatizado utilizando un script en Bash. Este proceso incluye los siguientes pasos clave:

- 1. Preparación de Directorios y Listas de Modelos, Meses, y Años:** Se definieron las rutas a los directorios donde se encuentran los archivos de predicción (hindcast) y observación (forecast). Además, se configuraron listas de modelos de predicción, meses del año, años de hindcast (1991 y 1992) y forecast (2023), y los lead times correspondientes (de 1 a 6 meses).
- 2. Configuración y Ejecución del Proceso de Verificación:** Se copian y configuran los archivos necesarios para ejecutar el Climate Predictability Tool (CPT). Luego, para cada combinación de modelo, año de hindcast y lead time, se construyen las rutas a los archivos de predicción y forecast, y se ejecuta el CPT para calcular los estadísticos.
- 3. Cálculo de Estadísticos:** Los estadísticos calculados incluyen desviación estándar, correlaciones de Pearson y Spearman, 2AFC, proporción de varianza, relación de varianza, sesgo medio, RMSE, MAE, tasa de aciertos, hit-skill score, LEPS, Gerrity, ROC-below, ROC-above, y ACC. Cada estadístico se guarda en un archivo de salida específico dentro del directorio correspondiente.

- 4. Automatización del Proceso:** El script permite la iteración sobre múltiples modelos y lead times, asegurando que se generen todos los archivos necesarios para la verificación estadística de cada combinación de parámetros. Esto garantiza una cobertura exhaustiva y organizada del análisis.

2.3 . Descripción Detallada de las Métricas Estadísticas

Aquí se incluye la descripción detallada de cada métrica con sus fórmulas, como se detalla a continuación:

- 1. Desviación Estándar (SD):** Mide la dispersión de un conjunto de datos en relación con su media.

$$SD = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

Donde x_i son los valores de los datos, N es el número total de datos, y μ es la media de los datos.

- 2. Correlación de Pearson (Pearson):** Mide la fuerza y la dirección de la relación lineal entre dos variables.

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}$$

Donde x_i y y_i son los valores de las dos variables, y \bar{x} y \bar{y} son sus medias.

- 3. Correlación de Spearman (Spearman):** Mide la relación monotónica entre dos variables mediante la correlación entre sus rangos.

$$\rho = 1 - \frac{6 \sum d_i^2}{N(N^2 - 1)}$$

Donde d_i es la diferencia entre los rangos de cada par de variables, y N es el número total de pares.

- 4. Área Bajo la Curva ROC de dos categorías (2AFC):** Mide la habilidad de un modelo para discriminar entre dos categorías.

$$2AFC = \frac{\text{Positivos Verdaderos} + \text{Falsos Positivos}}{2}$$

Es el área bajo la curva ROC para dos categorías.

- 5. Proporción de Varianza Explicada (Porc_varianza):** Mide la proporción de la varianza total de las observaciones explicada por el modelo.

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

Donde y_i son los valores observados, \hat{y}_i son las predicciones, \bar{y} es la media de los valores observados.

6. **Razón de Varianza (Varianza_ratio):** Compara la varianza de las predicciones con la varianza de las observaciones.

$$\text{Varianza Ratio} = \frac{\text{Varianza de las predicciones}}{\text{Varianza de las observaciones}}$$

7. **Desviación Media (Mean_bias):** Mide el sesgo sistemático en las predicciones.

$$\text{Bias} = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)$$

8. **Raíz del Error Cuadrático Medio (RMSE):** Mide la magnitud promedio del error entre las predicciones y las observaciones.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}$$

9. **Error Absoluto Medio (MAE):** Mide el error absoluto promedio entre las predicciones y las observaciones.

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|$$

10. **Índice de Acierto (Hit):** Mide la proporción de veces que el modelo predice correctamente un evento.

$$\text{Hit Rate} = \frac{\text{Aciertos}}{\text{Total de Observaciones}}$$

11. **Índice de Habilidad de Acierto (Hit Skill Score, HSS):** Compara el índice de acierto del modelo con un modelo aleatorio.

$$\text{HSS} = \frac{\text{Aciertos} - \text{Esperados por Azar}}{\text{Total de Observaciones} - \text{Esperados por Azar}}$$

12. **Índice de Habilidad LEPS (LEPS):** Mide la habilidad del modelo para prever eventos extremos.

$$LEPS = \frac{1}{N} \sum_{i=1}^N \text{Función de Pérdida } (y_i, \hat{y}_i)$$

La función de pérdida penaliza las diferencias entre eventos observados y predichos.

13. **Índice de Gerrity (Gerrity):** Evalúa la habilidad para predecir correctamente categorías ordenadas.

$$\text{Índice de Gerrity} = \frac{\sum_k W_k O_k}{\sum_k W_k}$$

Donde W_k es el peso asignado a cada categoría k y O_k son los aciertos en la categoría.

14. **Curva ROC por debajo (ROCbel):** Evalúa la habilidad de un modelo para discriminar eventos por debajo de un umbral.

$$ROCbel = \frac{\text{AUC por debajo del Umbral}}{\text{Total de Casos}}$$

Donde:

- **AUCbel** es el **Área Bajo la Curva (AUC)** específica para los casos por debajo del umbral.
- **Total de Casos** se refiere al número total de instancias o eventos analizados.

Área Bajo la Curva (AUCbel):

- El **AUCbel** mide el desempeño del modelo al predecir correctamente los eventos por debajo del umbral.
- Es un valor entre 0 y 1. Un valor más cercano a 1 indica que el modelo discrimina bien estos eventos, mientras que un valor c

15. **Curva ROC por encima (ROCabo):** Similar al ROCbel, pero aplicado a eventos por encima de un umbral.

$$ROCabo = \frac{\text{AUC por encima del Umbral}}{\text{Total de Casos}}$$

16. **Correlación Canónica (CC):** Mide la relación entre las predicciones y las observaciones considerando múltiples variables.

$$ACC = Cov(y, \hat{y}) / (\sigma_y, \sigma_{\hat{y}})$$

Donde Cov es la covarianza y σ es la desviación estándar.

2.4 Interpretación de Resultados

En este capítulo, se presentan y explican los principales estadísticos utilizados para evaluar el desempeño de los modelos climáticos en la predicción de fenómenos a lo largo del año. Estos 16 estadísticos han sido **agrupados en diferentes categorías** para facilitar su análisis, destacando patrones y tendencias específicas en la capacidad predictiva de cada modelo. El análisis se enfoca especialmente en estos grupos, ya que los resultados **abarcaban toda la región de Colombia**, lo que implica una evaluación integral y representativa de las predicciones en diversas condiciones climáticas

A continuación, se describen las categorías principales y los estadísticos agrupados, junto con su interpretación, en relación con los resultados presentados en **Capítulo 3**.

2.4.1 Correlaciones (pearson, spearman, cc):

Las medidas de correlación evalúan la fuerza y dirección de la relación entre las predicciones y las observaciones. Estas incluyen la correlación de Pearson, la correlación de Spearman y el Coeficiente de Correlación de Anomalías (ACC).

Interpretación general:

Rango típico: -1 a 1

Valores cercanos a 1 o -1 indican una fuerte relación

Valores cercanos a 0 indican una relación débil

2.4.2 Errores y Desviaciones (rmse, MAE):

Estas medidas cuantifican la magnitud de los errores de predicción. Incluyen el Error Cuadrático Medio (RMSE) y el Error Absoluto Medio (MAE). Proporcionan una medida directa de la precisión del modelo y RMSE es más sensible a errores grandes que MAE

Interpretación general:

- Valores más bajos indican mejor precisión
- Unidades son las mismas que la variable predicha

2.4.3 Estadísticos de Varianza y Desviación Estándar (var_ratio, p_var, SD):

Estos estadísticos evalúan cómo los modelos capturan la **variabilidad** de las precipitaciones observadas:

- **Relación de Varianza (var_ratio):** Compara la varianza entre predicciones y observaciones.
- **Proporción de Varianza Explicada (p_var):** Mide la fracción de la variabilidad observada capturada por el modelo.
- **Desviación Estándar (SD):** Evalúa la dispersión de las predicciones.

Interpretación general:

- **Valores cercanos a 1 en var_ratio:** Indican que el modelo refleja bien la variabilidad.
- **Valores altos de p_var:** Sugerencia de que el modelo captura adecuadamente la variabilidad.
- **Valores bajos de SD:** Reflejan predicciones más estables.

2.4.4 Habilidad de Pronóstico (hit, hss): Estas métricas evalúan la capacidad del modelo para predecir correctamente eventos específicos:

- **Hit Rate:** Proporción de aciertos en las predicciones.
- **Hit Skill Score (HSS):** Mide la habilidad del modelo en comparación con una predicción basada en climatología o azar.

Interpretación general:

- Rango típico: 0 a 1 para tasa de aciertos, $-\infty$ a 1 para HSS
- Valores más altos indican mejor habilidad predictiva

2.4.5 Evaluación de Consistencia y Discriminación en las Predicciones Climáticas: Mean Bias, 2AFC, LEPS y Gerrity

Este grupo de métricas mide la **consistencia** y la **capacidad del modelo para discriminar entre diferentes categorías o eventos**.

1. **Mean Bias:** Evalúa el **sesgo promedio** entre las predicciones y los valores observados.
 - **Bias positivo:** El modelo tiende a sobreestimar.
 - **Bias negativo:** El modelo tiende a subestimar.**Interpretación General:** Valores cercanos a cero indican predicciones más equilibradas.
2. **2AFC (Two-Alternative Forced Choice):** Mide la habilidad del modelo para **diferenciar entre dos posibles resultados binarios**.
Interpretación General:
 - **Valores altos:** Indican una buena capacidad de discriminación.
 - **Valores cercanos a 0:** Sugieren un desempeño similar al azar.
3. **LEPS (Linear Error in Probability Space):** Evalúa qué tan bien el modelo predice **categorías basadas en probabilidades**.
Interpretación General:
 - **Positivo:** Indica que el modelo supera al azar.
 - **Cercano a 0:** Desempeño similar al azar.
 - **Negativo:** Peor que el azar.

4. **Gerrity Score:** Mide la capacidad del modelo para **clasificar eventos ordenados o categóricos**.

Interpretación General:

- **Valores positivos:** Indican una mayor habilidad predictiva.
- **Valores cercanos a 0:** Sugieren que el modelo no tiene ventaja sobre el azar.
- **Valores negativos:** Reflejan un desempeño deficiente.

2.4.6 Evaluadores de Discriminación 2 (roc_bel, roc_above): Estas medidas se enfocan específicamente en la capacidad del modelo para discriminar eventos por encima o por debajo de ciertos umbrales. Incluyen el ROC (Receiver Operating Characteristic) para eventos por debajo y por encima de umbrales específicos.

Interpretación general:

- Valores por encima de 0.5 indican habilidad de discriminación mejor que el azar
- Valores más cercanos a 1 indican mejor capacidad de discriminación

En predicciones climáticas:

- Particularmente útiles para evaluar la habilidad del modelo en predecir eventos extremos
- Permiten evaluar el rendimiento del modelo en diferentes partes de la distribución de la variable climática

3. Análisis de Resultados

3.1 Análisis Integrado de Correlaciones Canónica, Pearson y Spearman

Este análisis integra los resultados de la correlación canónica (CC), junto con Pearson y Spearman, considerando la variabilidad por modelo, mes y lead time.

3.1.1 Análisis por Correlación Canónica (CC)

Rendimiento de los Modelos a lo largo de los Meses:

En la Figura 1(arriba) de comparación de **Coefficiente de Correlación (CC)** por modelo y mes, **CFSv2** y **CMC2** destacan por su rendimiento consistente a lo largo del año, con valores altos de CC que reflejan su buena capacidad predictiva. Los modelos **NMME** y **CMC1** muestran más variabilidad, especialmente en meses específicos.

En la Figura 1(abajo), que evalúa el CC por lead time para cada modelo, se puede observar que el rendimiento del modelo tiende a disminuir a medida que el **lead time** aumenta, lo cual es un comportamiento esperado debido a la incertidumbre creciente con el pronóstico a largo plazo. Modelos como **cfsv2** y **nasa** muestran una **degradación menos pronunciada** del CC con el aumento del lead time, sobresaliendo en predicciones a largo plazo.

Para los **lead times cortos** (1 a 3), los valores de CC son generalmente altos para la mayoría de los modelos, mientras en los **lead times más largos** (4 a 6), se nota una caída significativa en el rendimiento de algunos modelos, particularmente para los modelos **CMC1** y **NMME**.

3.1.2 Análisis de Correlaciones Pearson y Spearman

3.1.3 Comparación General para las correlaciones

Los patrones de Pearson y Spearman son consistentes con el análisis de CC, aunque se observan ligeras diferencias en meses como septiembre y diciembre, sugiriendo la influencia de relaciones no lineales. Los modelos **CFSv2** y **CMC2** sobresalen por sus altos valores positivos de correlación, los más altos durante los primeros meses (enero, marzo, junio) y hacia el final del año (noviembre-diciembre), con rangos entre **0.3 y 0.6**; mientras que **NMME** y **CMC1** muestran más variabilidad, especialmente en lead times largos. Los valores de los índices de correlación son más altos en el **primer trimestre del año (enero-marzo)**, en contraste con valores cercanos a cero en **abril, mayo, septiembre y noviembre** debido a la alta dispersión.

3.2 Análisis de errores absolutos medios (MAE) y errores cuadráticos medios (RMSE) por modelo, mes y lead time

Las Figuras 4 y 5 muestran el desempeño de los modelos en términos de MAE y RMSE por lead time y mes, permitiendo identificar patrones de precisión y variabilidad a lo largo del año.

3.2.1 Error absoluto medio (MAE). En los primeros meses del año (Figura 4), modelos como **NMME, NASA y GFDL** mantienen valores bajos en lead times cortos (1 a 3), destacando su buena capacidad predictiva. Sin embargo, **CMC1 y CFSv2** presentan mayores errores en meses críticos como Abril y Mayo.

En la segunda mitad del año (Figura 5), se observa un aumento en los errores durante Julio y Agosto, con cierta recuperación en Noviembre y Diciembre por parte de NASA y GFDL. NMME y CMC1 muestran alta dispersión en Octubre y Noviembre, afectando su consistencia.

3.2.2 Error cuadrático medio (RMSE). El RMSE refleja tendencias similares al MAE, pero revela una mayor sensibilidad a errores grandes. En la Figura 4, la mayor dispersión ocurre en Abril y Mayo, con CFSv2 mostrando mayor estabilidad que CMC1.

En la Figura 5, los errores grandes son más frecuentes en Septiembre y Octubre, lo que sugiere desafíos adicionales en estos meses para varios modelos.

3.2.3 Tendencias generales para MAE y RMSE

- A medida que el lead time aumenta, la precisión de los modelos disminuye, especialmente en NMME y CMC1, que pierden precisión en horizontes largos (6-9). Los modelos NASA, GFDL y CFSv2 mantienen mejor consistencia en lead times largos, destacando por su estabilidad en predicciones extendidas.
- Los modelos CFSv2 y GFDL sobresalen por su estabilidad a lo largo del año, especialmente en lead times cortos y medianos, mientras que NMME y CMC1 presentan mayor variabilidad.
- La dispersión de errores es menor en los primeros meses del año (Figura 4) en comparación con la segunda mitad (Figura 5), donde la variabilidad climática en meses como Septiembre y Octubre afecta significativamente la precisión de los modelos.
- Lead Times: Los lead times cortos (1-3) ofrecen mayor precisión, mientras que solo modelos como NASA y CFSv2 mantienen un rendimiento aceptable en horizontes largos (6-9).

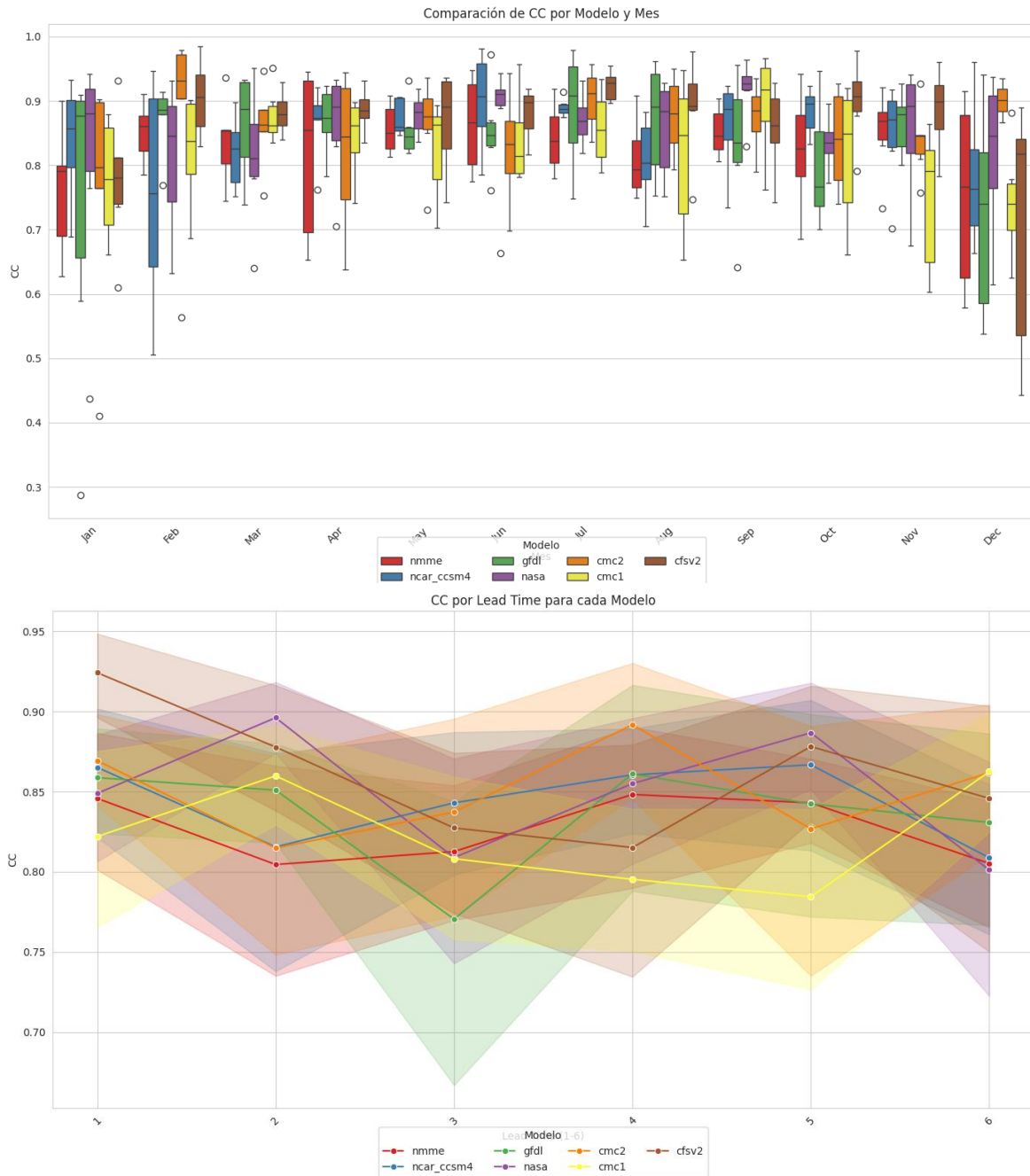


Figura 1. Comparación de la Correlación Canónica (CC) por Modelo y Mes (arriba, y Lead Time (abajo)

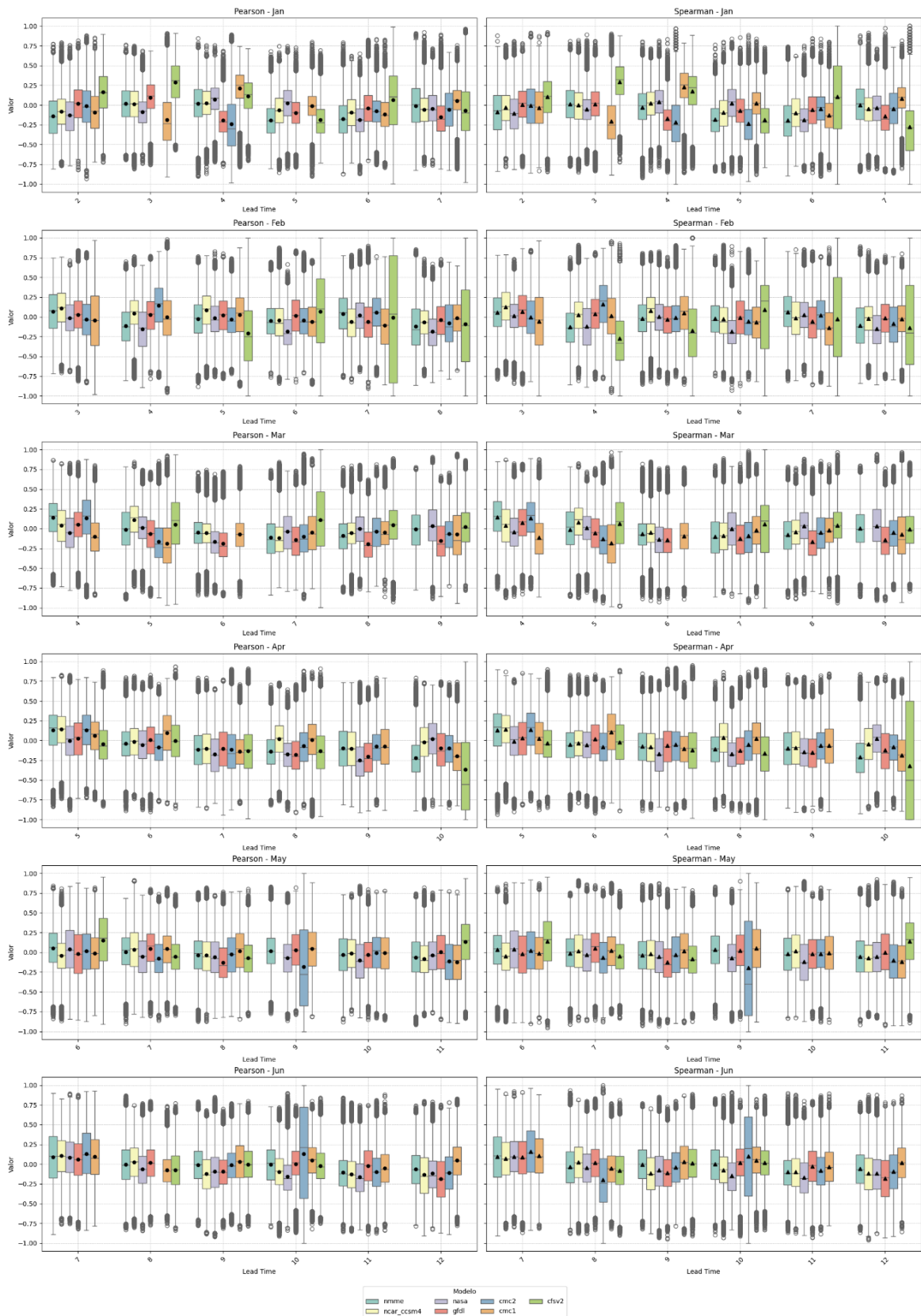


Figura 2. Correlaciones Pearson y Spearman por Modelo y Lead Time – Primer Semestre (Enero-Junio)

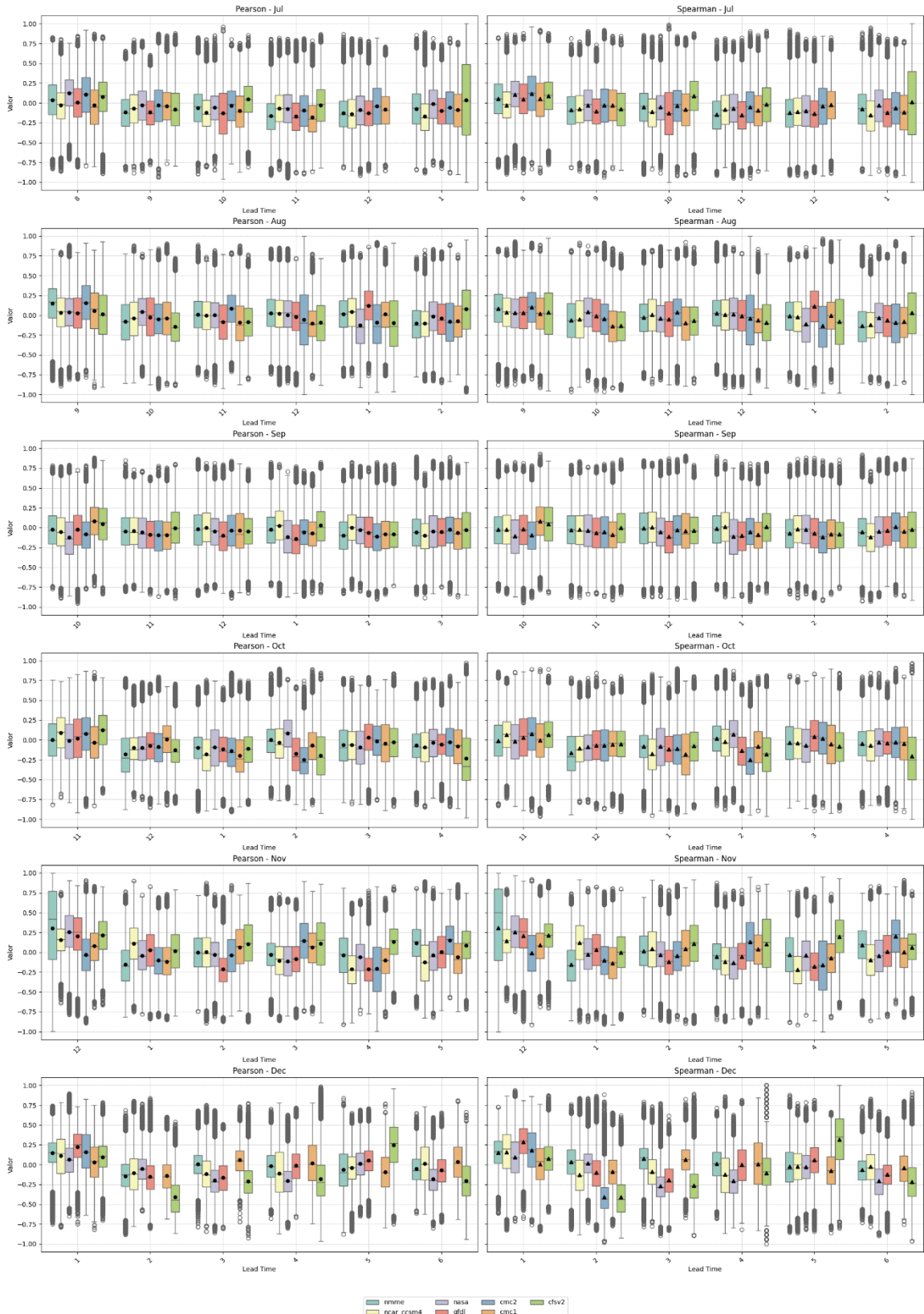


Figura 3. Correlaciones Pearson y Spearman por Modelo y Lead Time – Segundo Semestre (Julio-Diciembre)

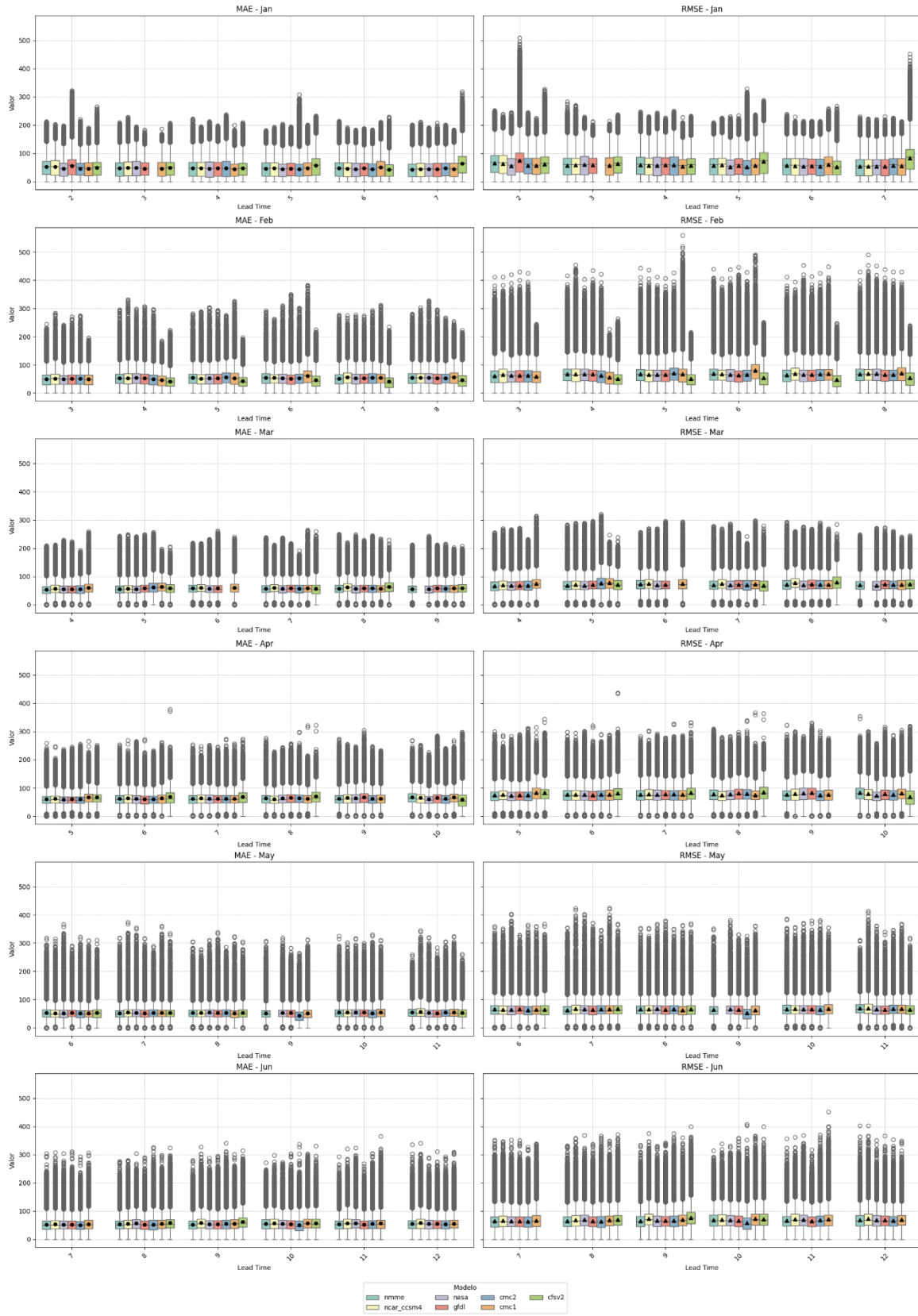


Figura 4. Desempeño de Modelos en MAE y RMSE durante el Primer Semestre (Enero-Junio).

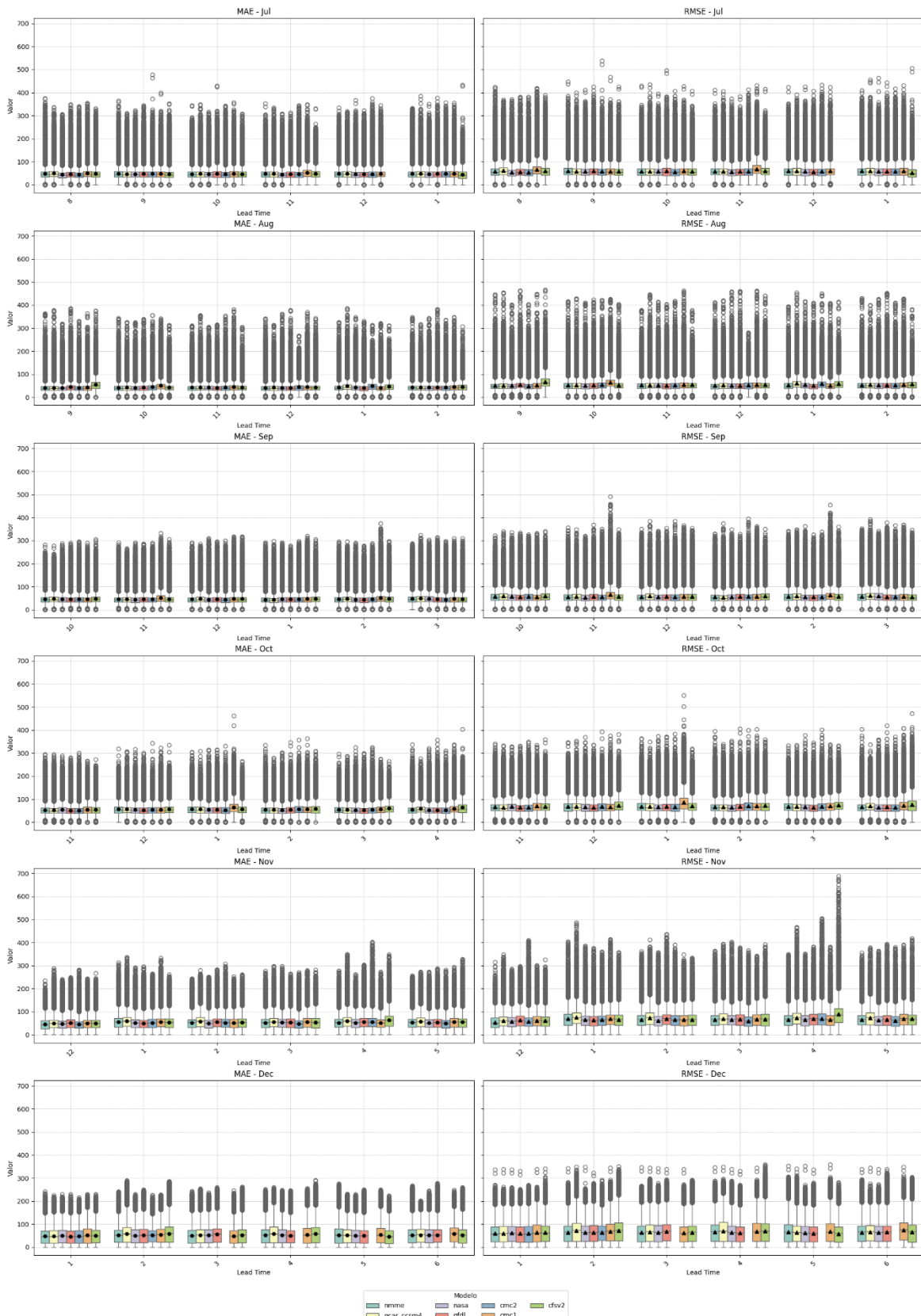


Figura 5. Desempeño de Modelos en MAE y RMSE durante el Segundo Semestre (Julio-Diciembre).

3.3 Análisis de los Estadísticos de Varianza y Desviación Estándar por Modelo y Mes

Hasta este punto, los primeros lead times han demostrado las mayores correlaciones y menores errores. Dado que estos estadísticos muestran poca variabilidad entre los diferentes horizontes, este análisis se centra en un promedio mensual, facilitando así la interpretación de los resultados.

Este análisis se enfoca hacia **la desviación estándar (SD)**, **la proporción de varianza explicada (p_var)** y **el ratio de varianza (var_ratio)**, métricas para comprender cómo los modelos capturan la variabilidad de las precipitaciones pronosticadas y cuán consistentes son en sus predicciones a lo largo del año. A continuación, se presenta la interpretación detallada de cada uno de los mapas de calor.

3.3.1 Relación de Varianza (var_ratio): El ratio de varianza (var_ratio) compara la variabilidad predicha por el modelo con la observada, ayudando a identificar subestimaciones o sobreestimaciones de la variabilidad. Así, de la Figura 6, se destaca que **CFSv2** y **CMC2** muestran algunas sobreestimaciones significativas de la varianza en **enero, junio y diciembre**, lo que podría indicar una mayor incertidumbre o sensibilidad en las predicciones para esos meses. **NMME**, **NCAR_CCSM4**, y **NASA** presentan un comportamiento más **estable**, pero con una tendencia a **subestimar la variabilidad**, lo que podría limitar la capacidad de capturar eventos extremos.

Enero y junio son meses con ratios de varianza más altos en varios modelos, lo que sugiere **mayor incertidumbre** o condiciones más difíciles de predecir. En contraste, meses como **febrero, abril y noviembre** muestran ratios más bajos, reflejando predicciones más consistentes.

3.3.2 Proporción de Varianza Explicada (p_var): La proporción de varianza explicada (p_var) mide cuánto de la variabilidad observada es capturada por el modelo (Figura 7).

El modelo **CFSv2** destaca en **Enero y Diciembre**, capturando de manera eficiente la variabilidad de las precipitaciones. El modelo **CMC2** presenta picos significativos en **Abril y Octubre**, lo que sugiere una buena capacidad para explicar las fluctuaciones del clima en esos meses.

Los meses de **Junio y Octubre** muestran más dispersión, indicando que los modelos enfrentan mayores desafíos en la representación de la amplitud de las fluctuaciones durante estos periodos.

3.3.3 Desviación Estándar (SD). Mide la variabilidad típica de las predicciones del modelo, proporcionando una evaluación de la consistencia en las proyecciones. De la Figura 8, se resalta que los modelos **CFSv2** y **GFDL** sobresalen con **baja SD** en meses como **Enero, Mayo y Diciembre**, lo que indica **predicciones más estables y menos dispersas**, mientras, **CMC2** y **NCAR_CESM4** presentan **alta SD** en **Abril y Agosto**, reflejando mayor incertidumbre en sus predicciones durante estos meses.

De **Julio a Septiembre**, algunos modelos como **CMC1 y NMME** muestran valores más elevados, posiblemente influenciados por fenómenos estacionales. Por su parte, **Abril** se destaca como un mes de alta variabilidad en varios modelos, lo que indica un desafío común en capturar los patrones climáticos de ese periodo.

3.3.4 Tendencias generales para ratio y proporción de la varianza y desviación estándar.

- Los modelos CFSv2 y GFDL se destacan por su baja desviación estándar (SD), un ratio de varianza cercano a 1 y una alta proporción de varianza explicada (p_var) en meses como Enero, Mayo y Diciembre, lo que refleja que sus predicciones son tanto estables como precisas al capturar la variabilidad de las precipitaciones. En contraste, CMC2 y NCAR_CESM4 muestran alta SD y valores atípicos en el ratio de varianza, lo que sugiere que estos modelos subestiman o sobreestiman la variabilidad en ciertos meses, especialmente en Junio y Agosto
- Abril, Agosto y Octubre presentan alta SD y menor proporción de varianza explicada (p_var) en varios modelos, reflejando los desafíos de capturar los patrones climáticos durante estos periodos.
- El ratio de varianza (var_ratio) más bajo en Octubre para varios modelos indica que este es un mes difícil para predecir la variabilidad del clima, probablemente debido a que climáticamente es un mes de altos volúmenes de lluvia para todo el país.

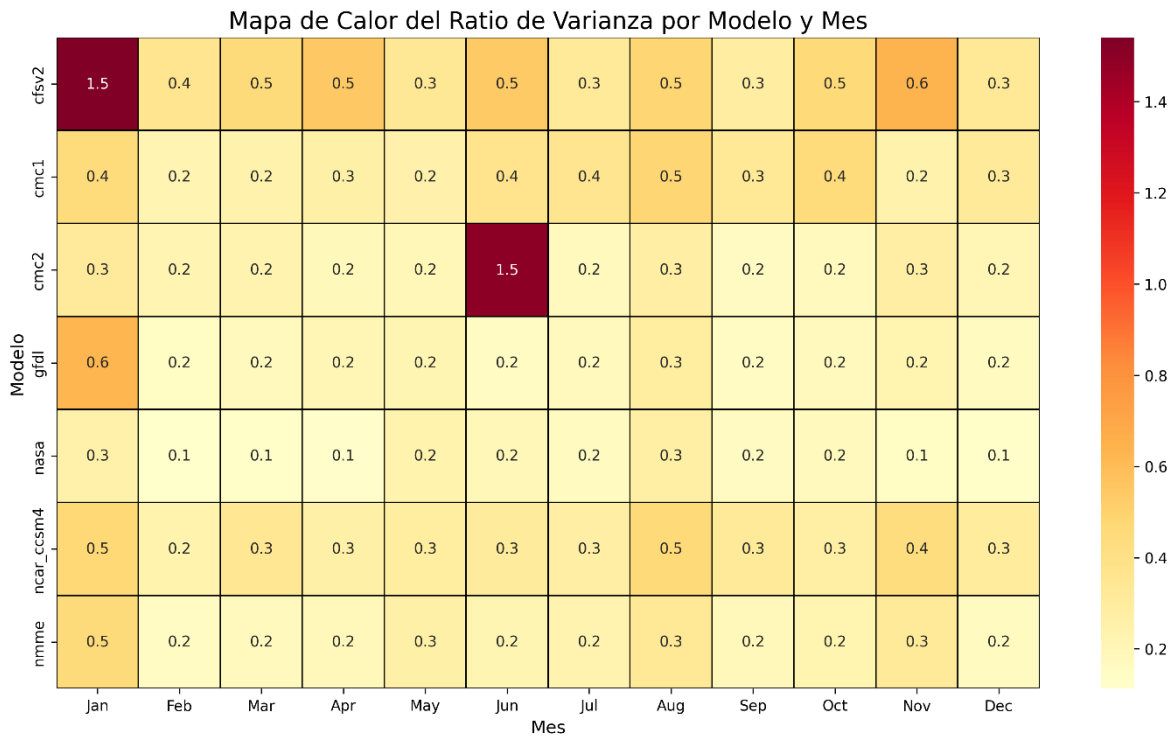


Figura 6. Mapa de Calor del Ratio de Varianza (var_ratio) por Modelo y Mes

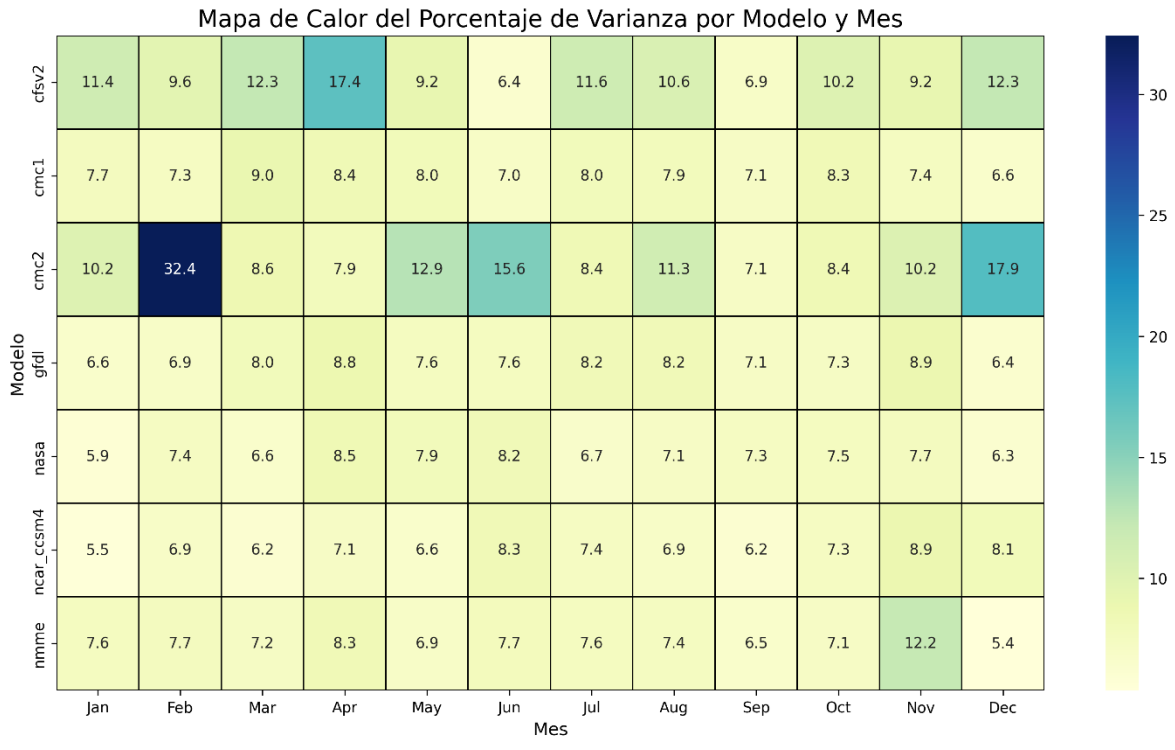


Figura 7. Mapa de Calor del Porcentaje de Varianza Explicada (p_{var}) por Modelo y Mes

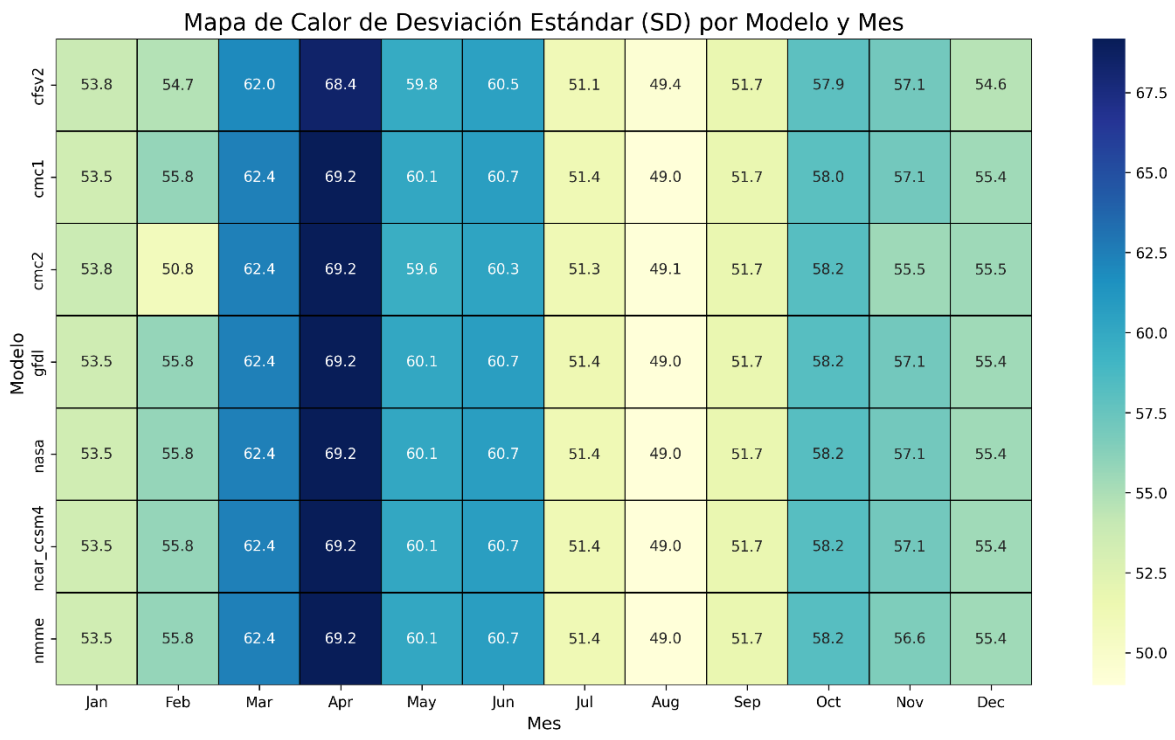


Figura 8 Mapa de Calor de Desviación Estándar (SD) por Modelo y Mes

3.4 Análisis sobre Tasa de Aciertos (Hit Rate) y Habilidad Predictiva (HSS) por Mes y Lead Time

Las Figuras 9 y 10 muestran el desempeño de los modelos en términos de **Hit Rate (Tasa de Aciertos)** y **Hit Skill Score (HSS)** por lead time y mes, permitiendo evaluar la precisión en la predicción de eventos específicos y la capacidad de los modelos para superar las predicciones de referencia, como la climatología.

3.4.1 Hit Rate. En los primeros meses (Figura 9), modelos como **NASA, GFDL y NMME** alcanzan valores más altos en **lead times cortos (1-3)**, destacando su capacidad predictiva. Sin embargo, en **Mayo y Junio**, la precisión disminuye notablemente en los modelos **CMC1 y CMC2**.

En la segunda mitad del año (Figura 10), se observa estabilidad en **Julio y Agosto**, especialmente en los modelos **NASA y GFDL**. Sin embargo, en los meses de transición, como **Octubre y Noviembre**, los valores de las tasas de acierto tienden a caer para **NMME y CMC1**.

3.4.2 HSS. En los primeros meses (Figura 9), **CFSv2 y GFDL** mantienen valores positivos, superando las predicciones basadas en climatología, especialmente en **lead times cortos y medianos**. En la segunda mitad del año (Figura 10), los modelos **NASA y CFSv2** logran mantener un rendimiento relativamente alto incluso en **lead times largos (6-9)**, pero la mayoría de los modelos presentan valores más bajos o negativos en **Septiembre, Octubre y Noviembre**, indicando dificultades en la precisión durante esos meses.

3.4.3 Tendencias generales para Hit Rate y HSS. A medida que aumenta el **lead time**, la precisión de los modelos tiende a disminuir, con **NMME y CMC1** presentando mayor degradación. Los modelos **NASA, GFDL y CFSv2** muestran un rendimiento más estable, destacándose en lead times largos (6-9), donde mantienen una mejor capacidad predictiva en comparación con los demás modelos.

En cuanto a consistencia, los modelos **NASA, GFDL y CFSv2** sobresalen como los modelos más consistentes tanto en hit rate como en HSS, especialmente en los primeros meses del año y en lead times cortos.

Meses Críticos: Mayo, Septiembre, Octubre y Noviembre representan mayores desafíos, con valores más bajos tanto en hit rate como en HSS, afectando la precisión de los modelos.

En resumen, **NASA, GFDL y CFSv2** destacan por su capacidad predictiva y estabilidad, mientras que **NMME y CMC1** muestran mayor variabilidad y una pérdida de precisión en meses de transición y en horizontes largos.

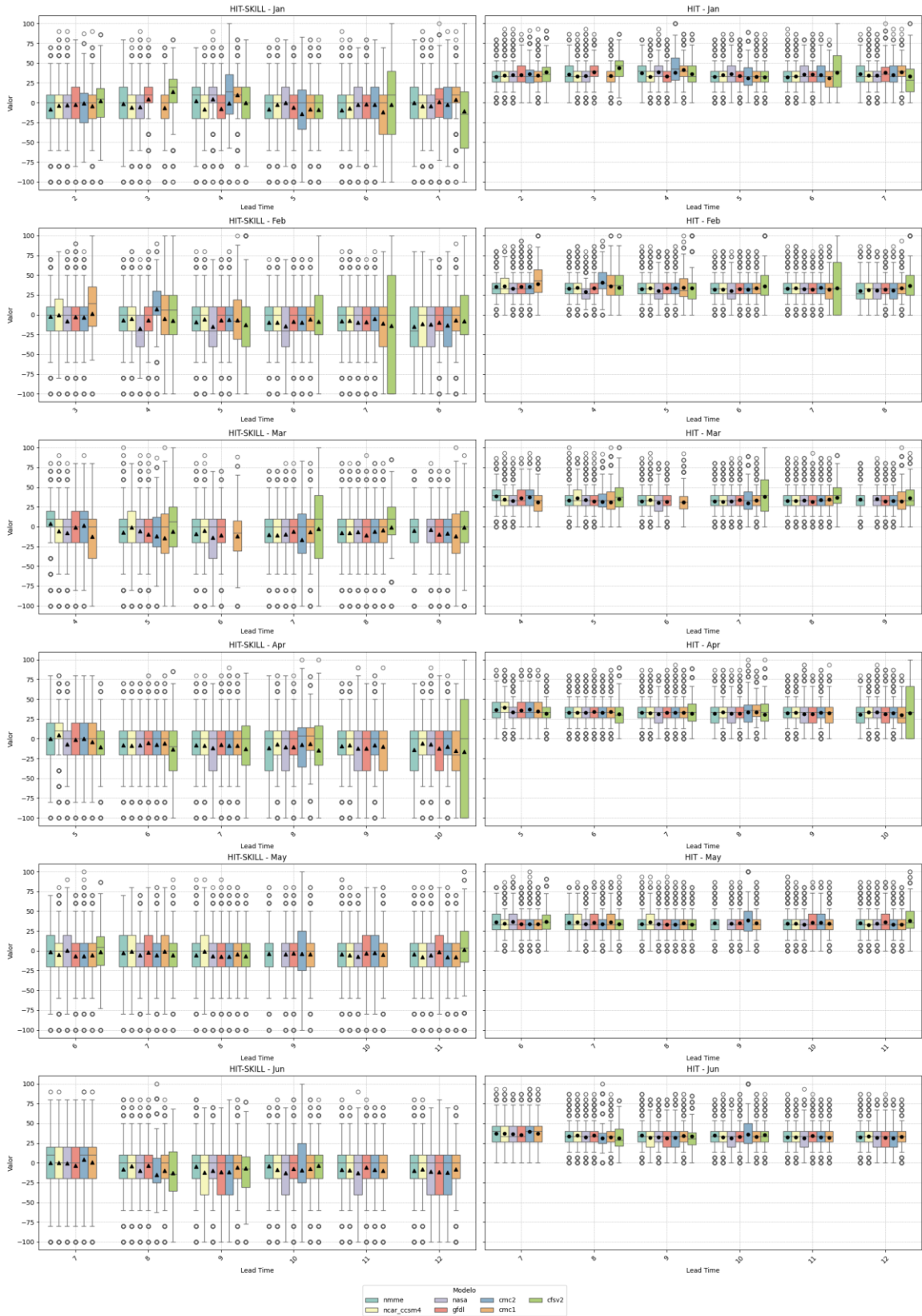


Figura 9. Comparativo de Habilidad de Acierto (HSS) y Tasa de Aciertos (Hit Rate) por Modelo, Mes y Lead Time (Enero - Junio)

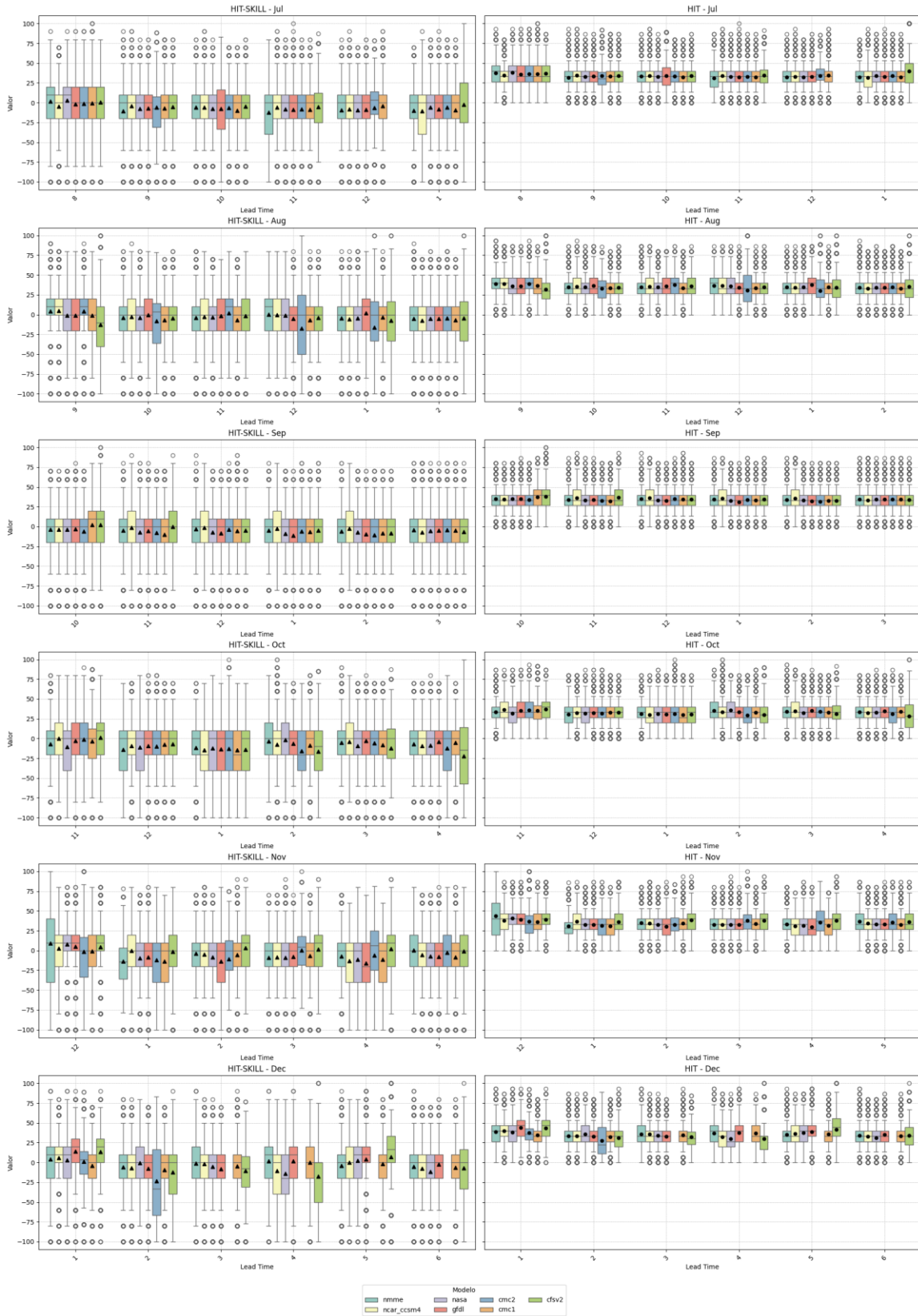


Figura 10. Comparativo de Habilidad de Acierto (HSS) y Tasa de Aciertos (Hit Rate) por Modelo, Mes y Lead Time (Julio - Diciembre).

3.5 Evaluación de Consistencia y Discriminación en las Predicciones Climáticas: Análisis de Mean Bias, 2AFC, LEPS y Gerrity

Este análisis combina las métricas de **consistencia** (mean bias) con los **evaluadores de discriminación** (2AFC, LEPS y Gerrity) para ofrecer una visión completa del rendimiento de los modelos climáticos.

Mientras que el **mean bias** ayuda a identificar **sesgos sistemáticos** en las predicciones al evaluar si el modelo tiende a sobrestimar o subestimar los valores observados, las **métricas de discriminación** como **2AFC, LEPS y Gerrity** miden la **capacidad del modelo para distinguir entre diferentes estados o eventos climáticos**. Estas métricas son especialmente útiles para evaluar el rendimiento en la predicción de **eventos extremos o poco frecuentes**.

3.5.1 Mean Bias: Identificación del Sesgo Promedio en las Predicciones.

La gráfica de boxplot presenta la distribución del Mean Bias para diferentes modelos climáticos a lo largo de los meses del año (Figura 11), permitiendo evaluar cómo cada modelo se desvía, en promedio, de los valores observados.

En relación a los meses de **Enero, Julio y Diciembre**, se observa una mayor dispersión en los valores del Mean Bias, con predicciones que varían entre ± 200 . Esto indica una mayor **incertidumbre** en las predicciones de los modelos durante estos periodos. Por su parte, meses como **Marzo, Abril y Octubre**, presentan la dispersión más moderada, y la mayoría de los modelos presentan sesgos cercanos a **cero**. Esto sugiere que las predicciones en estos meses son más **equilibradas**, con menos sobre o subestimación.

Los modelos **CFSv2 y GFDL**: Tienden a tener sesgos cercanos a cero en varios meses, lo que refleja un rendimiento más **estable**, mientras, los modelos **CMC1 y CMC2** muestran mayor dispersión en meses como **Mayo y Octubre**. Los modelos **NMME y NCAR_CESM4** muestran valores extremos tanto positivos como negativos en meses como **Febrero y Junio**, lo que indica predicciones menos consistentes y mayor variabilidad.

En general, los modelos muestran menor variabilidad en los primeros lead times (1-3), reflejando mayor precisión en las predicciones. Sin embargo, en los lead times largos (6-9), la dispersión del bias incrementa significativamente, especialmente en modelos como NMME y CMC2, lo que sugiere una mayor incertidumbre en sus estimaciones.

Tendencias y Conclusiones

- **Modelos Consistentes: GFDL y CFSv2** destacan por su sesgo medio bajo y estable, sugiriendo que sus predicciones son más confiables a lo largo del año.
- **Meses Críticos: Mayo y Octubre** se identifican como meses con mayor **dispersión** en el sesgo medio, indicando que los modelos enfrentan dificultades adicionales para capturar el valor de la precipitación en estos periodos.

- **Sobre y Subestimación:** Varios modelos muestran un patrón de sobreestimación en meses como **Julio** y **Agosto** (sesgos positivos) y subestimación en **Junio** y **Octubre** (sesgos negativos).

3.5.2 Análisis del 2AFC: Capacidad del Modelo para Diferenciar Eventos Binarios

El 2AFC (Two-Alternative Forced Choice) evalúa la capacidad del modelo para distinguir entre dos posibles resultados (Figura 12). Valores más altos reflejan mejor desempeño predictivo.

Mejor Desempeño: En **junio** y **noviembre**, los modelos como **GFDL** y **CFSv2** muestran **valores altos**, especialmente en **lead times cortos (1-3)**, indicando mayor precisión en predicciones a corto plazo.

Mayor Incertidumbre: En **enero** y **agosto**, se observa mayor **dispersión**, sobre todo en **lead times largos (6-9)**, lo que refleja dificultad para predecir en estos periodos con horizontes extendidos.

Tendencias por Modelo:

Los modelos **CFSv2** y **GFDL** mantienen **consistencia** en varios meses y en distintos lead times, lo que destaca su **capacidad predictiva estable**.

Los modelos **CMC2** y **NCAR_CESM4** presentan mayor **variabilidad**, especialmente en **horizontes largos**, mostrando menor precisión en estos escenarios.

Los modelos muestran **desempeño moderado** en la discriminación de eventos, con mejores resultados en **lead times cortos**. Los modelos **CFSv2** y **GFDL** sobresalen por su estabilidad en varios meses y horizontes de tiempo, mientras que otros modelos enfrentan **mayores desafíos** en predicciones a largo plazo.

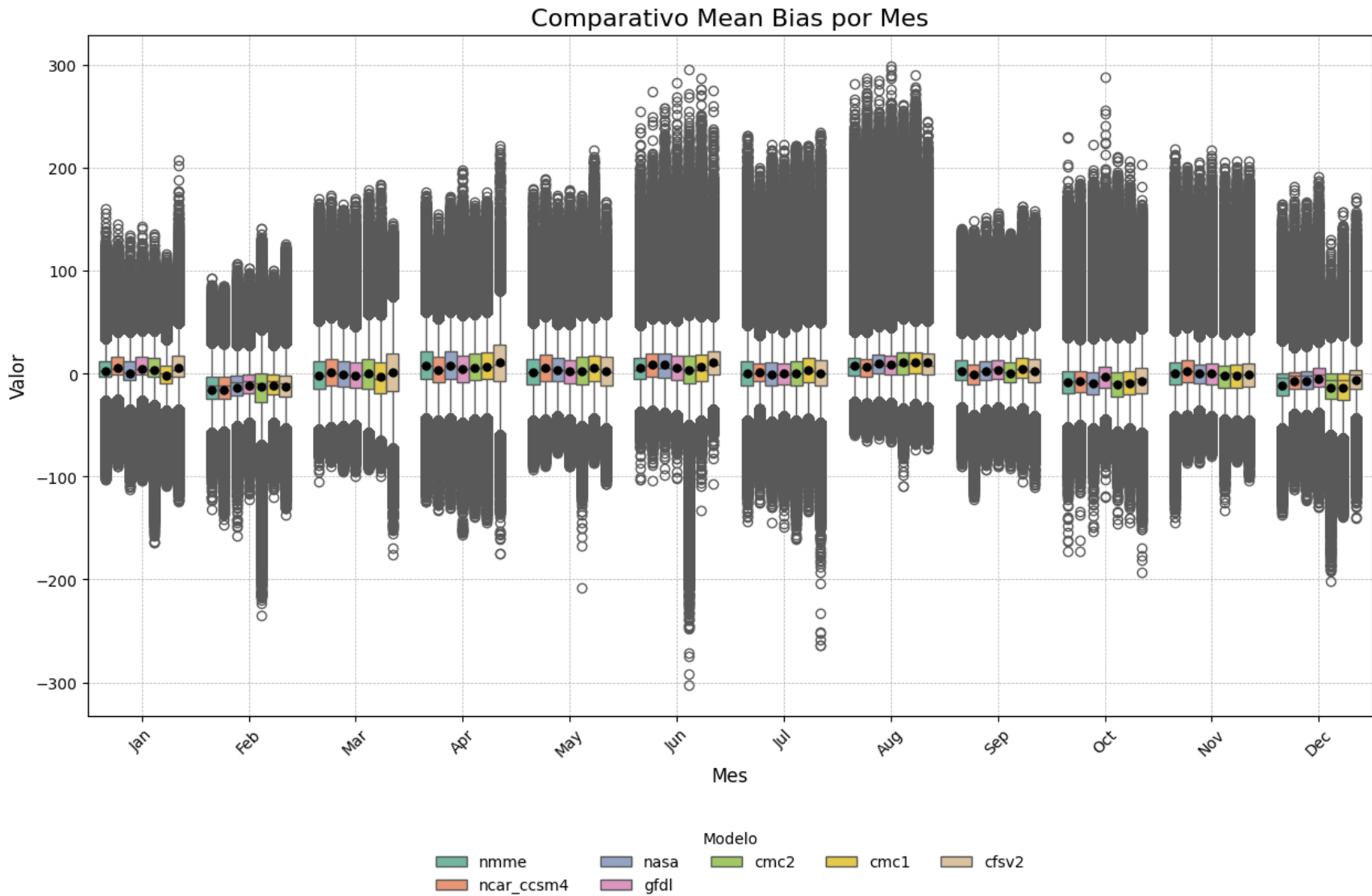


Figura 11. Comparativo del Sesgo por Mes y Lead Time.

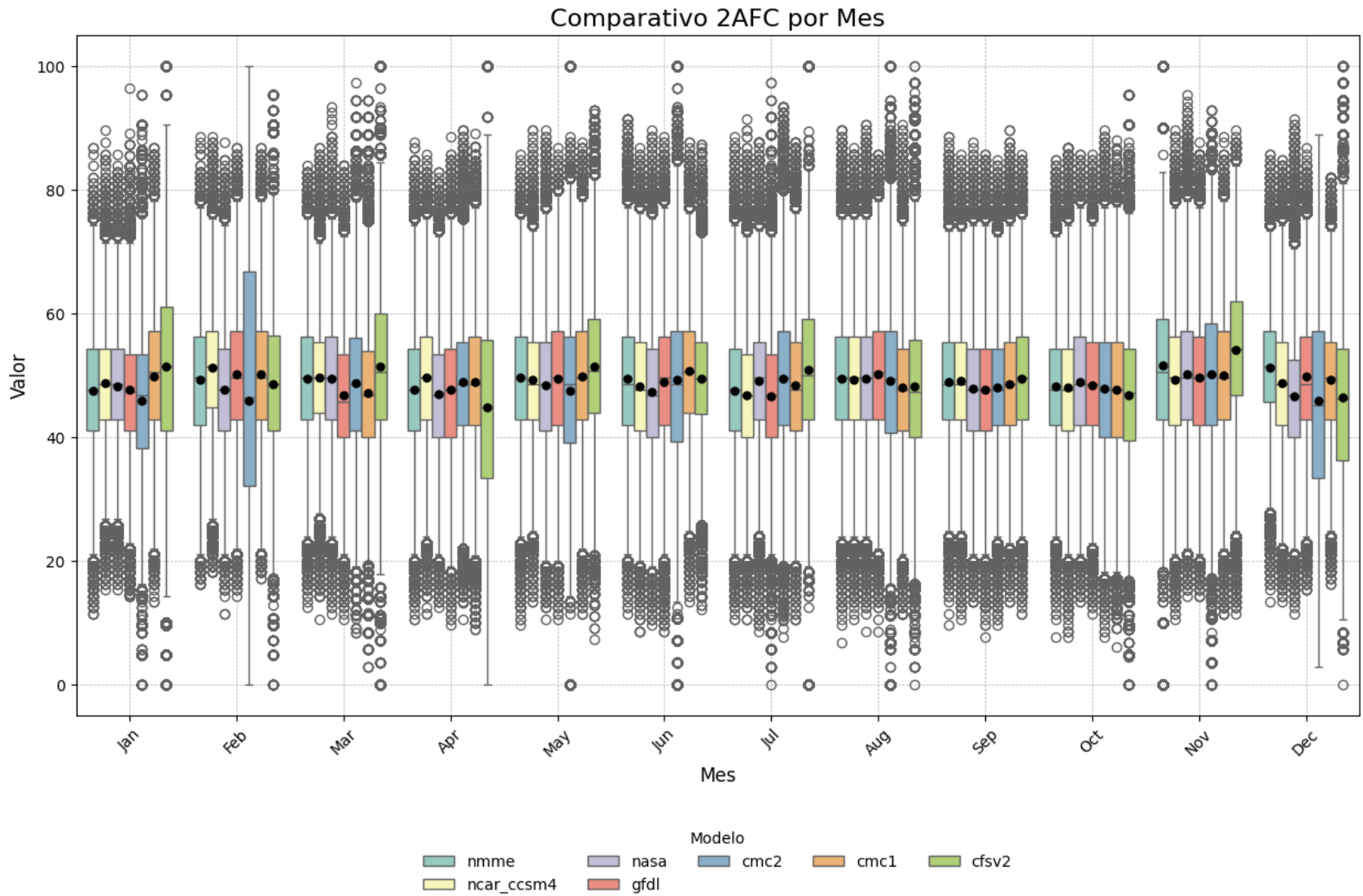


Figura 12. Comparativo del estadístico 2AFC por Mes y Lead Time

3.5.3 Análisis de la Precisión de las Predicciones en el Espacio de Probabilidad (LEPS).

El LEPS (Lear Error in Probability Space) evalúa la capacidad del modelo para predecir categorías de eventos según su probabilidad (Figura 13).

De acuerdo con la Figura 13, muestra que los modelos **CFSv2, GFDL y NASA** mantienen valores positivos en LEPS durante varios meses, mostrando **buena capacidad predictiva** incluso en periodos complejos, lo que los hace más confiables para predicciones estacionales.

Mayo, Octubre y Diciembre presentan mayor dispersión, con valores negativos y puntos atípicos que alcanzan -75, lo que indica mayor incertidumbre en estos periodos, particularmente en **lead times largos (6-9)**. En Junio y Agosto, la concentración de valores cercanos a cero sugiere que las predicciones para estos meses tienen poca ventaja sobre el azar, especialmente en los **horizontes más extendidos**.

Se mantiene la tendencia de mayor precisión en los lead times cortos, dado los valores positivos de LEPS, mientras que en lead times largos se observa un incremento de la dispersión, especialmente para los modelos NMME y CMC2.

Tendencias y Conclusiones

- **Modelos Estables:** Los modelos CFSv2, GFDL y NASA destacan por su capacidad predictiva, manteniendo valores positivos de LEPS tanto en lead times cortos como largos.
- **Meses y Lead Times Críticos:** Mayo, Octubre y Diciembre, junto con los lead times largos, representan mayores desafíos, con menor precisión y mayor variabilidad en las predicciones.
- **Evaluación Global:** Los meses intermedios del año (Junio a Agosto) presentan menos dispersión, lo que sugiere menor dificultad para predecir en estos periodos. Sin embargo, NMME y NCAR_CESM4 requieren ajustes para mejorar su consistencia, especialmente en horizontes largos.

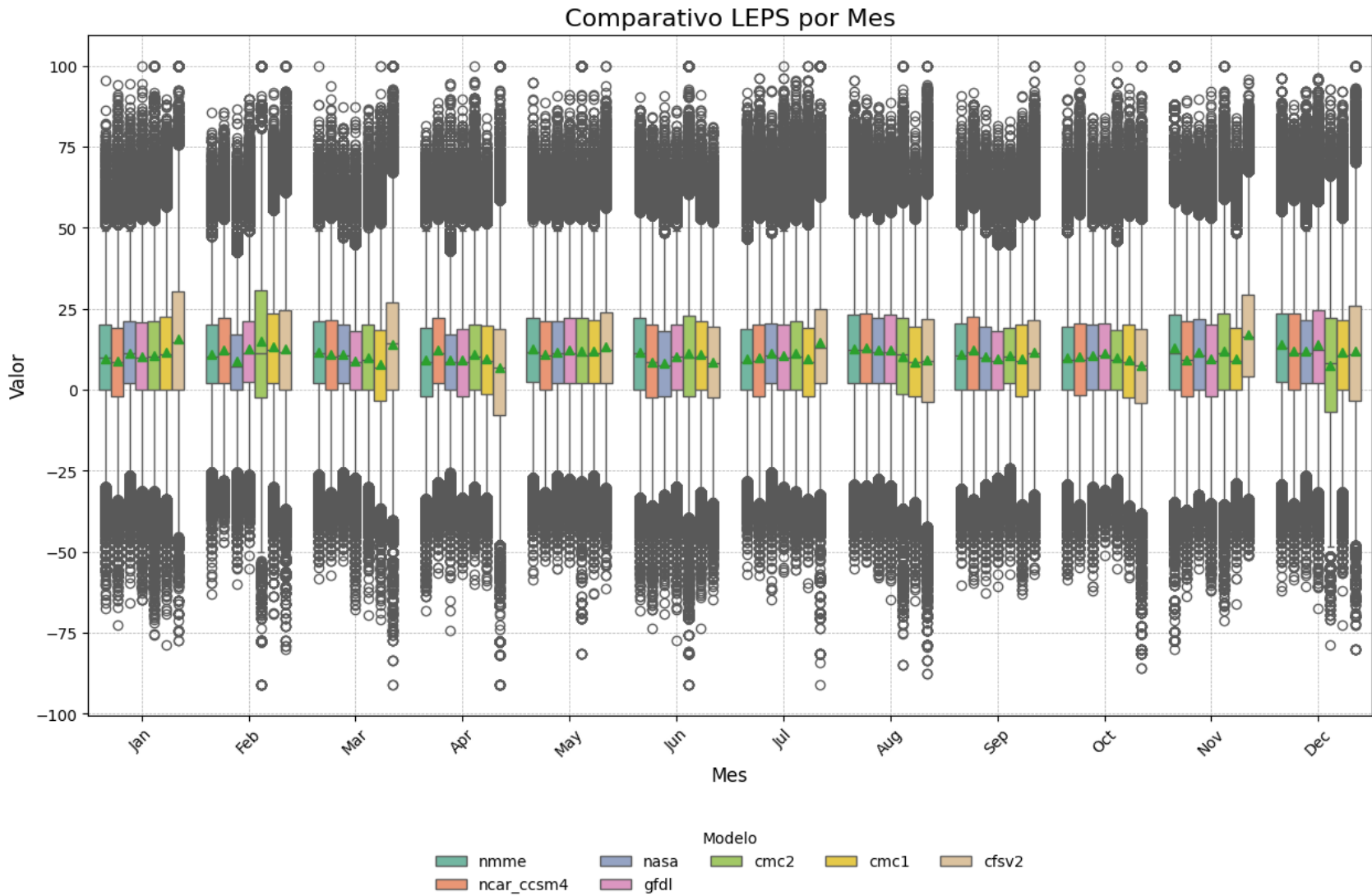


Figura 13. Comparativo del estadístico LEPS por Mes y Lead Time

3.5.4 Análisis del Estadístico Habilidad de Clasificación en Eventos Ordenados (Gerrity Score) por Mes y Modelo

El **estadístico Gerrity** es una métrica diseñada para evaluar la habilidad de los modelos en la clasificación de eventos ordinales o categóricos, siendo particularmente útil en predicciones climáticas, donde se busca medir qué tan bien un modelo asigna categorías específicas, así:

Como se observa en la Figura 14, la mayoría de los modelos muestran **valores positivos moderados** a lo largo del año, lo que refleja una capacidad aceptable para clasificar categorías correctamente. Los modelos **CFSv2 y GFDL** mantienen valores estables cercanos a cero o ligeramente positivos, reflejando predicciones equilibradas y consistentes. En contraste, el modelo **CMC2** muestra una mayor variabilidad, con meses donde destaca y otros donde su desempeño se acerca a cero. Los modelos **NMME y NCAR_CESM4** presentan más dispersión, con predicciones erráticas, especialmente en meses como **Febrero y Octubre**.

Algunos meses muestran una **dispersión significativa**, lo que indica una menor consistencia en las predicciones. **Enero, Mayo y Octubre** se destacan por la presencia de valores extremos que alcanzan ± 100 , lo que evidencia desafíos en estos periodos. En cambio, **Junio y Julio** presentan menor dispersión, con valores más concentrados alrededor de cero, lo que sugiere que las predicciones durante estos meses son más estables y equilibradas.

Tendencias y Conclusiones

- **Meses Críticos:** Enero, Mayo y Octubre representan desafíos para los modelos debido a la alta dispersión en sus predicciones.
- **Modelos Consistentes:** Los modelos **CFSv2 y GFDL** sobresalen por mantener un desempeño más estable a lo largo del año, incluso en diferentes lead times.
- **Impacto del Lead Time:** Los **lead times cortos** muestran mayor precisión en las predicciones, mientras que los **lead times largos** presentan mayor incertidumbre, requiriendo ajustes en modelos como NMME y CMC2.

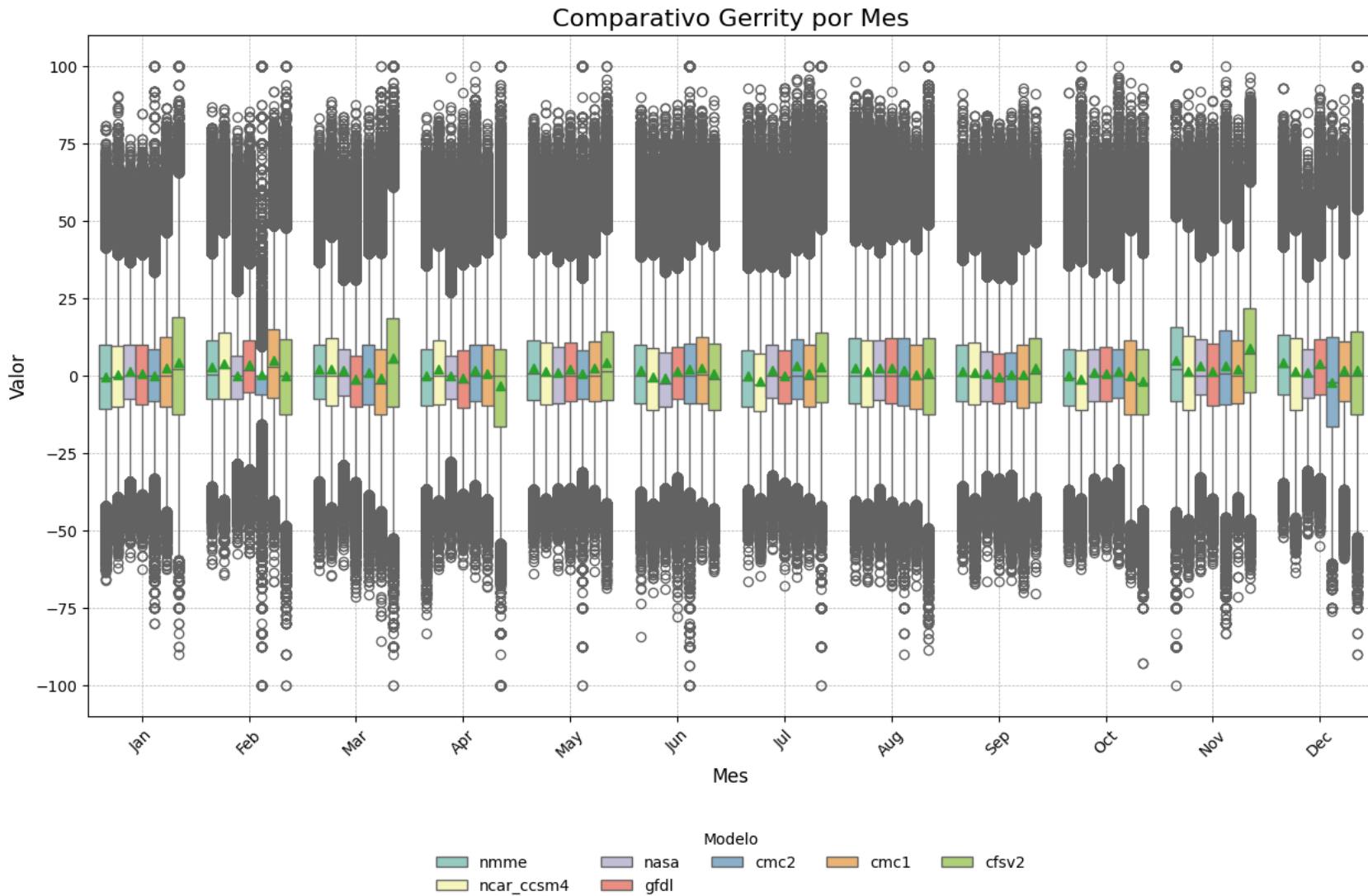


Figura 14. Comparativo del estadístico Gerrity Score por Mes y Lead Time

3.6 Análisis Resumido de los ROC (ROC Above y ROC Below) por Mes y Lead Time

Los gráficos muestran el desempeño de los modelos utilizando dos métricas de **ROC** (Figura 15 y 16):

- **ROC Below (ROCBEL):** Mide la habilidad del modelo para anticipar eventos por debajo de la climatología.
- **ROC Above (ROCABO):** Evalúa la capacidad del modelo para predecir eventos superiores de la climatología.

Desempeño por Lead Time:

- Los **lead times cortos (1-3)** tienden a mostrar mejores valores de ROC, lo que indica mayor precisión predictiva a corto plazo.
- En los **lead times largos (6-9)**, algunos modelos como **CMC1 y NMME** presentan una disminución de los estadísticos para ambas categorías (Below y Above), especialmente en meses de mayor variabilidad.

Modelos Destacados:

- **CFSv2 y GFDL** mantienen un desempeño relativamente alto en ambos indicadores del ROC a lo largo del año, destacándose tanto en predicciones por encima como por debajo de la climatología.
- **CMC2** muestra valores más inestables, especialmente en los meses de transición como abril y octubre, lo que refleja mayor dificultad en la precisión predictiva.

Meses Críticos:

- **Julio y agosto** presentan mayor dispersión en los valores ROC, lo que sugiere que la variabilidad climática estacional genera desafíos en las predicciones.
- **Enero y noviembre** muestran menor dispersión, con mejores valores en la mayoría de los modelos, indicando estabilidad en las condiciones climáticas durante estos meses.

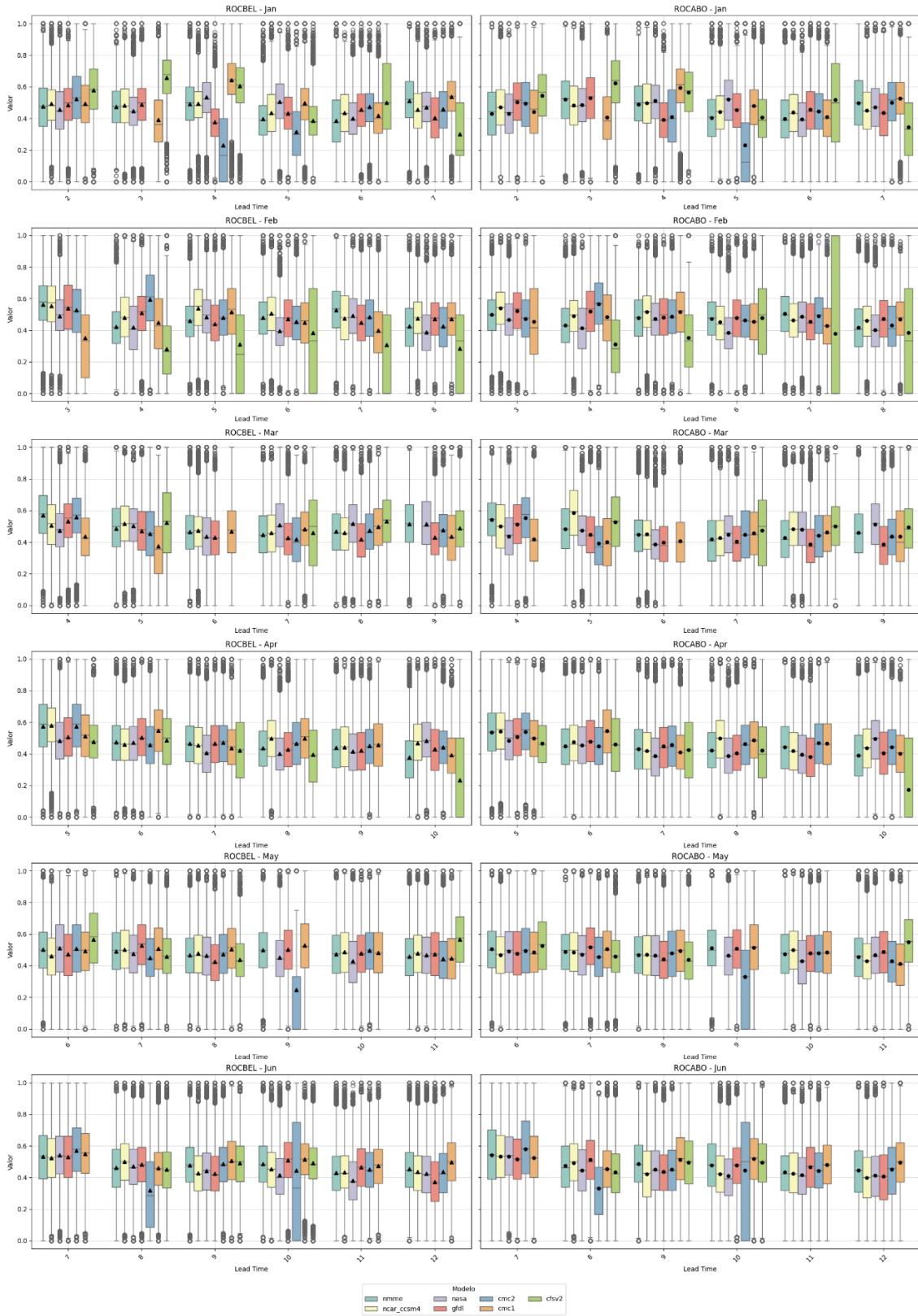


Figura 15. Comparativo de ROC Below (ROCBEL) y ROC Above (ROCABO) por Mes y Lead Time (Enero - Junio)

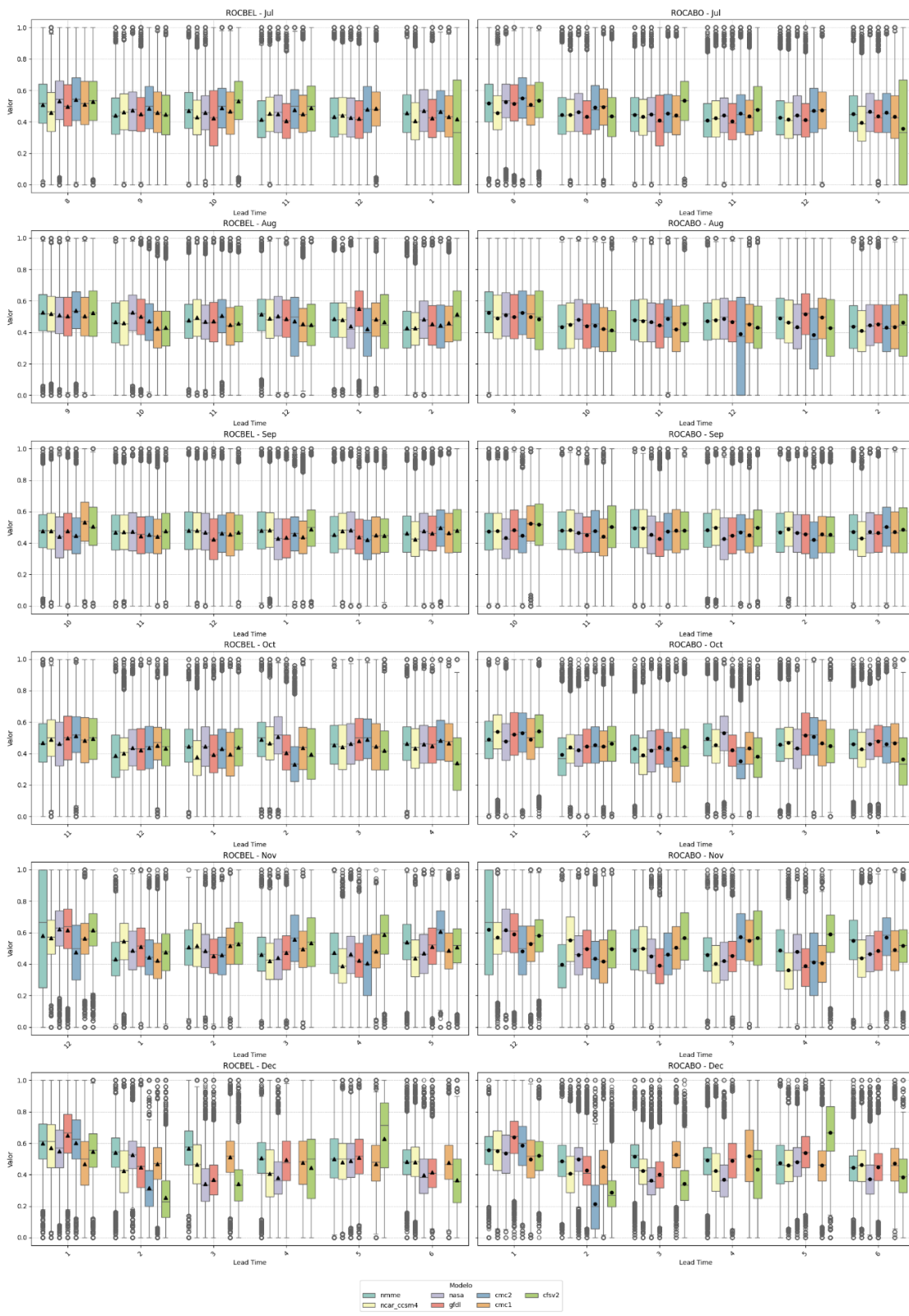


Figura 16. Comparativo de ROC Below (ROCBEL) y ROC Above (ROCABO) por Mes y Lead Time (Julio - Diciembre)

Conclusiones

De acuerdo con el análisis de la correlación canónica y los estadísticos de Pearson y Spearman, sugiere que los modelos **CFSv2** y **CMC2** son los más fiables para la predicción tanto estacional como por lead time, con una menor degradación del rendimiento en horizontes largos. Por otro lado, **NMME** y **CMC1** enfrentan más dificultades, especialmente en meses dentro de la barrera de la primavera y en predicciones a largo plazo.

Las métricas de error (RMSE y MAE) confirman que **CFSv2**, **NASA** y **GFDL** son las opciones más estables para la predicción de precipitación, especialmente en los primeros meses del año. Estos modelos mantienen su precisión en lead times largos, mientras que **NMME** y **CMC1** requieren ajustes adicionales para mejorar su consistencia, especialmente en periodos de alta variabilidad de la precipitación como abril y octubre, en gran parte del territorio colombiano.

El análisis del **Mean Bias** muestra que **CFSv2** y **GFDL** son los más consistentes a lo largo del año, con sesgos bajos y estables. En contraste, **NMME** y **CMC2** presentan mayor dispersión, especialmente en meses críticos como mayo y octubre. Esto resalta la necesidad de ajustar estos modelos para mejorar su precisión en meses clave.

El análisis del 2AFC refleja la **capacidad de los modelos para discriminar entre categorías de eventos**. En los primeros meses del año y hacia noviembre, modelos como **CFSv2** y **GFDL** mostraron valores altos de 2AFC, lo que indica una **alta habilidad discriminativa** en esos periodos, especialmente en **lead times cortos (1-3 meses)**.

Métricas como **HSS**, **LEPS** y **Gerrity Score** destacan la capacidad de discriminación de los modelos para eventos extremos. **CFSv2** y **GFDL** se destacan por mantener una habilidad predictiva superior, mientras que **CMC2** y **NCAR_CESM4** muestran mayor variabilidad en sus predicciones, especialmente en horizontes largos y en meses de transición

Recomendaciones

- **Predicciones a corto plazo (Lead times 1-3):**

Se recomienda utilizar cualquier modelo, dado que la mayoría muestra un rendimiento adecuado en estos horizontes.

CFSv2 y NASA destacan por su capacidad de ofrecer predicciones más estables y precisas en estos periodos.

- **Predicciones a largo plazo (Lead times 4-6):**

Para horizontes más largos, se recomienda utilizar modelos como CFSv2 o GFDL, que han demostrado mantener su precisión con menor degradación del rendimiento.

Evitar la dependencia exclusiva en NMME o CMC1 para periodos largos, especialmente en meses de alta variabilidad como octubre.

- **Ajustes y calibración:**

Realizar ajustes en los modelos CMC1, CMC2 y NMME para reducir la variabilidad en meses críticos (abril, mayo, octubre).

Incorporar técnicas de calibración para mejorar la capacidad de predicción de estos modelos en eventos extremos.

- **Uso operativo de la CPT:**

Integrar los modelos más estables en la CPT para mejorar las predicciones operativas, especialmente en periodos críticos como la transición entre estaciones.

Monitorear continuamente el rendimiento de los modelos y realizar ajustes iterativos para mejorar la confiabilidad y utilidad de las predicciones.

Referencias

Manrique-Suñén,, A., Gonzalez-Reviriego, N., Torralba, V., Cortesi, N., & Doblas-Reyes, F. J. (2020). *Choices in the verification of S2S forecasts and their implications for climate services*. Obtenido de *Monthly Weather Review*, 148(10), 3995–4008: <https://doi.org/10.1175/MWR-D-20-0067.1>

OMM, O. M. (2023). *WMO Library*. Obtenido de *Guidance on verification of operational seasonal climate forecasts (WMO-No. 1279)*: <https://library.wmo.int/es/records/item/56227-guidance-on-verification-of-operational-seasonal-climate-forecasts>