



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA CIVIL

**EVALUACIÓN DE LA METODOLOGÍA ESP PARA LA GENERACIÓN DE  
PRONÓSTICOS DE CAUDALES DE DESHIELO EN CUENCAS DE CHILE  
CENTRAL**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL

**DIEGO AGUSTÍN ARAYA REYDET**

PROFESOR GUÍA:  
PABLO MENDOZA ZÚÑIGA

MIEMBROS DE LA COMISIÓN:  
EDUARDO MUÑOZ CASTRO  
JAMES MCPHEE TORRES

SANTIAGO DE CHILE  
2022

RESUMEN DE LA MEMORIA PARA OPTAR  
AL TÍTULO DE INGENIERO CIVIL  
POR: **DIEGO AGUSTÍN ARAYA REYDET**  
FECHA: 2022  
PROF. GUÍA: PABLO MENDOZA ZÚÑIGA

## **EVALUACIÓN DE LA METODOLOGÍA ESP PARA LA GENERACIÓN DE PRONÓSTICOS DE CAUDALES DE DESHIELO EN CUENCAS DE CHILE CENTRAL**

Los pronósticos de caudales de deshielo son vitales para la gestión de recursos hídricos, operación de embalses, la determinación de disponibilidad hídrica para la agricultura, entre otros. Por lo tanto, es de gran interés desarrollar y evaluar métodos que permitan estimar caudales futuros. La presente investigación busca evaluar el desempeño de la metodología de predicción de caudales por conjuntos (ESP, por sus siglas en inglés), mediante la realización de pronósticos en retrospectiva, utilizando estructuras de modelos hidrológicos alternativas (GR4J, GR5J, GR6J y TUW), que cuentan con diferentes niveles de complejidad. El objetivo es pronosticar caudales de deshielo (periodo de septiembre/octubre a marzo) en 22 cuencas de Chile central (latitudes comprendidas entre 28°S y 37°S, aproximadamente), empleando siete funciones objetivo (FO) para calibrar y validar los modelos hidrológicos, y seis tiempos de inicialización para emitir los pronósticos (1° de mayo a 1° de octubre). Además, se realiza un proceso de evaluación de los pronósticos, empleando métricas que cuantifican sus propiedades probabilísticas y determinísticas; y un proceso de comparación con los pronósticos oficiales de la Dirección General de Aguas (DGA) de Chile.

Los resultados del proceso de calibración y validación muestran que el modelo TUW es el que mejor simula la serie completa de volúmenes diarios y estacionales, para todas las FO, aunque tiende a sobrestimar los volúmenes de magnitud media y subestimar los de magnitud baja. Con respecto a los pronósticos estacionales ESP, los resultados indican que, a medida que se emiten más cerca de septiembre/octubre, el error probabilístico y la dispersión del conjunto disminuyen. Se obtienen mejores pronósticos en las cuencas del centro de la zona de estudio (latitudes comprendidas entre 30°S y 34°S), las cuales cuentan con un alto flujo base, asociado a una respuesta más lenta, y con mayor acumulación de nieve. Las FO que mejores resultados entregan son las que minimizan los volúmenes estacionales (EVD<sub>Sep</sub> y EVD<sub>Oct</sub>), aunque no se desempeñan satisfactoriamente a resoluciones temporales más bajas (e.g., pronósticos mensuales o diarios). Con respecto a los modelos hidrológicos, TUW es el que entrega los mejores pronósticos, mejorando, en promedio, a los modelos GR en un 9% en cuanto a propiedades probabilísticas. En relación a los pronósticos emitidos en las cinco cuencas DGA, utilizando la mediana de los conjuntos, se supera el desempeño de la DGA en las cuencas Grande 1 y Maipo, en cuanto al sesgo y al error medio, mientras que en el resto de las cuencas, los resultados son cercanos a los oficiales.

Finalmente, el desempeño de los pronósticos estacionales ESP se condiciona fuertemente a (i) el tiempo de inicialización, (ii) la elección de la función objetivo, (iii) las características hidroclimáticas de las cuencas, y (iv) la elección del modelo hidrológico. De esta forma, la metodología se valida en la zona de estudio, y tiene el potencial de ser utilizada en Chile a un bajo costo computacional, generando pronósticos con buenas propiedades probabilísticas.

# Agradecimientos

En primer lugar, agradezco a mi familia por estar siempre. Una familia que es poco común, maravillosa, y capaz de entregar infinitas cantidades de amor. A mi hermana, mis primos, mis tíos, mi abuela, mis madres putativas y todos quienes han sido una constante en mi vida. Un agradecimiento especial a mi mamá y papá por el apoyo, el esfuerzo, los múltiples sacrificios y el amor que siempre me han entregado. Son los mejores de la vida.

Agradezco a los grandes amigos que hice en la U: Jorge, compañero de mil acontecimientos desde aquella lejana inducción mechona en 2015, imposible concebir la universidad sin tu constante compañía, apoyo, dedicación y tozudez que me permitieron evolucionar como persona. Gracias también por facilitarme el computador verdecito que hizo posible esta memoria. Karen, pilar fundamental desde primer año, especialmente en el apartado emocional, hemos crecido un montón y nadie puede negarlo. Sin tu constante apoyo y consejos que siempre fueron verdad quizás donde estaría. Orgulloso de ser tu amigo. Al Lalo, eterno compañero de millones de trabajos, y de las ventanas infinitas, nada hubiera sido igual sin ti. Jaco, el magíster más pro de todos, quien me enseñó firmeza, constancia y, especialmente, mucho amor propio. Gracias, en general, por tantos momentos, por la soportación, los infinitos aprendizajes, por tratar de darme el gusto siempre, el apañe hasta tarde en la U, las comidas de los viernes, las papitas, los mojitos, las caminatas por Estación Central, los cafés, las sonrisas, y mil cosas más. Sin ustedes nada hubiera sido posible, son los mejores también.

Agradezco al Alex, al Paul y al Chino, mis amigos extranjeros favoritos, a quienes conocí mejor en la recta final de sus magísteres, y aunque faltó más tiempo para compartir, valió cada segundo vivido. Al Joshi y al Bryan, quienes siempre han estado cuando los he necesitado. A la Javi y la Flo, el elenco estable que quedó del colegio. Y a todos los amigos que han ido y venido a lo largo de la carrera, quienes, de todas maneras, han influido en mi vida.

Un agradecimiento especial al Gary y la Arya, quienes llegaron a mi vida justo en el momento en que empecé la universidad, y me acompañaron día y noche, haciéndome feliz con sus locuras. Ojalá sean eternos.

Agradezco, por último, a mis profesores de la memoria: el profe Pablo, el Edu, el profe James, y también a la profe Kathy, siempre disponibles para ayudarme, aconsejarme y escucharme. Su pasión por lo que hacen es de las cosas que más admiro y más me quedarán.

Que vengan muchas más experiencias, desafíos, viajes y proyectos, para seguir creciendo juntos. Que nunca se acabe la alegría.

# Tabla de Contenido

<b>1. Introducción</b>	<b>1</b>
1.1. Introducción y motivación . . . . .	1
1.2. Objetivos . . . . .	2
1.3. Organización del documento . . . . .	2
<b>2. Revisión de la literatura</b>	<b>3</b>
2.1. Introducción general y objetivos . . . . .	3
2.2. Métodos de generación de pronósticos estacionales . . . . .	3
2.2.1. Enfoque estadístico . . . . .	3
2.2.2. Enfoque dinámico . . . . .	4
2.2.3. Enfoque híbrido . . . . .	4
2.3. Predicción de caudales por conjuntos . . . . .	5
2.4. Pre- y post-procesamiento, incorporación de información climática y variantes de ESP . . . . .	7
2.4.1. Pre-procesamiento . . . . .	7
2.4.2. Post-procesamiento . . . . .	9
2.4.3. Modificación de series climáticas . . . . .	10
2.5. Evaluación de conjuntos de pronósticos . . . . .	11
2.5.1. Evaluación determinística . . . . .	12
2.5.2. Evaluación probabilística . . . . .	12
<b>3. Descripción de la zona de estudio y datos utilizados</b>	<b>13</b>
<b>4. Metodología</b>	<b>17</b>
4.1. Descripción general . . . . .	17
4.2. Modelos utilizados . . . . .	18
4.2.1. Modelos GR . . . . .	18
4.2.2. Modelo TUW . . . . .	20
4.3. Estrategia de calibración . . . . .	22
4.4. Generación de pronósticos a través de ESP . . . . .	24
4.5. Evaluación de los pronósticos . . . . .	24
<b>5. Resultados</b>	<b>26</b>
5.1. Calibración y verificación . . . . .	26
5.1.1. Análisis de las cuencas DGA . . . . .	30
5.2. Generación y evaluación de los conjuntos de los pronósticos . . . . .	32
5.2.1. Pronósticos de volúmenes de deshielo de septiembre a marzo . . . . .	34
5.2.1.1. Mejores CRPSS . . . . .	37

5.2.1.2.	Cuencas DGA . . . . .	38
5.2.2.	Pronósticos de volúmenes de deshielo de octubre a marzo . . . . .	41
5.2.2.1.	Mejores CRPSS . . . . .	45
5.2.2.2.	Cuencas DGA . . . . .	46
5.3.	Análisis de desempeño . . . . .	49
5.3.1.	Desempeño según la clasificación de las observaciones . . . . .	49
5.3.2.	Desempeño según descriptores hidroclimáticos y físicos . . . . .	50
5.4.	Pronósticos en las cuencas DGA . . . . .	53
5.4.1.	Volúmenes estacionales . . . . .	53
5.4.2.	Volúmenes mensuales . . . . .	55
<b>6.</b>	<b>Conclusiones y recomendaciones</b>	<b>58</b>
6.1.	Conclusiones del estudio . . . . .	58
6.2.	Recomendaciones . . . . .	59
	<b>Bibliografía</b>	<b>61</b>
	<b>Anexos</b>	<b>66</b>
A.1.	Complemento ESP periodo de septiembre a marzo . . . . .	66
A.1.1.	Análisis por función objetivo de los pronósticos ESP . . . . .	66
A.1.2.	Mejores índices de confiabilidad $\alpha$ y $R^2$ . . . . .	69
A.1.3.	Análisis en detalle de los pronósticos ESP . . . . .	72
A.2.	Complemento ESP periodo de octubre a marzo . . . . .	82
A.2.1.	Análisis por función objetivo de los pronósticos ESP . . . . .	82
A.2.2.	Mejores índices de confiabilidad $\alpha$ y $R^2$ . . . . .	85
A.2.3.	Análisis en detalle de los pronósticos ESP . . . . .	88
A.3.	Complemento análisis de los pronósticos ESP a resolución temporal mensual de las cuencas DGA . . . . .	98

# Índice de Tablas

2.1.	Resumen de estudios sobre la metodología ESP. . . . .	6
2.2.	Resumen de estudios sobre incorporación de información climática en ESP. . . . .	10
3.1.	Descripción de los atributos presentados en la Tabla 3.2. . . . .	16
3.2.	Características hidroclimáticas (calculadas para el periodo de 1987/04 a 2020/03), físicas (obtenidas de CAMELS-CL) y flujo base (obtenido de Ladson et al., 2013) de las 22 cuencas en estudio. . . . .	16
4.1.	Descripción y rango de los parámetros de los modelos GR. . . . .	20
4.2.	Descripción de los parámetros y rangos utilizados para la calibración del modelo TUW. . . . .	22
4.3.	Funciones objetivo para calibrar los modelos. . . . .	23
4.4.	Métricas utilizadas para evaluar los pronósticos (Adaptado de Mendoza et al., 2017). . . . .	24
5.1.	Comparación del promedio de las métricas de los pronósticos emitidos en septiembre para cada modelo. Se destaca la diferencia entre los dos mejores modelos (generalmente los modelos TUW y GR6J). El <i>p-value</i> hace referencia al nivel de significancia de la diferencia entre las medias de los conjuntos que componen cada métrica. Cabe destacar que el promedio no es necesariamente el mejor indicador para evaluar el rendimiento, y sólo se utiliza en este caso para mostrar de manera sencilla el desempeño general de los modelos y sus diferencias. . . . .	37
5.2.	Comparación del promedio de las métricas de los pronósticos emitidos en octubre para cada modelo. Se destaca la diferencia entre los dos mejores modelos (TUW y GR6J). El <i>p-value</i> hace referencia al nivel de significancia de la diferencia entre las medias de los conjuntos que componen cada métrica. . . . .	45
5.3.	Evaluación del pronóstico estacional de la cuenca Grande 1. . . . .	53
5.4.	Evaluación del pronóstico estacional de la cuenca Choapa. . . . .	53
5.5.	Evaluación del pronóstico estacional de la cuenca Aconcagua. . . . .	53
5.6.	Evaluación del pronóstico estacional de la cuenca Maipo. . . . .	53
5.7.	Evaluación del pronóstico estacional de la cuenca Claro 2. . . . .	53
5.8.	Evaluación del pronóstico mensual de la cuenca Grande 1. . . . .	57
5.9.	Evaluación del pronóstico mensual de la cuenca Choapa. . . . .	57
5.10.	Evaluación del pronóstico mensual de la cuenca Aconcagua. . . . .	57
5.11.	Evaluación del pronóstico mensual de la cuenca Maipo. . . . .	57
5.12.	Evaluación del pronóstico mensual de la cuenca Claro 2. . . . .	57

# Índice de Figuras

2.1.	Esquema del método ESP. Fuente: Schaake (1978). . . . .	5
2.2.	La estrella roja es el valor CRPS del resultado de ejecutar ESP con el número completo de miembros. Los círculos azules son los valores CRPS óptimos teóricos para diferentes tamaños de conjuntos. $\Delta$ CRPS es todo el espacio de mejora posible teórico. Fuente: Sabzipour et al. (2021). . . . .	8
3.1.	Ubicación de las cuencas en estudio. . . . .	13
3.2.	(a) Coeficientes de Pardé de las cuencas en estudio para el periodo de 1987/04 - 2020/03, (b) Cantidad de años con observaciones de caudal en el mismo periodo de estudio. . . . .	14
3.3.	Características hidroclimáticas (calculadas para el periodo 1987/04 - 2020/03) y físicas (obtenidas de CAMELS-CL; Álvarez-Garreton et al., 2018) de las cuencas en estudio. . . . .	15
4.1.	Esquema general de la metodología, con énfasis en el método ESP y la evaluación de los pronósticos. . . . .	17
4.2.	Estructura de los modelos GR4J, GR5J y GR6J (adaptado de Pushpalatha et al., 2011). Los círculos dan cuenta de las forzantes del modelo, los cuadros grises corresponden a estanques, y el resto corresponde a flujos. . . . .	19
4.3.	Estructura del modelo TUW (adaptado de Neri et al., 2020). Los círculos dan cuenta de las forzantes del modelo, los cuadros grises corresponden a estanques, y el resto corresponde a flujos. . . . .	21
5.1.	Resultados del proceso de calibración-validación de las 22 cuencas en estudio (ordenadas de norte a sur, donde Pulido es la cuenca más norteña), para los cuatro modelos utilizados y la función objetivo KGE. Notar que los paneles (a), (b) y (c) no están en la misma escala. . . . .	26
5.2.	Resultados proceso de calibración-validación de las 22 cuencas en estudio, para los cuatro modelos utilizados y para la función objetivo SKGE. Notar que los paneles (a), (b) y (c) no están en la misma escala y que el desempeño de los modelos se presenta como la diferencia entre el KGE obtenido al evaluar el desempeño de los parámetros calibrados con SKGE y el KGE obtenido al evaluar el desempeño de los parámetros calibrados con KGE ( $\Delta$ KGE). . . . .	27
5.3.	eCDF de los resultados de la calibración y validación de las cuencas en estudio para todas las FO. Notar que los paneles (a) a (e) están positivamente orientados. Los paneles (f) y (g) están negativamente orientados, lo que implica que mientras más a la izquierda se encuentren las curvas, los modelos se desempeñan mejor. . . . .	28
5.4.	eCDF de los resultados de la calibración y validación de las cuencas en estudio para todas las FO, en términos de KGE. . . . .	29

5.5.	eCDF de los resultados de la calibración y validación de las cuencas en estudio para todas las FO, en términos de EVDSep normalizado (nEVDSep) por la desviación estándar, para evitar que cuencas con grandes volúmenes sesguen los resultados. . . . .	29
5.6.	Métricas e índices hidrológicos para los modelos GR4J y TUW calibrados con KGE. <b>SKGE de KGE</b> se refiere a Split KGE calculado con los parámetros calibrados con KGE; <b>sesgo de escorrentía</b> corresponde al sesgo total entre las simulaciones y observaciones (en ese orden); <b>sesgo SD</b> , es el sesgo entre la desviación estándar de la serie completa de la escorrentía media diaria simulada y observada; y <b>sesgo segmento alto, medio y bajo</b> hacen referencia al sesgo entre aquellos segmentos (probabilidad de excedencia menor a 2%, entre 20 y 70% y sobre 70%, respectivamente) de la curva de duración de los volúmenes diarios simulados y observados. Cabe destacar que, para KGE y SKGE el valor óptimo es 1, mientras que para nEVDSep y los sesgos el óptimo es 0. . . . .	30
5.7.	Métricas e índices hidrológicos para los modelos GR4J y TUW calibrados con SKGE. . . . .	31
5.8.	Métricas e índices hidrológicos para los modelos GR4J y TUW calibrados con EVDSep. . . . .	32
5.9.	Pronósticos de volumen de septiembre a marzo para la cuenca Maipo en el Manzano, utilizando el modelo GR4J, la función objetivo EVDSep, y tiempos de inicialización 1° mayo, 1° de julio y 1° de septiembre. Cada <i>boxplot</i> se compone de 32 miembros que constituyen el conjunto de pronósticos. La barra central es la mediana del conjunto, los bordes de la caja corresponden al primer y tercer cuartil, y los extremos a los valores mínimos y máximos que se encuentran dentro del rango borde + intercuartil (diferencia entre el tercer y primer cuartil). . . . .	33
5.10.	Comparación de los valores de CRPSS de los pronósticos ESP de septiembre a marzo, obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y modelo hidrológico. Cada <i>boxplot</i> se compone del CRPSS de las 22 cuencas en estudio. El <i>p-value</i> indica la probabilidad de que la media del mejor conjunto sea significativamente similar a la del segundo mejor. Vale decir, un bajo <i>p-value</i> indica que las medias son significativamente diferentes. El círculo coloreado indica cual función objetivo de cada modelo entrega el mejor CRPSS, mientras que el cuadrado indica cual es el modelo con la media más alta de entre todos. Notar que los paneles no se encuentran en la misma escala. . . . .	34
5.11.	Comparación de los valores del índice $\alpha$ de los pronósticos ESP de septiembre a marzo, obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y modelo hidrológico. Notar que los paneles cuentan con la misma escala. . . . .	35
5.12.	Comparación de los valores de $R^2$ de los pronósticos ESP de septiembre a marzo, obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y modelo. Notar que los paneles no se encuentran en la misma escala. . . . .	36
5.13.	Distribución espacial de los mejores CRPSS del periodo de septiembre a marzo (paneles de la izquierda), junto a las funciones objetivo que los generan (paneles de la derecha), para todas las cuencas, los modelos, y cuatro tiempos de inicialización. . . . .	38

5.14.	Comparación de los valores de CRPSS de las cuencas DGA por función objetivo, tiempo de inicialización y modelo. La banda gris define los límites de confianza al 95 % de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> . La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de septiembre. . . . .	39
5.15.	Comparación de los valores del índice $\alpha$ de las cuencas DGA por función objetivo, tiempo de inicialización y modelo. La banda gris define los límites de confianza al 95 % de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> . La línea discontinua demarca el máximo índice $\alpha$ obtenido para algún modelo el 1° de septiembre (tiempo de inicialización que no genera el mayor valor de $\alpha$ , pero que coincide con la inicialización de los pronósticos oficiales de la DGA). . . .	40
5.16.	Comparación de los valores de $R^2$ de las cuencas DGA por función objetivo, tiempo de inicialización y modelo. La banda gris define los límites de confianza al 95 % de la mejor FO de cada modelo obtenidos mediante <i>bootstrapping</i> . La línea discontinua demarca el máximo $R^2$ obtenido para algún modelo el 1° de septiembre. . . . .	41
5.17.	Comparación de los valores de CRPSS de los pronósticos ESP de octubre a marzo, obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y modelo hidrológico. Cada <i>boxplot</i> se compone del CRPSS de las 22 cuencas en estudio. El <i>p-value</i> indica la probabilidad de que la media del mejor conjunto sea significativamente similar a la del segundo mejor. Vale decir, un bajo <i>p-value</i> indica que las medias son significativamente diferentes. El círculo coloreado indica cual función objetivo de cada modelo entrega el mejor CRPSS, mientras que el cuadrado indica cual es el modelo con la media más alta de entre todos. Notar que los paneles no se encuentran en la misma escala. . . . .	42
5.18.	Comparación de los valores del índice $\alpha$ de los pronósticos ESP de octubre a marzo, obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y modelo. Notar que los paneles cuentan con la misma escala . .	43
5.19.	Comparación de los valores de $R^2$ de los pronósticos ESP de octubre a marzo, obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y modelo. Notar que los paneles no se encuentran en la misma escala. . . . .	44
5.20.	Distribución espacial de los mejores CRPSS del periodo de septiembre a marzo, junto a las funciones objetivo que los generan, para todos los modelos y tiempos de inicialización del 1° de julio al 1° de octubre. . . . .	46
5.21.	Comparación de los valores de CRPSS de las cuencas DGA por función objetivo, tiempo de inicialización y modelo. La banda gris define los límites de confianza al 95 % de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> . La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de octubre. . . . .	47
5.22.	Comparación de los valores del índice $\alpha$ de las cuencas DGA por función objetivo, tiempo de inicialización y modelo. La banda gris define los límites de confianza al 95 % de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> . La línea discontinua demarca el máximo índice $\alpha$ obtenido para algún modelo el 1° de octubre (tiempo de inicialización que no genera el mayor valor de $\alpha$ , pero que coincide con los pronósticos más hábiles). . . . .	48

5.23.	Comparación de los valores de $R^2$ de las cuencas DGA por función objetivo, tiempo de inicialización y modelo. La banda gris define los límites de confianza al 95 % de la mejor FO de cada modelo obtenidos mediante <i>bootstrapping</i> . La línea discontinua demarca el máximo $R^2$ obtenido para algún modelo el 1° de octubre. . . . .	49
5.24.	Análisis de desempeño del CRPS de los pronósticos al clasificar el volumen estacional observado en húmedo, normal y seco. La métrica nCRPS corresponde al CRPS normalizado por el volumen observado de septiembre a marzo de cada año, y varía entre $[0, \infty]$ , con 0 el valor óptimo. Además, cada <i>boxplot</i> se compone de todos los años húmedos/normales/secos de las 22 cuencas en estudio. . . .	50
5.25.	Relación entre el agrupamiento de cuencas y los descriptores de las 22 cuencas en estudio, para el modelo GR4J, la función objetivo KGE, periodo de septiembre a marzo y todos los tiempos de inicialización. El <i>p-value</i> indica si la diferencia entre las medias de cada grupo es significativa (menor <i>p-value</i> , mayor significancia). <b>F. Base</b> es el flujo base; <b>% Nieve</b> es la fracción de precipitación que cae como nieve; <b>Elevación</b> es la elevación media de las cuencas; <b>I. Aridez</b> es el índice de aridez; <b>Escorrentía</b> es el coeficiente de escorrentía; y <b>Variabilidad</b> es la variación inter-anual de la escorrentía. . . . .	51
5.26.	Relación entre el agrupamiento de cuencas y los descriptores de las 22 cuencas en estudio, para el modelo TUW, la función objetivo EVDOct, periodo de octubre a marzo y tiempos de inicialización del 1° de junio al 1° de octubre. <b>F. Base</b> es el flujo base; <b>% Nieve</b> es la fracción de precipitación que cae como nieve; <b>Elevación</b> es la elevación media de las cuencas; <b>I. Aridez</b> es el índice de aridez; <b>Escorrentía</b> es el coeficiente de escorrentía; y <b>Variabilidad</b> es la variación inter-anual de la escorrentía. . . . .	52
5.27.	Pronóstico determinístico de las cuencas DGA, inicializado el 1° de septiembre. $R_{ESP}^2$ y $R_{DGA}^2$ son las correlaciones entre la mediana del conjunto de pronósticos y las observaciones, y entre el pronóstico oficial de la DGA y las observaciones, respectivamente. El <i>p-value</i> indica el nivel de significancia de la correlación. Las observaciones corresponden al volumen de septiembre a marzo, y la correlación se realiza en la ventana de años en que la DGA tenga pronósticos (i.e., la correlación en Grande 1 se hace solo con las observaciones a partir de 2007, ya que desde ese año es oficialmente pronosticada). . . . .	54
5.28.	Comparación de los valores de CRPSS, índice $\alpha$ y coeficiente de determinación $R^2$ de los pronósticos ESP mensuales de septiembre a marzo, obtenidos con distintas funciones objetivo, agrupados por modelo e inicializados en septiembre. Las tres métricas fueron obtenidas de manera similar a como se muestra en la Figura 5.9, pero con el volumen desagregado por mes. Notar que los paneles no cuentan con la misma escala, y cada <i>boxplot</i> se compone de las cinco cuencas DGA en estudio. Los puntos coloreados indican las FO que entregan el máximo y mínimo valor de cada índice, y el nivel de significancia entre la diferencia de sus medias. Cabe destacar que el coeficiente de determinación se genera utilizando la mediana de los pronósticos ESP. . . . .	56

A.1.	Comparación de los valores de CRPSS de los pronósticos ESP, obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y función objetivo, para el periodo de septiembre a marzo. Cada <i>boxplot</i> se compone del CRPSS de las 22 cuencas en estudio. El <i>p-value</i> indica la probabilidad de que la media del mejor conjunto sea significativamente similar a la del segundo mejor. Vale decir, un bajo <i>p-value</i> indica que las medias son significativamente diferentes. El círculo coloreado indica cual función objetivo de cada modelo entrega el mejor CRPSS, mientras que el cuadrado indica cual es el modelo con la media más alta de entre todos. Notar que los paneles no se encuentran en la misma escala. . . . .	67
A.2.	Comparación de los valores del índice de confiabilidad $\alpha$ de los pronósticos ESP, obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y función objetivo, para el periodo de septiembre a marzo. Cada <i>boxplot</i> se compone del CRPSS de las 22 cuencas en estudio. El <i>p-value</i> indica la probabilidad de que la media del mejor conjunto sea significativamente similar a la del segundo mejor. Vale decir, un bajo <i>p-value</i> indica que las medias son significativamente diferentes. El círculo coloreado indica cual función objetivo de cada modelo entrega el mejor CRPSS, mientras que el cuadrado indica cual es el modelo con la media más alta de entre todos. Notar que los paneles no se encuentran en la misma escala. . . . .	68
A.3.	Comparación de los valores de $R^2$ de los pronósticos ESP, obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y función objetivo, para el periodo de septiembre a marzo. Cada <i>boxplot</i> se compone del CRPSS de las 22 cuencas en estudio. El <i>p-value</i> indica la probabilidad de que la media del mejor conjunto sea significativamente similar a la del segundo mejor. Vale decir, un bajo <i>p-value</i> indica que las medias son significativamente diferentes. El círculo coloreado indica cual función objetivo de cada modelo entrega el mejor CRPSS, mientras que el cuadrado indica cual es el modelo con la media más alta de entre todos. Notar que los paneles no se encuentran en la misma escala. . . . .	69
A.4.	Distribución espacial de los mejores índices de confiabilidad $\alpha$ del periodo de septiembre a marzo (paneles de la izquierda), junto a las funciones objetivo que los generan (paneles de la derecha), para todas las cuencas, los modelos, y cuatro tiempos de inicialización. . . . .	70
A.5.	Distribución espacial de los mejores $R^2$ del periodo de septiembre a marzo (paneles de la izquierda), junto a las funciones objetivo que los generan (paneles de la derecha), para todas las cuencas, los modelos, y cuatro tiempos de inicialización. . . . .	71
A.6.	Comparación de los valores de CRPSS de las cuencas 1 a 8, por función objetivo, tiempo de inicialización y modelo, para el periodo de septiembre a marzo. La banda gris define los límites de confianza al 95 % de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de septiembre. . . . .	73
A.7.	Comparación de los valores de CRPSS de las cuencas 9 a 15, por función objetivo, tiempo de inicialización y modelo, para el periodo de septiembre a marzo. La banda gris define los límites de confianza al 95 % de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de septiembre. . . . .	74

A.8.	Comparación de los valores de CRPSS de las cuencas 16 a 22, por función objetivo, tiempo de inicialización y modelo, para el periodo de septiembre a marzo. La banda gris define los límites de confianza al 95 % de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de septiembre. . . . .	75
A.9.	Comparación de los valores del índice de confiabilidad $\alpha$ de las cuencas 1 a 8, por función objetivo, tiempo de inicialización y modelo, para el periodo de septiembre a marzo. La banda gris define los límites de confianza al 95 % de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de septiembre. . . . .	76
A.10.	Comparación de los valores del índice de confiabilidad $\alpha$ de las cuencas 9 a 15, por función objetivo, tiempo de inicialización y modelo, para el periodo de septiembre a marzo. La banda gris define los límites de confianza al 95 % de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de septiembre. . . . .	77
A.11.	Comparación de los valores del índice de confiabilidad $\alpha$ de las cuencas 16 a 22, por función objetivo, tiempo de inicialización y modelo, para el periodo de septiembre a marzo. La banda gris define los límites de confianza al 95 % de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de septiembre. . . . .	78
A.12.	Comparación de los valores de $R^2$ de las cuencas 1 a 8, por función objetivo, tiempo de inicialización y modelo, para el periodo de septiembre a marzo. La banda gris define los límites de confianza al 95 % de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de septiembre. . . . .	79
A.13.	Comparación de los valores de $R^2$ de las cuencas 9 a 15, por función objetivo, tiempo de inicialización y modelo, para el periodo de septiembre a marzo. La banda gris define los límites de confianza al 95 % de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de septiembre. . . . .	80
A.14.	Comparación de los valores de $R^2$ de las cuencas 16 a 22, por función objetivo, tiempo de inicialización y modelo, para el periodo de septiembre a marzo. La banda gris define los límites de confianza al 95 % de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de septiembre. . . . .	81
A.15.	Comparación de los valores de CRPSS de los pronósticos (Oct - Mar) obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y función objetivo. . . . .	83
A.16.	Comparación de los valores del índice $\alpha$ de los pronósticos (Oct - Mar) obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y función objetivo. . . . .	84
A.17.	Comparación de los valores de $R^2$ de los pronósticos (Oct - Mar) obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y función objetivo. . . . .	85

A.18.	Distribución espacial de los mejores índices de confiabilidad $\alpha$ del periodo de octubre a marzo (paneles de la izquierda), junto a las funciones objetivo que los generan (paneles de la derecha), para todas las cuencas, los modelos, y cuatro tiempos de inicialización. . . . .	86
A.19.	Distribución espacial de los mejores $R^2$ del periodo de octubre a marzo (paneles de la izquierda), junto a las funciones objetivo que los generan (paneles de la derecha), para todas las cuencas, los modelos, y cuatro tiempos de inicialización. . . . .	87
A.20.	Comparación de los valores de CRPSS de las cuencas 1 a 8, por función objetivo, tiempo de inicialización y modelo, para el periodo de octubre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de octubre. . . . .	89
A.21.	Comparación de los valores de CRPSS de las cuencas 9 a 15, por función objetivo, tiempo de inicialización y modelo, para el periodo de octubre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de octubre. . . . .	90
A.22.	Comparación de los valores de CRPSS de las cuencas 16 a 22, por función objetivo, tiempo de inicialización y modelo, para el periodo de octubre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de octubre. . . . .	91
A.23.	Comparación de los valores del índice de confiabilidad $\alpha$ de las cuencas 1 a 8, por función objetivo, tiempo de inicialización y modelo, para el periodo de octubre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de octubre. . . . .	92
A.24.	Comparación de los valores del índice de confiabilidad $\alpha$ de las cuencas 9 a 15, por función objetivo, tiempo de inicialización y modelo, para el periodo de octubre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de octubre. . . . .	93
A.25.	Comparación de los valores del índice de confiabilidad $\alpha$ de las cuencas 16 a 22, por función objetivo, tiempo de inicialización y modelo, para el periodo de octubre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de octubre. . . . .	94
A.26.	Comparación de los valores de $R^2$ de las cuencas 1 a 8, por función objetivo, tiempo de inicialización y modelo, para el periodo de octubre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de octubre. . . . .	95

A.27.	Comparación de los valores de $R^2$ de las cuencas 9 a 15, por función objetivo, tiempo de inicialización y modelo, para el periodo de octubre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de octubre. . . . .	96
A.28.	Comparación de los valores de $R^2$ de las cuencas 16 a 22, por función objetivo, tiempo de inicialización y modelo, para el periodo de octubre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante <i>bootstrapping</i> con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de octubre. . . . .	97
A.29.	Comparación de los valores de CRPSS de los pronósticos a resolución temporal mensual de las cuencas DGA por función objetivo, tiempo de inicialización y modelo. . . . .	98
A.30.	Comparación de los valores del índice $\alpha$ de los pronósticos a resolución temporal mensual de las cuencas DGA por función objetivo, tiempo de inicialización y modelo. . . . .	99
A.31.	Comparación de los valores de $R^2$ de los pronósticos a resolución temporal mensual de las cuencas DGA por función objetivo, tiempo de inicialización y modelo. . . . .	100

# Capítulo 1

## Introducción

### 1.1. Introducción y motivación

Los pronósticos estacionales de caudal son importantes y beneficiosos en muchos sectores, incluida la gestión de recursos hídricos, operación de embalses, planificación de eventos extremos como crecidas o sequías y la determinación de disponibilidad hídrica para cultivos agrícolas (Slater y Villarini, 2018). Por lo tanto, es de gran interés desarrollar y evaluar metodologías que permitan estimar los caudales futuros con gran exactitud.

En la mayoría de los casos, la generación de pronósticos de caudal involucra el uso de modelos hidrológicos, los cuales son forzados con distintas variables meteorológicas (como precipitación, temperatura del aire, velocidad del viento, entre otros). Existe una gran cantidad de modelos hidrológicos para diversos propósitos (e.g., pronósticos estacionales, planificación de riego, control de inundaciones). Un ejemplo corresponde a la familia de modelos conceptuales de precipitación-escorrentía *Génie Rural* (GR; Perrin et al., 2003; Pushpalatha et al., 2011), cuya entrada son los registros históricos de precipitación y evapotranspiración potencial del sistema a modelar. Dichos modelos se caracterizan por ser fácilmente implementables y por su limitado requerimiento de datos (Coron et al., 2017).

No obstante, los modelos son una representación simplificada de la realidad, lo que inherentemente conlleva múltiples incertezas que deben ser abordadas para generar un pronóstico de calidad. Según Liu y Gupta (2007), el límite del sistema, la elección del modelo, la calibración de parámetros, la selección de condiciones iniciales y la calidad de las observaciones hidrometeorológicas representan las cinco mayores fuentes de incertidumbre en la modelación hidrológica. En este contexto, el método *Ensemble Streamflow Prediction* (ESP; Twedt et al., 1977; Day, 1985) incorpora al pronóstico la incertidumbre de la meteorología a través de los registros climáticos históricos. El resultado es un conjunto de escenarios que permite derivar la probabilidad de ocurrencia futura de caudales estacionales (Gobena y Gan, 2010). Este enfoque es utilizado operativamente en países como Estados Unidos (e.g. Franz et al., 2003) y el Reino Unido (e.g. Prudhomme et al., 2017), y diversos estudios han probado su efectividad en cuencas dominadas por el deshielo en países como el Reino Unido, Irlanda y Suecia (e.g. Harrigan et al., 2018; Donegan et al., 2020; Girons Lopez et al., 2021).

Por otro lado, el régimen hidrológico del centro y sur de Chile se caracteriza por contar con

una marcada estación pluvial (junio a agosto) y presentar acumulación de nieve en cuencas con cabecera en la cordillera de los Andes, donde los caudales de primavera-verano se deben principalmente al derretimiento del manto nival (Rubio-Álvarez y McPhee, 2010).

La principal motivación de este estudio es evaluar una metodología diferente a la que actualmente emplea la Dirección General de Aguas (DGA) para realizar sus pronósticos de deshielo en Chile, la cual se basa en modelos estadísticos alimentados con mediciones de altura de agua (e.g. equivalente en agua nieve, precipitación, caudal). Una ventaja de los modelos dinámicos, respecto a los estadísticos, es que entregan diversos flujos, como el equivalente en agua de nieve (SWE) o la evaporación real. Además, debido a que muchas actividades dependen de la adecuada predicción de caudales de deshielo, es importante contar con diversas herramientas y metodologías para predecirlos.

Así, el presente estudio se enfoca en 22 cuencas del centro de Chile, cuyos caudales de primavera-verano se deben principalmente al deshielo. Se busca evaluar el desempeño de la metodología ESP, utilizando estructuras de modelos alternativas, mediante la realización de pronósticos en retrospectiva (o *hindcasting*), y comparar su habilidad predictiva con aquellos pronósticos generados por la DGA.

## 1.2. Objetivos

### General

Evaluar la aplicabilidad de la metodología ESP para la generación de pronósticos de caudales de deshielo en cuencas del centro y sur de Chile.

### Específicos

- Seleccionar cuencas con un registro de datos continuo y extenso para configurar, calibrar y validar la familia de modelos hidrológicos GR y el modelo TUW.
- Implementar el método ESP para generar pronósticos con diferentes tiempos de inicialización, utilizando parámetros provenientes de distintos enfoques de calibración.
- Evaluar el desempeño de los pronósticos con respecto a los caudales observados en las estaciones fluviométricas de las cuencas, y compararlos con los pronósticos emitidos por la DGA.

## 1.3. Organización del documento

El Capítulo 2 presenta la revisión bibliográfica, con énfasis en el método ESP. En el Capítulo 3 se describe la zona de estudio. En el Capítulo 4 se presenta la metodología del estudio, donde se explican brevemente los modelos utilizados, las funciones objetivo empleadas para calibrarlos, la generación de pronósticos mediante ESP, y las métricas de evaluación usadas. El Capítulo 5 presenta los resultados obtenidos y su discusión. Finalmente, el Capítulo 6 incluye las conclusiones del estudio y recomendaciones para próximos trabajos.

# Capítulo 2

## Revisión de la literatura

### 2.1. Introducción general y objetivos

La generación de pronósticos de caudal estacional lleva aproximadamente un siglo de desarrollo. Algunas de las primeras predicciones utilizaban técnicas gráficas para relacionar las profundidades de nieve, medidas manualmente en las cabeceras de las cuencas, con los volúmenes de deshielo (Wood et al., 2018). En la actualidad, existen diversos enfoques, entre los cuales el método *Ensemble Streamflow Prediction* (ESP) se destaca por tener un bajo costo computacional y, por lo tanto, ser ampliamente utilizado en la comunidad científica (Harrigan et al., 2018).

El objetivo de esta sección es presentar los principales enfoques para generar pronósticos de caudales estacionales de caudal, haciendo un énfasis en el método ESP, sus principales variantes y métodos de verificación y/o evaluación a través de métricas determinísticas y probabilísticas.

### 2.2. Métodos de generación de pronósticos estacionales

Existen tres enfoques para generar pronósticos de caudales: (i) estadísticos, que buscan desarrollar relaciones empíricas entre los caudales estacionales, observaciones in-situ en la cuenca, y variables climáticas de larga escala (e.g., Grantz et al., 2005; Mendoza et al., 2014); (ii) dinámicos, que utilizan datos meteorológicos para forzar modelos hidrológicos (e.g., Slater y Villarini, 2018; Sabzipour et al., 2021); e (iii) híbridos, que combinan los dos ya mencionados (e.g., Wood y Schaake, 2008).

#### 2.2.1. Enfoque estadístico

Las técnicas basadas en este enfoque incluyen métodos de regresión, métodos multivariados y/o modelos de series de tiempo para establecer relaciones empíricas entre predictandos (generalmente caudal) y datos observados o modelados de precipitación, temperatura, índices climáticos de gran escala, patrones de teleconexión climática, entre otros (Slater y Villarini, 2018; Mendoza et al., 2017).

Este enfoque es utilizado actualmente por el Servicio de Conservación de Recursos Naturales (NRCS, por sus siglas en inglés) para realizar pronósticos estacionales de caudal en el oeste de Estados Unidos, donde el deshielo suele dominar el ciclo anual de escorrentía (Garen et al., 2011; Pagano et al., 2004). Los datos utilizados para llevar a cabo sus previsiones son el equivalente en agua de nieve (SWE), la precipitación, el caudal, e índices climáticos.

En el contexto nacional, la DGA dispone de diversos modelos estadísticos para generar sus pronósticos de deshielo en Chile. Para proyectar el volumen de la temporada, se realizan muestreos del manto nival en 15 estaciones nivométricas que posee la DGA en la zona mediterránea del país, además de considerar las lluvias entre abril y agosto, y el caudal observado en agosto. Luego, dicho volumen es analizado, ajustado según juicio experto y distribuido para cada mes, dependiendo de las condiciones existentes en la cuenca y de la temporada prevista (DGA, comunicación personal, 2021).

### **2.2.2. Enfoque dinámico**

Los modelos dinámicos simulan los procesos hidrológicos de las cuencas, incluyendo la partición de la precipitación entre escorrentía y evapotranspiración; la acumulación y derretimiento de la nieve; flujos evapotranspirativos; generación de escorrentía y flujo base. Generalmente, se fuerzan con datos de salida de modelos de circulación general (GCM), aunque, también pueden ser forzados con condiciones hidrológicas iniciales y datos meteorológicos históricos, como es el caso del método ESP (Day, 1985). Para aplicar este enfoque, se pueden utilizar tanto modelos conceptuales (como la familia de modelos GR; Coron et al., 2017) como modelos de base física (como el modelo *Variable Infiltration Capacity*, VIC; Liang et al., 1994).

El Servicio Meteorológico Nacional de Estados Unidos (NWS, por sus siglas en inglés) realiza simulaciones dinámicas de caudal en el oeste de Estados Unidos, mediante la implementación del modelo hidrológico *Sacramento Soil Moisture Accounting* (SAC-SMA; Burnash et al., 1973), forzándolo con un conjunto de series meteorológicas históricas mediante el método ESP (Pagano et al., 2004; Mendoza et al., 2017).

### **2.2.3. Enfoque híbrido**

Si bien varias agencias utilizan los enfoques estadístico o dinámico de manera excluyente, los modelos estadístico-dinámicos tienen múltiples fortalezas: son computacionalmente eficientes, pueden integrarse y aprender de una amplia selección de datos de entrada no estacionarios (pronósticos de modelos climáticos, series de tiempo de observaciones in-situ, y teleconexiones, por mencionar algunos) y pueden beneficiarse del progreso estadístico reciente (Pagano et al., 2004). El post-procesamiento estadístico de un pronóstico dinámico también es considerado como un enfoque híbrido (Mendoza et al., 2017).

## 2.3. Predicción de caudales por conjuntos

La predicción de caudal por conjuntos (*Ensemble Streamflow Prediction*; ESP, por sus siglas en inglés) es un método que genera múltiples simulaciones en base a un modelo hidrológico forzado con datos meteorológicos de observaciones históricas. El enfoque supone que las condiciones hidrológicas iniciales (IHCs) y el modelo son perfectos, y que los eventos meteorológicos que ocurrieron en el pasado son representativos de la incertidumbre de aquellos que ocurrirán en el futuro. Dicho de otra manera, toda la incertidumbre de los resultados se asocia a la meteorología y no a posibles errores en la estimación de las IHCs ni a la estructura del modelo (Day, 1985; Mendoza et al., 2017).

El origen del método se remonta al Laboratorio de Investigaciones Hidrológicas del NWS, que inició un proyecto para desarrollar un programa de ESP en 1975 (Wood et al., 2016). El propósito era desarrollar un procedimiento preciso y eficiente capaz de predecir el volumen del caudal con diferentes tiempos de adelanto, que van desde pocos días hasta varios meses, incorporando objetivamente la incertidumbre meteorológica en los pronósticos para cuencas de régimen nival en el oeste de Estados Unidos (Day, 1985; Donegan et al., 2020).

El método ESP tradicional produce  $N$  predicciones de escorrentía para el periodo de interés, dados  $N$  años de datos meteorológicos observados. Luego, la incertidumbre sobre el pronóstico se estima a partir de la función de densidad de probabilidad (FDP) construida a partir de los  $N$  miembros del conjunto, bajo el supuesto de que éstos son equiprobables (Gobena y Gan, 2010).

En la Figura 2.1 se presenta un esquema del funcionamiento del método ESP. Para inicializar el modelo, se recurre a un proceso de *spin-up*, que consiste en forzar el modelo hidrológico con cierta cantidad de observaciones para obtener IHCs estables y, luego, se realizan los pronósticos. Normalmente interesa un intervalo de tiempo acotado, y no el pronóstico completo (e.g., del pronóstico de un año hidrológico, sólo interesan los meses de septiembre a marzo).

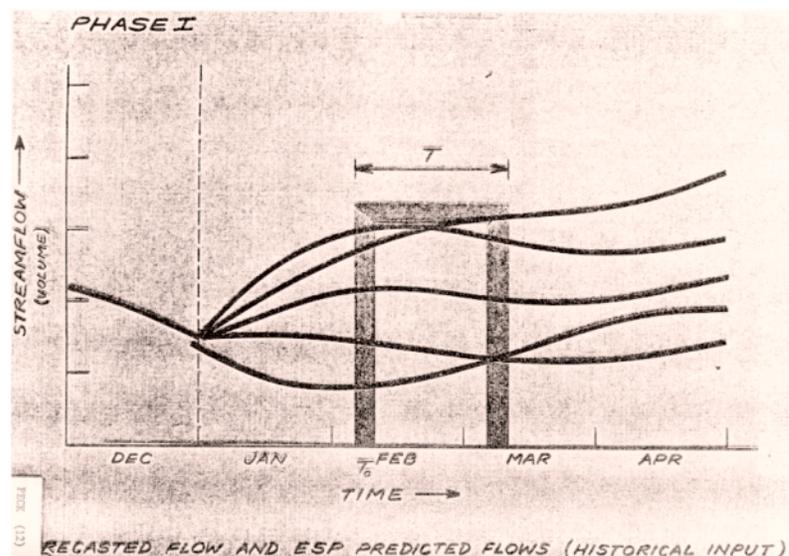


Figura 2.1: Esquema del método ESP. Fuente: Schaake (1978).

La ventaja de utilizar enfoques de predicción por conjuntos es que se genera una variedad de resultados posibles. Así, el método puede mostrar qué tan probables son los diferentes escenarios en los próximos días o meses y qué tan útiles pueden ser los pronósticos (Pechlivanidis et al., 2020). Además, los pronósticos por conjuntos contienen más información que un pronóstico determinístico, pues permiten obtener una caracterización idealizada de la incertidumbre, enriqueciendo el proceso de toma de decisiones (Renner et al., 2009).

Por otro lado, algunas limitaciones del método recaen en que, si la cantidad de años es pequeña, el rango de variación de la FDP podría ser pequeño y sesgado; o si la cuenca presenta regímenes variados, la FDP puede presentar mucha dispersión. Además, suponer que cada año es equiprobable no considera el conocimiento de posibles teleconexiones climáticas que permitan descartar eventos poco probables (Gobena y Gan, 2010). Asimismo, la dependencia de los pronósticos en eventos meteorológicos históricos hace que sea imposible capturar respuestas hidrológicas a eventos no registrados (Girons Lopez et al., 2021).

En varios países se utiliza el método ESP para el pronóstico estacional de caudales, incluyendo EE. UU., Rusia, República Checa, Eslovaquia, Alemania y Suecia, por mencionar algunos (Wood et al., 2016). Además, en la actualidad se siguen realizando estudios de verificación, comparación, e incluso de *benchmarking* (análisis y evaluación de algún producto respecto a una línea base) sobre esta metodología. Algunos ejemplos relevantes para esta memoria se resumen en la Tabla 2.1.

Tabla 2.1: Resumen de estudios sobre la metodología ESP.

Estudio	Ubicación y modelo	Algoritmo de optimización	Funciones objetivo	Principales resultados
<b>Wood y Lettenmaier (2008)</b>	Oeste de EE.UU, utilizando el modelo VIC	No relevante para este estudio	No relevante para este estudio	La importancia relativa de las IHCs varía durante el año, según el tiempo de inicialización y el de adelanto
<b>Harrigan et al. (2018)</b>	314 cuencas del Reino Unido, utilizando el modelo GR4J (Perrin et al., 2003)	Algoritmo incorporado en el paquete “airGR” de R (Coron et al., 2017), basado en búsqueda local	Eficiencia de Kling-Gupta modificada (Gupta et al., 2009; Kling et al., 2012)	ESP es mejor que la climatología promedio; su habilidad decae cuando aumenta el tiempo de adelanto; y se desempeña mejor en cuencas de respuesta lenta
<b>Donegan et al. (2020)</b>	46 cuencas de Irlanda, utilizando el modelo GR4J	Esquema evolutivo mediante la exploración de regiones prometedoras con búsqueda local	Eficiencia de Kling-Gupta no paramétrica (Pool et al., 2018)	ESP es mejor que la climatología promedio; su habilidad decae cuando aumenta el tiempo de adelanto; y se desempeña mejor en cuencas de respuesta lenta

Estudio	Ubicación y modelo	Algoritmo de optimización	Funciones objetivo	Principales resultados
<b>Muñoz-Castro et al. (2020)</b>	13 cuencas de alta montaña en Chile central, utilizando el modelo GR4J acoplado al módulo nival CemaNeige (Valéry et al., 2014)	Algoritmo de optimización global <i>Shuffled Complex Evolution</i> (SCE; Duan et al., 1992)	Eficiencias de Kling-Gupta y de Nash-Sutcliffe (Nash y Sutcliffe, 1970), Split Kling-Gupta (Fowler et al., 2018) y la raíz del error cuadrático medio	ESP mejora cuando la contribución nival tiene mayor importancia y cuando se inicializa más próximo al 1 de octubre, debido al mayor conocimiento de las IHCs
<b>Girons Lopez et al. (2021)</b>	39.493 cuencas de Suecia, utilizando el modelo S-HYPE (Lindström et al., 2010)	Ajuste manual de los parámetros mediante juicio experto y evaluación visual	Eficiencias de Nash-Sutcliffe y de Kling-Gupta	Los pronósticos son hábiles en tiempos de adelanto de hasta 3 meses. Esta habilidad es menor en cuencas de respuesta rápida, y mejora en cuencas de respuesta lenta

## 2.4. Pre- y post-procesamiento, incorporación de información climática y variantes de ESP

Estudios anteriores demuestran que los resultados que entrega ESP pueden ser inconsistentes o estar segados (e.g., Kang et al., 2010; Gobena y Gan, 2010; Sabzipour et al., 2021). Las principales fuentes de incertidumbre están asociadas tanto a la información meteorológica utilizada para generar los conjuntos, como a la estructura del modelo hidrológico, los parámetros y las condiciones iniciales (Kang et al., 2010). Algunas formas de abordar estas limitantes son mediante esquemas de pre- y post-procesamiento, la incorporación de información climática, y/o la mejor estimación en las IHCs (Mendoza et al., 2017).

### 2.4.1. Pre-procesamiento

Sabzipour et al. (2021) realizaron una breve revisión de diversos métodos para escoger los años meteorológicos observados (miembros del conjunto). El objetivo es acercarse a la máxima habilidad teórica de un pronóstico ESP clásico, para un tiempo de adelanto de 90 días. Para calcular dicha habilidad, utilizaron un algoritmo de optimización genético, el cual, mediante *hindcasting*, busca la combinación de miembros que entreguen el mejor *Continuous Ranked Probability Score* (CRPS; Gneiting y Raftery, 2007; métrica probabilística de error

entre la curva de densidad de probabilidad de las predicciones y las observaciones, con valor ideal 0). Cabe destacar que el proceso de *hindcasting* utiliza todos los años que cuenten con datos meteorológicos, excluyendo el año que busca ser pronosticado.

La Figura 2.2 muestra un ejemplo ilustrativo de la optimización genética para diferentes tamaños de conjunto. La tendencia indica que a medida que disminuye el tamaño del conjunto, mejoran los valores de CRPS.

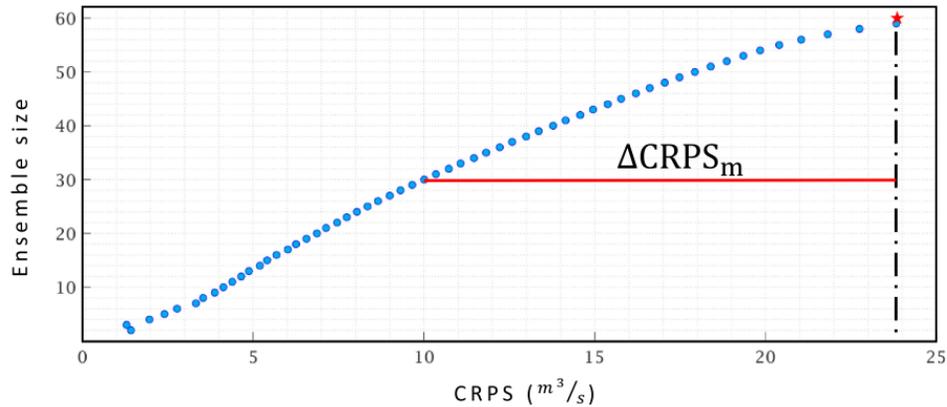


Figura 2.2: La estrella roja es el valor CRPS del resultado de ejecutar ESP con el número completo de miembros. Los círculos azules son los valores CRPS óptimos teóricos para diferentes tamaños de conjuntos.  $\Delta\text{CRPS}_m$  es todo el espacio de mejora posible teórico. Fuente: Sabzipour et al. (2021).

La dificultad, por lo tanto, es encontrar aquellos miembros que generan un mejor pronóstico futuro. Para esto, los autores probaron dos métodos:

- **Tabla de contingencia:** en este método, se categorizan los valores de precipitación en húmedos, normales y secos. Luego, se categorizan los registros de escorrentía y se determinan las probabilidades de que la temporada de escorrentía sea, por ejemplo, húmeda, dado que la temporada de la precipitación es seca. Finalmente, se elabora una tabla de contingencia con todas las probabilidades y, en función del pronóstico y categorización de la meteorología futura (precipitación en el ejemplo) se escoge un conjunto de miembros que cumplan con dicha categorización.
- **Método analógico:** este método supone que los miembros del conjunto de la meteorología histórica más adecuados son aquellos similares a la meteorología ocurrida inmediatamente antes de la fecha de pronóstico. Por lo tanto, se calcula la diferencia de algún estadístico (media, suma, máximo, entre otros) del predictor de cada miembro del conjunto (por ejemplo, precipitación), con el estadístico del predictor del periodo inmediatamente anterior al pronóstico, se ordenan ascendentemente, y se escogen los  $m$  mejores miembros.

Si bien los resultados no son ideales, las metodologías permiten descartar miembros improbables, y en algunos casos, mejorar la habilidad de pronóstico del método ESP. No obstante,

sigue existiendo espacio para mejorar. Los autores proponen, además, probar la eliminación de miembros utilizando índices climáticos y diversos predictores.

Cabe destacar que existen más métodos de pre-procesamiento ampliamente utilizados, incluyendo el procesamiento estadístico de pronósticos meteorológicos (e.g., Zalachori et al., 2012, Messner et al., 2014); enfoques de corrección mediante mapeo de cuantiles, según el clima esperado del periodo de pronóstico (e.g., Shi et al., 2008); o el reemplazo de miembros del conjunto por aquellos más probables (e.g., Schaake et al., 2007). No obstante, Sabzipour et al. (2021) critica estas metodologías, ya que emplean todos los años meteorológicos y no eliminan aquellos improbables.

## 2.4.2. Post-procesamiento

Los pronósticos subestacionales a estacionales pueden procesarse para corregir inconsistencias, sesgos, y agregar información. Existen diversos enfoques, algunos de los cuales se presentan a continuación:

1. **Corrección del sesgo:** Hashino et al. (2007) propusieron tres métodos para corregir el sesgo:
  - **Corrección de sesgo de evento:** este método supone que el mismo sesgo existe para cada simulación, por lo que se debe calcular un factor de error ( $Y_{obs}/Y_{sim}$ ) para corregir los pronósticos.
  - **Regresión:** consiste en ajustar una regresión que elimina el sesgo al reemplazar el volumen simulado con el valor esperado del caudal observado, dado el volumen simulado. Los coeficientes del modelo de regresión se estiman utilizando los caudales observados y simulados del registro histórico.
  - **Mapeo por cuantiles:** este método utiliza las distribuciones de probabilidad empírica de los caudales observados y simulados para corregir los sesgos del volumen simulado.
2. **Método de ponderación:** Hamlet y Lettenmaier (1999) demostraron que el uso de los índices climáticos El Niño Oscilación del Sur (ENSO, por sus siglas en inglés) y Oscilación Decadal del Pacífico (PDO, por sus siglas en inglés) para seleccionar ciertos miembros del conjunto mejora la habilidad de los pronósticos en una cuenca de Norteamérica en primavera-verano. Este enfoque asigna un peso igual a todos los miembros de la categoría que coincide con las señales climáticas, y un peso igual a cero para los miembros no coincidentes.

Werner et al. (2004) utilizaron índices climáticos para ponderar miembros del conjunto generados por ESP. En particular, El Niño-3.4 es aquel índice que muestra mejor correlación en cuencas del río Colorado, EE. UU. El esquema que asigna mayor peso a miembros del conjunto que más correlación presentan con el índice utilizado mostró un mejor desempeño, demostrando el valor potencial de utilizar esta técnica de post-procesamiento para pronósticos estacionales de caudal, especialmente en regiones cuya meteorología local presenta teleconexiones con las fases ENSO.

3. **Conjunto de modelos:** en este enfoque, que también puede ser considerado como una variante del ESP clásico, Bohn et al. (2010) combinaron, mediante promedio simple y regresión lineal múltiple, tres modelos hidrológicos: VIC, SAC-SMA y Noah (Chen et al., 1996). Los autores llevaron a cabo su estudio en tres cuencas de régimen nival en Estados Unidos. En términos de la raíz del error cuadrático medio, los enfoques que utilizaron múltiples modelos se desempeñaron mejor que aquellos que utilizaron modelos individuales.

### 2.4.3. Modificación de series climáticas

A diferencia de los estudios de Hamlet y Lettenmaier (1999) y Werner et al. (2004), los índices climáticos pueden ser incorporados de manera diferente, al modificar la metodología ESP clásica, tal y como se realizó en los estudios resumidos en la Tabla 2.2.

Tabla 2.2: Resumen de estudios sobre incorporación de información climática en ESP.

Estudio	Resumen
Šípek y Daňhelka (2015)	Los autores modificaron el método clásico ESP al reemplazar la entrada de series históricas observadas por series sintéticas, las cuales están restringidas según los patrones y variables climáticas de gran escala, y son generadas mediante LARS-WG (Semenov, 2007), un generador de clima estocástico. Los resultados fueron, en general, mejores en comparación con la versión clásica del ESP hasta con un mes de adelanto. Además, este método reduce la dispersión de los trazos de escorrentía pronosticados, mejorando, a la vez, la tasa de aciertos de eventos observados. No obstante, la calidad de los resultados podría verse influenciada por la cantidad de datos disponibles.
Gobena y Gan (2010)	Los autores desarrollaron una técnica para incorporar al método ESP pronósticos de temperatura y precipitación, con tiempos de adelanto de 0 a 3 meses, mediante dos modelos de predicción numérica. La técnica involucra una reducción de escala y desagregación temporal de los pronósticos mensuales que entregan los modelos numéricos, mediante el método <i>K-nearest neighbour</i> (KNN; Lall y Sharma, 1996), donde se selecciona un año del periodo histórico que esté más cerca del año de pronóstico en términos del índice climático PDO. Si bien los resultados no fueron mayormente positivos, los autores asociaron la incertidumbre a una insuficiente calibración del modelo utilizado, ya que tanto el ESP modificado, como los ESP clásicos que utilizaron como <i>benchmark</i> no entregaron pronósticos de calidad.

Estudio	Resumen
Beckers et al. (2016)	Los autores generaron una metodología en la cual el método ESP se condiciona a índices ENSO (el índice multivariado de ENSO entregó mejores correlaciones) en dos pasos. En primer lugar, se selecciona una serie de trazas históricas del conjunto ESP completo en función de la similitud entre el valor del índice ENSO en el año histórico y su valor en el momento del pronóstico. En el segundo paso, un remuestreador estocástico de clima condicionado a ENSO genera miembros de conjunto adicionales. Estos trazos remuestreados compensan la reducción de tamaño del conjunto en el primer paso y mejoran la habilidad de pronóstico de ESP. En dos de las tres estaciones utilizadas, se observa un mejoramiento de entre 5 a 10 % en la habilidad del método ESP para tiempos de adelanto superiores a 2 meses.

## 2.5. Evaluación de conjuntos de pronósticos

La evaluación y/o verificación es esencial para cualquier sistema de pronósticos, pues permite caracterizar su desempeño y compararlo con otros estándares (Franz et al., 2003). Además, permite analizar pronósticos desde diversas perspectivas (e.g., caudales bajos, caudales altos, volúmenes, velocidad de respuesta de la cuenca) y evaluar qué tan adecuados serían para diferentes aplicaciones (Murphy, 1993; Renner et al., 2009; Bradley et al., 2019).

Un pronóstico es “bueno” dependiendo de su objetivo y del punto de vista del evaluador. Murphy (1993) identificaron tres cualidades que un buen pronóstico debe tener: (i) **consistencia**, lo que implica que el pronóstico es coherente con el mejor juicio de un pronosticador basado en su conocimiento; (ii) la **calidad**, que es la capacidad de un pronóstico de predecir adecuadamente según ciertos criterios objetivos; y (iii) **valor**, que se refiere a cuán valioso es el pronóstico para tomar decisiones.

Mientras que la consistencia y el valor son cualidades subjetivas, la **calidad** de un pronóstico se puede evaluar objetivamente comparando los eventos pronosticados con las observaciones, y tiene diferentes atributos, entre los cuales se destacan (Murphy, 1993):

- **Sesgo (*bias*):** describe la diferencia entre las medias de los pronósticos y de las observaciones. Por ejemplo, si el promedio de los pronósticos es consistentemente mayor que el promedio de las observaciones, entonces el pronóstico tiene un sesgo positivo.
- **Precisión (*accuracy*):** mide la diferencia promedio entre los pronósticos y las observaciones. Esta diferencia se denomina error y, mientras más pequeño sea, más preciso es el pronóstico.
- **Habilidad (*skill*):** detalla la precisión de un pronóstico respecto a un pronóstico de referencia (e.g., el promedio de la hidroclimatología histórica, distinto modelo, u otro). Los pronósticos que son más precisos que la referencia se dice que son “hábiles” (*skillful*).

- **Fiabilidad (*reliability*):** corresponde a la concordancia entre valores pronosticados y observados (e.g., valores medios, cuantiles y métricas determinísticas similares).
- **Resolución (*resolution*):** indica si los resultados son efectivamente diferentes para distintos pronósticos emitidos, y si estos entregan información valiosa y utilizable.
- **Agudeza (*sharpness*):** define el grado de variabilidad de los pronósticos, y su tendencia a predecir valores extremos.

Se recomienda usar una amplia variedad de métricas para evaluar los diferentes aspectos de la calidad de un pronóstico, ya que una sola no permite describir todos los atributos asociados a éste (Bradley et al., 2019).

Por otro lado, la metodología ESP genera pronósticos probabilísticos que buscan cuantificar una incertidumbre idealizada a partir de un conjunto de resultados posibles. Sin embargo, a partir de estas probabilidades se puede obtener un pronóstico determinístico –i.e., pronóstico de valor único, sin declaración de incertidumbre (Wilks, 2011)– al tomar una medida de resumen, como por ejemplo la media o mediana del conjunto (Bradley et al., 2019).

De esta forma, según su finalidad y el tipo de pronóstico, las evaluaciones se pueden clasificar en determinísticas y probabilísticas:

### 2.5.1. Evaluación determinística

Este tipo de evaluación incluye métricas clásicas para verificar pronósticos determinísticos, como por ejemplo raíz del error cuadrático medio (RMSE), eficiencia de Nash-Sutcliffe (NSE), sesgo, coeficiente de correlación de Pearson ( $r$ ), el coeficiente de determinación ( $R^2$ ), entre otros. No obstante, el uso exclusivo de estas métricas hace que se pierda gran cantidad de información (Franz et al., 2003), especialmente en el caso del método ESP.

### 2.5.2. Evaluación probabilística

Los pronósticos probabilísticos se han popularizado en el último tiempo (Jordan et al., 2019), y existen distintas métricas para evaluar su calidad, las cuales asignan a los pronósticos una puntuación numérica basada en la distribución predicha y en los eventos observados (Gneiting y Raftery, 2007). Bradley et al. (2008) propusieron y resumieron algunas métricas de puntaje, como por ejemplo CRPS, *Proper Linear Score* (PLS; Wilson et al., 1999), y sus respectivas versiones *skill score* (i.e., evaluación con respecto a una referencia).

Sin embargo, resumir la calidad de un conjunto de pronósticos en algunas métricas es imposible debido a su naturaleza multifacética. Así, cualquier métrica inevitablemente oculta información importante sobre otros atributos (Bradley et al., 2008). Existen otras herramientas que cuantifican la fiabilidad y la agudeza del conjunto de pronósticos, como los índices  $\alpha$ ,  $\xi$  y  $\pi$  (Renard et al., 2010), mientras que los histogramas de rango son útiles para determinar su fiabilidad y diagnosticar errores en su media y dispersión (Hamill, 2001), aunque su interpretación no siempre es trivial.

# Capítulo 3

## Descripción de la zona de estudio y datos utilizados

La Figura 3.1 presenta la delimitación de las 22 cuencas que componen la zona de estudio, ubicadas entre las regiones de Atacama y Bío-Bío. Dichas cuencas se seleccionaron según los siguientes criterios: (i) índice de intervención humana menor a 0,05 (i.e., menos del 5% del caudal anual asignado a derechos de agua superficiales), (ii) no contar con grandes embalses, (iii) no contar con derechos de agua no consuntivos que devuelvan sus aguas abajo de la estación fluviométrica que delimita la cuenca, (iv) contar con una componente de deshielo en sus caudales (Figura 3.2) y (v) tener 25 o más años válidos (i.e., menos del 25% de sus datos faltantes) de observaciones de caudal en el periodo de abril de 1987 a marzo de 2020.

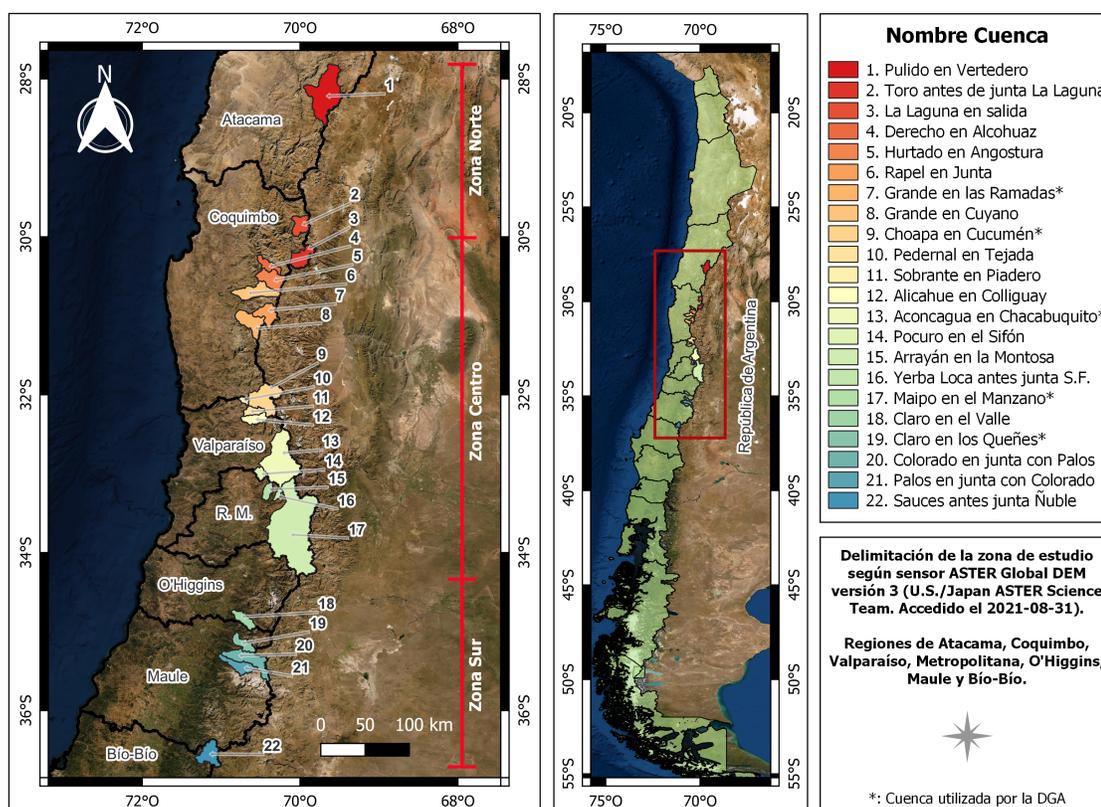


Figura 3.1: Ubicación de las cuencas en estudio.

La cuenca número 17, delimitada por la estación fluviométrica “río Maipo en el Manzano”, fue agregada a la zona de estudio sin cumplir los criterios de intervención, ya que es importante para el abastecimiento de agua en la región Metropolitana. De esta forma, 5 de las 19 cuencas que utiliza la DGA para realizar sus pronósticos anuales de deshielo son parte del estudio.

Se utilizaron datos a resolución temporal diaria de (i) caudal observado, (ii) precipitación, (iii) temperatura media del aire y (iv) evapotranspiración potencial, obtenidos desde la base de datos CAMELS-CL. Esta plataforma recupera los datos públicos de caudal desde las plataformas en línea de la Dirección Meteorológica de Chile (DMC) y la DGA. Los datos de precipitación y temperatura media son derivados a partir del producto grillado del CR2MET versión 2.0, a una resolución espacial de  $0,05^\circ$  y resolución temporal diaria, los que son recortados y promediados a escala de cuenca. Por último, la evapotranspiración potencial se obtiene a partir de la fórmula propuesta por Hargreaves y Samani (1985), y es promediada a escala de cuenca.

Adicionalmente, se utilizaron datos de elevación a partir del modelo de elevación digital del mapa topográfico *ASTER Global DEM versión 3.0* (U.S./Japan Aster Science Team) para generar curvas hipsométricas de las cuencas. También, se emplearon datos de caudal medio mensual rellenados (Quevedo Leiva, 2021) exclusivamente para la evaluación los pronósticos; y los datos de los pronósticos oficiales de la DGA, obtenidos de su portal en línea.

En la Figura 3.2 se presentan los coeficientes de Pardé de las 22 cuencas en estudio, y la distribución espacial de la cantidad de años válidos de observaciones de caudal en el periodo de estudio. Se aprecia como las cuencas, en general, cuentan con un régimen nivo-pluvial, donde los máximos de escorrentía ocurren entre octubre y diciembre. Además, las cuencas presentan una buena disponibilidad de años, teniendo en cuenta que el máximo es 33.

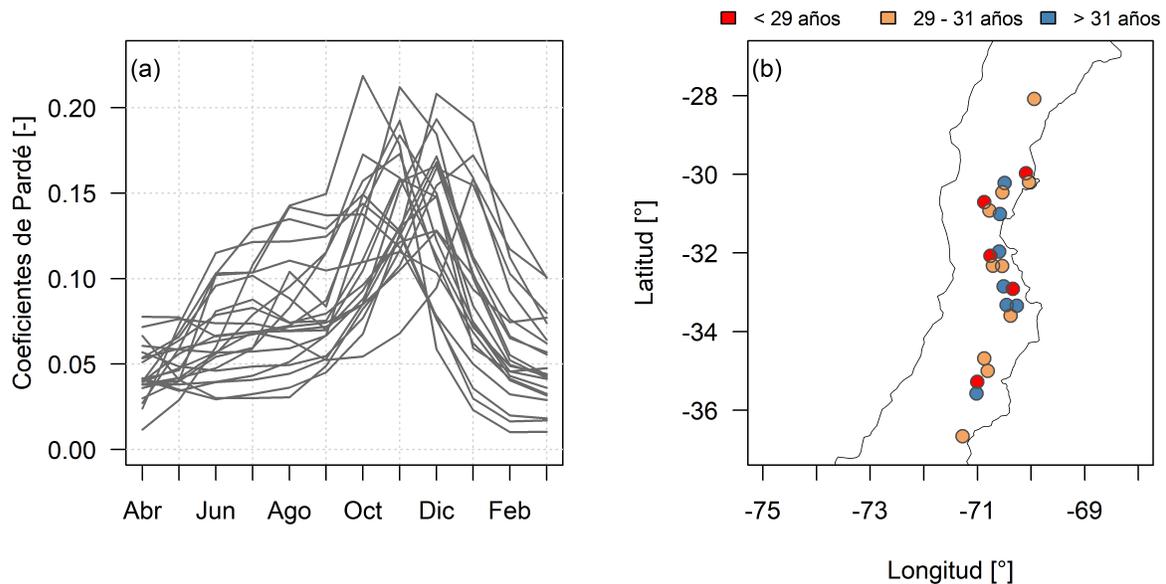


Figura 3.2: (a) Coeficientes de Pardé de las cuencas en estudio para el periodo de 1987/04 - 2020/03, (b) Cantidad de años con observaciones de caudal en el mismo periodo de estudio.

En la Figura 3.3 se muestran las principales características hidroclimáticas y físicas de las cuencas en estudio. Se observan grandes contrastes entre las cuencas del norte y del sur. Hacia el sur, aumentan los valores de caudal medio anual y precipitación media anual. La fracción de precipitación que cae como nieve es mayor en las cuencas del centro y del norte, mientras que en las cuencas del sur es baja. El índice de aridez disminuye hacia las cuencas del sur. Finalmente, la elevación media de las cuencas disminuye hacia el sur de manera consistente con la topografía cordillerana.

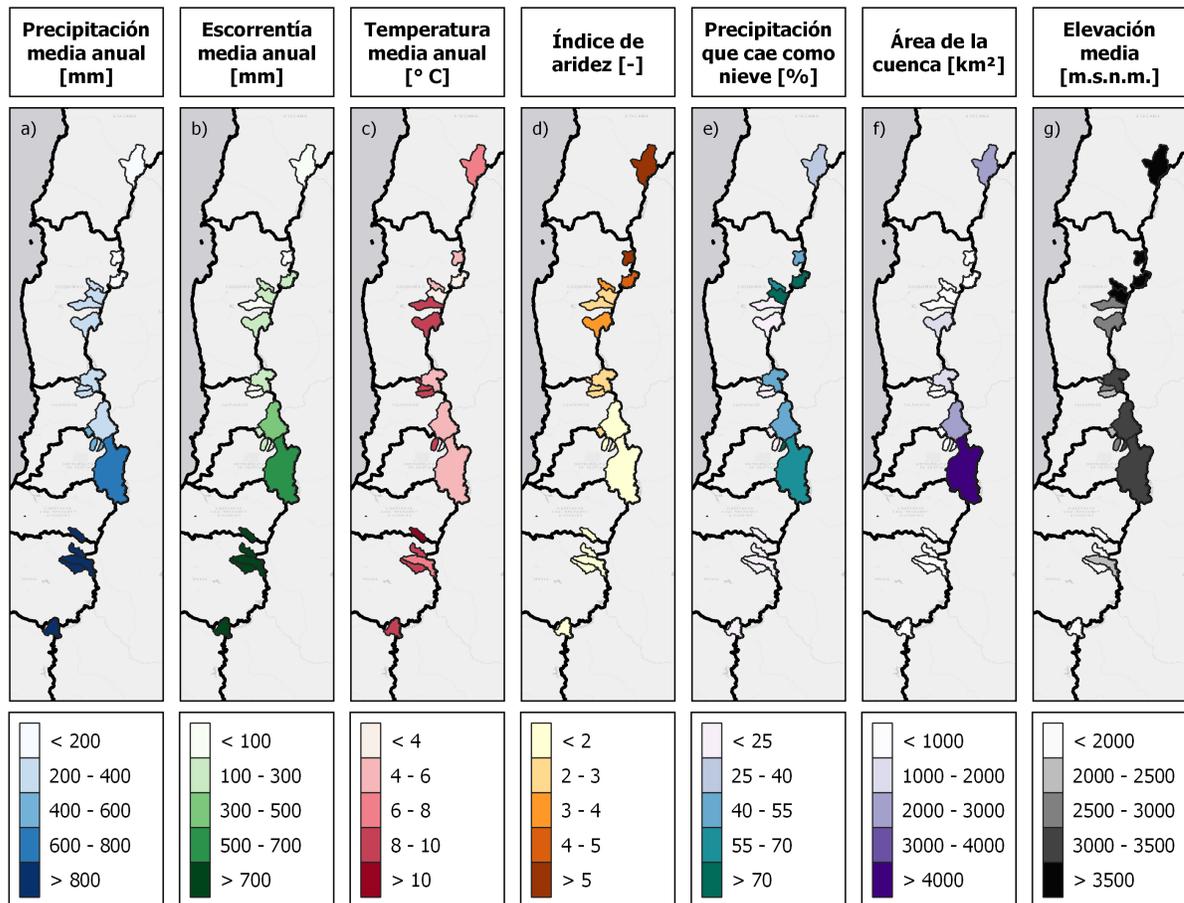


Figura 3.3: Características hidroclimáticas (calculadas para el periodo 1987/04 - 2020/03) y físicas (obtenidas de CAMELS-CL; Álvarez-Garretón et al., 2018) de las cuencas en estudio.

Además, en las Tablas 3.1 y 3.2 se detallan los valores de las características hidroclimáticas y físicas de las cuencas, agregando atributos adicionales como el coeficiente de escorrentía anual (fracción de precipitación que escurre como escorrentía), el coeficiente de variación inter-anual de la escorrentía (relación entre la desviación estándar y la media), y el flujo base de la cuenca (proporción entre el flujo base medio diario y escorrentía media diaria).

Tabla 3.1: Descripción de los atributos presentados en la Tabla 3.2.

Atributo	Descripción	Unidades
PMA	Precipitación media anual	mm
EMA	Escorrentía media anual	mm
TMA	Temperatura media anual	°C
IA	Índice de aridez	-
CE	Coficiente de escorrentía anual	-
FN	Fracción de precipitación que cae como nieve	%
CVE	Coficiente de variación de la escorrentía anual	-
Área	Área de la cuenca	km <sup>2</sup>
EM	Elevación media	m.s.n.m.
FB	Flujo base	-

Tabla 3.2: Características hidroclimáticas (calculadas para el periodo de 1987/04 a 2020/03), físicas (obtenidas de CAMELS-CL) y flujo base (obtenido de Ladson et al., 2013) de las 22 cuencas en estudio.

Código	Cuenca	PMA	EMA	TMA	IA	CE	FN	CVE	Área	EM	FB
3414001	Pulido	126	20	6,3	7,0	0,2	36	0,5	2022	3566	0,83
4302001	Toro	133	43	4,3	5,6	0,3	50	0,4	467	3905	-
4301002	Laguna	156	104	2,6	4,3	0,7	71	0,7	558	4275	0,77
4311001	Derecho	222	107	4,6	3,2	0,5	64	1,0	338	3544	0,81
4501001	Hurtado	251	128	3,9	2,4	0,5	70	0,8	672	3724	0,86
4522002	Rapel	313	71	8,7	3,0	0,2	12	1,3	821	2661	0,67
4511002	Grande 1	284	202	6,7	2,9	0,7	34	0,9	569	3098	0,78
4513001	Grande 2	277	147	8,8	3,6	0,7	8	1,0	1287	2726	0,75
4703002	Choapa	318	236	6,0	2,7	0,7	53	0,7	1132	3142	0,75
5101001	Pedernal	355	89	9,6	2,8	0,2	3	1,1	81	2454	-
5100001	Sobrante	397	90	8,3	2,3	0,3	13	1,1	241	2610	0,78
5200001	Alicahue	391	94	9,1	2,5	0,3	7	1,1	348	2403	0,78
5410002	Aconcagua	390	448	5,3	1,9	1,1	52	0,5	2113	3178	0,76
5411001	Pocuro	445	124	11,1	2,5	0,3	1	0,8	181	2006	0,72
5722001	Arrayán	560	207	8,1	1,6	0,4	18	0,6	216	2513	0,75
5721001	Yerba Loca	456	278	3,3	1,2	0,6	72	0,4	110	3424	-
5710001	Maipo	764	605	5,0	0,9	0,8	59	0,3	4839	3181	0,81
6027001	Claro 1	1297	702	10,7	0,8	0,6	1	0,6	349	1605	0,63
7103001	Claro 2	1568	1527	8,9	0,6	1,0	8	0,4	354	1857	0,70
7112001	Colorado	1650	1526	7,1	0,5	0,9	23	0,4	878	2288	0,73
7115001	Palos	1743	1532	8,1	0,5	0,9	14	0,3	490	1973	0,80
8104001	Sauces	1752	1581	8,3	0,6	0,9	10	0,4	607	1683	0,68

# Capítulo 4

## Metodología

### 4.1. Descripción general

Para emitir pronósticos de volúmenes de deshielo se buscó un conjunto de cuencas con suficiente disponibilidad de datos. Luego, se seleccionaron los modelos hidrológicos para representar la dinámica de las cuencas, cuyos parámetros fueron calibrados y validados. Para incorporar a los pronósticos la incertidumbre de la meteorología se implementó la metodología ESP. Finalmente, y para emitir un pronóstico respaldado, se realizó una evaluación para caracterizar y cuantificar su calidad.

En la Figura 4.1 se presenta un esquema de la metodología general llevada a cabo para la emisión de pronósticos de volúmenes de deshielo.

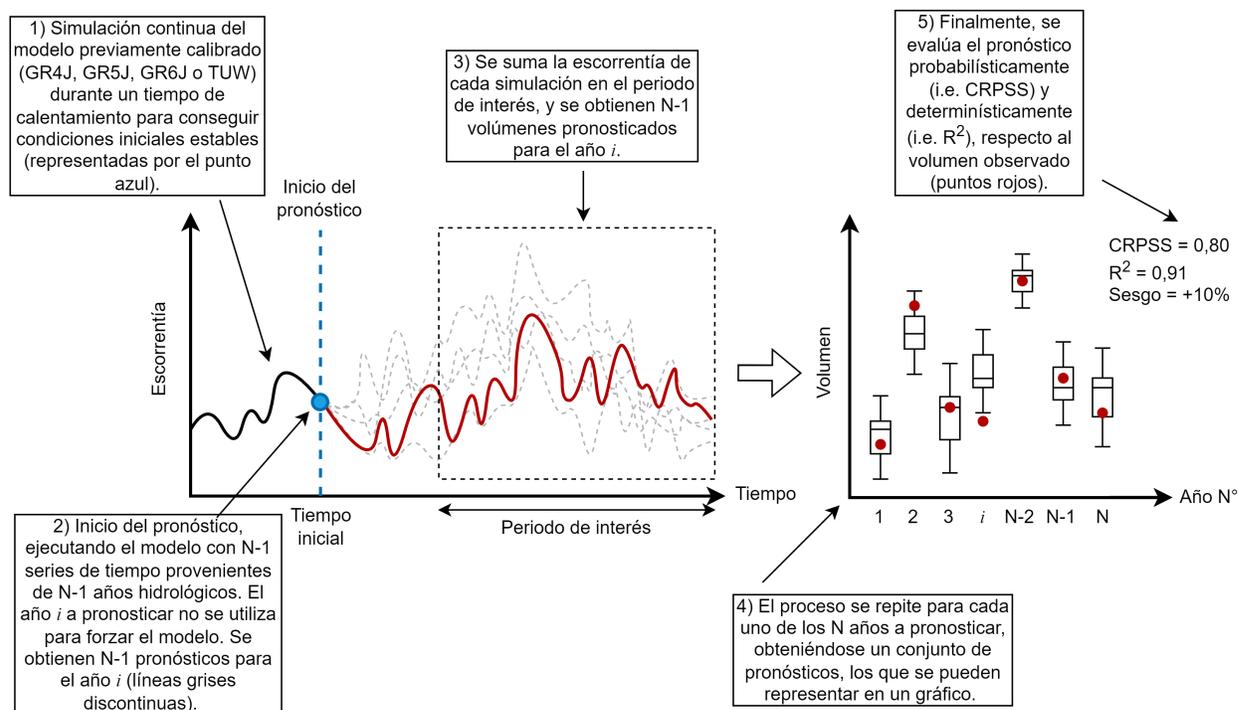


Figura 4.1: Esquema general de la metodología, con énfasis en el método ESP y la evaluación de los pronósticos.

El periodo de estudio va desde abril de 1987 hasta marzo de 2020, es decir, 33 años a ser pronosticados. Este periodo fue escogido con el fin de contar con una suficiente cantidad de años para llevar a cabo el proceso de calentamiento, y de garantizar que las cuencas cuenten con una mínima cantidad de observaciones de caudal, ya que varias no cuentan con datos antes de 1985.

En este estudio, se consideraron dos periodos para generar los pronósticos: (i) septiembre a marzo (periodo en el cual se realizan los pronósticos oficiales en Chile por parte de la DGA) y (ii) octubre a marzo. Además, se emplearon cinco tiempos de inicialización para pronosticar el volumen de escorrentía de septiembre a marzo: 1° de mayo, 1° de junio, 1° de julio, 1° de agosto y 1° de septiembre, y para octubre a marzo se agrega el 1° de octubre.

En las siguientes secciones se describen los modelos utilizados, el proceso de calibración, la emisión de pronósticos mediante ESP y su evaluación.

## 4.2. Modelos utilizados

### 4.2.1. Modelos GR

Los modelos *Génie Rural* (GR) son modelos conceptuales y concentrados de precipitación-escorrentía, originalmente desarrollados por el Instituto Nacional de Investigaciones para la Agricultura, la Alimentación y el Medio Ambiente (INRAE, por sus siglas en francés), y modificados por diversos autores para cubrir distintas necesidades. Estos modelos buscan representar matemáticamente la dinámica de los procesos hidrológicos mediante el uso de estanques, sin hacer referencia directa a expresiones basadas en la física. Requieren como entrada las series de precipitación (mm/día) y evapotranspiración potencial (mm/día).

Se utilizaron tres modelos GR a escala diaria, cuyos parámetros fueron calibrados según las observaciones de caudal del sistema a modelar:

- **GR4J (Perrin et al., 2003)**: consiste en dos estanques, uno de almacenamiento y otro de enrutamiento, y cuatro parámetros libres ( $X_1$ ,  $X_2$ ,  $X_3$  y  $X_4$ ) que controlan los almacenamientos y los flujos.
- **GR5J (Pushpalatha et al., 2011)**: incorpora una función mejorada de intercambio subterráneo, al agregar un parámetro libre  $X_5$  que permite revertir la dirección de los intercambios durante el año.
- **GR6J (Pushpalatha et al., 2011)**: suma un nuevo estanque exponencial que permite una mejor simulación de volúmenes bajos, y que está controlado por el parámetro libre  $X_6$ .

Los tres modelos están acoplados al módulo de nieves *CemaNeige* para mejorar la modelación de volúmenes de deshielo (Valéry et al., 2014). Este módulo está basado en un enfoque de derretimiento grado-día, y requiere como datos de entrada la precipitación diaria (mm), la temperatura media diaria (°C) y la curva hipsométrica para distribuir en diferentes bandas

de elevación las forzantes climáticas. Cuenta con un estanque de almacenamiento nival y dos parámetros libres ( $CN_1$  y  $CN_2$ ), que también deben ser calibrados según observaciones de caudal. A partir de este punto, cada vez que se haga referencia a los modelos GR se sobreentiende que incluyen al módulo *CemaNeige*.

La estructura de los modelos GR se muestra en la Figura 4.2. La descripción de los parámetros de los tres modelos y sus rangos se encuentran en la Tabla 4.1. Para mayores detalles en la formulación de los modelos, véase Perrin et al. (2003) y Pushpalatha et al. (2011).

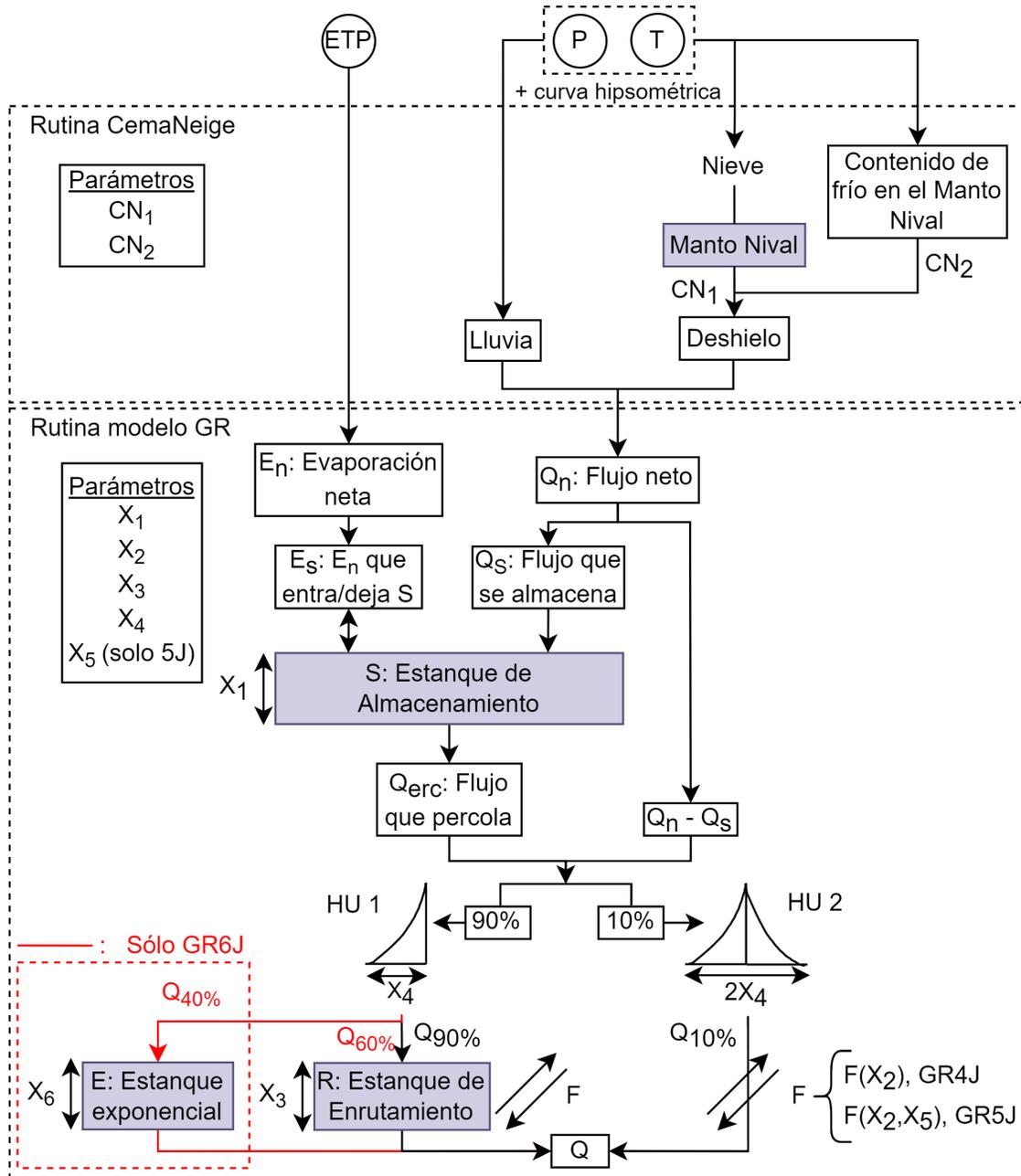


Figura 4.2: Estructura de los modelos GR4J, GR5J y GR6J (adaptado de Pushpalatha et al., 2011). Los círculos dan cuenta de las forzantes del modelo, los cuadros grises corresponden a estanques, y el resto corresponde a flujos.

Tabla 4.1: Descripción y rango de los parámetros de los modelos GR.

Parámetro	Descripción	Unidad	GR4J	GR5J	GR6J
$X_1$	Capacidad máxima del estanque de almacenamiento	mm	0 – 4000	0 – 1600	0 – 1500
$X_2$	Coefficiente de intercambio subterráneo	mm/d	-35 – 35	-4 – 4	-1 – 3
$X_3$	Capacidad máxima del estanque de enrutamiento	mm	0 – 2000	0 – 1000	0 – 1000
$X_4$	Tiempo base de los hidrogramas unitarios	d	0,5 – 20	0,5 – 30	0,5 – 20
$X_5$	Umbral del intercambio subterráneo	-	-	0 – 2	-1 – 2
$X_6$	Capacidad máxima del estanque exponencial	mm	-	-	0 – 200
$CN_1$	Coefficiente de ponderación del estado térmico del manto nival	-	0 – 1	0 – 1	0 – 1
$CN_2$	Coefficiente de derretimiento grado-día	mm/°C/d	0 – 10	0 – 10	0 – 10

La diferencia entre los rangos de los parámetros comunes de los modelos se debe a que sus estructuras son diferentes, y los valores se compensan por la incorporación de otros parámetros y/o estanques. Cabe destacar que los resultados al evaluar los modelos GR5J y GR6J con los rangos del GR4J (y anulando sus parámetros adicionales,  $X_5$  y  $X_6$ ), no dan igual al modelo GR4J.

#### 4.2.2. Modelo TUW

El *TUWmodel*, es un modelo conceptual y concentrado, desarrollado por la Universidad Tecnológica de Viena (TUW, por sus siglas en alemán). Al igual que los modelos GR, busca representar la dinámica hidrológica a través de estanques e intercambios de flujos a escala de cuenca, y requiere como entrada las series de precipitación (mm/día), temperatura diaria (°C) y evapotranspiración potencial (mm/día). Su estructura sigue la del modelo HBV (Bergström, 1976) y se puede observar en la Figura 4.3.

El modelo consiste en cuatro módulos principales: (i) rutina de nieve basada en un enfoque de derretimiento grado-día y cuenta con cinco parámetros libres; (ii) rutina de humedad del suelo con tres parámetros libres; (iii) rutina de respuesta, que cuenta con un almacenamiento superior y otro inferior, donde ocurren las respuestas rápidas y muy rápidas, y las lentas, respectivamente, y que cuenta con 5 parámetros libres; y (iv) rutina de enrutamiento, donde los flujos de salida son administrados según una función de transferencia triangular, y cuenta con 2 parámetros libres.

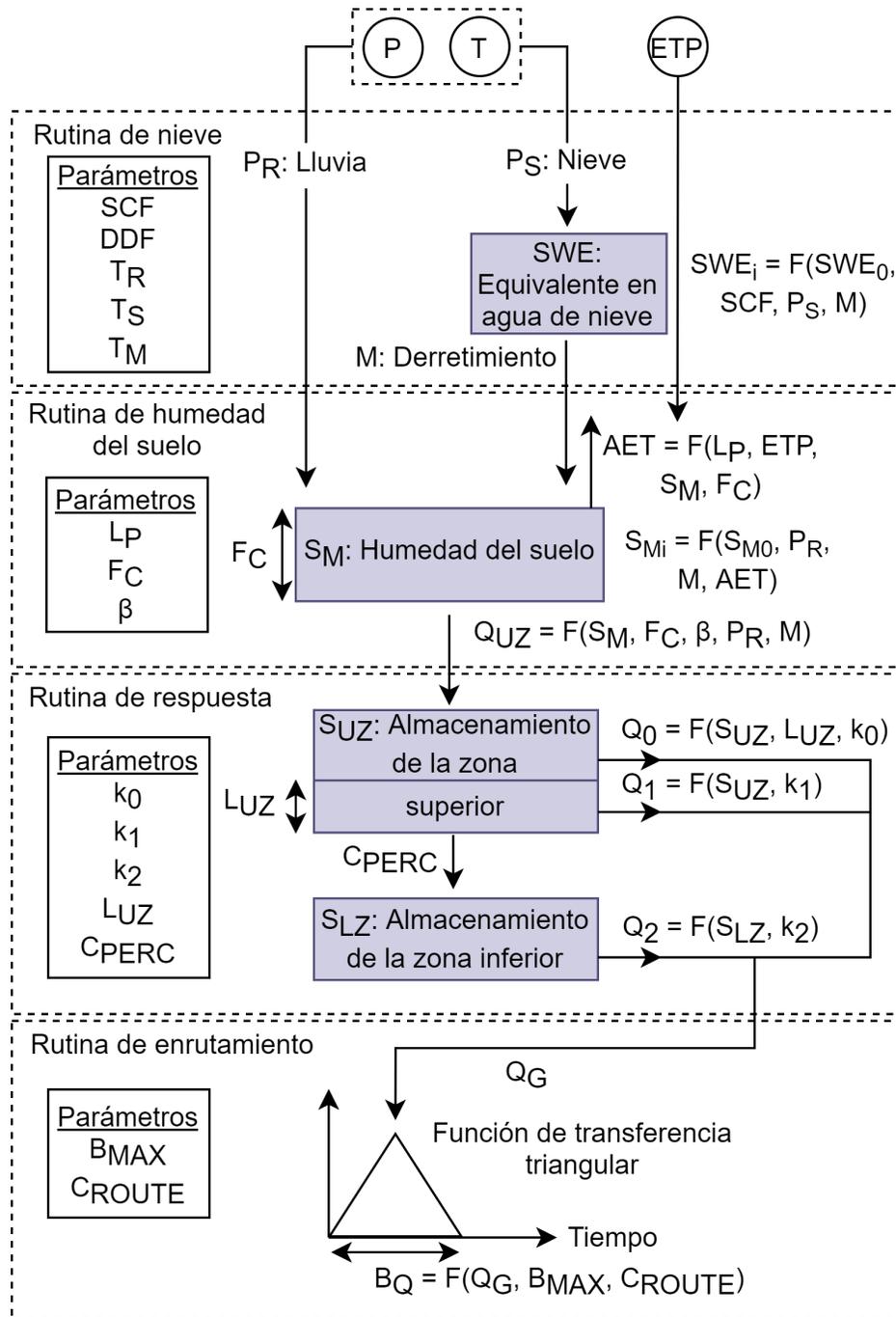


Figura 4.3: Estructura del modelo TUW (adaptado de Neri et al., 2020). Los círculos dan cuenta de las forzantes del modelo, los cuadros grises corresponden a estanques, y el resto corresponde a flujos.

La descripción de los parámetros del modelo se encuentra en la Tabla 4.2, junto a los rangos utilizados para su calibración. Si bien, se suele fijar el parámetro  $B_{MAX}$ , se optó por dejarlo libre ya que no hay referencias de un valor adecuado para cuencas chilenas. Por lo tanto, el modelo cuenta con 15 parámetros libres a ser calibrados. Para mayores detalles de la formulación de este modelo, véase el apéndice A del artículo de Parajka et al. (2007).

Tabla 4.2: Descripción de los parámetros y rangos utilizados para la calibración del modelo TUW.

Parámetro	Descripción	Unidad	Rango
$SCF$	Coefficiente de corrección nival	–	0,5 – 2
$DDF$	Coefficiente de derretimiento grado-día	mm/°C/d	0 – 5
$T_R$	Umbral sobre el cual la precipitación es lluvia	°C	1 – 5
$T_S$	Umbral bajo el cual la precipitación es nieve	°C	-3 – 1
$T_M$	Umbral sobre el cual comienza el derretimiento	°C	-2 – 4
$L_P$	Coefficiente que limita la evaporación potencial	-	0 – 1
$F_C$	Capacidad de campo	mm	0 – 1000
$\beta$	Parámetro no lineal para la producción de escorrentía	-	0 – 20
$k_0$	Coef. de almacenamiento para respuestas muy rápidas	d	0 – 2
$k_1$	Coefficiente de almacenamiento para respuestas rápidas	d	2 – 30
$k_2$	Coefficiente de almacenamiento para respuestas lentas	d	30 – 500
$L_{UZ}$	Umbral a partir del cual comienza la respuesta	mm	1 - 100
$C_{PERC}$	Tasa de percolación constante	mm/d	0 – 10
$B_{MAX}$	Tiempo base máximo para flujos lentos	d	0 – 30
$C_{ROUTE}$	Parámetro de escalamiento	d <sup>2</sup> /mm	0 – 50

### 4.3. Estrategia de calibración

La calibración de los parámetros de los modelos se llevó a cabo mediante el algoritmo de optimización global *Shuffled Complex Evolution* (SCE; Duan et al., 1992), que se basa en cuatro conceptos para lograr una optimización efectiva:

1. Combinación de enfoques probabilísticos y deterministas.
2. Agrupamiento (*clustering*).
3. Evolución sistemática en la dirección de la mejora global de un agrupamiento de puntos que abarcan el espacio de parámetros completo.
4. Evolución competitiva.

De esta forma, el algoritmo se encarga de minimizar una función objetivo (i.e., métrica de desempeño entre observaciones y simulaciones), modificando en cada iteración los parámetros de los modelos hasta cumplir con un criterio de convergencia.

El periodo de calibración seleccionado abarca los años hidrológicos de 1994/95 a 2012/13, intervalo de tiempo que asegura la representación de un amplio rango de condiciones hidrológicas. Los ocho años previos se utilizaron para el calentamiento de los modelos (i.e., garantizar condiciones iniciales estables). La validación del modelo se llevó a cabo en dos periodos diferentes: (i) 1987/88 - 1993/94, que incluye diversas condiciones hidroclimáticas, y (ii) 2013/14 - 2019/20, que se caracteriza por ser un periodo seco.

El rango óptimo de parámetros (Tablas 4.1 y 4.2) fue obtenido a partir de múltiples ensayos tipo prueba y error, calibrando los modelos con las funciones objetivo KGE, NSE y RMSE a una resolución temporal diaria. La necesidad de encontrar rangos de parámetros diferentes a los existentes en las referencias nace al considerar que estas son representativas de sus zonas de estudio, y no necesariamente de las cuencas ubicadas dentro del dominio de estudio.

Una vez obtenido el rango óptimo de parámetros, se calibraron los modelos utilizando siete funciones objetivo diferentes, cada una enfocada en ajustar alguna característica en particular de la serie de esorrentía. En la Tabla 4.3 se describen estas métricas.

Tabla 4.3: Funciones objetivo para calibrar los modelos.

Notación	Descripción	Función objetivo
KGE	Métrica basada en mínimos cuadrados, que optimiza simultáneamente la correlación de Pearson, la variabilidad (desviación estándar) y el sesgo (media) de la serie completa. Suele resultar en una mejor representación de los volúmenes altos.	KGE
NSE	Métrica basada en mínimos cuadrados. Busca minimizar el error de la serie completa, respecto a una referencia (promedio de los volúmenes). Suele resultar en una mejor representación de los volúmenes altos.	NSE
KGE+qKGE	Combinación lineal que asigna el mismo peso a los volúmenes altos y bajos al optimizar la inversa de los volúmenes.	$0,5 \cdot KGE(Q) + 0,5 \cdot KGE(1/Q)$
KGE+logNSE	Combinación lineal de dos funciones diferentes, una se enfoca en los volúmenes altos y la otra en los bajos al aplicar una transformación logarítmica.	$0,5 \cdot KGE(Q) + 0,5 \cdot NSE(\log(Q))$
SKGE	Métrica que calcula KGE para cada año hidrológico y luego los promedia. Permite reducir la variabilidad anual, generando una mejor representación de las diversas condiciones hidrológicas.	Split KGE
EVDSep	Métrica que busca minimizar el error del volumen del periodo de septiembre a marzo. Se suman los volúmenes diarios de dicho periodo, para cada año hidrológico, y se minimiza el RMSE de la serie.	RMSE
EVDOct	Idéntica que la anterior, pero para el periodo de octubre a marzo.	RMSE

Cabe destacar que todas las funciones objetivo fueron utilizadas para pronosticar los volúmenes de septiembre a marzo y de octubre a marzo, aunque EVDOct es exclusiva para el segundo periodo.

## 4.4. Generación de pronósticos a través de ESP

Para generar los pronósticos de volúmenes de deshielo se utilizó la metodología ESP en retrospectiva, es decir, mediante la emisión de pronósticos para años pasados. La siguiente descripción es válida para los cuatro modelos hidrológicos empleados:

1. Para cada año hidrológico  $n$  a pronosticar se forzó el modelo durante 8 años hasta el tiempo de inicialización  $t_i$  para estimar las condiciones iniciales (IHCs).
2. Se ejecutó el modelo con las forzantes de los 32 años restantes (no se utilizan las forzantes del año a pronosticar), hasta el 31 de marzo, generando un conjunto de 32 pronósticos para el año  $n$ . El punto de partida corresponde a las IHCs asociadas al tiempo de inicialización  $t_i$  del año  $n$ .
3. El periodo de interés corresponde a la época de deshielo (1 de septiembre a 31 de marzo, o 1 de octubre a 31 de marzo, según corresponda), por lo que se sumaron los volúmenes diarios asociados a dichos periodos y se descartaron los volúmenes restantes.
4. Se repitió este proceso para cada modelo, para cada cuenca, para cada año a pronosticar, y para cada tiempo de inicialización. De esta forma, para cada cuenca y tiempo de inicialización, se tienen 33 años pronosticados, cada uno compuesto de 32 conjuntos.

## 4.5. Evaluación de los pronósticos

La evaluación de los pronósticos se realizó de manera probabilística (utilizando los conjuntos de pronósticos) y determinística (utilizando la mediana de los conjuntos como pronóstico único). La Tabla 4.4 resume las métricas utilizadas.

Tabla 4.4: Métricas utilizadas para evaluar los pronósticos (Adaptado de Mendoza et al., 2017).

Nombre	Expresión	Descripción
Coficiente de correlación	$r = \frac{\sum_{i=1}^N (q_{m,i} - \bar{q}_m)(o_i - \bar{o})}{\sqrt{\sum_{i=1}^N (q_{m,i} - \bar{q}_m)^2} \sqrt{\sum_{i=1}^N (o_i - \bar{o})^2}}$	Métrica determinística que varía entre $[-1, 1]$ , con un valor perfecto de 1. Mide la relación lineal entre la simulación ( $q_{m,i}$ ) y la observación ( $o_i$ ).
Sesgo porcentual	$\% \text{Bias} = 100 \cdot \frac{\sum_{i=1}^N (q_{m,i} - o_i)}{\sum_{i=1}^N o_i}$	Métrica determinística que varía entre $(-\infty, \infty)$ , con un valor perfecto de 0. Mide la diferencia entre el promedio de las simulaciones y las observaciones.
Raíz del error cuadrático medio	$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (q_{m,i} - o_i)^2}$	Métrica determinística que varía entre $[0, \infty)$ , con un valor perfecto de 0. Mide la dispersión promedio de las simulaciones respecto a la observaciones.

Nombre	Expresión	Descripción
Puntaje de habilidad probabilística continua	$\text{CRPSS} = 1 - \frac{\overline{\text{CRPS}_{sim}}}{\overline{\text{CRPS}_{ref}}}$ $\text{CRPS} = \frac{1}{N} \cdot \sum_{i=1}^N \int_{-\infty}^{\infty} [F(q) - F_0(q)]^2 dq$ $F_0(q) = \begin{cases} 0, & \text{si } q < 0 \\ 1, & \text{si } q \leq 0 \end{cases}$	Métrica probabilística que varía entre $(-\infty, 1]$ , con un valor perfecto de 1. Mide la habilidad del CRPS de las simulaciones respecto al CRPS de referencia (i.e., la hidroclimatología promedio). CRPS cuantifica la diferencia entre la función de distribución acumulada empírica (eCDF) de la simulación y la eCDF de la observación.
Índice de confiabilidad $\alpha$	$\alpha = 1 - 2 \left[ \frac{1}{N} \sum_{i=1}^N  P_i(o_i) - U(o_i)  \right]$	Métrica probabiística que varía entre $[0, 1]$ , con un valor perfecto de 1. Cuantifica la proximidad entre la eCDF de los $p$ -values de cada simulación, asociada a su observación $i$ , con la CDF de una distribución uniforme.

En la Tabla 4.4,  $q_{m,i}$  es la mediana del conjunto de pronósticos para el año  $i$ ;  $\overline{q_m}$  es el promedio de las medianas de los conjuntos;  $o_i$  es la observación del año  $i$ ;  $\bar{o}$  es el promedio de las observaciones;  $P_i(o_i)$  es la probabilidad de no excedencia asociada a  $o_i$  usando los pronósticos de conjunto;  $\overline{\text{CRPS}}$  es el promedio de los CRPS de todos los años; y  $U(o_i)$  es la probabilidad de no excedencia asociada a  $o_i$  usando una distribución uniforme  $U [0, 1]$ .

Los intervalos de confianza para las métricas de verificación se crearon utilizando *bootstrap* con reemplazo (Wilks, 2011). En cada paso se volvieron a muestrear 33 pares de pronósticos de conjunto y observaciones de la distribución original. Este proceso se repitió 1000 veces y, luego, todas las métricas se calcularon para cada remuestreo y se ordenaron para obtener límites de confianza del 95 %.

Para el análisis de desempeño de los pronósticos, por un lado, se clasificaron los años en húmedos/normales/secos para las 22 cuencas en estudio, mediante la separación en terciles del volumen estacional observado, para identificar relaciones entre el CRPS y la temporada de cada año hidrológico. Por otro lado, para encontrar relaciones con las características físicas e hidrológicas de las cuencas, se realizó un *clustering* o agrupamiento, mediante el algoritmo *k-means clustering* (Hartigan y Wong, 1979). Este método tiene como objetivo la partición de las cuencas en grupos, donde cada una pertenece al grupo cuyo valor medio es más cercano. Este agrupamiento se realizó utilizando CRPSS, el índice  $\alpha$  y el  $R^2$  de los pronósticos de cada cuenca (i.e., la entrada del algoritmo es una matriz de 22 cuencas x 3 métricas), donde cada métrica se normalizó por separado para no sesgar el proceso. Los  $p$ -values reportados indican el nivel de significancia estadística entre las diferencias de las medias de los grupos, y se obtuvieron a partir de la aplicación de una prueba t (*t-test*).

# Capítulo 5

## Resultados

### 5.1. Calibración y verificación

En la Figura 5.1 se muestra el resultado de la calibración y validación de las 22 cuencas en estudio para la función objetivo KGE. Se observa que el modelo TUV presenta un mejor desempeño respecto a los tres modelos GR para todos los periodos y, en particular, para el de calibración. Con respecto a los modelos GR, todos se desempeñan de manera similar en los periodos de calibración y validación 1; no obstante, en el periodo de validación 2, el modelo GR4J es superior.

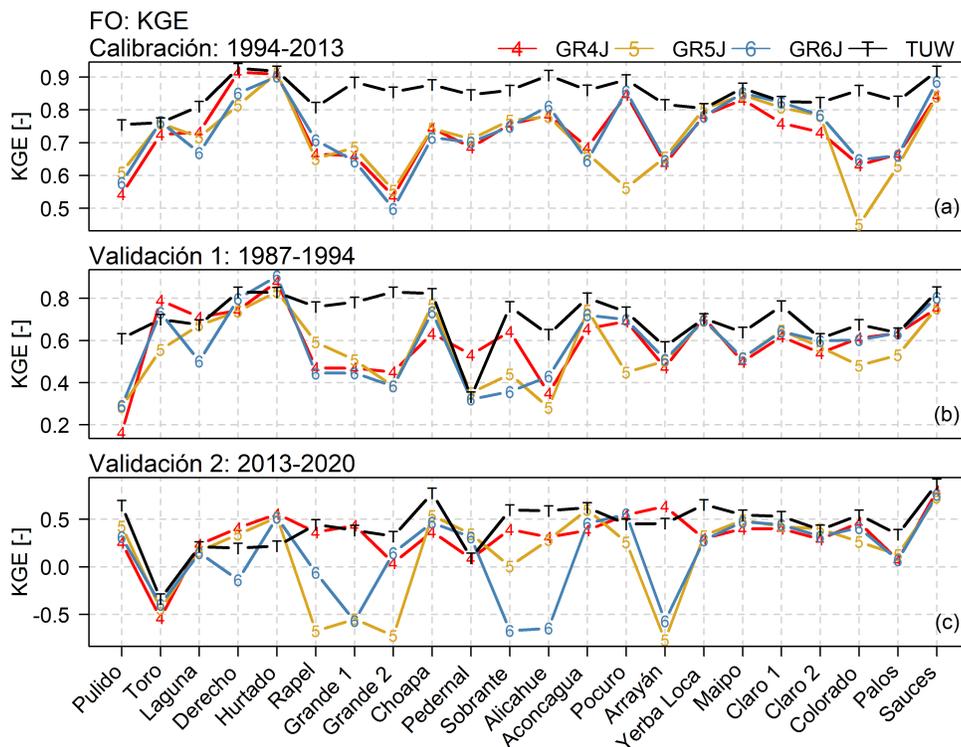


Figura 5.1: Resultados del proceso de calibración-validación de las 22 cuencas en estudio (ordenadas de norte a sur, donde Pulido es la cuenca más norteña), para los cuatro modelos utilizados y la función objetivo KGE. Notar que los paneles (a), (b) y (c) no están en la misma escala.

Por otro lado, los modelos GR5J y GR6J presentan grandes deficiencias en algunas cuencas para el periodo de validación 2 (e.g., Toro, Rapel, Grande 1, Grande 2). Al observar los componentes de la función KGE y compararlos con el modelo TUV –que se desempeña mejor en aquellas cuencas– los modelos GR5J y GR6J fallan notablemente en la representación de la variabilidad y el sesgo de los volúmenes simulados.

En la Figura 5.2 se presenta el resultado de la calibración y validación para la función objetivo SKGE, en términos de  $\Delta KGE$ . Vale decir, la diferencia entre SKGE evaluada con KGE (i.e., se calibró con SKGE, y con sus parámetros se evaluó el desempeño utilizando la función KGE de la serie completa) y KGE evaluada con KGE. Se observa que, en el periodo de calibración el desempeño empeora al calibrar con SKGE; en el periodo de validación 1, el desempeño es variable, mejorando en algunas cuencas y en otras empeorando; y en el periodo de validación 2, las cuencas que presentaban un bajo desempeño al calibrar con KGE (Figura 5.1), mejoran enormemente su desempeño. Esta mejora se debe a la naturaleza de SKGE, que permite generar una mejor transferencia temporal de los parámetros a los periodos de validación al representar de mejor manera las condiciones hidrológicas de cada año.

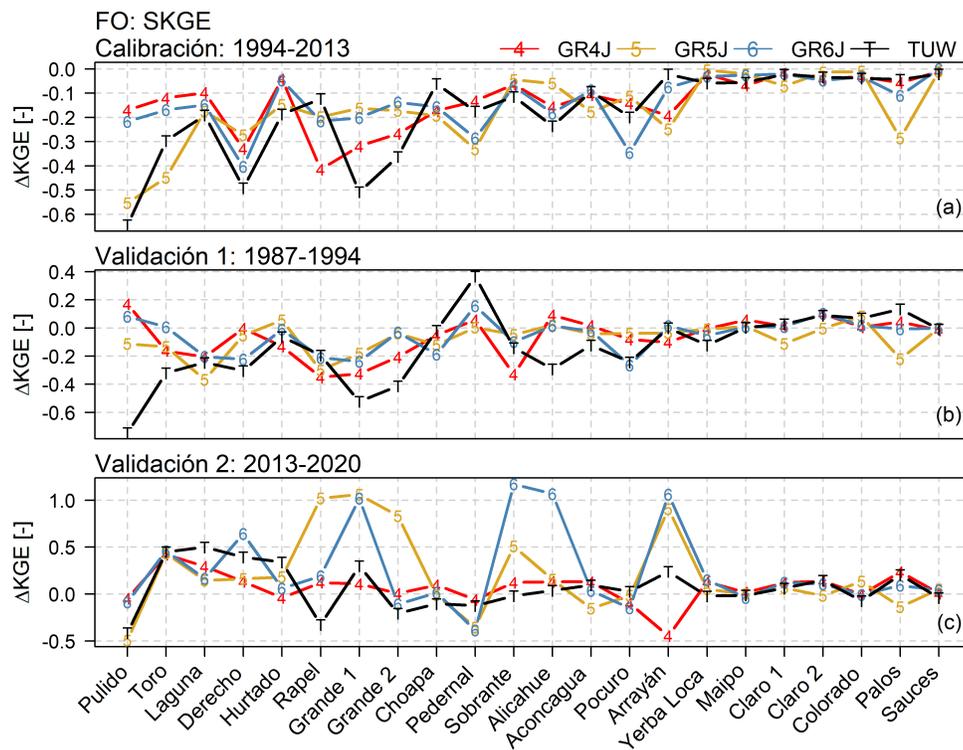


Figura 5.2: Resultados proceso de calibración-validación de las 22 cuencas en estudio, para los cuatro modelos utilizados y para la función objetivo SKGE. Notar que los paneles (a), (b) y (c) no están en la misma escala y que el desempeño de los modelos se presenta como la diferencia entre el KGE obtenido al evaluar el desempeño de los parámetros calibrados con SKGE y el KGE obtenido al evaluar el desempeño de los parámetros calibrados con KGE ( $\Delta KGE$ ).

En la Figura 5.3 se presenta la función de distribución acumulada empírica (eCDF, por sus siglas en inglés) de los resultados de la calibración y validación para todas las funciones

objetivo. El panel (a) es el equivalente de la Figura 5.1, donde también es posible notar que el modelo TUW es el que mejor desempeño presenta. Es directo observar que, para todas las funciones objetivo y todos los periodos el modelo TUW es superior a los GR, salvo en algunos tramos bajos de la eCDF en los periodos de validación.

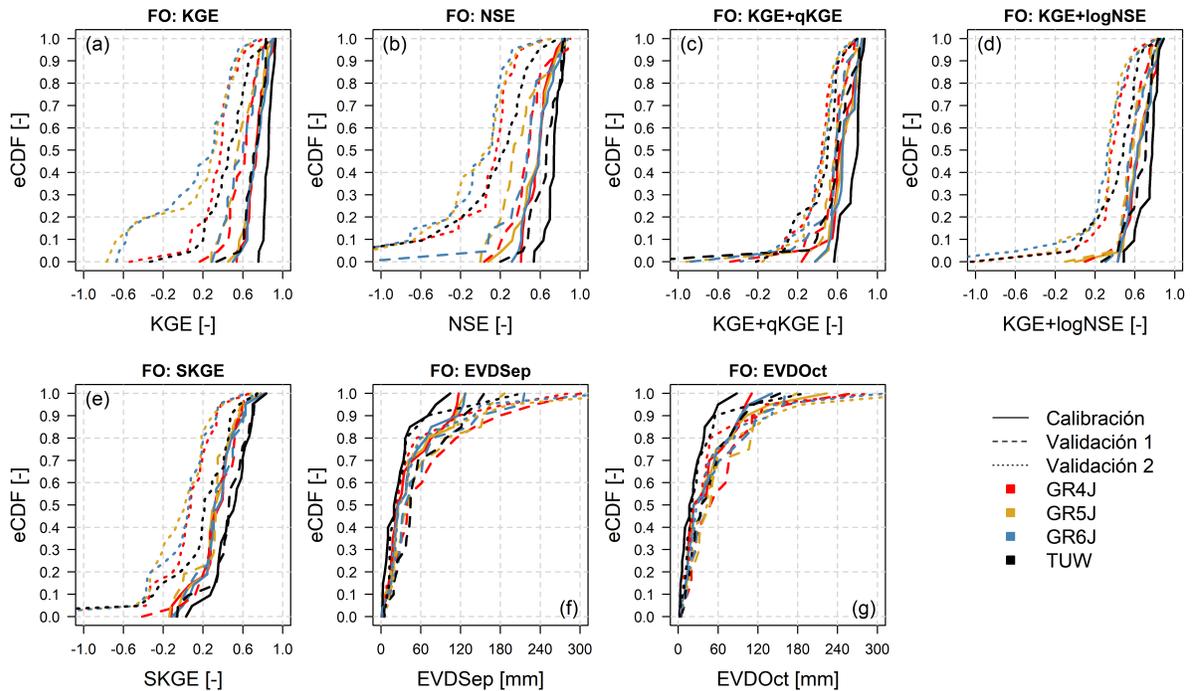


Figura 5.3: eCDF de los resultados de la calibración y validación de las cuencas en estudio para todas las FO. Notar que los paneles (a) a (e) están positivamente orientados. Los paneles (f) y (g) están negativamente orientados, lo que implica que mientras más a la izquierda se encuentren las curvas, los modelos se desempeñan mejor.

En la Figura 5.4 se presenta la eCDF de los resultados de la calibración y validación de todas las funciones objetivo en términos de KGE. Es directo notar la gran diferencia de desempeño en validación 2 entre KGE y SKGE. Además, es interesante observar que KGE+logNSE presenta un comportamiento similar a SKGE, ya que ambas FO se enfocan, también, en los caudales bajos.

Por otro lado, es conveniente recordar que las funciones objetivo EVDSep y EVDOct minimizan el volumen estacional del periodo de deshielo. De esta forma, es esperable que el desempeño sea bajo al comparar el rendimiento con FO que evalúan el periodo completo a escala diaria (e.g., KGE). En estas FO, TUW es superior sólo en el tramo alto de las eCDF, mientras que, bajo 0,3 en EVDSep y 0,5 en EVDOct, otros modelos se desempeñan mejor.

Finalmente, en la Figura 5.5 se observan los resultados del proceso de calibración en términos del volumen de septiembre a marzo normalizado (nEVDSep). Es directo notar que la función EVDSep es la que presenta mejores resultados para el periodo de calibración. Sin embargo, es interesante notar que KGE+qKGE, KGE+logNSE y SKGE presentan resultados similares e incluso mejores que EVDSep para validación 1 y validación 2. Esto se debe a que

dichas métricas se enfocan, también, en los volúmenes más bajos, por lo tanto, en época de sequía (validación 2), es esperable que generen un mejor desempeño que una métrica enfocada principalmente en volúmenes altos (EVDSep/RMSE).

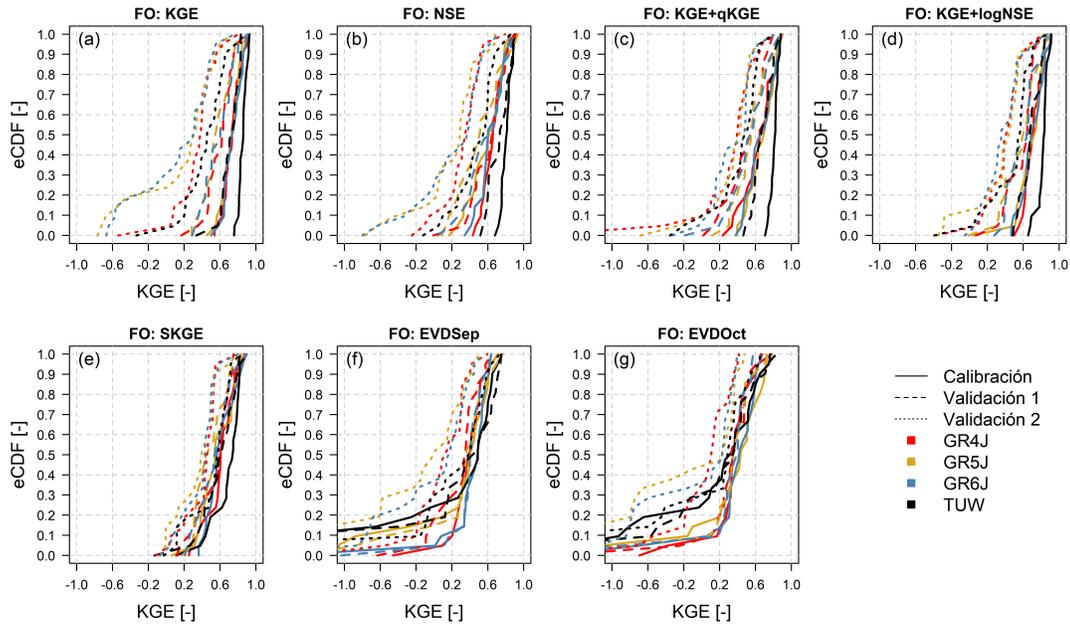


Figura 5.4: eCDF de los resultados de la calibración y validación de las cuencas en estudio para todas las FO, en términos de KGE.

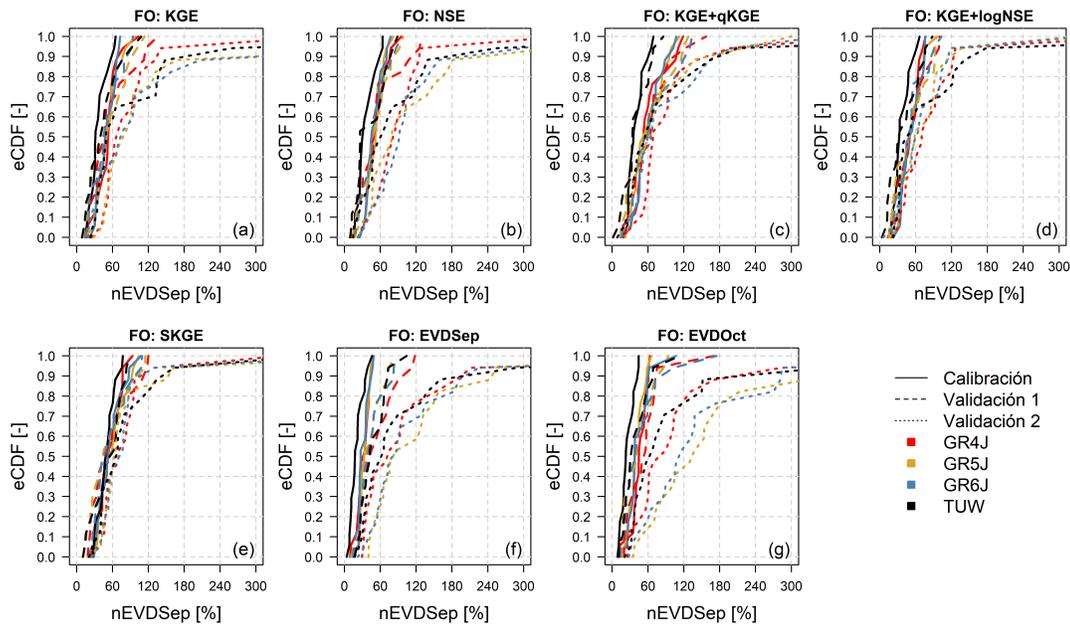


Figura 5.5: eCDF de los resultados de la calibración y validación de las cuencas en estudio para todas las FO, en términos de EVDSep normalizado (nEVDSep) por la desviación estándar, para evitar que cuencas con grandes volúmenes sesguen los resultados.

### 5.1.1. Análisis de las cuencas DGA

En este estudio se incluyen cinco cuencas que utiliza la DGA para emitir sus pronósticos (Figura 3.1). Debido a su distribución homogénea, se considera que son representativas de la zona de estudio, por lo que resulta conveniente realizarles un análisis de desempeño según algunas métricas clave (KGE, SKGE y EVDSep) y algunos índices hidrológicos. Sólo se presentan los resultados de los dos mejores modelos en cuanto a calibración y validación (GR4J y TUW).

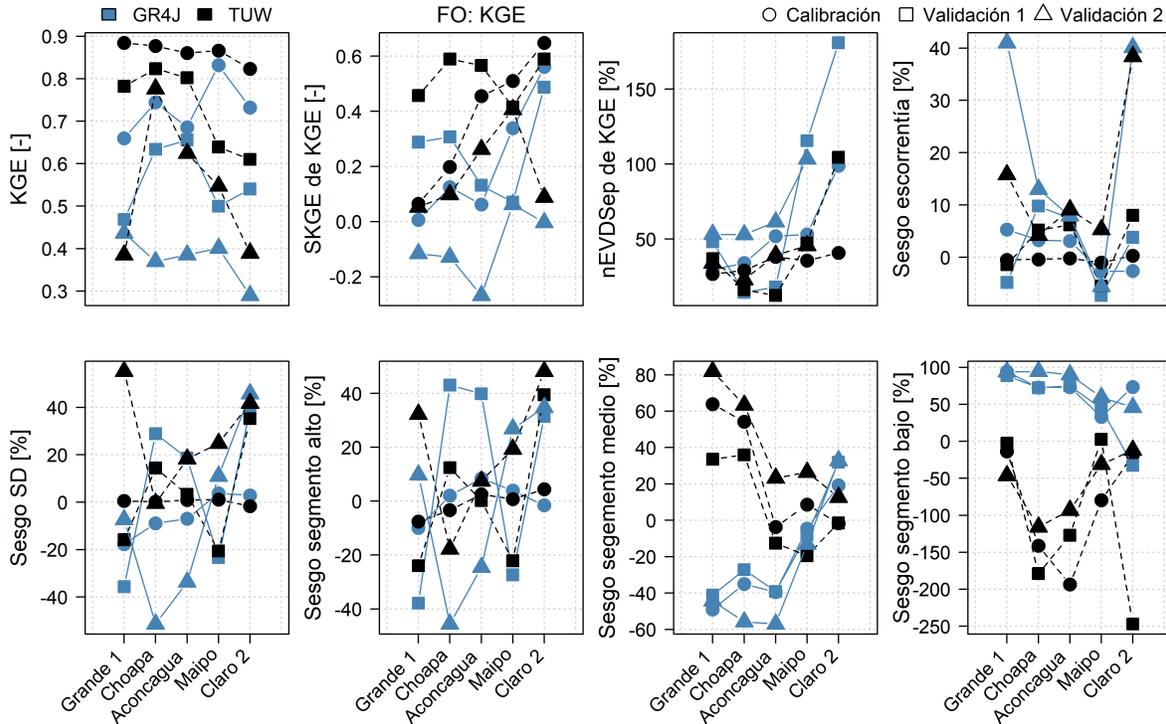


Figura 5.6: Métricas e índices hidrológicos para los modelos GR4J y TUW calibrados con KGE. **SKGE de KGE** se refiere a Split KGE calculado con los parámetros calibrados con KGE; **sesgo de escorrentía** corresponde al sesgo total entre las simulaciones y observaciones (en ese orden); **sesgo SD**, es el sesgo entre la desviación estándar de la serie completa de la escorrentía media diaria simulada y observada; y **sesgo segmento alto, medio y bajo** hacen referencia al sesgo entre aquellos segmentos (probabilidad de excedencia menor a 2 %, entre 20 y 70 % y sobre 70 %, respectivamente) de la curva de duración de los volúmenes diarios simulados y observados. Cabe destacar que, para KGE y SKGE el valor óptimo es 1, mientras que para nEVDSep y los sesgos el óptimo es 0.

Las métricas e índices de las cuencas DGA, para los modelos GR4J y TUW calibrados con KGE, se presentan en la Figura 5.6. De manera general, el modelo TUW es superior en KGE, SKGE, nEVDSep, en el sesgo de escorrentía, en el sesgo de desviación estándar y en el sesgo de caudales altos. La diferencia más importante se observa al comparar los nEVDSep, mostrando que TUW logra representar mejor los volúmenes estacionales.

Con respecto al sesgo en el segmento medio, TUV tiende a sobreestimar la escorrentía, mientras que GR4J la tiende a subestimar, no obstante, GR4J logra un mejor desempeño general. Por último, en el sesgo del segmento bajo, en general, TUV subestima en gran medida la escorrentía, mientras que GR4J la sobreestima, pero a magnitudes menores. Nuevamente, para este índice, el modelo GR4J se desempeña mejor que el modelo TUV.

En la Figura 5.7, se muestran las métricas e índices de las cuencas DGA para los modelos GR4J y TUV calibrados con SKGE. Nuevamente, para KGE, SKGE, nEVDSEp, sesgo de escorrentía, desviación estándar y caudales altos, el modelo TUV es superior con algunas salvedades. Sin embargo, es interesante notar como ambos modelos mejoran la representación de los caudales medios y bajos, donde los modelos se encuentran más igualados en cuanto a calidad de desempeño.

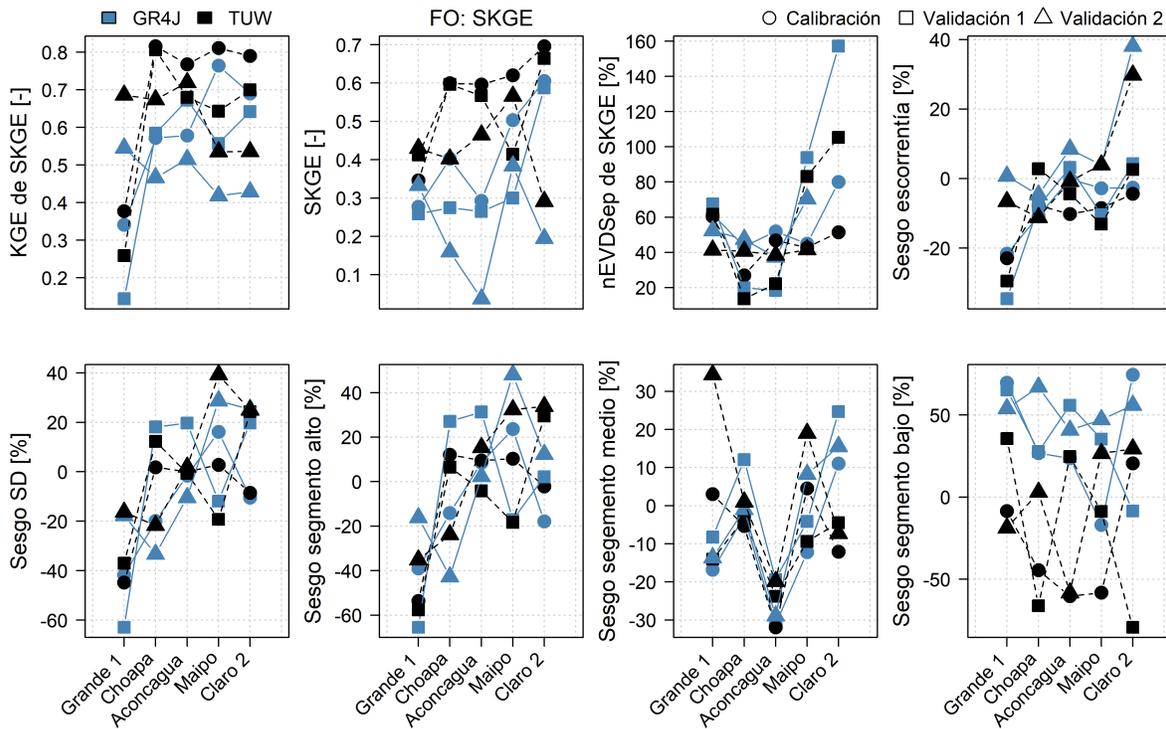


Figura 5.7: Métricas e índices hidrológicos para los modelos GR4J y TUV calibrados con SKGE.

Por último, la Figura 5.8 muestra las métricas e índices de las cuencas DGA para los modelos GR4J y TUV calibrados con la métrica estacional EVDSEp. Para los primeros cinco indicadores, y siguiendo la tendencia de las anteriores dos métricas, el modelo TUV presenta un mejor desempeño general. En el sesgo de caudales altos, ambos modelos se desempeñan de manera similar, mientras que en el sesgo del segmento medio y bajo el modelo GR4J es ampliamente superior.

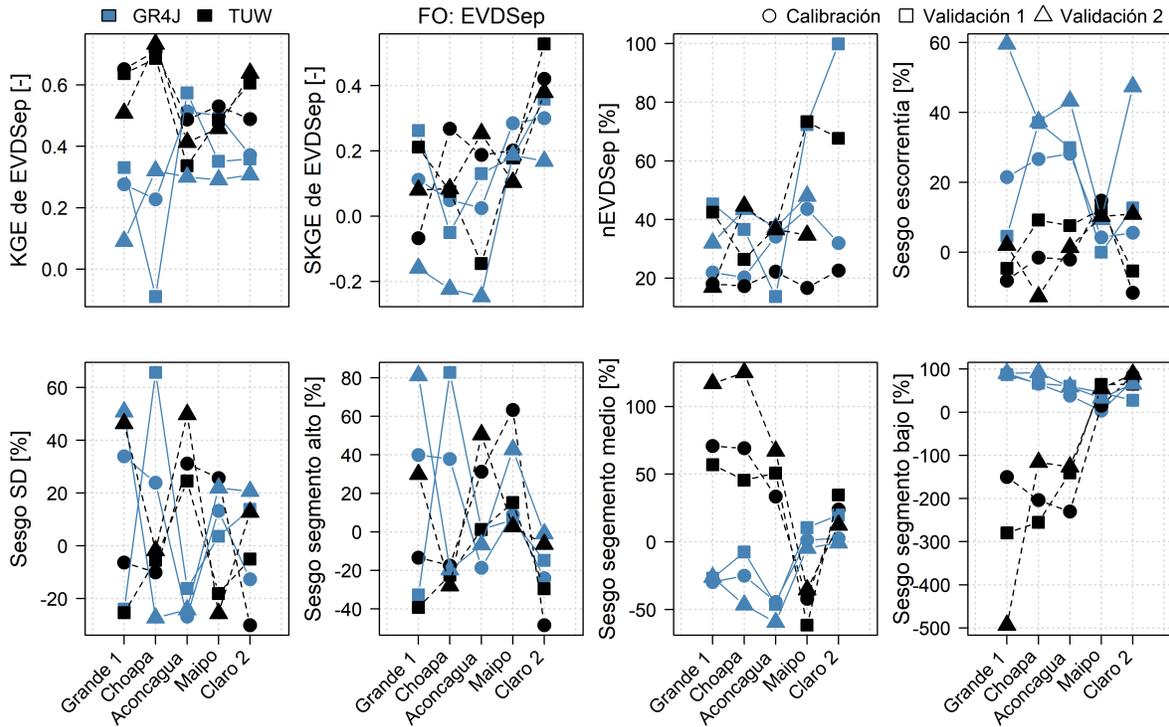


Figura 5.8: Métricas e índices hidrológicos para los modelos GR4J y TUW calibrados con EVDSep.

## 5.2. Generación y evaluación de los conjuntos de los pronósticos

A partir de la calibración de los parámetros de los modelos, se generan pronósticos estacionales para las 22 cuencas en estudio, con cuatro modelos diferentes, siete funciones objetivo, seis tiempos de inicialización, y dos periodos (septiembre a marzo y octubre a marzo), lo que arroja un total de 6.336 pronósticos (288 por cuenca).

La Figura 5.9 ilustra el proceso de generación de pronósticos ESP, mostrando resultados para el volumen de septiembre a marzo de la cuenca Maipo en el Manzano. El volumen se presenta en millones de metros cúbicos ( $Mm^3$ ).

Pronóstico volumen de Sep-Mar de Maipo En El Manzano.  
FO: EVDSep. Modelo: GR4J. Inicialización: 1° de Mayo.

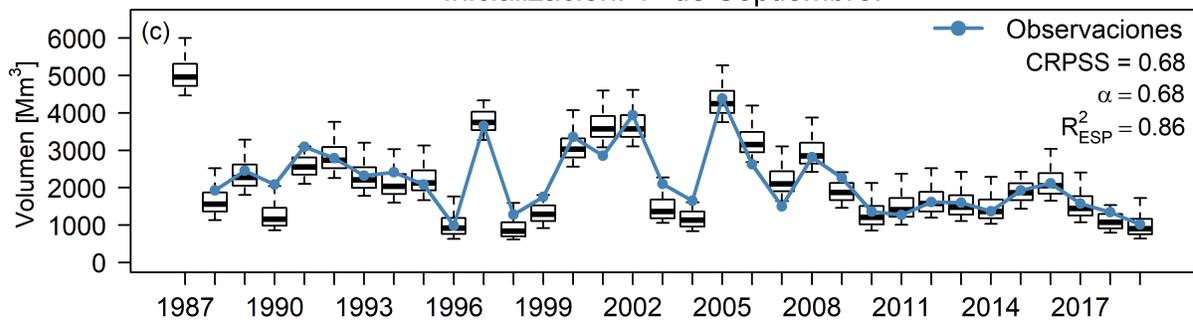
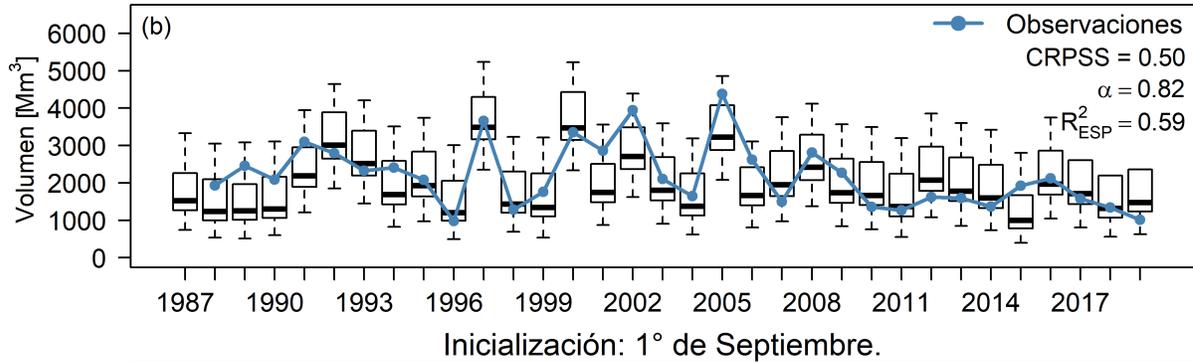
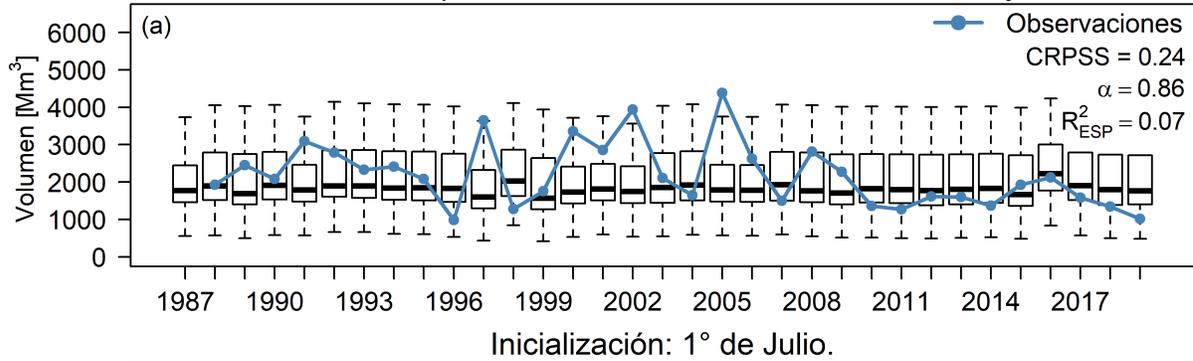


Figura 5.9: Pronósticos de volumen de septiembre a marzo para la cuenca Maipo en el Manzano, utilizando el modelo GR4J, la función objetivo EVDSep, y tiempos de inicialización 1° mayo, 1° de julio y 1° de septiembre. Cada *boxplot* se compone de 32 miembros que constituyen el conjunto de pronósticos. La barra central es la mediana del conjunto, los bordes de la caja corresponden al primer y tercer cuartil, y los extremos a los valores mínimos y máximos que se encuentran dentro del rango borde + intercuartil (diferencia entre el tercer y primer cuartil).

Con respecto a las métricas, el CRPSS y el índice  $\alpha$  reportados se calculan según lo expuesto en la sección 4.5, lo que arroja un único valor por pronóstico. El coeficiente de determinación  $R^2$  corresponde a la correlación de Pearson (sección 4.5) al cuadrado. Los índices CRPSS y  $R^2$  mejoran, y el índice de confiabilidad  $\alpha$  disminuye su valor, a medida que se inicializa el pronóstico más cerca del periodo de deshielo.

### 5.2.1. Pronósticos de volúmenes de deshielo de septiembre a marzo

La Figura 5.10 presenta el CRPSS de los pronósticos agrupados por modelo y tiempo de inicialización. Se observa que los CRPSS mejoran a medida que se inicializa más cerca del 1° de septiembre (notar la escala de los paneles). Esto se debe a un mayor conocimiento de las condiciones hidrológicas iniciales, lo que provoca que la incertidumbre total del pronóstico se deba mayormente a las forzantes y en menor medida a las IHCs (e.g., Harrigan et al., 2018; Donegan et al., 2020).

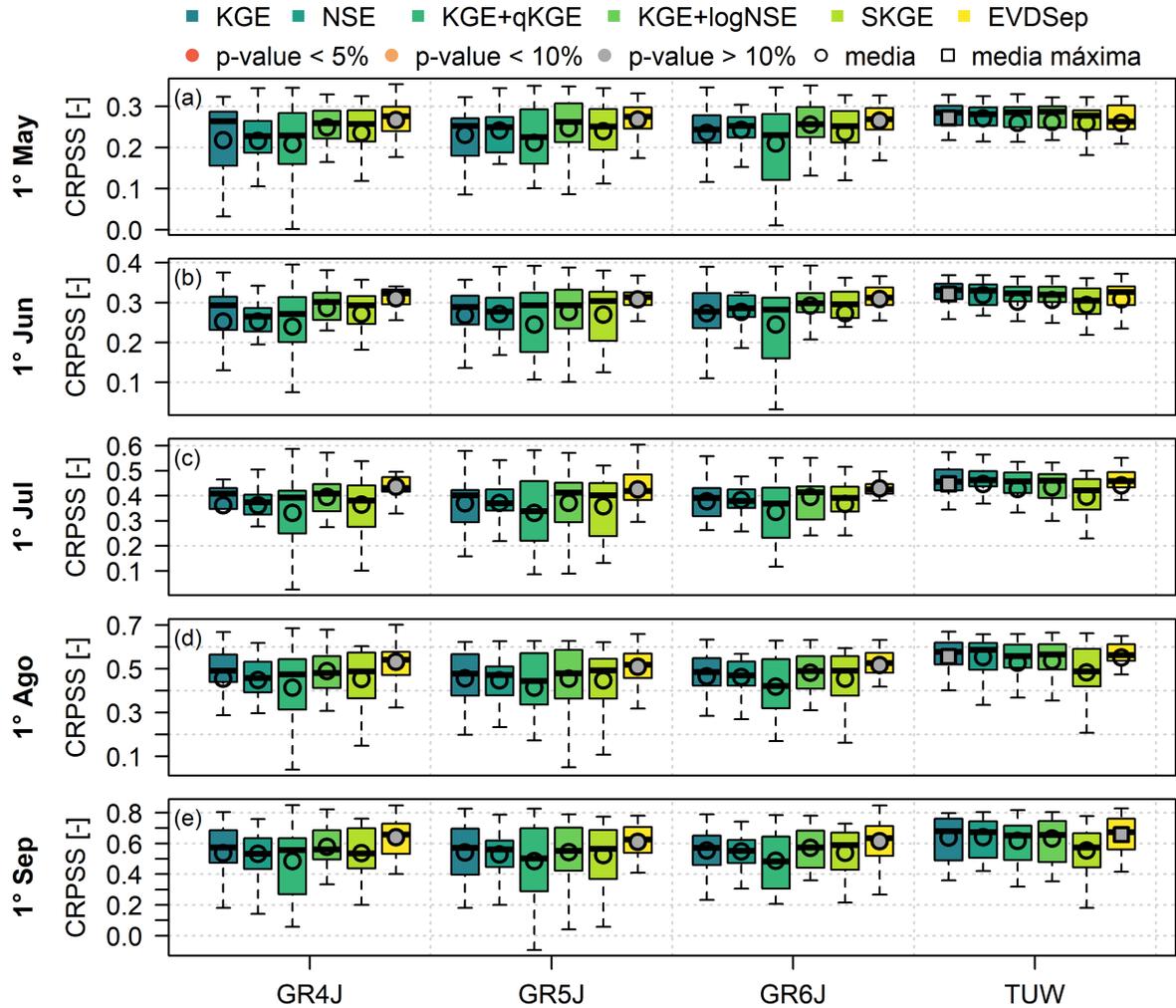


Figura 5.10: Comparación de los valores de CRPSS de los pronósticos ESP de septiembre a marzo, obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y modelo hidrológico. Cada *boxplot* se compone del CRPSS de las 22 cuencas en estudio. El *p-value* indica la probabilidad de que la media del mejor conjunto sea significativamente similar a la del segundo mejor. Vale decir, un bajo *p-value* indica que las medias son significativamente diferentes. El círculo coloreado indica cual función objetivo de cada modelo entrega el mejor CRPSS, mientras que el cuadrado indica cual es el modelo con la media más alta de entre todos. Notar que los paneles no se encuentran en la misma escala.

Por otro lado, la función objetivo que mejores valores de CRPSS entrega es EVDSep para todos los modelos GR. Sin embargo, para el modelo TUW esta función sólo es mejor el 1° de septiembre, mientras que, para el resto de tiempos de inicialización, KGE entrega mayores valores. La FO KGE+qKGE es la que entrega los peores valores de CRPSS para los modelos GR, presentando gran dispersión e, incluso, valores negativos para el modelo GR5J el 1° de septiembre. Para el modelo TUW, la FO SKGE se observa como la más deficiente. Cabe señalar que este último modelo es el que entrega mayores valores de CRPSS para todos los tiempos de inicialización, aunque estas superioridades no son estadísticamente significativas.

La Figura 5.11 presenta el mismo análisis, pero para el índice de confiabilidad  $\alpha$ . En este caso, la métrica disminuye a medida que el tiempo de inicialización se acerca al inicio del periodo de deshielo. Esto se debe a la disminución progresiva de la dispersión del conjunto de pronósticos (i.e., los *boxplot* se hacen más pequeños), lo que provoca que la mayoría de las observaciones se encuentren fuera del conjunto de pronósticos y los *p-values* se alejen de la forma de una distribución uniforme U [0, 1] (e.g., Renard et al., 2010; Mendoza et al., 2017).

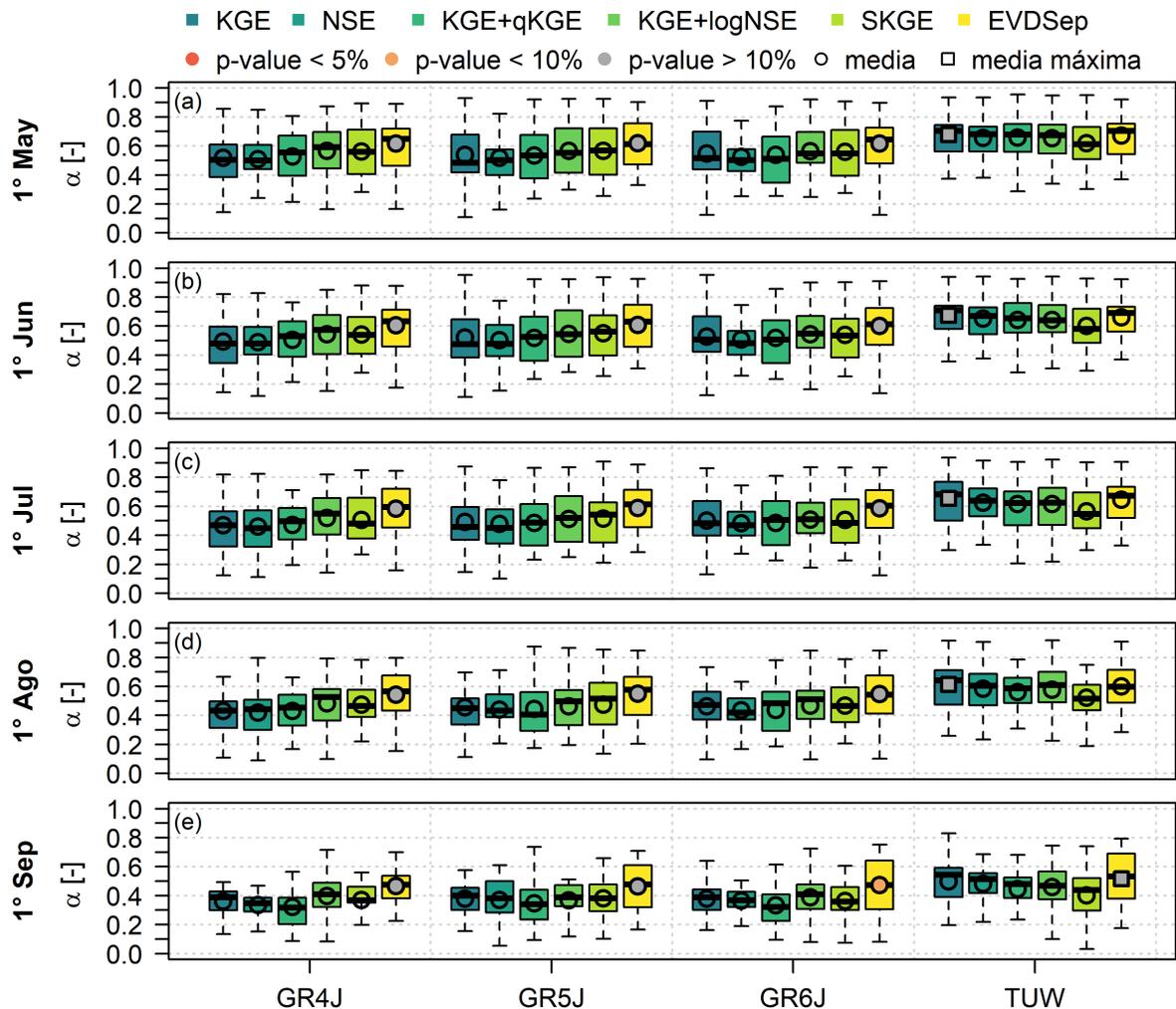


Figura 5.11: Comparación de los valores del índice  $\alpha$  de los pronósticos ESP de septiembre a marzo, obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y modelo hidrológico. Notar que los paneles cuentan con la misma escala.

La función objetivo que mejor confiabilidad  $\alpha$  entrega es EVDSep para todos los modelos GR. Sin embargo, para el modelo TUW esta función sólo es mejor el 1° de septiembre; para el resto de tiempos de inicialización, KGE entrega mayores  $\alpha$ . Nuevamente, el modelo TUW entrega las mejores confiabilidades, pero se debe tener en cuenta que estas superioridades no son estadísticamente significativas respecto a los otros modelos.

Con respecto al  $R^2$  (Figura 5.12), aumenta cuando se aproxima el inicio del periodo de deshielo (notar la escala de los paneles), y los valores son bastante homogéneos entre modelos y funciones objetivo, especialmente a partir del 1° de julio. En la familia de modelos GR, la función EVDSep es la que entrega mejores  $R^2$ , seguida por KGE+logNSE. Al utilizar el modelo TUW, las funciones NSE y KGE entregan mejores valores. Cabe destacar que los modelos GR4J y TUW entregan los valores más altos de  $R^2$  en cada tiempo de inicialización (símbolo cuadrado), sin embargo, estas superioridades no son estadísticamente significativas.

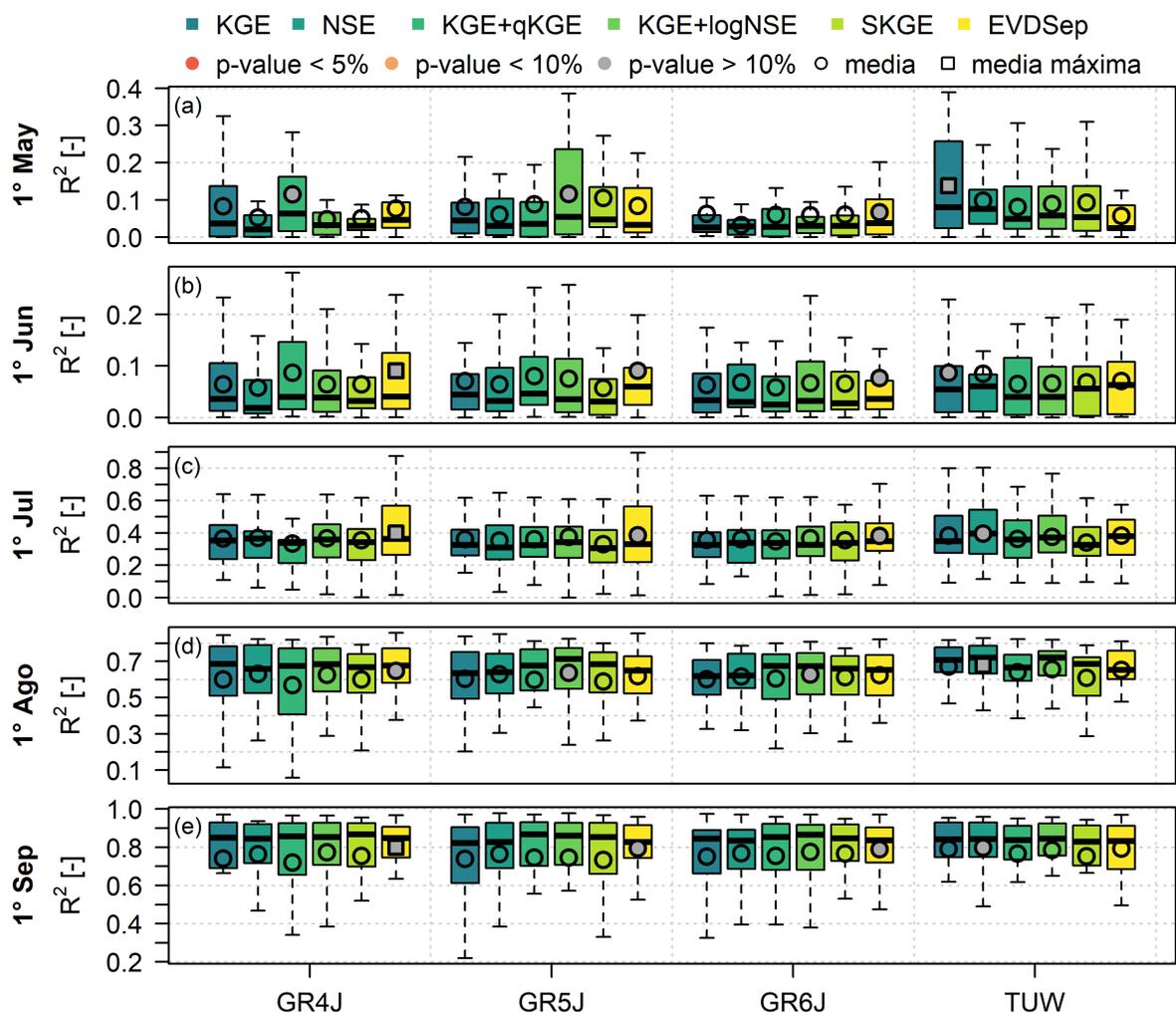


Figura 5.12: Comparación de los valores de  $R^2$  de los pronósticos ESP de septiembre a marzo, obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y modelo. Notar que los paneles no se encuentran en la misma escala.

En la Tabla 5.1 se muestra el promedio de cada una de las métricas de todos los pronósticos

inicializados en septiembre, para todas las cuencas y funciones objetivo. Es claro observar que el modelo TUW es el que entrega los valores de las métricas más altos, siendo la diferencia con respecto al segundo mejor modelo estadísticamente significativa para CRPSS y el índice de confiabilidad  $\alpha$ . Además, es posible notar que la familia de modelos GR entrega valores similares, no destacándose ninguno por sobre otro.

Tabla 5.1: Comparación del promedio de las métricas de los pronósticos emitidos en septiembre para cada modelo. Se destaca la diferencia entre los dos mejores modelos (generalmente los modelos TUW y GR6J). El *p-value* hace referencia al nivel de significancia de la diferencia entre las medias de los conjuntos que componen cada métrica. Cabe destacar que el promedio no es necesariamente el mejor indicador para evaluar el rendimiento, y sólo se utiliza en este caso para mostrar de manera sencilla el desempeño general de los modelos y sus diferencias.

Métrica	GR4J	GR5J	GR6J	TUW	Dif. dos mejores [%]	<i>p-value</i>
CRPSS	0,55	0,54	0,55	0,62	12,87	0.02
$\alpha$	0,37	0,38	0,38	0,47	21,41	0.01
$R^2$	0,76	0,75	0,77	0,78	1,72	0.11

Al analizar el desempeño general en base a las tres métricas, la función objetivo EVDSep es la que tiende a generar un mejor desempeño en todos los modelos, especialmente en los modelos GR. El modelo TUW es más versátil, y se desempeña consistentemente bien para todas las FO, además, presenta los mejores desempeños para todas las métricas y tiempos de inicialización, salvo algunas excepciones en  $R^2$ . Sin embargo, resulta conveniente realizar una comparación, de carácter particular, entre los modelos con aquellas funciones objetivo que entreguen los mejores CRPSS en cada cuenca.

En el Anexo A, sección A.1, se presenta el mismo análisis, pero agrupado por función objetivo. El resultado general muestra que el modelo TUW es superior al resto de los modelos al comparar la misma métrica (e.g., KGE del modelo TUW con KGE de los modelos GR), siendo esta superioridad estadísticamente significativa para algunas FO, como KGE y NSE.

### 5.2.1.1. Mejores CRPSS

En la Figura 5.13 se muestran los mejores CRPSS de los pronósticos ESP y su distribución espacial, por modelo y tiempo de inicialización.

Se observa que el CRPSS aumenta hacia el 1° de septiembre, y que en las cuencas de la zona centro es más alto que en el resto de las cuencas. Los resultados indican que todos los modelos se desempeñan de manera similar, aunque TUW, en septiembre, es el que entrega más cuencas (4) con  $CRPSS \geq 0,8$ ; seguido por el modelo GR4J con dos cuencas. En agosto, TUW sigue siendo superior, con la mayor cantidad de cuencas con  $CRPSS \geq 0,6$ .

Con respecto a las funciones objetivo, EVDSep es la que genera los valores de CRPSS más altos, especialmente en los modelos GR. En el modelo TUW, las funciones se encuentran más distribuidas, aunque predominan EVDSep, KGE y NSE.

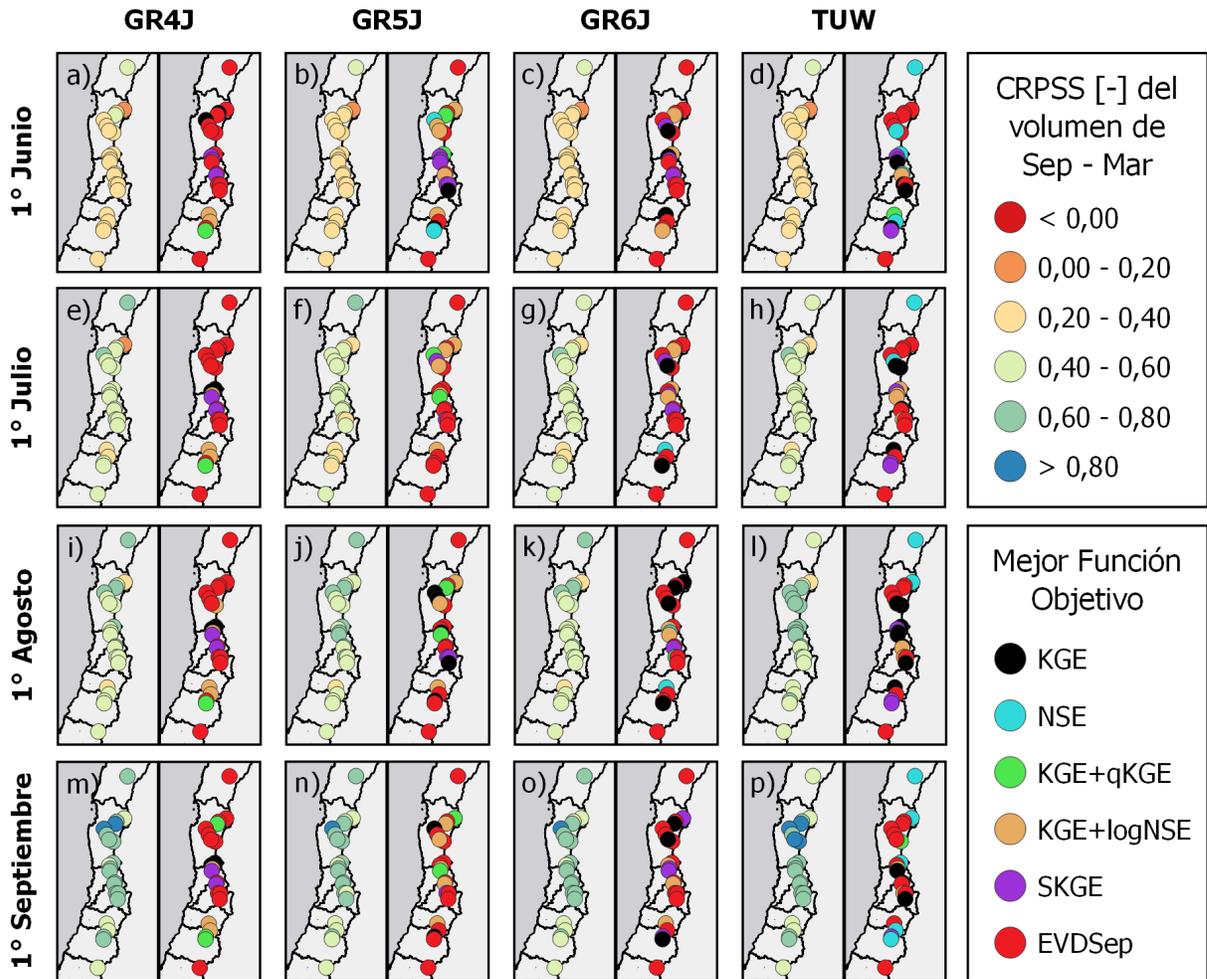


Figura 5.13: Distribución espacial de los mejores CRPSS del periodo de septiembre a marzo (paneles de la izquierda), junto a las funciones objetivo que los generan (paneles de la derecha), para todas las cuencas, los modelos, y cuatro tiempos de inicialización.

En el Anexo A, sección A.2, se muestran los mejores índices  $\alpha$  y  $R^2$ . El resultado general indica que el modelo TUV es ligeramente superior a los modelos GR, y que la función objetivo estacional EVD Sep es la que genera los valores de las métricas más altos. En promedio, TUV entrega mejoras en CRPSS, índice  $\alpha$  y  $R^2$  de un 6%, 9% y 4%, respectivamente, con respecto al mejor de los modelos GR.

### 5.2.1.2. Cuencas DGA

Las Figuras 5.14, 5.15 y 5.16, muestran la evolución del CRPSS, índice  $\alpha$  y  $R^2$ , para las cuencas DGA, según modelo, función objetivo y tiempos de inicialización. Para el resto de las cuencas véase el Anexo A, sección A.3.

Con respecto al CRPSS (Figura 5.14), se observa como aumenta su valor hacia el 1° de septiembre, y como se va degradando hacia las cuencas del sur (Grande 1, más al norte, Claro 2, más al sur). En general, EVD Sep se destaca como la función objetivo que entrega

los mejores valores de CRPSS. Es interesante notar que, teniendo como referencia el modelo y la función objetivo que generan el mejor CRPSS (siendo TUW el mejor en la mayoría de los casos), la correcta elección de la FO en los demás modelos consigue generar un CRPSS al nivel del mejor. Esta situación se aprecia con claridad en la cuenca Aconcagua donde, para los modelos GR, existe una gran dispersión de la métrica según la FO, pero EVDSep logra generar un CRPSS similar al del modelo TUW.

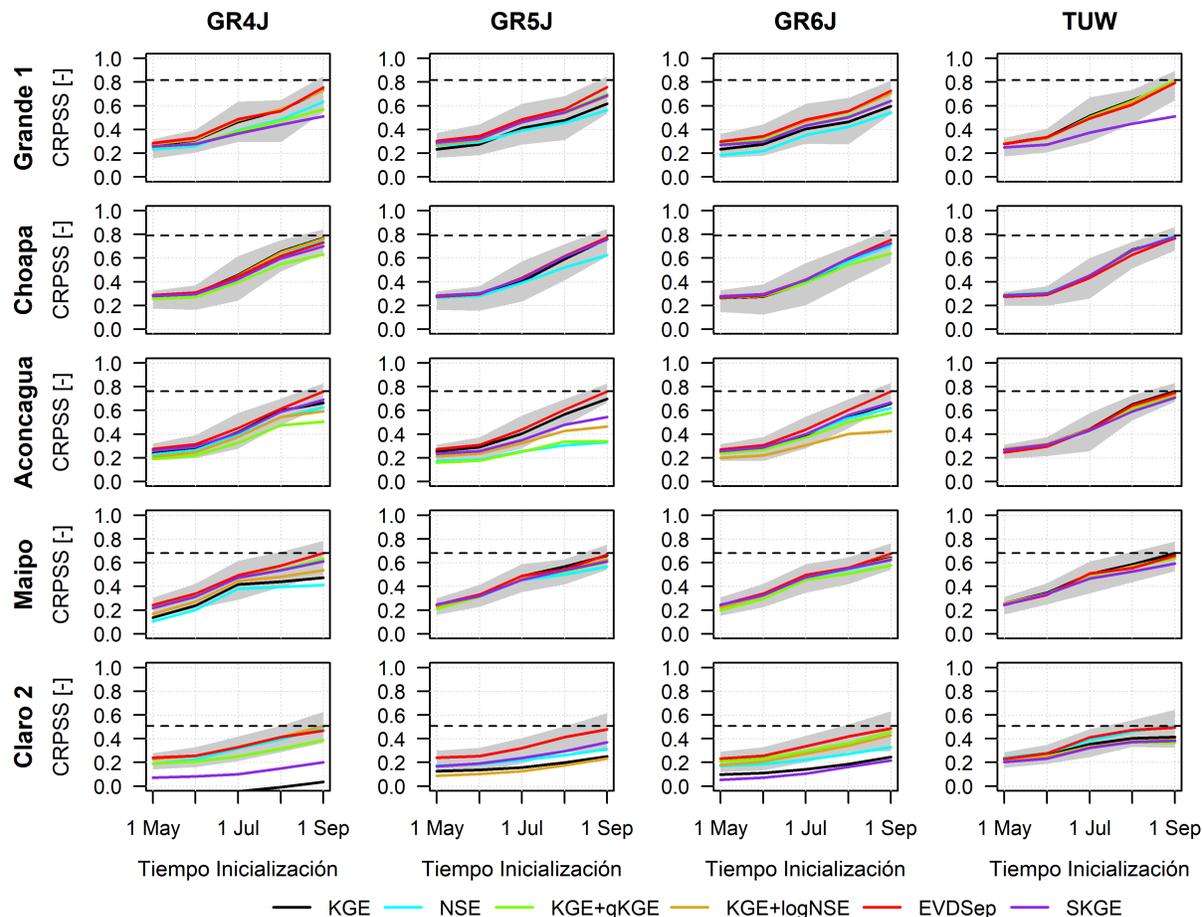


Figura 5.14: Comparación de los valores de CRPSS de las cuencas DGA por función objetivo, tiempo de inicialización y modelo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping*. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de septiembre.

Con respecto al índice de confiabilidad  $\alpha$  (Figura 5.15), su valor disminuye hacia el 1° de septiembre, y es consistentemente más alto en las cuencas centrales (Choapa, Aconcagua y Maipo). Nuevamente se destaca EVDSep como la función objetivo óptima, la cual, en general, presenta mayor estabilidad a través de los diferentes tiempos de inicialización. Si se compara la mejor combinación de modelo y su función objetivo (siendo TUW el mejor en la mayoría de los casos), con la mejor FO de los demás modelos, la diferencia es más significativa en comparación con CRPSS.

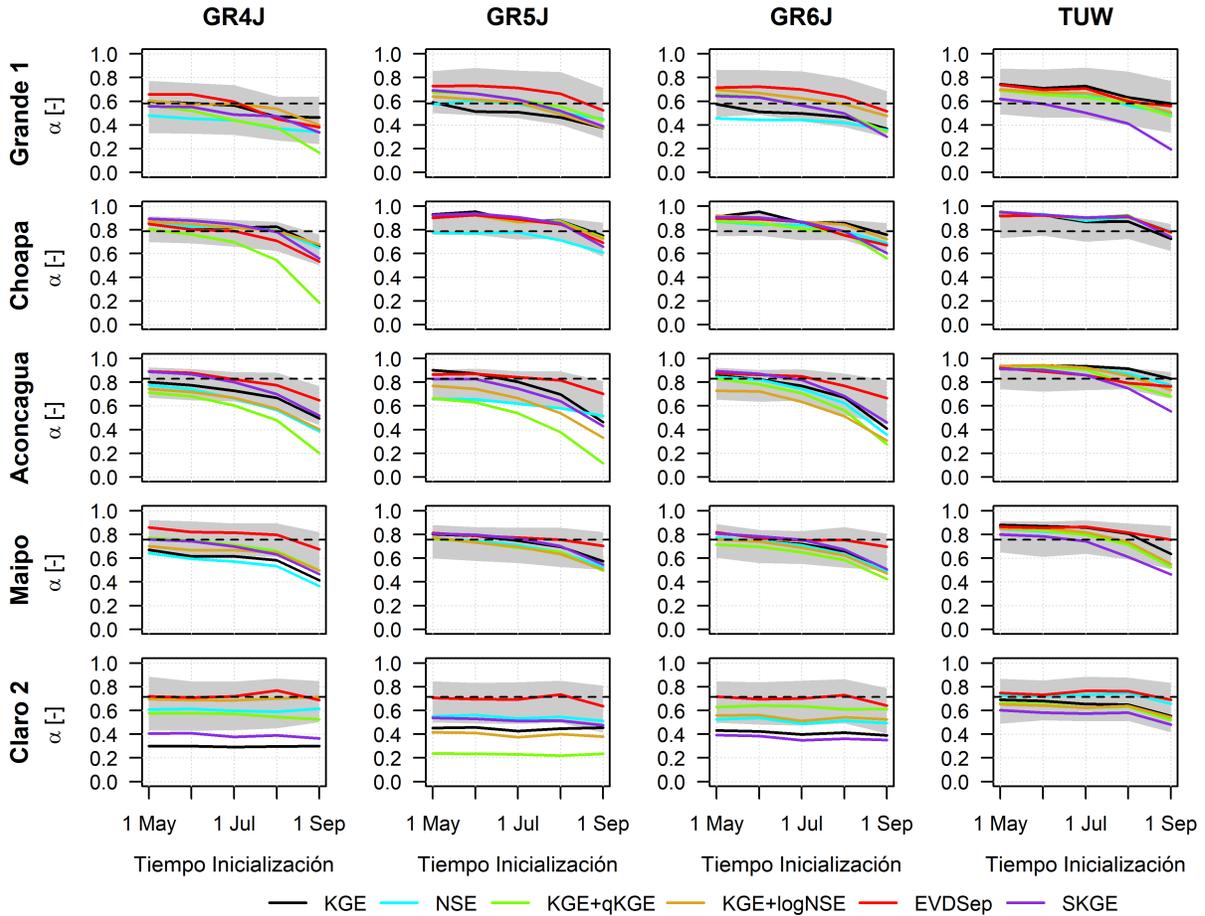


Figura 5.15: Comparación de los valores del índice  $\alpha$  de las cuencas DGA por función objetivo, tiempo de inicialización y modelo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping*. La línea discontinua demarca el máximo índice  $\alpha$  obtenido para algún modelo el 1° de septiembre (tiempo de inicialización que no genera el mayor valor de  $\alpha$ , pero que coincide con la inicialización de los pronósticos oficiales de la DGA).

Los resultados de la Figura 5.16 muestran que el  $R^2$  presenta un comportamiento similar al de CRPSS, aumentando hacia el 1° de septiembre y degradándose hacia las cuencas del sur. No obstante, en algunos casos disminuye su valor entre el 1° de mayo y 1° de junio, para luego ascender a medida que se aproxima el 1° de septiembre. Este comportamiento se debe a que en mayo existe gran incertidumbre en las IHCs, donde no se producen acumulaciones de agua con poder predictivo hasta julio o agosto. De esta forma, la mediana de los conjuntos no presenta variabilidad (ver Figura 5.9, panel (a)). A medida que se avanza en los tiempos de inicialización, la mediana aumenta su variabilidad, pero no necesariamente se incrementa la correlación con las observaciones, por lo que existe la posibilidad de que las medianas de mayo se correlacionen mejor a las observaciones (lo que no implica un mejor ajuste) que las medianas de junio. A partir de julio la incertidumbre de las IHCs disminuye considerablemente y el pronóstico se correlaciona mejor.

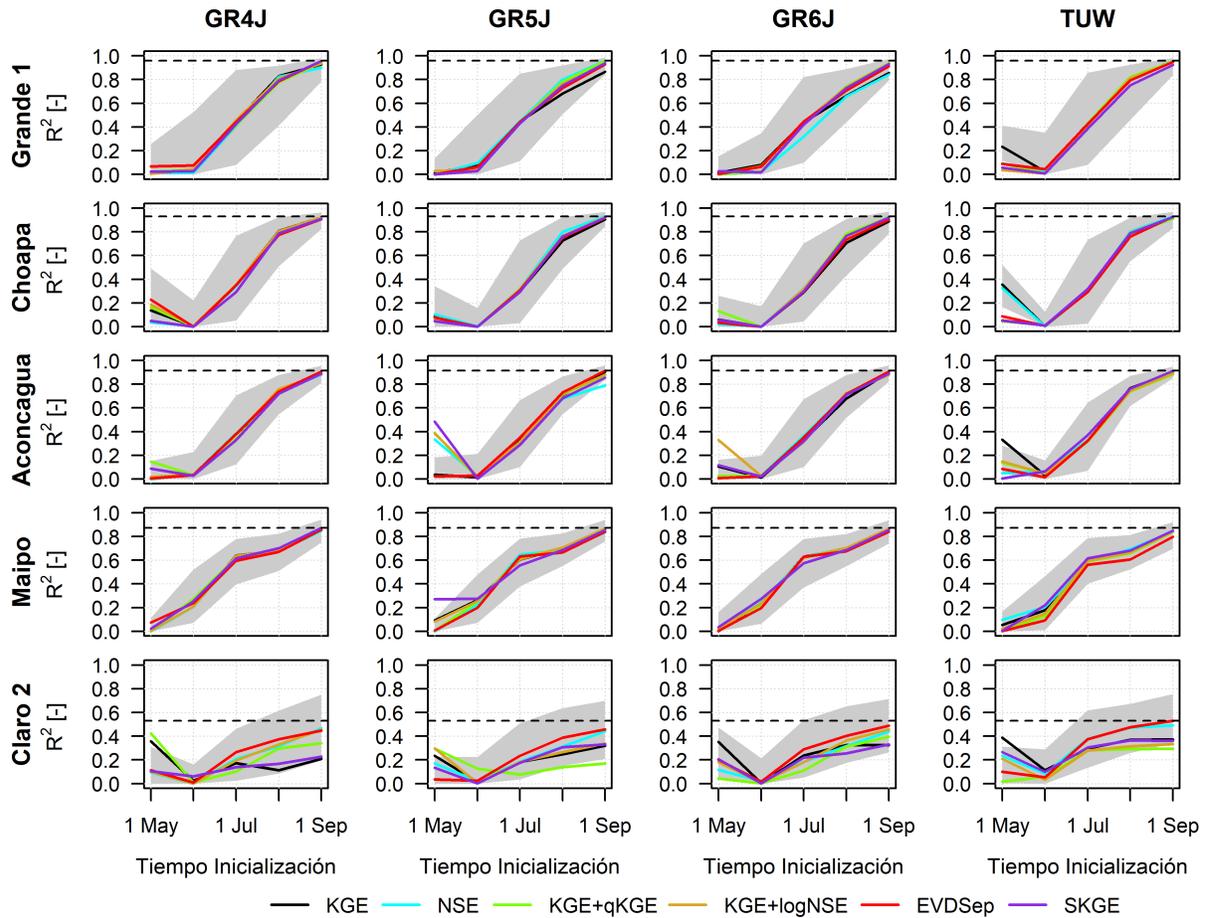


Figura 5.16: Comparación de los valores de  $R^2$  de las cuencas DGA por función objetivo, tiempo de inicialización y modelo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo obtenidos mediante *bootstrapping*. La línea discontinua demarca el máximo  $R^2$  obtenido para algún modelo el 1° de septiembre.

Finalmente, al comparar el modelo y la función objetivo que generan el mejor  $R^2$  con la mejor FO de los demás modelos, no hay diferencias significativas. Es decir, todos los modelos generan valores de  $R^2$  similares el 1° de septiembre. Además, se observa que la elección de la FO no afecta demasiado el comportamiento de esta métrica.

### 5.2.2. Pronósticos de volúmenes de deshielo de octubre a marzo

La Figura 5.17 presenta el CRPSS de los pronósticos agrupados por modelo y tiempo de inicialización. Se observa que la métrica mejora a medida que el pronóstico se inicializa más cerca del 1° de octubre (notar la escala de los paneles), y que el modelo TUW, en general, es el que presenta menos dispersión y entrega mayores valores de CRPSS.

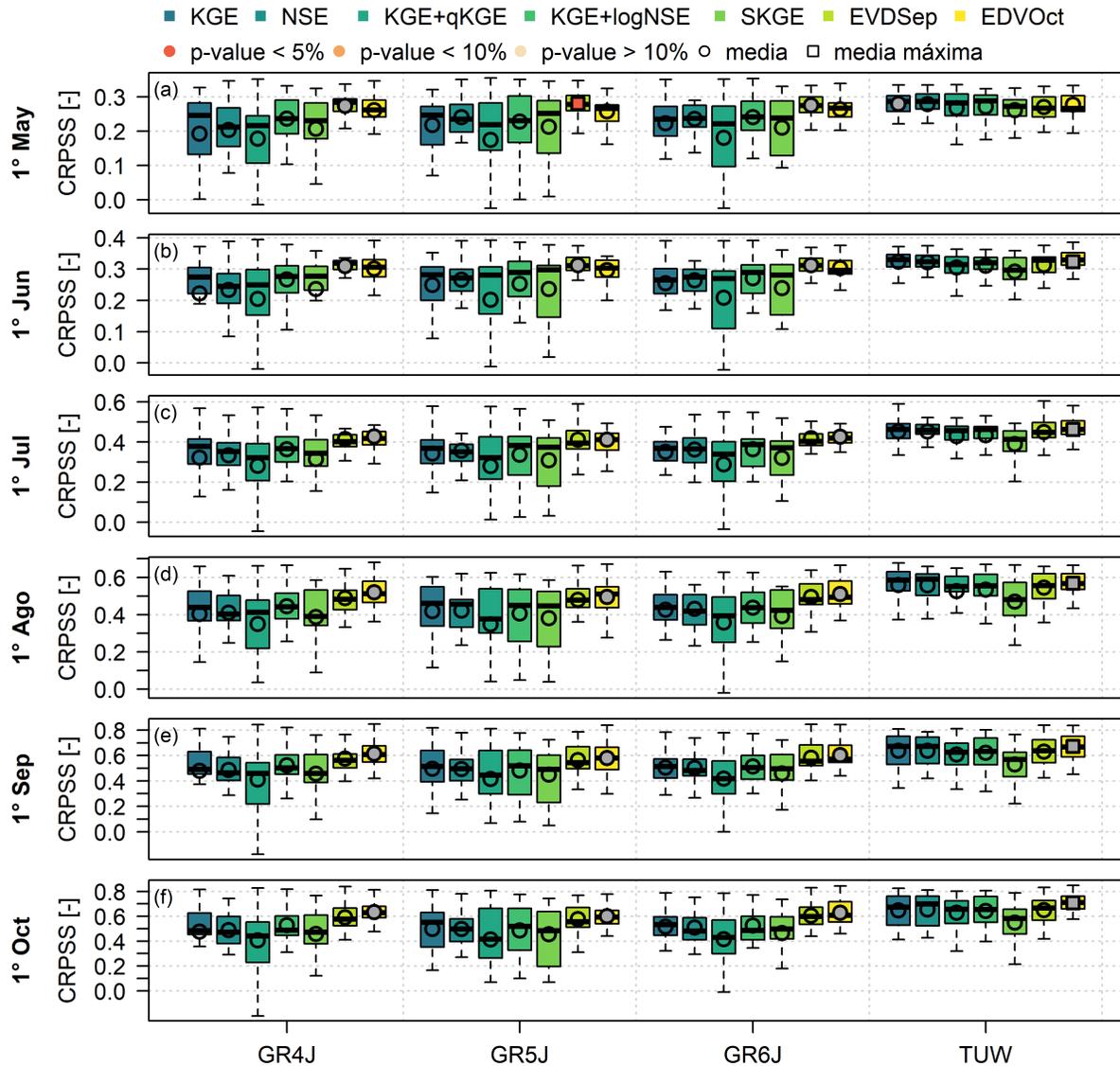


Figura 5.17: Comparación de los valores de CRPSS de los pronósticos ESP de octubre a marzo, obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y modelo hidrológico. Cada *boxplot* se compone del CRPSS de las 22 cuencas en estudio. El *p-value* indica la probabilidad de que la media del mejor conjunto sea significativamente similar a la del segundo mejor. Vale decir, un bajo *p-value* indica que las medias son significativamente diferentes. El círculo coloreado indica cual función objetivo de cada modelo entrega el mejor CRPSS, mientras que el cuadrado indica cual es el modelo con la media más alta de entre todos. Notar que los paneles no se encuentran en la misma escala.

La función objetivo que mejores CRPSS entrega es EVDOct, para todos los modelos inicializados entre el 1° de julio y el 1° de octubre. Para los modelos GR, en mayo y junio, EVDSep toma mayor relevancia, en tanto que, para el modelo TUV, KGE y EVDOct entregan mejores valores. Las funciones KGE+qKGE y SKGE son las que entregan los peores valores de CRPSS, presentando gran dispersión y, particularmente para KGE+qKGE, valores negativos en el modelo GR4J en múltiples tiempos de inicialización.

Con respecto al índice de confiabilidad  $\alpha$  (Figura 5.18), su valor disminuye a medida que el tiempo de inicialización se aproxima al 1° de octubre. Es interesante notar que, para el 1° de octubre, la función objetivo EVDOct es la que entrega mayor confiabilidad para todos los modelos, pero para el resto de los tiempos de inicialización, EVDSep es generalmente la mejor para los modelos GR, mientras que KGE es la mejor para el modelo TUW. Este último modelo es el que entrega las mayores confiabilidades para todos los tiempos de inicialización.

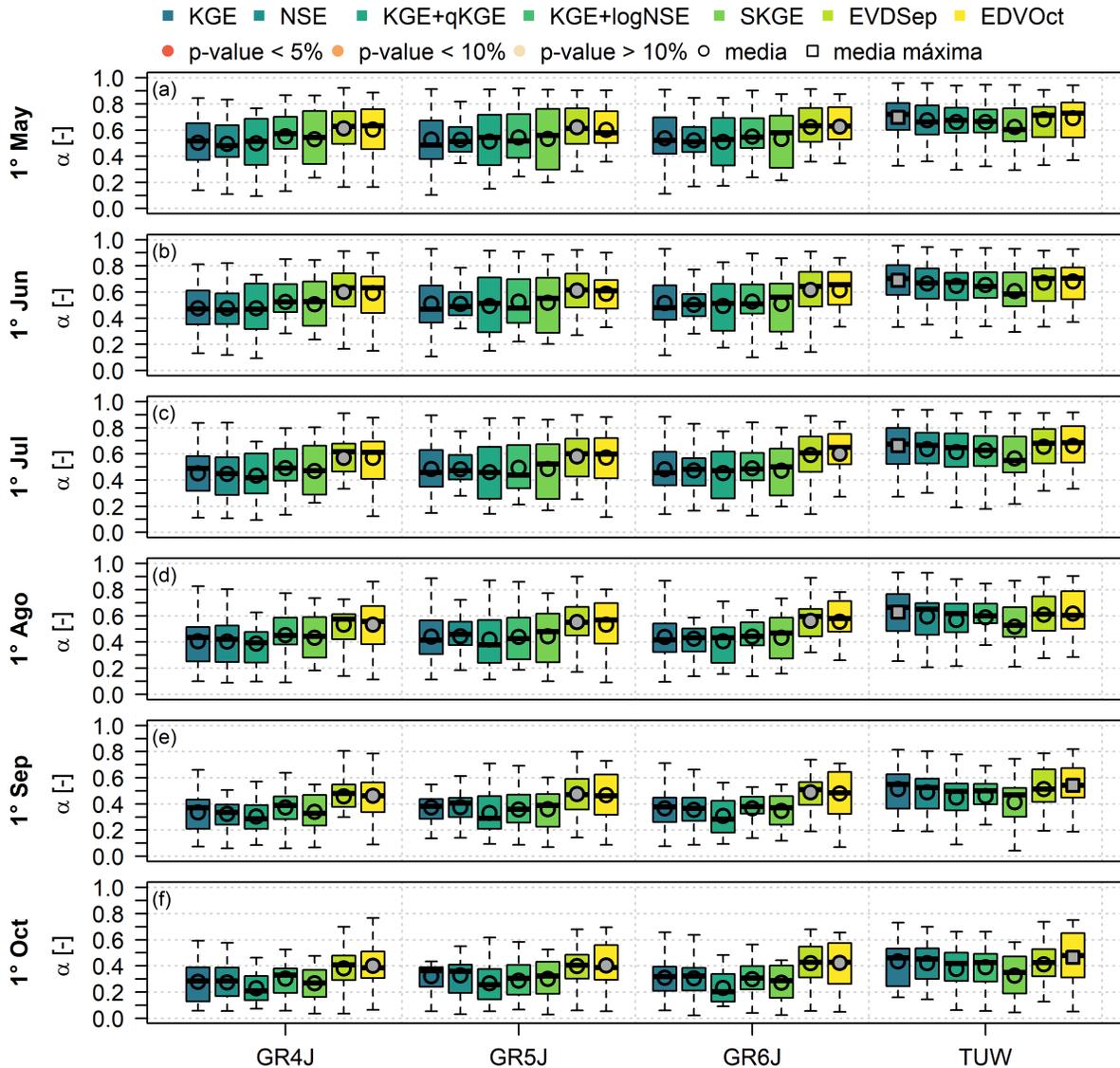


Figura 5.18: Comparación de los valores del índice  $\alpha$  de los pronósticos ESP de octubre a marzo, obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y modelo. Notar que los paneles cuentan con la misma escala

Finalmente, los resultados de  $R^2$  se presentan en la Figura 5.19, el cual mejora al acercarse al 1° de octubre. EVDOct es la función objetivo que mejores valores entrega, seguida por NSE. Si bien el comportamiento de la métrica es bastante homogéneo, el modelo TUW es el que, en general, entrega los mayores  $R^2$ .

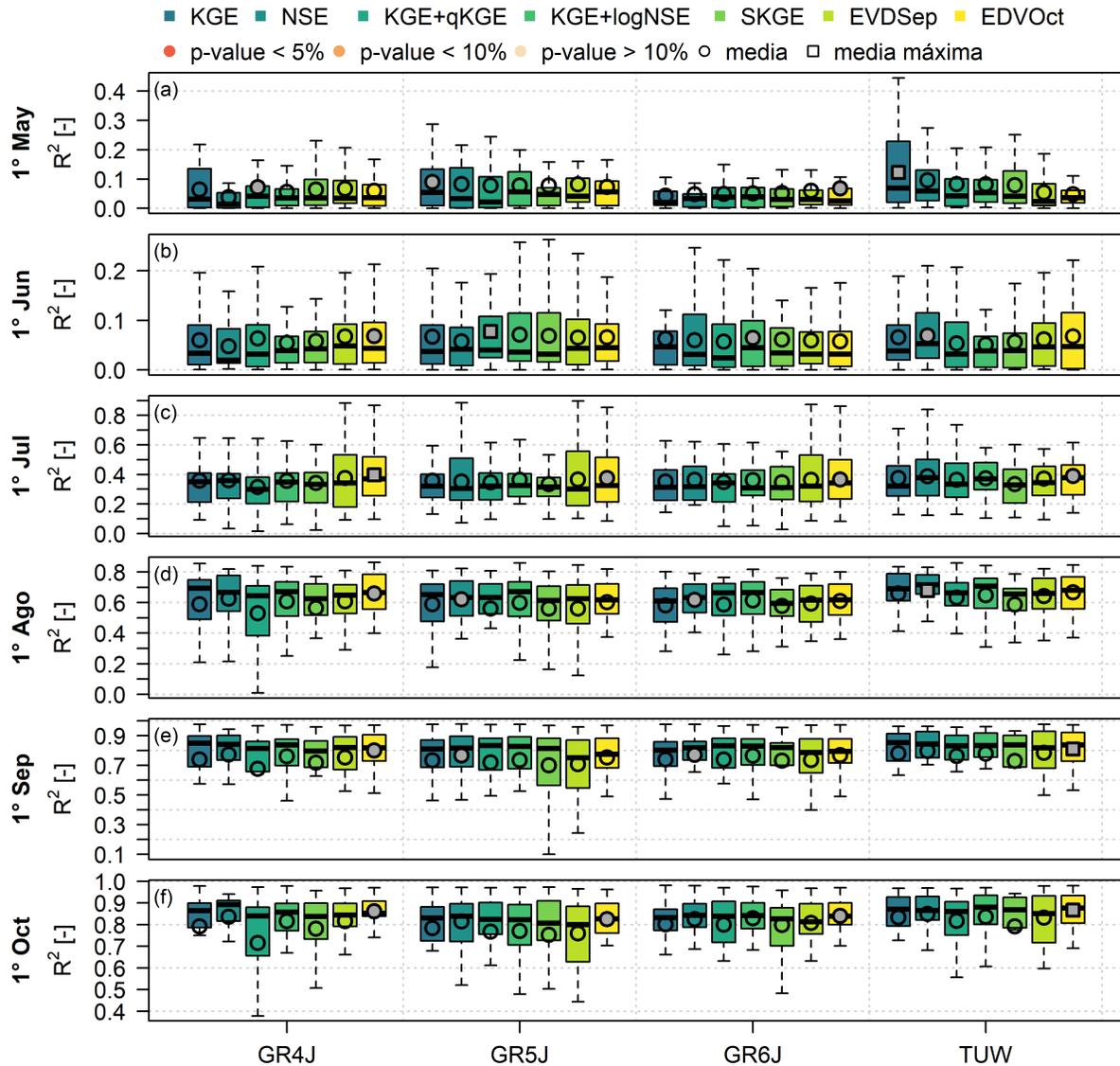


Figura 5.19: Comparación de los valores de  $R^2$  de los pronósticos ESP de octubre a marzo, obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y modelo. Notar que los paneles no se encuentran en la misma escala.

Al analizar el desempeño general en base a las tres métricas, se obtiene que la mejor función objetivo es EVDOct. Sin embargo, EVDSep también presenta buenos desempeños en CRPSS y en el índice  $\alpha$  para los modelos GR. Además, NSE se destaca en la métrica  $R^2$ .

Con respecto a los modelos, TUW nuevamente presenta los mejores resultados e, incluso, aumenta la brecha existente con la familia de modelos GR en cuanto a CRPSS, en comparación con el periodo de septiembre a marzo. De todas formas, se realiza una comparación, de carácter particular, entre los modelos con aquellas funciones objetivo que entreguen los mejores CRPSS en cada cuenca.

En la Tabla 5.2 se muestra el promedio de cada una de las métricas de todos los pronósticos inicializados en octubre, para todas las cuencas y funciones objetivo. Al igual que en el

pronóstico de septiembre a marzo, el modelo TUW es el que entrega los valores de las métricas más altos, siendo la diferencia con respecto al segundo mejor modelo estadísticamente significativa para CRPSS y el índice de confiabilidad  $\alpha$ . Además, la familia de modelos GR entrega valores similares, no destacándose ninguno por sobre otro.

Tabla 5.2: Comparación del promedio de las métricas de los pronósticos emitidos en octubre para cada modelo. Se destaca la diferencia entre los dos mejores modelos (TUW y GR6J). El *p-value* hace referencia al nivel de significancia de la diferencia entre las medias de los conjuntos que componen cada métrica.

Métrica	GR4J	GR5J	GR6J	TUW	Dif. dos mejores [%]	<i>p-value</i>
CRPSS	0,51	0,50	0,52	0,64	22,37	0.00
$\alpha$	0,30	0,33	0,32	0,40	23,14	0.01
R <sup>2</sup>	0,80	0,78	0,81	0,83	2,19	0.17

En el Anexo B, sección B.1, se presenta el mismo análisis de las tres métricas, pero agrupado por función objetivo. El resultado general muestra que el modelo TUW es superior al resto de los modelos al comparar la misma métrica, siendo esta superioridad estadísticamente significativa para todas las FO, menos para EVDSep y EVDOct.

### 5.2.2.1. Mejores CRPSS

En la Figura 5.20 se muestran los mejores CRPSS de los pronósticos y su distribución espacial, por modelo y tiempo de inicialización. Se observa que el desempeño de las cuencas de la zona sur es similar al resto de las cuencas para el 1° de octubre, lo que implica que, respecto a septiembre-marzo, el desempeño de estas cuencas es mejor al pronosticar en octubre-marzo.

Al analizar los modelos, es interesante notar que, en octubre, TUW entrega cinco cuencas con  $\text{CRPSS} \geq 0,8$ , notablemente más que GR4J que entrega sólo dos, y el resto de modelos GR sólo una. Además, inicializar en septiembre aún entrega buenos desempeños, con tres cuencas con  $\text{CRPSS} \geq 0,8$  para TUW, y dos para GR4J, aunque se pierde desempeño en las cuencas del sur. En agosto, nuevamente el modelo TUW es el único que mantiene varias cuencas con  $\text{CRPSS} \geq 0,6$ , y pierde menos rendimiento en las cuencas de la zona sur.

Con respecto a las funciones objetivo, EVDOct y EVDSep son las que generan los valores de CRPSS más altos. Es interesante notar que en las cuencas de la zona sur, predominan estas funciones estacionales, mientras que en las demás zonas es más variable.

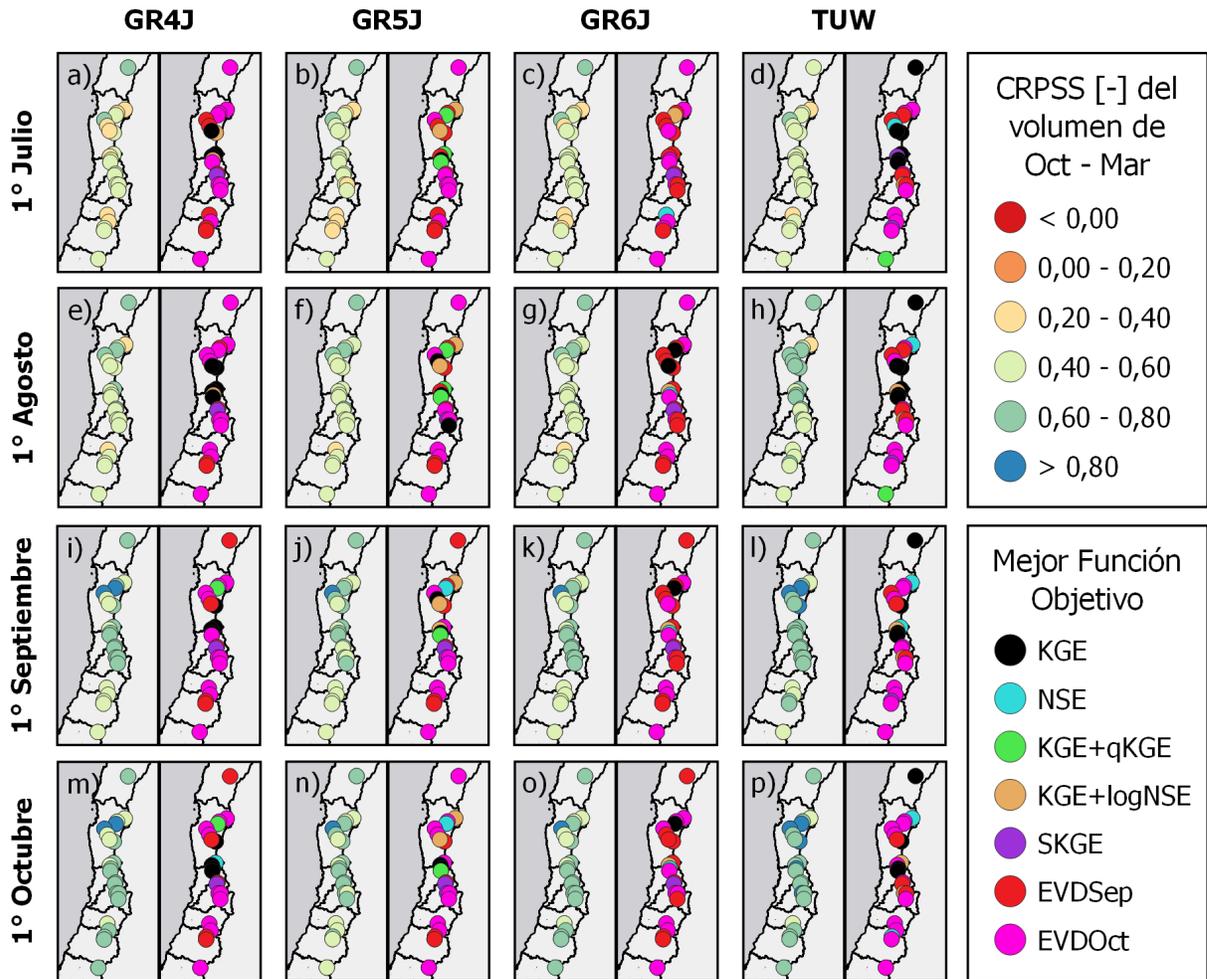


Figura 5.20: Distribución espacial de los mejores CRPSS del periodo de septiembre a marzo, junto a las funciones objetivo que los generan, para todos los modelos y tiempos de inicialización del 1° de julio al 1° de octubre.

En el Anexo B, sección B.2, se muestran los mejores índices  $\alpha$  y  $R^2$ . El resultado general indica que el modelo TUV es ligeramente superior a los modelos GR, y que las funciones objetivo EVDSep y EVDOct son las que generan los valores de las métricas más altos. En promedio, TUV entrega mejoras en CRPSS, índice  $\alpha$  y  $R^2$  de un 10%, 8% y 3%, respectivamente, con respecto al mejor de los modelos GR.

#### 5.2.2.2. Cuencas DGA

Las Figuras 5.21, 5.22 y 5.23, muestran la evolución del CRPSS, índice  $\alpha$  y  $R^2$ , para las cuencas DGA, según modelo, función objetivo y tiempos de inicialización. Para el resto de las cuencas véase el Anexo B, sección B.3.

Con respecto al CRPSS (Figura 5.21), se observa como aumenta su valor hacia el 1° de octubre, y como se va degradando hacia las cuencas del sur. En general, EVDSep y EVDOct

se destacan como las funciones objetivo que entregan los mejores valores de CRPSS. Al igual como ocurre en el periodo de septiembre a marzo, la correcta elección de la FO en los diferentes modelos consigue generar un CRPSS al nivel del modelo que genera el mejor valor de CRPSS (siendo TUW el mejor en la mayoría de los casos). En la cuenca Grande 1 no se cumple esta situación, ya que el modelo TUW es altamente superior, sin embargo, la elección de la FO sigue siendo relevante.

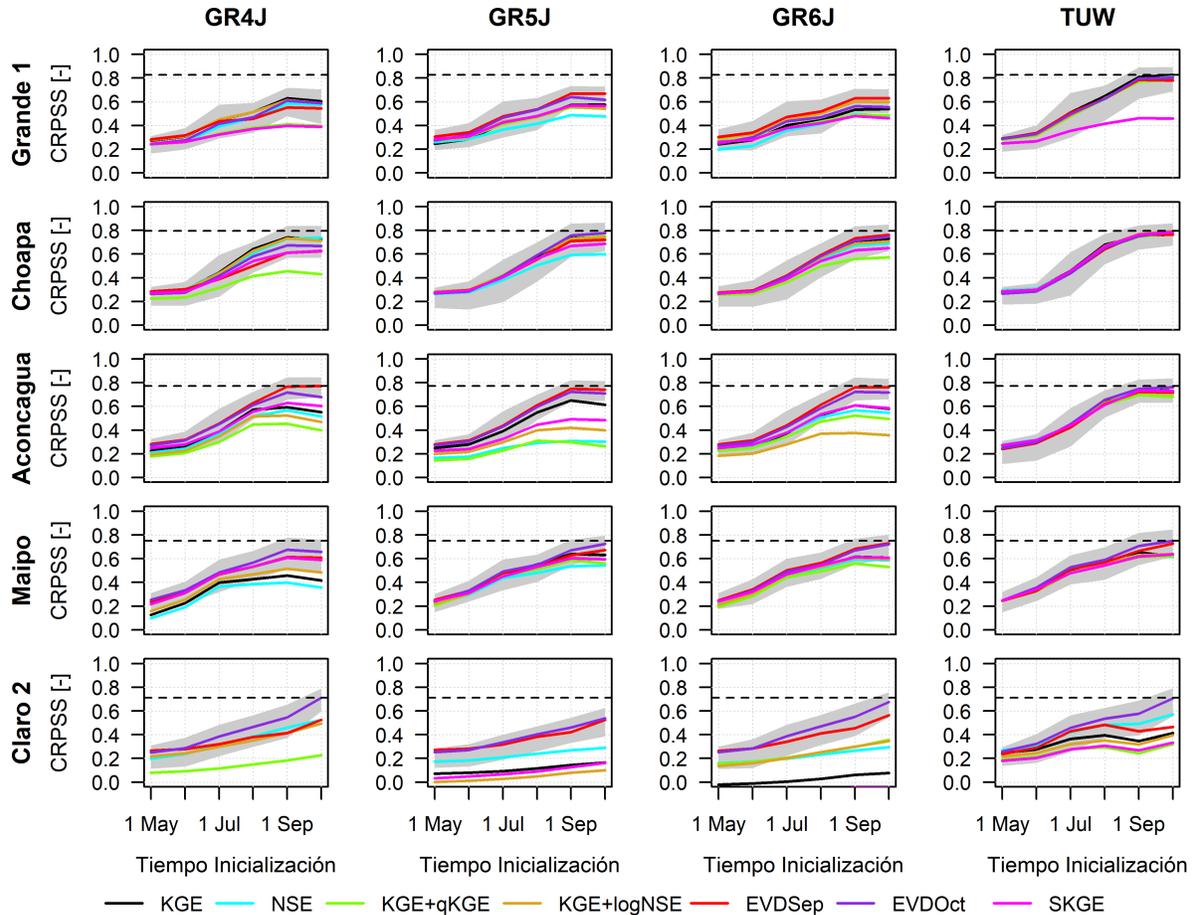


Figura 5.21: Comparación de los valores de CRPSS de las cuencas DGA por función objetivo, tiempo de inicialización y modelo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping*. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de octubre.

En relación al índice de confiabilidad  $\alpha$  (Figura 5.22), su valor disminuye hacia el 1° de octubre, y es consistentemente más alto en las cuencas centrales (Choapa, Aconcagua y Maipo). Nuevamente se destacan EVDSep y EVDOct como las funciones objetivo óptimas, las cuales, en general, presentan mayor estabilidad a través de los diferentes tiempos de inicialización. Si se compara la mejor combinación de modelo y su función objetivo (siendo TUW el mejor en la mayoría de los casos), con la mejor FO de los demás modelos, la diferencia es leve, por lo que es de gran importancia escoger una correcta FO.

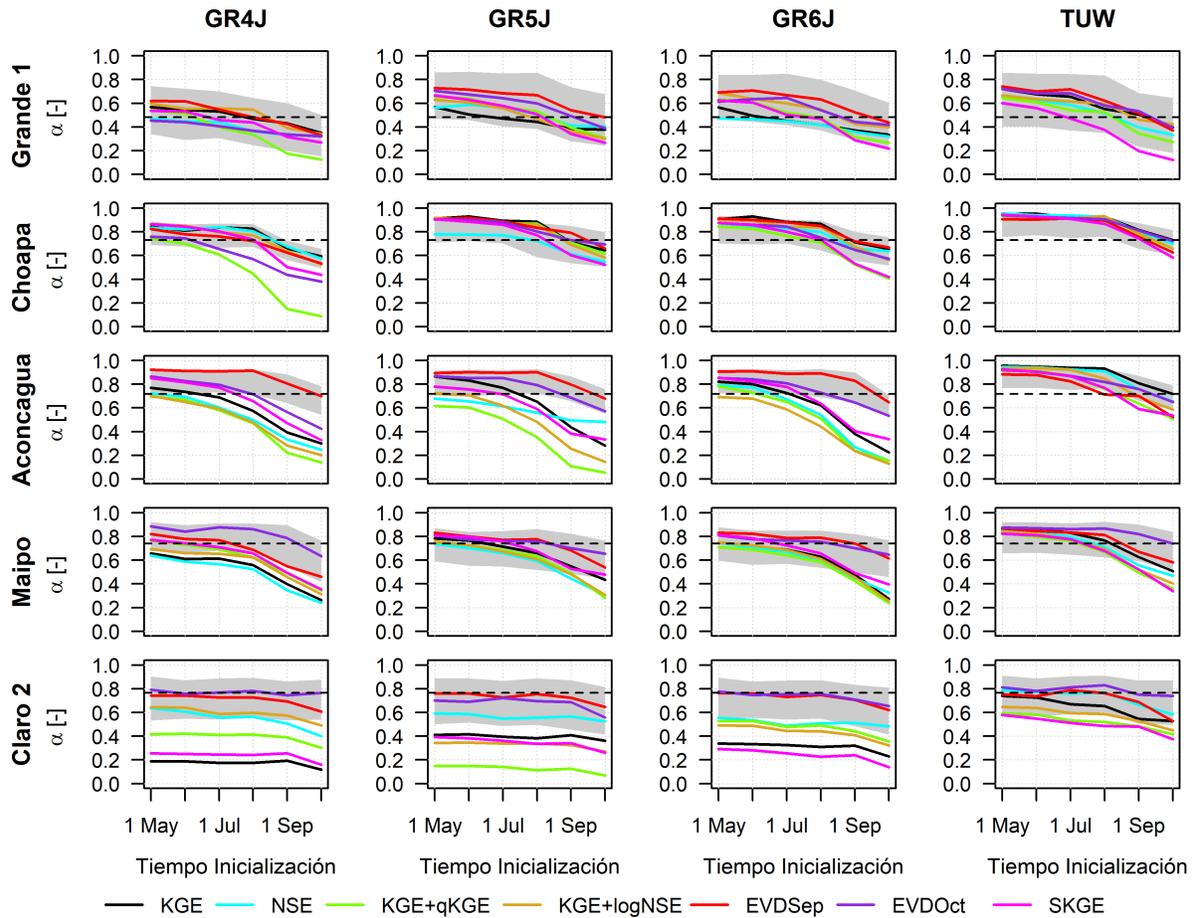


Figura 5.22: Comparación de los valores del índice  $\alpha$  de las cuencas DGA por función objetivo, tiempo de inicialización y modelo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping*. La línea discontinua demarca el máximo índice  $\alpha$  obtenido para algún modelo el 1° de octubre (tiempo de inicialización que no genera el mayor valor de  $\alpha$ , pero que coincide con los pronósticos más hábiles).

Los resultados de la Figura 5.23 muestran que el  $R^2$  presenta un comportamiento similar al de CRPSS, aumentando hacia el 1° de octubre y degradándose hacia las cuencas del sur. Al comparar el modelo y la función objetivo que generan el mejor  $R^2$  con la mejor FO de los demás modelos, no hay diferencias significativas. Es decir, todos los modelos generan valores de  $R^2$  similares el 1° de octubre. Además, se observa que la elección de la FO no afecta demasiado el comportamiento de esta métrica, excepto en la cuenca Claro 2.

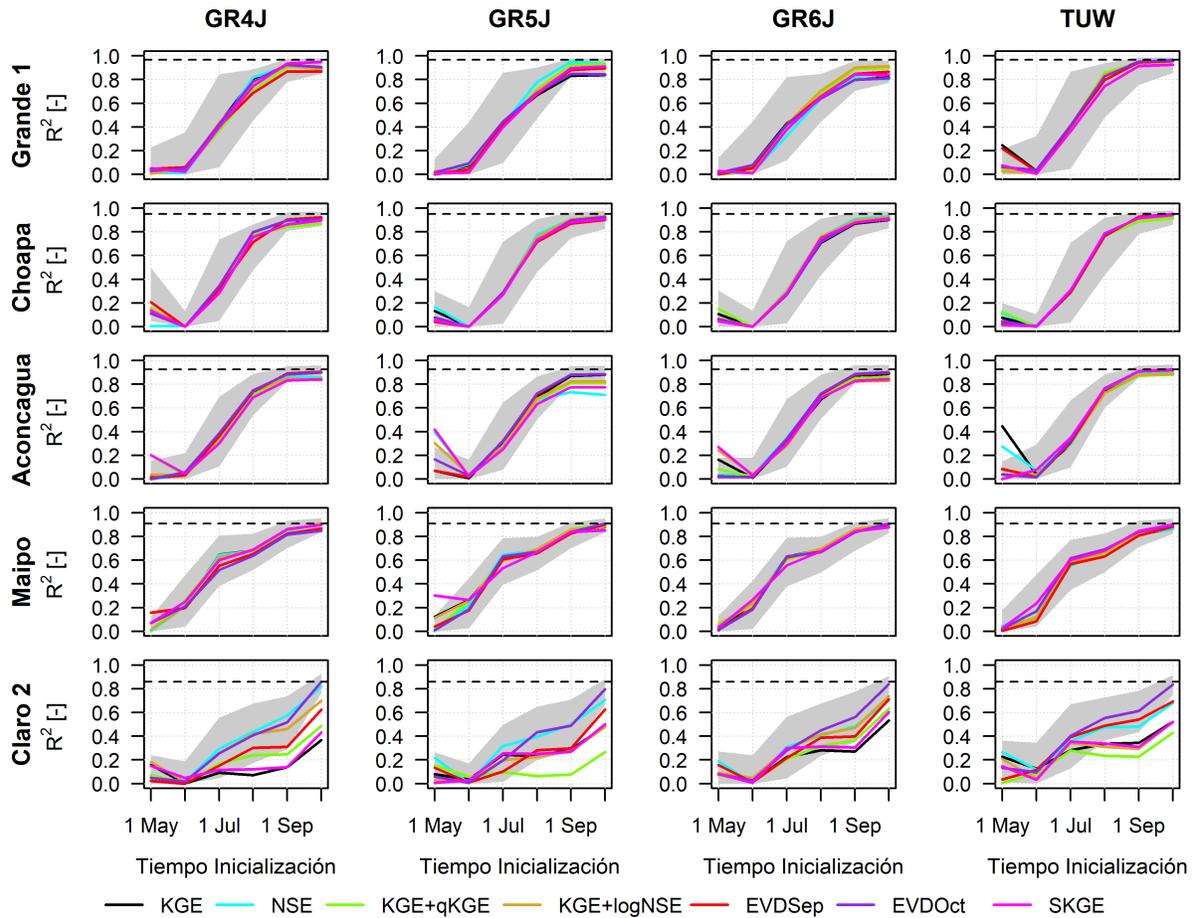


Figura 5.23: Comparación de los valores de  $R^2$  de las cuencas DGA por función objetivo, tiempo de inicialización y modelo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo obtenidos mediante *bootstrapping*. La línea discontinua demarca el máximo  $R^2$  obtenido para algún modelo el 1° de octubre.

### 5.3. Análisis de desempeño

Para complementar los resultados de los pronósticos, se llevaron a cabo dos análisis: el primero busca evaluar el desempeño de los pronósticos según la clasificación de las observaciones en años húmedos, normales y secos, utilizando la métrica CRPS, mientras que el segundo pretende encontrar relaciones entre el desempeño de los modelos y algunos descriptores hidroclimáticos y físicos de las cuencas (Tabla 3.2).

#### 5.3.1. Desempeño según la clasificación de las observaciones

En la Figura 5.24 se observa que, utilizando la función objetivo EVDsep –que involucra el uso de RMSE, métrica enfocada en valores altos– se obtiene un buen desempeño del CRPS en años clasificados como húmedos y normales según el volumen de agua observado, pero empeo-

ra considerablemente en años secos. Además, es interesante notar que el modelo TUW es el que mejor CRPS presenta en años secos. Por otro lado, si se utiliza la función KGE+logNSE, métrica enfocada en caudales altos y bajos, el desempeño del CRPS en años secos mejora notablemente, especialmente al inicializar el 1° de julio, 1° de agosto y 1° de septiembre. No obstante, en mayo y junio vuelve a empeorar el desempeño en años secos, aunque sigue siendo ligeramente mejor que EVDSep. Por último, para la función KGE+logNSE, el modelo TUW es el que presenta peores CRPS en periodos secos, con respecto a los modelos GR.

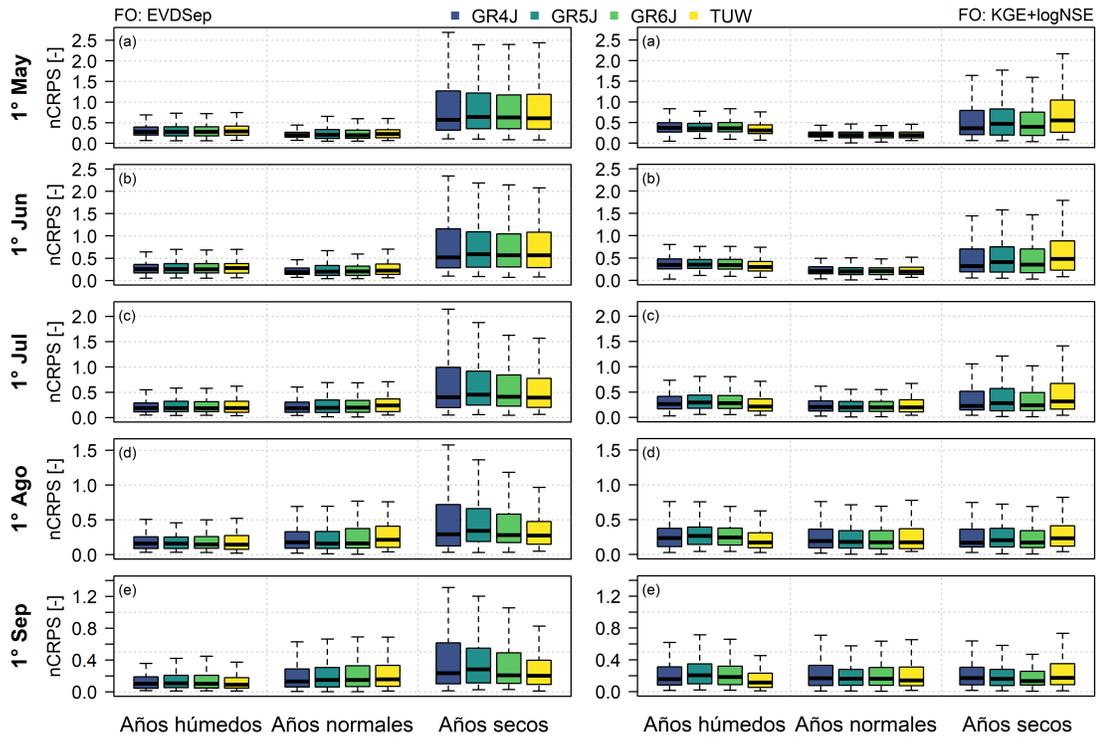


Figura 5.24: Análisis de desempeño del CRPS de los pronósticos al clasificar el volumen estacional observado en húmedo, normal y seco. La métrica nCRPS corresponde al CRPS normalizado por el volumen observado de septiembre a marzo de cada año, y varía entre  $[0, \infty]$ , con 0 el valor óptimo. Además, cada *boxplot* se compone de todos los años húmedos/normales/secos de las 22 cuencas en estudio.

### 5.3.2. Desempeño según descriptores hidroclimáticos y físicos

La Figura 5.25 muestra el resultado del agrupamiento para el modelo GR4J calibrado con KGE, para el periodo de deshielo de septiembre a marzo, y para todos los tiempos de inicialización. Las cuencas delta son aquellas cuyo promedio de las métricas es más alto que el promedio de las cuencas omega. Así, para el 1° de agosto y 1° de septiembre, las cuencas con mejor desempeño se ubican en la zona central de la zona de estudio. Para estos tiempos de inicialización, existe una relación significativa entre las mejores cuencas, un flujo base alto y una alta variabilidad inter-anual de la escorrentía. Además, para el 1° de septiembre existe una relación entre la fracción de nieve y las mejores cuencas. Para el resto de tiempos de inicialización, no hay relaciones claras, además, la significancia estadística (valores de *p-values*

altos) de las medias de los agrupamientos es baja.

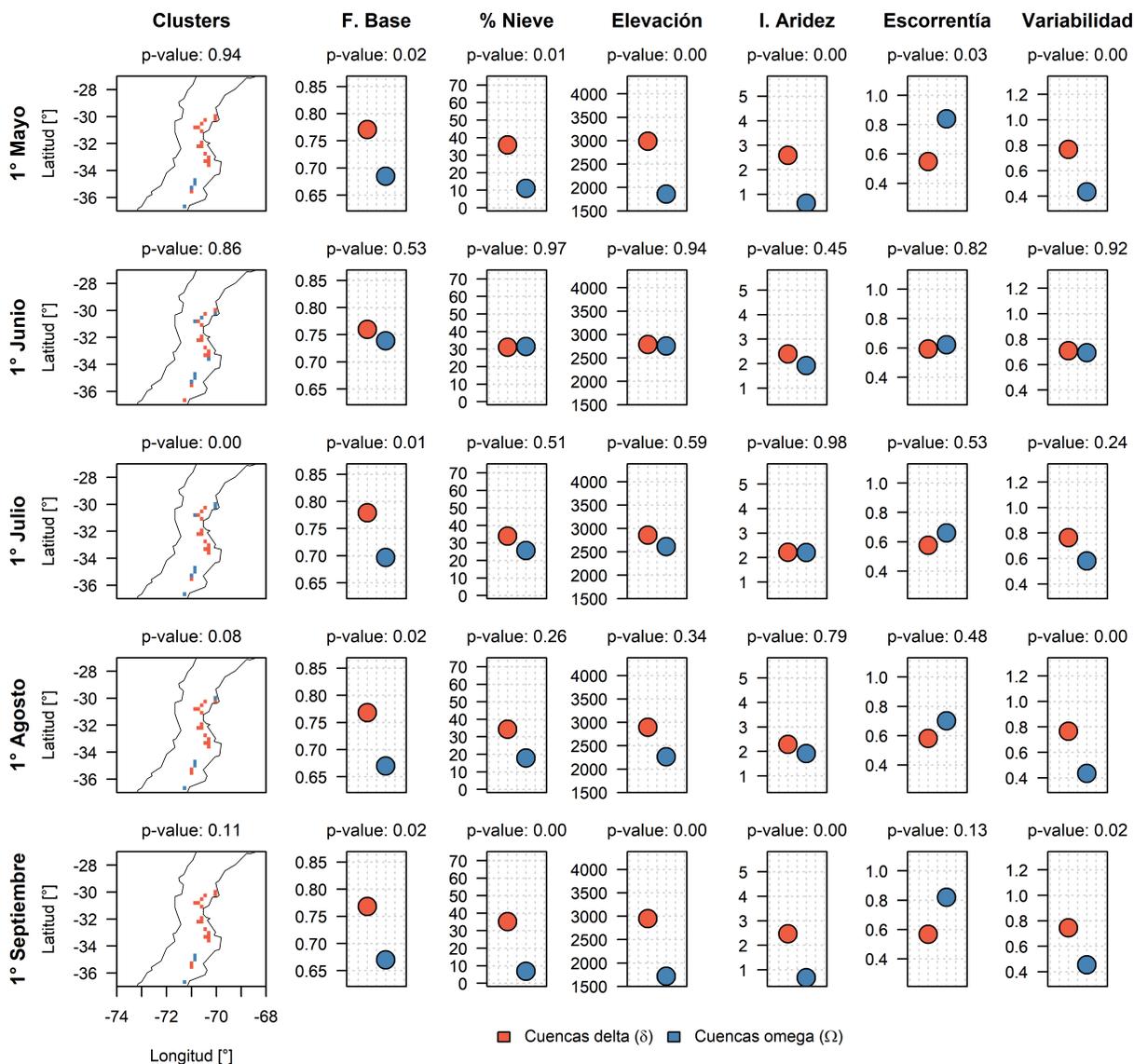


Figura 5.25: Relación entre el agrupamiento de cuencas y los descriptores de las 22 cuencas en estudio, para el modelo GR4J, la función objetivo KGE, periodo de septiembre a marzo y todos los tiempos de inicialización. El *p-value* indica si la diferencia entre las medias de cada grupo es significativa (menor *p-value*, mayor significancia). **F. Base** es el flujo base; **% Nieve** es la fracción de precipitación que cae como nieve; **Elevación** es la elevación media de las cuencas; **I. Aridez** es el índice de aridez; **Escorrentía** es el coeficiente de escorrentía; y **Variabilidad** es la variación inter-anual de la escorrentía.

En la Figura 5.26, se observa el mismo agrupamiento, pero para el modelo TUV, la función objetivo EVDOct y el periodo de octubre a marzo. Para los tiempos de inicialización 1° de octubre al 1° de agosto, se observan las mismas relaciones entre las cuencas delta y el flujo base y la variabilidad inter-anual de la escorrentía que en la Figura 5.25. Además, entre septiembre y agosto toma mayor relevancia la relación con la fracción de nieve. Para el resto de tiempos de inicialización, el agrupamiento de cuencas es poco significativo (i.e., las medias

de los grupos de cuencas son similares), por lo que es esperable que las relaciones existentes también lo sean.

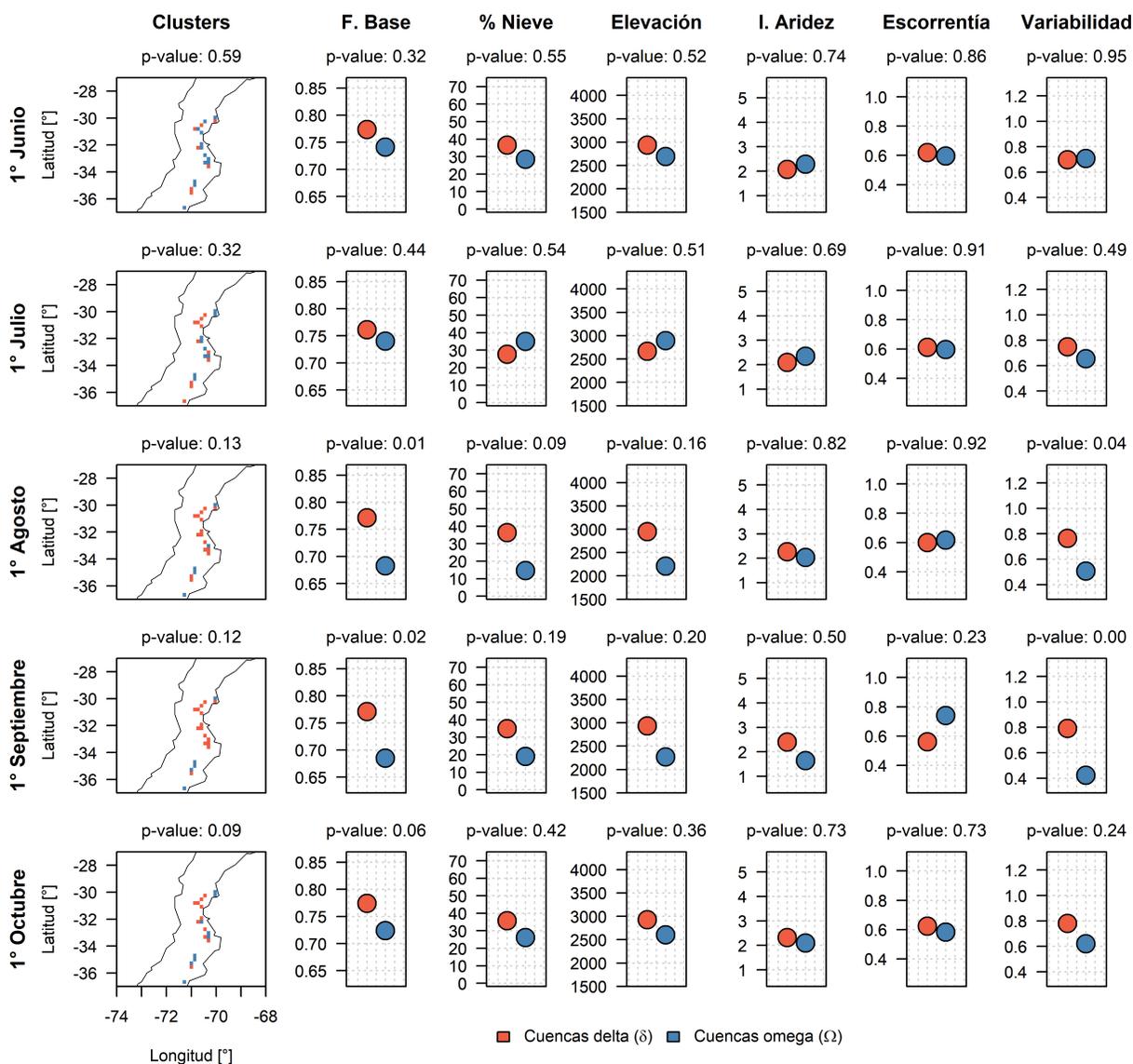


Figura 5.26: Relación entre el agrupamiento de cuencas y los descriptores de las 22 cuencas en estudio, para el modelo TUW, la función objetivo EVDOct, periodo de octubre a marzo y tiempos de inicialización del 1° de junio al 1° de octubre. **F. Base** es el flujo base; **% Nieve** es la fracción de precipitación que cae como nieve; **Elevación** es la elevación media de las cuencas; **I. Aridez** es el índice de aridez; **Escorrentía** es el coeficiente de escorrentía; y **Variabilidad** es la variación inter-anual de la escorrentía.

Resulta interesante notar que, para modelos, funciones objetivo, y periodos de evaluación diferentes, hay correlaciones que persisten. Existe un mejor desempeño de los pronósticos en cuencas con mayor flujo base (i.e., respuesta más lenta), relación que también reportaron Harrigan et al. (2018). También, a mayor coeficiente de variabilidad (i.e., cuencas con escorrentías anuales más variables), mejores resultados se observan. Por último, cuencas con una mayor fracción de precipitación que cae como nieve también se desempeñan mejor, situación

que se relaciona con una respuesta más lenta de las cuencas (e.g., Girons Lopez et al., 2021).

## 5.4. Pronósticos en las cuencas DGA

### 5.4.1. Volúmenes estacionales

En la Figura 5.27 se presentan los pronósticos de volumen estacional para las cuencas DGA por modelo (se omite el GR5J). Los pronósticos fueron generados con IHCs del 1° de septiembre (al igual que la DGA) y la función objetivo utilizada corresponde a la mejor obtenida para la serie completa de cada cuenca y modelo, según el  $R^2$  y el CRPSS.

Además, en las Tablas 5.3 a 5.7 se presenta el  $R^2$ , el sesgo porcentual, el error cuadrático medio (RMSE; en millones de metros cúbicos) y el error absoluto medio (MAE; en millones de metros cúbicos) para complementar la evaluación de los pronósticos.

Tabla 5.3: Evaluación del pronóstico estacional de la cuenca Grande 1.

Grande 1	DGA	GR4J	GR6J	TUW
$R^2$	0,72	0,82	0,60	0,81
Sesgo	46,0	-20,9	-3,3	-13,0
RMSE	41	18	27	19
MAE	28	15	18	13

Tabla 5.4: Evaluación del pronóstico estacional de la cuenca Choapa.

Choapa	DGA	GR4J	GR6J	TUW
$R^2$	0,95	0,92	0,90	0,93
Sesgo	4,6	-6,4	1,5	-3,8
RMSE	42	53	56	49
MAE	29	41	42	38

Tabla 5.5: Evaluación del pronóstico estacional de la cuenca Aconcagua.

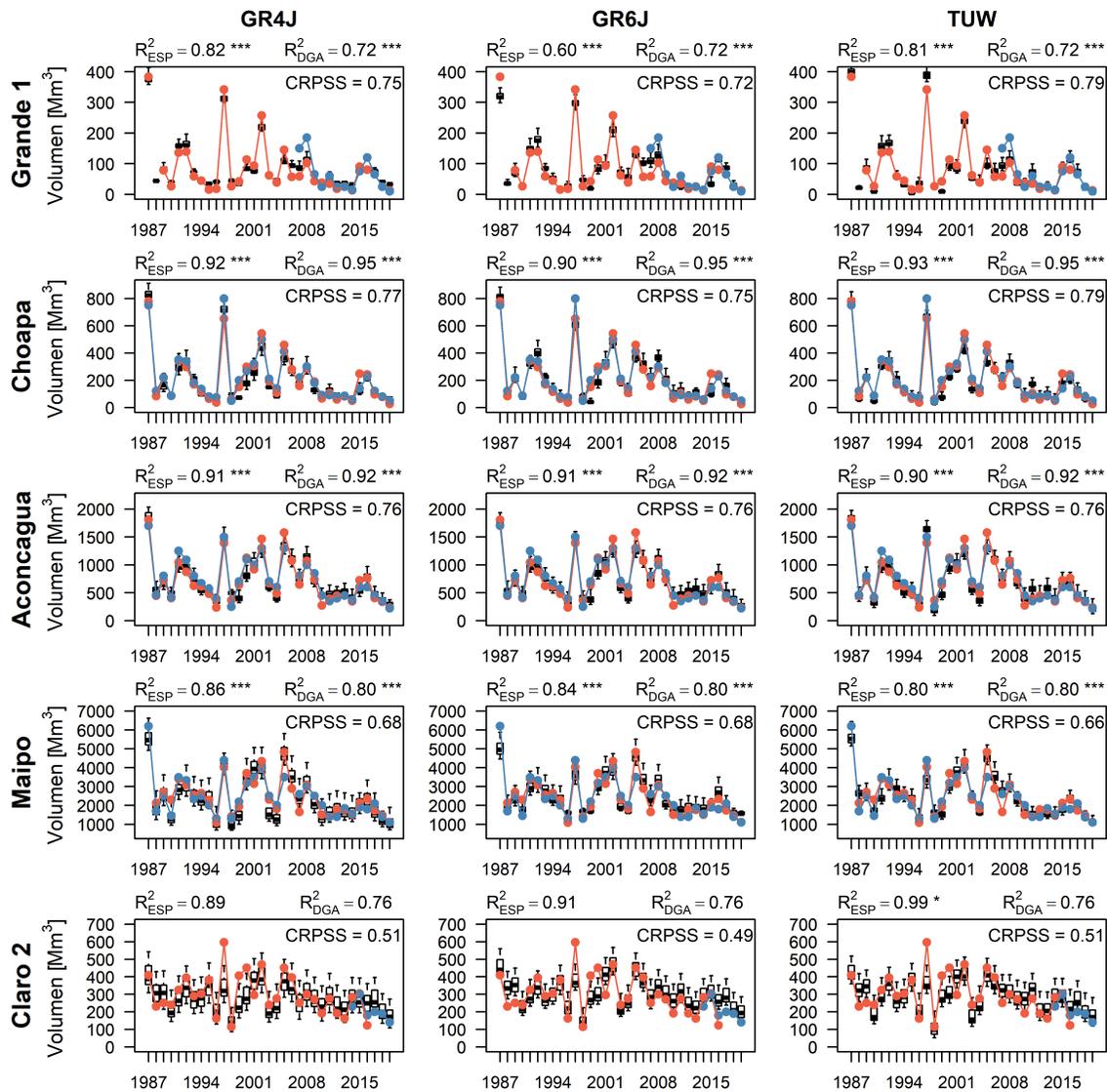
Aconcagua	DGA	GR4J	GR6J	TUW
$R^2$	0,92	0,91	0,91	0,90
Sesgo	3,3	0,3	-0,2	-3,0
RMSE	118	126	129	129
MAE	96	102	103	103

Tabla 5.6: Evaluación del pronóstico estacional de la cuenca Maipo.

Maipo	DGA	GR4J	GR6J	TUW
$R^2$	0,80	0,86	0,84	0,80
Sesgo	-2,7	-7	-0,9	-2,4
RMSE	424	406	372	414
MAE	311	307	323	308

Tabla 5.7: Evaluación del pronóstico estacional de la cuenca Claro 2.

Claro 2	DGA	GR4J	GR6J	TUW
$R^2$	0,76	0,89	0,91	0,99
Sesgo	-0,2	-6,1	-15,5	-5,8
RMSE	45	80	77	58
MAE	38	68	46	47



● Obs ● DGA □ ESP \*\*\*: p-value < 1% \*\*: p-value < 5% \*: p-value < 10%

Figura 5.27: Pronóstico determinístico de las cuencas DGA, inicializado el 1° de septiembre.  $R^2_{ESP}$  y  $R^2_{DGA}$  son las correlaciones entre la mediana del conjunto de pronósticos y las observaciones, y entre el pronóstico oficial de la DGA y las observaciones, respectivamente. El *p-value* indica el nivel de significancia de la correlación. Las observaciones corresponden al volumen de septiembre a marzo, y la correlación se realiza en la ventana de años en que la DGA tenga pronósticos (i.e., la correlación en Grande 1 se hace solo con las observaciones a partir de 2007, ya que desde ese año es oficialmente pronosticada).

Con respecto a Grande 1 (Tabla 5.3), la DGA emite pronósticos para esta cuenca desde el año 2007. Se observa que los pronósticos ESP emitidos con los modelos GR4J y TUW presentan mejores indicadores que la DGA, mientras que el pronóstico del modelo GR6J sólo es inferior en  $R^2$ . Además, este último presenta el menor sesgo.

En las cuencas Choapa y Aconcagua (Tablas 5.4 y 5.5), se observa que la DGA emite un pronóstico satisfactorio, presentando métricas superiores a casi todos los pronósticos ESP. Sin embargo, estos tienden a mejorar a la DGA en cuanto al sesgo porcentual. Además, los pronósticos generados con los modelos TUW en Choapa, y GR6J en Aconcagua, son los que entregan métricas más cercanas a los pronósticos DGA en cuanto a  $R^2$ , RMSE y MAE.

Con respecto a la cuenca del Maipo (Tabla 5.6), todos los pronósticos ESP igualan o mejoran a la DGA en cuanto a  $R^2$ . Además, los pronósticos emitidos con los modelos GR6J y TUW la mejoran en el sesgo y en el RMSE. El pronóstico DGA es levemente superior a ESP en términos de MAE.

Para la cuenca Claro 2 (5.7) la DGA emite pronósticos a partir de 2014, por lo que hay pocos valores para comparar (razón por la cual la correlación no es estadísticamente significativa). La DGA logra mejores valores en todas las métricas, menos en  $R^2$ . El pronóstico con el modelo TUW es el que más se aproxima al de la DGA, pero aún así presenta errores altos.

En cuanto a los modelos, y a pesar de que en el análisis general el modelo TUW entrega mejores indicadores para los pronósticos ESP en comparación con el resto de los modelos examinados, la diferencia es leve al pronosticar individualmente las cuencas utilizando la mejor función objetivo posible. Sin ir más lejos, el modelo GR6J es el que entrega pronósticos con el menor sesgo.

Finalmente, se debe destacar el valor agregado de la metodología ESP por sobre la de la DGA: se obtiene la probabilidad de ocurrencia de los caudales futuros; se entregan diversos flujos como el equivalente de agua de nieve o la evaporación/evapotranspiración; la descripción temporal del periodo completo de los flujos, entre otros.

#### 5.4.2. Volúmenes mensuales

La DGA, emite pronósticos de caudales medios mensuales, mediante la distribución del volumen estacional en los siete meses que dura el periodo de deshielo (septiembre a marzo). De esta forma, resulta de interés analizar el desempeño de los modelos a resolución temporal mensual para comparar con mayor detalle el producto DGA con la metodología ESP y, además, evaluar de manera parcial las diferencias entre funciones objetivo.

En la Figura 5.28 se presentan los tres índices utilizados para evaluar los pronósticos mensuales de las cinco cuencas DGA, inicializados en septiembre y agrupados por modelo. Se observa que la función objetivo NSE genera los valores de CRPSS y  $R^2$  más altos para los modelos GR4J y TUW, mientras que, SKGE entrega los valores más bajos de CRPSS y EVD-Sep los más bajos de  $R^2$  para los mismos modelos. Además, para el modelo GR6J, la función objetivo KGE+KGE entrega los valores del índice  $\alpha$  más altos, mientras que EVD-Sep entrega los valores más bajos. Sólo se destacan las diferencias que son estadísticamente significativas debido a la cantidad reducida de cuencas evaluadas.

En el Anexo C se presenta el detalle de los tres índices para cada cuenca DGA. En general,

se aprecia que NSE es la función objetivo que mejores CRPSS y  $R^2$  entrega, mientras que EVDSep y SKGE son las que peores valores entregan en los mismos índices.

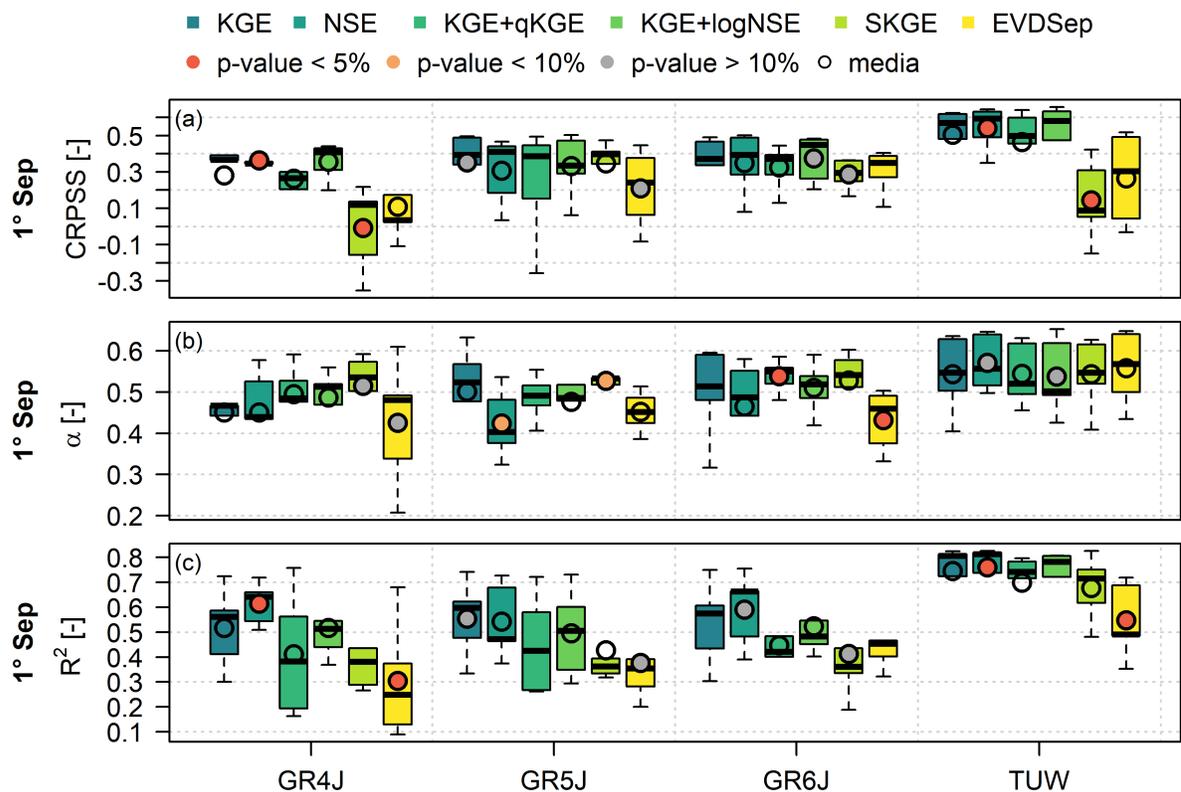


Figura 5.28: Comparación de los valores de CRPSS, índice  $\alpha$  y coeficiente de determinación  $R^2$  de los pronósticos ESP mensuales de septiembre a marzo, obtenidos con distintas funciones objetivo, agrupados por modelo e inicializados en septiembre. Las tres métricas fueron obtenidas de manera similar a como se muestra en la Figura 5.9, pero con el volumen desagregado por mes. Notar que los paneles no cuentan con la misma escala, y cada *boxplot* se compone de las cinco cuencas DGA en estudio. Los puntos coloreados indican las FO que entregan el máximo y mínimo valor de cada índice, y el nivel de significancia entre la diferencia de sus medias. Cabe destacar que el coeficiente de determinación se genera utilizando la mediana de los pronósticos ESP.

En las Tablas 5.8 a 5.12 se compara el pronóstico oficial de volúmenes mensuales de la DGA con los pronósticos mensuales ESP. Se utilizaron las funciones objetivo NSE y EVDSep, ya que la primera es la que mejores valores de CRPSS y  $R^2$  entrega, y la segunda es una de las que entrega los peores valores (aunque a resolución temporal estacional es de las mejores). Se puede notar que los pronósticos emitidos por la DGA son superiores a aquéllos emitidos con el método ESP para  $R^2$ , RMSE y MAE, en todas las cuencas, excepto Grande 1. Es interesante observar que, en general, NSE entrega mejores valores que EVDSep en todos los indicadores, excepto en el sesgo, donde los pronósticos emitidos con EVDSep son menos sesgados, incluso, que los de la DGA. Además, el modelo TUW entrega el mejor desempeño general con respecto a los modelos GR.

El análisis sugiere que la función objetivo EVDSep es deficiente para pronosticar volú-

menes a resolución temporal más fina (recordando que EVDSep es utilizada para calibrar minimizando el volumen estacional), lo que ya se suponía en el análisis de los índices hidrológicos (Figura 5.6). No obstante, entrega pronósticos menos sesgados, lo que se debe, también, a su naturaleza estacional.

Tabla 5.8: Evaluación del pronóstico mensual de la cuenca Grande 1.

Grande 1	DGA	GR4J		GR5J		GR6J		TUW	
FO	-	NSE	EVDSep	NSE	EVDSep	NSE	EVDSep	NSE	EVDSep
R <sup>2</sup>	0,58	0,33	0,55	0,58	0,50	0,12	0,44	0,59	0,63
Sesgo	43,3	14,6	27,8	12,2	9,1	-52,0	8,1	2,7	14,0
RMSE	7,6	5,2	4,6	4,3	5,1	7,2	5,4	4,4	4,5
MAE	4,1	3,8	3,4	3,2	3,0	5,2	3,2	2,6	3,2

Tabla 5.9: Evaluación del pronóstico mensual de la cuenca Choapa.

Choapa	DGA	GR4J		GR5J		GR6J		TUW	
FO	-	NSE	EVDSep	NSE	EVDSep	NSE	EVDSep	NSE	EVDSep
R <sup>2</sup>	0,90	0,67	0,09	0,68	0,29	0,67	0,47	0,82	0,69
Sesgo	4,4	-11,4	4,3	-16,1	-2,8	-15,8	-1,2	-7,0	-5,4
RMSE	11,2	20,8	49,1	21,8	34,0	21,4	27,2	15,2	20,1
MAE	7,1	12,0	20,3	12,4	16,5	11,8	14,5	9,5	12,4

Tabla 5.10: Evaluación del pronóstico mensual de la cuenca Aconcagua.

Aconcagua	DGA	GR4J		GR5J		GR6J		TUW	
FO	-	NSE	EVDSep	NSE	EVDSep	NSE	EVDSep	NSE	EVDSep
R <sup>2</sup>	0,86	0,64	0,68	0,47	0,66	0,67	0,60	0,82	0,36
Sesgo	3,0	-20,6	-1,3	-19,7	-1,6	-22,2	-1,3	-10,5	-0,5
RMSE	30,9	53,6	46,5	64,8	48,2	53,2	51,7	36,4	86,2
MAE	21,1	35,8	31,9	42,0	31,7	34,5	33,6	26,7	56,1

Tabla 5.11: Evaluación del pronóstico mensual de la cuenca Maipo.

Maipo	DGA	GR4J		GR5J		GR6J		TUW	
FO	-	NSE	EVDSep	NSE	EVDSep	NSE	EVDSep	NSE	EVDSep
R <sup>2</sup>	0,79	0,73	0,25	0,73	0,20	0,76	0,45	0,74	0,49
Sesgo	-3,4	-24,4	-8,0	-16,5	-1,4	-16,5	-2,4	-14,1	-5,9
RMSE	94,9	136,3	217,6	124,7	205,0	118,6	157,1	113,7	152,2
MAE	63,5	105,9	157,2	85,8	148,1	80,6	103,5	81,2	99,3

Tabla 5.12: Evaluación del pronóstico mensual de la cuenca Claro 2.

Claro 2	DGA	GR4J		GR5J		GR6J		TUW	
FO	-	NSE	EVDSep	NSE	EVDSep	NSE	EVDSep	NSE	EVDSep
R <sup>2</sup>	0,72	0,38	0,39	0,34	0,42	0,35	0,40	0,55	0,51
Sesgo	-4,6	-2,3	17,0	10,0	21,8	6,7	17,5	7,6	-2,4
RMSE	9,8	14,5	19,4	15,2	18,9	14,8	16,8	13,4	15,8
MAE	7,3	11,4	14,3	13,2	14,1	12,6	13,2	9,6	11,8

# Capítulo 6

## Conclusiones y recomendaciones

En el presente estudio se calibraron y validaron cuatro modelos hidrológicos conceptuales en 22 cuencas de Chile central (latitudes comprendidas entre 28°S y 37°S, aproximadamente), con el objetivo de evaluar el método de predicción de caudales por conjuntos (ESP, por sus siglas en inglés) en la generación de pronósticos de caudales de deshielo. Para conseguir el objetivo, se emplearon siete funciones objetivo y seis tiempos de inicialización. Además, para cuantificar objetivamente el desempeño de los pronósticos se utilizaron métricas probabilísticas (CRPSS y el índice de confiabilidad  $\alpha$ ) y determinísticas (coeficiente de determinación  $R^2$ , sesgo y el error medio cuadrático).

### 6.1. Conclusiones del estudio

De manera general, los resultados de este estudio indican que el método ESP presenta valores de CRPSS e índice  $\alpha$  que dependen del tiempo de inicialización, la función objetivo utilizada para calibrar, las características hidroclimáticas de las cuencas y del modelo seleccionado.

Con respecto al CRPSS, sus valores decaen rápidamente a medida que el instante de inicialización se aleja del comienzo del periodo de deshielo. Este descenso es similar para los periodos de septiembre-marzo y octubre-marzo. Esto quiere decir que las condiciones hidrológicas iniciales (IHCs) adquieren menos relevancia hacia el inicio del año hidrológico, ya que existe menos agua acumulada en las cuencas. El modelo TUW es el que genera los valores de CRPSS más altos, con valores un 13% y un 22% (en promedio) más altos que el mejor modelo GR en los periodos de septiembre-marzo y octubre-marzo, respectivamente. Sin embargo, si se utiliza la FO que genera, en cada modelo, los mejores valores de CRPSS, la diferencia entre los modelos disminuye, obteniéndose que TUW es superior al mejor modelo GR en un 6% y un 10%, respectivamente. Esta situación sugiere que la elección de la función objetivo es más relevante que la elección del modelo.

En cuanto a la confiabilidad  $\alpha$  de ESP, ésta decae a medida que el instante de inicialización se aleja del comienzo del periodo de deshielo. En este caso la elección de la función objetivo es fundamental ya que, en general, la mejor FO genera una pérdida de confiabilidad menos pronunciada, contrarrestando la pérdida asociada al tiempo de inicialización. La elección

del modelo, en un inicio, parece ser de gran relevancia, ya que TUW genera, en promedio, valores de confiabilidad un 21 % y 23 % más altos que el mejor modelo GR en los periodos de septiembre-marzo y octubre-marzo, respectivamente. No obstante, al utilizar la función objetivo que, en cada modelo, genera los valores de confiabilidad más altos, esta diferencia disminuye hasta llegar a un 9 % y 8 %, respectivamente.

Las mejores funciones objetivo corresponden a las que buscan minimizar los errores en volúmenes estacionales (EVD<sub>Sep</sub> y EVD<sub>Oct</sub>), especialmente en la familia de modelos GR. Incluso, los valores de CRPSS más altos ( $\geq 0,8$ ) y las confiabilidades más altas en la época de deshielo, se generan, en su mayoría, con estas funciones. Sin embargo, esta superioridad no se refleja al realizar un pronóstico a resolución temporal mensual. En general, EVD<sub>Sep</sub> genera volúmenes mensuales de baja correlación con las observaciones, con respecto a las funciones objetivo que minimizan los caudales diarios, como por ejemplo NSE. Cabe señalar que esta última función objetivo y KGE también generan valores altos de CRPSS y del índice  $\alpha$  en el modelo TUW.

La metodología ESP se desempeña mejor en las cuencas del centro de la zona de estudio, asociadas a una mayor acumulación de nieve, de respuesta más lenta (un alto índice de flujo base) y con mayor variabilidad inter-anual en sus volúmenes.

En relación a las cuencas DGA, los pronósticos de caudal estacional ESP presentan mejor desempeño que los pronósticos de la DGA en las cuencas Grande 1 y Maipo, mientras que en el resto de las cuencas los resultados son similares y comparables. El modelo GR6J es el que produce los resultados menos sesgados. Por otro lado, considerando que se utiliza simplemente la mediana del conjunto para emitir el pronóstico, los resultados pueden ser mejores si se emplean técnicas de pre- y post-procesamiento. Además, los modelos entregan otros flujos que pueden ser de gran interés, como la evaporación o el SWE.

Con respecto a la complejidad de los modelos, TUW cuenta con 15 parámetros libres y entrega mejores resultados. No obstante, dentro de la familia de modelos GR, que cuentan con seis a ocho parámetros libres, no hay claridad respecto a la relación complejidad/desempeño. Si bien, pareciera que el modelo GR6J es el que entrega mejores resultados, las diferencias son bajas o nulas en la práctica, especialmente al comparar el rendimiento de pronósticos generados con las mejores funciones objetivo. Por lo tanto, no hay suficiente información para asegurar que un modelo más complejo genera mejores resultados. De cualquier manera, se muestra que los modelos conceptuales, desde el más simple (GR4J, seis parámetros) hasta el más complejo (TUW) logran generar buenos pronósticos ESP, especialmente en la zona centro de la zona de estudio (latitudes comprendidas entre 30°S y 34°S). Por lo tanto, esta metodología tiene el potencial de ser utilizada operacionalmente en Chile a un bajo costo computacional.

## 6.2. Recomendaciones

Considerando que una métrica que busca minimizar los errores de los volúmenes estacionales da buenos resultados, se recomienda aplicar una métrica a resolución temporal diaria, como KGE, sólo en el periodo de deshielo, o ponderada para darle mayor peso a dicho pe-

riodo. Si se desea generar un pronóstico de caudales medios mensuales, se podrían aplicar funciones objetivo en dicha resolución temporal.

En el estudio se mostró que los modelos hidrológicos conceptuales tienen un gran potencial operacional para emitir pronósticos confiables. Sin embargo, no hay claridad sobre la influencia de la utilización de modelos más o menos complejos en la metodología ESP, o si las estructuras de modelos son las adecuadas para esta zona del mundo. Por lo tanto, se recomienda evaluar la sensibilidad de la metodología ESP utilizando diferentes estructuras de modelos, para cuantificar, en cierta medida, si se obtienen buenos resultados por las razones correctas.

En relación a ESP, el método estima la incertidumbre total mediante la utilización de los registros climáticos históricos, ignorando el resto de fuentes de incertidumbre. Se recomienda el uso de técnicas de asimilación de datos para generar un conjunto de IHCs, lo que permite cuantificar su incertidumbre y mejorar su confiabilidad. En la misma línea, los esquemas de pre- y post-procesamiento se encargan de abordar limitantes del método ESP como la incertidumbre no cuantificada, o sesgos sistemáticos que puedan existir. Por lo tanto, se recomienda experimentar con la selección o ponderación de ciertos conjuntos climáticos más probables mediante el uso de índices climáticos; o probar métodos de corrección del sesgo. Teniendo en cuenta que utilizar la mediana de todos los conjuntos puede generar un pronóstico sesgado, estos esquemas tienen un alto potencial de mejoramiento. Por otro lado, se podría modificar la metodología empleada en este estudio, y evaluar el valor agregado de la utilización de conjuntos de pronósticos climáticos (en vez de la climatología histórica), lo que permitiría, en cierta medida, trabajar con la enorme incertidumbre climática que existe al inicio del año hidrológico.

# Bibliografía

- Beckers, J. V., Weerts, A. H., Tijdeman, E., y Welles, E. (2016). ENSO-conditioned weather resampling method for seasonal ensemble streamflow prediction. *Hydrology and Earth System Sciences*, 20(8), 3277–3287.
- Bergström, S. (1976). *Development and application of a conceptual runoff model for scandinavian catchments*.
- Bohn, T. J., Sonessa, M. Y., y Lettenmaier, D. P. (2010). Seasonal hydrologic forecasting: Do multimodel ensemble averages always yield improvements in forecast skill? *Journal of Hydrometeorology*, 11(6), 1358–1372.
- Bradley, A., Demargne, J., y Franz, K. (2008). *How Do We Know Whether Seasonal Climate Forecasts are Any Good?* In: Troccoli A., Harrison M., Anderson D.L.T., Mason S.J. (Ed) *Seasonal Climate: Forecasting and Managing Risk*. NATO Science Series, Vol 82. Springer, Dordrecht.
- Bradley, A., Demargne, J., y Franz, K. (2019). *Attributes of Forecast Quality*. In: Duan Q., Pappenberger F., Wood A., Cloke H., Schaake J. (Ed) *Handbook of Hydrometeorological Ensemble Forecasting* (849-892). Springer, Berlin, Heidelberg.
- Burnash, R. J., Ferral, R. L., y McGuire, R. A. (1973). *A generalized streamflow simulation system: Conceptual modeling for digital computers*. Joint Federal and State River Forecast Center Tech. Rep., U.S. Department of Commerce, National Weather Service and State of California, Department of Water Resources.
- Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H. L., Koren, V., Duan, Q. Y., Ek, M., y Betts, A. (1996). Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *Journal of Geophysical Research Atmospheres*, 101(D3), 7251–7268.
- Coron, L., Thirel, G., Delaigue, O., Perrin, C., y Andréassian, V. (2017). The suite of lumped GR hydrological models in an R package. *Environmental Modelling and Software*, 94, 166–171.
- Day, G. N. (1985). Extended Streamflow Forecasting Using NWSRFS. *Journal of Water Resources Planning and Management*, 111(2), 157–170.
- DGA, comunicación personal. (2021). Solicitado a través de Ley de Transparencia el 15 de abril de 2021.
- Donegan, S., Murphy, C., Harrigan, S., Broderick, C., Golian, S., Knight, J., Matthews, T., Prudhomme, C., Quinn, D. F., Scaife, A. A., Stringer, N., y Wilby, R. L. (2020). Conditioning Ensemble Streamflow Prediction with the North Atlantic Oscillation improves skill at longer lead times. *[Preprint]*. *[Citado el 15 de mayo de 2021]*.

- Duan, Q., Sorooshian, S., y Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, 28(4), 1015–1031.
- Fowler, K., Peel, M., Western, A., y Zhang, L. (2018). Improved Rainfall-Runoff Calibration for Drying Climate: Choice of Objective Function. *Water Resources Research*, 54(5), 3392–3408.
- Franz, K. J., Hartmann, H. C., Sorooshian, S., y Bales, R. (2003). Verification of National Weather Service Ensemble Streamflow Predictions for water supply forecasting in the Colorado River Basin. *Journal of Hydrometeorology*, 4(6), 1105–1118.
- Garen, D. C., Abramovich, R., Julander, R., Kaiser, R., Lea, J., McClure, R., y Tama, R. (2011). Chapter 7. Water Supply Forecasting. *National Engineering Handbook*, 1–19.
- Girons Lopez, M., Crochemore, L., y G. Pechlivanidis, I. (2021). Benchmarking an operational hydrological model for providing seasonal forecasts in Sweden. *Hydrology and Earth System Sciences*, 25(3), 1189–1209.
- Gneiting, T., y Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378.
- Gobena, A. K., y Gan, T. Y. (2010). Incorporation of seasonal climate forecasts in the ensemble streamflow prediction system. *Journal of Hydrology*, 385(1-4), 336–352.
- Grantz, K., Rajagopalan, B., Clark, M., y Zagona, E. (2005). A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts. *Water Resources Research*, 41(10), 1–14.
- Gupta, H. V., Kling, H., Yilmaz, K. K., y Martinez, G. F. (2009). Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1), 80-91.
- Hamill, T. M. (2001). Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review*, 129(3), 550–560.
- Hamlet, A. F., y Lettenmaier, D. P. (1999). Columbia River Streamflow Forecasting Based on ENSO and PDO Climate Signals. *Journal of Water Resources Planning and Management*, 125(6), 333–351.
- Hargreaves, G. H., y Samani, Z. A. (1985). Reference crop evapotranspiration from temperature. *Applied engineering in agriculture*, 1(2), 96–99.
- Harrigan, S., Prudhomme, C., Parry, S., Smith, K., y Tanguy, M. (2018). Benchmarking ensemble streamflow prediction skill in the UK. *Hydrology and Earth System Sciences*, 22(3), 2023–2039.
- Hartigan, J. A., y Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), 100–108.
- Hashino, T., Bradley, A. A., y Schwartz, S. S. (2007). Evaluation of bias-correction methods for ensemble streamflow volume forecasts. *Hydrology and Earth System Sciences*, 11(2), 939–950.
- Jordan, A., Krüger, F., y Lerch, S. (2019). Evaluating probabilistic forecasts with scoring rules. *Journal of Statistical Software*, 90(2014).

- Kang, T. H., Kim, Y. O., y Hong, I. P. (2010). Comparison of pre- and post-processors for ensemble streamflow prediction. *Atmospheric Science Letters*, 11(2), 153–159.
- Kling, H., Fuchs, M., y Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, 424-425, 264-277.
- Lall, U., y Sharma, A. (1996). A Nearest Neighbor Bootstrap For Resampling Hydrologic Time Series. *Water Resources Research*, 32, 679-693.
- Liang, X., Lettenmaier, D. P., Wood, E. F., y Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*, 99(D7).
- Lindström, G., Pers, C., Rosberg, J., Strömqvist, J., y Arheimer, B. (2010). Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. *Hydrology Research*, 41(3-4), 295–319.
- Liu, Y., y Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research*, 43(7), 1–18.
- Mendoza, P. A., Rajagopalan, B., Clark, M. P., Cortés, G., y McPhee, J. (2014). A robust multimodel framework for ensemble seasonal hydroclimatic forecasts. *Water Resources Research*, 50(7), 6030–6052.
- Mendoza, P. A., Wood, A. W., Clark, E., Rothwell, E., Clark, M. P., Nijssen, B., Brekke, L. D., y Arnold, J. R. (2017). An intercomparison of approaches for improving operational seasonal streamflow forecasts. *Hydrology and Earth System Sciences*, 21(7), 3915–3935.
- Messner, J. W., Mayr, G. J., Zeileis, A., y Wilks, D. S. (2014). Heteroscedastic extended logistic regression for postprocessing of ensemble guidance. *Monthly Weather Review*, 142(1), 448–456.
- Muñoz-Castro, E., Hernandez, D., Mendoza, P., y Vargas, X. (2020). Comparación de métodos de ensemble forecasting aplicados al pronóstico de volúmenes de deshielo en Chile Central. *XXIV Congreso Chileno de Ingeniería Hidráulica*.
- Murphy, A. H. (1993). What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting. *Weather and Forecasting*, 8(2), 281-293.
- Nash, J., y Sutcliffe, J. (1970). River flow forecasting through conceptual models part i — a discussion of principles. *Journal of Hydrology*, 10(3), 282-290.
- Neri, M., Parajka, J., y Toth, E. (2020, 02). Importance of the information content in the study area when regionalising rainfall-runoff model parameters: the role of nested catchments and gauging station density. Preprint.
- Pagano, T., Garen, D., y Sorooshian, S. (2004). Evaluation of official western U.S. seasonal water supply outlooks, 1922-2002. *Journal of Hydrometeorology*, 5(5), 896–909.
- Parajka, J., Merz, R., y Blöschl, G. (2007). Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 austrian catchments. *Hydrological Processes*, 21(4), 435-446.
- Pechlivanidis, I. G., Crochemore, L., Rosberg, J., y Bosshard, T. (2020). What Are the Key Drivers Controlling the Quality of Seasonal Streamflow Forecasts? *Water Resources*

- Perrin, C., Michel, C., y Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279(1-4), 275–289.
- Pool, S., Vis, M., y Seibert, J. (2018). Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency. *Hydrological Sciences Journal*, 63(13-14), 1941-1953.
- Prudhomme, C., Hannaford, J., Harrigan, S., Boorman, D., Knight, J., Bell, V., Jackson, C., Svensson, C., Parry, S., Bachiller-Jareno, N., Davies, H., Davis, R., Mackay, J., McKenzie, A., Rudd, A., Smith, K., Bloomfield, J., Ward, R., y Jenkins, A. (2017). Hydrological Outlook UK: an operational streamflow and groundwater level forecasting system at monthly to seasonal time scales. *Hydrological Sciences Journal*, 62(16), 2753-2768.
- Pushpalatha, R., Perrin, C., Le Moine, N., Mathevet, T., y Andréassian, V. (2011). A downward structural sensitivity analysis of hydrological models to improve low-flow simulation. *Journal of Hydrology*, 411(1), 66-76.
- Quevedo Leiva, M. (2021). Metodología para el relleno de registros fluviométricos diarios en cuencas chilenas cercanas al régimen natural. Memoria para optar al Título de Ingeniero Civil. Departamento de Ingeniería Civil, Universidad de Chile.
- Renard, B., Kavetski, D., Kuczera, G., Thyer, M., y Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, 46(5).
- Renner, M., Werner, M. G., Rademacher, S., y Sprockereef, E. (2009). Verification of ensemble flow forecasts for the River Rhine. *Journal of Hydrology*, 376(3-4), 463–475.
- Rubio-Álvarez, E., y McPhee, J. (2010). Patterns of spatial and temporal variability in streamflow records in south central Chile in the period 1952-2003. *Water Resources Research*, 46(5), 1–16.
- Sabzipour, B., Arsenault, R., y Brissette, F. (2021). Evaluation of the potential of using subsets of historical climatological data for ensemble streamflow prediction (ESP) forecasting. *Journal of Hydrology*, 595, 125656.
- Schaake, J. (1978). The National Weather Service extended streamflow prediction techniques: Description and applications during 1977. En *3rd annual climate diagnostics workshop*.
- Schaake, J., Demargne, J., Hartman, R., Mullusky, M., Welles, E., Wu, L., Herr, H., Fan, X., y Seo, D. J. (2007). Precipitation and temperature ensemble forecasts from single-value forecasts. *Hydrology and Earth System Sciences Discussions*, 4, 655–717.
- Semenov, M. A. (2007). Development of high-resolution UKCIP02-based climate change scenarios in the UK. *Agricultural and Forest Meteorology*, 144(1), 127-138.
- Shi, X., Wood, A. W., y Lettenmaier, D. P. (2008). How essential is hydrologic model calibration to seasonal streamflow forecasting? *Journal of Hydrometeorology*, 9(6), 1350–1363.
- Šípek, V., y Daňhelka, J. (2015). Modification of input datasets for the Ensemble Streamflow Prediction based on large-scale climatic indices and weather generator. *Journal of*

*Hydrology*, 528, 720–733.

- Slater, L. J., y Villarini, G. (2018). Enhancing the Predictability of Seasonal Streamflow With a Statistical-Dynamical Approach. *Geophysical Research Letters*, 45(13), 6504–6513.
- Twedt, T. M., Schaake Jr, J. C., y Peck, E. L. (1977). National Weather Service extended streamflow prediction [USA]. En *Proceedings Western Snow Conference*.
- Valéry, A., Andréassian, V., y Perrin, C. (2014). ‘As simple as possible but not simpler’: What is useful in a temperature-based snow-accounting routine? Part 2 - Sensitivity analysis of the CemaNeige snow accounting routine on 380 catchments. *Journal of Hydrology*, 517, 1176–1187.
- Werner, K., Brandon, D., Clark, M., y Gangopadhyay, S. (2004). Climate index weighting schemes for NWS ESP-based seasonal volume forecasts. *Journal of Hydrometeorology*, 5(6), 1076–1090.
- Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). Academic press.
- Wilson, L. J., Burrows, W. R., y Lanzinger, A. (1999). A strategy for verification of weather element forecasts from an ensemble prediction system. *Monthly Weather Review*, 127(6), 956–970.
- Wood, A. W., y Lettenmaier, D. P. (2008). An ensemble approach for attribution of hydrologic prediction uncertainty. *Geophysical Research Letters*, 35(14), 1–5.
- Wood, A. W., Pagano, T., y Roos, M. (2016). Tracing The Origins of ESP. Recuperado el 14 de abril de 2021, de HEPEX Blog: <https://hepex.inrae.fr/tracing-the-origins-of-esp>.
- Wood, A. W., Sankarasubramanian, A., y Mendoza, P. (2018). Seasonal Ensemble Forecast Post-processing. *Handbook of Hydrometeorological Ensemble Forecasting*, 1–27.
- Wood, A. W., y Schaake, J. C. (2008). Correcting errors in streamflow forecast ensemble mean and spread. *Journal of Hydrometeorology*, 9(1), 132–148.
- Zalachori, I., Ramos, M.-H., Garçon, R., Mathevet, T., y Gailhard, J. (2012). Statistical processing of forecasts for hydrological ensemble prediction: a comparative study of different bias correction strategies. *Advances in Science and Research*, 8(1), 135–141.

# Anexos

## A.1. Complemento ESP periodo de septiembre a marzo

### A.1.1. Análisis por función objetivo de los pronósticos ESP

En las Figuras A.1, A.2, A.3 se presentan el CRPSS, el índice de confiabilidad  $\alpha$  y el  $R^2$  agrupados por función objetivo y tiempo de inicialización, para el volumen de septiembre a marzo.

Con respecto al CRPSS, se observa que el modelo TUV entrega mayores valores para todas las funciones objetivo y los tiempo de inicialización, salvo para EVDSep el 1° de mayo y 1° de junio. Sin embargo, para KGE, NSE y KGE+qKGE, el modelo TUV es significativamente mejor que el resto para todo tiempo de inicialización. Es interesante notar que, para el 1° de septiembre, la función EVDSep y el modelo TUV entregan los mejores CRPSS, mientras que para el resto de los tiempos de inicialización, KGE y el modelo TUV son superiores.

En relación al índice  $\alpha$ , el modelo TUV entrega mayores valores para todo tiempo de inicialización y todas las funciones objetivo. Además, para las funciones KGE, NSE, KGE+qKGE y KGE+logNSE, la superioridad del modelo TUV es estadísticamente significativa. Asimismo, y al igual como ocurre en CRPSS, la función EVDSep y el modelo TUV entregan los valores más altos de confiabilidad para el 1° de septiembre, mientras que para el resto de los tiempos de inicialización, KGE y el modelo TUV es la mejor combinación.

Sobre el  $R^2$ , se observa que ninguna diferencia es estadísticamente significativa, aunque el modelo TUV es el que, generalmente, entrega los valores más altos. Sin embargo, la función EVDSep y el modelo GR4J entregan los valores de  $R^2$  más altos para los tiempos de inicialización 1° de junio, 1° de julio y 1° de septiembre. Para el resto de los tiempos de inicialización, la combinación de KGE con el modelo TUV entregan los valores más altos.

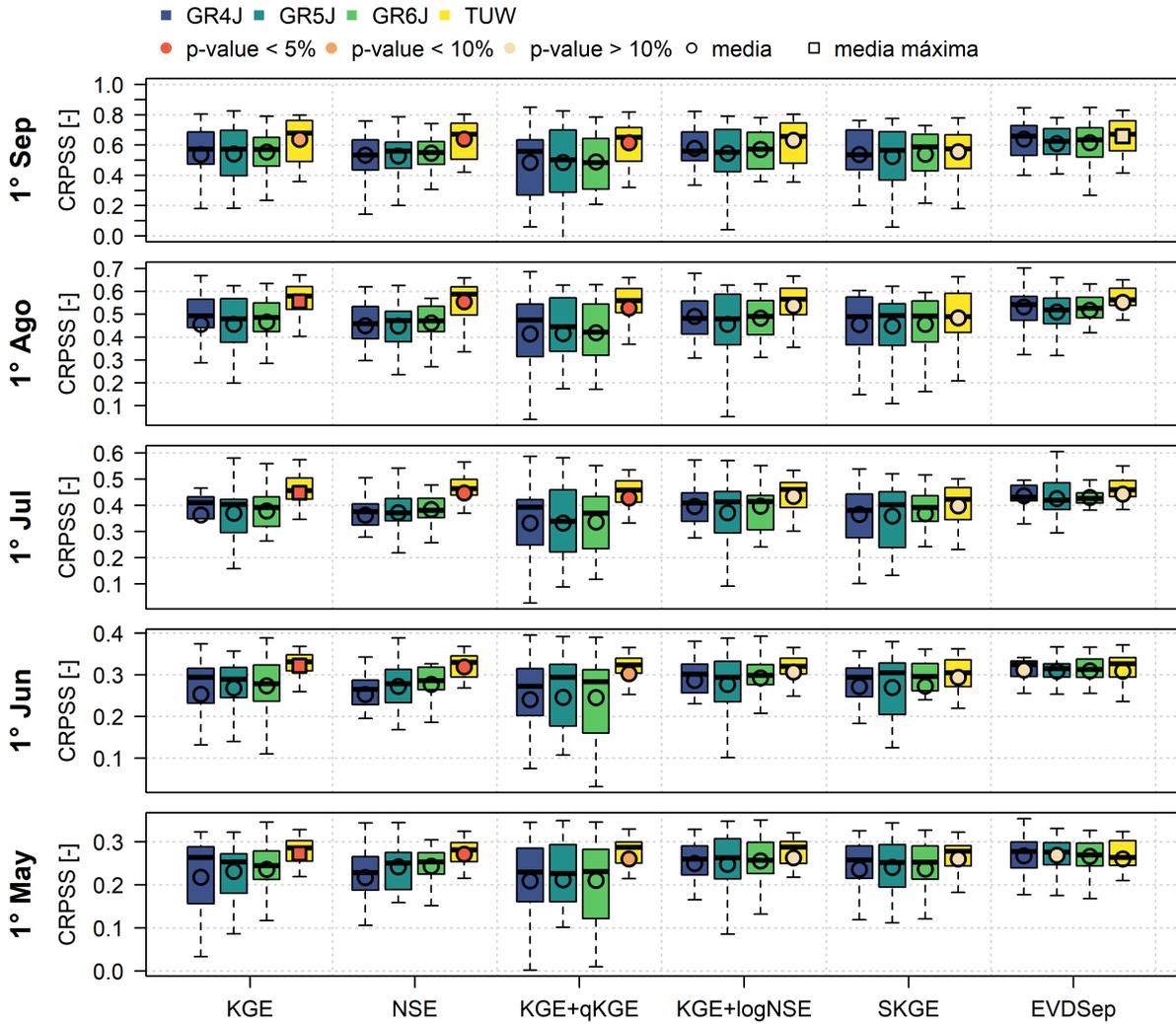


Figura A.1: Comparación de los valores de CRPSS de los pronósticos ESP, obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y función objetivo, para el periodo de septiembre a marzo. Cada *boxplot* se compone del CRPSS de las 22 cuencas en estudio. El *p-value* indica la probabilidad de que la media del mejor conjunto sea significativamente similar a la del segundo mejor. Vale decir, un bajo *p-value* indica que las medias son significativamente diferentes. El círculo coloreado indica cual función objetivo de cada modelo entrega el mejor CRPSS, mientras que el cuadrado indica cual es el modelo con la media más alta de entre todos. Notar que los paneles no se encuentran en la misma escala.

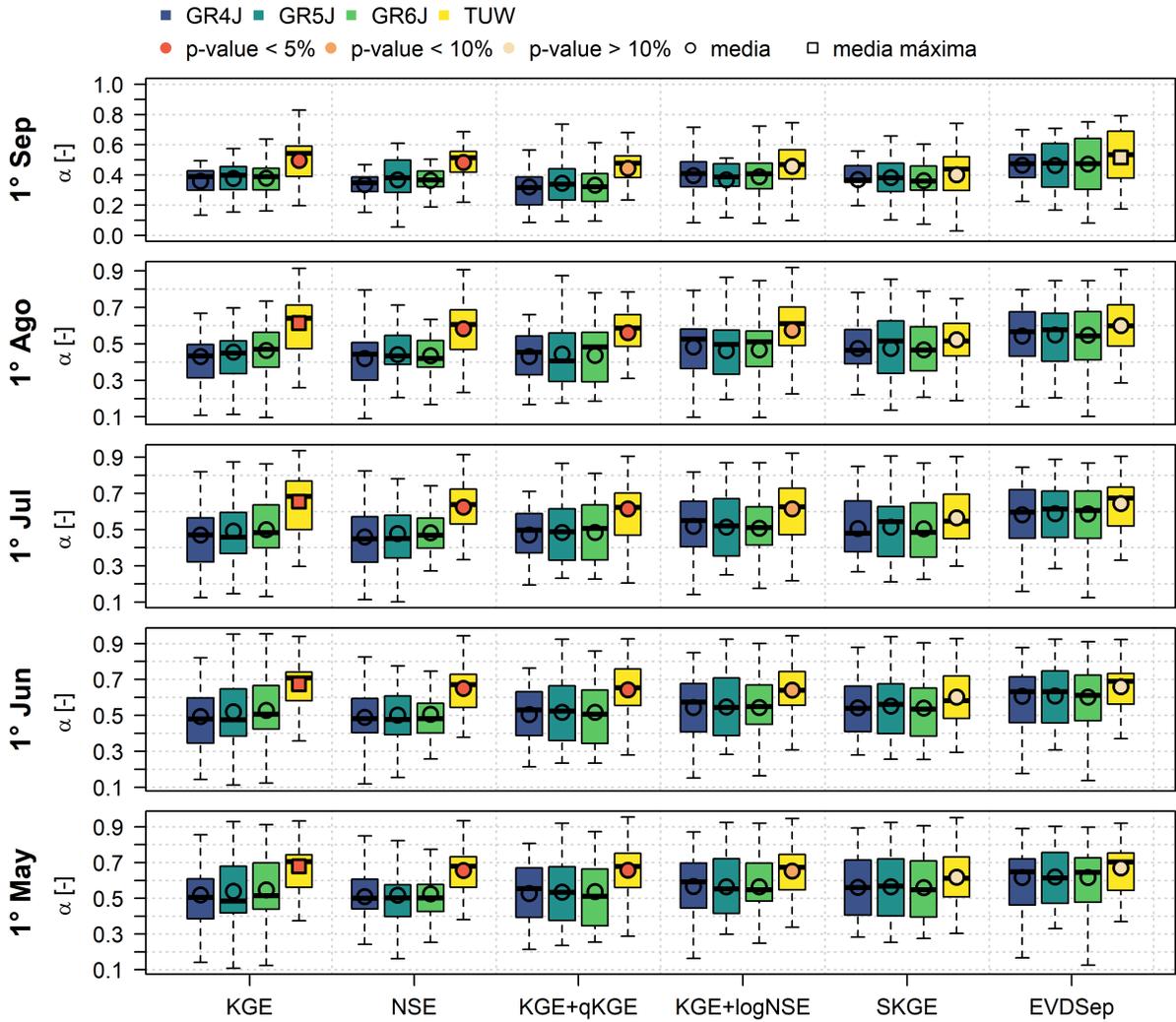


Figura A.2: Comparación de los valores del índice de confiabilidad  $\alpha$  de los pronósticos ESP, obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y función objetivo, para el periodo de septiembre a marzo. Cada *boxplot* se compone del CRPSS de las 22 cuencas en estudio. El *p-value* indica la probabilidad de que la media del mejor conjunto sea significativamente similar a la del segundo mejor. Vale decir, un bajo *p-value* indica que las medias son significativamente diferentes. El círculo coloreado indica cual función objetivo de cada modelo entrega el mejor CRPSS, mientras que el cuadrado indica cual es el modelo con la media más alta de entre todos. Notar que los paneles no se encuentran en la misma escala.

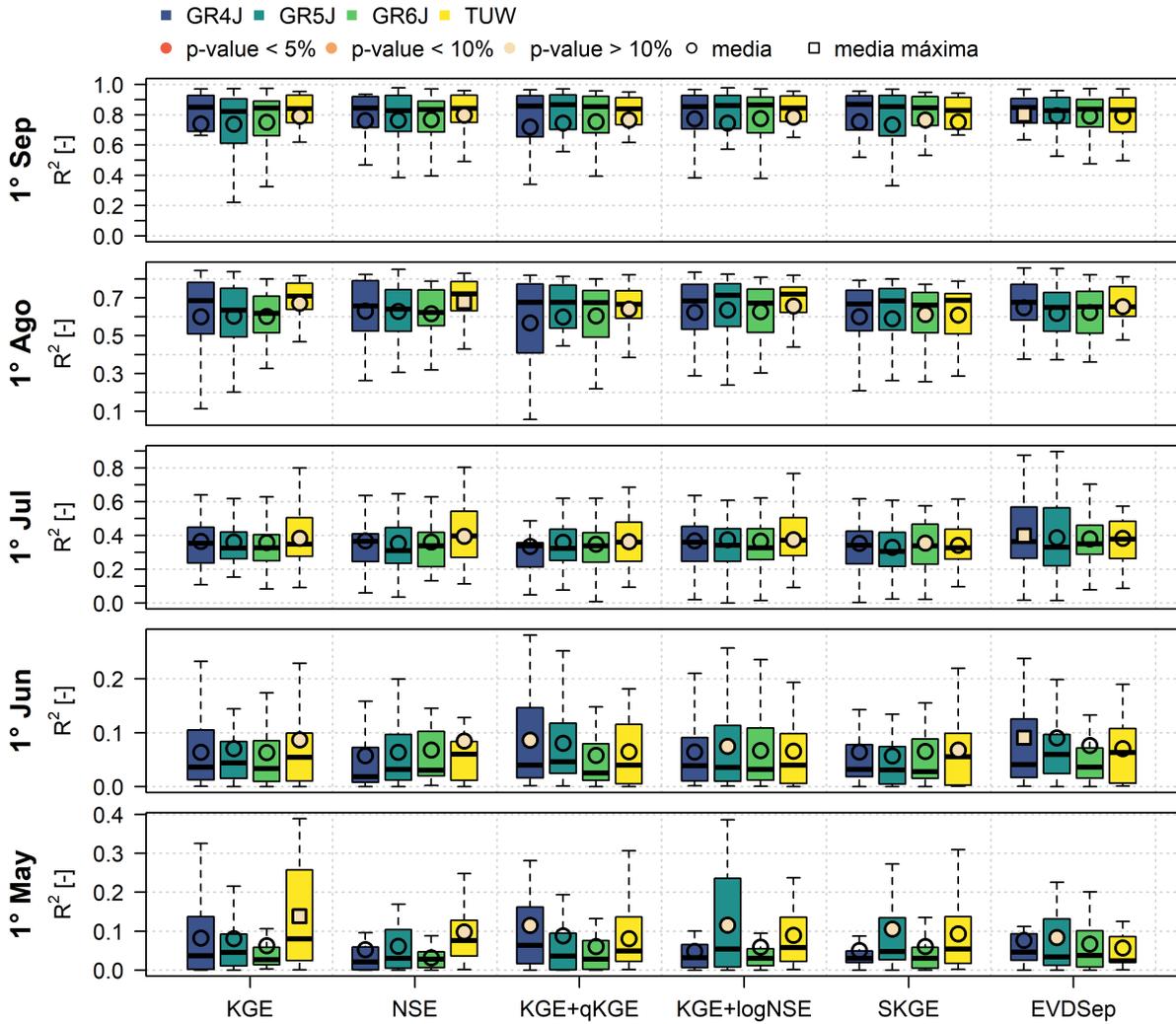


Figura A.3: Comparación de los valores de  $R^2$  de los pronósticos ESP, obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y función objetivo, para el periodo de septiembre a marzo. Cada *boxplot* se compone del CRPSS de las 22 cuencas en estudio. El *p-value* indica la probabilidad de que la media del mejor conjunto sea significativamente similar a la del segundo mejor. Vale decir, un bajo *p-value* indica que las medias son significativamente diferentes. El círculo coloreado indica cual función objetivo de cada modelo entrega el mejor CRPSS, mientras que el cuadrado indica cual es el modelo con la media más alta de entre todos. Notar que los paneles no se encuentran en la misma escala.

### A.1.2. Mejores índices de confiabilidad $\alpha$ y $R^2$

En la Figura A.4 se muestran los mejores índices  $\alpha$  de los pronósticos ESP y su distribución espacial, por modelo y tiempo de inicialización, para el periodo de septiembre a marzo. Se aprecia que el índice disminuye su valor hacia el 1° de septiembre. Los resultados indican que los modelos GR se desempeñan de manera similar, mientras que el modelo TUV se presenta como el mejor modelo. En junio, TUV es el que entrega más cuencas (3) con un índice  $\alpha$

$\geq 0,9$ ; seguido por el modelo GR6J con dos cuencas. En julio y agosto, TUV sigue siendo superior, con dos cuencas con un índice  $\alpha \geq 0,9$  y la mayor cantidad de cuencas con un índice  $\alpha \geq 0,7$ . En septiembre, todos los modelos presentan resultados similares.

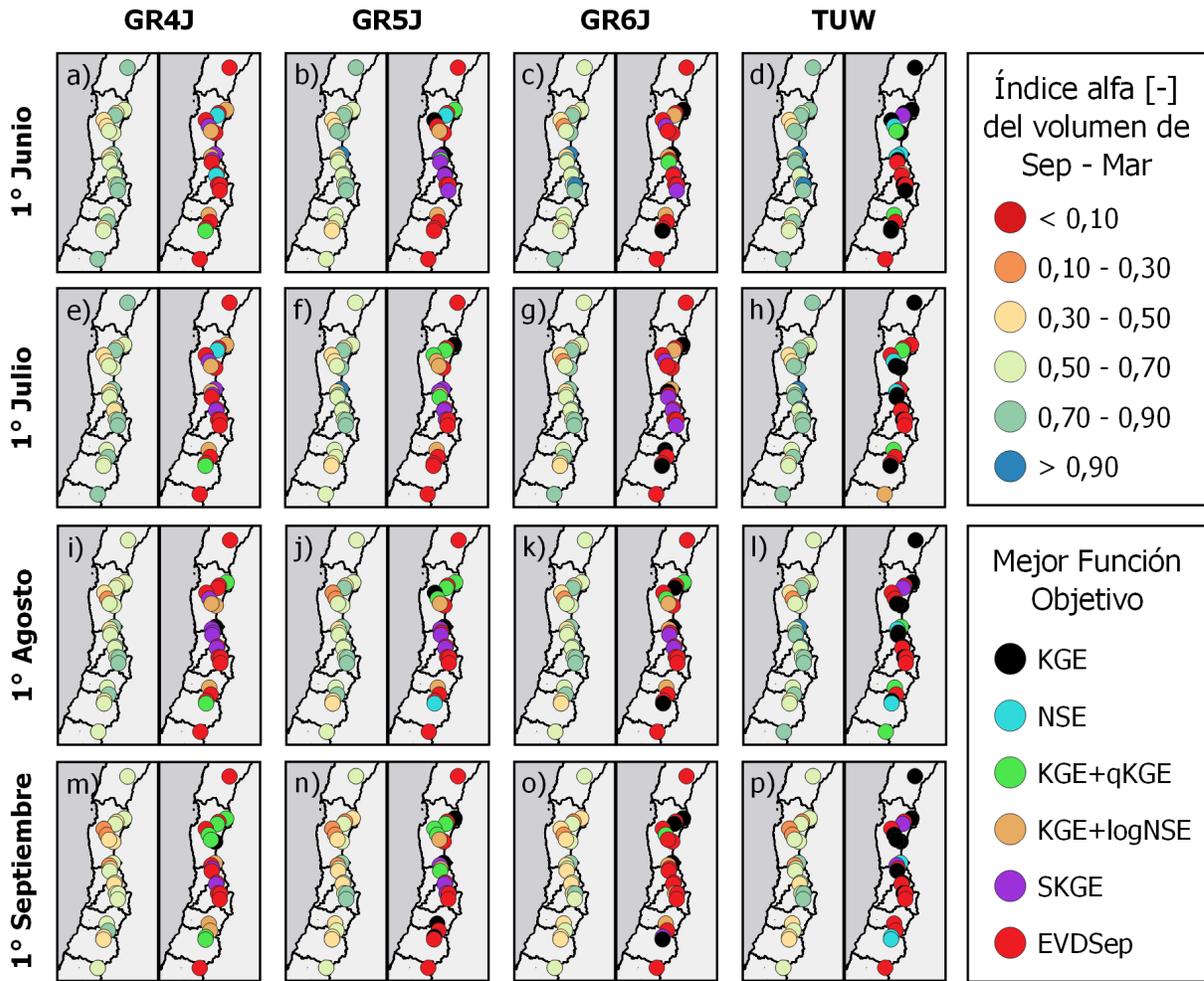


Figura A.4: Distribución espacial de los mejores índices de confiabilidad  $\alpha$  del periodo de septiembre a marzo (paneles de la izquierda), junto a las funciones objetivo que los generan (paneles de la derecha), para todas las cuencas, los modelos, y cuatro tiempos de inicialización.

Con respecto a las funciones objetivo, EVDSep es la que genera los valores de confiabilidad más altos, especialmente en los modelos GR. En el modelo TUV, las funciones se encuentran más distribuidas, aunque predominan EVDSep, KGE y NSE.

En la Figura A.5 se muestran los mejores índices  $R^2$  de los pronósticos ESP y su distribución espacial, por modelo y tiempo de inicialización, para el periodo de septiembre a marzo. Se aprecia que la métrica aumenta su valor hacia el 1° de septiembre. Los resultados indican que los modelos GR entregan valores de  $R^2$  extremadamente similares. La única diferencia apreciable se encuentra en las cuencas del sur, donde el modelo TUV entrega valores de  $R^2$  ligeramente mayores que los modelos GR.

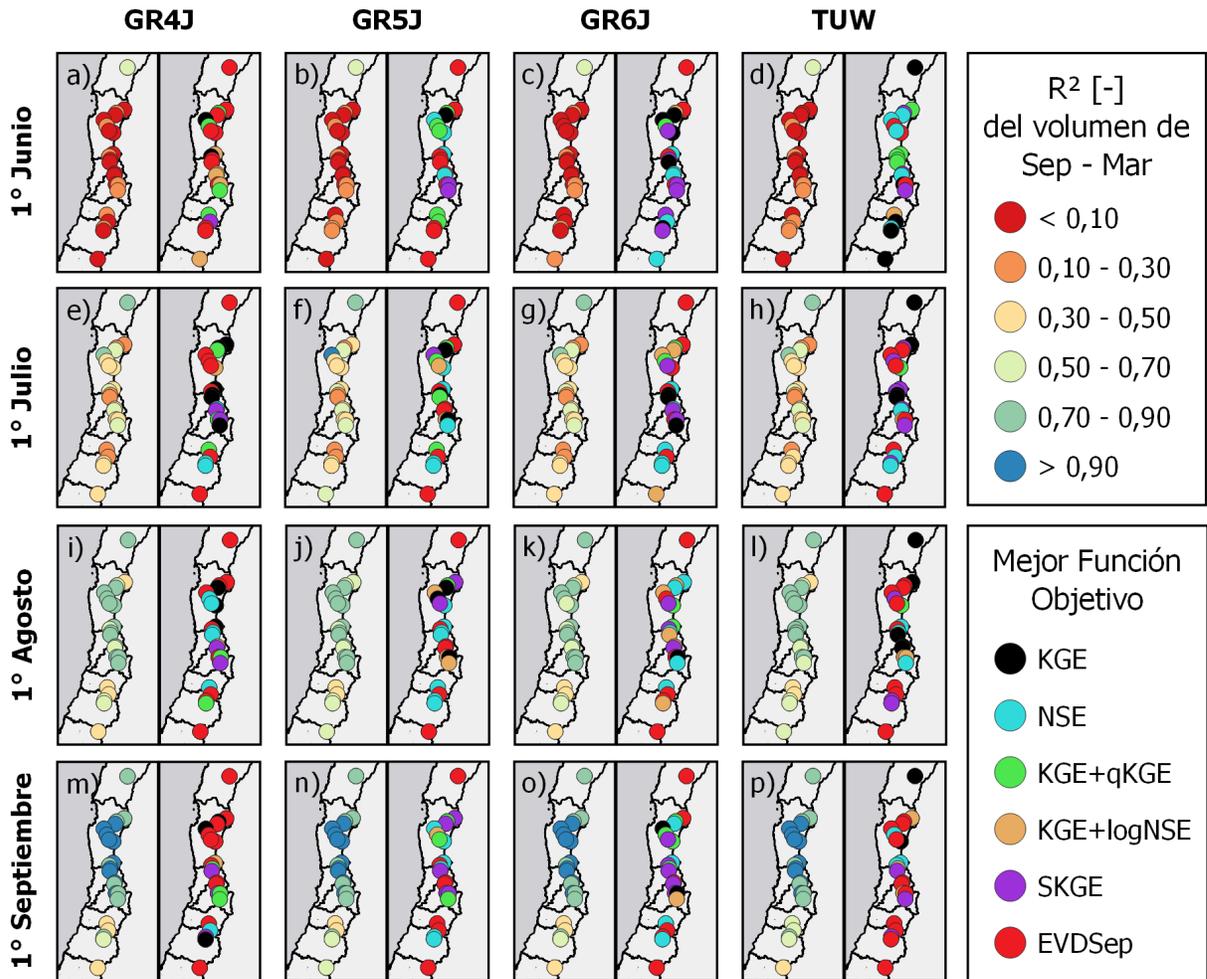


Figura A.5: Distribución espacial de los mejores  $R^2$  del periodo de septiembre a marzo (paneles de la izquierda), junto a las funciones objetivo que los generan (paneles de la derecha), para todas las cuencas, los modelos, y cuatro tiempos de inicialización.

Con respecto a las funciones objetivo, no se observan patrones claros para determinar alguna que se destaque por sobre las demás. De todas formas, esta métrica presenta un comportamiento estable para todos los modelos y las funciones objetivos, por lo que es esperable que ninguna FO se destaque por sobre otra.

### A.1.3. Análisis en detalle de los pronósticos ESP

En las Figuras A.6 a A.14 se presenta el detalle de las métricas CRPSS, índice  $\alpha$  y  $R^2$  para las 22 cuencas en estudio, según modelo, tiempo de inicialización y función objetivo, para el volumen de septiembre a marzo.

Se presentan, además, las bandas asociadas al 95% de confianza para la mejor función objetivo de cada modelo. La línea discontinua muestra el valor máximo de cada métrica obtenida para el mejor modelo el 1° de septiembre, con el fin de facilitar la comparación entre las funciones óptimas de cada modelo, y determinar la importancia relativa de la elección de la FO y/o el modelo.

Con respecto al CRPSS, en general, se observa que la correcta elección de la función objetivo en los modelos consigue generar un CRPSS al nivel del CRPSS generado por el mejor modelo (siendo TUW el mejor en la mayoría de los casos). Existen algunas excepciones como en la cuenca Grande 1, Grande 2, Pedernal y Sobrante, en que la elección del modelo tiene mayor relevancia.

En relación al índice  $\alpha$ , la elección del modelo adquiere mayor relevancia, especialmente al considerar que TUW es el que entrega mayores valores de confiabilidad al inicializar el 1° de septiembre. No obstante, específicamente en los modelos GR, sigue siendo altamente relevante escoger una correcta FO.

Sobre el  $R^2$ , todos las funciones objetivo y los modelos se comportan similar. No obstante, en algunas cuencas tiene gran relevancia la elección de la FO (e.g. Pulido, Pedernal, Claro 1, Claro 2 y Sauces). Siguiendo esta línea, y tal como ocurre en CRPSS, la correcta elección de la función objetivo en los modelos consigue generar un  $R^2$  al nivel del  $R^2$  generado por el mejor modelo.

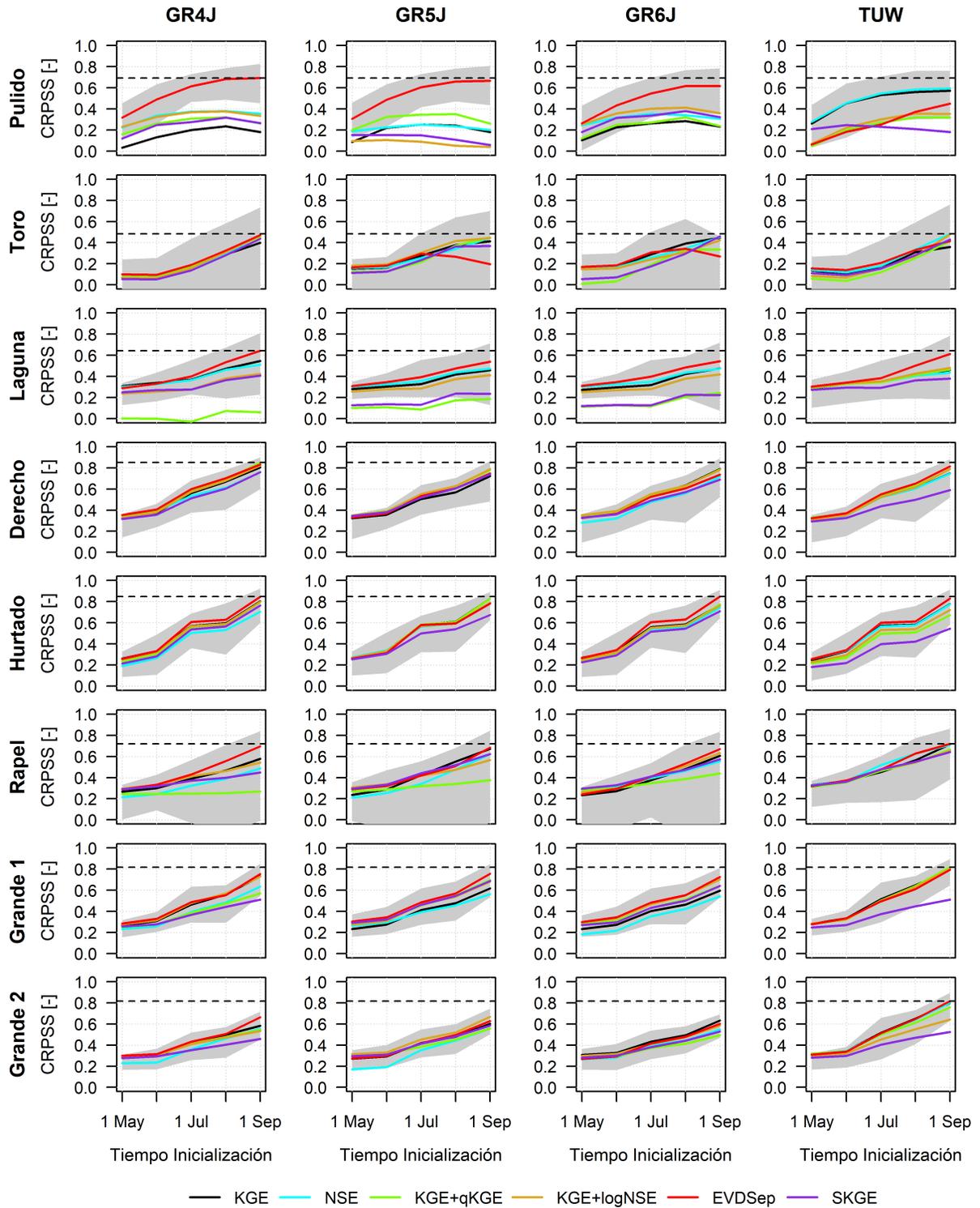


Figura A.6: Comparación de los valores de CRPSS de las cuencas 1 a 8, por función objetivo, tiempo de inicialización y modelo, para el periodo de septiembre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping* con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de septiembre.

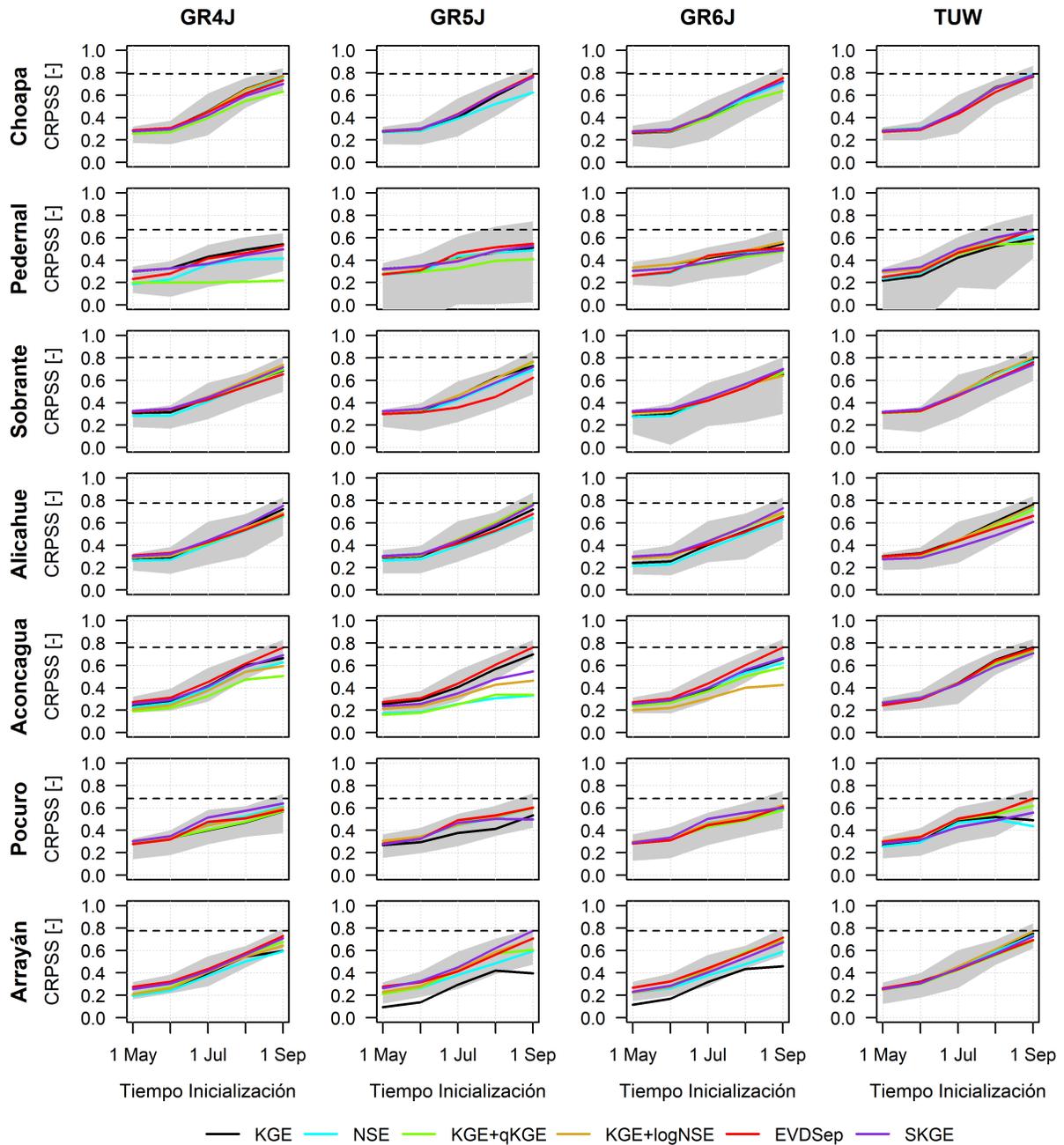


Figura A.7: Comparación de los valores de CRPSS de las cuencas 9 a 15, por función objetivo, tiempo de inicialización y modelo, para el periodo de septiembre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping* con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de septiembre.

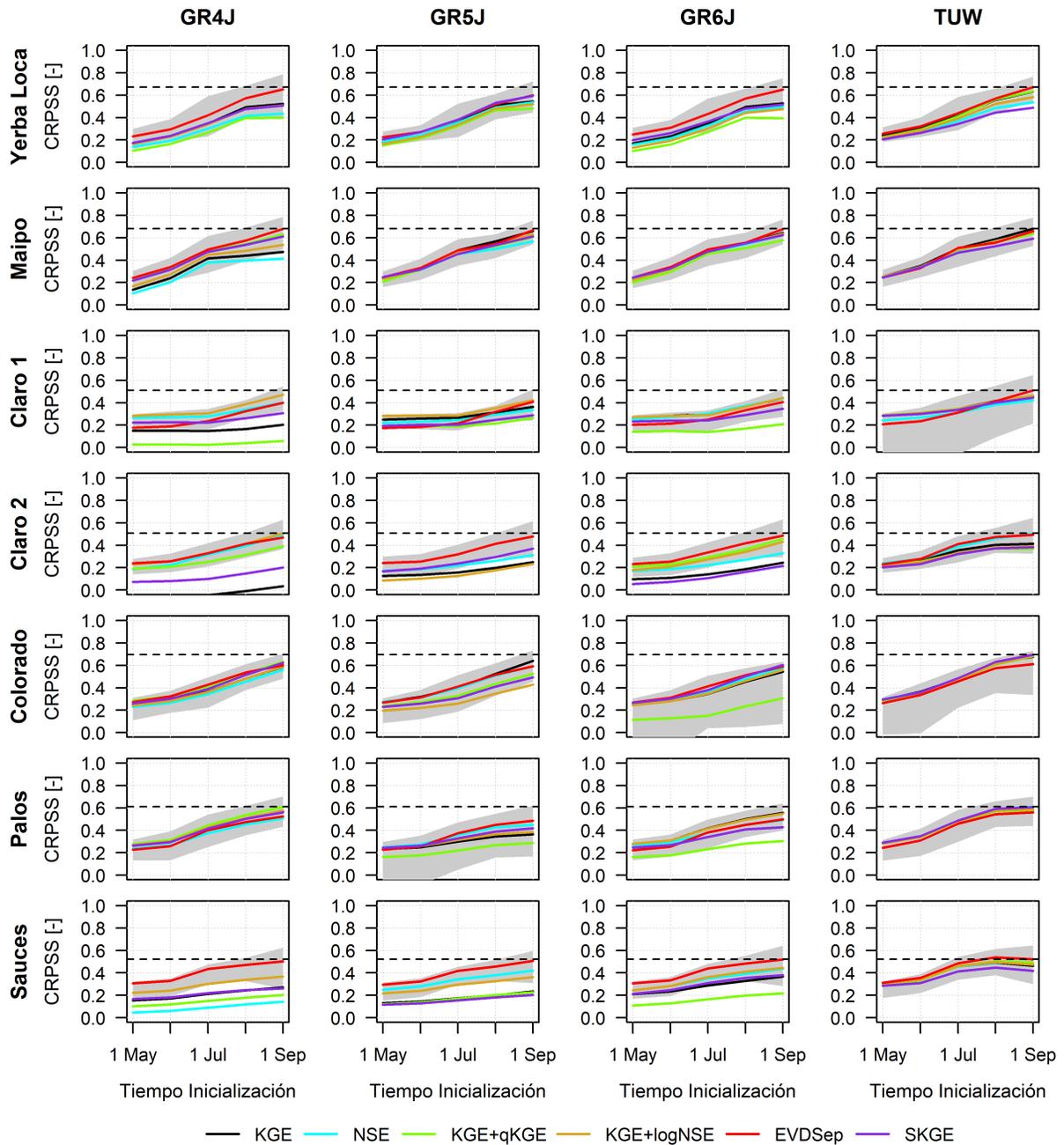


Figura A.8: Comparación de los valores de CRPSS de las cuencas 16 a 22, por función objetivo, tiempo de inicialización y modelo, para el periodo de septiembre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping* con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de septiembre.

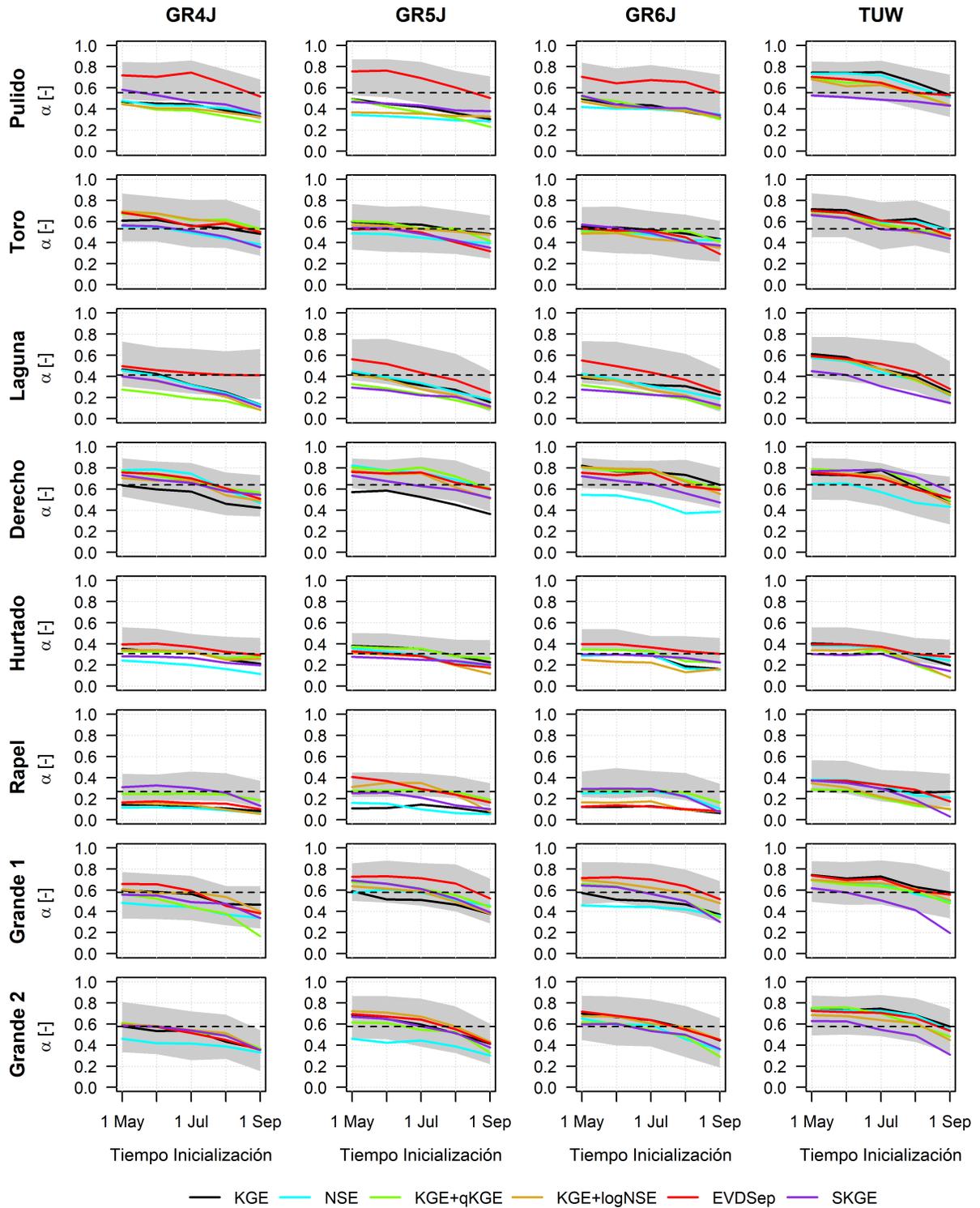


Figura A.9: Comparación de los valores del índice de confiabilidad  $\alpha$  de las cuencas 1 a 8, por función objetivo, tiempo de inicialización y modelo, para el periodo de septiembre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping* con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de septiembre.

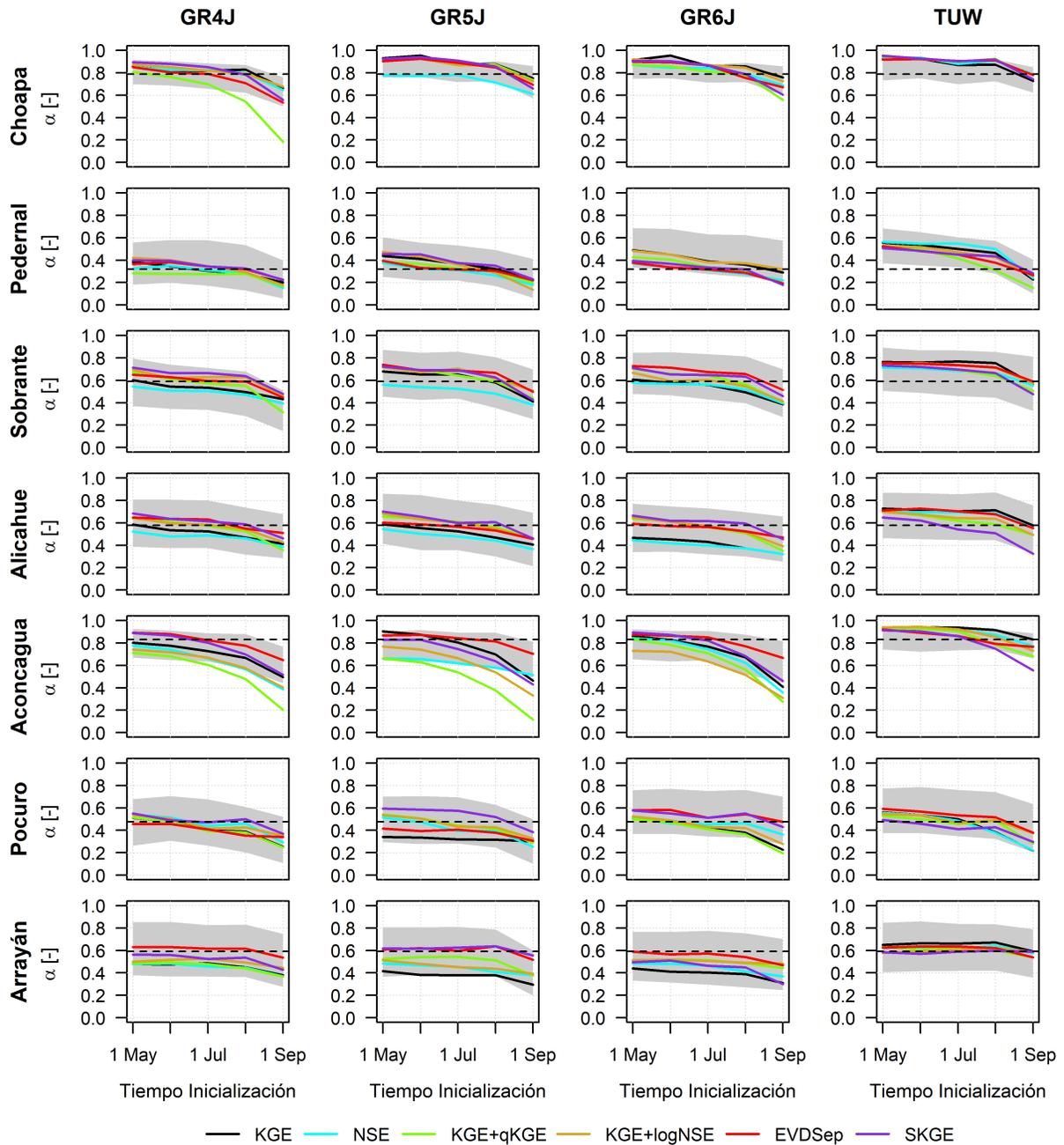


Figura A.10: Comparación de los valores del índice de confiabilidad  $\alpha$  de las cuencas 9 a 15, por función objetivo, tiempo de inicialización y modelo, para el periodo de septiembre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping* con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de septiembre.

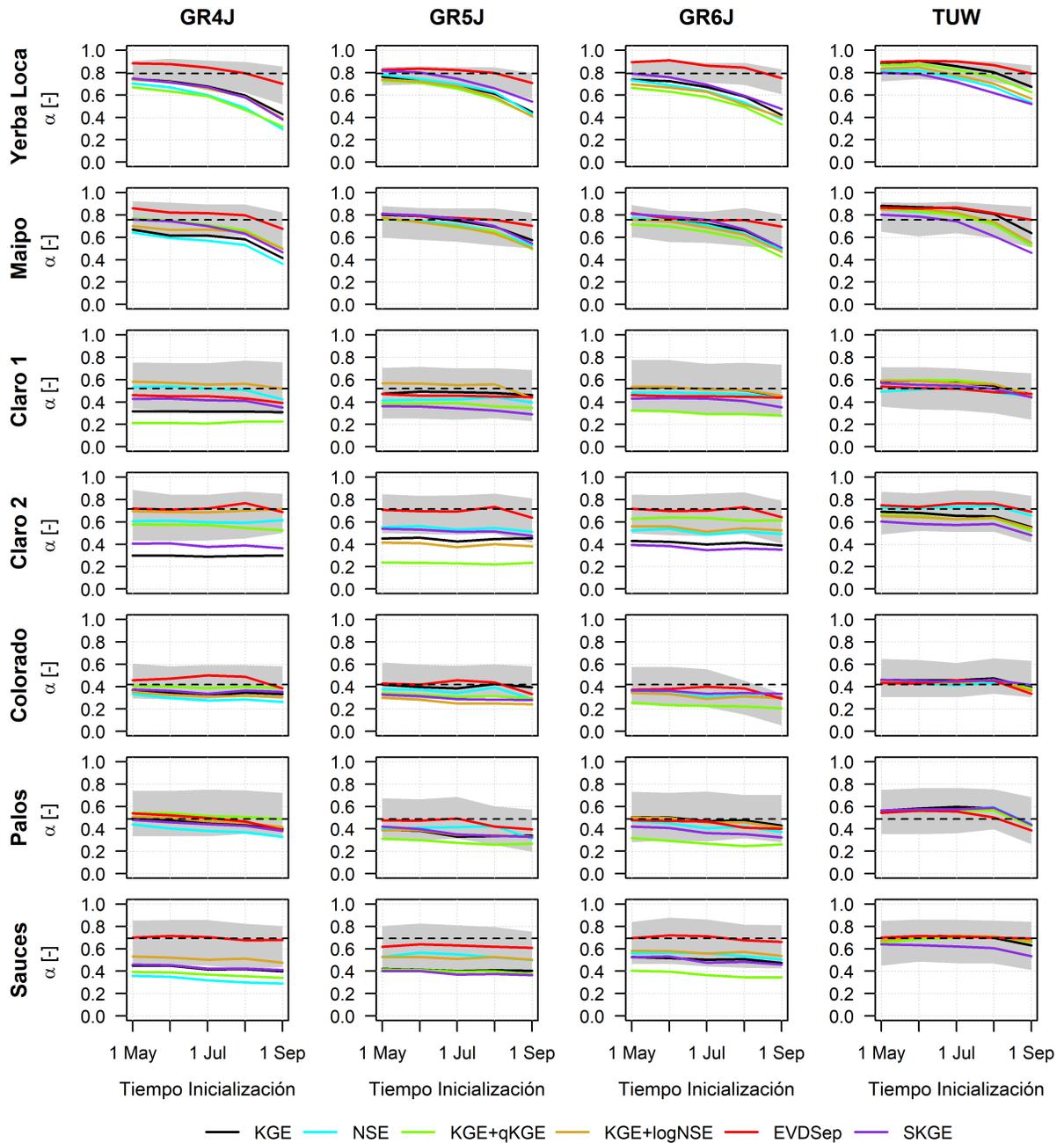


Figura A.11: Comparación de los valores del índice de confiabilidad  $\alpha$  de las cuencas 16 a 22, por función objetivo, tiempo de inicialización y modelo, para el periodo de septiembre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping* con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de septiembre.

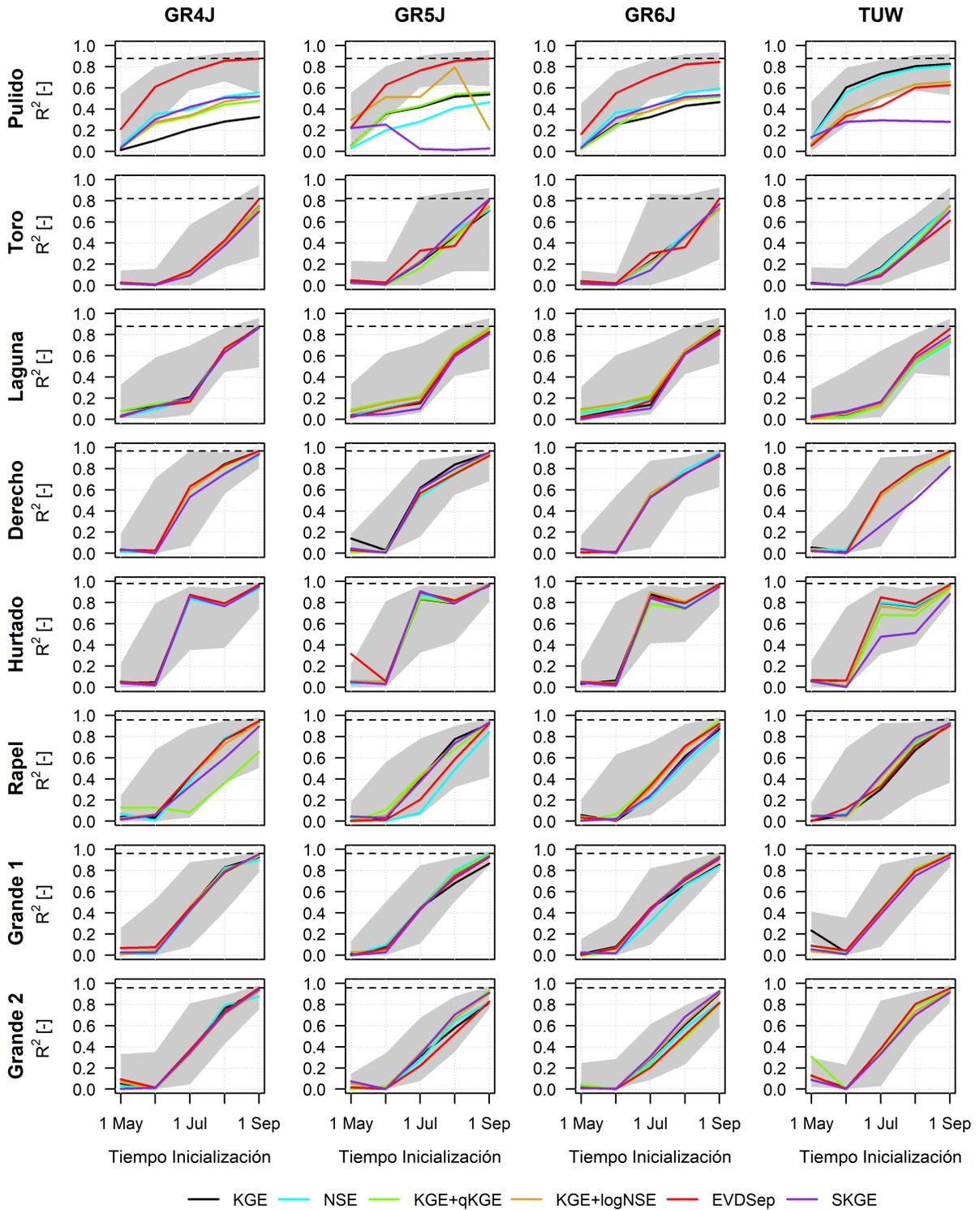


Figura A.12: Comparación de los valores de  $R^2$  de las cuencas 1 a 8, por función objetivo, tiempo de inicialización y modelo, para el periodo de septiembre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping* con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de septiembre.

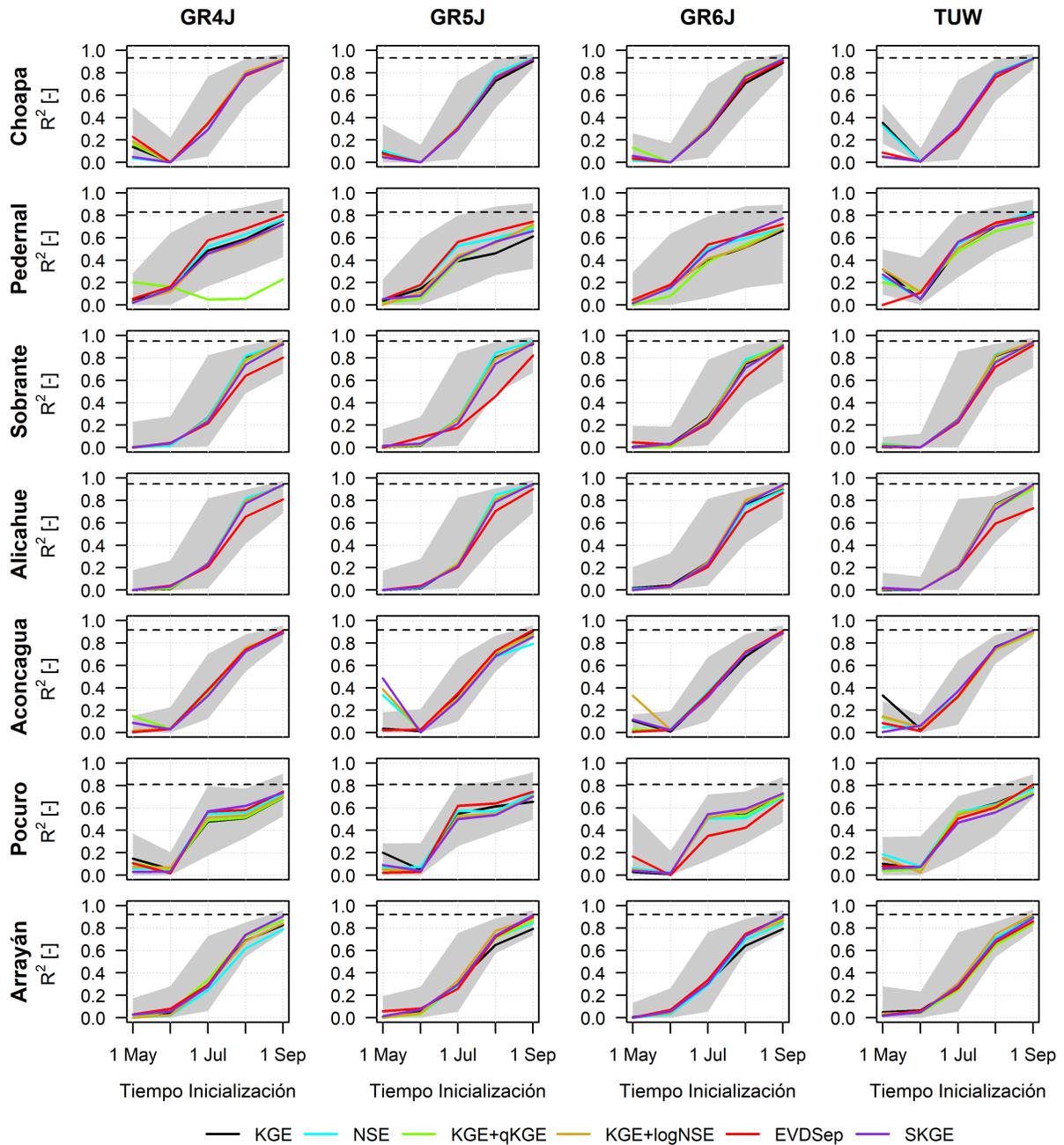


Figura A.13: Comparación de los valores de  $R^2$  de las cuencas 9 a 15, por función objetivo, tiempo de inicialización y modelo, para el periodo de septiembre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping* con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de septiembre.

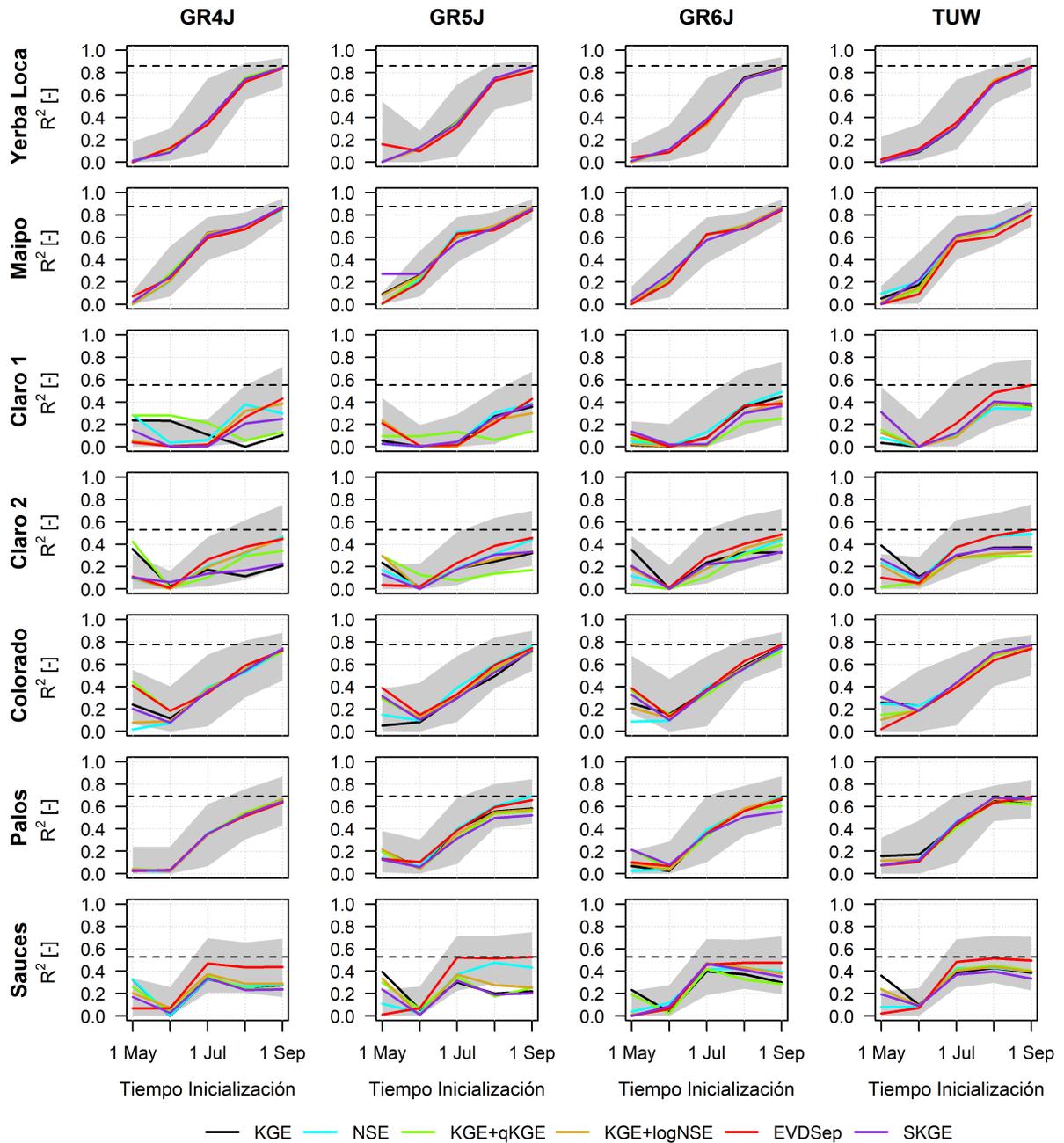


Figura A.14: Comparación de los valores de  $R^2$  de las cuencas 16 a 22, por función objetivo, tiempo de inicialización y modelo, para el periodo de septiembre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping* con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de septiembre.

## A.2. Complemento ESP periodo de octubre a marzo

### A.2.1. Análisis por función objetivo de los pronósticos ESP

En las Figuras A.15, A.16, A.17 se presentan el CRPSS, el índice de confiabilidad  $\alpha$  y el  $R^2$  agrupados por función objetivo y tiempo de inicialización, para el volumen de octubre a marzo.

Con respecto al CRPSS, se observa que el modelo TUW entrega mayores valores para todas las funciones objetivo y los tiempo de inicialización, salvo para EVDSep el 1° de mayo. Además, esta superioridad es estadísticamente significativa para la mayoría de los tiempos de inicialización y las FO. Es interesante notar que, para los tiempos de inicialización 1° de junio al 1° de octubre, la función EVDOct y el modelo TUW entregan los mejores CRPSS, mientras que para el 1° de mayo, el modelo GR5J y la función EVDSep es la mejor combinación.

En relación al índice  $\alpha$ , el modelo TUW entrega mayores valores para todo tiempo de inicialización y todas las funciones objetivo. Además, para las funciones KGE, NSE, KGE+qKGE y KGE+logNSE, la superioridad del modelo TUW es estadísticamente significativa. Asimismo, la función EVDOct y el modelo TUW entregan los mejores valores de confiabilidad para el 1° de septiembre y el 1° de octubre, mientras que para el resto de los tiempos de inicialización, KGE y el modelo TUW es la combinación que entrega los valores más altos.

Sobre el  $R^2$ , se observa que ninguna diferencia es estadísticamente significativa, aunque el modelo TUW es el que, generalmente, entrega los valores más altos. Además, la función EVDOct y el modelo TUW entregan los valores de  $R^2$  más altos para los tiempos de inicialización 1° de septiembre, 1° de octubre. Para el resto de los tiempos de inicialización, existen diferentes combinaciones de FO y modelos que entregan los mayores valores.

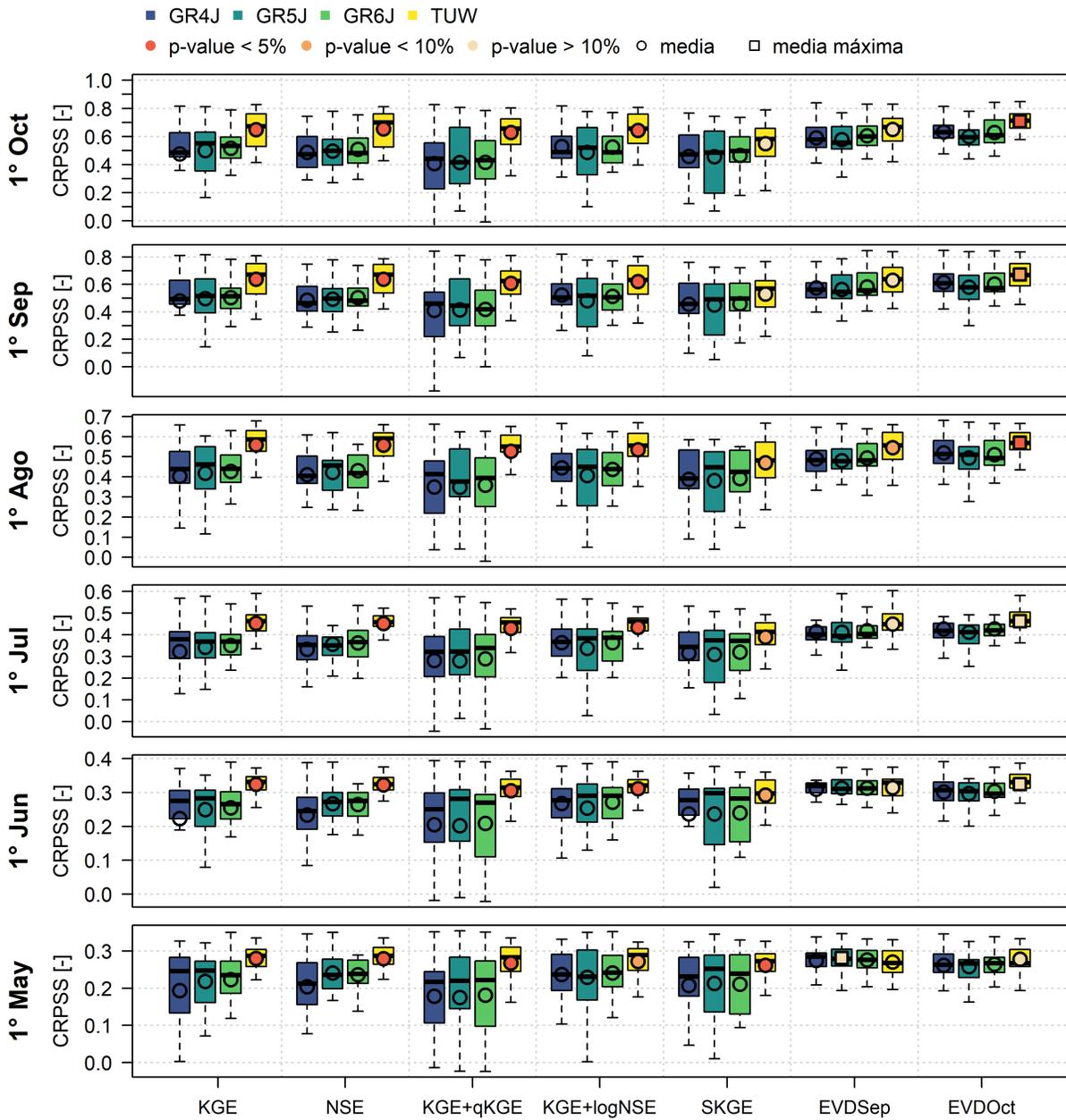


Figura A.15: Comparación de los valores de CRPSS de los pronósticos (Oct - Mar) obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y función objetivo.

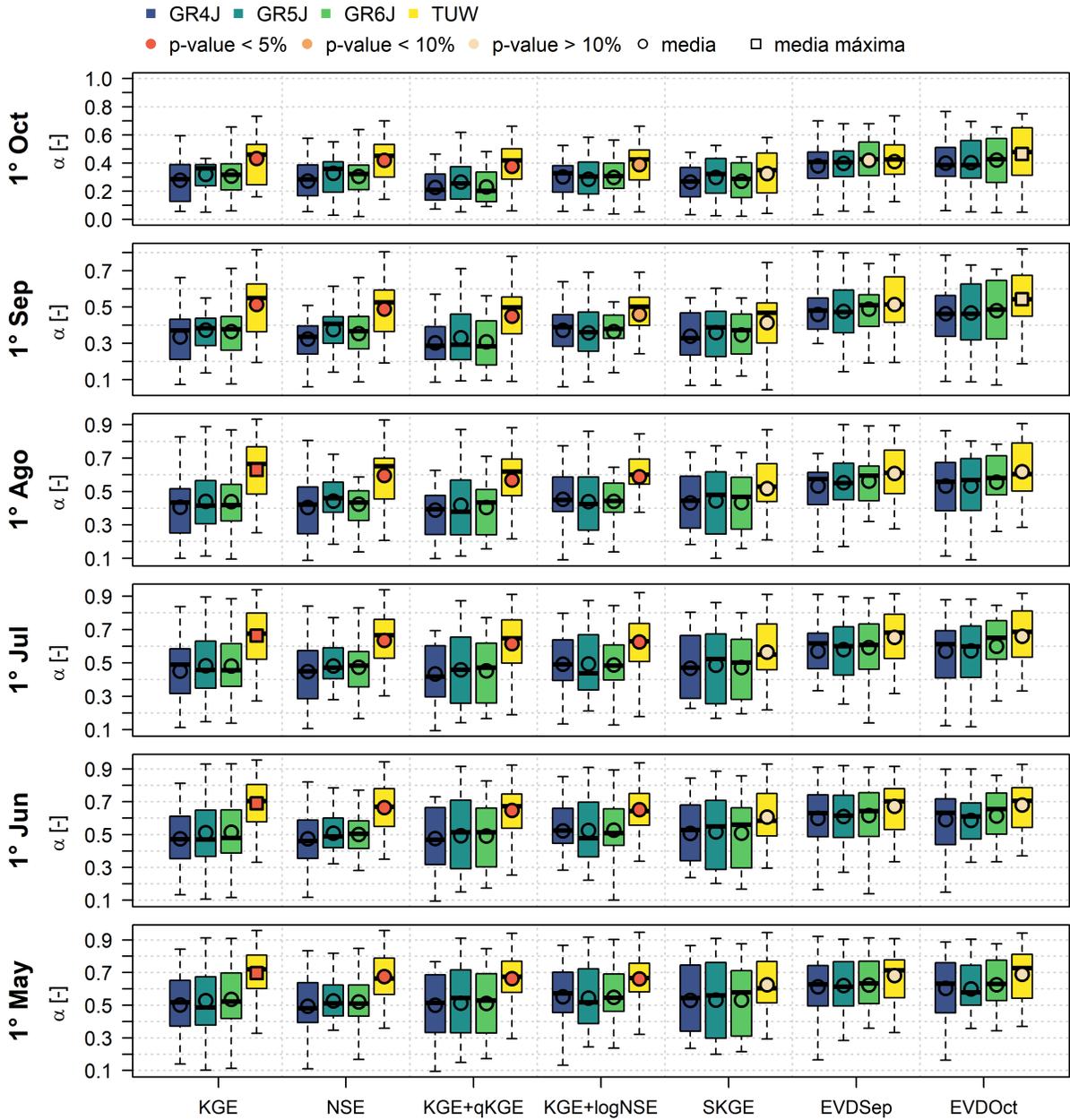


Figura A.16: Comparación de los valores del índice  $\alpha$  de los pronósticos (Oct - Mar) obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y función objetivo.

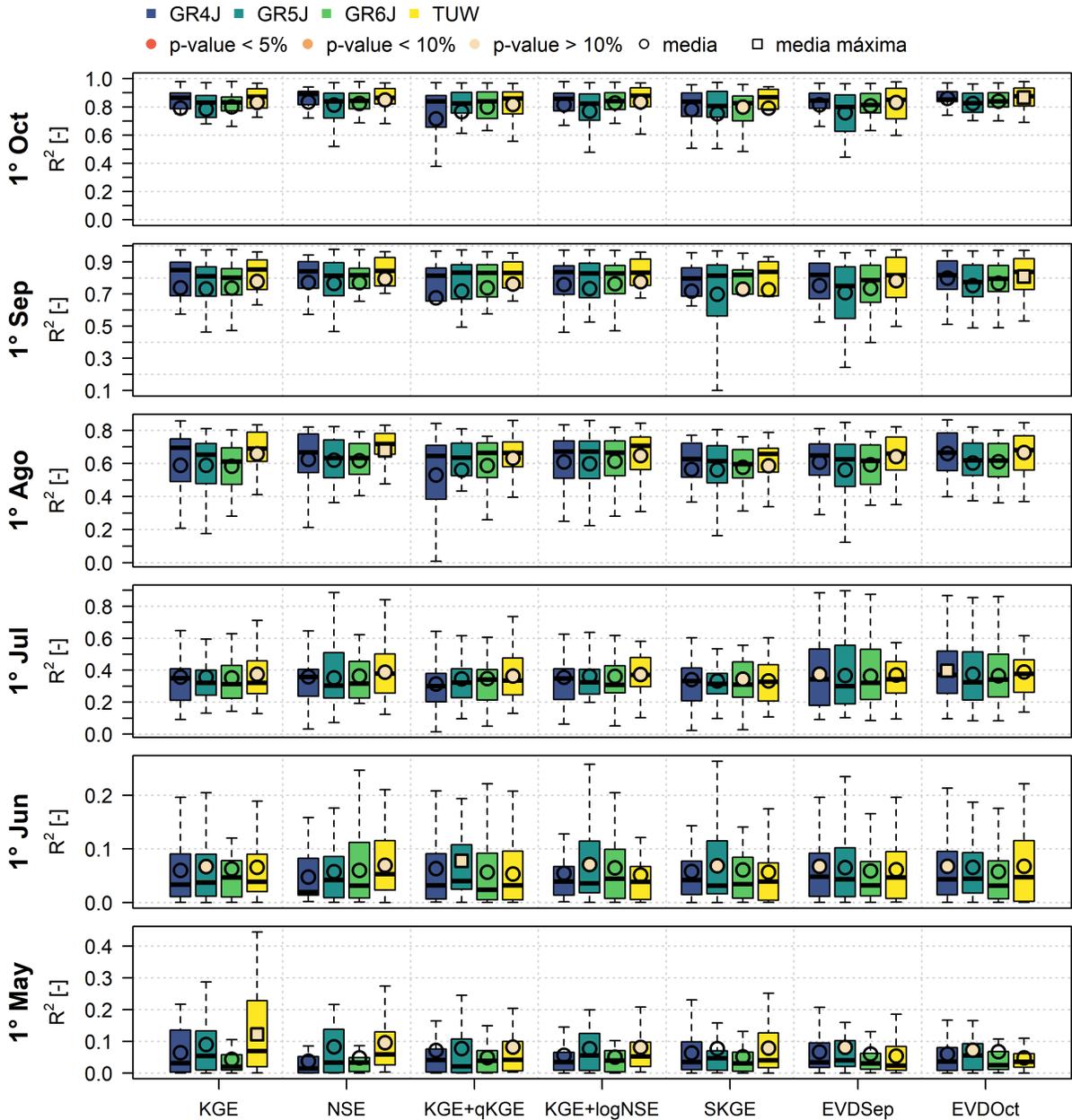


Figura A.17: Comparación de los valores de  $R^2$  de los pronósticos (Oct - Mar) obtenidos con distintas funciones objetivo, agrupados por tiempo de inicialización y función objetivo.

### A.2.2. Mejores índices de confiabilidad $\alpha$ y $R^2$

En la Figura A.18 se presentan los mejores índices  $\alpha$  de los pronósticos ESP y su distribución espacial, por modelo y tiempo de inicialización, para el periodo de octubre a marzo. Se aprecia que el índice disminuye su valor hacia el 1° de octubre. Los resultados muestran que los modelos GR se desempeñan de manera similar, mientras que el modelo TUW se presenta como el mejor modelo. En julio y agosto, TUW entrega dos cuencas con un índice  $\alpha \geq 0,9$ ,

mientras que los modelos GR solo entregan una. En septiembre y octubre, TUV sigue siendo superior, con varias cuencas con un índice  $\alpha \geq 0,7$ .

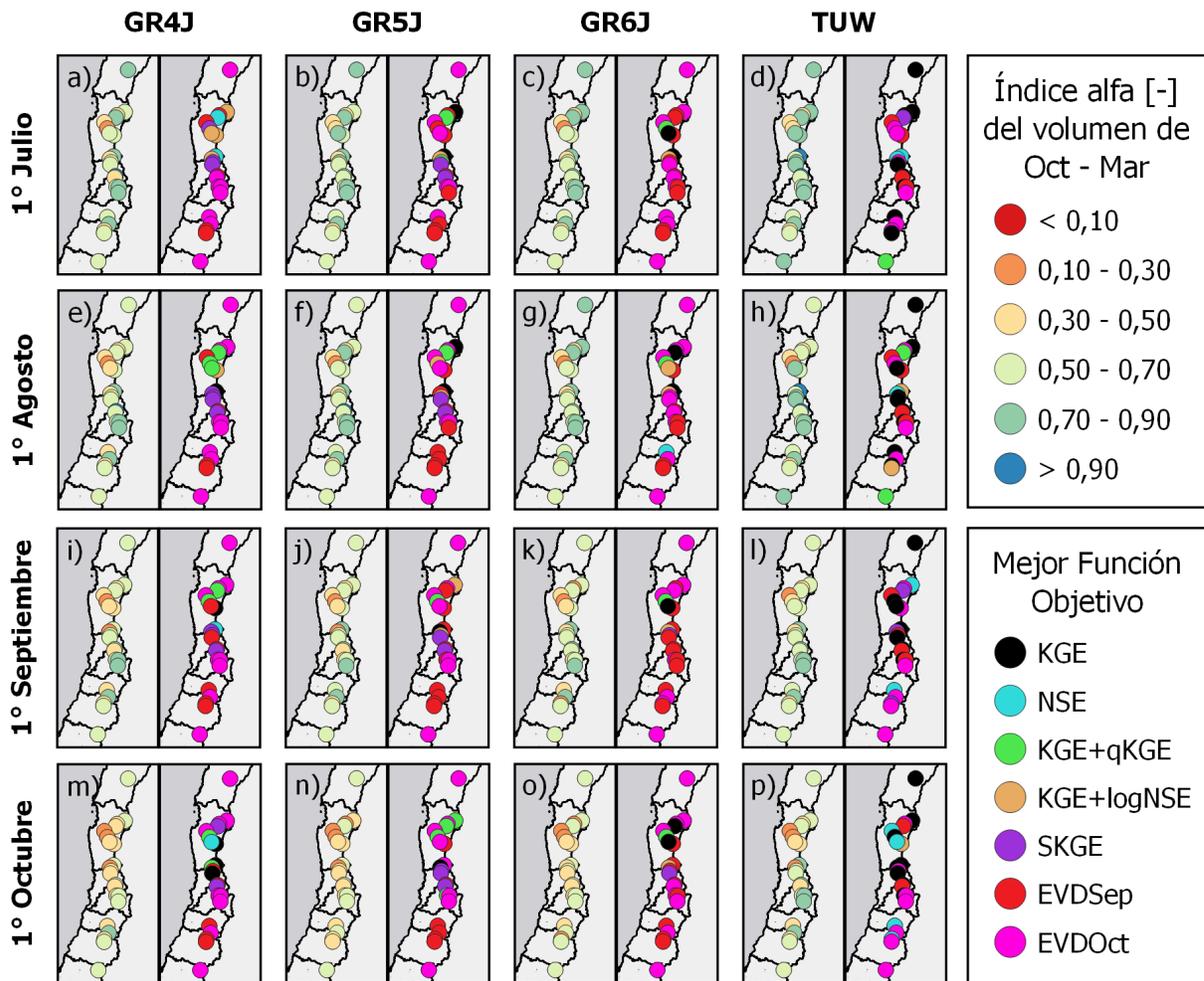


Figura A.18: Distribución espacial de los mejores índices de confiabilidad  $\alpha$  del periodo de octubre a marzo (paneles de la izquierda), junto a las funciones objetivo que los generan (paneles de la derecha), para todas las cuencas, los modelos, y cuatro tiempos de inicialización.

Con respecto a las funciones objetivo, hay una clara superioridad de las funciones estacionales (EVDSep y EVDOct) para todos los modelos, las que generan los valores de confiabilidad más altos. En el modelo TUV, KGE también entrega valores altos de confiabilidad

En la Figura A.19 se presentan los mejores índices  $R^2$  de los pronósticos ESP y su distribución espacial, por modelo y tiempo de inicialización, para el periodo de octubre a marzo. Se aprecia que la métrica aumenta su valor hacia el 1° de octubre. Los resultados indican que todos los modelos entregan valores de  $R^2$  similares. Sólo el modelo TUV se destaca por entregar valores ligeramente más altos en algunas cuencas del centro, lo que se puede apreciar con mayor claridad en agosto y septiembre.

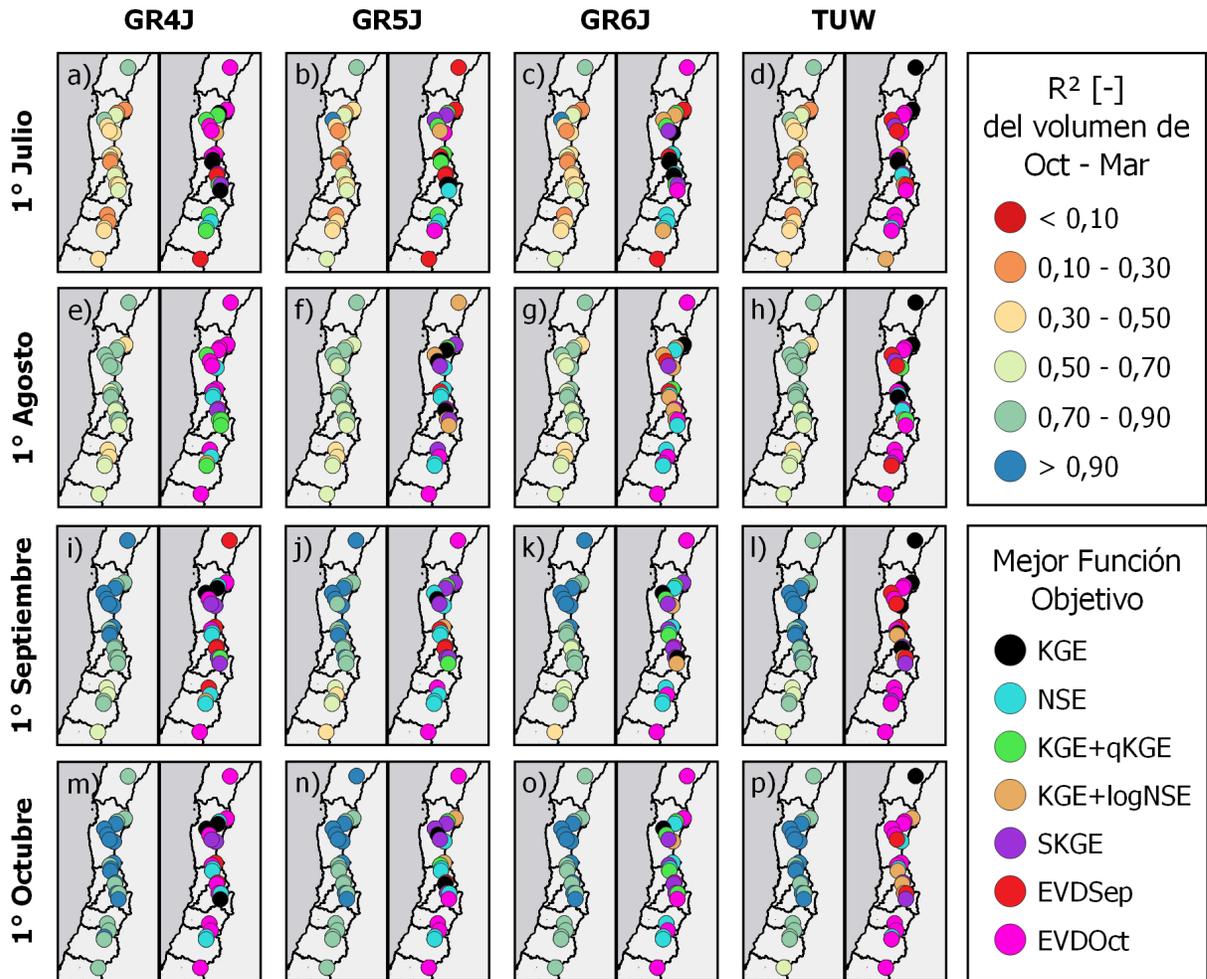


Figura A.19: Distribución espacial de los mejores  $R^2$  del periodo de octubre a marzo (paneles de la izquierda), junto a las funciones objetivo que los generan (paneles de la derecha), para todas las cuencas, los modelos, y cuatro tiempos de inicialización.

Con respecto a las funciones objetivo, las funciones estacionales (EVDSep y EVDOct) tienden a dominar en el modelo TUV, sin embargo, para el resto de los modelos, no se observan patrones claros.

### A.2.3. Análisis en detalle de los pronósticos ESP

En las Figuras A.20 a A.28 se presenta el detalle de las métricas CRPSS, índice  $\alpha$  y  $R^2$  para las 22 cuencas en estudio, según modelo, tiempo de inicialización y función objetivo, para el volumen de octubre a marzo.

Se presentan, además, las bandas asociadas al 95% de confianza para la mejor función objetivo de cada modelo. La línea discontinua muestra el valor máximo de cada métrica obtenida para el mejor modelo el 1° de octubre, con el fin de facilitar la comparación entre las funciones óptimas de cada modelo, y determinar la importancia relativa de la elección de la FO y/o el modelo.

Con respecto al CRPSS, en general, sigue siendo más importante la elección de la función objetivo, especialmente en los modelos GR, ya que existe alta variabilidad entre FO. No obstante, la elección del modelo toma más relevancia en comparación con el periodo de septiembre a marzo.

En relación al índice de confiabilidad  $\alpha$ , la elección del modelo pierde relevancia respecto al periodo de septiembre a marzo. El 1° de octubre los modelos comparten confiabilidades similares, por lo que es altamente relevante escoger una correcta FO, especialmente en los modelos GR.

Sobre el  $R^2$ , todas las funciones objetivo y los modelos se comportan similar. Sin embargo, en algunas cuencas tiene gran relevancia la elección de la FO (e.g. Pulido, Rapel, Sobrante, Claro 1, Claro 2 y Sauces). En la misma línea, la correcta elección de la función objetivo en los modelos consigue generar un  $R^2$  al nivel del  $R^2$  generado por el mejor modelo.

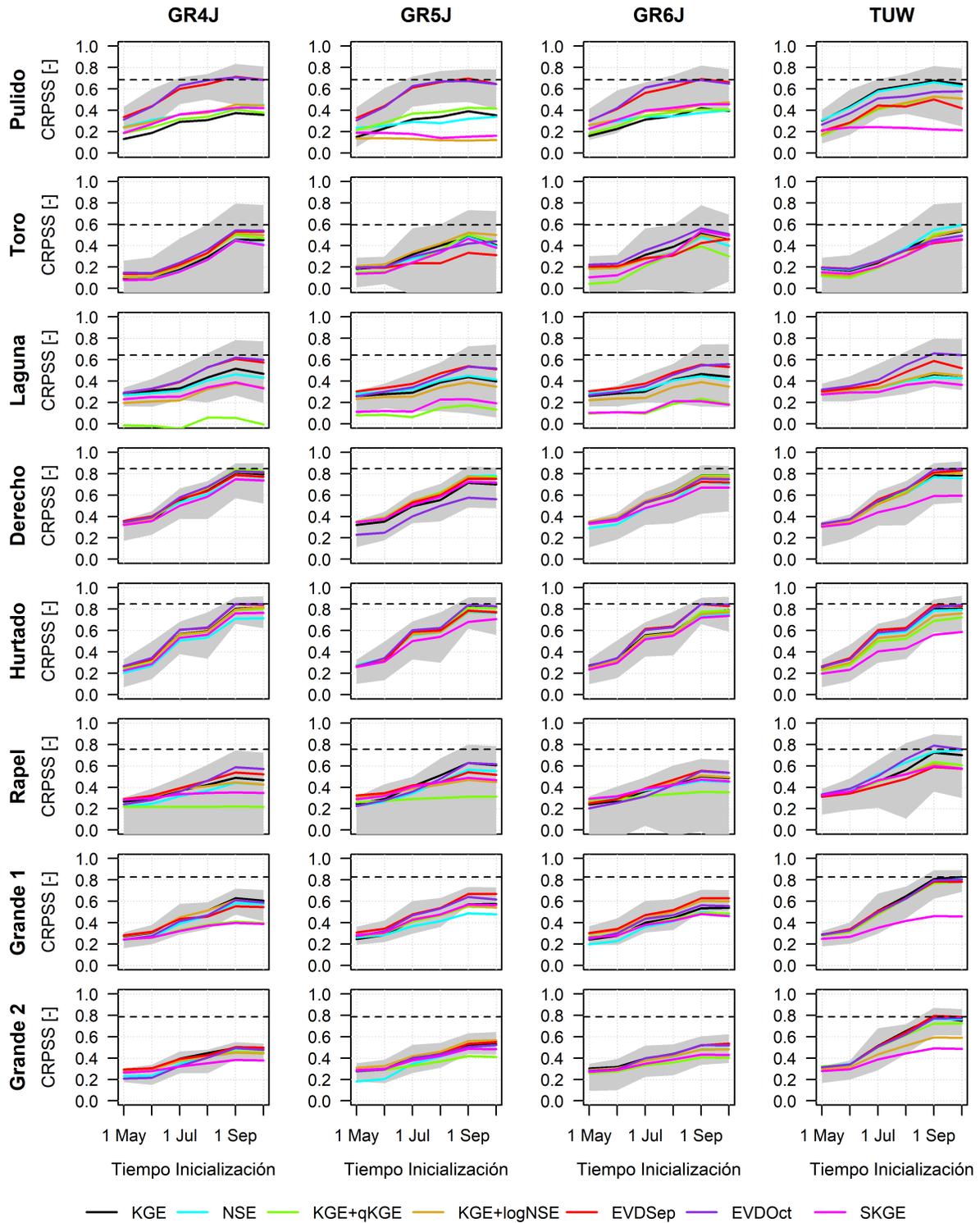


Figura A.20: Comparación de los valores de CRPSS de las cuencas 1 a 8, por función objetivo, tiempo de inicialización y modelo, para el periodo de octubre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping* con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de octubre.

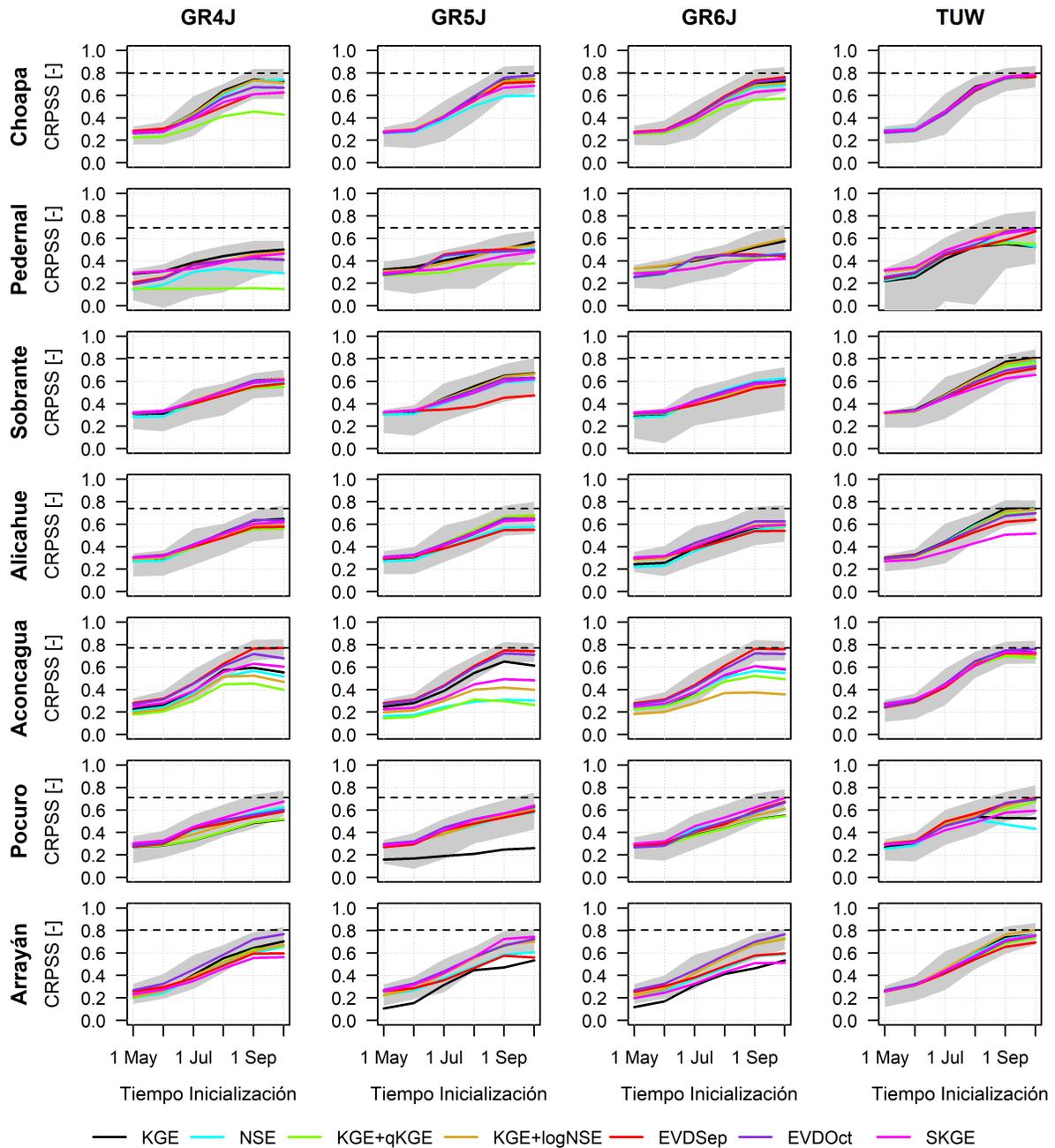


Figura A.21: Comparación de los valores de CRPSS de las cuencas 9 a 15, por función objetivo, tiempo de inicialización y modelo, para el periodo de octubre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping* con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de octubre.

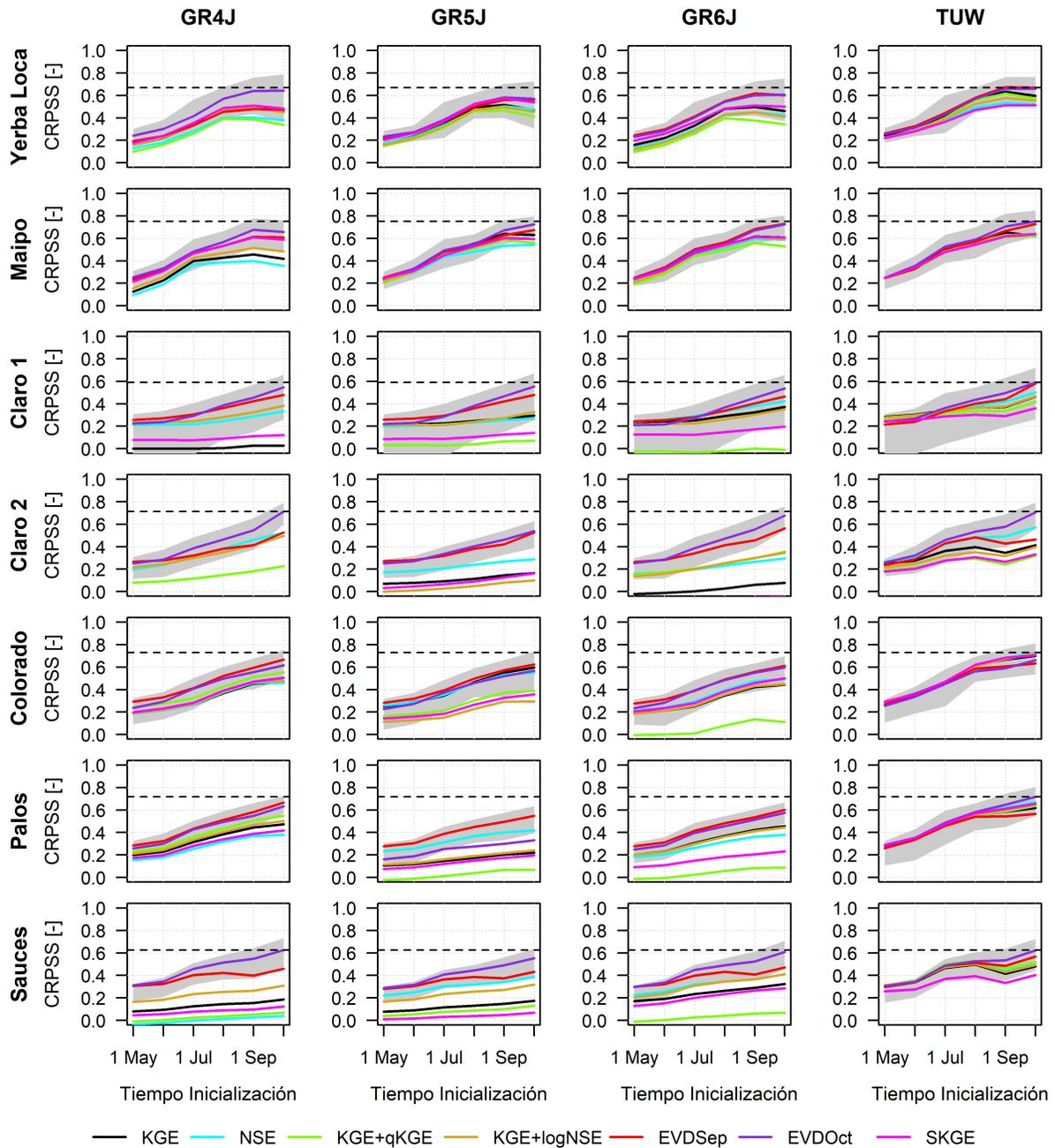


Figura A.22: Comparación de los valores de CRPSS de las cuencas 16 a 22, por función objetivo, tiempo de inicialización y modelo, para el periodo de octubre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping* con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de octubre.

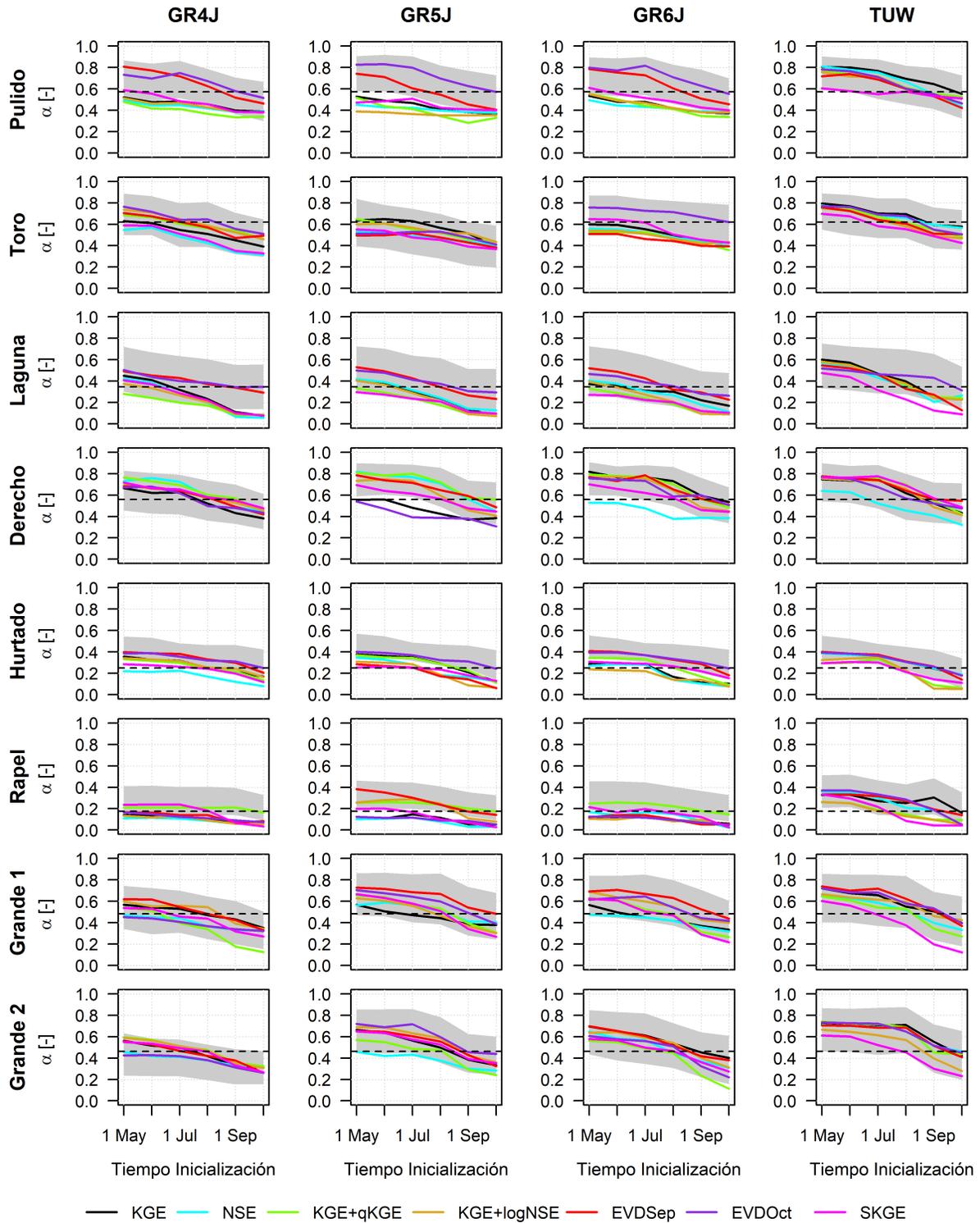


Figura A.23: Comparación de los valores del índice de confiabilidad  $\alpha$  de las cuencas 1 a 8, por función objetivo, tiempo de inicialización y modelo, para el periodo de octubre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping* con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de octubre.

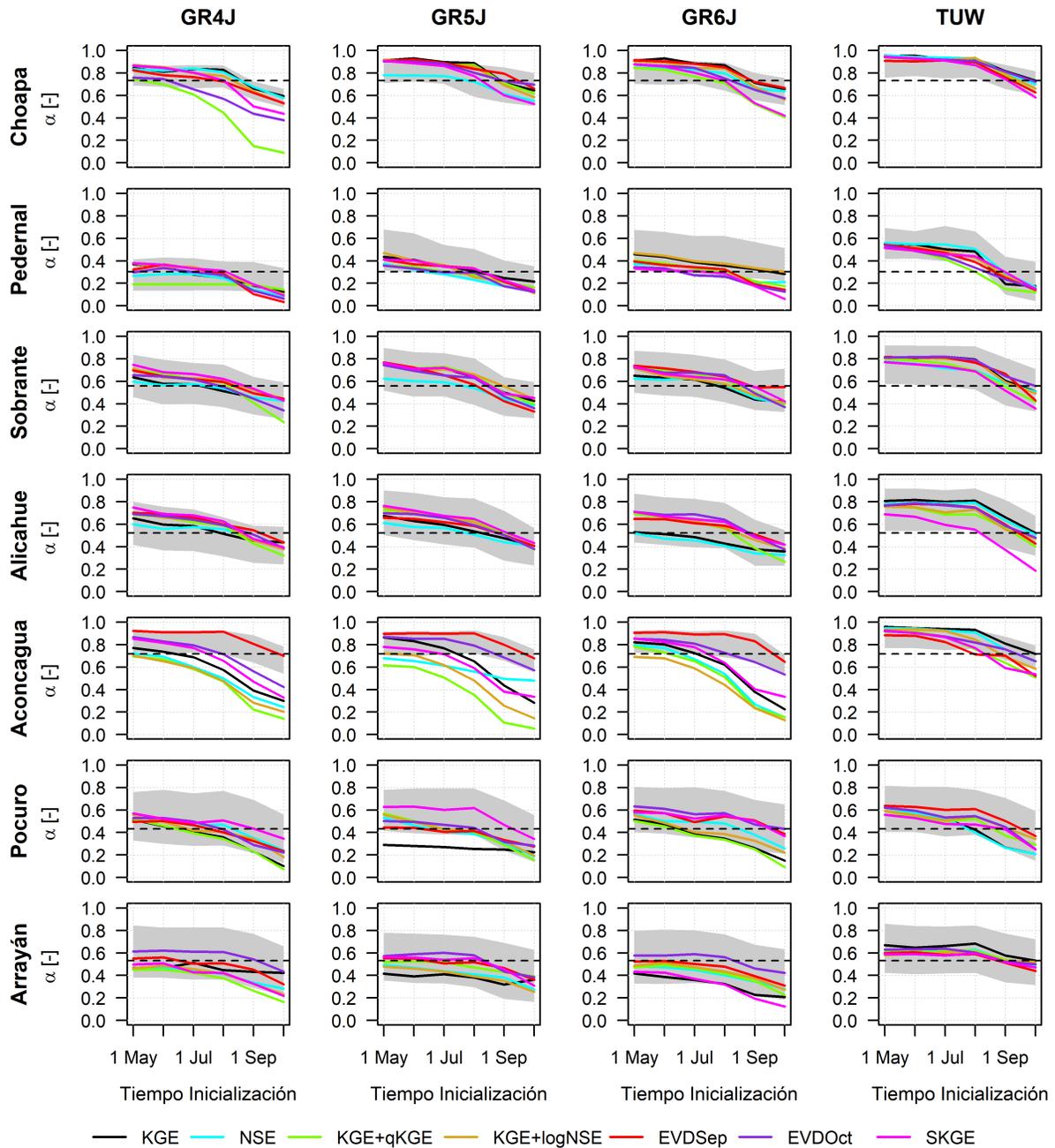


Figura A.24: Comparación de los valores del índice de confiabilidad  $\alpha$  de las cuencas 9 a 15, por función objetivo, tiempo de inicialización y modelo, para el periodo de octubre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping* con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de octubre.

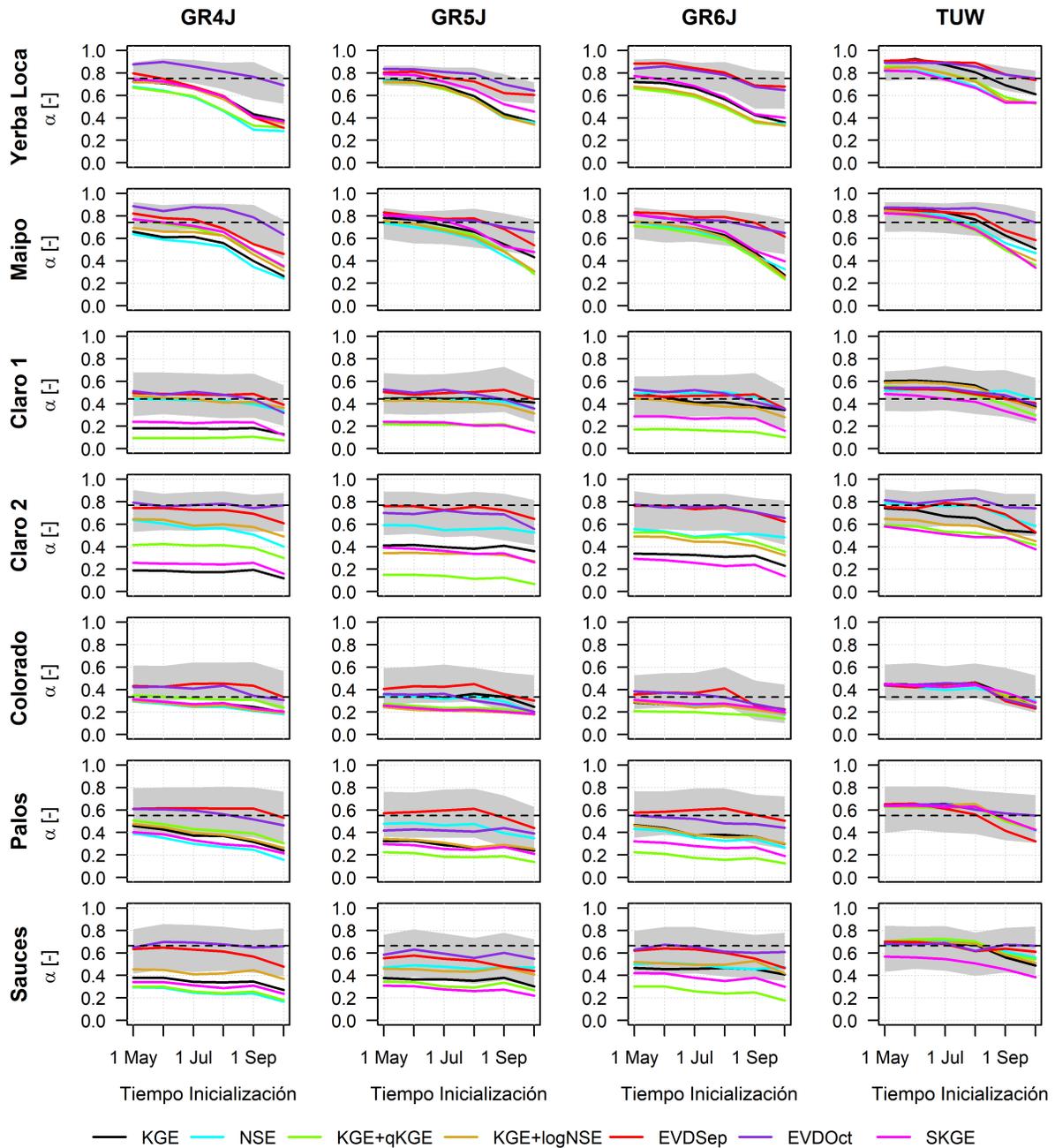


Figura A.25: Comparación de los valores del índice de confiabilidad  $\alpha$  de las cuencas 16 a 22, por función objetivo, tiempo de inicialización y modelo, para el periodo de octubre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping* con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de octubre.

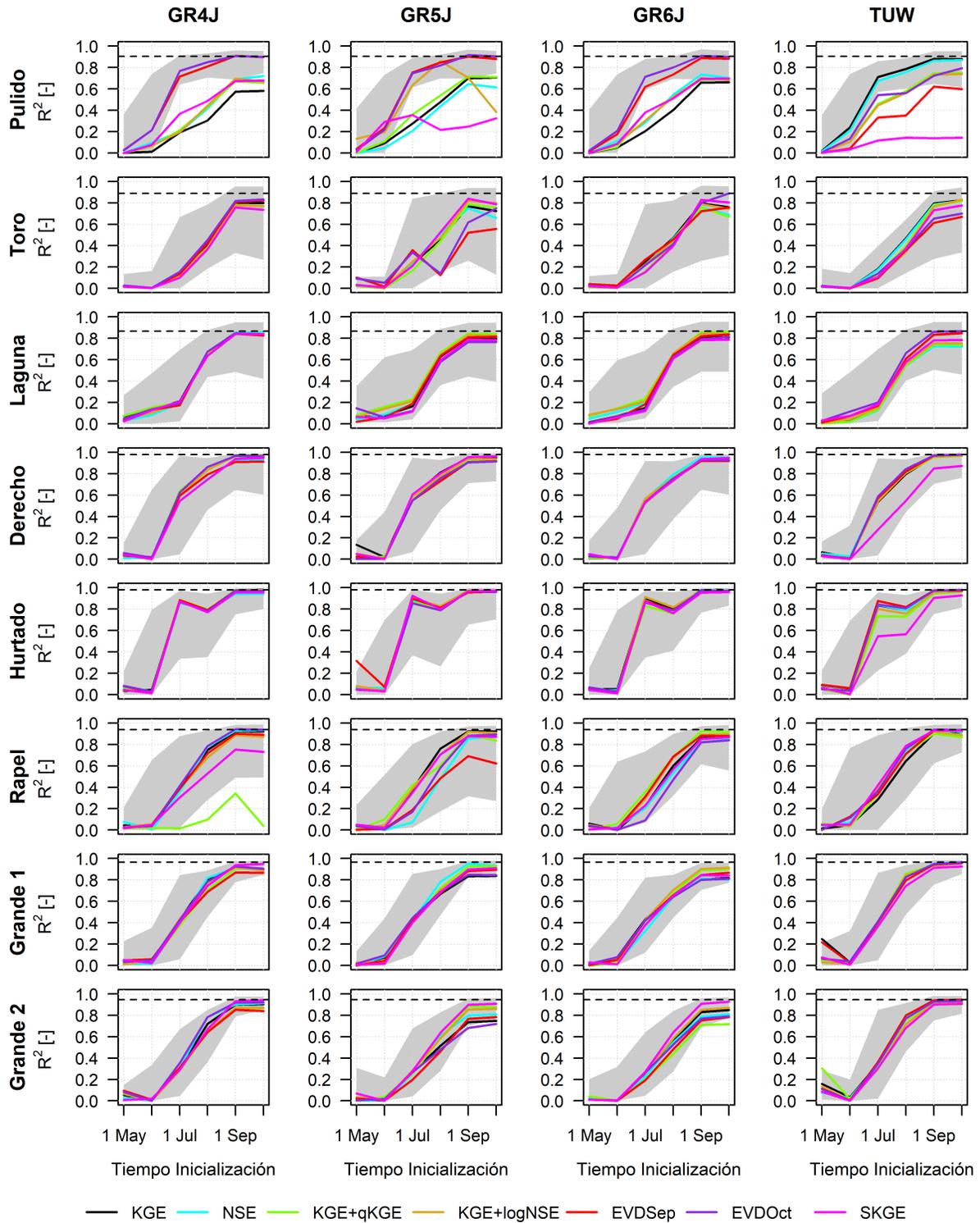


Figura A.26: Comparación de los valores de  $R^2$  de las cuencas 1 a 8, por función objetivo, tiempo de inicialización y modelo, para el periodo de octubre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping* con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de octubre.

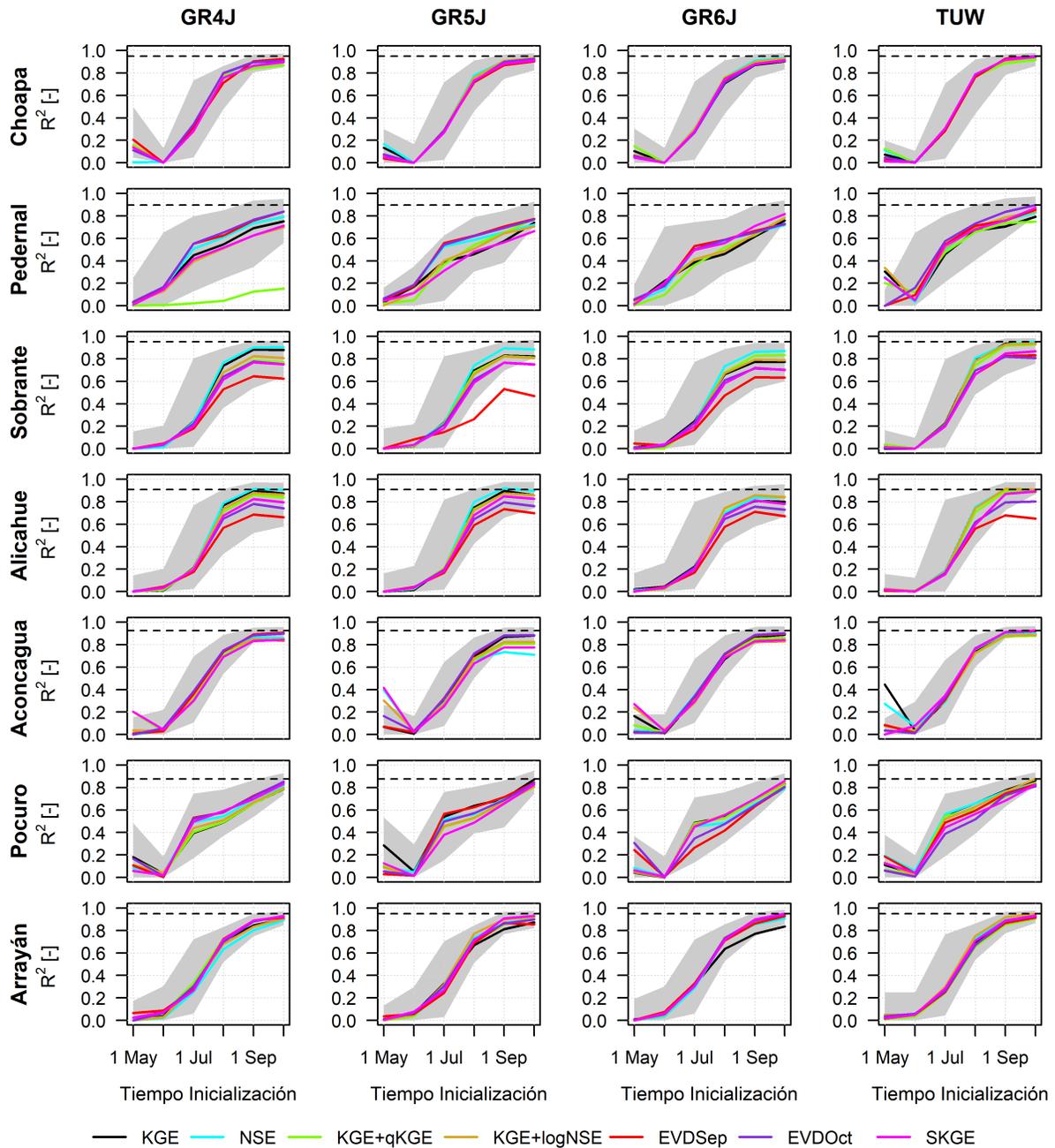


Figura A.27: Comparación de los valores de  $R^2$  de las cuencas 9 a 15, por función objetivo, tiempo de inicialización y modelo, para el periodo de octubre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping* con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de octubre.

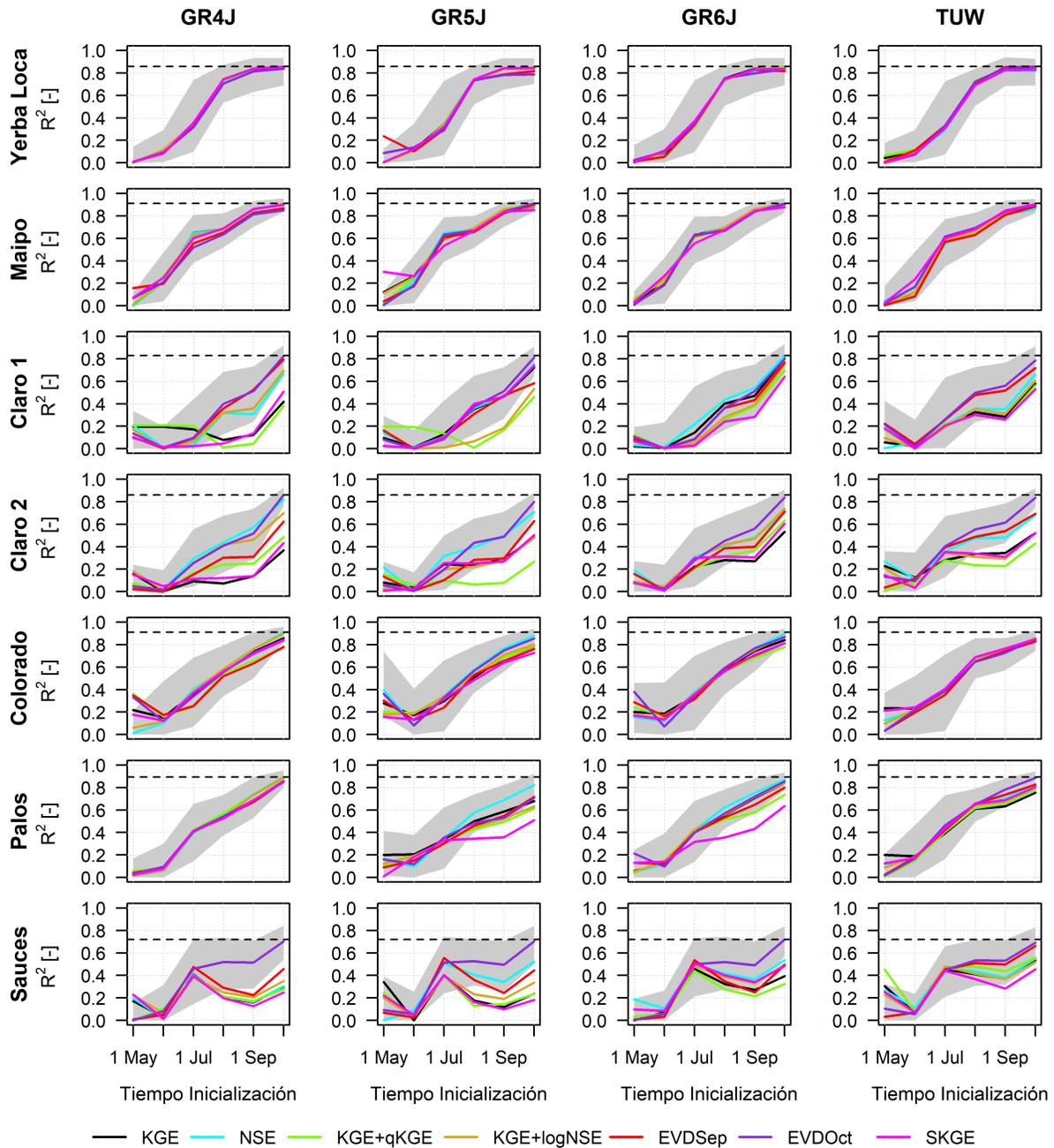


Figura A.28: Comparación de los valores de  $R^2$  de las cuencas 16 a 22, por función objetivo, tiempo de inicialización y modelo, para el periodo de octubre a marzo. La banda gris define los límites de confianza al 95% de la mejor FO de cada modelo, obtenidos mediante *bootstrapping* con reemplazo. La línea discontinua demarca el máximo CRPSS obtenido para algún modelo el 1° de octubre.

### A.3. Complemento análisis de los pronósticos ESP a resolución temporal mensual de las cuencas DGA

En la Figura A.29 se muestra la evolución temporal del índice CRPSS de los pronósticos a resolución temporal mensual de las cuencas DGA. Se observa que los valores son más bajos que a resolución temporal estacional, existiendo algunas funciones objetivo que generan valores negativos. En general, EVDSep y SKGE son las que generan los valores más bajos, y NSE la que genera los más altos.

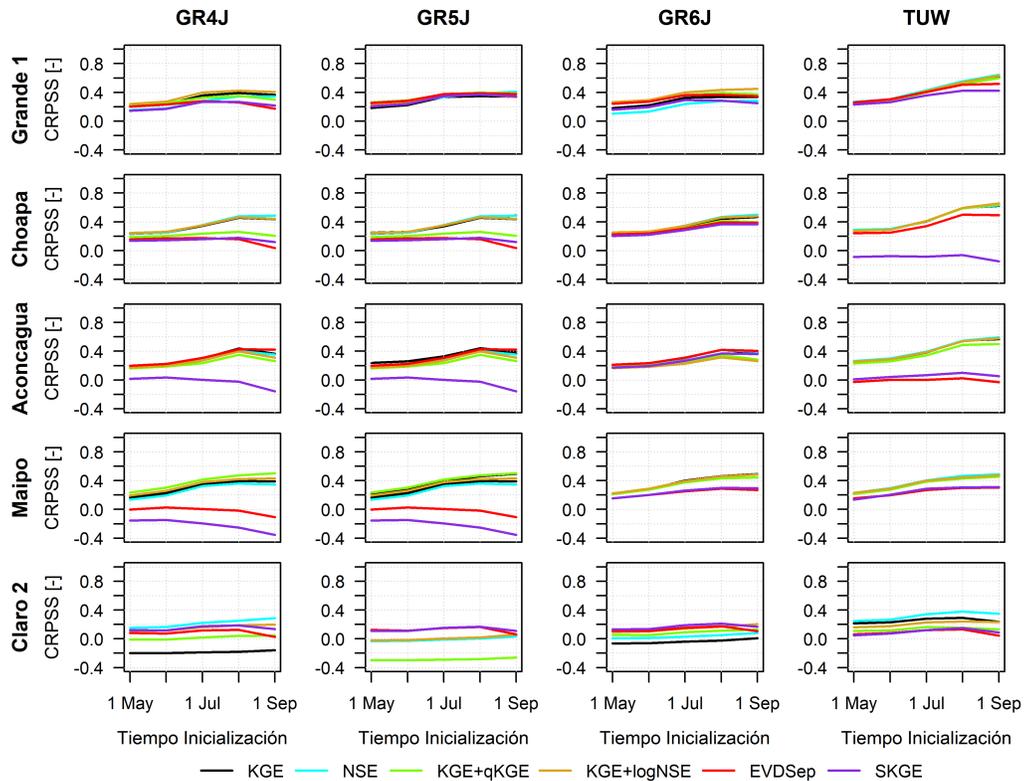


Figura A.29: Comparación de los valores de CRPSS de los pronósticos a resolución temporal mensual de las cuencas DGA por función objetivo, tiempo de inicialización y modelo.

Con respecto al índice  $\alpha$  (Figura A.30, EVDSep es una de las funciones objetivo que genera los valores más bajos, mientras que no hay claridad en la función que genera los valores más altos. En general, el comportamiento es similar en todos los modelos.

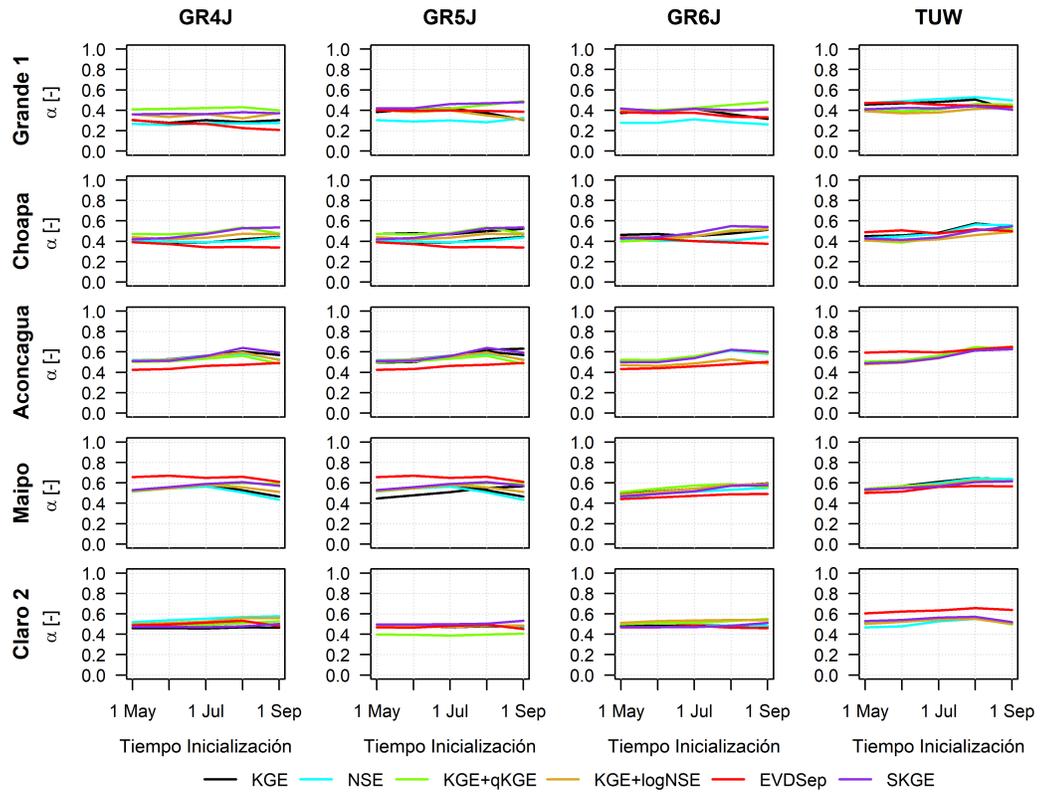


Figura A.30: Comparación de los valores del índice  $\alpha$  de los pronósticos a resolución temporal mensual de las cuencas DGA por función objetivo, tiempo de inicialización y modelo.

Finalmente, en la Figura A.31 se muestra el coeficiente de determinación y su evolución temporal. Es posible notar que NSE es la función objetivo que, en general, entrega los valores más altos, mientras que EVDSEp los valores más bajos.

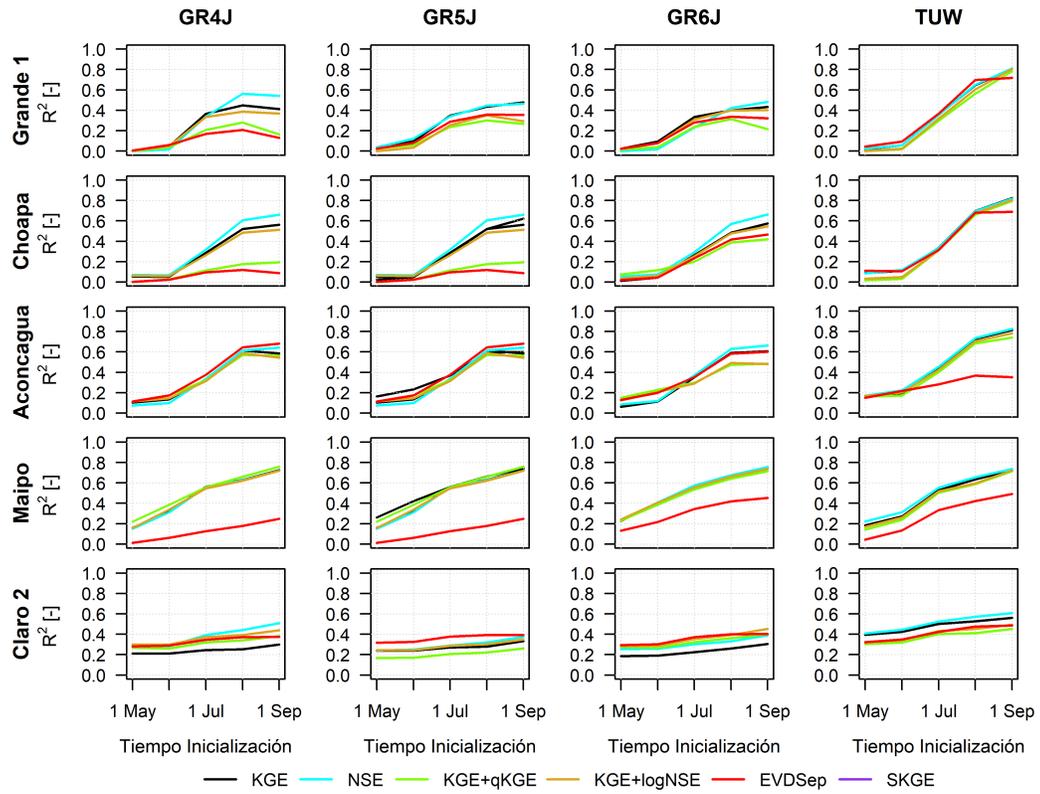


Figura A.31: Comparación de los valores de  $R^2$  de los pronósticos a resolución temporal mensual de las cuencas DGA por función objetivo, tiempo de inicialización y modelo.