



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

ASIMILACIÓN DE OBSERVACIONES IN SITU E INFORMACIÓN SATELITAL PARA ESTIMACIONES DE BALANCE HÍDRICO EN CHILE CONTINENTAL

TESIS PARA OPTAR AL GRADO DE MAGISTER EN CIENCIAS DE LA INGENIERÍA
MENCION RECURSOS Y MEDIO AMBIENTE HÍDRICO

EDUARDO ALBERTO FERNANDO MUÑOZ CASTRO

PROFESOR GUÍA
PABLO MENDOZA ZÚÑIGA

MIEMBROS DE LA COMISIÓN
GONZALO CORTÉS SORUCO
XIMENA VARGAS MESA

ESTE TRABAJO HA SIDO PARCIALMENTE FINANCIADO POR ANID A TRAVÉS DE LA BECA MAGISTER NACIONAL AÑO 2020
FOLIO N°22200599 Y APOYADO POR LA INFRAESTRUCTURA DE SUPERCÓMPUTO DEL NLHPC (ECM-02)

SANTIAGO DE CHILE
2021

RESUMEN DE LA TESIS PARA OPTAR AL
GRADO DE: Magister en Ciencias de la Ingeniería
mención Recursos y Medio Ambiente Hídrico
POR: Eduardo Muñoz Castro
FECHA: Julio 2021
PROFESOR GUÍA: Pablo Mendoza Zúñiga

ASIMILACIÓN DE OBSERVACIONES IN SITU E INFORMACIÓN SATELITAL PARA ESTIMACIONES DE BALANCE HÍDRICO EN CHILE CONTINENTAL

Los modelos hidrológicos son herramientas útiles para estudiar diferentes procesos e interacciones del ciclo hidrológico. No obstante, la escasez de observaciones in-situ es siempre una limitación para configurar y validar estos modelos, los cuales corresponden a un conjunto de hipótesis sobre el ciclo hidrológico. Los avances en percepción remota posicionan a observaciones derivadas a partir de este tipo de técnicas como una alternativa para mejorar la disponibilidad de información. A pesar de ello, tanto observaciones in-situ como remotas conllevan errores intrínsecos, propios de su medición y/o estimación. De este modo, además de modelos imperfectos, se tienen observaciones imperfectas, a partir de las cuales se busca generar conocimiento hidrológico. La asimilación de datos o *data assimilation* (DA) busca combinar de manera óptima la información proveniente de observaciones con aquella proveniente de modelos hidrológicos para mejorar sus estimaciones y reducir la incertidumbre asociada a ellas.

El presente estudio tuvo por objetivo evaluar el beneficio de implementar asimilación de datos en estimaciones de balance hídrico a escala de cuenca en Chile continental. Para ello, se implementó el Filtro de Kalman de conjuntos (EnKF) acoplado al modelo hidrológico *Variable Infiltration Capacity* (VIC), espacialmente distribuido de 0,05° lat-lon, en 12 cuencas a través de un gradiente hidroclimático en Chile central. Se configuraron experimentos en base a la perturbación de las forzantes de VIC (precipitación, velocidad de viento, temperatura y humedad relativa) y cuatro casos según los estados perturbados y observaciones asimiladas: i) C1.E1, perturbación de humedad de suelo (SM) inicial y asimilación de escorrentía (Q), ii) C2.E1, perturbación de SM y equivalente de nieve en agua (SWE) inicial, y asimilación de Q, iii) C2.E2, perturbación de SM y SWE iniciales, y asimilación de SWE y cobertura de nieve (SC) agregados a escala de cuenca, y iv) C2.E3, perturbación de SM y SWE iniciales, y asimilación de Q, SWE y SC.

Los resultados obtenidos muestran que 1) la mejor configuración es aquella donde se asimila Q, SC y SWE (i.e., C2.E3), donde el RMSE promedio de los conjuntos de las 12 cuencas mejora, en promedio, $3,76 \pm 4,58$ mm/d en términos de la escorrentía media diaria y $80,79 \pm 54,12$ mm para el SWE diario promedio a escala de cuenca respecto al caso sin asimilación; 2) la asimilación de observaciones concentradas permite corregir sesgos en flujos y estados del sistema, pero no permite ajustar patrones espaciales de las variables de interés; y 3) la asimilación de datos tiene impactos en todas las variables de estado y flujos simulados, principalmente en la escorrentía directa y flujo base. Asimismo, estos impactos se traducen en cambios en el balance hidrológico a nivel anual.

Finalmente, el presente estudio abre el debate sobre las implicancias de la asimilación de datos en la estimación del balance hídrico en cuencas con componente nival. A su vez, se evidencia la necesidad de evaluar la consistencia hidrológica de los resultados de asimilación utilizando observaciones no asimiladas y mejorar la comprensión de la relación estimación a priori – observación – estimación a posteriori.

*A mi abuela, Zoila, a mi tata, Oscar, y a mi primo, Keko...
Quienes desde aquella dimensión inmensa e infinita de la eternidad
están disfrutando y celebrando este momento...
Esto es para ustedes, los amo*

“Carpe diem quam minimum credula postero memento mori...”



AGRADECIMIENTOS

Resulta imposible comenzar estas líneas y, más aún, dar cualquier paso en mi andar por esta vida, sin antes agradecer y reconocer a mis padres, Oscar y Agripina, quienes con su cariño, amor y sacrificio me han permitido cumplir cada uno de mis sueños. A mi hermano, Oscar, mi Rorro, mi referente, quién (a su modo) me ha entregado grandes enseñanzas de la vida y momentos de antología que espero siempre podamos repetir. A mi pareja, amiga y compañera, Camila, por su amor, apoyo y compañía en estos 10 años en los cuales hemos crecido caminando juntos y tomados de la mano por el camino de la vida. A mis tíos, tías y primos, quienes pese a las diferencias que puedan existir, siempre están ahí, ya sea en las malas o para compartir un buen vino, un buen asado o una buena conversación cruzada en la sobremesa. A quienes ya tuvieron que partir, gracias por las enseñanzas y los buenos momentos, las llevo en el corazón.

A mis amigos Guillermo Vásquez, Matías Escudero, Carolina Pérez y Sofia Peña, quienes desde el primer año de universidad me han acompañado, celebrando los éxitos y motivándome en los momentos difíciles. Quienes a pesar de la distancia física que trajo consigo la pandemia de Covid19, han sabido estar ahí, preocupados con un mensaje, un meme o un juego en línea conectados por llamada telefónica. A Daniela Barrios y Manuel Cortés, con quienes a pesar de haber tomado caminos diferentes y no poder compartir o hablar con frecuencia, el cariño, confianza y amistad perdura.

A Jacqueline Suárez (Jaqui), por ser simplemente ella, por las buenas vibras que entrega cada día, por la dedicación que pone en su trabajo, reflejando el verdadero espíritu del servidor público, por ser una persona maravillosa de corazón infinito. A Nicolás Vásquez (futuro académico, no tengo duda de eso), por su enorme voluntad y disposición a enseñar, resolver alguna duda o simplemente tomar una cerveza (o ron). A Natalia Bustamante, por la compañía durante los almuerzos y la buena conversación que allí siempre se dio. A Nery Budde por sus buenas energías y apoyo en los cursos que compartimos como cuerpo docente. A Alvaro Torrent, por el apoyo informático y soporte técnico con el PC de la oficina. A las tías Verito, Isabel y Carmen, por mantener nuestro espacio de trabajo impecable y su enorme calidad humana. A los amigos de la oficina 340 de la división de RHMA: Alonso Mejías, Roberto Jaco, Borja Mir, Felipe Durán y Fabiola Pinto. Por la buena onda, las conversaciones procastinando, los temas musicales que escuchamos, los cafecitos, las papas fritas aceitosas y más.

Al Doctor Pablo Mendoza, profesor, investigador y músico. Gracias por confiar y creer en mi, por guiarme en esta investigación, por tu enorme paciencia para enseñar, por tu humildad. Gracias por brindarme tu amistad, por aquellos karaokes y cervezas arreglando el mundo y discutiendo sobre hidrología. Gracias por permitirme compartir contigo diferentes cursos de pre y postgrado (por favor no me jubiles) y seguir aprendiendo con ellos. Gracias por eso y mucho más.

A la profesora Ximena Vargas, por todas las oportunidades que me ha dado, por desafiarme a siempre dar más, por las conversaciones filosóficas y/o hidrológicas tomando un cafecito (presenciales hasta marzo de 2020 y luego remotas), por sus consejos y la sabiduría que transmite. Sin lugar a duda el ejemplo patente de un maestro y guía. ¡No nos deje nunca!

A Gonzalo Cortés, por aceptar ser parte de esta comisión y por sus acuciosos comentarios y observaciones que sirvieron para complementar la investigación. Espero algún día nos podamos conocer en persona y poder conversar sobre DA y cualquier otro tema que pueda surgir.

En fin, gracias a todos por ser parte de mi vida. Con muchos estaré en deuda eternamente, pero tengan por seguro que siempre los llevaré en mi corazón y les honraré a través de mis acciones, pues si hoy soy quién soy y he llegado hasta aquí, es gracias a ustedes.

TABLA DE CONTENIDO

1	INTRODUCCIÓN	1
1.1	Motivación	1
1.2	Avances recientes en asimilación de datos.....	1
1.3	Potencial de la asimilación en problemas de disponibilidad hídrica	2
1.4	Objetivos	3
1.5	Estructura del documento	3
2	ZONA DE ESTUDIO	4
2.1	Selección de cuencas de estudio.....	4
2.2	Caracterización general de las cuencas seleccionadas.....	4
3	DATOS DISPONIBLES	8
3.1	Registros fluviométricos	8
3.2	Observaciones relacionadas al manto nival	9
3.3	Evapotranspiración (ET) y humedad de suelo (SM).....	10
4	METODOLOGÍA	12
4.1	Modelo Hidrológico Variable Infiltration Capacity (VIC).....	13
4.1.1	Descripción del modelo hidrológico	13
4.1.2	Forzantes meteorológicas	14
4.1.3	Calibración y diagnóstico del modelo hidrológico	14
4.2	Asimilación de datos – Ensemble Kalman Filter (EnKF)	17
4.3	Error del modelo y generación de conjuntos (<i>ensembles</i>)	19
4.3.1	Perturbación de forzantes y estados	19
4.4	Interpretación y verificación de conjuntos	22
4.5	Evaluación de la asimilación de datos.....	24
4.5.1	Evaluación de flujos y estados agregados a escala de cuenca.....	24
4.5.2	Evaluación respecto a observaciones de referencia no asimiladas.....	25
4.6	Configuración de experimentos de asimilación de datos	26
4.7	Incorporación del error de la observación	28
4.8	Resumen de errores de modelación y observación adoptados.....	28
5	RESULTADOS.....	30
5.1	Calibración del modelo hidrológico.....	30
5.2	Habilidad del ensemble generado	33
5.3	Resultados de asimilación de datos	35
5.3.1	Evaluación de resultados agregados a escala de cuenca	35

5.3.2	Evaluación de resultados espacialmente distribuidos.....	41
5.3.3	Implicancias de la asimilación de datos en la caracterización del año hidrológico 2000/2001	44
6	COMENTARIOS Y CONCLUSIONES.....	48
	BIBLIOGRAFÍA	51
	ANEXO A: RESULTADOS DE LA CALIBRACIÓN DEL MODELO HIDROLÓGICO VIC EN LAS CUENCAS DE ESTUDIO	I
	ANEXO B: CONSIDERACIONES ADICIONALES	XI
	ANEXO C: COMPARACIÓN DE CONFIGURACIONES ADOPTADAS PARA LA ASIMILACIÓN DE DATOS.....	XV

ÍNDICE DE TABLAS

Tabla 2.1: Atributos fisiográficos de las cuencas de las 12 cuencas de estudio.....	6
Tabla 2.2: Características hidroclimáticas de las 12 cuencas de estudio. Se incluye el promedio anual en el periodo 1985-2015, con el rango ([min. – máx.]) asociado, del índice de aridez (IA), precipitación (P), temperatura media (Tm) y escorrentía (Q).....	7
Tabla 4.1: Subperiodos de tiempo definidos para la calibración y validación del modelo hidrológico	15
Tabla 4.2: Métricas basadas en procesos aplicadas para el diagnóstico de la calibración del modelo hidrológico VIC.....	16
Tabla 4.3: Estadísticos y métricas para la interpretación de conjuntos. El subíndice i representa cada paso de tiempo.	22
Tabla 4.4: Criterios para la verificación de conjuntos.	22
Tabla 4.5: Criterios de eficiencia definidos para la evaluación de las simulaciones OL y DA.	24
Tabla 4.6: Configuración de experimentos de asimilación.	27
Tabla 4.7: Resumen de errores definidos para el ejercicio de asimilación de datos con EnKF. .	29
Tabla 5.1: Cambios en la partición P-ET-Q al asimilar datos según las diferentes configuraciones propuestas respecto al caso sin asimilación, desagregados por cuenca.	47

ÍNDICE DE FIGURAS

- Figura 2.1: Ubicación geográfica de las cuencas de estudio y patrones hidrológicos estacionales promedio en el periodo 1985-2015. (A) Mapa de la ubicación geográfica de las 12 cuencas de estudio. (B-D) Panel superior: precipitación mensual promedio en el periodo 1985-2015 promediada entre las cuencas contenidas en el subdominio. Las barras de color negro representan valores máximos y mínimos de precipitación promedio para el subconjunto de cuencas (i.e., rango), y la línea segmentada corresponde al promedio (equivalente a barras azules). En rojo se indica la temperatura media anual promedio y su rango. Panel inferior: Coeficientes de Pardé anuales promedio en el periodo 1985-2015 entre las cuencas que componen el subdominio indicado (línea azul continua) y rango de variación (zona difuminada).....5
- Figura 3.1: Curvas de descarga oficiales en el periodo 1985-2015 utilizadas por la Dirección General de Aguas (DGA) para el control fluviométrico. Ejemplo asociado a la estación (A) Estero Derecho en Alcohuaz – BNA 4311001, (B) Río Choapa en Cuncumén – BNA 4703002, (C) Estero Arrayán en la Montosa – BNA 57220001, y (D) Río Sauces antes junta con Ñuble – BNA 8104001. Líneas continuas corresponden a las curvas de descarga estimada por la DGA, mientras que los trazos segmentados son una extrapolación realizada para ajustar visualmente los gráficos reportados (i.e., las curvas no tienen los mismos rangos altimétricos y de caudal).9
- Figura 3.2: Ejemplo de los valores de referencia derivados de la interpolación de GLEAM v3.2a a la grilla de CR2Met. (A) Distribución espacial de la evapotranspiración (con resolución 0,05° lat-lon) para el año hidrológico 2000/2001. (B) Acercamiento (*zoom in*) a la figura del panel A y ajuste en la escala de colores que da cuenta de la evapotranspiración estimada para el año hidrológico 2000/2001 en la cuenca Estero Arrayán en la Montosa distribuida espacialmente a 0,05° lat-lon. (C) Serie de tiempo de la evapotranspiración y humedad de suelo (SM) de la zona radicular agregada a nivel de la cuenca Estero Arrayán en la Montosa (01/04/2000 – 31/03/2001).....11
- Figura 4.1: Esquema metodológico del estudio.....12
- Figura 4.2: Arquitectura del modelo hidrológico Variable Infiltration Capacity (VIC).13
- Figura 4.3: Esquema metodológico de la estimación de matrices de covarianza de forzantes y estados del sistema. Ejemplo aplicado a una cuenca que ocupa nueve (9) pixeles de la grilla espacial con la cual se configura el modelo hidrológico.21
- Figura 4.4: Ejemplo de interpretación de histogramas de rango.23
- Figura 4.5: Diagrama de flujo del proceso de asimilación de datos con EnKF.27
- Figura 4.6: Esquema de la metodología empleada para la agregación de estados iniciales, distribuidos a nivel de subpíxel y píxel, a escala de cuenca y posterior propagación de estados actualizados desde la escala de cuenca a la escala de subpíxel. Ejemplo aplicado al caso de la actualización de humedad de suelo. En el caso del equivalente de nieve en agua (SWE), sólo existe una variable que se manipula de manera análoga a cualquier capa de suelo.28
- Figura 5.1: Resultados de calibración del modelo VIC en la cuenca Estero Arrayán en La Montosa – BNA 5722001. Se muestran las simulaciones para el periodo 01/01/1979-31/12/2015, y el color gris muestra el periodo de calibración utilizado. (A) Caudal medio diario observado y simulado. (B) Caudal medio mensual observado y simulado. (C) Caudal medio diario promedio anual observado y simulado. (D) Ubicación geográfica de la cuenca. (E) Balance hidrológico adimensional simulado; los puntos pequeños corresponden al balance anual, mientras que el punto de mayor tamaño y con borde rojo corresponde al promedio en el

periodo 1985-2015. (F) Curva de duración de caudales medios diarios. (G) Cobertura de nieve agregada a escala de cuenca. (H) Equivalente de nieve en agua agregado a escala de cuenca.....	30
Figura 5.2: Resultados del KGE y sus componentes obtenidos de la calibración (01/04/1985-31/03/2005) y validación (01/04/2005-31/04/2015) del modelo VIC en las cuencas de estudio. Cada caja agrupa 12 valores (i.e., un valor por cuenca), mientras que los subperiodos seco y húmedo representan déficit o superávit de por lo menos un 20% de la precipitación anual respecto al promedio 1985-2015.	31
Figura 5.3: Evaluación del desempeño del modelo VIC en las cuencas de estudio en el periodo de calibración (01/04/1985-31/03/2005) y validación (01/04/2005-31/04/2015), a partir de índices de consistencia hidrológica.....	32
Figura 5.4: Histograma de rango del ensemble sin asimilación (OL) generado a partir de la perturbación de la condición inicial de humedad en las tres capas de suelo configuradas en VIC (Caso 1). En cada gráfico, se indica el código BNA de la cuenca de estudio, el índice de discrepancia y los valores asociados a consistencia del ensemble e indistinguibilidad estadística normalizada. La línea roja segmentada representa una distribución de rangos uniformes.....	34
Figura 5.5: Histograma de rango del ensemble sin asimilación (OL) generado a partir de la perturbación de la condición inicial de humedad en las tres capas configuradas en VIC y el SWE inicial (Caso 2). En cada gráfico, se indica el código BNA de la cuenca de estudio, el índice de discrepancia y los valores asociados a consistencia del ensemble e indistinguibilidad estadística normalizada. La línea roja segmentada representa una distribución de rangos uniformes.	35
Figura 5.6: Resultados de escorrentía media diaria a escala de cuenca según las configuraciones con y sin asimilación de datos en la cuenca Estero Arrayán en La Montosa.....	36
Figura 5.7: Resultados de SWE a escala de cuenca y resolución temporal diaria según las configuraciones con y sin asimilación de datos en la cuenca Estero Arrayán en La Montosa.	36
Figura 5.8: Resultados de la fSCA a escala de cuenca y resolución temporal diaria según las configuraciones con y sin asimilación de datos en la cuenca Estero Arrayán en La Montosa.	37
Figura 5.9: Evaluación del valor agregado, expresado en términos de la raíz del error cuadrático medio normalizado (NRMSE), de asimilar datos en las estimaciones de escorrentía (Q), equivalente de nieve en agua (SWE) y cobertura de nieve (SC o fSCA) con respecto a la situación sin asimilación (año hidrológico 2000/01). La línea roja segmentada representa el valor en que los resultados sin asimilación y con asimilación son equivalentes. Notar que el eje Y se ha truncado en 1.5 para facilitar la comparación entre esquemas de asimilación.	38
Figura 5.10: Ganancia normalizada de la asimilación de datos respecto al caso OL en la estimación de escorrentía total media diaria para el año hidrológico 2000/01. La línea roja segmentada representa el valor en que los resultados sin asimilación y con asimilación son equivalentes. El valor máximo en el eje X se ha truncado en 1.5 con fines comparativos.	39
Figura 5.11: Ganancia normalizada de la asimilación de datos respecto al caso OL en la estimación de SWE para el año hidrológico 2000/01. La línea roja segmentada representa el valor en que los resultados sin asimilación y con asimilación son equivalentes. El valor máximo en el eje X se ha truncado en 1.5 con fines comparativos.	39
Figura 5.12: Ganancia de asimilar datos en la estimación de flujos y estados, a la escala de cuenca, no considerados como observaciones asimilables. La línea roja segmentada	

representa el valor en que los resultados sin asimilación y con asimilación son equivalentes. La extensión de las barras representa el percentil del 25% y 75% asociado a las diferencias estimadas para cada miembro evaluado a nivel diario en el periodo dado por el año hidrológico 2000/01. El punto al centro de cada una de las barras representa la mediana del conjunto.....41

Figura 5.13: Ejemplo de los impactos de la asimilación de datos en los montos de escorrentía anual distribuidos a 0,05° latitud-longitud. En los diferentes paneles (A-D), se muestran los resultados obtenidos para la cuenca Estero Arrayán en La Montosa según las diferentes configuraciones. El punto rojo indica la ubicación de la estación fluviométrica que da nombre a la cuenca.42

Figura 5.14: Ganancia normalizada de asimilar datos en la representación de patrones espaciales de evapotranspiración real (ET), sublimación (SUB), humedad de suelo superficial (SM_{sup}) y humedad de suelo en la zona radicular (SM_{root}). La línea segmentada roja representa el caso en que no hay ganancia. El valor máximo en el eje X se ha truncado en 1.5 con fines comparativos.42

Figura 5.15: Ganancia normalizada de asimilar datos en las componentes de la métrica E_{sp} respecto al caso sin asimilación. Se incluyen resultados para ET (izquierda) y SM_{root} (derecha). La línea roja segmentada representa el valor en que los resultados sin asimilación y con asimilación son equivalentes. El valor máximo en el eje X se ha truncado en 1.5 con fines comparativos.43

Figura 5.16: Diferencias porcentuales en flujos y estados derivados de VIC entre las configuraciones de asimilación de datos y los casos sin asimilación. La línea roja segmentada representa el caso en que no hay diferencias entre las simulaciones con y sin asimilación de datos. La extensión de las barras representa el percentil del 25% y 75% asociado a las diferencias estimadas para cada miembro evaluado a nivel diario en el periodo dado por el año hidrológico 2000/01. El punto al centro de cada una de las barras representa la mediana del ensemble.45

Figura 5.17: Estimación de balance hídrico para el año hidrológico 2000/2001. Las barras de error corresponden al percentil 25 y 75% derivado de los miembros del ensemble asociado a cada cuenca, mientras que el símbolo representa la mediana. (A-D) Resultados obtenidos del ejercicio de asimilación según las diferentes configuraciones propuestas.....46

1 INTRODUCCIÓN

1.1 Motivación

Las estimaciones de disponibilidad hídrica son esenciales para un buen manejo del agua y, por lo tanto, para asegurar la producción agrícola e industrial, suministro a la población y conservación de los ecosistemas locales. En este sentido, los modelos hidrológicos son herramientas que sirven de apoyo al estudio del ciclo del agua (e.g., Alvarez-Garreton et al., 2015; Noh et al., 2014), más aún cuando no se cuenta con observaciones para caracterizar ciertos flujos o estados de interés. No obstante, al ser inherentemente una simplificación de la realidad, los modelos tienen asociadas diferentes fuentes de incertidumbre, que afectan directamente su desempeño (e.g., Butts et al., 2004; Döll et al., 2016; Müller Schmied et al., 2014). Ejemplo de ello son errores en las forzantes (e.g., Elsner et al., 2014; McMillan et al., 2011), incertidumbre estructural (i.e., procesos representados, estructura de ecuaciones y parametrizaciones, conectividad espacial; e.g., Gupta et al. 2012; Clark et al., 2015) y paramétrica (e.g., Vrugt et al., 2003), inconsistencias entre observaciones utilizadas para calibrar el modelo (e.g., Duethmann et al., 2020; Montanari and Di Baldassarre, 2013), errores en estados/condiciones iniciales (e.g., Li et al., 2009), y errores residuales (e.g., Göttinger and Bárdossy, 2008; Samadi et al., 2018). Por lo tanto, resulta clave identificar, comprender, cuantificar y reducir la incertidumbre a fin de contar con resultados precisos y confiables (Liu and Gupta, 2007).

Entre las componentes que contribuyen a la incertidumbre en las estimaciones hidrológicas, una de las más importantes corresponde a errores en estados/condiciones iniciales. Diferentes estudios indican que, independiente del objetivo de modelación (i.e., consistencia con las propiedades físicas del sistema), una buena estimación de las variables de estado es clave para asegurar el buen comportamiento de un modelo (Moradkhani et al., 2005b, 2005a; Rakovec et al., 2015). Lo anterior se sustenta en que los estados (e.g., humedad de suelo, equivalente de nieve en agua, etc.) capturan la dinámica del sistema y regulan los diferentes procesos que ocurren en el mismo. Asimismo, Liu & Gupta (2007) establecen que la estimación de estados en sistemas dinámicos es un proceso donde la información adquirida desde observaciones, y acumulada en el tiempo por el modelo, se propaga a todas las variables de estado.

1.2 Avances recientes en asimilación de datos

Para reducir la incertidumbre en estimaciones hidrológicas se requiere, sin duda, de mejores modelos hidrológicos y una mayor cantidad y mejor calidad de datos hidrológicos. Pese a que los modelos hidrológicos se encuentran en constante actualización, incorporando mejoras en la representación de los procesos hidrológicos, interacción con la atmósfera, entre otros (e.g., Clark et al., 2015; Hamman et al., 2018; Mendoza et al., 2015), siguen siendo representaciones imperfectas de la realidad. Por otro lado, si bien existe una escasez de observaciones hidrológicas in-situ en Chile (Mendoza et al., 2012; Vásquez et al., 2021), especialmente en alta montaña (Ragettli et al., 2014), los avances en percepción remota (e.g., Entekhabi et al., 2010; Huffman et al., 2019; Kerr et al., 2001; Mu et al., 2011; Tapley et al., 2004), y sus proyecciones futuras respecto a este tipo de observaciones (e.g., Lettenmaier et al., 2015; McCabe et al.,

2017), surgen como una alternativa para mejorar la disponibilidad de información espacialmente distribuida y (semi) continua en el tiempo. Tomando en cuenta lo anterior, se requiere del desarrollo de técnicas para combinar la información proveniente de diferentes fuentes y asimilarla mejor (e.g., Alvarez-Garreton et al., 2016; Clark et al., 2006; De Lannoy et al., 2006; Margulis et al., 2006, 2002; Pauwels and De Lannoy, 2009; Weerts and El Serafy, 2006). Dichas técnicas se conocen como Asimilación de Datos o “*Data Assimilation*” (DA) (e.g., Liu and Gupta, 2007; Reichle, 2008), y buscan combinar la información de modelos hidrológicos imperfectos y observaciones imperfectas para cuantificar/reducir la incertidumbre en las estimaciones de diferentes variables,

Durante la última década, varios autores han explorado la posibilidad de asimilar observaciones en modelos hidrológicos con diferentes propósitos (e.g., Girotto et al., 2016; Kumar et al., 2016; Rakovec et al., 2015; Revilla-Romero et al., 2016; Zhang et al., 2017). Muchos de ellos apuntan a actualizar estados del sistema (e.g., humedad de suelo, equivalente de nieve en agua, etc.) como medio para mejorar la representación de un determinado proceso hidrológico. Más aún, la aplicación de la asimilación de datos para actualizar estados de modelos hidrológicos ha sido estudiada en extenso, y con diferentes modelos (Rasmussen et al., 2015). Así, se reconoce la necesidad de contar con técnicas que permitan cuantificar y reducir la incertidumbre asociada a la estimación de estados, los cuales favorezcan un mejor entendimiento de los procesos hidrológicos. A partir de lo anterior, se podría mejorar, por ejemplo, la gestión de los recursos hídricos en diferentes esferas (e.g., disponibilidad, amenazas asociadas a eventos extremos, pronósticos operacionales, entre otros).

Entre los estados que regulan la respuesta hidrológica a distintas escalas, resaltan el estado del manto nival y la humedad de suelo. Por un lado, la cantidad de agua almacenada en el manto nival – i.e., equivalente de nieve en agua (SWE – *Snow Water Equivalent*) – juega un rol importante por el almacenamiento estacional de agua en lugares donde las precipitaciones se concentran en el invierno, y regular la generación de escorrentía (Freudiger et al., 2017). Por otro lado, la humedad de suelo (SM – *Soil Moisture*), entendida como la cantidad de agua contenida en las capas superiores del suelo, representa la interacción directa del suelo con la atmósfera a partir de la evapotranspiración, y la partición de la lluvia entre infiltración y escorrentía superficial. Ahora bien, se debe tener en consideración que, a pesar de representar sólo el 0,0012% del agua disponible en la Tierra donde, a su vez, sólo el 0,05% de ella es agua dulce, la humedad de suelo es importante porque, además de vincular los ciclos del agua y energía, controla la ocurrencia de sequías y crecidas (De Lannoy et al., 2019).

1.3 Potencial de la asimilación en problemas de disponibilidad hídrica

Caracterizar la evolución temporal de los almacenamientos de agua, y los flujos asociados, no sólo contribuye a una mejor comprensión de las interacciones entre los procesos climáticos e hidrológicos, sino que también apoya la toma de decisiones. En términos prácticos, una forma de realizar dicha caracterización corresponde a las estimaciones de balance hídrico, las cuales dan cuenta de la partición de los flujos de precipitación (P), evapotranspiración (ET) y escorrentía (Q), y cómo estas afectan el balance de masa en un sistema hidrológico. La necesidad de contar con estimaciones de balance hídrico ha motivado diferentes estudios a nivel mundial. Como precisa Vásquez

et al. (2021), esfuerzos ligados a caracterizar la partición anual entre precipitación, evapotranspiración, escorrentía, o caracterizaciones estacionales, incluso diarias, se reportan en la literatura internacional con el objetivo de caracterizar este balance. A pesar de lo anterior, aunque se reportan algunos estudios de asimilación de datos (e.g., Bulygina and Gupta, 2009; Lohmann et al., 2004; Pan et al., 2008; Pan and Wood, 2006) orientados a evaluar el desempeño de modelos en estimaciones de balance de agua a nivel de unidad de almacenamiento (*water budget*) y los flujos asociados, no se encuentran estudios alusivos al valor agregado de la asimilación de datos en estimaciones de balance hídrico a escala de cuenca.

En Chile, a pesar de existir problemas de gestión hídrica en algunas zonas del país (e.g., Petorca; Muñoz et al., 2020) y, más aún, una persistente condición de sequía desde 2010 (Garreaud et al., 2019, 2017), no se ha explorado el potencial de las técnicas de asimilación para mejorar estimaciones de disponibilidad hídrica. Más aún, los únicos estudios con dominio en Chile donde se han aplicado técnicas de asimilación de datos a flujos o estados hidrológicos, se concentran en mejorar pronósticos (Mendoza et al., 2012) o caracterizaciones del manto nival (Cortés and Margulis, 2017). Así, hasta la fecha, no se cuenta con estudios a gran escala que permitan evaluar la utilidad de la asimilación de datos en la reducción de la incertidumbre asociada a estimaciones de balance hidrológico en Chile continental.

1.4 Objetivos

El presente estudio tiene por objetivo explorar potenciales mejoras en las estimaciones de balance hidrológico a partir de la cuantificación y reducción de la dispersión en las simulaciones hidrológicas realizadas con el modelo hidrológico *Variable Infiltration Capacity* (VIC; Liang et al., 2004, 1996; Wood et al., 1992). Para ello, se actualizan las condiciones iniciales de humedad de suelo y equivalente de nieve en agua a partir de la asimilación de datos mediante *Ensemble Kalman Filter* (EnKF; Evensen et al., 1994; Burgers et al. 1998), aplicado de manera secuencial y sincrónica. Así, se espera cuantificar el efecto de la incorporación de este enfoque en las estimaciones de 1) balance hídrico en un determinado periodo de análisis, y 2) flujos y estados hidrológicos. En términos específicos, se definen los siguientes puntos de interés:

- Implementar el esquema de asimilación de datos *Ensemble Kalman Filter* (EnKF) con el modelo hidrológico VIC.
- Evaluar el efecto de la asimilación de datos en la reducción de incertidumbre en estimaciones de flujos y estados hidrológicos.
- Evaluar diferentes configuraciones de generación de escenarios de simulación.

1.5 Estructura del documento

El resto del documento se organiza en 5 capítulos. En el Capítulo 2, se realiza una descripción de la zona de estudio; en el Capítulo 3 se presentan los datos disponibles para el desarrollo de la investigación; en el Capítulo 4 se resume la metodología empleada; en el Capítulo 5 se presentan los resultados de estudio y, finalmente, en el Capítulo 6 se resumen las principales conclusiones derivadas de la investigación y algunas ideas de trabajo futuro.

2 ZONA DE ESTUDIO

2.1 Selección de cuencas de estudio

En este estudio, se definen los siguientes criterios para la selección de cuencas:

1. **Régimen natural o cercano:** se exige un bajo o nulo grado de intervención humana en aguas superficiales. Para ello, se consideran las cuencas definidas en régimen natural en el proyecto Actualización del Balance Hídrico Nacional (DGA, 2017), las cuales cumplen con i) la suma de derechos consuntivos superficiales menor al 5% del caudal medio anual de la cuenca, y ii) la restitución de extracciones asociadas a derechos no consuntivos, en caso de existir bocatomas importantes, se ubica aguas arriba del control fluviométrico de la cuenca.
2. **Observaciones de caudal medio diario en el periodo 1985-2015:** se filtran como candidatas aquellas cuencas que tengan al menos el 90% de los días en el periodo 1985-2015 con información.
3. **Completitud a nivel mensual de registros de caudal medio diario:** se exige que los meses en el periodo 1985-2015 tengan al menos el 90% de los días con información (i.e., a lo más pueden faltar tres días).
4. **Completitud a nivel anual de registros de caudal medio diario:** entendiendo como año completo aquel que tiene todos sus meses completos, se filtran como candidatas aquellas cuencas que tengan al menos el 90% de los años en el periodo 1985-2015 completos (i.e., a lo más tres años pueden estar incompletos).

En base a los criterios definidos anteriormente, se seleccionan 12 cuencas distribuidas, como se muestra en el panel A de la Figura 2.1, entre las regiones¹ de Coquimbo y Maule.

2.2 Caracterización general de las cuencas seleccionadas

Algunas de las cuencas seleccionadas se asemejan en cuanto a ciertos atributos geomorfológicos y fisiográficos (Tabla 2.1) y/o climáticos (Tabla 2.2), lo que, a su vez, se expresa en similitudes hidrológicas entre ellas (B-D; Figura 2.1).

En términos de cobertura de suelo (Tabla 2.1), se debe notar que, al ser todas cuencas de alta montaña, predomina el suelo desnudo y, en algunos casos, matorrales o bosque nativo. Asimismo, en forma coherente a los patrones latitudinales de la cordillera de Los Andes, el rango de elevación comienza a disminuir a medida que se avanza al sur del dominio de estudio. Por ejemplo, en la cuenca 4311001 (zona B; Figura 2.1) se tiene un rango de elevación que va desde los 1.647 a 5.540 m.s.n.m., mientras que la cuenca 8104001 (zona D; Figura 2.1) exhibe un rango de elevación que va desde 684 a 2.669 m.s.n.m. En cuanto a la pendiente media de las diferentes cuencas, se tiene que varía entre 222 y 295 m/km, lo cual da cuenta valores similares entre ellas. Respecto al área de cada cuenca, hay tres que destacan por superar los 1.000 km², las cuales, según su

¹ División territorial político-administrativa de Chile, la cual actualmente considera 16 regiones.

código BNA asociado, corresponden a: 5710001, 5410002, y 4703002. El resto de las cuencas se encuentra, en su mayoría (5 de 9), por debajo de los 400 km².

En términos hidroclimáticos (Tabla 2.2), se aprecia un marcado gradiente latitudinal, con precipitaciones y escorrentía (índice de aridez) que aumentan (disminuye) a medida que se avanza latitudinalmente. Por ejemplo, en las cuencas ubicadas en la zona B (Figura 2.1) predomina el clima de tundra y semiárido de lluvia invernal, con montos anuales promedio en torno a los 280 mm/año, mientras que, a medida que se avanza latitudinalmente (zonas C y D; Figura 2.1) cambia a mediterráneo frío y/o de altura con lluvia invernal, con montos de precipitación de hasta 1.928 mm/año. Adicionalmente, el coeficiente de variación de la precipitación anual (i.e., razón entre la desviación estándar y promedio) varía entre 26%-68%, con menores valores (i.e., menor variabilidad interanual de la precipitación) hacia el sur.

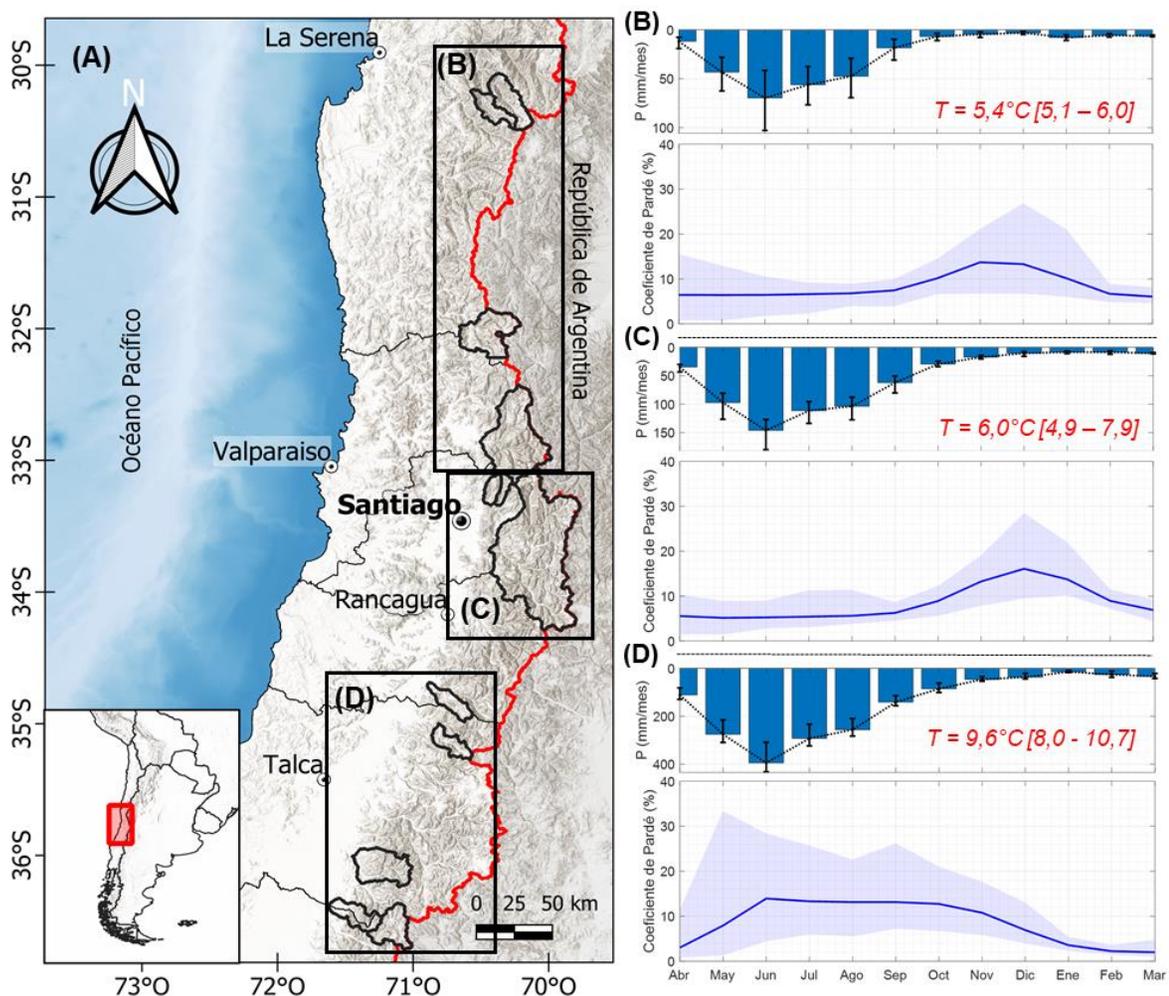


Figura 2.1: Ubicación geográfica de las cuencas de estudio y patrones hidrológicos estacionales promedio en el periodo 1985-2015. (A) Mapa de la ubicación geográfica de las 12 cuencas de estudio. (B-D) Panel superior: precipitación mensual promedio en el periodo 1985-2015 promediada entre las cuencas contenidas en el subdominio. Las barras de color negro representan valores máximos y mínimos de precipitación promedio para el subconjunto de cuencas (i.e., rango), y la línea segmentada corresponde al promedio (equivalente a barras azules). En rojo se indica la temperatura media anual promedio y su rango. Panel inferior: Coeficientes de Pardé anuales promedio en el periodo 1985-2015 entre las cuencas que componen el subdominio indicado (línea azul continua) y rango de variación (zona difuminada).

Tabla 2.1: Atributos fisiográficos de las cuencas de las 12 cuencas de estudio.

Nombre de la cuenca	Código BNA del control fluv.	Latitud* (°S)	Longitud* (°O)	Área (km ²)	Elevación mediana [mín. - máx.] (m.s.n.m.)	Pendiente media (m/km)	Cobertura de suelo primaria [secundaria]	
							Tipo	Porcentaje cubierto (%)
Estero Derecho En Alcohuz	4311001	30,221	70,493	338	3.716 [1.647 – 5.540]	255,4	Suelo desnudo [Matorrales]	69,0% [27,6%]
Río Cochiguaz En El Peñón	4313001	30,122	70,434	675	3.709 [1.338 – 5.275]	265,6	Suelo desnudo [Matorrales]	66,5% [28,7%]
Río Choapa En Cuncumén	4703002	31,967	70,594	1.132	3.322 [1.153 – 5.038]	230,9	Suelo desnudo [Matorrales]	63,6% [28,7%]
Río Aconcagua En Chacabuquito	5410002	32,850	70,509	2.113	3.322 [936 – 5.924]	279,9	Suelo desnudo [Matorrales]	66,2% [22,2%]
Río Maipo En El Manzano	5710001	33,594	70,379	4.839	3.226 [872 – 6.550]	276,0	Suelo desnudo [Pastizales]	69,5% [13,0%]
Estero Yerba Loca Antes Junta San Francisco	5721001	33,341	70,364	110	3.540 [1.348 – 5.431]	295,5	Suelo desnudo [Pastizales]	69,6% [12,3%]
Estero Arrayan En La Montosa	5722001	33,326	70,456	216	2545 [1.050 – 3.829]	242,2	Suelo desnudo [Matorrales]	50,4% [29,5%]
Río Claro En El Valle	6027001	34,687	70,874	349	1.608 [5.38 – 3.046]	222,2	Matorrales [Bosque nativo]	35,4% [26,7%]
Río Claro En Los Queñes	7103001	34,998	70,809	354	1.796 [662 – 4.077]	256,2	Suelo desnudo [Matorrales]	36,4% [32,9%]
Río Perquilauquén en San Manuel	7330001	36,376	71,623	502	1.129 [265 - 2230]	223,7	Bosque nativo [Matorrales]	48,2% [30,0%]
Río Achibueno En La Recova	7354002	36,003	71,442	894	1.273 [301 – 3.197]	249,0	Bosque nativo [Matorrales]	41,2% [26,2%]
Río Sauces Antes Junta Con Nuble	8104001	36,665	71,274	607	1.694 [684 – 2.669]	245,1	Suelo desnudo [Matorrales]	49,1% [19,8%]

*Coordenadas geográficas relativas a la ubicación del control fluviométrico que delimita la cuenca de estudio (i.e., salida u "outlet").

Fuente: Atributos obtenidos de la base de datos CAMELS-CL (Alvarez-Garretón et al., 2018)

Tabla 2.2: Características hidroclimáticas de las 12 cuencas de estudio. Se incluye el promedio anual en el periodo 1985-2015, con el rango ([min. – máx.]) asociado, del índice de aridez (IA), precipitación (P), temperatura media (Tm) y escorrentía (Q).

Nombre de la cuenca	Código BNA del control fluv.	Clasificación climática**	Régimen hidrológico	IA (-)	P (mm/año)	Tm (°C)	Q (mm/año)
Estero Derecho En Alcohuz	4311001	Clima de tundra de lluvia invernal (72%); clima semiárido de lluvia invernal (28%)	Nival	4,72 [1,0 - 12,5]	213 [57 - 654]	5,1 [4,4 - 5,9]	84 [24 - 401]
Río Cochiguaz En El Peñón	4313001	Clima de tundra de lluvia invernal (70%); clima semiárido de lluvia invernal (30%)	Nival	6,10 [1,3 - 14,8]	175 [52 - 568]	5,4 [4,6 - 6,1]	117 [36 - 448]
Río Choapa En Cuncumén	4703002	Clima de tundra de lluvia invernal (50%); clima mediterráneo de lluvia invernal de altura (49%)	Nival	3,38 [1,0 - 7,1]	326 [122 - 814]	6,0 [5,3 - 6,8]	260 [58 - 836]
Río Aconcagua En Chacabucuito	5410002	Clima de tundra de lluvia invernal (50%); clima mediterráneo de lluvia invernal de altura (27%)	Nival	2,30 [0,8 - 6,2]	408,7 [123 - 886]	5,2 [4,4 - 5,9]	459 [166 - 1033]
Río Maipo En El Manzano	5710001	Clima de tundra de lluvia invernal (49%); clima mediterráneo frío (29%)	Nival	1,08 [0,5 - 2,6]	796 [286 - 1429]	4,9 [4,2 - 5,7]	683 [354 - 1212]
Estero Yerba Loca Antes Junta San Francisco	5721001	Clima de tundra de lluvia invernal (47%); clima mediterráneo frío de lluvia invernal (28%)	Nival	1,18 [0,5 - 2,4]	541 [229 - 1089]	5,2 [4,4 - 5,9]	195 [88 - 459]
Estero Arrayan En La Montosa	5722001	Clima mediterráneo frío de lluvia invernal (42%); clima mediterráneo de lluvia invernal de altura (39%)	Nival	1,74 [0,7 - 3,5]	583 [253 - 1195]	7,9 [7,0 - 8,5]	210 [54 - 684]
Río Claro En El Valle	6027001	Clima mediterráneo de lluvia invernal de altura (46%); clima mediterráneo frío de lluvia invernal (32%)	Nivo-pluvial	0,88 [0,5 - 2,0]	1319 [543 - 2255]	10,7 [9,9 - 11,2]	762 [144 - 1550]
Río Claro En Los Queñes	7103001	Clima mediterráneo frío de lluvia invernal (43%); clima mediterráneo de lluvia invernal de altura (41%)	Pluvio-nival	0,67 [0,4 - 1,5]	1623 [654 - 2735]	8,9 [8,0 - 9,4]	1481 [563 - 2976]
Río Perquillauquén en San Manuel	7330001	Clima mediterráneo de lluvia invernal de altura (58%); clima mediterráneo de lluvia invernal (42%)	Pluvial	0,63 [0,4 - 1,4]	1928 [854 - 2796]	10,3 [9,5 - 10,7]	1782 [438 - 3167]
Río Achibueno En La Recova	7354002	Clima mediterráneo de lluvia invernal (37%); clima mediterráneo de lluvia invernal de altura (36%)	Pluvial	0,6 [0,4 - 1,5]	1877 [750 - 2857]	10,0 [9,2 - 10,5]	1351 [342 - 2742]
Río Saucos Antes Junta Con Ñuble	8104001	Clima mediterráneo de lluvia invernal de altura (60%); clima mediterráneo frío de lluvia invernal (37%)	Nivo-pluvial	0,65 [0,4 - 1,5]	1813 [749 - 2726]	8,0 [7,1 - 8,5]	1560 [430 - 2685]

**Zonas Climáticas por cuenca según clasificación de Köppen-Geiger escala 1:1.500.500, recuperadas de la Infraestructura de Datos Geoespaciales (IDE), Ministerio de Bienes Nacionales de Chile (<https://www.ide.cl/index.php/clima-y-atmosfera/item/1556-zonas-climaticas-de-chile-segun-koepfen-geiger-escala-1-1-500-000>).

3 DATOS DISPONIBLES

3.1 Registros fluviométricos

Las cuencas seleccionadas para el estudio se encuentran controladas por estaciones fluviométricas administradas por la Dirección General de Aguas de Chile (DGA). Dichos registros son de carácter público y pueden ser adquiridos directamente a través de la plataforma web de la DGA (<https://dga.mop.gob.cl/>), o bien desde bases de datos desarrolladas por instituciones académicas, como por ejemplo el Explorador Climático (<http://explorador.cr2.cl/>) o el Explorador de cuencas de CAMELS-CL (<http://camels.cr2.cl/>), ambas administradas por el Centro de Ciencias del Clima y Resiliencia – CR2.

Los registros de caudal en las estaciones de la red de monitoreo nacional se obtienen a partir de curvas de descarga, las cuales vinculan, a través de propiedades hidráulicas del escurrimiento, la altura de escurrimiento con el caudal pasante. Así, dada una sección de control (natural, natural modificada o artificial) y ciertas propiedades hidráulicas del tramo, las curvas de descarga son ajustadas esporádicamente a partir de aforos. La ecuación (3.1) muestra la forma genérica de una curva de descarga en la cual se adopta un modelo hidráulico del tipo Manning (Stowhas, 2017), donde se tiene que el caudal pasante (Q) depende de la pérdida de carga o pendiente del eje hidráulico (J), la rugosidad de la sección (n), el perímetro mojado (Ω) y el radio hidráulico (R_h). Lo anterior implica que el caudal pasante puede ser caracterizado a partir de la altura de escurrimiento (H), la pérdida de carga (J), la rugosidad (n) y la geometría de la sección, tal como se indica en la ecuación (3.2).

$$Q = \frac{\sqrt{J}}{n} \cdot \Omega \cdot R_h^{2/3} \quad (3.1)$$

$$Q = f(H, J, n, \text{geometría}) \quad (3.2)$$

De este modo, la curva de descarga sólo será invariante si ninguna de las propiedades hidráulicas cambia en el tiempo, lo cual, dada la dinámica fluvial (e.g., arrastre de sedimentos y embancamientos) resulta complejo de asegurar a priori, por lo que se requiere de mediciones periódicas de pares Q - H para validar o actualizar las curvas. En este sentido, cambios en la curva de descarga de una estación no se pueden atribuir directamente a errores en las estimaciones de caudal. La Figura 3.1 muestra, a modo de ejemplo, las curvas de descarga en el periodo 1985-2015 asociadas a la estación Estero Derecho en Alcohuaz (panel A), Río Choapa en Cuncumén (panel B), Estero Arrayán en La Montosa (panel C), y Río Sauces antes junta con Ñuble (panel D). Allí se observa que, en esos 30 años de análisis, entre 7 y 34 curvas de descarga han sido utilizadas para estimar el caudal a partir de los registros limnimétricos de cada estación. Ahora bien, dichos cambios no se pueden asociar, necesariamente, a errores en la estimación de caudales, sino que pueden estar asociados a cambios en las propiedades hidráulicas del tramo de control o cambios en la geometría de la sección por embancamiento u otros. Lo anterior considerando que las curvas de descarga se construyen con aforos que se realizan en períodos donde los caudales son bajos o normales, y no durante crecidas, por lo que las extrapolaciones de las curvas de descarga a caudales muy bajos o altos

pueden tener errores importantes. Clark et al. (2008) reconocen esta limitación en el análisis de las curvas de descarga, enfatizando además que la similitud entre curvas no implica menor incertidumbre, sino que puede estar asociada a falta de registros en el tramo (e.g., caudales altos).

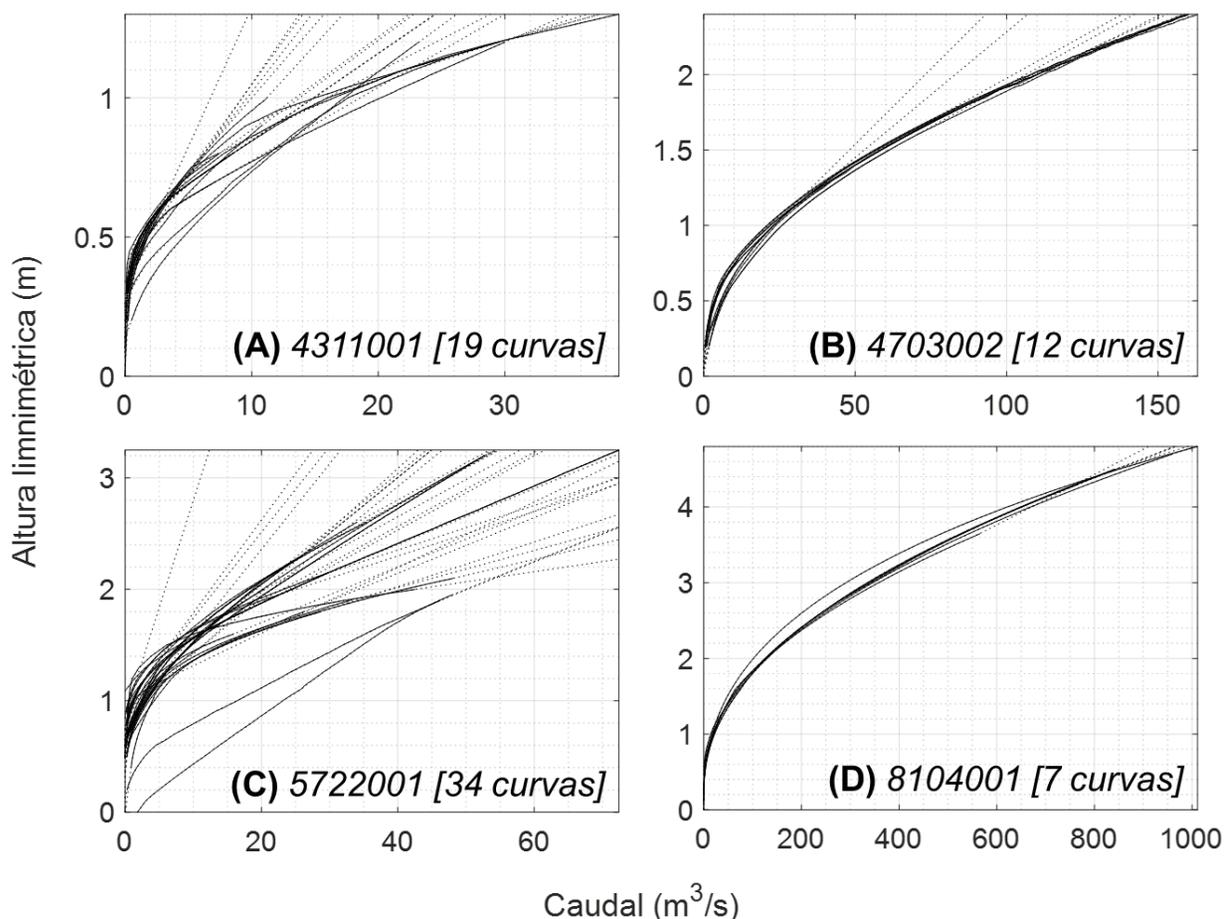


Figura 3.1: Curvas de descarga oficiales en el periodo 1985-2015 utilizadas por la Dirección General de Aguas (DGA) para el control fluviométrico. Ejemplo asociado a la estación (A) Estero Derecho en Alcohuaz – BNA 4311001, (B) Río Choapa en Cuncumén – BNA 4703002, (C) Estero Arrayán en la Montosa – BNA 5722001, y (D) Río Sauces antes junta con Ñuble – BNA 8104001. Líneas continuas corresponden a las curvas de descarga estimada por la DGA, mientras que los trazos segmentados son una extrapolación realizada para ajustar visualmente los gráficos reportados (i.e., las curvas no tienen los mismos rangos altimétricos y de caudal).

En consecuencia, dada la dificultad que implica caracterizar los errores en las observaciones de caudal a través de la curva de descarga asociada a la estación (e.g., Figura 3.1), se debe adoptar un enfoque alternativo al análisis de las curvas de descarga. Por ejemplo, Clark et al. (2008), consideraron el error de la observación proporcional a su magnitud y con distribución normal. Cabe mencionar que este punto será abordado nuevamente en la sección 4.7.

3.2 Observaciones relacionadas al manto nival

La nieve de los Andes centrales corresponde a la principal fuente de agua fresca para los valles allí emplazados, donde habita más del 40% de la población chilena (INE,

2017). Así, la caracterización del estado del manto nival resulta clave para una mejor predictibilidad hidrológica (e.g., Mendoza et al., 2014, 2020; Cornwell et al., 2016). Tomando en cuenta lo anterior, resulta de interés para el presente estudio la incorporación de algunas observaciones, espacialmente representativas, que den cuenta del estado del manto nival a resolución temporal diaria. Así, se utiliza el producto de fracción de cobertura nival (SC, SCA o fSCA) diaria a resolución horizontal de 500 metros de MODIS Terra y Aqua (Hall et al., 2002), disponibles en la base de datos del *National Snow and Ice Data Center* (NSIDC; <https://nsidc.org/>), los cuales son agregados espacialmente a escala de cuenca.

Además, se utiliza el producto de equivalente de nieve en agua (SWE) de Cortés y Margulis (2017), disponible en la base de datos CAMELS-CL (Alvarez-Garretón et al., 2018) a la escala de cuenca y a resolución temporal diaria en el periodo 01/04/1984 – 31/03/2015 (i.e., 31 años hidrológicos). Dicho producto corresponde a un reanálisis donde se asimilaron datos de fSCA, derivados de productos satelitales, en simulaciones generadas por el modelo SSiB3, forzado por MERRA, y acoplado a un modelo de curva de agotamiento del manto nival (Cortés and Margulis, 2017). La metodología ha sido testeada utilizando observaciones en Sierra Nevada (Margulis et al., 2016) y en Los Andes centrales (Cortés et al., 2016), reportando errores menores a 0,01 m en las estimaciones a posteriori de SWE. A pesar ser una ‘observación’ escasa en Chile, donde la red de estaciones SNOTEL es precaria, y entendiendo la naturaleza y limitaciones del producto de Cortés y Margulis (2017), se utiliza como referencia observacional para evaluar el valor agregado que aportaría la disponibilidad de datos de SWE a la escala de cuenca.

3.3 Evapotranspiración (ET) y humedad de suelo (SM)

Para evaluar la humedad de suelo y evapotranspiración simulados con el modelo VIC, se utilizan los valores recuperados de *Global Land Evaporation Amsterdam Model – GLEAM v3.5a* (Martens et al., 2017). El modelo, forzado por diferentes reanálisis como, por ejemplo, ERA5 y MSWEP, estima en forma separada diferentes componentes de la evaporación terrestre o evapotranspiración (e.g., transpiración, sublimación, evaporación desde suelo desnudo, etc.), basándose en observaciones satelitales que son asimiladas para corregir errores en las forzantes. Además de componentes asociadas a la evapotranspiración, el modelo genera salidas intermedias como, por ejemplo, humedad de suelo en la zona radicular, humedad de suelo superficial, entre otros. Así, tomando en cuenta que los valores de evapotranspiración y salidas intermedias de GLEAM provienen de un modelo, en este estudio se contrastan los patrones de distribución espacial de dicho producto con los obtenidos con VIC.

Los valores adquiridos desde GLEAM (<https://www.gleam.eu/>) comprenden el periodo 1980-2020, a resolución temporal diaria, y espacialmente distribuidos a 0,25° latitud-longitud. Para comparar este producto con los resultados del modelo VIC, se interpolan linealmente los valores de GLEAM al centroide de los pixeles VIC (i.e., grilla CR2Met). Si bien existen técnicas que permiten realizar un escalamiento más sofisticado, tomando en cuenta la información recuperada será utilizada sólo como referencia, se opta por el método antes mencionados por su bajo costo computacional y fácil implementación. En la Figura 3.2 se ilustran algunos resultados derivados de GLEAM

v3.5a a nivel distribuido, interpolado linealmente a una resolución de $0,05^\circ$, para una sección de Chile continental (panel A), para la cuenca Estero Arrayán en la Montosa (panel B), y algunas variables agregadas a escala de cuenca para el año hidrológico 2000/01 asociadas a la misma cuenca (panel C).

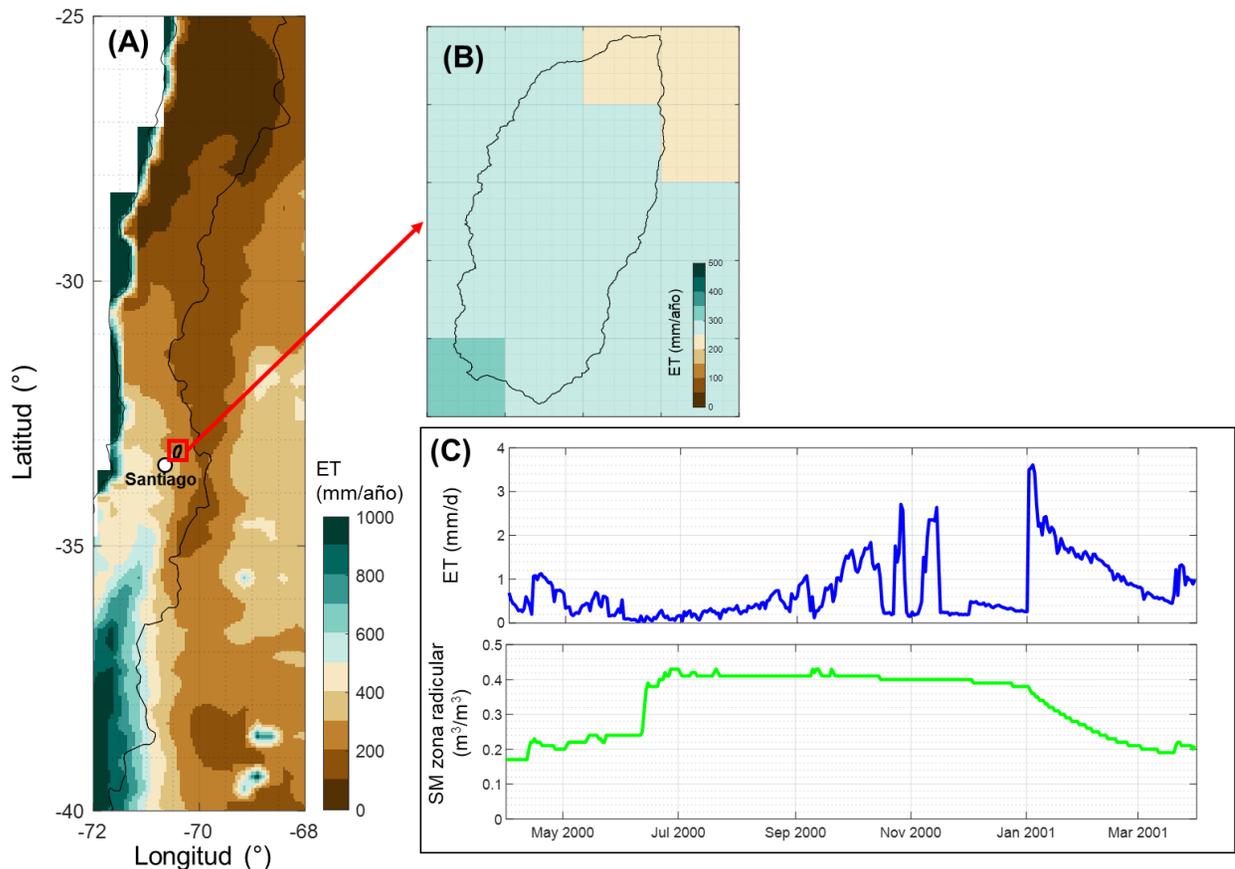


Figura 3.2: Ejemplo de los valores de referencia derivados de la interpolación de GLEAM v3.2a a la grilla de CR2Met. (A) Distribución espacial de la evapotranspiración (con resolución $0,05^\circ$ lat-lon) para el año hidrológico 2000/2001. (B) Acercamiento (*zoom in*) a la figura del panel A y ajuste en la escala de colores que da cuenta de la evapotranspiración estimada para el año hidrológico 2000/2001 en la cuenca Estero Arrayán en la Montosa distribuida espacialmente a $0,05^\circ$ lat-lon. (C) Serie de tiempo de la evapotranspiración y humedad de suelo (SM) de la zona radicular agregada a nivel de la cuenca Estero Arrayán en la Montosa (01/04/2000 – 31/03/2001).

Cabe recalcar que no se aplica ninguna corrección a los valores obtenidos desde GLEAM, por lo cual pueden existir sesgos propios del modelo y otros incorporados al momento de interpolar linealmente los valores a la grilla de interés. Sin perjuicio de lo anterior, los datos sirven como una referencia de la señal que debiese tener, por ejemplo, la evapotranspiración o humedad de suelo en la zona radicular.

4 METODOLOGÍA

En este capítulo se describen la metodología empleada y los esquemas de generación de conjuntos o *ensembles* utilizados. Así, se comienza con la descripción del modelo hidrológico VIC, las variables de entrada requeridas para generar las simulaciones y la presentación de *Shuffled Complex Evolution* (SCE-UA), algoritmo utilizado para la calibración de 14 parámetros del modelo. Posteriormente, se define el método de asimilación de datos implementado – *Ensemble Kalman Filter* (EnKF; Evensen et al., 1994; Burgers et al. 1998) – y se proponen algunas métricas de interpretación y evaluación de conjuntos. Finalmente, se plantean los esquemas de generación de conjuntos empleados para los experimentos de asimilación y se definen los errores del modelo y las observaciones, definidos a priori para el presente estudio en base a tanteo. La Figura 4.1 esquematiza, a grandes rasgos, la metodología empleada para el desarrollo de este estudio.

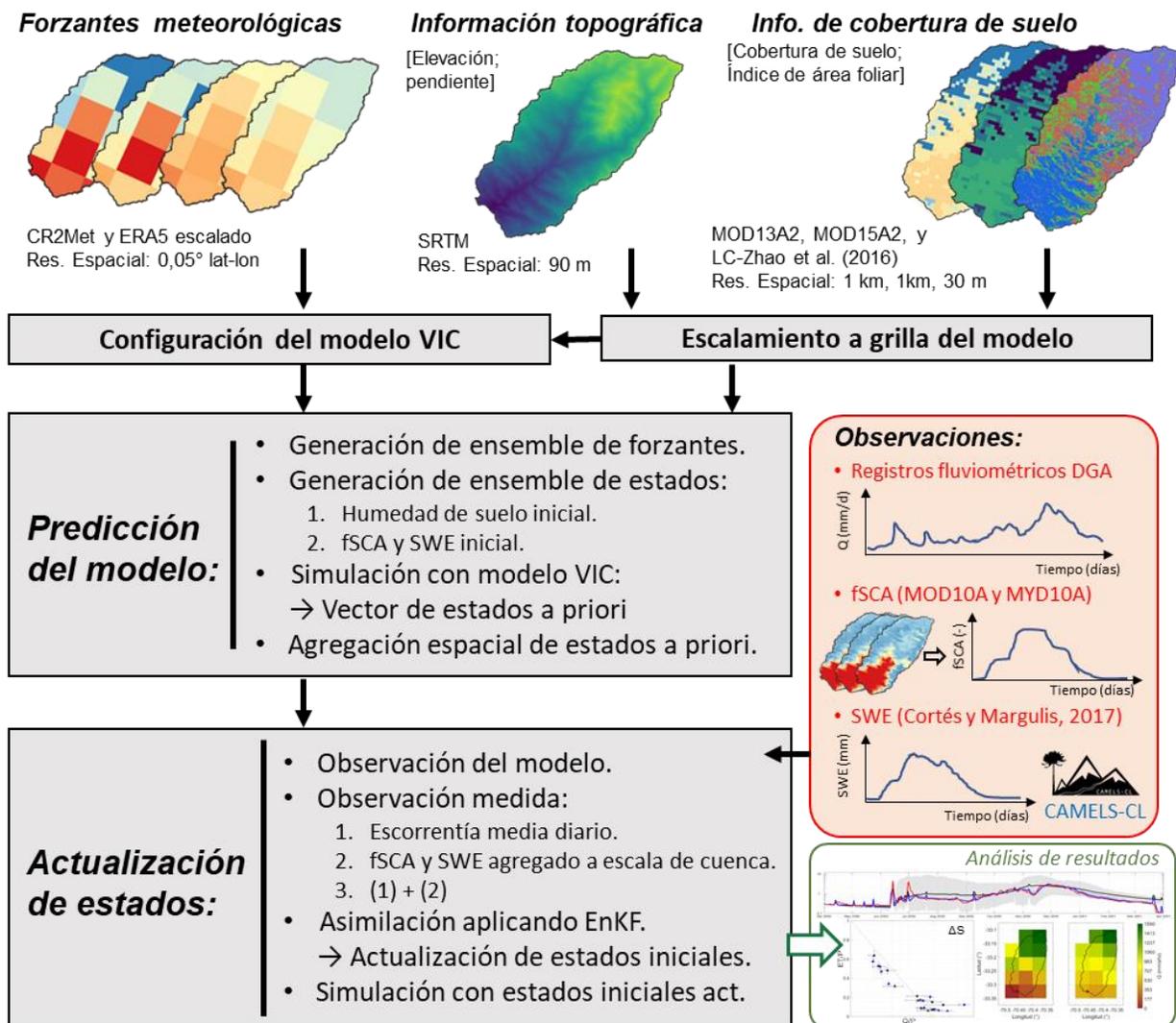


Figura 4.1: Esquema metodológico del estudio.

4.1 Modelo Hidrológico Variable Infiltration Capacity (VIC)

4.1.1 Descripción del modelo hidrológico

El modelo hidrológico corresponde al *Variable Infiltration Capacity* (VIC). VIC (Figura 4.2) es un modelo de *Land Surface*² (LSM) de macro-escala (unidades de modelación representadas por pixeles mayores a un kilómetro), motivación física (i.e., busca representar procesos de forma realista y explícita), y semi-distribuido. Este modelo también fue utilizado en el proyecto de Actualización del Balance Hídrico Nacional (DGA, 2017), por lo que su uso en el presente estudio permite evaluar el valor agregado de implementar asimilación de datos, ofreciendo oportunidades de mejora en la estimación de flujos y estados hidrológicos.

Para representar la heterogeneidad a escala de subpíxel, el modelo permite definir diferentes tipos de cobertura vegetal en cada píxel, además de bandas de elevación para dar cuenta del efecto orográfico sobre la precipitación y temperatura. Cada tipo de cobertura de suelo es representado mediante su índice de área foliar (*Leaf Area Index* - LAI), la resistencia del follaje y el porcentaje de raíces en las distintas capas de la columna de suelo. Así, en el presente estudio, para mantener coherencia metodológica con los estudios DGA (2019a, 2019b, 2018, 2017), no se utilizan bandas de elevación, se mantienen las coberturas de suelo y sus propiedades, se mantienen los módulos de balance energético y suelo congelado apagados, y se resta el aporte de origen glacial (i.e., nieve y hielo en zonas glaciadas) a los valores modelados por VIC. En resumen, sólo se modifican las forzantes y los parámetros identificados como sensibles (recalibrados).

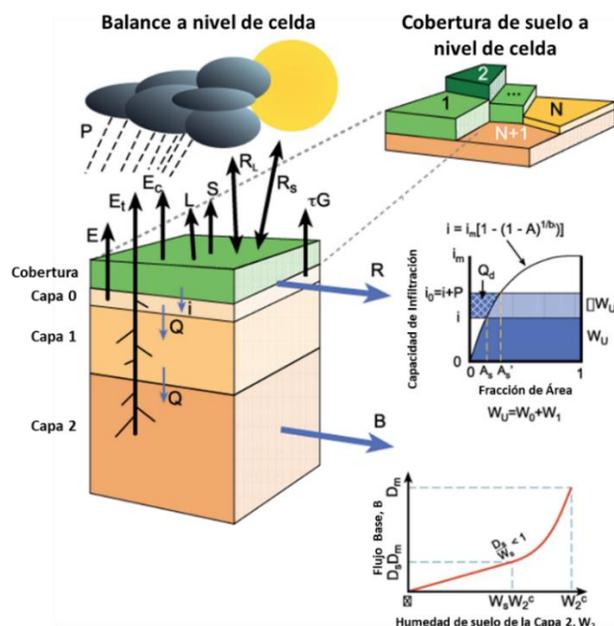


Figura 4.2: Arquitectura del modelo hidrológico Variable Infiltration Capacity (VIC).

(Adaptado de <https://vic.readthedocs.io/en/master/Overview/ModelOverview/>)

² Los *Land Surface Models* (LSM) son modelos que permiten simular y cerrar el balance de los flujos de energía (e.g., radiación solar, intercambio de calor sensible) y agua (e.g., transpiración, escorrentía) en la superficie terrestre y los estados asociados a la misma (e.g., humedad de suelo, nieve, vegetación, etc.).

4.1.2 Forzantes meteorológicas

Siguiendo el enfoque propuesto en DGA (2017), se utilizan como forzantes meteorológicas las series tri-horarias de precipitación, velocidad del viento, temperatura media y humedad relativa. Para obtener los valores precipitación y temperatura, se utiliza como base el producto grillado CR2Met versión 2.0 (DGA, 2019a), el cual incluye precipitación y temperaturas extremas (i.e., máxima y mínima) a resolución temporal diaria y espacialmente distribuida a 0,05° latitud-longitud para Chile continental. Dicho producto se combina con la distribución temporal agregada a nivel tri-horario derivada del reanálisis atmosférico ERA5 (C3S, 2017). En resumen, se considera el valor diario asociado a CR2Met v2.0 y la distribución intradiaria de ERA5. En cuanto a la velocidad del viento y la humedad relativa, ambas son obtenidas del reanálisis atmosférico ERA5 y escaladas a la grilla CR2Met según lo propuesto en DGA (2019a).

4.1.3 Calibración y diagnóstico del modelo hidrológico

La calibración de los parámetros del modelo hidrológico VIC se realiza a resolución temporal diaria utilizando el algoritmo de calibración global “*Shuffled Complex Evolution*” (Duan and Gupta, 1992). A partir de un enfoque de ponderadores aplicados de manera uniforme en las celdas de cada cuenca (Pokhrel and Gupta, 2010), se recalibran 14 parámetros identificados como los más sensibles para las estimaciones del índice de Kling-Gupta (Gupta et al., 2009) según DGA (2019b). Dichos parámetros son:

- 4 parámetros asociados a la infiltración.
- 4 parámetros de nieve asociados al albedo de nieve fresca, decaimiento del albedo y la rugosidad de la nieve.
- 2 parámetros asociados al paso de agua de una capa de suelo a otra.
- 1 parámetro asociado a la profundidad de cada capa de suelo (i.e., 3 en total).
- 1 parámetro asociado a la temperatura máxima a la cual puede llover o nevar.

En términos prácticos, el algoritmo SCE-UA se basa en el “aprendizaje” obtenido de la superficie de respuesta de una función objetivo, y se puede resumir en seis pasos:

1) Se definen S sets o conjuntos en el espacio de parámetros. De acuerdo a los N parámetros a calibrar (dimensionalidad del problema) y el número de grupos P definidos por el modelador, Duan et al. (1992) definen el tamaño óptimo de la población como $S = P \times M$, donde $M = 2 \times N + 1$. En el presente estudio, $P = 10$ y $N = 14$, por lo que $S = 10 \times (2 \times 14 + 1) = 290$.

2) Se realizan simulaciones con el modelo a partir de los S sets de parámetros, y se evalúa la función que se espera minimizar. Si se tiene un problema de maximización, es necesario reformularlo para que se deba minimizar (e.g, $\max(R^2)$ equivale a $\min(-R^2)$).

3) Los conjuntos de parámetros son ordenados de manera alternada en P grupos, de M elementos cada uno, según el valor de función objetivo. Es decir, el primer grupo tendrá la mejor función objetivo, luego la décimo primera mejor función, etc., mientras el segundo grupo tendrá la segunda mejor función objetivo, luego la décimo segunda función objetivo, y así sucesivamente.

- 4) Cada grupo genera nuevas muestras en el subespacio de parámetros y simulaciones a partir de ellas, comenzado un proceso denominado “evolución competitiva”.
- 5) Se evalúa el desempeño de los nuevos conjuntos de parámetros muestreados y se combinan con los ya existentes en cada grupo, reemplazando aquellos que tengan un valor de función objetivo mayor (recordar que el método minimiza).
- 6) Se evalúa el criterio de convergencia del algoritmo. Si converge, se termina la operación, de lo contrario se vuelve a comenzar el proceso, pero con la población S actualizada.

Por otra parte, se debe precisar que, con el fin de obtener una condición inicial estable en los almacenamientos del modelo al comienzo del periodo de calibración, se define un periodo de calentamiento (i.e., el modelo corre sin ser evaluado) de 6 años. El resto del periodo se divide entre calibración y validación. Cabe precisar que este último subperiodo tiene por finalidad evaluar el desempeño del modelo en condiciones hidroclimáticas diferentes a las cuales fue calibrado (i.e., habilidad predictiva; Klemeš, 1986). En la Tabla 4.1 se resumen los subperiodos mencionados, al igual que las fechas de inicio y final, y el número de años asociados a cada uno.

Tabla 4.1: Subperiodos de tiempo definidos para la calibración y validación del modelo hidrológico

Periodo	Inicio	Final	N° de años hidrológicos
Calentamiento	01-04-1979	31-03-1985	6
Calibración	01-04-1985	31-03-2005	20
Validación	01-04-2005	31-03-2015	10

En el presente estudio, los caudales medios diarios observados y simulados son contrastados utilizando como función objetivo la eficiencia de Kling-Gupta (KGE; Gupta et al., 2009, ecuación (4.1)). Este índice busca minimizar la distancia euclidiana entre la correlación (r ; ajuste temporal) y el sesgo en la desviación estándar (α ; ajuste de variabilidad) y promedio (β ; ajuste de volumen) simulado respecto a la observación.

$$KGE = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} ; \text{ con } \alpha = \frac{\sigma_S}{\sigma_O} \text{ y } \beta = \frac{\bar{S}}{\bar{O}} \quad (4.1)$$

Donde:

- r : Coeficiente de correlación de Pearson.
- \bar{O} : Promedio asociado a la serie de tiempo de la variable observada.
- \bar{S} : Promedio asociado a la serie de tiempo de la variable simulada.
- σ_O : Desviación estándar asociada a la serie de tiempo de la variable observada.
- σ_S : Desviación estándar asociada a la serie de tiempo de la variable simulada.

Para mayor información respecto a la configuración del algoritmo de calibración, así como de 14 los parámetros calibrados, se incluyen más detalles en el Anexo A del presente estudio.

Para diagnosticar el desempeño del modelo en términos de la representación de procesos hidrológicos, se proponen las métricas indicadas en la Tabla 4.2, las cuales mantienen la nomenclatura de la ec. (4.1). Algunas de ellas (4/7), corresponden a las propuestas por Van Werkhoven et al. (2008): i) error en la partición precipitación-escorrentía (Q/P), ii) error cuadrático medio (RMSE), iii) RMSE con transformación de Box-Cox, y iv) error en la pendiente del segmento mediodo la curva de duración (SFDCE). A su vez, se evalúa el error asociado al centro temporal (CT; Cortés et al., 2011) de la curva de variación estacional promedio (CVE), la raíz del error cuadrático medio de los coeficientes de Pardé (Pardé, 1940, 1933) y el coeficiente de correlación de Pearson (r) de la curva de variación estacional promedio. Cabe mencionar que las métricas se normalizan y redefinen en sentido positivo a modo de facilitar su interpretación. Así, valores cercanos (lejanos) a 1 (i.e., valor óptimo) representan un mejor (peor) desempeño. No obstante, su interpretación se debe realizar en términos del sesgo relativo a la referencia según la cual se normaliza la métrica.

Tabla 4.2: Métricas basadas en procesos aplicadas para el diagnóstico de la calibración del modelo hidrológico VIC.

Notación	Nombre	Ecuación	Valor Óptimo
Error Q/P	Error normalizado del coeficiente de escorrentía anual	$Error \frac{Q}{P} = 1 - \frac{abs\left(\frac{\bar{S}}{\bar{P}} - \frac{\bar{O}}{\bar{P}}\right)}{\bar{O}/\bar{P}}$	1
rRMSE	Raíz del error cuadrático medio normalizada	$rRMSE = 1 - \frac{\sqrt{\frac{1}{T} \sum_{t=1}^T (S_t - O_t)^2}}{\bar{O}}$	1
rRMSE Box-Cox	Raíz del error cuadrático medio normalizada con transformación de Box-Cox aplicada a los caudales.	$rRMSE \text{ Box - Cox} = rRMSE(Z_S, Z_O)$ $Z_Q = \frac{(1 + Q)^\lambda - 1}{\lambda}$	1
SFDCE	Error absoluto normalizado en la pendiente del segmento medio de la curva de duración (i.e., entre los perc. 30% y 70%)	$SFDCE = 1 - \frac{abs\left(\frac{S_{70} - S_{30}}{40} - \frac{O_{70} - O_{30}}{40}\right)}{(O_{70} - O_{30})/40}$	1
Error CT	Error normalizado en el centroide o centro de temporalidad de la curva de variación estacional.	$Error \text{ CT} = 1 - \frac{CT_S - CT_O}{CT_O}$ $CT_Q = \frac{\sum_{i=1}^{12} Q_i \cdot t_i}{\sum_{i=1}^{12} Q_i} \text{ (meses)}$	1
Pardé RMSE	Raíz del error cuadrático medio normalizado de los coeficientes de Pardé	$Pardé \text{ RMSE} = 1 - \sqrt{\frac{1}{12} \sum_{t=1}^{12} \frac{(Pardé_s - Pardé_o)^2}{Pardé_o}}$ $Pardé_{Q,i} = \frac{Q_i}{\sum_{i=1}^{12} Q_i} \text{ (meses)}$	1

Adicionalmente, en la Tabla 4.2 el operador “ $\overline{(\cdot)}$ ” representa el promedio temporal de la variable en cuestión, mientras que los subíndices del tipo X_{70} y X_{30} representan los percentiles 70% y 30% de la curva de duración. Por otro lado, el parámetro “ λ ”, que define la transformación de Box-Cox, se estima a modo de asegurar que la suma de la asimetría y kurtosis absolutas de la variable transformada sea cercana a cero. Cabe mencionar que en el análisis se consideran años hidrológicos (i.e., abril a marzo) y que las métricas se aplican para evaluar el desempeño del modelo en términos de la escurrentía media diaria.

4.2 Asimilación de datos – Ensemble Kalman Filter (EnKF)

El problema de asimilación de datos o *data assimilation* (DA) tiene por objetivo mejorar la estimación de, por ejemplo, estados de modelos imperfectos a partir de observaciones imperfectas. En consecuencia, resulta clave entender el funcionamiento del modelo desde la perspectiva espacio-estado. En general, se puede plantear una ecuación de la forma presentada en (4.2) para representar en forma discreta la función de transferencia de estados, donde el operador $f(\cdot)$ representa la función que gobierna la evolución temporal del vector de estados x (i.e., un modelo hidrológico), y que depende de la condición inicial de los mismos, las forzantes “ u ” y el vector de parámetros θ^f . En términos prácticos, $f(\cdot)$ vincula las salidas del modelo y el espacio-estado (e.g., humedad del suelo). Además, se debe reconocer la naturaleza imperfecta del modelo, por lo que se agrega el error aditivo “ ω ”, el cual se denomina “error del modelo”.

$$x_t = f(x_{t-1}, u_t, \theta^f) + \omega_t \quad (4.2)$$

De este modo, el objetivo de la asimilación de datos radica en reducir el error de modelo para mejorar la consistencia del estado y/o flujo estimado (Moradkhani et al., 2019). A su vez, se debe tener en cuenta que los estados del sistema condicionan los flujos observados en cada paso de tiempo por lo que, al igual que el operador de estados, se define el operador de observaciones $h(\cdot)$, el cual permite vincular el espacio de estados con el espacio de observaciones (e.g., caudal). Como se aprecia en la ecuación (4.3), el vector de observaciones depende del vector de estados “ x ” y del vector de parámetros “ θ^h ”. En forma análoga a (4.2), se considera un error aditivo asociado a la observación, el cual puede ser entendido como la diferencia para recuperar el valor “real” o “verdadero” (i.e., medido) de la observación (Moradkhani et al. 2019).

$$y_t = h(x_t, \theta^h) + \varepsilon_t \quad (4.3)$$

En términos simples, el problema de asimilación de datos es el proceso de mejoramiento los estados del sistema reduciendo la diferencia entre la observación del modelo y el valor real o medido, dada la disponibilidad de dicha medición (i.e., los estados quedan condicionados por las observaciones). Así, el problema se puede expresar, en su forma más general, según el teorema de Bayes aplicado al problema de estimación de los estados de un modelo, condicionado por las observaciones disponibles:

$$p(x_{1:t}|y_{1:t}, u_{1:t}, \theta^f, \theta^h) \propto p_h(y_{1:t}|x_{1:t}, \theta^h)p_f(x_{1:t}|u_{1:t}, \theta^f) \quad (4.4)$$

En la ecuación anterior, p_h y p_f corresponden a la función de densidad de probabilidad (FDP o PDF por sus siglas en inglés) del espacio de estados y observaciones

respectivamente. Tomando en cuenta la forma en que se aborda la solución de la ecuación (4.4), se pueden distinguir dos enfoques de asimilación de datos: i) suavizados (*smoothers*), y ii) filtros (*filters*). En el primer caso, se resuelve completamente el problema de asimilación propuesto en la ecuación (4.4), a partir de la minimización de una función de costo para actualizar estados dentro de una ventana de tiempo, mientras que, en el segundo, se realiza una aproximación markoviana del sistema dinámico modelado, la cual se expresa explícitamente a través de la probabilidad a priori (*prior probability*) del estado “ p_f ” (Moradkhani et al., 2019), y se asimilan observaciones “pasadas”. Un ejemplo de aplicación para ambos métodos serían los reanálisis y pronósticos operacionales, respectivamente (Rakovec et al., 2015). Así, en los filtros se supone que una observación en el tiempo “ t ” tendrá efectos sólo hasta los estados observados en el tiempo “ $t-s$ ”, con $s>0$. Tomando en cuenta las simplificaciones que conllevan los métodos basados en filtros y el menor costo computacional asociado, en el presente estudio se utiliza dicho enfoque suponiendo, a su vez, que la influencia de las observaciones en el tiempo “ t ” es sólo hasta el paso de tiempo anterior (i.e., “ $t-1$ ”), según lo cual la expresión (4.4) se puede reescribir como (4.5).

$$p(x_{1:t}|y_{1:t}, u_{1:t}, \theta^f, \theta^h) \propto p_h(y_t|x_t, \theta^h)p_f(x_t|x_{t-1}, u_t, \theta^f, y_{1:(t-1)}) \quad (4.5)$$

El método de asimilación implementado en el presente estudio corresponde al *Ensemble Kalman Filter* (EnKF; Evensen, 2009) o Filtro de Kalman de Conjuntos, cuya idoneidad para trabajar con modelos geofísicos no-lineales ya ha sido demostrada en estudios anteriores (e.g., Emerick and Reynolds, 2012; Ibaceta et al., 2020; Rasmussen et al., 2015; Siripatana et al., 2019). El método se puede dividir principalmente en dos pasos: 1) predicción del modelo, y 2) actualización de estados. El primer paso corresponde a la generación del vector de estados a priori a partir de la función de transferencia “ f ” y el vector de observaciones simuladas, generado a partir del operador de observaciones “ h ”. Así, las ecuaciones (4.6) y (4.7) muestran la forma general en que se plantean el vector de estados a priori y observaciones a priori, respectivamente:

$$\hat{x}_{t,i}^- = f(\hat{x}_{t-1,i}^+, u_{t,i}, \theta) + \omega_{t,i} \quad (4.6)$$

$$\hat{y}_{t,i} = h(\hat{x}_{t,i}^-, \theta) \quad (4.7)$$

$$y_{t,i} = y_t + \varepsilon_{t,i} \quad (4.8)$$

Se debe notar que, al vector de estados, se agrega un término asociado al error del modelo, el cual será descrito y profundizado en la sección 4.3. Luego, dada una observación “ y_t ”, se inicia el paso de actualización de estados. Finalmente, como se muestra en la expresión (4.9), se realiza la actualización del vector de estados (o cualquier variable de interés según sea el caso), donde “ K ” corresponde a la ganancia de Kalman y se calcula según la expresión (4.10). Cabe mencionar que la notación del tipo “ C_{ab} ” hace referencia a la matriz de covarianza entre la variable “ a ” y “ b ”.

$$\hat{x}_{t,i}^+ = \hat{x}_{t,i}^- + K(y_{t,i} - \hat{y}_{t,i}) \quad (4.9)$$

$$K = C_{xy}(C_{yy} + \Sigma_0)^{-1} \quad (4.10)$$

4.3 Error del modelo y generación de conjuntos (*ensembles*)

El *Ensemble Kalman Filter* (EnKF) utiliza la varianza entre los múltiples miembros del conjunto (*ensemble members*) para cuantificar el error del modelo, los cuales se generan por la perturbación estocástica de diferentes componentes del modelo (Clark et al., 2008). Así, dichas perturbaciones dan cuenta de la incertidumbre intrínseca a cada una de las componentes perturbadas (e.g., condición inicial, forzantes, estructura). Resulta importante notar que la incertidumbre o error del modelo puede ser tratada de manera individual (i.e., desagregada por fuente de incertidumbre) o concentrada como un término de “incertidumbre total”.

En el caso puntual del presente estudio, el error del modelo se trabaja desagregado por componente, bajo el supuesto de que las deficiencias en las simulaciones, cuantificadas a partir de las diferencias entre la escorrentía observada y simulada en la salida de cada cuenca, se explican principalmente por errores en las forzantes y condiciones iniciales de humedad de suelo y SWE. De este modo, se debe definir un esquema a través del cual se pueda perturbar individualmente cada componente de interés, con valores que permitan generar un “buen” conjunto.

Las secciones que siguen a continuación resumen la metodología empleada para perturbar las forzantes y condiciones iniciales de humedad de suelo y SWE del modelo VIC en cada paso de tiempo (i.e., secuencialmente), así como el algoritmo empleado para estimar los errores del modelo y observaciones, y la definición de aquello que se entenderá como “buen conjunto”.

4.3.1 Perturbación de forzantes y estados

Por la naturaleza de cada una de las forzantes meteorológicas de VIC (i.e., precipitación, velocidad del viento, temperatura media, y humedad relativa), y a modo de no agregar sesgos artificiales, el error se define en términos multiplicativos o aditivos según sea el caso. Así, la perturbación de la precipitación, velocidad de viento y humedad relativa, se define según un término multiplicativo, como se muestra en las expresiones (4.11) y (4.12), las que ejemplifican el caso de la precipitación. Allí, el término “ ε_P ” representa el error asociado a la precipitación y “ u_P ” un número aleatorio que sigue una distribución uniforme.

$$P' = \varphi_P \cdot P \quad (4.11)$$

$$\varphi_P = (1 - \varepsilon_P) + 2u_P\varepsilon_P \quad (4.12)$$

En las ecuaciones anteriores, se tienen N valores de u_P , que corresponde al número de miembros del conjunto definido para el estudio. Para el caso de la temperatura, se adopta un enfoque aditivo como se muestra en las expresiones (4.13) y (4.14), donde “ ε_T ” muestra el error asociado a la temperatura y “ u_T ” un número aleatorio que sigue una distribución uniforme.

$$T' = T + \varphi_T \quad (4.13)$$

$$\varphi_T = -\varepsilon_T + 2u_T\varepsilon_T \quad (4.14)$$

En consecuencia, las forzantes se perturban según cuatro términos de error: ε_P (error en precipitación), ε_u (error en viento), ε_T (error en temperatura), y ε_{HR} (error en humedad relativa). Se debe notar que, para el caso de las forzantes, se aplica una perturbación espacialmente homogénea (i.e., el mismo número aleatorio por variable para todos los píxeles del dominio) a modo de preservar los patrones de distribución espacial obtenidos por construcción. A su vez, se supone que no existe correlación temporal entre las perturbaciones generadas a las forzantes.

Por otro lado, el error del modelo se caracteriza a partir de la condición inicial de la simulación para un determinado tiempo “t”. El supuesto en la formulación del esquema radica en que el caudal observado en el tiempo “t” depende principalmente del estado observado al final de la simulación en el paso de tiempo anterior (además de las forzantes del modelo). Así, a partir de las expresiones (4.15) y (4.16), y la magnitud del error “ ε_S ”, se genera el conjunto de condiciones iniciales. Es importante precisar que el enfoque propuesto para la perturbación de estados se aplica a la humedad de suelo, desagregada por capa, y al SWE inicial según sea el caso.

$$S'_0 = S_0 + \varphi_S \quad (4.15)$$

$$\varphi_S = -\varepsilon_S |S_0(t) - S_0(t-1)| + 2u_S \varepsilon_S |S_0(t) - S_0(t-1)| \quad (4.16)$$

El término “u” que acompaña a las ecuaciones (4.12), (4.14) y (4.16) proviene de una distribución uniforme, que se estima a partir de la probabilidad acumulada de una distribución normal estándar “s” utilizando la transformación definida en la ecuación (4.17), donde “*erfc()*” es la función de error complementario y “u” varía entre cero y uno.

$$u = \frac{1}{2} \operatorname{erfc} \left(\frac{s}{\sqrt{2}} \right) \quad (4.17)$$

Considerando la covariabilidad que puede existir entre variables, el muestreo normal estándar “s”, que da origen el término “u”, se obtiene a través de una distribución normal multivariada con media cero y covarianza igual a la matriz de correlación entre forzantes o estados según sea el caso de análisis (i.e., ambas fuentes de incertidumbre se analizan por separado). Lo anterior, permite preservar la correlación que existe entre las variables al momento de generar los números aleatorios “u”. Así, como se esquematiza en la Figura 4.3 para una cuenca de nueve píxeles (i.e., $N_p = 9$), tomando como ejemplo el caso de las forzantes (i.e., precipitación (P), velocidad del viento (u), temperatura (T) y humedad relativa (HR)), para cada paso de tiempo “t=j” se procede de la siguiente forma:

1. Se construyen $N_v = 4$ vectores “ V_j ” de nueve (9) elementos (i.e., número de píxeles) asociados a cada una de las forzantes (i.e., $j = 1, 2, 3, 4$).
2. (i) Se estima el coeficiente de correlación de Pearson (r) entre los vectores “ V_j ” (i.e., $r_{(P,P)}$, $r_{(P,u)}$, etc.). (ii) A partir de los coeficientes de correlación, se construye la matriz simétrica “ r_C ” con dimensiones $N_v \times N_v$. (iii) Se define la matriz “S”, la cual sigue una distribución normal multivariada de dimensión $\text{ndim} = N_v$ con media cero y varianza “ r_C ”, donde las columnas mantienen el orden en que se asignan los r a la matriz r_C .

Con la distribución de la matriz “S”, se genera una muestra de N números aleatorios, donde N el tamaño del conjunto. En consecuencia, para este ejemplo, se obtiene una matriz de dimensión N x 4, donde las columnas (i.e., variables) mantienen la correlación original entre sí, y los elementos siguen una distribución normal. Así, se aplica a “S” la ec. (4.17), según lo cual se obtiene la matriz $U = [u_{V1} \dots u_{VN}]$, de dimensiones N x 4 en el ejemplo. Dicha matriz cumple con la propiedad de mantener la correlación entre columnas, pero los valores se distribuyen uniformemente. Se debe notar que cada fila de la matriz U representa un escenario de perturbación de las forzantes del modelo. Por ejemplo, para el primer escenario, por construcción, el elemento (1,1) de la matriz corresponderá a u_p , (1,2) a u_u , (1,3) a u_T y (1,4) a u_{HR} . Las filas que siguen (i.e., miembros del conjunto) mantienen la notación. El procedimiento se repite en forma análoga para los estados del modelo (i.e., humedad de suelo en las capas 1, 2 y 3, y SWE), donde se toman en cuenta los valores obtenidos durante la simulación determinística (i.e., con forzantes originales) para estimar las matrices de correlación en cada paso de tiempo.

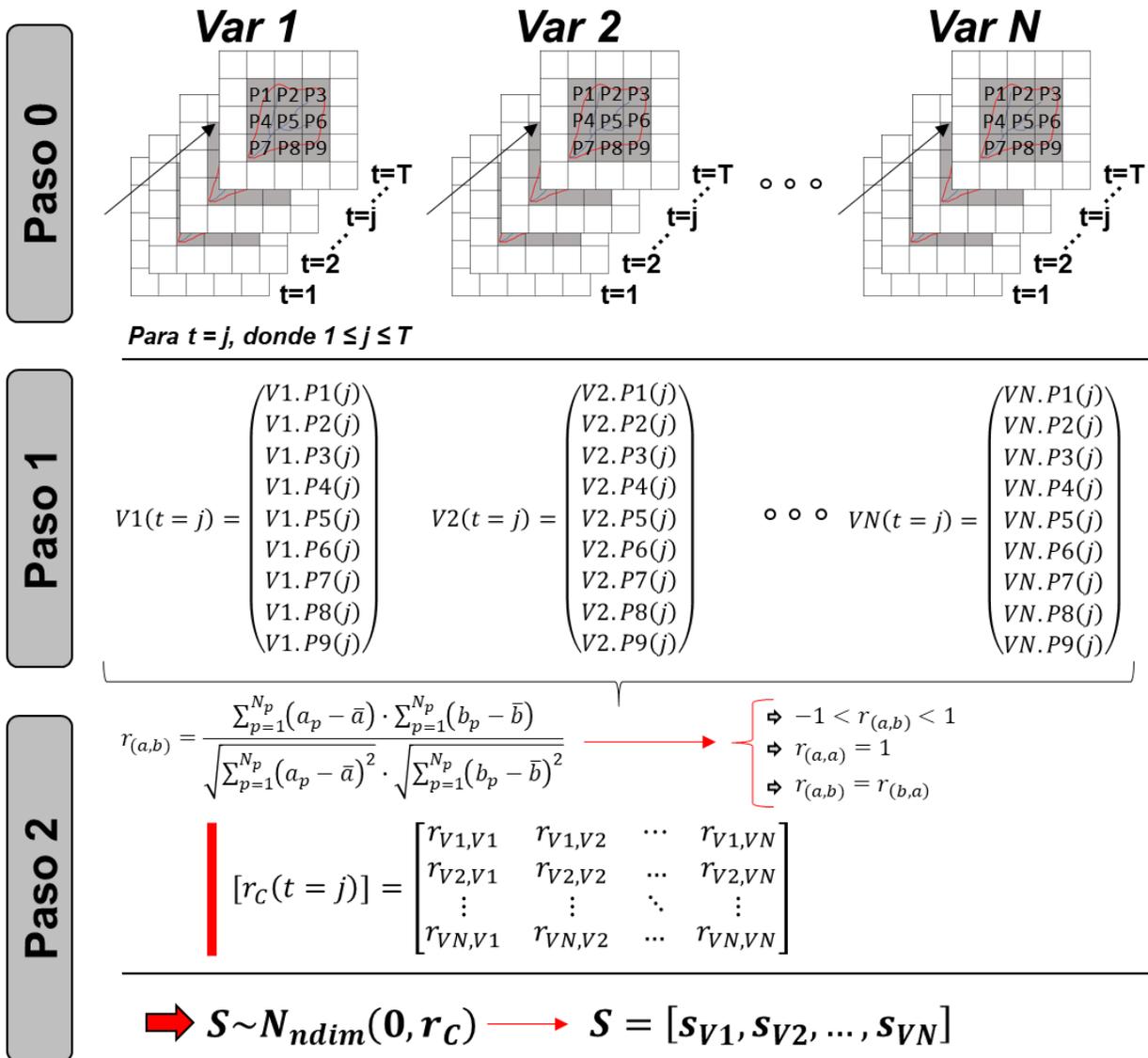


Figura 4.3: Esquema metodológico de la estimación de matrices de covarianza de forzantes y estados del sistema. Ejemplo aplicado a una cuenca que ocupa nueve (9) píxeles de la grilla espacial con la cual se configura el modelo hidrológico.

4.4 Interpretación y verificación de conjuntos

En estadística, existen diferentes indicadores para caracterizar un conjunto de datos, como por ejemplo el promedio, la dispersión, asimetría, entre otros (Tabla 4.3). En el caso de los *ensembles* o conjuntos, dichos criterios se pueden aplicar para cada paso de tiempo “i” para facilitar la interpretación de los miembros del conjunto.

Tabla 4.3: Estadísticos y métricas para la interpretación de conjuntos. El subíndice i representa cada paso de tiempo.

Nombre	Fórmula	Descripción
Promedio del conjunto	$\bar{x}_i = \frac{1}{N} \sum_{k=1}^N \hat{x}_{i,k}$	Medida de la tendencia central de la variable estudiada.
<i>Ensemble Spread</i> o Dispersión del conjunto	$ensp_i = \frac{1}{N} \sum_{k=1}^N (\hat{x}_{i,k} - \bar{x}_i)^2$	Medida de la dispersión o variación respecto al promedio de los N miembros del conjunto
Error cuadrático medio del conjunto	$mse_i = \frac{1}{N} \sum_{k=1}^N (\hat{x}_{i,k} - y_i)^2$	Medida estadística de la diferencia cuadrática de los miembros del conjunto y el valor observado de la variable de interés. Una forma análoga del estadístico corresponde al <i>rmse</i> (<i>root mean square error</i>), donde se aplica raíz cuadrada al <i>mse</i> .
Ensemble Skill o Habilidad del conjunto	$ensk_i = (\bar{x}_i - y_i)^2$	Medida estadística de la diferencia cuadrática entre el promedio del conjunto (i.e., valor esperado) y el valor observado para la variable de interés

Además, diferentes autores (e.g., Alvarez-Garreton et al., 2014; De Lannoy et al., 2006; Pauwels and De Lannoy, 2009) proponen criterios para la verificación de los conjuntos, entre los cuales destacan la consistencia e indistinguibilidad estadística (Tabla 4.4). En consecuencia, se define el “buen conjunto” como aquel que muestra resultados de consistencia estadística e indistinguibilidad normalizada cercanos a la unidad.

Tabla 4.4: Criterios para la verificación de conjuntos.

Nombre	Fórmula	Descripción
Consistencia estadística del conjunto	$\frac{\langle ensk \rangle}{\langle ensp \rangle} = \frac{\frac{1}{T} \sum_{i=1}^T ensk_i}{\frac{1}{T} \sum_{i=1}^T ensp_i}$	Valor óptimo igual a 1. Valores mayores a 1 indican presencia de sesgo o poca dispersión del conjunto.
Razón del promedio temporal de la raíz de la habilidad del conjunto y el promedio temporal del <i>rmse</i> de cada miembro	$\frac{\langle \sqrt{ensk} \rangle}{\langle \sqrt{mse} \rangle} = \frac{\frac{1}{T} \sum_{i=1}^T \sqrt{ensk_i}}{\frac{1}{T} \sum_{i=1}^T \sqrt{mse_i}}$	Valor óptimo igual a $\sqrt{(N+1)/2N}$ para asegurar indistinguibilidad estadística de los miembros del conjunto.

En términos prácticos, la consistencia se puede entender como una medida de la razón entre cuánto se aleja el promedio del conjunto respecto a la observación y cuán disperso son los miembros del conjunto respecto a su promedio (i.e., coherencia entre

los errores conjunto-observación y conjunto-conjunto). Así, valores menores (mayores) a 1 estarán asociados a conjuntos con una dispersión mayor (menor) al sesgo de su promedio respecto a la observación. Por otro lado, para evaluar cuánto se acercan los miembros del conjunto a las observaciones, se utiliza el criterio de indistinguibilidad estadística, el cual corresponde a una medida de la razón entre cuánto se aleja el promedio del conjunto de la observación y cuánto se aleja cada miembro del mismo respecto a la observación. En algunos estudios (e.g., Alvarez-Garretón et al., 2015), el criterio antes mencionado se divide por el valor óptimo (i.e., $\sqrt{(N + 1)/2N}$, con N igual al tamaño del conjunto), para redefinir el óptimo igual a la unidad (i.e., 1), mientras que valores mayores o menores podrían asociarse a sesgos.

Para evaluar la confiabilidad (“*reliability*”) del conjunto, en forma complementaria a los estadísticos de interpretación (Tabla 4.3) y criterios de verificación (Tabla 4.4), se utilizan diagramas de Talagrand, también conocidos como histogramas de rango (Hamill, 2001). La Figura 4.4 esquematiza posibles resultados que se pueden obtener con los histogramas de rango y su interpretación. Así, cuanto más plano sea el histograma de rango (i.e., todas las barras muestran la misma frecuencia), mejor podría ser el conjunto. No obstante, por sí solo, el histograma de rango no entrega información sobre la habilidad del conjunto o conjunto, por lo que un histograma plano (i.e., rangos uniformes) no asegura que los resultados sean “buenos”. En dicho sentido, los histogramas de rango serán útiles para concluir sobre la distribución de los miembros del conjunto y detectar posibles sesgos.

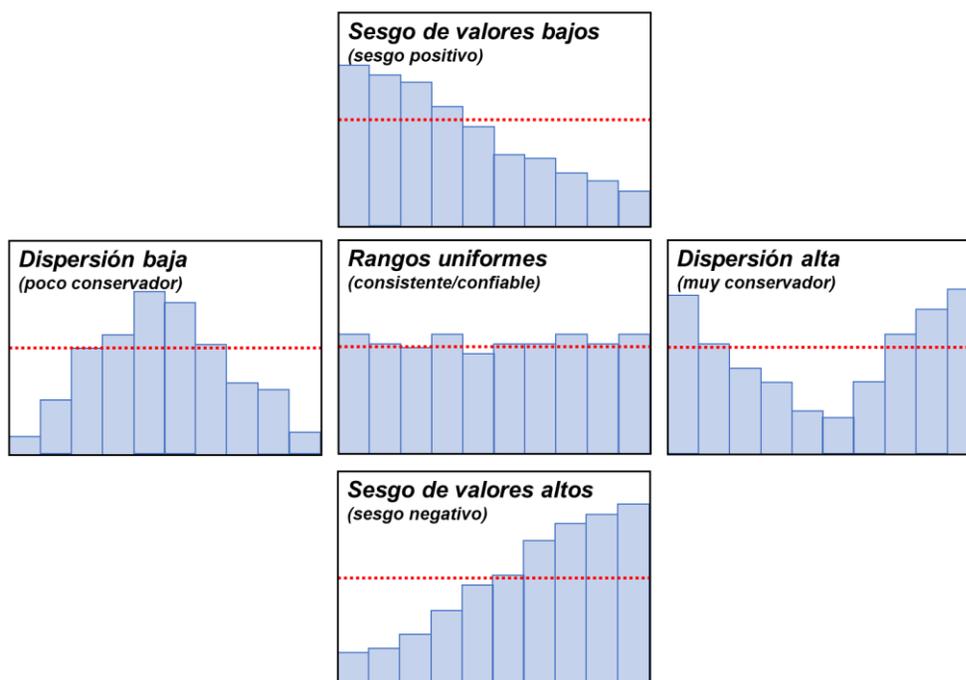


Figura 4.4: Ejemplo de interpretación de histogramas de rango.

De manera complementaria, se puede aplicar el índice de confiabilidad (RI – *Reliability Index*; Delle Monache et al. 2006) como se muestra en la ecuación (4.18), donde “ N_{Rangos} ” corresponde al número de intervalos definidos para construir el histograma, “ $conteo_i$ ” es el número de veces en que el evento observado cae dentro del i -ésimo rango, “ N_{puntos} ” es la suma de “ $conteo_i$ ”, para $i=1, \dots, N_{Rangos}$.

$$RI = \sum_{i=1}^{N_{Rangos}} \left| \frac{\text{conteo}_i}{N_{puntos}} - \frac{1}{N_{Rangos}} \right| \times 100 \quad (4.18)$$

Dada su formulación, valores de RI menores implican que los rangos están más cerca de representar frecuencias asociadas a la confiabilidad perfecta. No obstante, el RI no provee información asociada a la forma del histograma de rango, por lo cual no reemplaza su aplicación (Delle Monache et al. 2006). Para evitar confusiones semánticas sobre la interpretación del índice RI, el cual podría asociarse erróneamente a índices de confiabilidad (e.g., BS; Murphy, 1986) o diagramas de confiabilidad (Wilks, 2011), se adopta el término Índice de Discrepancia (DI – *Discrepancy Index*; Mendoza et al., 2015).

4.5 Evaluación de la asimilación de datos

Considerando los datos disponibles para contrastar los resultados derivados la simulación sin asimilación (*open loop*, OL) y con asimilación (DA), y su naturaleza (i.e., observación in-situ/remota, resultados de modelo, etc.), se presenta la metodología para evaluar variables simuladas tanto agregadas a escala de cuenca como distribuidas.

4.5.1 Evaluación de flujos y estados agregados a escala de cuenca

Para evaluar los resultados de simulaciones OL y DA se utilizan cuatro (4) métricas derivadas de la serie de valores diarios (Tabla 4.5). Cada métrica se calcula para cada miembro de conjunto de tamaño N. Así, para una determinada métrica “X”, la notación “mX” hará referencia al promedio de X entre los N miembros del conjunto.

Tabla 4.5: Criterios de eficiencia definidos para la evaluación de las simulaciones OL y DA.

Notación	Nombre	Ecuación	Valor Óptimo; Rango
KGE	Eficiencia de Kling-Gupta (Gupta et al., 2009)	$KGE = 1 - \sqrt{(r - 1)^2 + \left(\frac{\sigma_S}{\sigma_O} - 1\right)^2 + \left(\frac{\bar{S}}{\bar{O}} - 1\right)^2}$	1 ; (-∞, 1]
NSE	Eficiencia de Nash-Sutcliffe (Nash and Sutcliffe, 1970)	$NSE = 1 - \frac{\sum_{t=1}^T (S_t - O_t)^2}{\sum_{t=1}^T (O_t - \bar{O})^2}$	1 ; (-∞, 1]
LNSE	NSE con valores en escala log.	$LNSE = 1 - \frac{\sum_{t=1}^T (\ln(S_t) - \ln(O_t))^2}{\sum_{t=1}^T (\ln(O_t) - \ln(\bar{O}))^2}$	1 ; (-∞, 1]
RMSE	Raíz del error cuadrático medio	$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (S_t - O_t)^2}$	0 ; [0, ∞)

Donde:

- r : Coeficiente de correlación de Pearson.
- O_t : Variable observada en el tiempo t .
- S_t : Variable simulada en el tiempo t .

- \bar{O} : Promedio asociado a la serie de tiempo de la variable observada.
 \bar{S} : Promedio asociado a la serie de tiempo de la variable simulada.
 σ_o : Desviación estándar asociada a la serie de tiempo de la variable observada.
 σ_s : Desviación estándar asociada a la serie de tiempo de la variable simulada.
 T : Extensión temporal del registro (i.e., número de pasos de tiempo evaluados).

Con el objetivo de homologar la interpretación de las métricas presentadas en la Tabla 4.5, se redefine el RMSE a partir de la normalización de la raíz del error cuadrático medio por el promedio observado (\bar{O}). Así, el RMSE se reescribe según la ecuación (4.19), donde la métrica “ $rRMSE$ ” toma valores entre infinito negativo y uno (1; óptimo).

$$rRMSE = 1 - \frac{RMSE}{\bar{O}} \quad (4.19)$$

Para evaluar el valor agregado de aplicar DA respecto a las simulaciones en OL, se utiliza la diferencia en la raíz del error cuadrático medio normalizado (NRMSE) definido en Álvarez-Garretón et al. (2015), el cual se escribe según la ecuación (4.20). Así, valores menores a 1 indican mejoras en los resultados respecto al caso base (OL).

$$NRMSE = \frac{\frac{1}{N} \sum_{i=1}^N \sqrt{\frac{1}{T} \sum_{t=1}^T (S_t^{DA} - O_t)^2}}{\frac{1}{N} \sum_{i=1}^N \sqrt{\frac{1}{T} \sum_{t=1}^T (S_t^{OL} - O_t)^2}} = \frac{mRMSE^{DA}}{mRMSE^{OL}} \quad (4.20)$$

Adicionalmente, se propone evaluar la ganancia normalizada del resto de las métricas a partir de la expresión (4.21), donde M_{Opt} corresponde al valor óptimo que puede adoptar la métrica M (e.g., $M_{Opt} = 1$ para KGE). En consecuencia, la ecuación (4.21) apunta a evaluar cuánto mejora la métrica respecto a su mejora potencial y se interpreta en forma análoga al NRMSE.

$$M^* = \frac{M_{Opt} - \frac{1}{N} \sum_{i=1}^N M_i^{DA}}{M_{Opt} - \frac{1}{N} \sum_{i=1}^N M_i^{OL}} = \frac{M_{Opt} - mM^{DA}}{M_{Opt} - mM^{OL}} \quad (4.21)$$

Salvo la métrica LNSE, por la transformación logarítmica que involucra, el resto de los criterios presentados en la Tabla 4.5 son evaluados para la cobertura de nieve fraccional (fSCA) y equivalente de nieve en agua (SWE) agregados a escala de cuenca.

4.5.2 Evaluación respecto a observaciones de referencia no asimiladas

A pesar de los avances en técnicas de percepción remota, los productos derivados de imágenes satelitales aún presentan sesgos y problemas relacionados a la escala y continuidad espaciotemporal de los registros. Es por ello que, con el ánimo de contrastar “señales” asociadas a, por ejemplo, evapotranspiración, humedad de suelo superficial, y sublimación, se utiliza la métrica propuesta por Dembelé et al. (2020b):

$$E_{SP} = 1 - \sqrt{(r_s - 1)^2 + (\gamma - 1)^2 + (\alpha - 1)^2} \quad (4.22)$$

La ecuación (4.24) es similar al KGE (Gupta et al., 2009), pero considera términos que apuntan a evaluar la señal simulada respecto a la observada (i.e., resultados no dependen del monto). Así, el término “ r_s ” corresponde al coeficiente de correlación de Spearman, definido a partir de la ecuación (4.23), “ γ ” es la razón de variabilidad (i.e., la razón entre los coeficientes de variación), definida según (4.24), y “ α ” el término de coincidencia de la señal espacial, calculado como se presenta en (4.25).

$$r_s = 1 - \frac{6 \sum_1^n d_i^2}{n(n^2 - 1)} \quad (4.23)$$

$$\gamma = \frac{\sigma_S / \mu_S}{\sigma_O / \mu_O} \quad (4.24)$$

$$\alpha = 1 - E_{RMS}(Z_{X_S}, Z_{X_O}) \quad (4.25)$$

La ecuación (4.25) se calcula a partir de la raíz del error cuadrático medio de los valores estandarizados según (4.26), el cual se estima según la expresión (4.27).

$$Z_X = \frac{X - \mu_X}{\sigma_X} \quad (4.26)$$

$$E_{RMS}(X_S, X_O) = \sqrt{\frac{1}{n} \sum_1^n (X_S - X_O)^2} \quad (4.27)$$

Asimismo, para comparar los patrones de variabilidad espacial simulados por VIC con aquéllos derivados de productos satelitales, se utiliza el análisis temporal de píxeles propuesto por Dembelé et al. (2020a), pero modificado para incorporar el indicador E_{SP} . Así, la métrica queda definida según (4.28), donde “ X ” corresponde al número de píxeles por fila e “ Y ” al número de píxeles por columna. En otras palabras, la dimensión del producto espacialmente distribuido sería $[X \times Y]$.

$$\phi_{PW} = 1 - \left[\frac{1}{X \times Y} \sum_{\substack{1 \leq i \leq X \\ 1 \leq j \leq Y \\ \forall t \in T}} E_{SP}(S^{i,j}, O^{i,j}) \right] \quad (4.28)$$

En la expresión (4.28), cada píxel (i, j) es evaluado aplicando la métrica E_{SP} y luego se obtiene el promedio espacial asociado.

4.6 Configuración de experimentos de asimilación de datos

Gillijns et al. (2006) reconocen la naturaleza heurística³ de la definición del número de miembros requeridos para implementar EnKF (i.e., tamaño de conjunto o ensemble). A su vez, precisan que, en la literatura, se recomienda un tamaño de conjunto entre 50 y

³ Manera de buscar la solución de un problema mediante métodos no rigurosos en términos científicos, como por ejemplo tanteo, reglas empíricas, documentos o fuentes históricas, etc.

100 miembros. Considerando el costo computacional que implica correr el modelo VIC (aproximadamente 0,882 s/día/píxel/miembro), se define un tamaño de *conjunto* de 50 miembros, y las configuraciones propuestas se resumen en la Tabla 4.6.

Tabla 4.6: Configuración de experimentos de asimilación.

ID	Perturbación	Observación asimilada	Estado actualizado
E1.C1	Forzantes y condición inicial de humedad de suelo	Escorrentía media diaria (Q)	Condición inicial de humedad de suelo
E2.C1	Forzantes, condición inicial de: humedad de suelo, equivalente de nieve en agua (SWE) y fracción de cobertura nival (fSCA)	Escorrentía media diaria (Q)	Condición inicial de humedad de suelo, equivalente de nieve en agua y fracción de cobertura nival.
E2.C2	Forzantes, condición inicial de: humedad de suelo, equivalente de nieve en agua (SWE) y fracción de cobertura nival (fSCA)	Equivalente de nieve en agua (SWE) y fracción de cobertura nival (fSCA)	Condición inicial de humedad de suelo, equivalente de nieve en agua y fracción de cobertura nival.
E2.C3	Forzantes, condición inicial de: humedad de suelo, equivalente de nieve en agua (SWE) y fracción de cobertura nival (fSCA)	Escorrentía media diaria (Q), equivalente de nieve en agua (SWE) y fracción de cobertura nival (fSCA)	Condición inicial de humedad de suelo, equivalente de nieve en agua y fracción de cobertura nival.

Así, a partir de cada una de las configuraciones indicadas en la Tabla 4.6, se sigue el diagrama de flujo de la Figura 4.5 donde, luego de la perturbación inicial de estados iniciales y forzantes, y la simulación de dichos escenarios, se evalúa la disponibilidad de observaciones y se inicia el proceso de asimilación de datos.

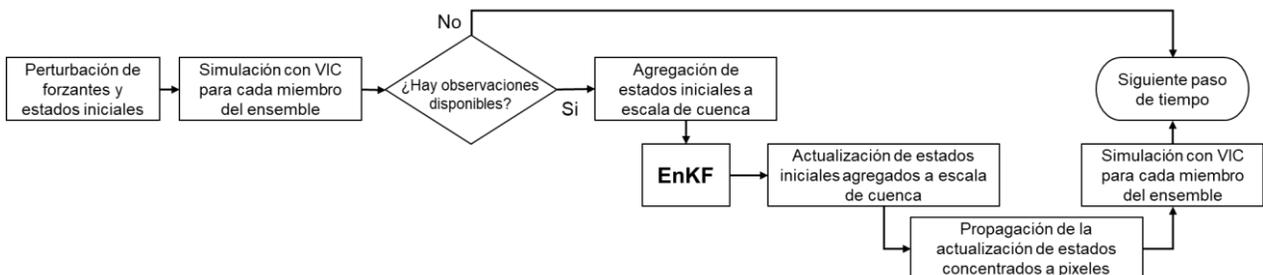


Figura 4.5: Diagrama de flujo del proceso de asimilación de datos con EnKF.

En la Figura 4.5 se indica un proceso de “agregación de estados iniciales desde la escala de píxel a la escala de cuenca”, el cual se aplica para reducir la dimensionalidad del problema considerando que las observaciones asimiladas están en la escala de cuenca y que no hay un esquema de rastreo. De este modo, se sigue la metodología presentada en la Figura 4.6, la cual esquematiza el caso de la humedad de suelo (análogo para SWE). En primer lugar se estima, para cada píxel, la humedad de suelo promedio (SMX_p) en cada una de las “X” capas definidas (i.e., tres), a partir de la suma ponderada de la fracción de cobertura de suelo ($frac_{p,c}$) y el valor de humedad de suelo asociada al mismo ($SMX_{p,c}$). Luego, con el valor de humedad promedio por capa y por píxel, se procede a estimar la humedad de suelo promedio por capa a escala de cuenca, la cual depende de la fracción que ocupa la cuenca en el píxel p (i.e., $frac_p$) y la humedad de

suelo promedio de la capa X del mismo (i.e., SMX_p). Así, se actualiza la humedad de suelo de cada una de las tres capas a la escala de cuenca mediante EnKF, y se estima un factor de actualización de estados a priori, según el cual se propaga la corrección desde la escala de cuenca hasta el nivel de subpíxel.

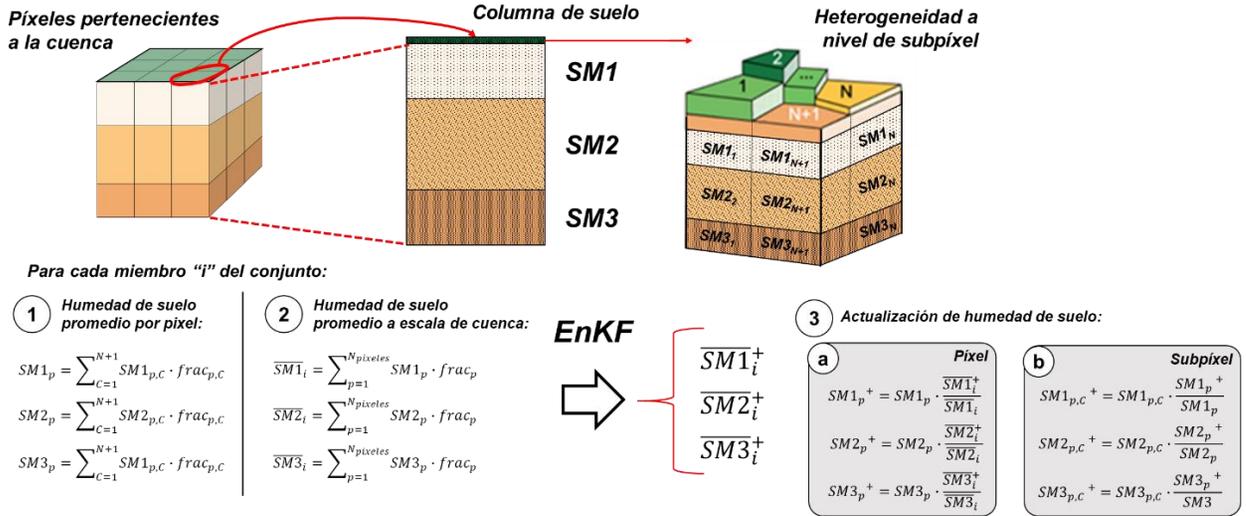


Figura 4.6: Esquema de la metodología empleada para la agregación de estados iniciales, distribuidos a nivel de subpíxel y píxel, a escala de cuenca y posterior propagación de estados actualizados desde la escala de cuenca a la escala de subpíxel. Ejemplo aplicado al caso de la actualización de humedad de suelo. En el caso del equivalente de nieve en agua (SWE), sólo existe una variable que se manipula de manera análoga a cualquier capa de suelo.

4.7 Incorporación del error de la observación

A partir del enfoque propuesto por Clark et al. (2008), se define el error de la observación utilizando un enfoque aditivo (ecuación (4.29)). A su vez, el término asociado al error (φ_Q) se supone normal y linealmente dependiente de la magnitud de la observación Q (ecuación 4.32). De este modo, basta con definir el término " ε_Q " para estimar el error de la observación, lo cual no es trivial si se busca atribuir a errores instrumentales, ya que no todos los factores que condicionan la observación se pueden controlar de manera continua (e.g., embancamientos).

$$Q' = Q + \varphi_Q \quad (4.29)$$

$$\varphi_Q \sim N(0, \sigma_Q^2) ; \quad \sigma_Q^2 = (\varepsilon_Q \cdot Q)^2 \quad (4.30)$$

En consecuencia, para estimar el valor asociado a " ε_Q ", se propone adoptar un enfoque de calibración de errores (e.g., Li et al., 2014) o definir un valor a priori (Clark et al., 2008).

4.8 Resumen de errores de modelación y observación adoptados

Los tiempos de cómputo asociados a las simulaciones con VIC y los diferentes miembros del conjunto limitan las alternativas disponibles. Por ejemplo, para simular un año hidrológico en una cuenca de 14 píxeles como Estero Arrayán en la Montosa en

modo Open-Loop (OL; i.e., corridas sin asimilación de datos), con un tamaño de conjunto de 50 miembros, paralelizado en 14 nodos de simulación, y utilizando la infraestructura del NLHPC (<https://www.nlhpc.cl/>), el tiempo de cómputo es de, aproximadamente, 4-5 horas. Además, el tiempo anterior se duplica para las simulaciones en modo asimilación de datos (DA). En base a lo anterior, se definió el año hidrológico 2000/2001 como periodo de prueba. Por otro lado, se realizaron pruebas con distintos valores asociados a los errores del modelo y observaciones obtenidos de la literatura (e.g., Clark et al., 2008; Li et al., 2014; Mendoza et al., 2012) para obtener, por tanteo, un conjunto de valores que permitiera, a lo menos, contener la serie de observaciones (i.e., que dispersión del conjunto contenga la serie observada). De este modo, se definen los valores presentados en la Tabla 4.7, los cuales son aplicados por igual a las diferentes cuencas de estudio. En la tabla se resume, a su vez, la fuente de incertidumbre y el tipo de perturbación asociada a cada una de las variables empleadas para generar el *ensemble* o conjunto de simulación.

Tabla 4.7: Resumen de errores definidos para el ejercicio de asimilación de datos con EnKF.

Fuente de Incertidumbre	Variable	Notación	Valor	Perturbación	Tipo*
Forzante	Precipitación	ϵ_P	0,50	Multiplicativa	$U(-\epsilon_P, \epsilon_P)$
	Velocidad del viento	ϵ_u	0,20	Multiplicativa	$U(-\epsilon_u, \epsilon_u)$
	Temperatura	ϵ_T	0,20	Aditiva	$U(-\epsilon_T, \epsilon_T)$
	Humedad relativa	ϵ_{HR}	0,20	Multiplicativa	$U(-\epsilon_{HR}, \epsilon_{HR})$
Condición Inicial	Humedad de suelo capa 1 (SM1)	ϵ_{SM1}	2,00	Aditiva	$U(-\epsilon_{SM1} \Delta S , \epsilon_{SM1} \Delta S)$
	Humedad de suelo capa 2 (SM2)	ϵ_{SM2}	2,00	Aditiva	$U(-\epsilon_{SM2} \Delta S , \epsilon_{SM2} \Delta S)$
	Humedad de suelo capa 3 (SM3)	ϵ_{SM3}	2,00	Aditiva	$U(-\epsilon_{SM3} \Delta S , \epsilon_{SM3} \Delta S)$
	Equivalente de nieve en agua (SWE)	ϵ_{SWE}	2,00	Aditiva	$U(-\epsilon_{SWE} \Delta S , \epsilon_{SWE} \Delta S)$
Flujo observado	Escorrentía	ϵ_Q	0,15	Aditiva	$N(0, [\epsilon_Q Q]^2)$

*U() y N() denotan una distribución uniforme y normal, respectivamente. $|\Delta S|$ corresponde a la variación absoluta de la condición inicial del estado S (i.e., SM1, SM2, SM3 o SWE según sea el caso) entre el tiempo t y t-1.

5 RESULTADOS

5.1 Calibración del modelo hidrológico

Una vez calibrados los parámetros del modelo, para cada cuenca del dominio de estudio se generan simulaciones en el periodo 1979-2015 con el fin de evaluar la calidad de las mismas. A modo de ejemplo, la Figura 5.1 muestra resultados para la cuenca Estero Arrayán en la Montosa (5722001), desde donde se desprende que, en general, el modelo replica patrones diarios, mensuales y anuales de caudal, y curva de duración. Sin embargo, se observan diferencias en valores altos de caudal medio diario en las diferentes escalas temporales ilustradas (paneles A-C). Los resultados obtenidos para el resto de las cuencas de estudio pueden ser encontrados en el Anexo A.

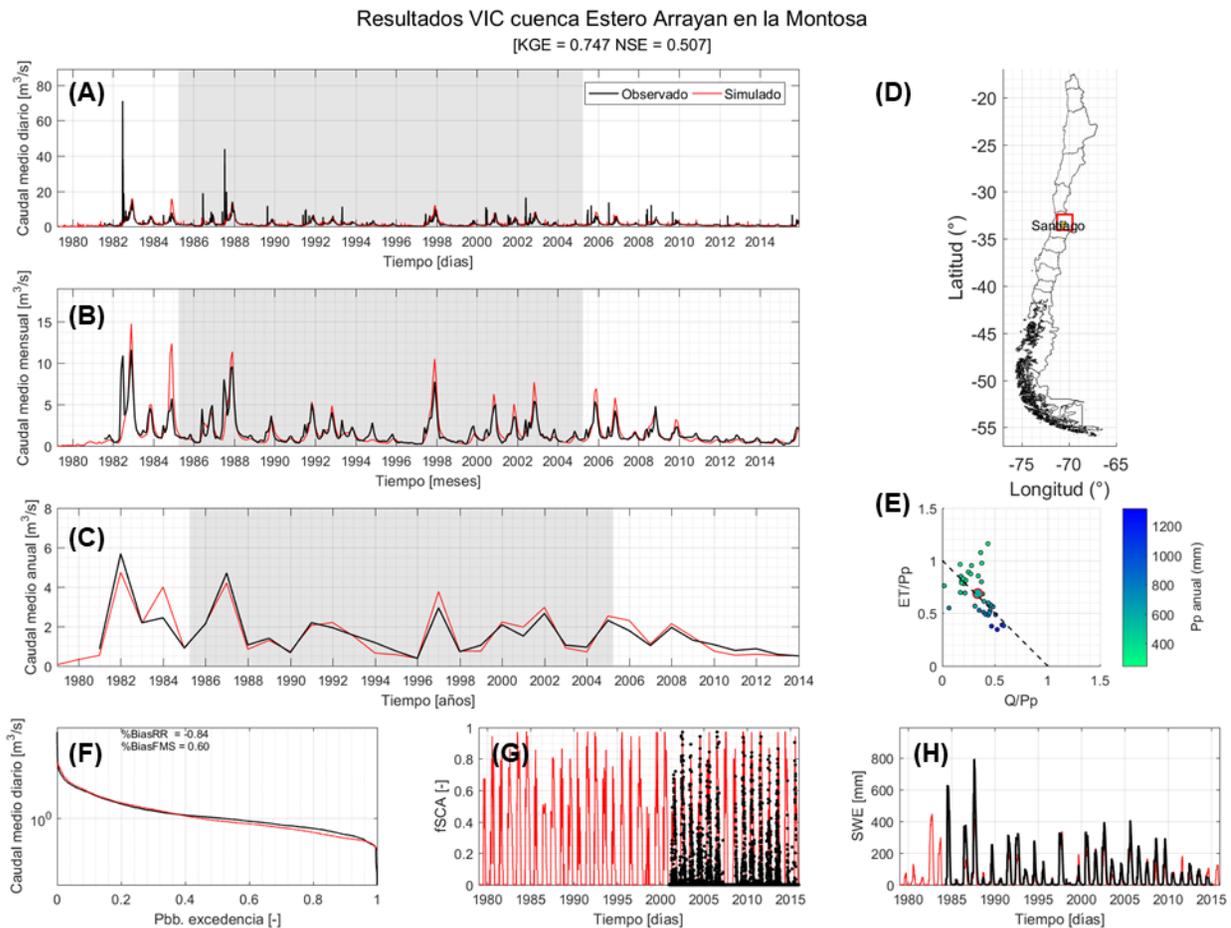


Figura 5.1: Resultados de calibración del modelo VIC en la cuenca Estero Arrayán en La Montosa – BNA 5722001. Se muestran las simulaciones para el periodo 01/01/1979-31/12/2015, y el color gris muestra el periodo de calibración utilizado. (A) Caudal medio diario observado y simulado. (B) Caudal medio mensual observado y simulado. (C) Caudal medio diario promedio anual observado y simulado. (D) Ubicación geográfica de la cuenca. (E) Balance hidrológico adimensional simulado; los puntos pequeños corresponden al balance anual, mientras que el punto de mayor tamaño y con borde rojo corresponde al promedio en el periodo 1985-2015. (F) Curva de duración de caudales medios diarios. (G) Cobertura de nieve agregada a escala de cuenca. (H) Equivalente de nieve en agua agregado a escala de cuenca.

La Figura 5.2 muestra los resultados de KGE y sus componentes, para las 12 cuencas en estudio, obtenidos para los periodos de calibración y validación (Tabla 4.1). Además, la evaluación es estratificada para diferentes condiciones de déficit o superávit de precipitación anual respecto al promedio en el periodo 1985-2015 (i.e., año seco, normal, húmedo). Para efectos prácticos, en este estudio se considera como un año seco (húmedo) aquel que presenta un déficit (superávit) de al menos 20% respecto al monto de precipitación anual promedio en el periodo 1985-2015. El umbral del 20% de define considerando la mitad del coeficiente de variación promedio entre las 12 cuencas, asociada al promedio de precipitación anual en el periodo 1985-2015.

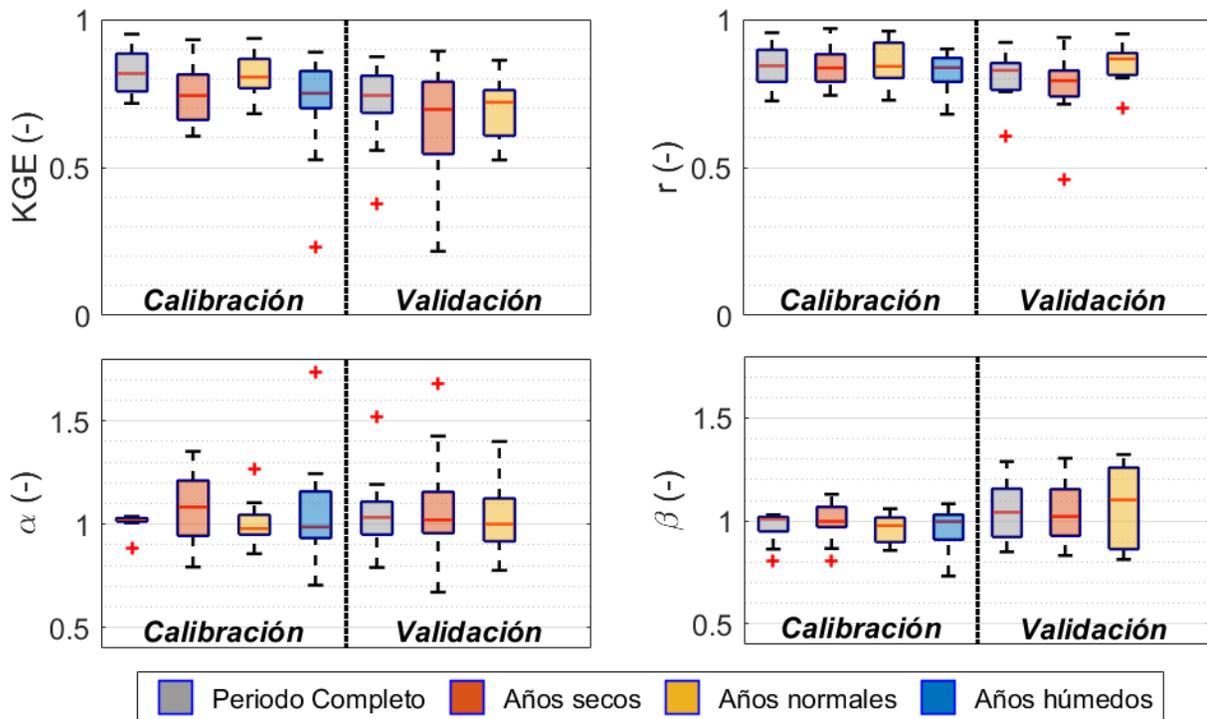


Figura 5.2: Resultados del KGE y sus componentes obtenidos de la calibración (01/04/1985-31/03/2005) y validación (01/04/2005-31/04/2015) del modelo VIC en las cuencas de estudio. Cada caja agrupa 12 valores (i.e., un valor por cuenca), mientras que los subperiodos seco y húmedo representan déficit o superávit de por lo menos un 20% de la precipitación anual respecto al promedio 1985-2015.

De la Figura 5.2, se desprende que todas las cuencas presentan resultados de KGE y correlación de Pearson mayores o iguales 0,7 (i.e., valores satisfactorios) para el periodo de calibración completo. Respecto a la componente de variabilidad (i.e., α), los valores se encuentran cercanos a 1, con una sola cuenca en 0,9 (i.e., la variabilidad simulada es 10% menor a la observada). A su vez, para la componente de ajuste de volumen (i.e., β), a pesar de que la mediana de las cuencas esté en 1, existe un sesgo negativo en torno al 20% (i.e., volumen simulado es 20% menor al observado).

En el periodo de validación, los resultados de KGE empeoran respecto al periodo de calibración (Figura 5.2). Si bien la componente de temporalidad r también empeora su valor, el cambio es de menor magnitud (e.g., mediana no cambia, sólo cambian los anchos del segundo y tercer cuartil) que el observado en las otras componentes (e.g., aumento en el tamaño de la caja y cambios en la mediana cercanos al 5%). Al desagregar

el análisis por años secos, normales y húmedos, en la Figura 5.2 se observa que el KGE del periodo completo (tanto para calibración como validación) muestra un mayor parecido al KGE de los años normales. La misma apreciación se desprende de la inspección visual de las componentes del KGE, aunque en este caso las componentes α y β para años normales en el periodo de validación también empeoran.

Adicionalmente, para complementar el diagnóstico de la calibración, en la Figura 5.3 se evalúan las métricas basadas en procesos hidrológicos indicadas en la Tabla 4.2, donde valores cercanos (lejanos) a 1 dan cuenta del óptimo.

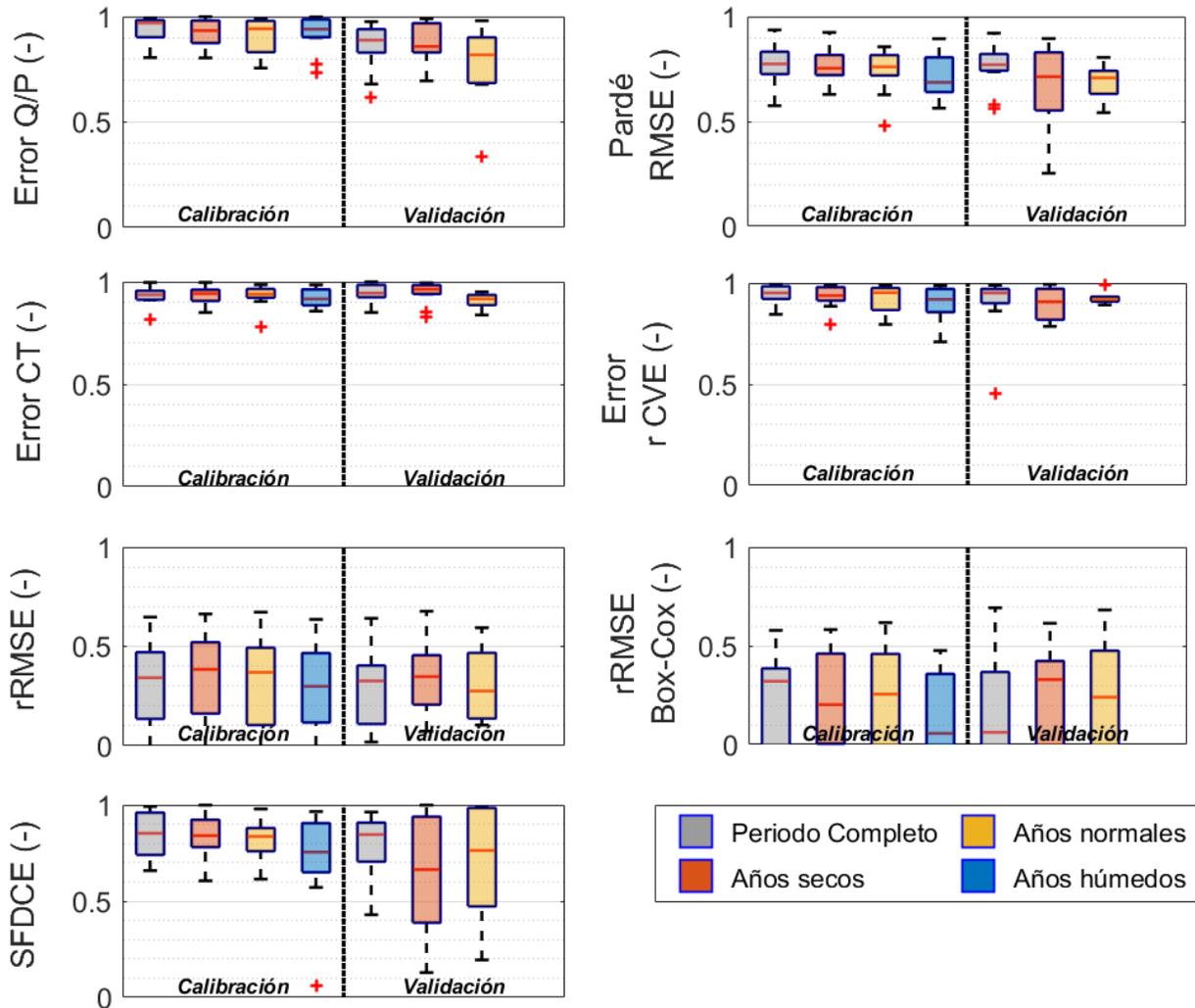


Figura 5.3: Evaluación del desempeño del modelo VIC en las cuencas de estudio en el periodo de calibración (01/04/1985-31/03/2005) y validación (01/04/2005-31/04/2015), a partir de índices de consistencia hidrológica.

De la Figura 5.3 se desprende que, en el periodo de calibración, los diferentes subperiodos analizados (i.e., periodo completo, años secos, normales y húmedos) muestran un desempeño similar. Tanto la partición Q/P, como los coeficientes de Pardé, temporalidad de la escorrentía (CT y r) y SFDCE presentan un desempeño que se podría calificar como bueno o satisfactorio, pues registran valores mayores a 0,5. El principal problema se detecta en la representación de los caudales altos (rRMSE) y en la representación de valores bajos de escorrentía (rRMSE Box-Cox). A su vez, se observa

que, en todos los casos, al evaluar el desempeño en el periodo de validación los resultados empeoran respecto a la calibración, lo cual sugiere potenciales problemas asociados a la transferibilidad temporal de parámetros y el supuesto de estacionariedad de los mismos (no explorado en este trabajo). De este modo, los resultados obtenidos concuerdan con los hallazgos derivados de la Figura 5.2, pues 1) los problemas en la simulación están de la mano con la representación de valores altos y bajos de escorrentía, y 2) la transferencia temporal de los parámetros calibrados (diferencias en desempeño en periodo de calibración y validación).

5.2 Habilidad del ensemble generado

En la Figura 5.4 y Figura 5.5 se presentan los histogramas de rango de los dos casos de estudio (i.e., perturbando sólo condiciones iniciales de humedad de suelo y humedad de suelo con equivalente de nieve en agua) sin asimilación de datos (i.e., Open-Loop), junto con las métricas de habilidad del ensemble evaluadas para cada cuenca (i.e., consistencia e indistinguibilidad) y el índice de discrepancia (i.e., DI; Delle Monache et al., 2006).

En la Figura 5.4 se observa que, en algunas cuencas (e.g., 4703002 y 5722001), la calidad del ensemble califica como óptima según las métricas de evaluación (cerca de 1), a pesar de no presentar un histograma de rango uniforme salvo para las cuencas 4703002 (DI = 21,9%) y 5722001 (DI = 51,5%), que son las que más se aproximan. El resto de las cuencas muestran sesgos negativos (i.e., mayores frecuencias en los intervalos más grandes) en sus histogramas de rango y alta dispersión. Dicha diferencia entre el análisis cualitativo y cuantitativo se puede atribuir a compensaciones temporales generadas al promediar los estadísticos que definen las métricas aplicadas. Por otro lado, las cuencas restantes presentan, en general, problemas asociados a sesgos negativos, con valores de consistencia e indistinguibilidad estadística normalizada menores que 1, e histogramas de rango con alta dispersión y sesgo en valores altos. El único resultado que difiere de lo anterior, al menos en términos de la consistencia, es para la cuenca 5410002, donde, si bien la indistinguibilidad normalizada es cercana al óptimo, la consistencia es más del doble del óptimo, es decir, el ensemble generado es poco disperso respecto al error promedio.

Al agregar el equivalente de nieve en agua inicial como estado a perturbar para generar el ensemble (Figura 5.5), el desempeño del ensemble empeora respecto al caso en que sólo se perturba el estado inicial de la humedad de suelo por capa, principalmente en las cuencas que antes presentaban un mejor desempeño (i.e., 4313001, 4703002, 5721001, y 5722001). A su vez, en las cuencas ubicadas en la zona C del dominio de estudio (Figura 2.1), el desempeño del ensemble no varía considerablemente. En casi todos los casos antes indicados, se generan sesgos negativos, los cuales se evidencian con valores de consistencia menores a 0,5 e histogramas de rango que ilustran ensembles con sesgo de valores altos y alta dispersión. En otras palabras, no cambian los resultados obtenidos en términos cualitativos a partir del histograma de rango, pero, en términos cuantitativos, los ensembles empeoran incluso en cerca de un 100%. Las únicas excepciones a la disminución generalizada en la calidad del ensemble recaen en las cuencas 5410002 y 4311001, las cuales mejoran, por ejemplo, la consistencia en cerca de un 92% y 30% respectivamente. Asimismo, el histograma de rango presenta

mejoras para la cuenca 5410002, lo cual se puede verificar con la disminución en un 29,6% del índice de discrepancia (DI). No obstante, a pesar de lo anterior, el histograma mantiene los sesgos negativos presentados en el primer caso (Figura 5.4).

Tomando en cuenta las diferencias entre los resultados de la Figura 5.4 y Figura 5.5, donde el único cambio radica en la perturbación de un estado adicional (i.e., SWE inicial), se desprende que, al cambiar la configuración de generación del ensemble (i.e., variables perturbadas), los errores asociados cambian. Lo anterior se puede atribuir a las implicancias explícitas de asignar un error a una cierta variable donde, a pesar de tratar la incertidumbre pseudo desagregada (i.e., el error sólo se atribuye a la variable en cuestión) incluye la incertidumbre de otras componentes que se interrelacionan con ella. Así, en el caso de agregar una nueva componente para caracterizar el error del modelo, los errores debiesen ajustarse a modo de redistribuir la incertidumbre general del sistema.

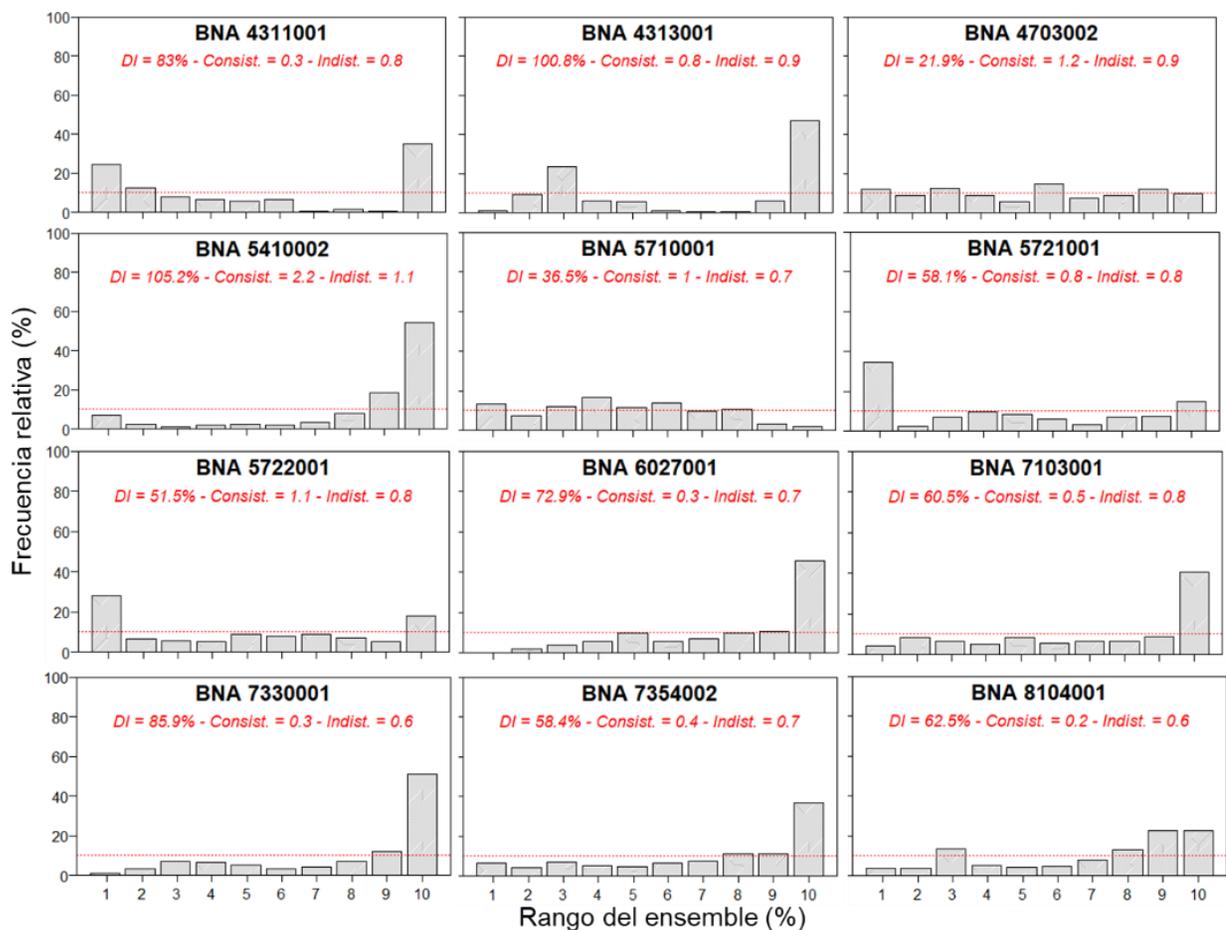


Figura 5.4: Histograma de rango del ensemble sin asimilación (OL) generado a partir de la perturbación de la condición inicial de humedad en las tres capas de suelo configuradas en VIC (Caso 1). En cada gráfico, se indica el código BNA de la cuenca de estudio, el índice de discrepancia y los valores asociados a consistencia del ensemble e indistinguibilidad estadística normalizada. La línea roja segmentada representa una distribución de rangos uniformes.

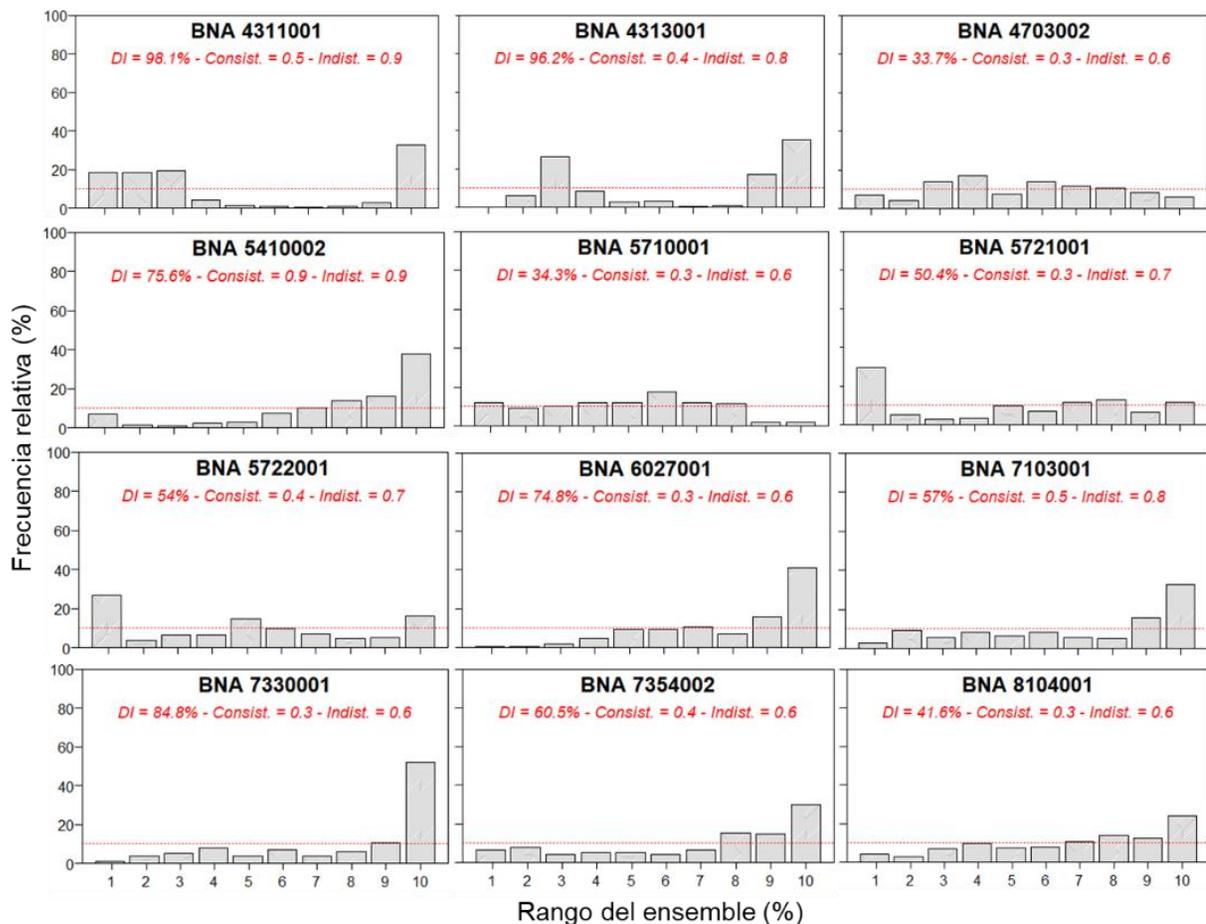


Figura 5.5: Histograma de rango del ensemble sin asimilación (OL) generado a partir de la perturbación de la condición inicial de humedad en las tres capas configuradas en VIC y el SWE inicial (Caso 2). En cada gráfico, se indica el código BNA de la cuenca de estudio, el índice de discrepancia y los valores asociados a consistencia del ensemble e indistinguibilidad estadística normalizada. La línea roja segmentada representa una distribución de rangos uniformes.

5.3 Resultados de asimilación de datos

5.3.1 Evaluación de resultados agregados a escala de cuenca

A partir de las diferentes configuraciones propuestas (Tabla 4.6), la Figura 5.6, Figura 5.8 y Figura 5.7 ilustran, a modo de ejemplo, los resultados obtenidos para el año hidrológico 2000/01 en términos de escorrentía media diaria, fracción de área cubierta por nieve y equivalente de nieve en agua respectivamente, para la cuenca Estero Arrayán en la Montosa (5722001).

De las figuras se desprende que, según el esquema propuesto, la asimilación de datos (DA) no genera mejoras en la cobertura de nieve simulada respecto al caso sin asimilación (OL). Para el caso de la escorrentía media diaria (Figura 5.6), se aprecia que, salvo para la configuración C2.E2 (i.e., sólo asimilación de fSCA y SWE), la dispersión del ensemble disminuye y se acerca a la observación. Además, se observa que no existen cambios visualmente detectables entre las configuraciones C1.E1, C2.E1 y C2.E3, pero que, al contrastar los cambios en términos de SWE (Figura 5.7), queda de manifiesto el valor agregado de asimilar observaciones del manto nival.

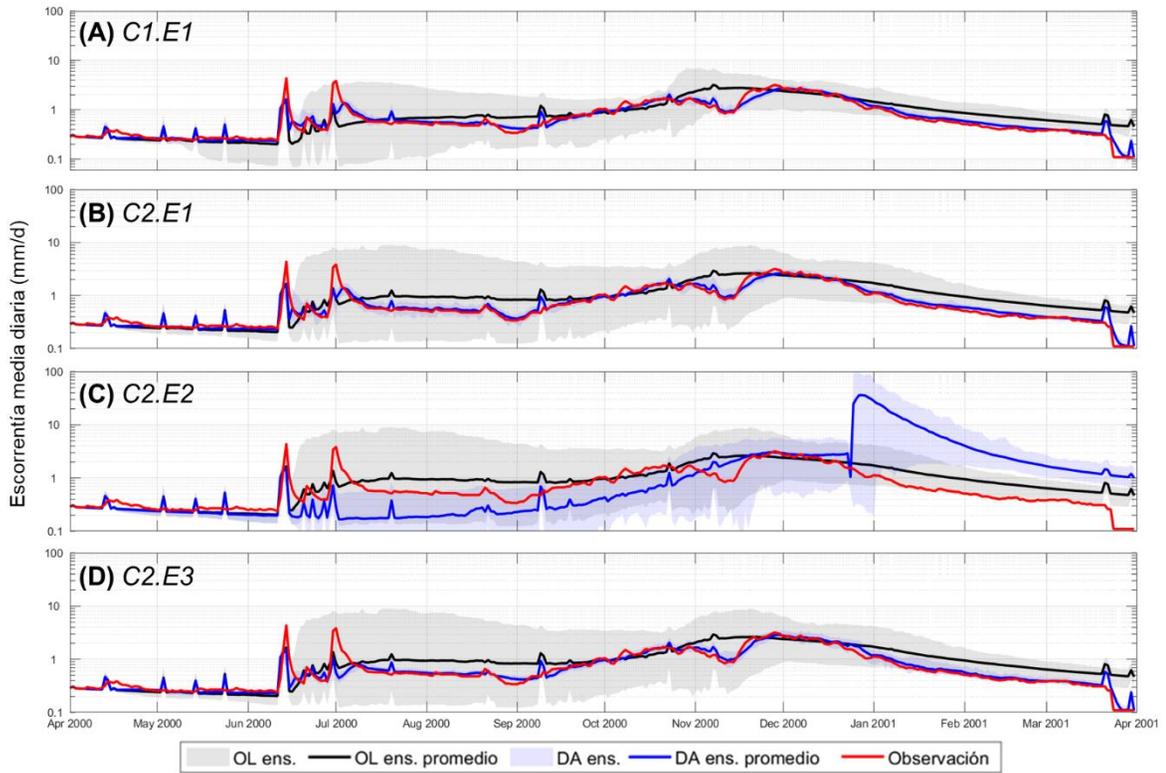


Figura 5.6: Resultados de escorrentía media diaria a escala de cuenca según las configuraciones con y sin asimilación de datos en la cuenca Estero Arrayán en La Montosa.

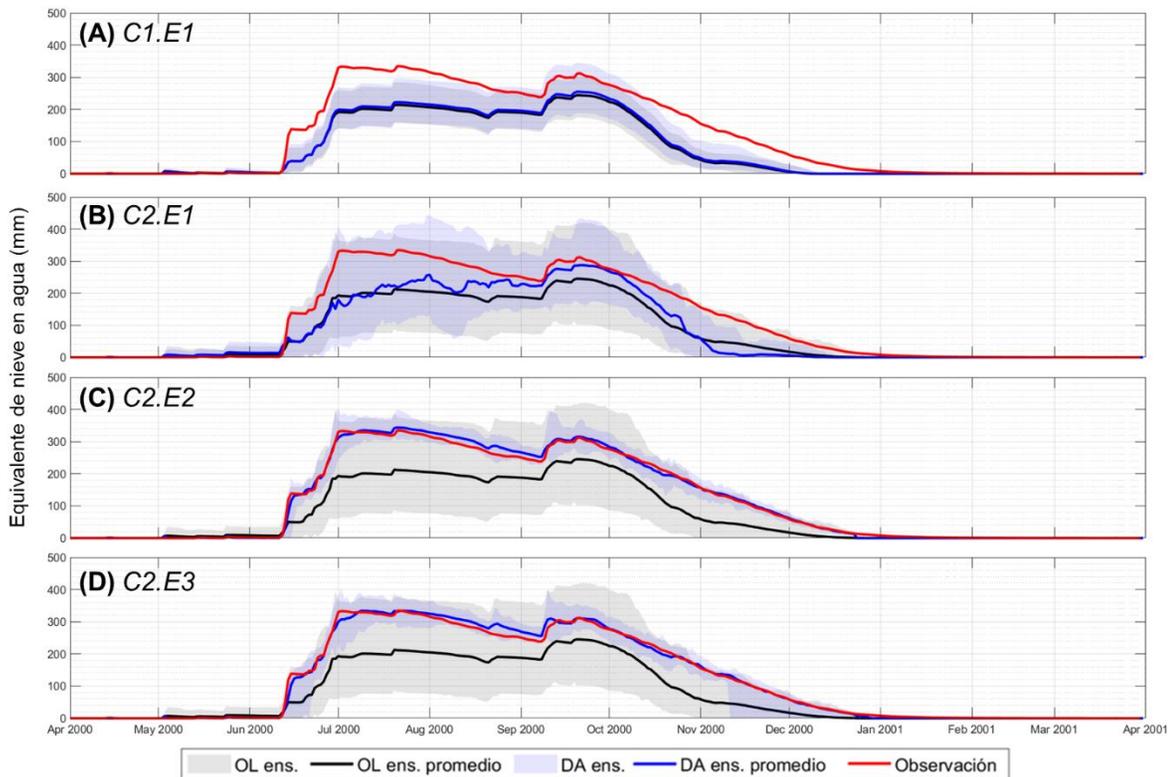


Figura 5.7: Resultados de SWE a escala de cuenca y resolución temporal diaria según las configuraciones con y sin asimilación de datos en la cuenca Estero Arrayán en La Montosa.

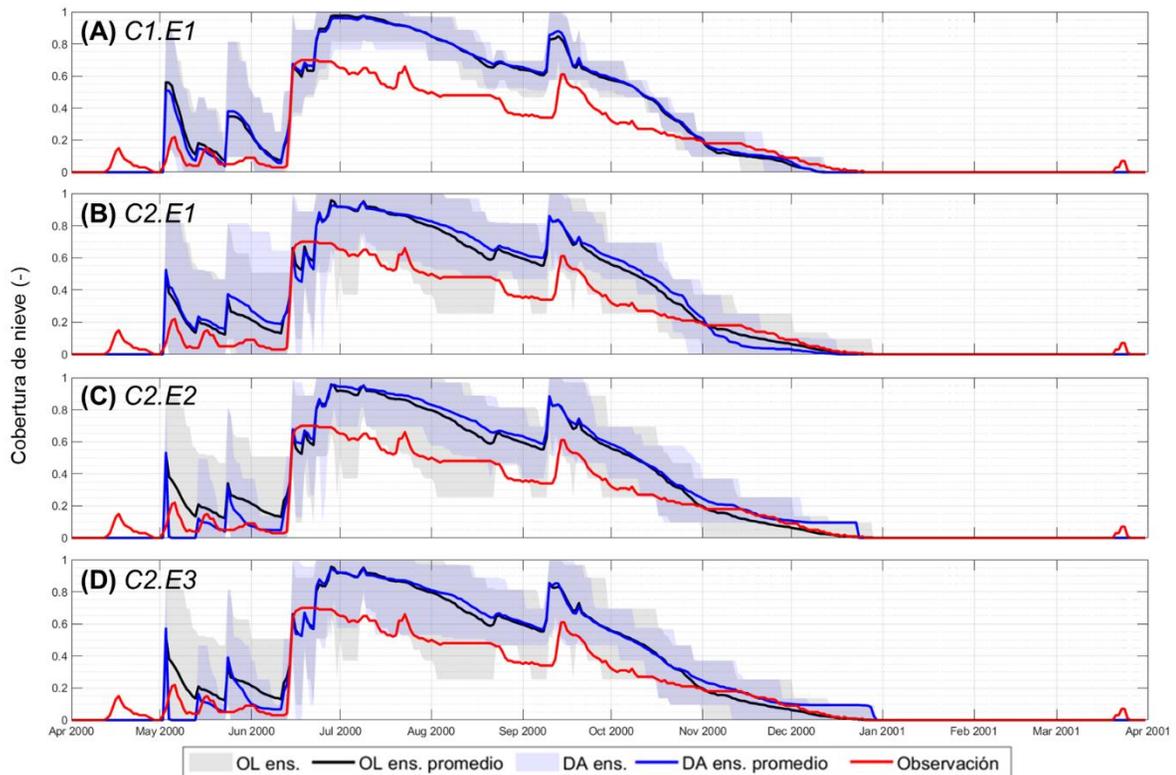


Figura 5.8: Resultados de la fSCA a escala de cuenca y resolución temporal diaria según las configuraciones con y sin asimilación de datos en la cuenca Estero Arrayán en La Montosa.

Para las configuraciones en que sólo se asimila escorrentía media diaria (i.e., C1.E1 y C2.E1), se observan diferencias en los montos, temporalidad y duración del equivalente de nieve en agua (SWE) simulado respecto a la referencia de Cortés y Margulis (2017). No obstante, en las configuraciones donde se asimila SWE y fSCA (i.e., C2.E2 y C2.E3), además de disminuir la dispersión del ensemble, la temporalidad, los montos y la duración del SWE mejoran. Lo anterior queda de manifiesto al comparar el valor promedio del ensemble asimilado, el cual coincide con el SWE de referencia.

El análisis anterior se repite para las diferentes cuencas del dominio de estudio, con lo cual se obtienen los resultados presentados en la Figura 5.9. Allí, se evalúa el valor agregado de la asimilación de datos (DA) respecto al caso sin asimilación (OL) aplicando la ecuación (4.20). Se desprende que, respecto a la simulación de escorrentía media diaria (panel superior), la configuración C2.E2 (i.e., sólo asimilación de fSCA y SWE) no tiene un valor agregado respecto al caso OL para las cuencas ubicadas en las zonas B y C del dominio de estudio (Figura 2.1), donde la escorrentía está dominada por la contribución nival. No obstante, se debe tener en cuenta que no se modificaron los errores del modelo respecto al caso C1.E1, sólo se agregó un nuevo estado inicial perturbado y el error asociado al mismo. En consecuencia, se puede estar alterando el efecto de la asimilación de dichas observaciones considerando que, además del SWE inicial, se actualiza la humedad de suelo. El resto de las configuraciones mejora, en todas las cuencas, las estimaciones de escorrentía media diaria entre el 10% y 90%, siendo la mejor configuración dependiente de cada cuenca. Por ejemplo, la mejor configuración en la cuenca 4311001 es C2.E1, mientras que en la cuenca 5721001 es C2.E3.

Con respecto a la estimación de SWE (panel medio, Figura 5.9), se observa que el caso C1.E1 no entrega diferencias con respecto al caso sin asimilación (i.e., no hay valor agregado en asimilar caudal para mejorar la estimación del SWE). Por otro lado, la configuración C2.E1 empeora las estimaciones de SWE para algunas cuencas (e.g., 4311001, 5710001, 8104001) y las mejora en otras (e.g., 4313001, 5722001), aunque no de forma considerable ($< 15\%$). En términos generales, se obtienen resultados similares para las configuraciones C2.E2 y C2.E3, con mejoras entre el 70-80% en el valor promedio a escala de cuenca.

En cuanto a la estimación de la cobertura de nieve (SC o fSCA; panel inferior, Figura 5.9), al igual que en la Figura 5.8, se observa que, en general, no existen cambios de gran magnitud (e.g., $>15\%$) entre los casos OL y DA en cuando a estimaciones de cobertura de nieve. La única excepción es la cuenca 7330001 donde, para la configuración C2.E2, el SC mejora un 20% respecto al caso OL.

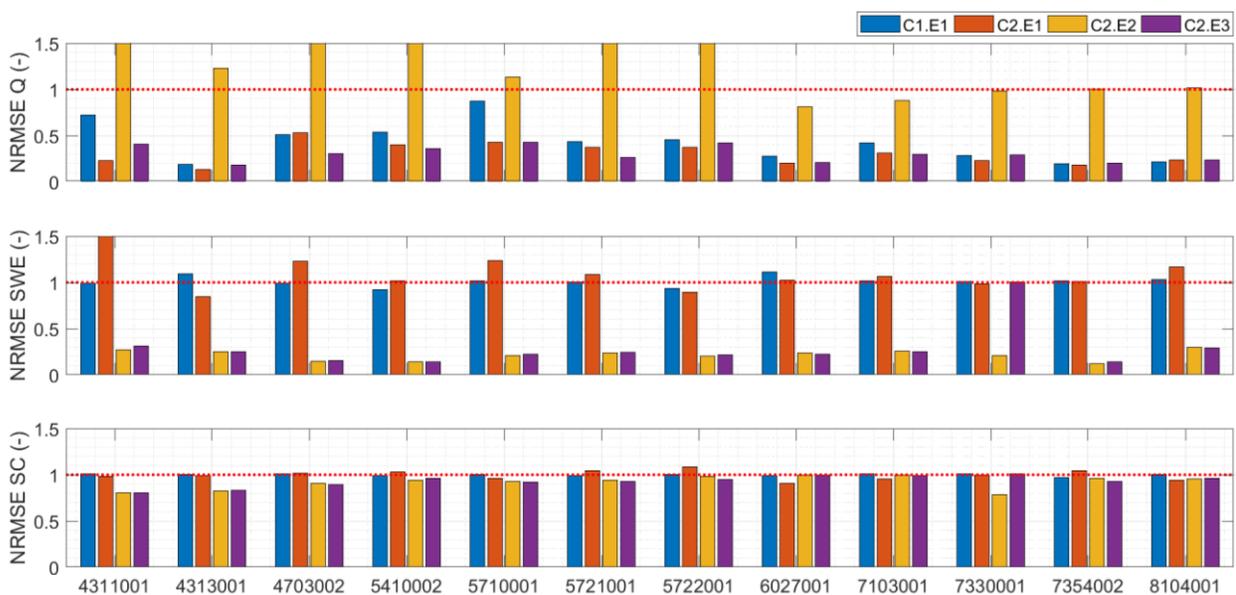


Figura 5.9: Evaluación del valor agregado, expresado en términos de la raíz del error cuadrático medio normalizado (NRMSE), de asimilar datos en las estimaciones de escorrentía (Q), equivalente de nieve en agua (SWE) y cobertura de nieve (SC o fSCA) con respecto a la situación sin asimilación (año hidrológico 2000/01). La línea roja segmentada representa el valor en que los resultados sin asimilación y con asimilación son equivalentes. Notar que el eje Y se ha truncado en 1.5 para facilitar la comparación entre esquemas de asimilación.

Tomando en cuenta que las variables con las mayores diferencias entre OL y DA corresponden a la escorrentía media diaria y SWE, se analiza la ganancia normalizada asociada a dichas variables en las Figura 5.10 y Figura 5.11, respectivamente. Cabe mencionar que la ganancia normalizada se calcula a partir de la ecuación (4.21) y de las métricas indicadas en la Tabla 4.5, evaluadas para cada uno de los $N = 50$ miembros del ensemble y posteriormente promediadas.

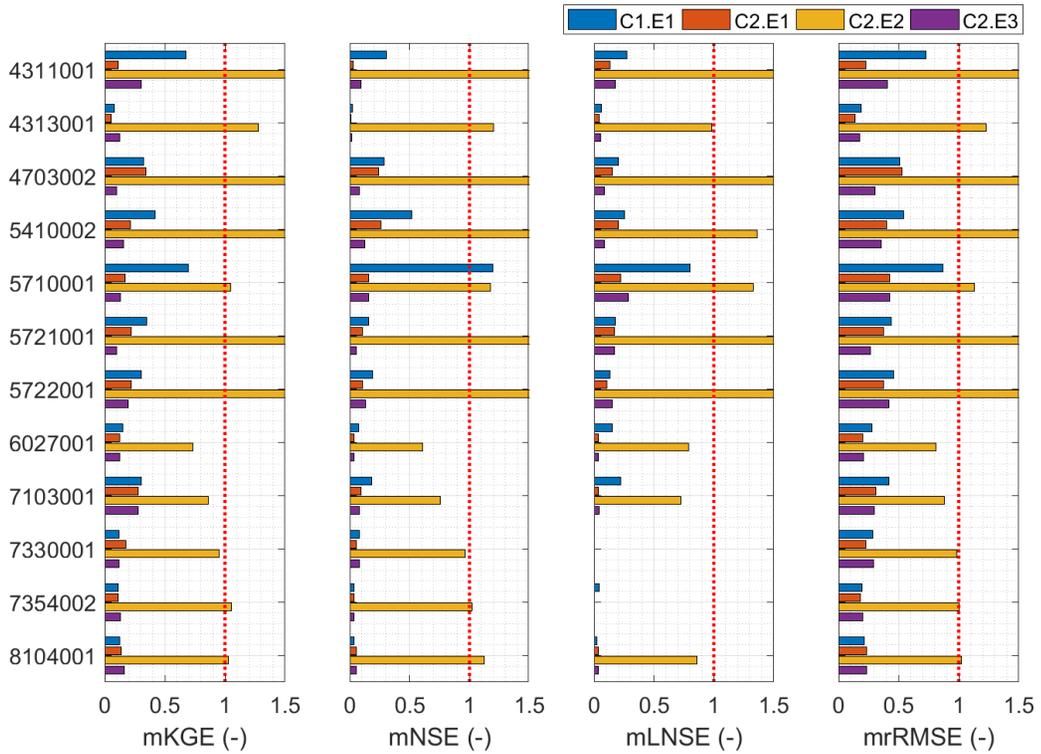


Figura 5.10: Ganancia normalizada de la asimilación de datos respecto al caso OL en la estimación de escorrentía total media diaria para el año hidrológico 2000/01. La línea roja segmentada representa el valor en que los resultados sin asimilación y con asimilación son equivalentes. El valor máximo en el eje X se ha truncado en 1.5 con fines comparativos.

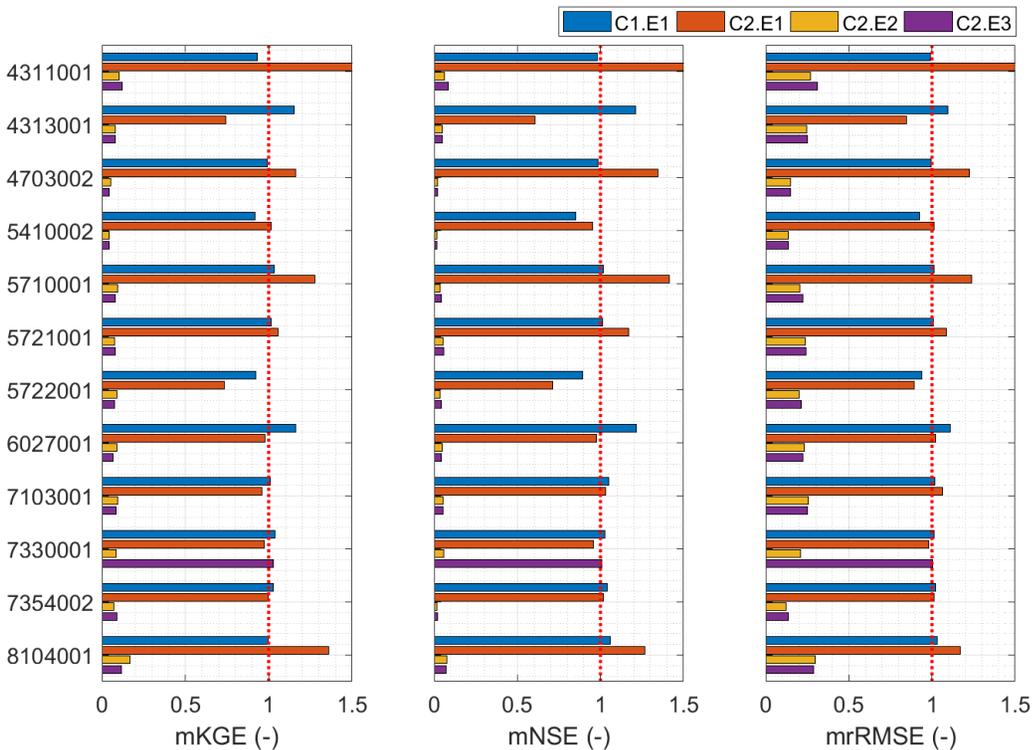


Figura 5.11: Ganancia normalizada de la asimilación de datos respecto al caso OL en la estimación de SWE para el año hidrológico 2000/01. La línea roja segmentada representa el valor en que los resultados sin asimilación y con asimilación son equivalentes. El valor máximo en el eje X se ha truncado en 1.5 con fines comparativos.

Para el caso de la escorrentía media diaria, los resultados presentados en la Figura 5.10 reafirman los hallazgos de la Figura 5.9 pues, excepto para la configuración C2.E2, se aprecian mejoras al asimilar observaciones de caudal y variables relativas al estado del manto nival. Por ejemplo, los resultados de KGE y rRMSE mejoran entre un 30-90% respecto al caso OL, mientras que el NSE y LNSE alcanzan mejoras cercanas al 100% (i.e., el valor asociado a DA es el óptimo de la métrica).

Del mismo modo, el SWE (Figura 5.11) no muestra mejoras considerables (> 15%) para la configuración C1.E1, presentando valores incluso peores al caso OL (e.g., 4313001, 6027001), lo cual se extiende para las tres métricas evaluadas. Por otro lado, la configuración C2.E1 tampoco logra mejoras generalizadas en todas las cuencas evaluadas. Así, las principales mejoras en la estimación de SWE vienen dadas por las configuraciones C2.E2 y C2.E3, superando incluso el 90% en algunas métricas (e.g., mKGE cuenca 5410002, mNSE en la cuenca 7354002). En resumen, los resultados derivados del análisis de SWE coinciden con la Figura 5.10, por lo que los hallazgos se pueden resumir en los siguientes puntos:

- Al analizar estimaciones de escorrentía media diaria, la asimilación de observaciones de dicha variable siempre mejora entre un 10-90% (dependiendo de la métrica de evaluación) los resultados respecto al caso sin asimilación. No obstante, lo anterior no conlleva mejoras en la estimación de SWE.
- Al analizar estimaciones de SWE agregadas a escala de cuenca, la asimilación de SWE y fSCA mejora, en general, entre un 70-90% (dependiendo de la métrica de evaluación) los resultados respecto al caso sin asimilación. No obstante, lo anterior no conlleva mejoras en la estimación de escorrentía media diaria.
- La asimilación conjunta de escorrentía media diaria, fSCA y SWE permite, en general, mejorar las estimaciones de escorrentía media diaria y SWE, disminuyendo la dispersión (i.e., incertidumbre) del ensemble generado.

Se debe precisar que la disminución de la dispersión del ensemble de la variable asimilada es un efecto que, por construcción, siempre se dará en el caso DA respecto a OL. Así, con el objetivo de evaluar el desempeño de los casos propuestos en base a “observaciones” de referencia no consideradas en la asimilación (ver sección 3.3), la Figura 5.12 ilustra la ganancia normalizada en las estimaciones de evapotranspiración real (ET), sublimación (SUB), humedad de suelo superficial⁴ (SM_{sup}) y humedad de la zona radicular⁵ (SM_{root}) recuperadas de GLEAM v3.5a y espacialmente agregada a escala de cuenca. Así, de la Figura 5.12 se desprende que, tanto para la sublimación como para la humedad de suelo superficial, no existen cambios evidentes entre los casos OL y DA, pues la mayoría de los resultados obtenidos se concentran en torno a la línea segmentada que indica dicha condición. No obstante, tanto ET en las cuencas pertenecientes al grupo B (Figura 2.1), como la humedad de suelo en la zona radicular en casi todas las cuencas, muestran variaciones de interés respecto al caso OL.

⁴ Para efectos prácticos de comparación, se supone que la primera capa de suelo en VIC (i.e., SM1) es una buena aproximación a la humedad de suelo superficial (SM_{sup}).

⁵ Considerando que la profundidad de la zona radicular en la zona de estudio va entre los 2-4 metros según la configuración del archivo de vegetación de VIC, se supone que la suma de la humedad de suelo en la capa 1 y 2 (i.e., SM1+SM2) de VIC es una buena aproximación a SM_{root} .

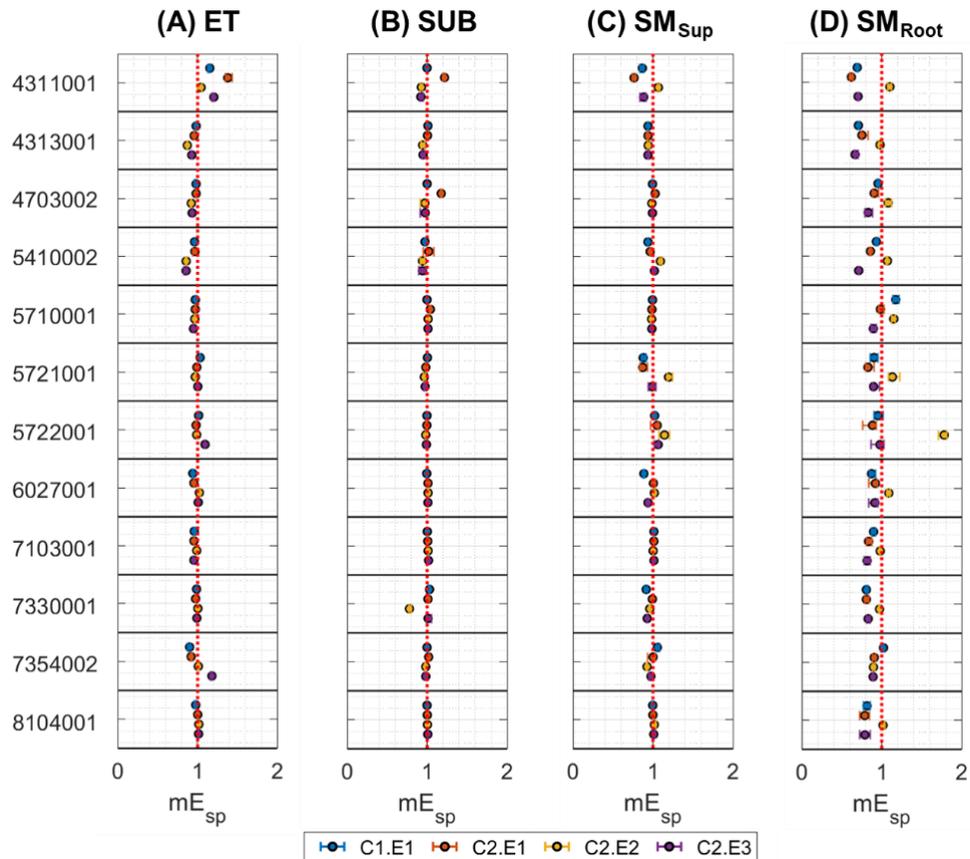


Figura 5.12: Ganancia de asimilar datos en la estimación de flujos y estados, a la escala de cuenca, no considerados como observaciones asimilables. La línea roja segmentada representa el valor en que los resultados sin asimilación y con asimilación son equivalentes. La extensión de las barras representa el percentil del 25% y 75% asociado a las diferencias estimadas para cada miembro evaluado a nivel diario en el periodo dado por el año hidrológico 2000/01. El punto al centro de cada una de las barras representa la mediana del conjunto.

De la Figura 5.12 se desprende que la ET empeora entre un 10-40% respecto al caso DA para la cuenca 4311001, mientras que, en el resto de las cuencas, se aprecian mejoras de hasta un 20% en la mediana del conjunto (e.g., 5410002, 7354002). Respecto al SM_{Root} , la mayoría de las cuencas reporta una ganancia de aplicar DA respecto a OL, con mejoras de entre un 10-40% (e.g., 4311001, 5410002, 8104001). No obstante, la configuración C2.E2 reporta resultados de asimilación peores respecto al caso OL en algunas cuencas (e.g., 5721001, 5722001).

5.3.2 Evaluación de resultados espacialmente distribuidos

A pesar de no contar con una configuración experimental de asimilación de datos que permita actualizar estados de manera espacialmente distribuida, la actualización agregada a escala de cuenca sí permite ajustar los montos de estados y flujos a nivel de píxel. A modo de ejemplo, la Figura 5.13 ilustra los cambios (i.e., OL vs. DA) en los montos anuales de escorrentía asociados al promedio del ensemble por píxel. Dicho ajuste supone una relación proporcional de la actualización concentrada respecto a los píxeles; es decir, la asimilación de datos mantiene la distribución espacial del caso OL, actualizando los montos espacialmente distribuidos. En consecuencia, resultados similares se obtienen para otros flujos y estados del modelo. Tomando en cuenta lo

anterior, la Figura 5.14 presenta los resultados obtenidos del análisis espacial de las estimaciones de evapotranspiración real (ET), sublimación (SUB), humedad de suelo superficial (SM_{Sup}) y humedad de la zona radicular (SM_{Root}) recuperadas de GLEAM v3.5a, respecto a las simulaciones con VIC. Cabe mencionar que, para realizar este análisis, se aplica la ecuación (4.28), detallada en la sección 4.5.2, la cual define la métrica E_{pw} .

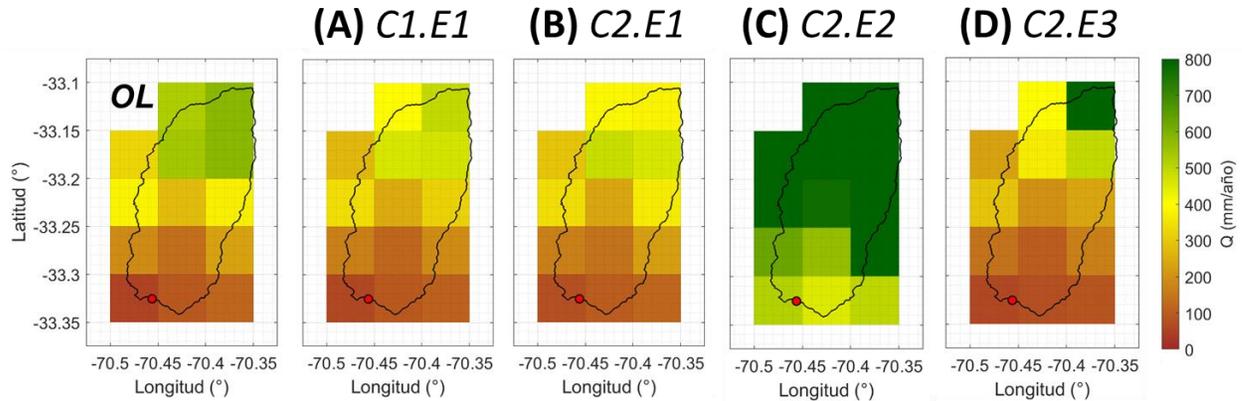


Figura 5.13: Ejemplo de los impactos de la asimilación de datos en los montos de escorrentía anual distribuidos a $0,05^\circ$ latitud-longitud. En los diferentes paneles (A-D), se muestran los resultados obtenidos para la cuenca Estero Arrayán en La Montosa según las diferentes configuraciones. El punto rojo indica la ubicación de la estación fluviométrica que da nombre a la cuenca.

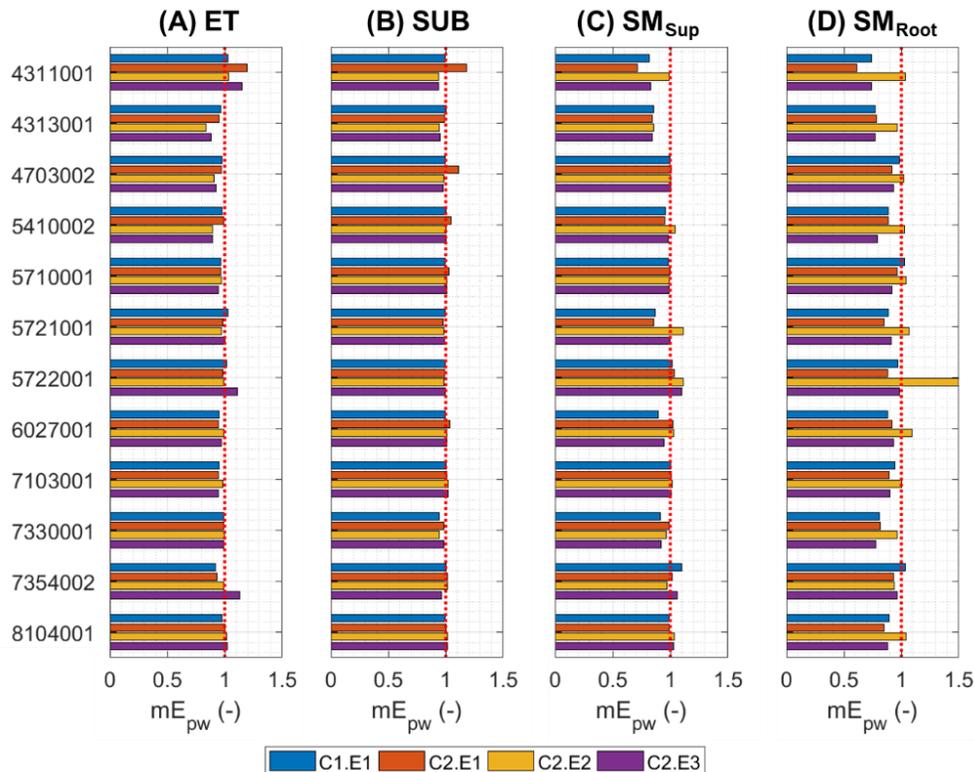


Figura 5.14: Ganancia normalizada de asimilación de datos en la representación de patrones espaciales de evapotranspiración real (ET), sublimación (SUB), humedad de suelo superficial (SM_{sup}) y humedad de suelo en la zona radicular (SM_{Root}). La línea segmentada roja representa el caso en que no hay ganancia. El valor máximo en el eje X se ha truncado en 1.5 con fines comparativos.

En la Figura 5.14 se observa que, al igual que en la Figura 5.12, no existen importantes ganancias con la asimilación de datos en algunas de las variables analizadas (e.g., SUB). No obstante, destacan ganancias cercanas al 10% en el caso de ET (e.g., 7354002), y 5-40% en el caso de SM_{Root} (e.g., 4313001, 7330001).

La Figura 5.15 desglosa algunas de las componentes del E_{pw} para identificar cuál es la que mejora con respecto al caso OL. Los resultados indican que la asimilación de datos generalmente mejora la razón de variabilidad (γ) de ET, principalmente para las cuencas ubicadas al norte del dominio (i.e., zona B y C, Figura 2.1), alcanzando valores entre 10-90% para algunas cuencas y configuraciones evaluadas (e.g., C2.E3, 4703002; C1.E1, 5710001). Dicho valor agregado se pierde en cuencas más australes (e.g., 7330001, 8104001). La correlación de Spearman (ρ) presenta mejoras en algunas cuencas y configuraciones (e.g., 10% para C1.E1 y C2.E1 en cuenca 7103001) pero, en general, los valores están en torno a la equivalencia entre el caso OL y DA. Por otro lado, el SM_{Root} muestra mejoras, principalmente, en la correlación de Spearman (hasta un 80% respecto al caso OL), la coincidencia de la señal espacial (α ; hasta un 60%) y, en algunas cuencas y configuraciones, en la razón de variabilidad (hasta un 40%).

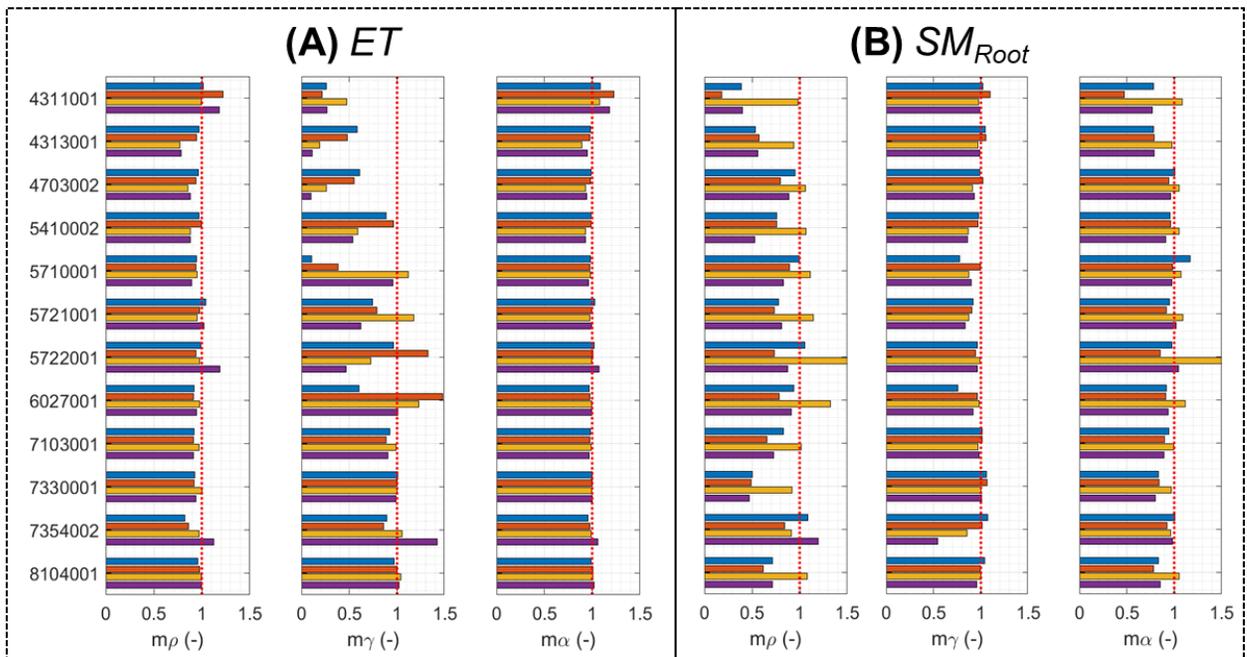


Figura 5.15: Ganancia normalizada de asimilar datos en las componentes de la métrica E_{sp} respecto al caso sin asimilación. Se incluyen resultados para ET (izquierda) y SM_{root} (derecha). La línea roja segmentada representa el valor en que los resultados sin asimilación y con asimilación son equivalentes. El valor máximo en el eje X se ha truncado en 1.5 con fines comparativos.

Tomando en cuenta las limitaciones de los experimentos realizados, resulta interesante obtener mejoras respecto a los patrones espaciales de algunas variables y estados modelados. No obstante, se debe considerar que los datos empleados como referencia (i.e., GLEAM) presentan sesgos que, a pesar de ser minimizados por la formulación de las funciones de evaluación, se mantienen. Ejemplo de ello pueden ser errores en la señal de GLEAM, errores inducidos por la interpolación a la grilla de modelación, entre otros. A pesar de ello, la comparación permite tener ideas sobre la coherencia que existe entre las simulaciones realizadas con VIC y la referencia.

5.3.3 Implicancias de la asimilación de datos en la caracterización del año hidrológico 2000/2001

5.3.3.1 Cambios en flujos y estados simulados

Para evaluar los impactos de la asimilación de datos en los diferentes flujos y estados simulados, se analizan los cambios porcentuales a nivel diario en el periodo de interés, agregados espacialmente a escala de cuenca, considerando como referencia el promedio del ensemble sin asimilación de datos (i.e., *Open-Loop*). En otras palabras, para cada miembro del ensemble, dada una variable “*i*”, se estima la diferencia entre el caso DA y OL, y se divide por el valor promedio del miembro OL (ecuación (5.1)).

$$\Delta i = 100 \times \frac{DA - OL}{OL} \quad (5.1)$$

Por otro lado, tomando en cuenta que no existe una correspondencia directa entre los escenarios definidos por los ensembles OL y DA, se utilizan los percentiles de la muestra (i.e., se asocia una probabilidad de no excedencia a cada miembro) para comparar en forma coherente los diferentes casos (i.e., se emparejan los miembros de acuerdo a su número de orden). A su vez, se cuantifican los cambios porcentuales antes mencionados para el año hidrológico 2000/01 completo. Las variables comparadas corresponden a escorrentía directa (QD), flujo base (FB), humedad del suelo por capa (i.e., SM1, SM2 y SM3), evapotranspiración real (ET), sublimación (SUB), evapotranspiración potencial de suelo saturado (PET), y recarga⁶ (R). Los resultados obtenidos para el año hidrológico completo se presentan en la Figura 5.16. De la Figura 5.16 se desprende que:

- Entre las cuencas 4311001 y 5721001, y 7354002, la asimilación de datos, en general, la escorrentía directa (QD) aumenta, principalmente cuando se adopta la configuración C2.E3, que logra ajustar las estimaciones de escorrentía total y SWE a las observaciones.
- El flujo base (FB) disminuye en algunas cuencas (e.g., 5710001) y aumenta en otras ubicadas, principalmente, en la parte sur del dominio (e.g., 7330001).
- La asimilación de datos no afecta en gran medida ($|\Delta| < 6\%$) los montos asociados a la humedad de suelo (SM) en las diferentes capas de suelo.
- La evapotranspiración real (ET) y sublimación (SUB) presentan diferencias de hasta un 40% y 20% respectivamente en comparación al caso OL. No se observa un patrón claro de variación entre las diferentes cuencas y configuraciones de asimilación.
- La recarga (R) aumenta respecto al caso OL cuando se aplica DA, siguiendo, por construcción, el patrón evidenciado en el flujo base.

En resumen, las variables que mayores diferencias presentan al momento de asimilar datos corresponden a la escorrentía directa (QD), flujo base (FB), evapotranspiración (ET), sublimación (SUB) y recarga (R).

⁶ Variable no modelada directamente por VIC, la cual es estimada a partir de una conceptualización del balance de masa en la última capa de suelo. Para mayor detalle, se refiere al lector al Anexo B.

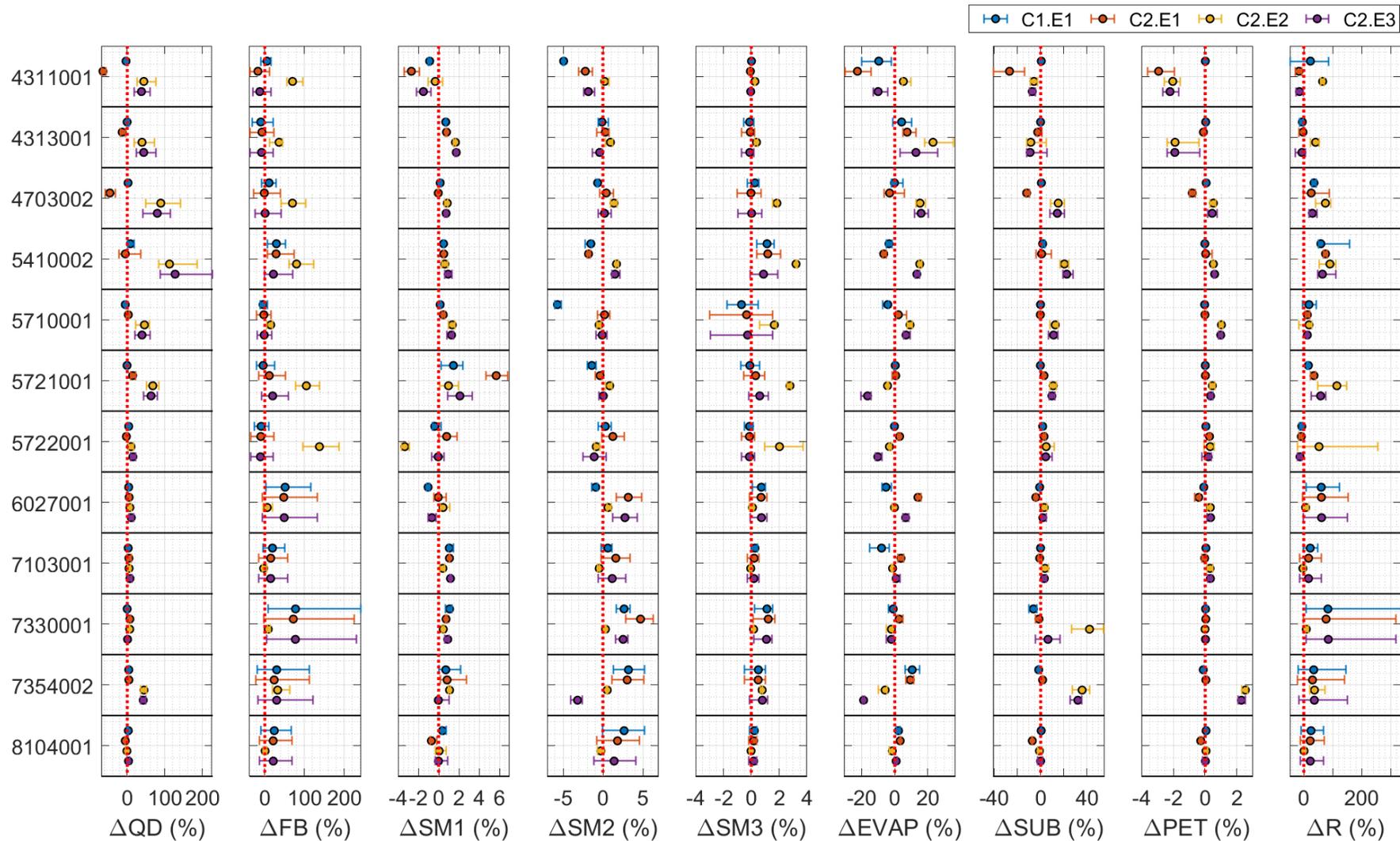


Figura 5.16: Diferencias porcentuales en flujos y estados derivados de VIC entre las configuraciones de asimilación de datos y los casos sin asimilación. La línea roja segmentada representa el caso en que no hay diferencias entre las simulaciones con y sin asimilación de datos. La extensión de las barras representa el percentil del 25% y 75% asociado a las diferencias estimadas para cada miembro evaluado a nivel diario en el periodo dado por el año hidrológico 2000/01. El punto al centro de cada una de las barras representa la mediana del ensemble.

5.3.3.2 Cambios en la estimación del balance hídrico

En términos simples, el balance hidrológico en una cuenca se puede estimar como se muestra en la ecuación (5.2), donde ΔS corresponde a la variación de almacenamiento en el tiempo Δt , \bar{P} es la precipitación media sobre la cuenca, \overline{ET} la evapotranspiración real promedio a escala de cuenca, y \bar{Q} la escorrentía. Más aún, para periodos de tiempo largos (i.e., climatológicos), y suponiendo que el flujo subterráneo neto es marginal, el balance a escala de cuenca se puede aproximar a cero (i.e., entradas al sistema son iguales a las salidas), con lo cual (5.2) se puede reescribir según la expresión (5.3).

$$\frac{\Delta S}{\Delta t} = \bar{P} - \overline{ET} - \bar{Q} \quad (5.2)$$

$$\frac{\overline{ET}}{\bar{P}} = 1 - \frac{\bar{Q}}{\bar{P}} \quad (5.3)$$

Si bien en el presente estudio sólo se modela un año hidrológico (i.e., no necesariamente se cumple el supuesto $\Delta S \approx 0$), en la Figura 5.17 se analizan los impactos de la asimilación de datos en el balance hídrico asociado.

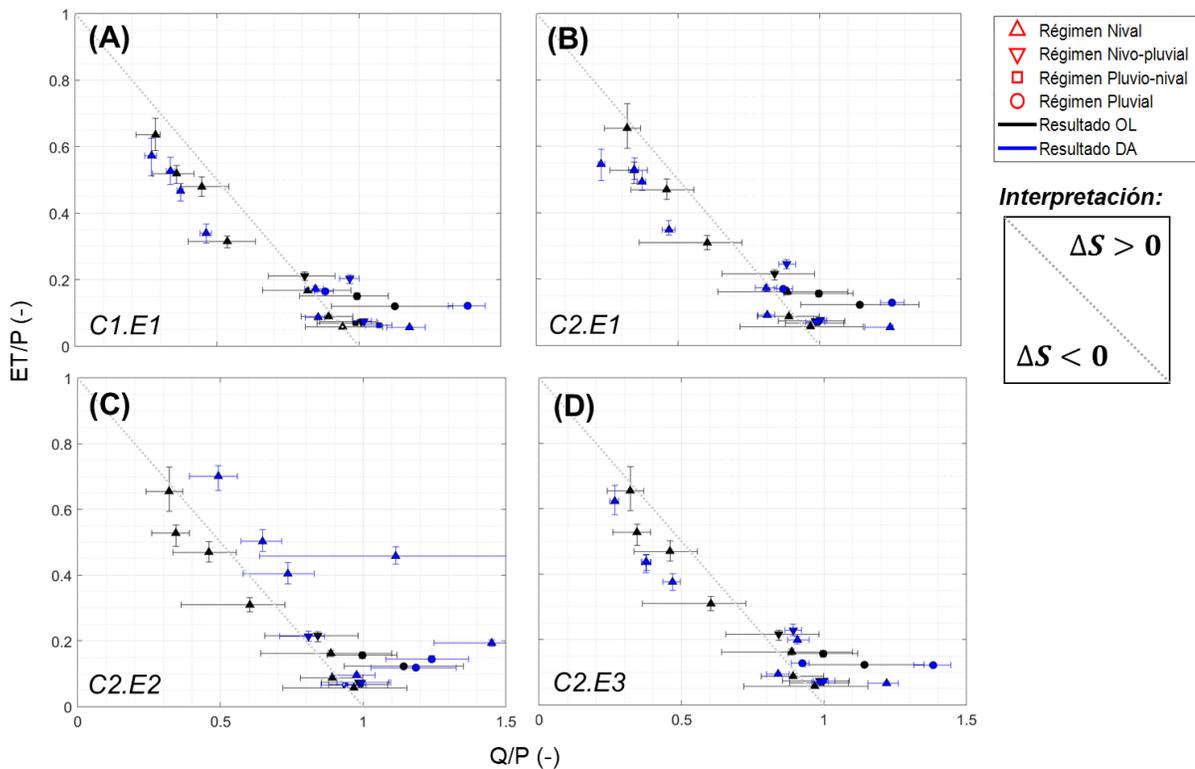


Figura 5.17: Estimación de balance hídrico para el año hidrológico 2000/2001. Las barras de error corresponden al percentil 25 y 75% derivado de los miembros del ensemble asociado a cada cuenca, mientras que el símbolo representa la mediana. (A-D) Resultados obtenidos del ejercicio de asimilación según las diferentes configuraciones propuestas.

De la Figura 5.17 se desprende que, en primer lugar, existe mayor dispersión en los montos de ET anual en cuencas áridas; dicha dispersión disminuye en cuencas húmedas (e.g., $Q/P > 0,5$), llegando a ser prácticamente nula en algunos puntos. Dicho resultado

es coherente entre las diferentes configuraciones evaluadas. En segundo lugar, se observa que la configuración C2.E2 (panel C) cambia el signo de la variación de almacenamiento asociada al balance hidrológico. En otras palabras, la cuenca pasa de un balance negativo a uno positivo al asimilar datos, lo cual se evidencia en casi todas las cuencas. No obstante, se debe considerar que los resultados presentados en secciones anteriores apuntan a que dicha configuración (i.e., C2.E2), al menos en lo que respecta a escorrentía, mantiene o incluso empeora el desempeño del modelo respecto al caso OL. El resto de las configuraciones (paneles A, B y D), a diferencia del panel C, no muestran cambios evidentes en el signo del almacenamiento (i.e., existe acuerdo). Para apoyar el análisis gráfico de la Figura 5.17, en la Tabla 5.1 se resumen los cambios en la partición P-ET-Q anual de las diferentes configuraciones de DA respecto a OL.

Tabla 5.1: Cambios en la partición P-ET-Q al asimilar datos según las diferentes configuraciones propuestas respecto al caso sin asimilación, desagregados por cuenca.

BNA	Régimen hidrológico	C1.E1		C2.E1		C2.E2		C2.E3	
		$\Delta ET/P$	$\Delta Q/P$						
4311001	Nival	-10%	-5%	-17%	-29%	7%	54%	-5%	-17%
4313001	Nival	8%	-14%	13%	-22%	31%	22%	22%	-22%
4703002	Nival	3%	3%	7%	-8%	19%	63%	22%	2%
5410002	Nival	-3%	25%	-4%	29%	18%	83%	15%	26%
5710001	Nival	-3%	-4%	5%	-9%	10%	10%	8%	-6%
5721001	Nival	1%	-6%	0%	0%	-5%	88%	-17%	9%
5722001	Nival	-3%	-17%	5%	-19%	-3%	142%	-7%	-18%
6027001	Nivo-pluvial	-4%	20%	14%	5%	-1%	-4%	6%	6%
7103001	Pluvio-nival	-9%	9%	4%	0%	-3%	-5%	2%	-1%
7330001	Pluvial	1%	23%	5%	10%	-3%	4%	-1%	21%
7354002	Pluvial	10%	-11%	9%	-13%	-8%	24%	-19%	-7%
8104001	Nivo-pluvial	3%	1%	5%	2%	-1%	1%	1%	1%

De la Tabla 5.1 se desprende que cada configuración de asimilación tiene efectos diferentes sobre el balance hídrico. En algunas configuraciones y cuencas, se observan similitudes en las variaciones obtenidas, pues si bien los montos cambian, el signo se mantiene para cada componente del balance. Por ejemplo, para las cuencas 4311001 y 4313001, la partición ET/P y Q/P mantiene su signo en todos los casos, salvo C2.E2. En otras palabras, se podría afirmar que existe un acuerdo entre los resultados. Sin embargo, en otras cuencas (e.g., 5722001), no existe tal acuerdo entre todas las configuraciones propuestas. Del mismo modo, resulta importante notar que, por ejemplo, la cuenca 8104001 no muestra cambios importantes (menores a $\pm 5\%$) en la partición P-ET-Q y, en consecuencia, en el balance hídrico asociado. Algo similar ocurre para las cuencas 7103001 y 5710001, donde los cambios antes mencionados no superan, en valor absoluto, el 10% del valor OL. En síntesis, la asimilación de datos puede tener diferentes impactos en el balance hídrico anual dependiendo tanto de la cuenca como de la configuración (i.e., variable asimilada). Así, la elección de una configuración óptima del esquema de DA, entendiendo la dinámica del sistema y caracterizando apropiadamente las fuentes de incertidumbre, resulta clave en favor de, efectivamente, reducir la incertidumbre asociada a flujos y estados hidrológicos y, a su vez, ser consistentes en esas estimaciones, obteniendo las respuestas correctas por las razones correctas.

6 COMENTARIOS Y CONCLUSIONES

El presente estudio tuvo por objetivo evaluar el beneficio de implementar asimilación de datos en estimaciones de flujos y estados hidrológicos en Chile continental. Para ello, se implementó secuencialmente, para cada día con observaciones en el periodo 01/04/2000 – 31/03/2001, el Filtro de Kalman de conjuntos (EnKF) acoplado al modelo hidrológico *Variable Infiltration Capacity* (VIC) en 12 cuencas con bajo nivel de intervención humana, ubicadas a lo largo de un gradiente hidroclimático entre los 30°S y 37°S. Los experimentos numéricos fueron diseñados en base a la perturbación de las forzantes del modelo (i.e., precipitación, velocidad de viento, temperatura y humedad relativa) y algunos estados iniciales (humedad de suelo por capa y equivalente de nieve en agua), generándose cuatro configuraciones de EnKF distintas. Así, a partir de los resultados obtenidos, se concluye lo siguiente:

1. Asimilar sólo Q no es suficiente para asegurar la disminución en la dispersión de otros flujos y estados del sistema. Por ejemplo, al asimilar sólo Q se empeoran en promedio para las 12 cuencas un 2% del SWE máximo las estimaciones de SWE diario. Por el contrario, al asimilar de manera conjunta Q, SWE y fSCA, se mejora tanto Q como SWE Y fSCA. En consecuencia, la incorporación de diferentes observaciones, orientadas a cuantificar diferentes procesos, permite reducir la dispersión en las estimaciones asociadas a los mismos y, con ello, al resto de los flujos y estados del sistema.
2. El esquema de asimilación permite, a pesar de no haber sido diseñado con ese fin, mejorar algunas propiedades espaciales como, por ejemplo, la razón de variabilidad, coincidencia de la señal espacial y correlación. Dichas mejoras se observan principalmente en la evapotranspiración y humedad de suelo en la zona radicular, variables que muestran mejoras de hasta un 90% y 80%, respectivamente, en la razón de variabilidad y correlación de Spearman.
3. La asimilación de datos tiene impactos en todas las variables de estado y flujos simulados. Las variables que mayores diferencias presentan al momento de asimilar datos corresponden a flujos del modelo, incluyendo la escorrentía directa (QD), flujo base (FB), evapotranspiración (ET), sublimación (SUB) y recarga (R), donde FB es el que exhibe las mayores diferencias respecto al caso OL.

Adicionalmente, existen algunos aspectos y hallazgos de interés derivados del presente estudio. Algunos de los más relevantes se presentan a continuación:

- La configuración C2.E2 (i.e., sólo asimilación de SC y SWE) no tiene un valor agregado en las estimaciones de escorrentía media diaria respecto al caso OL (ver Figura 5.9 y Figura 5.10) para las cuencas ubicadas en las zonas B y C del dominio de estudio (Figura 2.1).
- En el resto de las configuraciones, todas las cuencas reportan mejoras en términos de la escorrentía, que van entre el 10% y 90%. Lo anterior implica mejoras en el RMSE de, en promedio, $3,56 \pm 4,69$ mm/d a $3,76 \pm 4,58$ mm/d en la estimación de escorrentía media diaria en las 12 cuencas respecto al caso OL, lo que a su vez corresponde una

reducción promedio de entre $10 \pm 8 \%$ y $15 \pm 12 \%$ del caudal medio anual en cada cuenca para el año de estudio.

- Asimilar variables relacionadas al manto nival (i.e., SC y SWE) mejora entre un 70-90% las estimaciones de SWE. Lo anterior implica mejoras en el RMSE de, en promedio, 80 ± 54 mm en la estimación de SWE diario, agregado a escala de cuenca, para las 12 cuencas respecto al caso sin asimilación de SWE y fSCA. En términos de magnitud, la mejora corresponde a disminuir el error, en promedio, un $19 \pm 8 \%$ del SWE máximo de la temporada 2000/01 para cada cuenca.
- En ninguna de las configuraciones analizadas se mejora en gran medida la cobertura de nieve respecto a las observaciones MODIS. En efecto, los cambios no superan el 3% de la cobertura de nieve fraccional respecto al caso OL.
- La asimilación conjunta de Q, SC y SWE (i.e., C2.E3) permite mejorar, simultáneamente, las estimaciones de Q y SWE. En promedio para las 12 cuencas del dominio, el RMSE promedio de los ensemble mejora $3,76 \pm 4,58$ mm/d en términos de la escorrentía media diaria y $80,79 \pm 54,12$ mm para el SWE diario promedio a escala de cuenca. En consecuencia, se identifica como la mejor configuración de las testeadas, pues reduce el error de la escorrentía media diaria, en promedio, en un 15% [7 – 48%] del caudal medio anual, y el SWE diario en un 19% [9 – 35%] respecto al SWE máximo de la temporada.
- La metodología adoptada logra mejorar los patrones espaciales de las variables simuladas asimilando a escala de cuenca, a pesar de no estar diseñada para ello. Por ejemplo, la razón de variabilidad (γ) de ET, principalmente para las cuencas ubicadas al norte del dominio (i.e., zona B y C, Figura 2.1), alcanzando valores entre 10-90% para algunas cuencas y configuraciones evaluadas (e.g., C2.E3, 4703002; C1.E1, 5710001).
- La asimilación de datos tiene impactos sobre todas las salidas del modelo. Si bien, en términos de montos, algunas componentes no cambian en forma considerable (e.g., $\Delta SM1$, $\Delta SM2$ y $\Delta SM3 < \pm 6\%$; Figura 5.16), la asimilación impacta en la variabilidad y temporalidad de las mismas.
- Cada configuración de asimilación de datos tiene efectos diferentes sobre el balance hídrico (Figura 5.17). No obstante, aquellas que muestran un mejor desempeño en las estimaciones de Q (i.e., C1.E1, C2.E1, y C2.E3), muestran coherencia entre sí en cuanto al signo del balance.
- En algunas configuraciones y cuencas se observan similitudes en las variaciones obtenidas para el balance hídrico (Figura 5.17 y Tabla 5.1). Si bien los montos cambian, el signo se mantiene para cada componente del balance. Por el contrario, en otras configuraciones y cuencas no existe tal acuerdo. Más aún, en algunas cuencas no existen variaciones absolutas mayores al 5% (i.e., marginal).

En síntesis, dada la variabilidad de los resultados obtenidos entre las diferentes configuraciones de asimilación, resulta necesario definir cuál es la mejor. Para ello, en el presente estudio se aplicaron diferentes métricas de evaluación contra las cuales se compararon las observaciones disponibles respecto a las simulaciones con asimilación (DA) y sin asimilación (OL). Los resultados obtenidos de esta investigación indican que la mejor configuración corresponde a la C2.E3, es decir, cuando se asimila escorrentía

media diaria (Q), cobertura de nieve (SC o fSCA) y equivalente de nieve en agua (SWE), para actualizar la humedad de suelo inicial por capa y el equivalente de nieve en agua inicial.

Por otro lado, se debe tener en cuenta que el presente estudio cuenta con algunas limitaciones derivadas de los supuestos realizados para la configuración de los experimentos realizados. Así, surgen las siguientes ideas para mejorar el presente trabajo:

- Actualización de la distribución espacial de estados: en el presente estudio se trabajó en forma agregada la actualización de estados, por lo que, se propone adoptar un enfoque de actualización de estados distribuida. Así, dada una observación, se actualiza una matriz donde se incluyen los N estados y Z píxeles, similar a lo propuesto por Rakovec et al. (2015). No obstante, para ello se debe incorporar un esquema de rastreo de caudales, que permita conocer el caudal en cualquier punto de la red de drenaje, o bien extraer del conjunto de cuencas aquellas donde el tiempo de concentración es mucho mayor a 1 día (i.e., donde el supuesto de que toda el agua que – a través del flujo base y escorrentía directa – sale de un píxel, llega instantáneamente a la salida de la cuenca (i.e., *outlet*) sea inválido).
- Aplicar curva de agotamiento (*Snow Depletion Curve* – SDC) en la actualización de estados: tomando en cuenta que se perturba y actualiza el SWE inicial en cada paso de tiempo, se propone la incorporación de SDC que permitan llevar los cambios de SWE a cambios en la cobertura de nieve a nivel de cuenca y/o píxel.
- Incorporar observaciones de ET: tomando en cuenta que el desempeño del modelo mejoró considerablemente al asimilar Q, SWE y SC, se propone la incorporación de ET para mejorar la estimación de la partición ET/P y, con ello, el balance hídrico.
- Mejorar representación de errores de observación y modelo: en el presente estudio, se definieron valores de error a priori, los cuales no aseguran un ensemble óptimo. Así, se plantea incorporar un enfoque de calibración de errores a partir de una “Estimación máxima a posteriori” (MAP; Anexo B).

Finalmente, se concluye que, si bien la asimilación de datos permite corregir y actualizar los montos y temporalidad de los estados y flujos hidrológicos simulados, disminuyendo así la dispersión asociada, la evaluación del desempeño del modelo es clave para asegurar la consistencia del proceso de asimilación de datos. Para ello, se recomienda contrastar los resultados contra diferentes observaciones, incluyendo datos independientes del proceso de asimilación, a modo de asegurar la correcta representación de los procesos hidrológicos que gobiernan el sistema.

BIBLIOGRAFÍA

- Alvarez-Garreton, C., Mendoza, P.A., Pablo Boisier, J., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Puelma, C., Cortes, G., Garreaud, R., McPhee, J., Ayala, A., 2018. The CAMELS-CL dataset: Catchment attributes and meteorology for large sample studies-Chile dataset. *Hydrol. Earth Syst. Sci.* 22, 5817–5846. <https://doi.org/10.5194/hess-22-5817-2018>
- Alvarez-Garreton, C., Ryu, D., Western, A.W., Crow, W.T., Robertson, D.E., 2014. The impacts of assimilating satellite soil moisture into a rainfall-runoff model in a semi-arid catchment. *J. Hydrol.* 519, 2763–2774. <https://doi.org/10.1016/j.jhydrol.2014.07.041>
- Alvarez-Garreton, C., Ryu, D., Western, A.W., Crow, W.T., Su, C.H., Robertson, D.R., 2016. Dual assimilation of satellite soil moisture to improve streamflow prediction in data-scarce catchments. *Water Resour. Res.* 52, 5357–5375. <https://doi.org/10.1002/2015WR018429>
- Alvarez-Garreton, C., Ryu, D., Western, A.W., Su, C.H., Crow, W.T., Robertson, D.E., Leahy, C., 2015. Improving operational flood ensemble prediction by the assimilation of satellite soil moisture: Comparison between lumped and semi-distributed schemes. *Hydrol. Earth Syst. Sci.* 19, 1659–1676. <https://doi.org/10.5194/hess-19-1659-2015>
- Bulygina, N., Gupta, H., 2009. Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation. *Water Resour. Res.* 45, 1–20. <https://doi.org/10.1029/2007WR006749>
- Burgers, G., Van Leeuwen, P.J., Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.* 126, 1719–1724. [https://doi.org/10.1175/1520-0493\(1998\)126<1719:ASITEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2)
- Butts, M.B., Payne, J.T., Kristensen, M., Madsen, H., 2004. An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *J. Hydrol.* 298, 242–266. <https://doi.org/10.1016/j.jhydrol.2004.03.042>
- Clark, M.P., Nijssen, B., Lundquist, J.D., Kavetski, D., Rupp, D.E., Woods, R.A., Freer, J.E., Gutmann, E.D., Wood, A.W., Brekke, L.D., Arnold, J.R., Gochis, D.J., Rasmussen, R.M., 2015. A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resour. Res.* <https://doi.org/10.1002/2015WR017198>
- Clark, M.P., Rupp, D.E., Woods, R.A., Zheng, X., Ibbitt, R.P., Slater, A.G., Schmidt, J., Uddstrom, M.J., 2008. Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. *Adv. Water Resour.* 31, 1309–1324. <https://doi.org/10.1016/j.advwatres.2008.06.005>
- Clark, M.P., Slater, A.G., Barrett, A.P., Hay, L.E., McCabe, G.J., Rajagopalan, B., Leavesley, G.H., 2006. Assimilation of snow covered area information into hydrologic and land-surface models. *Adv. Water Resour.* 29, 1209–1221. <https://doi.org/10.1016/j.advwatres.2005.10.001>
- Copernicus Climate Change Service (C3S), 2017. ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate [WWW Document]. Copernicus Clim. Chang. Serv. Clim. Data Store. URL <https://cds.climate.copernicus.eu/cdsapp#!/home>
- Cortés, G., Giroto, M., Margulis, S., 2016. Snow process estimation over the extratropical Andes using a data assimilation framework integrating MERRA data and Landsat imagery. *Water Resour. Res.* 52, 2582–2600. <https://doi.org/10.1002/2015WR018376>
- Cortés, G., Margulis, S., 2017. Impacts of El Niño and La Niña on interannual snow accumulation in the Andes: Results from a high-resolution 31 year reanalysis. *Geophys. Res. Lett.* 44, 6859–6867. <https://doi.org/10.1002/2017GL073826>
- Cortés, G., Vargas, X., McPhee, J., 2011. Climatic sensitivity of streamflow timing in the extratropical western Andes Cordillera. *J. Hydrol.* 405, 93–109. <https://doi.org/10.1016/j.jhydrol.2011.05.013>
- De Lannoy, G.J.M., de Rosnay, P., Reichle, R.H., 2019. Soil Moisture Data Assimilation, in: *Handbook of Hydrometeorological Ensemble Forecasting*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 701–743. https://doi.org/10.1007/978-3-642-39925-1_32

- De Lannoy, G.J.M., Houser, P.R., Pauwels, V.R.N., Verhoest, N.E.C., 2006. Assessment of model uncertainty for soil moisture through ensemble verification. *J. Geophys. Res. Atmos.* 111. <https://doi.org/10.1029/2005JD006367>
- Delle Monache, L., Hacker, J.P., Zhou, Y., Deng, X., Stull, R.B., 2006. Probabilistic aspects of meteorological and ozone regional ensemble forecasts. *J. Geophys. Res. Atmos.* 111, 1–15. <https://doi.org/10.1029/2005JD006917>
- Dembélé, M., Ceperley, N., Zwart, S.J., Salvadore, E., Mariethoz, G., Schaefli, B., 2020a. Potential of satellite and reanalysis evaporation datasets for hydrological modelling under various model calibration strategies. *Adv. Water Resour.* 143. <https://doi.org/10.1016/j.advwatres.2020.103667>
- Dembélé, M., Hrachowitz, M., Savenije, H.H.G., Mariéthoz, G., Schaefli, B., 2020b. Improving the Predictive Skill of a Distributed Hydrological Model by Calibration on Spatial Patterns With Multiple Satellite Data Sets. *Water Resour. Res.* 56, 1–26. <https://doi.org/10.1029/2019WR026085>
- DGA, 2019a. Aplicación de la metodología de actualización del balance hídrico nacional en las cuencas de la parte sur de la Macrozona Austral e Isla de Pascua, SIT N° 444.
- DGA, 2019b. Aplicación de la metodología de actualización del balance hídrico nacional en la macrozona sur y parte norte de la macrozona Austral, SIT N° 441.
- DGA, 2018. Aplicación de La Metodología de Actualización del Balance Hídrico Nacional en las Cuencas de la Macrozona Norte y Centro.
- DGA, 2017. Actualización del Balance Hídrico Nacional, SIT N° 417, Ministerio de Obras Públicas, Dirección General de Aguas, División de Estudios y Planificación, Santiago, Chile. Realizado por: Universidad de Chile & Pontificia Universidad Católica de Chile.
- Döll, P., Douville, H., Güntner, A., Müller Schmied, H., Wada, Y., 2016. Modelling Freshwater Resources at the Global Scale: Challenges and Prospects. *Surv. Geophys.* 37, 195–221. <https://doi.org/10.1007/s10712-015-9343-1>
- Duan, Q., Gupta, V., 1992. Effective and Efficient Global Optimization for conceptual rainfall-runoff models. *Water Resour. Res.* 28, 1015–1031.
- Duethmann, D., Bloschl, G., Parajka, J., 2020. Why does a conceptual hydrological model fail to correctly predict discharge changes in response to climate change? *Hydrol. Earth Syst. Sci.* 24, 3493–3511. <https://doi.org/10.5194/hess-24-3493-2020>
- Elsner, M.M., Gangopadhyay, S., Pruitt, T., Brekke, L.D., Mizukami, N., Clark, M.P., 2014. How Does the Choice of Distributed Meteorological Data Affect Hydrologic Model Calibration and Streamflow Simulations? *J. Hydrometeorol.* 15, 1384–1403. <https://doi.org/10.1175/jhm-d-13-083.1>
- Emerick, A.A., Reynolds, A.C., 2012. History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. *Comput. Geosci.* 16, 639–659. <https://doi.org/10.1007/s10596-012-9275-5>
- Entekhabi, D., Njoku, E.G., O'Neill, P.E., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T.J., Johnson, J., Kimball, J., Piepmeier, J.R., Koster, R.D., Martin, N., McDonald, K.C., Moghaddam, M., Moran, S., Reichle, R., Shi, J.C., Spencer, M.W., Thurman, S.W., Tsang, L., Van Zyl, J., 2010. The Soil Moisture Active Passive (SMAP) Mission. *Proc. IEEE* 98, 704–716. <https://doi.org/10.1109/JPROC.2010.2043918>
- Evensen, G., 2009. Data assimilation: The ensemble kalman filter, *Data Assimilation (Second Edition): The Ensemble Kalman Filter*. <https://doi.org/10.1007/978-3-642-03711-5>
- Evensen, G., 1994. Inverse methods and data assimilation in nonlinear ocean models. *Phys. D Nonlinear Phenom.* 77, 108–129. [https://doi.org/10.1016/0167-2789\(94\)90130-9](https://doi.org/10.1016/0167-2789(94)90130-9)
- Freudiger, D., Kohn, I., Seibert, J., Stahl, K., Weiler, M., 2017. Snow redistribution for the hydrological modeling of alpine catchments. *Wiley Interdiscip. Rev. Water* 4, e1232. <https://doi.org/10.1002/wat2.1232>
- Garreaud, R.D., Alvarez-Garreton, C., Barichivich, J., Boisier, J.P., Christie, D., Galleguillos, M., LeQuesne, C., McPhee, J., Zambrano-Bigiarini, M., 2017. The 2010–2015 megadrought in central Chile: impacts on regional hydroclimate and vegetation. *Hydrol. Earth Syst. Sci.* 21,

- 6307–6327. <https://doi.org/10.5194/hess-21-6307-2017>
- Garreaud, R.D., Boisier, J.P., Rondanelli, R., Montecinos, A., Sepúlveda, H.H., Veloso-Aguila, D., 2019. The Central Chile Mega Drought (2010–2018): A climate dynamics perspective. *Int. J. Climatol.* 40, 421–439. <https://doi.org/10.1002/joc.6219>
- Gillijns, S., Mendoza, O.B., Chandrasekar, J., De Moor, B.L.R., Bernstein, D.S., Ridley, A., 2006. What is the ensemble Kalman filter and how well does it work? *Proc. Am. Control Conf.* 2006, 4448–4453. <https://doi.org/10.1109/acc.2006.1657419>
- Giroto, M., De Lannoy, G.J.M., Reichle, R.H., Rodell, M., 2016. Assimilation of gridded terrestrial water storage observations from GRACE into a land surface model. *Water Resour. Res.* 52, 4164–4183. <https://doi.org/10.1002/2015WR018417>
- Götzinger, J., Bárdossy, A., 2008. Generic error model for calibration and uncertainty estimation of hydrological models. *Water Resour. Res.* 44, 1–18. <https://doi.org/10.1029/2007wr006691>
- Gupta, H. V., Clark, M.P., Vrugt, J.A., Abramowitz, G., Ye, M., 2012. Towards a comprehensive assessment of model structural adequacy. *Water Resour. Res.* 48, 1–16. <https://doi.org/10.1029/2011WR011044>
- Gupta, H. V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* 377, 80–91.
- Hall, D.K., Riggs, G.A., Salomonson, V. V., Digirolamo, N.E., Bayr, K.J., 2002. MODIS snow-cover products 83, 181–194.
- Hamill, T.M., 2001. Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Weather Rev.* 129, 550–560. [https://doi.org/10.1175/1520-0493\(2001\)129<0550:IORHFV>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2)
- Hamman, J.J., Nijssen, B., Bohn, T.J., Gergel, D.R., Mao, Y., 2018. The variable infiltration capacity model version 5 (VIC-5): Infrastructure improvements for new applications and reproducibility. *Geosci. Model Dev.* 11, 3481–3496. <https://doi.org/10.5194/gmd-11-3481-2018>
- Huffman, G.J., Stocker, E.F., Bolvin, D.T., Nelkin, E.J., Tan, J., 2019. GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06, Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC). <https://doi.org/10.5067/GPM/IMERG/3B-HH/06>
- Ibaceta, R., Splinter, K.D., Harley, M.D., Turner, I.L., 2020. Enhanced Coastal Shoreline Modeling Using an Ensemble Kalman Filter to Include Nonstationarity in Future Wave Climates. *Geophys. Res. Lett.* 47, 1–12. <https://doi.org/10.1029/2020GL090724>
- INE, 2017. Resultados Censo 2017. Santiago, Chile.
- Kerr, Y.H., Waldteufel, P., Wigneron, J.P., Martinuzzi, J.M., Font, J., Berger, M., 2001. Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Trans. Geosci. Remote Sens.* 39, 1729–1735. <https://doi.org/10.1109/36.942551>
- Klemeš, V., 1986. Operational testing of hydrological simulation models. *Hydrol. Sci. J.* 31, 13–24. <https://doi.org/10.1080/02626668609491024>
- Kumar, S. V., Zaitchik, B.F., Peters-Lidard, C.D., Rodell, M., Reichle, R., Li, B., Jasinski, M., Mocko, D., Getirana, A., De Lannoy, G., Cosh, M.H., Hain, C.R., Anderson, M., Arsenault, K.R., Xia, Y., Ek, M., 2016. Assimilation of Gridded GRACE Terrestrial Water Storage Estimates in the North American Land Data Assimilation System. *J. Hydrometeorol.* 17, 1951–1972. <https://doi.org/10.1175/jhm-d-15-0157.1>
- Lettenmaier, D.P., Alsdorf, D., Dozier, J., Huffman, G.J., Pan, M., Wood, E.F., 2015. Inroads of remote sensing into hydrologic science during the WRR era. *Water Resour. Res.* 51, 7309–7342. <https://doi.org/10.1002/2015WR017616>
- Li, H., Luo, L., Wood, E.F., Schaake, J., 2009. The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting. *J. Geophys. Res. Atmos.* 114, 1–10. <https://doi.org/10.1029/2008JD010969>
- Li, Y., Ryu, D., Western, A.W., Wang, Q.J., Robertson, D.E., Crow, W.T., 2014. An integrated error parameter estimation and lag-aware data assimilation scheme for real-time flood

- forecasting. *J. Hydrol.* 519, 2722–2736. <https://doi.org/10.1016/j.jhydrol.2014.08.009>
- Liang, X., Lettenmaier, D.P., Wood, E.F., Burges, S.J., 2004. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res.* 99, 14415. <https://doi.org/10.1029/94jd00483>
- Liang, X., Wood, E.F., Lettenmaier, D.P., 1996. Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification. *Glob. Planet. Change* 13, 195–206. [https://doi.org/10.1016/0921-8181\(95\)00046-1](https://doi.org/10.1016/0921-8181(95)00046-1)
- Liu, Y., Gupta, H. V., 2007. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resour. Res.* 43, 1–18. <https://doi.org/10.1029/2006WR005756>
- Lohmann, D., Mitchell, K.E., Houser, P.R., Wood, E.F., Schaake, J.C., Robock, A., Cosgrove, B.A., Sheffield, J., Duan, Q., Luo, L., Higgins, R.W., Pinker, R.T., Tarpley, J.D., 2004. Streamflow and water balance intercomparisons of four land surface models in the North American Land Data Assimilation System project. *J. Geophys. Res. Atmos.* 109, 1–22. <https://doi.org/10.1029/2003jd003517>
- Margulis, S.A., Cortés, G., Girotto, M., Durand, M., 2016. A landsat-era Sierra Nevada snow reanalysis (1985-2015). *J. Hydrometeorol.* 17, 1203–1221. <https://doi.org/10.1175/JHM-D-15-0177.1>
- Margulis, S.A., Entekhabi, D., McLaughlin, D., 2006. Spatiotemporal Disaggregation of Remotely Sensed Precipitation for Ensemble Hydrologic Modeling and Data Assimilation. *J. Hydrometeorol.* 7, 511–533. <https://doi.org/10.1175/JHM492.1>
- Margulis, S.A., McLaughlin, D., Entekhabi, D., Dunne, S., 2002. Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment. *Water Resour. Res.* 38, 35-1-35–18. <https://doi.org/10.1029/2001wr001114>
- Martens, B., Miralles, D.G., Lievens, H., Van Der Schalie, R., De Jeu, R.A.M., Fernández-Prieto, D., Beck, H.E., Dorigo, W.A., Verhoest, N.E.C., 2017. GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geosci. Model Dev.* 10, 1903–1925. <https://doi.org/10.5194/gmd-10-1903-2017>
- McCabe, M.R., Alsdorf, D.E., Miralles, D.G., Uijlenhoet, R., Wagner, W., Lucieer, A., Houborg, R., Niko E. C. Verhoest, Trenton E. Franz, J.S., Gao, H., Wood, E.F., 2017. The Future of Earth Observation in Hydrology. *Hydrology and Earth System Sciences (under review)*. *Hydrol. Earth Syst. Sci. (under Rev.* 3879–3914.
- McMillan, H., Jackson, B., Clark, M., Kavetski, D., Woods, R., 2011. Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models. *J. Hydrol.* 400, 83–94. <https://doi.org/10.1016/j.jhydrol.2011.01.026>
- Mendoza, Pablo A, Clark, M.P., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G., Gupta, H., 2015. Are we unnecessarily constraining the agility of complex process-based models? *Water Resour. Res.* 51. <https://doi.org/10.1002/2014WR015820>
- Mendoza, P.A., McPhee, J., Vargas, X., 2012. Uncertainty in flood forecasting: A distributed modeling approach in a sparse data catchment. *Water Resour. Res.* 48. <https://doi.org/10.1029/2011WR011089>
- Mendoza, Pablo A., Rajagopalan, B., Clark, M.P., Ikeda, K., Rasmussen, R.M., 2015. Statistical Postprocessing of High-Resolution Regional Climate Model Output. *Mon. Weather Rev.* 143, 1533–1553. <https://doi.org/10.1175/MWR-D-14-00159.1>
- Montanari, A., Di Baldassarre, G., 2013. Data errors and hydrological modelling: The role of model structure to propagate observation uncertainty. *Adv. Water Resour.* 51, 498–504. <https://doi.org/10.1016/j.advwatres.2012.09.007>
- Moradkhani, H., Hsu, K.L., Gupta, H., Sorooshian, S., 2005a. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resour. Res.* 41, 1–17. <https://doi.org/10.1029/2004WR003604>
- Moradkhani, H., Nearing, G.S., Abbaszadeh, P., Pathiraja, S., 2019. Fundamentals of Data Assimilation and Theoretical Advances, in: *Handbook of Hydrometeorological Ensemble Forecasting*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 675–699.

- https://doi.org/10.1007/978-3-642-39925-1_30
- Moradkhani, H., Sorooshian, S., Gupta, H. V., Houser, P.R., 2005b. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Adv. Water Resour.* 28, 135–147. <https://doi.org/10.1016/j.advwatres.2004.09.002>
- Mu, Q., Zhao, M., Running, S.W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sens. Environ.* 115, 1781–1800. <https://doi.org/10.1016/j.rse.2011.02.019>
- Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F.T., Flörke, M., Döll, P., 2014. Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. *Hydrol. Earth Syst. Sci.* 18, 3511–3538. <https://doi.org/10.5194/hess-18-3511-2014>
- Muñoz, A.A., Klock-Barría, K., Alvarez-Garretón, C., Aguilera-Betti, I., González-Reyes, Á., Lastra, J.A., Chávez, R.O., Barría, P., Christie, D., Rojas-Badilla, M., Lequesne, C., 2020. Water crisis in petorca basin, Chile: The combined effects of a mega-drought and water management. *Water (Switzerland)* 12. <https://doi.org/10.3390/w12030648>
- Murphy, A.H., 1986. A New Decomposition of the Brier Score: Formulation and Interpretation. *Mon. Weather Rev.* 114, 2671–2673. [https://doi.org/10.1175/1520-0493\(1986\)114<2671:ANDOTB>2.0.CO;2](https://doi.org/10.1175/1520-0493(1986)114<2671:ANDOTB>2.0.CO;2)
- Nash, J., Sutcliffe, J., 1970. River flow forecasting through conceptual models part I - A discussion of principles. *J. Hydrol.* 10, 282–290.
- Noh, S.J., Rakovec, O., Weerts, A.H., Tachikawa, Y., 2014. On noise specification in data assimilation schemes for improved flood forecasting using distributed hydrological models. *J. Hydrol.* 519, 2707–2721. <https://doi.org/10.1016/j.jhydrol.2014.07.049>
- Pan, M., Wood, E.F., 2006. Data assimilation for estimating the terrestrial water budget using a constrained ensemble Kalman filter. *J. Hydrometeorol.* 7, 534–547. <https://doi.org/10.1175/JHM495.1>
- Pan, M., Wood, E.F., Wójcik, R., McCabe, M.F., 2008. Estimation of regional terrestrial water cycle using multi-sensor remote sensing observations and data assimilation. *Remote Sens. Environ.* 112, 1282–1294. <https://doi.org/10.1016/j.rse.2007.02.039>
- Pardé, M., 1940. Quelques notes sur l'hydrologie du Brahmapoutre, du Gange et de l'Indus. *Rev. Géographie Alp.* 28, 157–198.
- Pardé, M., 1933. Fleuves et rivières. *Collect. Arman. Colin. Sect. Géographie* fre no. 155.
- Pauwels, V.R.N., De Lannoy, G.J.M., 2009. Ensemble-based assimilation of discharge into rainfall-runoff models: A comparison of approaches to mapping observational information to state space. *Water Resour. Res.* 45, W08428. <https://doi.org/10.1029/2008WR007590>
- Pokhrel, P., Gupta, H. V., 2010. On the use of spatial regularization strategies to improve calibration of distributed watershed models. *Water Resour. Res.* 46, W01505. <https://doi.org/10.1029/2009WR008066>
- Ragettli, S., Cortés, G., Mcphee, J., Pellicciotti, F., 2014. An evaluation of approaches for modelling hydrological processes in high-elevation, glacierized Andean watersheds. *Hydrol. Process.* 28, 5674–5695. <https://doi.org/10.1002/hyp.10055>
- Rakovec, O., Weerts, A.H., Sumihar, J., Uijlenhoet, R., 2015. Operational aspects of asynchronous filtering for flood forecasting. *Hydrol. Earth Syst. Sci.* 19, 2911–2924. <https://doi.org/10.5194/hess-19-2911-2015>
- Rasmussen, J., Madsen, H., Jensen, K.H., Refsgaard, J.C., 2015. Data assimilation in integrated hydrological modeling using ensemble Kalman filtering: evaluating the effect of ensemble size and localization on filter performance. *Hydrol. Earth Syst. Sci.* 19, 2999–3013. <https://doi.org/10.5194/hess-19-2999-2015>
- Reichle, R.H., 2008. Data assimilation methods in the Earth sciences. *Adv. Water Resour.* 31, 1411–1418. <https://doi.org/10.1016/j.advwatres.2008.01.001>
- Revilla-Romero, B., Wanders, N., Burek, P., Salamon, P., de Roo, A., 2016. Integrating remotely sensed surface water extent into continental scale hydrology. *J. Hydrol.* 543, 659–670. <https://doi.org/10.1016/j.jhydrol.2016.10.041>

- Samadi, S., Tufford, D.L., Carbone, G.J., 2018. Estimating hydrologic model uncertainty in the presence of complex residual error structures. *Stoch. Environ. Res. Risk Assess.* 32, 1259–1281. <https://doi.org/10.1007/s00477-017-1489-6>
- Siripatana, A., Giraldi, L., Le Maître, O.P., Knio, O.M., Hoteit, I., 2019. Combining ensemble Kalman filter and multiresolution analysis for efficient assimilation into adaptive mesh models. *Comput. Geosci.* 23, 1259–1276. <https://doi.org/10.1007/s10596-019-09882-z>
- Stowhas, L., 2017. *Fundamentos de Hidrología Aplicada, Segunda Ed.* ed. Sello Editorial USM, Valparaíso.
- Tapley, B.D., Bettadpur, S., Watkins, M., Reigber, C., 2004. The gravity recovery and climate experiment: Mission overview and early results. *Geophys. Res. Lett.* 31, 1–4. <https://doi.org/10.1029/2004GL019920>
- van Werkhoven, K., Wagener, T., Reed, P., Tang, Y., 2008. Characterization of watershed model behavior across a hydroclimatic gradient. *Water Resour. Res.* 44, 1–16. <https://doi.org/10.1029/2007WR006271>
- Vásquez, N., Cepeda, J., Gómez, T., Mendoza, P.A., Lagos, M., Boisier, J.P., Álvarez-Garretón, C., Vargas, X., 2021. Catchment-Scale Natural Water Balance in Chile, in: *Water Resources of Chile*. pp. 189–208. https://doi.org/10.1007/978-3-030-56901-3_9
- Vrugt, J.A., Dekker, S.C., Bouten, W., 2003. Identification of rainfall interception model parameters from measurements of throughfall and forest canopy storage. *Water Resour. Res.* 39, 1–10. <https://doi.org/10.1029/2003WR002013>
- Weerts, A.H., El Serafy, G.Y.H., 2006. Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. *Water Resour. Res.* 42, 1–17. <https://doi.org/10.1029/2005WR004093>
- Wilks, D., 2011. *Statistical methods in the atmospheric sciences*. Academic Press.
- Wood, E.F., Lettenmaier, D.P., Zartarian, V.G., 1992. A land-surface hydrology parameterization with subgrid variability for general circulation models. *J. Geophys. Res.* 97, 2717–2728. <https://doi.org/10.1029/91JD01786>
- Yilmaz, K.K., Gupta, H. V, Wagener, T., 2008. A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resour. Res.* 44. <https://doi.org/10.1029/2007WR006716>
- Zhang, H., Hendricks Franssen, H.-J., Han, X., Vrugt, J.A., Vereecken, H., 2017. State and parameter estimation of two land surface models using the ensemble Kalman filter and the particle filter. *Hydrol. Earth Syst. Sci.* 21, 4927–4958. <https://doi.org/10.5194/hess-21-4927-2017>

ANEXO A: RESULTADOS DE LA CALIBRACIÓN DEL MODELO HIDROLÓGICO VIC EN LAS CUENCAS DE ESTUDIO

Para facilitar la reproducibilidad del presente estudio, en la Tabla A.1 se presentan los valores utilizados para configurar el algoritmo de optimización global *Shuffled Complex Evolution* (SCE-UA; Duan et al., 1992). El código de SCE-UA se encuentra implementado en lenguaje fortran, y se corrió con el apoyo de la infraestructura de supercomputo Guacolda-Leftraru del NLHPC (<https://www.nlhpc.cl/>).

Tabla A.1: Configuración del algoritmo de optimización global Shuffled Complex Evolution.

Parámetro	Descripción	Valor (-)
nopt	Número de parámetros a optimizar.	14
maxn	Máximo número de intentos antes de que finalice la optimización.	30000
kstop	Número de iteraciones de reordenamiento en que el valor debe cambiar en pcento (max=9).	9
pcento	Porcentaje de cambio.	0,05
ngs	Número de complejos en la población inicial.	10
npg	Número de puntos en cada complejo (=2*nopt+1).	29
nps	Número de puntos en cada sub-complejo (=nopt+1).	15
nspl	Número de iteraciones de evolución permitidos para cada complejo antes de ser reordenados (=2*nopt+1).	29
mings	Número mínimo de complejos necesarios (=ngs).	10
iniflg	Indicador (1/0) para incluir el punto inicial en la población.	1
iprint	Indicador (1/0) para imprimir información sobre el proceso de optimización.	0
iseed	Semilla inicial para generar los números aleatorios.	55

La Tabla A.2 incluye una descripción de los parámetros calibrados del modelo *Variable Infiltration Capacity* (VIC), además de los límites inferior (LB) y superior (UB) asociados a cada uno. Para mayor información respecto al modelo VIC, se refiere al lector a los informes del proyecto Actualización del Balance Hídrico Nacional (DGA, 2019b, 2019a, 2018, 2017), o al sitio web <https://vic.readthedocs.io/en/master/>.

Los parámetros obtenidos para cada cuenca, derivados del proceso de calibración con SCE-UA, así como el valor de la función objetivo asociada (i.e., KGE; Gupta et al., 2009) y el NSE (Nash & Sutcliffe, 1970), se presentan en la Tabla A.3.

Los resultados de calibración obtenidos para las 12 cuencas de estudio se presentan en forma gráfica desde la Figura A1 a A12. En cada figura, para los valores observados y simulados, se presenta una vista general de la ubicación espacial de la cuenca, la serie de valores diarios, mensuales y anuales, la curva de duración de caudal medio diario, la fracción de cobertura de nieve fraccional (fSCA) y el equivalente de nieve en agua (SWE) agregados a escala de cuenca a nivel diario. A su vez, se agrega una gráfica de la partición adimensional P-ET-Q calculada para cada año en el periodo 1979-2015 y el promedio (circulo de mayor tamaño) del periodo 1985-2015.

Finalmente, la Figura A.13 muestra, a través de gráficos de caja (i.e., *boxplots*), el desempeño general del modelo VIC evaluado a partir de diferentes métricas derivadas de la serie de valores diarios (e.g., NSE, Log-NSE, Log-R2) y la curva de duración (BiasFHV, BiasQ50, BiasQ95) observada y simulada. A su vez, se analizan por separado los periodos de calibración y validación, considerando el subperiodo completo y una clasificación de año seco, normal y húmedo análogo a lo presentado en 5.1.

Tabla A.2: Descripción de los parámetros calibrados y rango de búsqueda. X0 corresponde al valor inicial utilizado para correr VIC, mientras que LB (*lower bound*) y UB (*upper bound*) corresponden a los límites inferior y superior, respectivamente, asociados al parámetro de interés.

Parámetro	Descripción	X0	LB	UB
binfilt	Parámetro de la curva de infiltración variable (-).	0,529	0,005	0,9
Ds	Fracción de Dsmax donde comienza el flujo base no lineal (-).	0,791	0,05	1
Dsmax	Máxima velocidad del flujo base (mm/d).	7,286	5	200
Ws	Fracción de humedad de suelo máxima donde ocurre flujo base no lineal (-).	0,432	0,1	1
expt	Exponente n en la ecuación de Campbell para la conductividad hidráulica. Los valores deben ser mayores a 3 (-).	3,045	3	12
depth1	Espesor de la primera capa de suelo (m).	0,137	0,05	1,5
depth2	Espesor de la segunda capa de suelo (m).	0,619	0,3	4
depth3	Espesor de la tercera capa de suelo (m).	3,473	0,3	4
NEW_SNOW_ALB	Albedo de la nieve fresca (i.e., recién caída) (-).	0,85	0,8	0,92
SNOW_ALB_ACCUM_A	Parámetro de decaimiento del albedo de la nieve en el periodo de acumulación (-).	0,94	0,85	0,97
SNOW_ALB_THAW_A	Parámetro de decaimiento de la nieve en el periodo de deshielo (-).	0,82	0,75	0,92
Ksat	Conductividad hidráulica saturada de la capa de suelo (mm/d).	379,242	0	1000
MIN_RAIN_TEMP	Temperatura mínima a la cual puede ocurrir precipitación líquida (lluvia) (°C).	0	-2	2
snow_rough	Rugosidad de la superficie del manto nival (m).	0,0001	0,00001	0,02

Tabla A.3: Parámetros obtenidos del proceso de calibración con SCE-UA para cada cuenca, valor de la función objetivo de calibración KGE, y valor de NSE estimado en el periodo de calibración para el conjunto de parámetros óptimos.

Parámetro	Código BNA del control fluviométrico asociado a la cuenca calibrada											
	4311001	4313001	4703002	5410002	5710001	5721001	5722001	6027001	7103001	7330001	7354002	8104001
binfilt	0,074	0,077	0,202	0,143	0,147	0,01	0,039	0,001	0,008	0,008	0,006	0,026
Ds	0,488	0,354	0,424	0,63	0,332	0,612	0,502	0,685	0,75	0,794	0,703	0,615
Dsmax	119,236	66,728	76,847	70,342	67,246	55,394	147,56	184,507	144,652	107,473	180,349	206,964
Ws	0,612	0,608	0,642	0,865	0,733	0,884	0,707	0,453	0,216	0,229	0,386	0,151
expt	4,069	3,26	3,208	3,051	3,436	6,689	7,378	8,391	3,535	3,07	5,02	3,551
depth1	2,893	0,4	0,039	0,014	0,057	2,938	1,679	0,348	0,108	0,627	1,037	0,344
depth2	2,951	4,2	3,015	3,529	5,054	3,626	4,138	1,151	2,273	0,461	0,79	2,024
depth3	3,996	3,079	3,896	4,119	3,642	3,509	3,393	0,703	0,626	0,87	0,536	0,562
NEW_SNOW_ALB	0,874	0,923	0,915	0,93	0,916	0,758	0,882	0,871	0,954	0,458	0,751	0,941
SNOW_ALB_ACCUM_A	0,988	0,987	0,963	0,981	0,852	0,978	0,935	0,789	0,805	0,576	0,845	0,754
SNOW_ALB_THAW_A	0,85	0,86	0,911	0,855	0,871	0,869	0,906	0,909	0,917	0,779	0,831	0,92
Ksat	1265,75	1574,12	2381,15	2373,12	1877,05	2105,17	1673,94	2353,74	2717,79	2924,06	2693,16	1892,32
MIN_RAIN_TEMP	1,83	2,09	1,99	2,88	-0,04	-2,00	-2,62	1,41	-1,19	-3,45	-2,62	-2,44
snow_rough	0,02089	0,01355	0,01368	0,00018	0,00823	0,01160	0,01139	0,01821	0,00205	0,00003	0,01821	0,00003
Resultados Calibración con SCE-UA utilizando KGE como función objetivo												
KGE	0,947	0,953	0,92	0,793	0,887	0,757	0,749	0,807	0,777	0,718	0,844	0,847
NSE	0,895	0,911	0,841	0,724	0,758	0,477	0,508	0,60	0,561	0,637	0,68	0,699

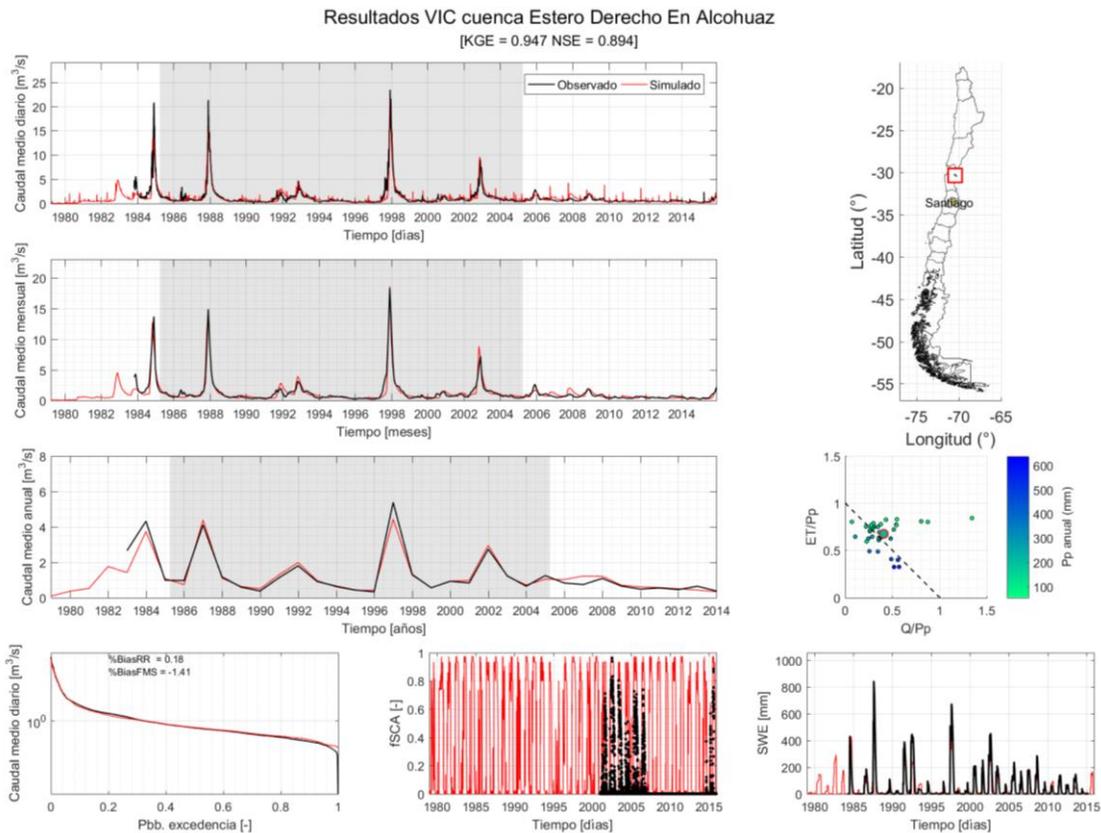


Figura A.1: Resultados de calibración para la cuenca Estero Derecho En Alcohuaz (4311001).

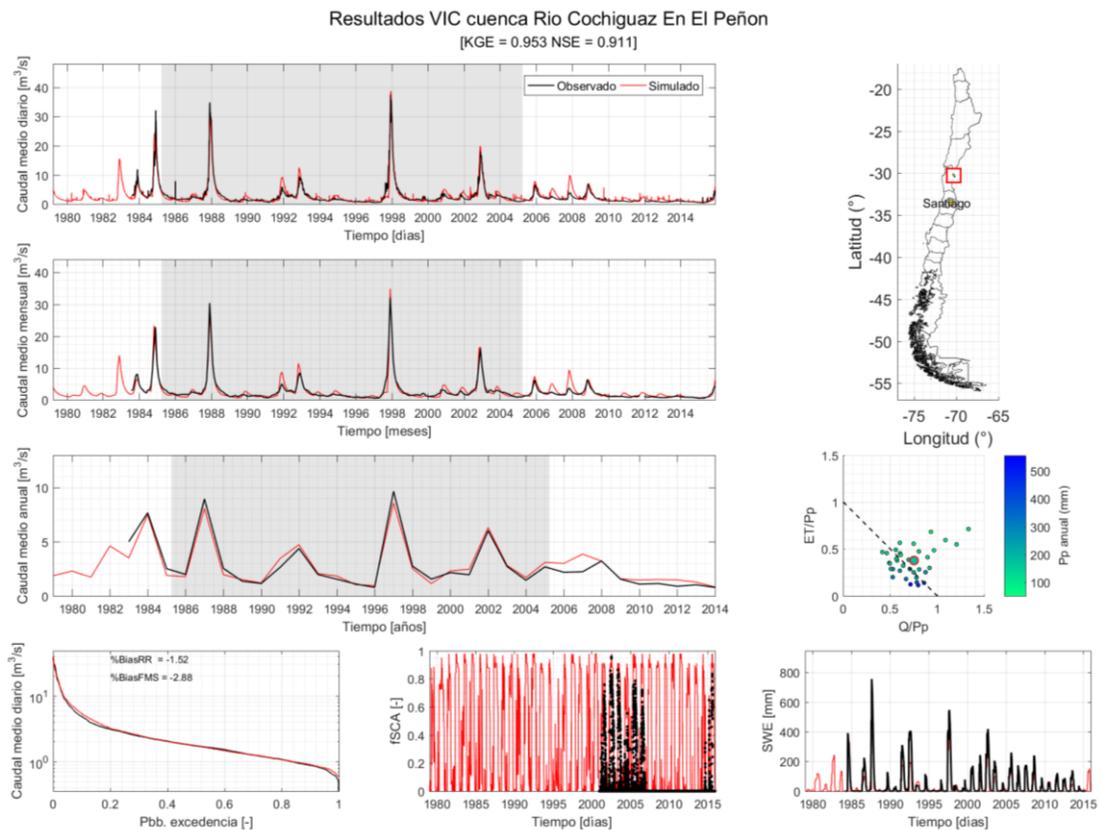


Figura A.2: Resultados de calibración para la cuenca Río Cochiguaz en el Peñón (4313001).

Resultados VIC cuenca Río Choapa En Cuncumen
[KGE = 0.92 NSE = 0.841]

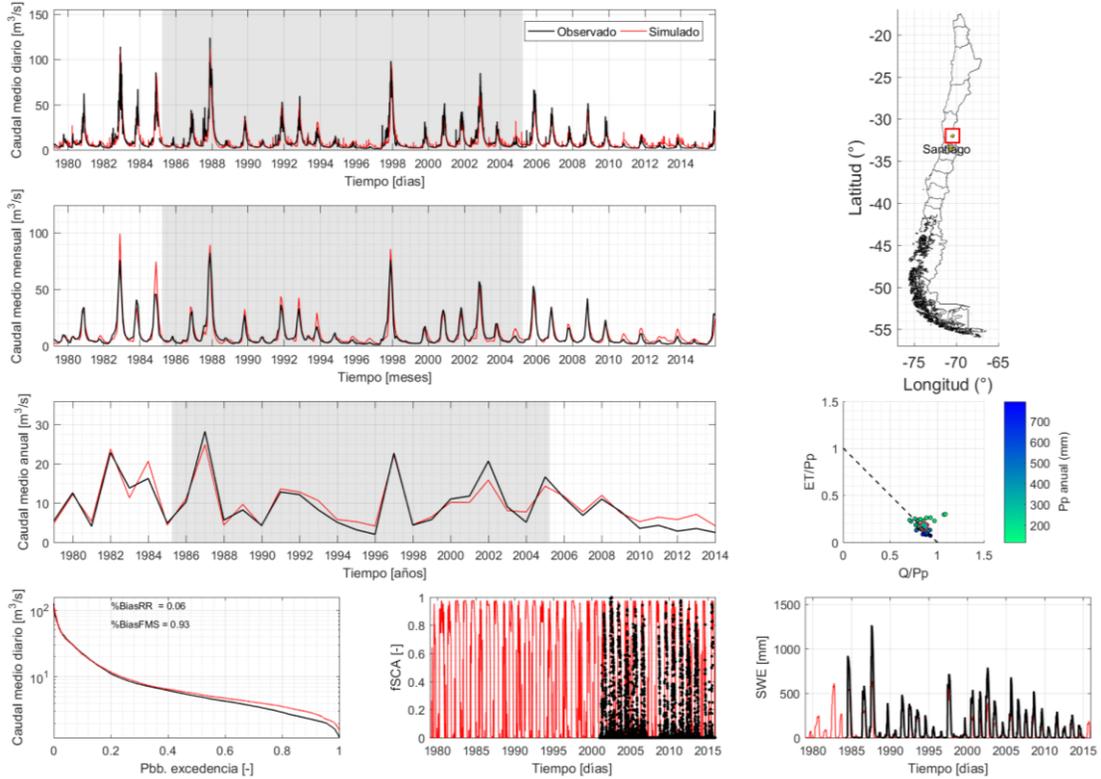


Figura A.3: Resultados de calibración para la cuenca Río Choapa en Cuncumen (4703002).

Resultados VIC cuenca Río Aconcagua En Chacabuquito
[KGE = 0.792 NSE = 0.726]

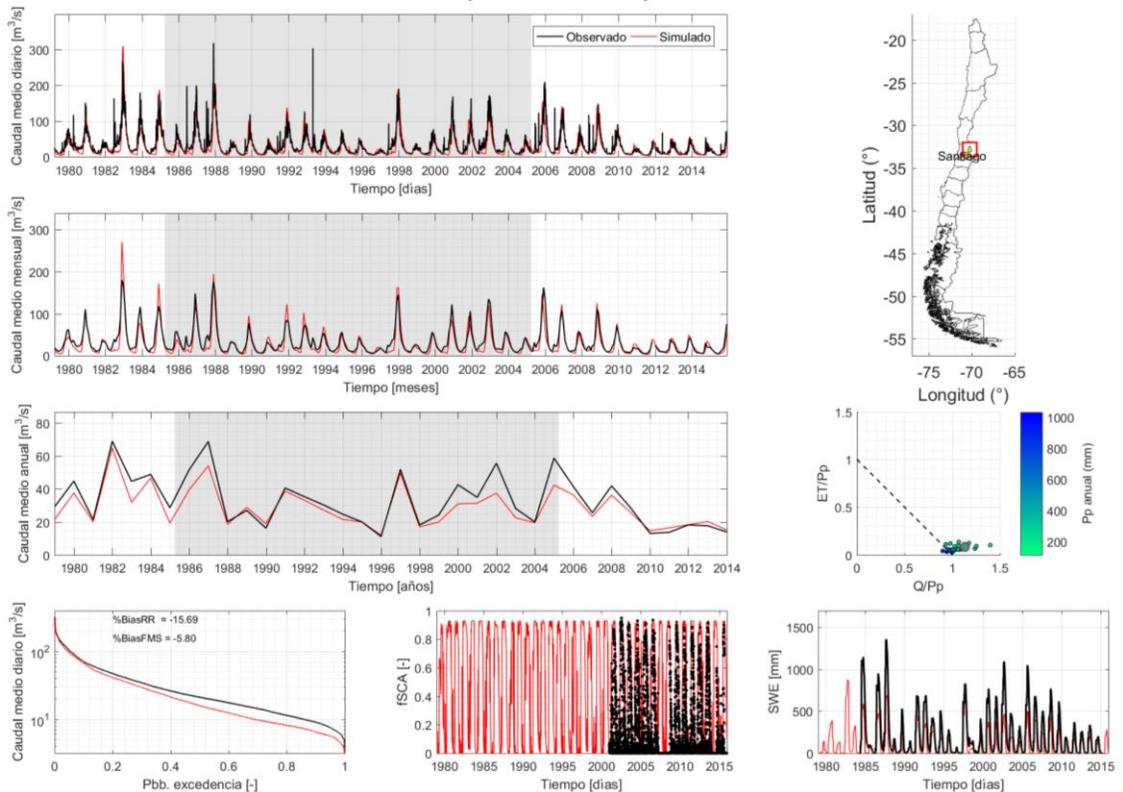


Figura A.4: Resultados de calibración para la cuenca Río Aconcagua en Chacabuquito (5410002).

Resultados VIC cuenca Río Maipo En El Manzano
[KGE = 0.877 NSE = 0.76]

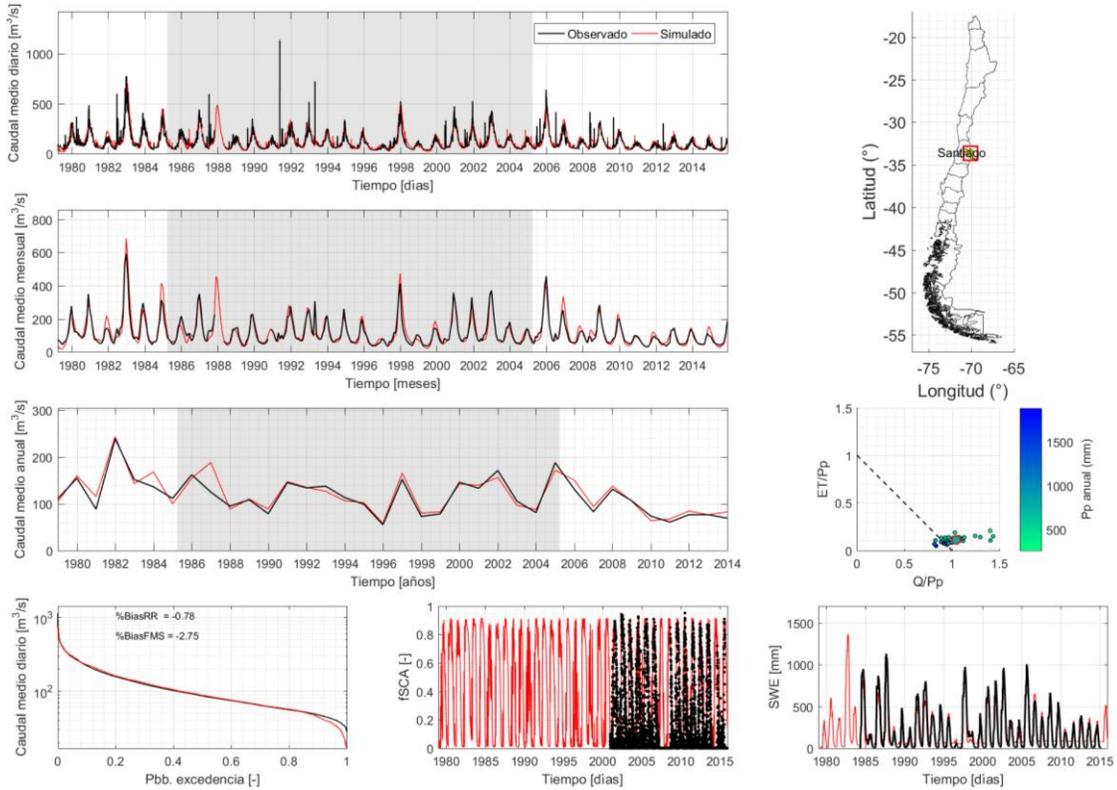


Figura A.5: Resultados de calibración para la cuenca Río Maipo en el Manzano (5710001).

Resultados VIC cuenca Estero Yerba Loca Antes Junta San Francisco
[KGE = 0.741 NSE = 0.475]

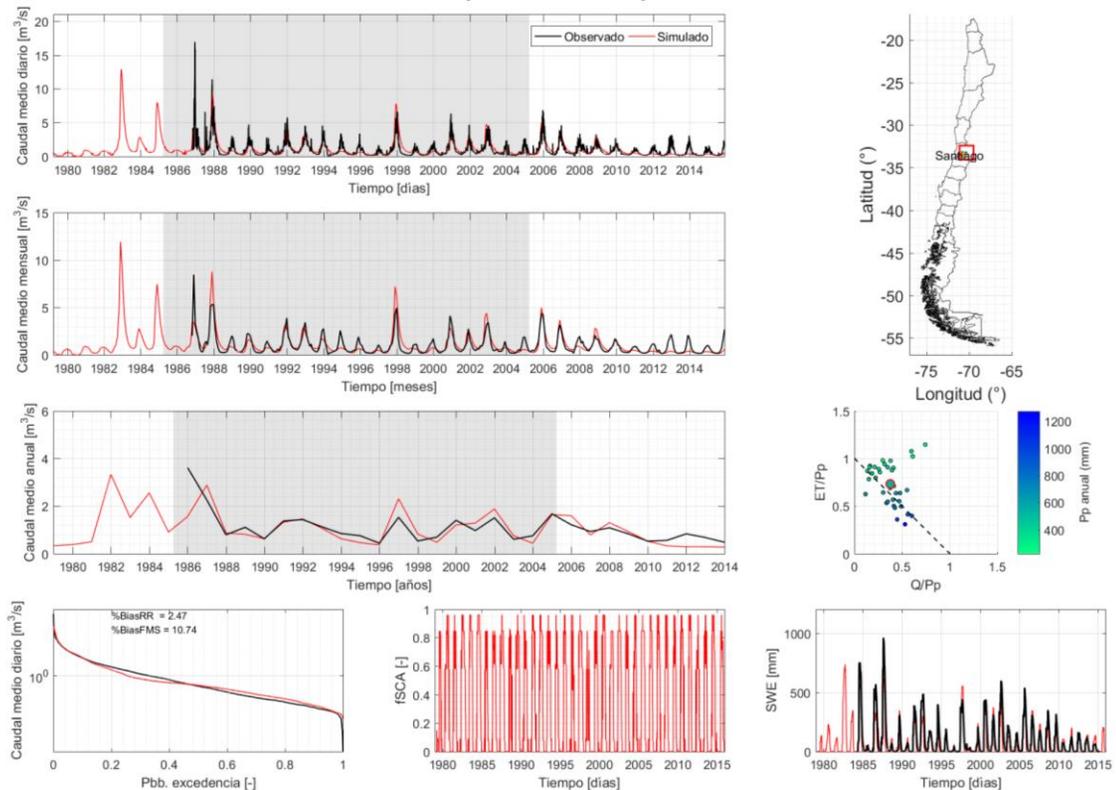


Figura A.6: Resultados de calibración para la cuenca Estero Yerba Loca (5721001).

Resultados VIC cuenca Estero Arrayan en la Montosa
[KGE = 0.747 NSE = 0.507]

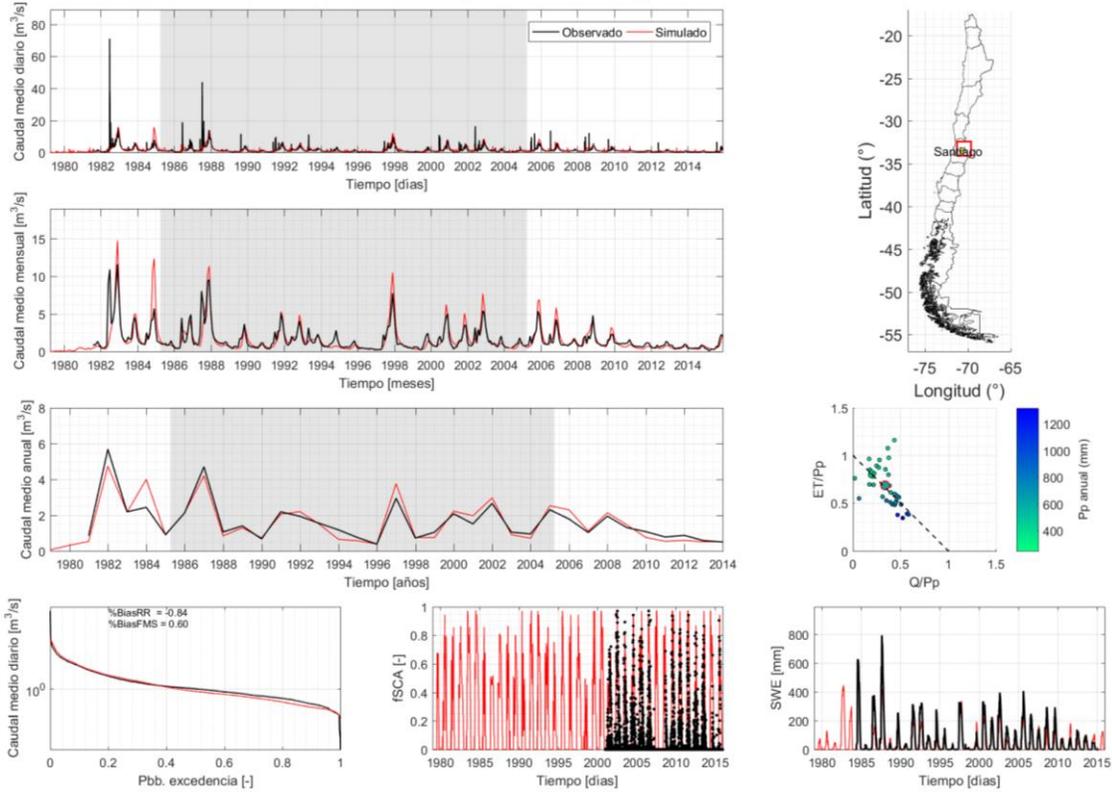


Figura A.7: Resultados de calibración para la cuenca Estero Arrayán en la Montosa (5722001).

Resultados VIC cuenca Río Claro En El Valle
[KGE = 0.801 NSE = 0.601]

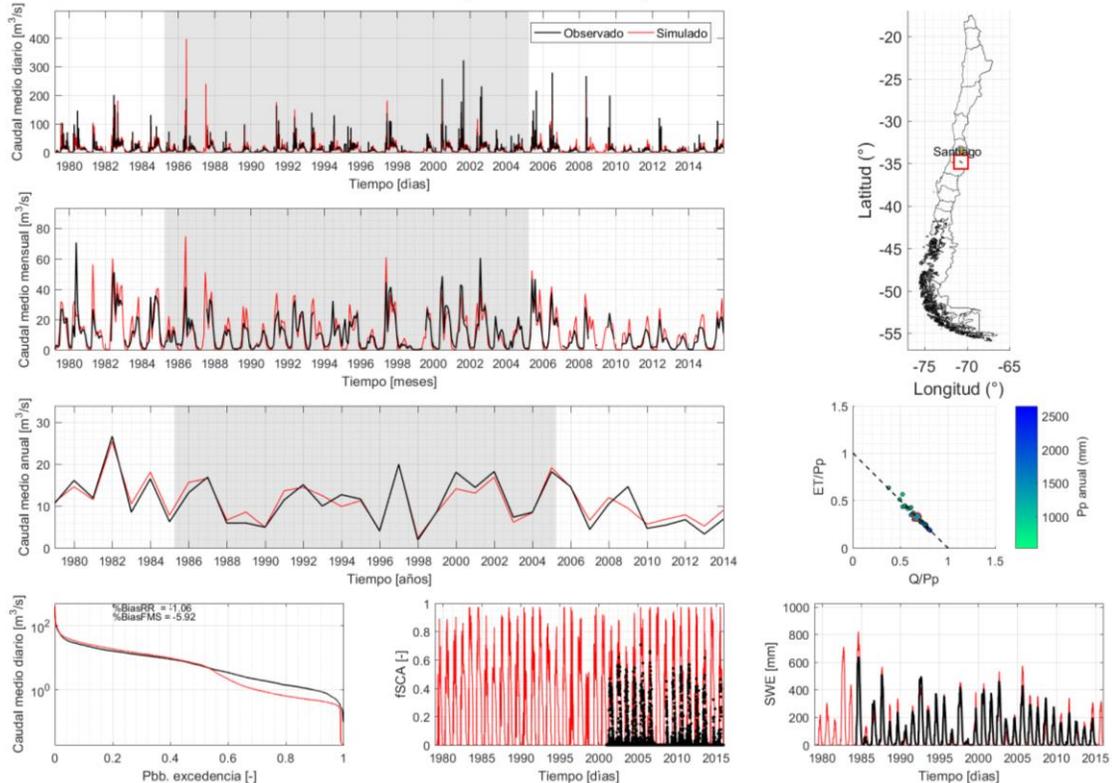


Figura A.8: Resultados de calibración para la cuenca Río Claro en el Valle (6027001).

Resultados VIC cuenca Río Claro En Los Queñes
[KGE = 0.777 NSE = 0.563]

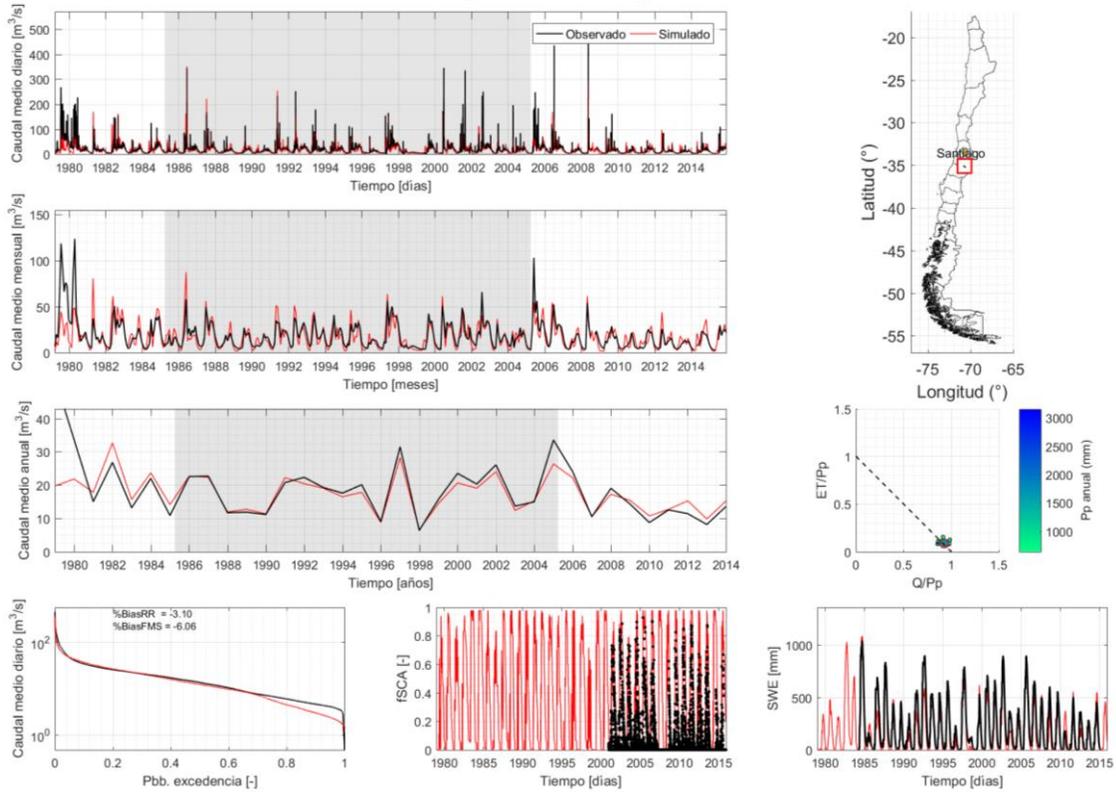


Figura A.9: Resultados de calibración para la cuenca Río Claro en los Queñes (7103001).

Resultados VIC cuenca Río Perquillauquén En San Manuel
[KGE = 0.715 NSE = 0.636]

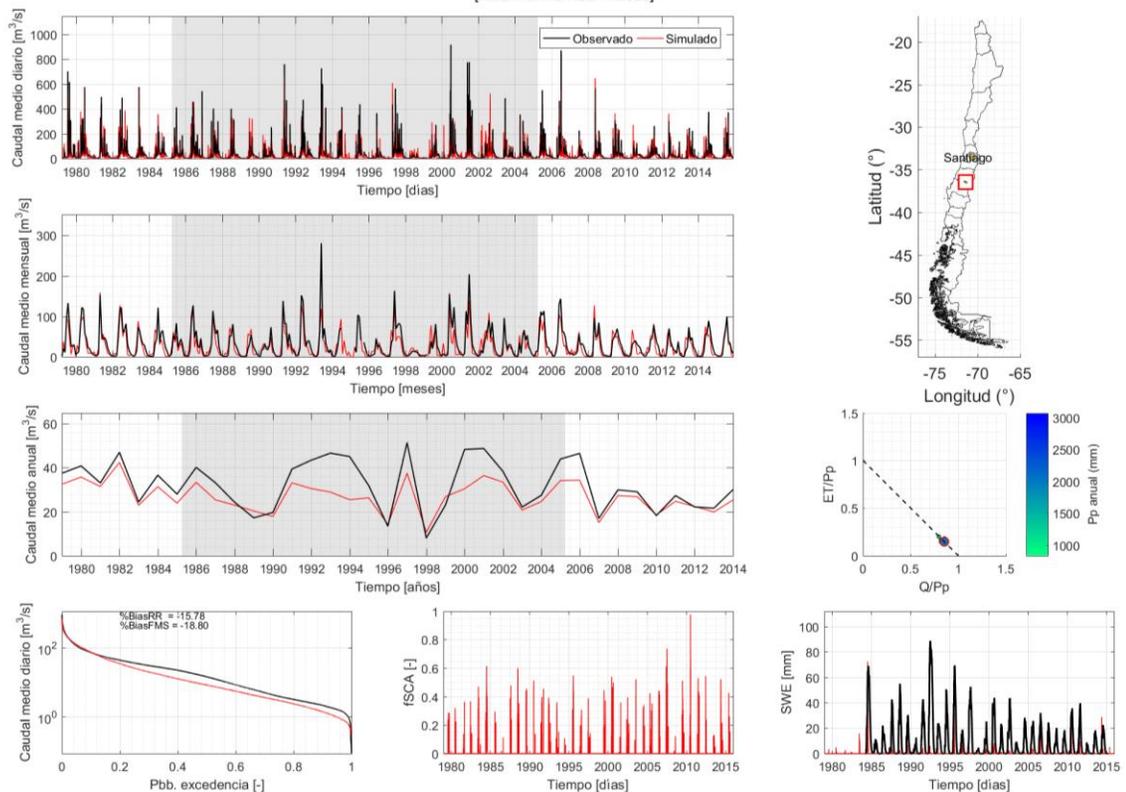


Figura A.10: Resultados de calibración para la cuenca Río Perquillauquén (7330001).

Resultados VIC cuenca Río Achibueno En La Recova
[KGE = 0.838 NSE = 0.681]

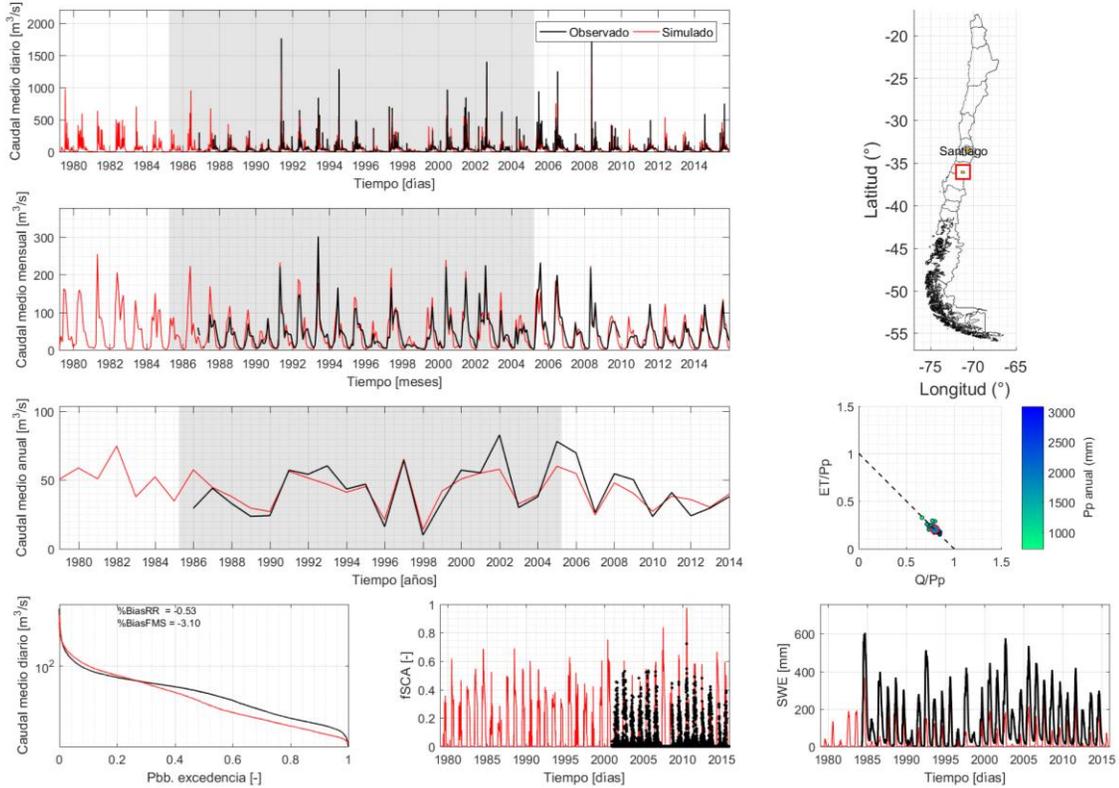


Figura A.11: Resultados de calibración para la cuenca Río Achibueno (7354002).

Resultados VIC cuenca Río Saucés Antes Junta Con Ñuble
[KGE = 0.847 NSE = 0.699]

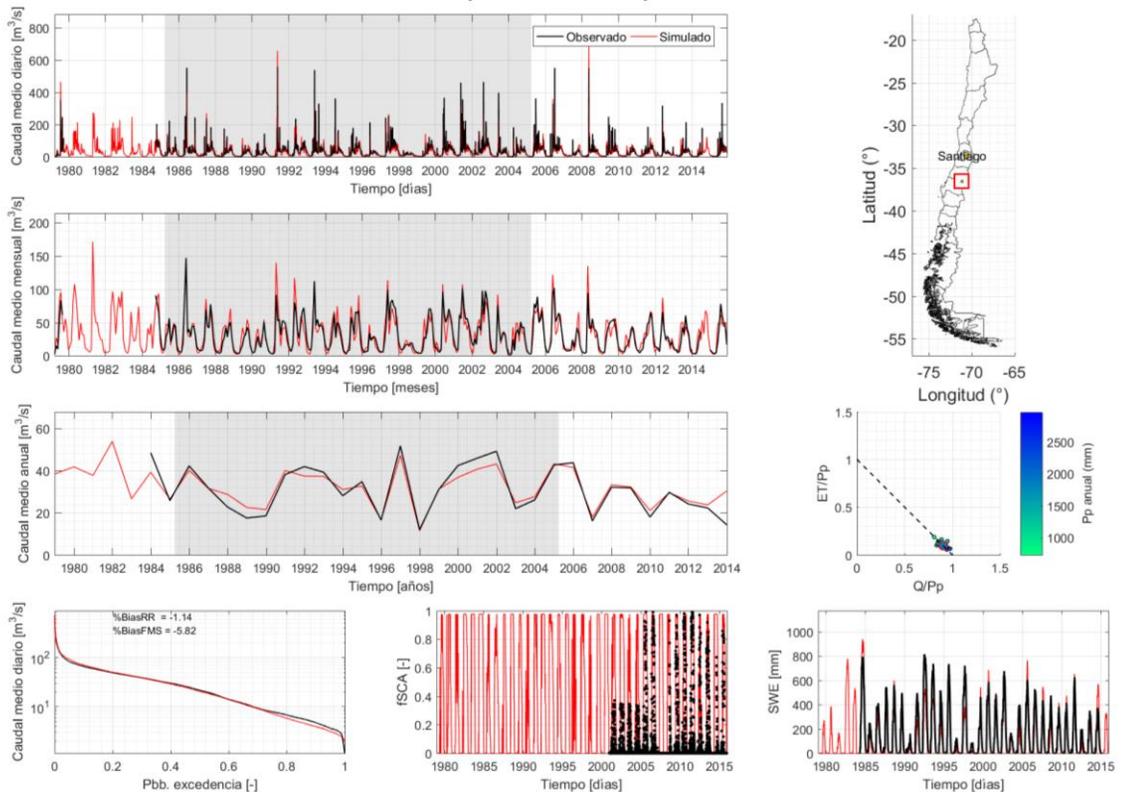


Figura A.12: Resultados de calibración para la cuenca Río Saucés (8104001).

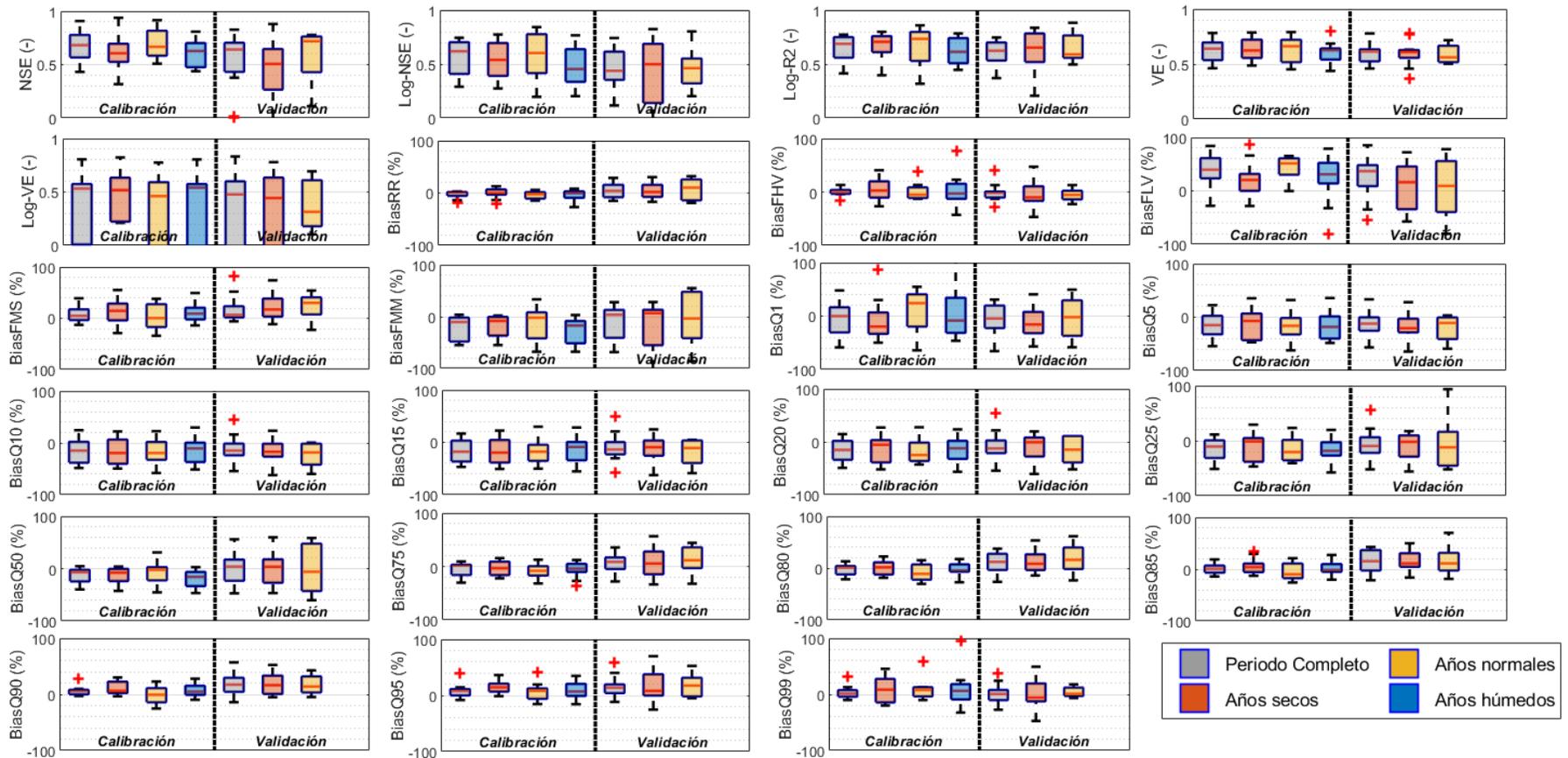


Figura A.13: Evaluación del desempeño del modelo VIC en el periodo de calibración (01/04/1985 a 31/03/2005) y validación (01/04/2005 a 31/03/2015). Todos los indicadores se presentan estratificados para el periodo completo, años secos, normales y húmedos. Las métricas NSE (Nash & Sutcliffe, 1970), Log-NSE (transformación logarítmica de NSE), Log-R2 (transformación logarítmica de coeficiente de determinación – R2), eficiencia volumétrica (VE) y su transformación logarítmica (Log-VE) se aplican sobre la serie de valores diarios (i.e., hidrograma). El resto de las métricas o índices se aplican sobre la curva de duración, donde BiasRR, BiasFHV, BiasFLV, BiasFMS y BiasFMM se definen según lo propuesto por Yilmaz et al. (2008), y los índices BiasQX, corresponde al sesgo del caudal con probabilidad de excedencia de X% simulado respecto a la observación.

ANEXO B: CONSIDERACIONES ADICIONALES

B.1 Estimación de límites físicos de la humedad de suelo

Para evitar inconsistencias en los resultados, se deben definir umbrales que delimiten lo físicamente posible. Así, al generar el *ensemble* se impone que no pueden existir valores de humedad de suelo negativos en ninguna capa y que el contenido de humedad en las capas de suelo no puede superar un valor máximo “ SM_{max} ”, el cual se puede estimar a partir del máximo espacio que puede ocupar el agua en la columna de suelo (Figura B.1). Dichos espacios vacíos o de poros se pueden caracterizar a través de la porosidad. Así, según la profundidad de cada capa en la que se discretiza la columna de suelo, y la porosidad asociada a cada una, se plantea la ecuación (B.1), según la cual se puede obtener el valor máximo de humedad de suelo para cada capa.

$$SM_{max} = \text{profundidad de suelo} \times \text{porosidad} \quad (\text{B.1})$$

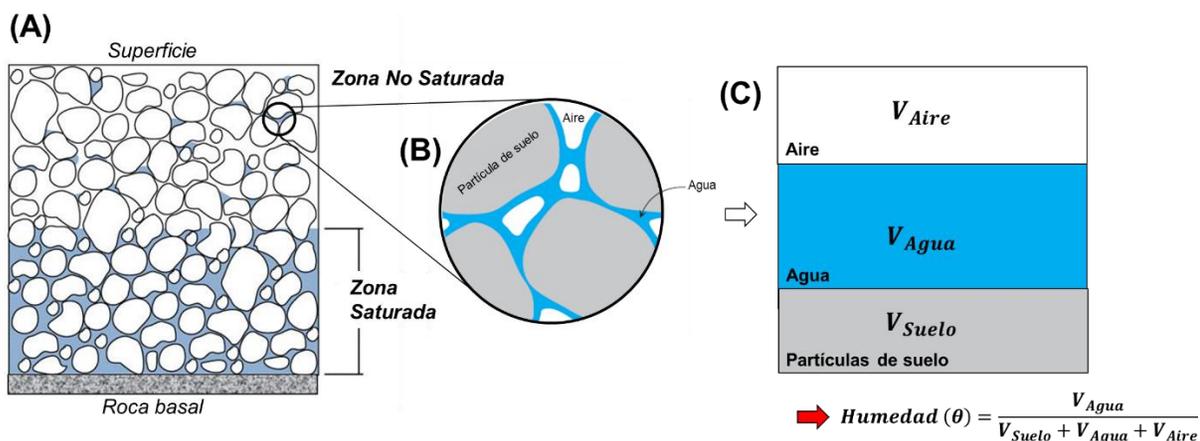


Figura B.1: (A) Esquema general de la columna de suelo, identificando las zonas no saturada (o zona de aireación o vadosa) y saturada. (B) Agua y aire en el espacio de poros del medio granular (adaptado de Fitts 2012). (C) Simplificación de la composición de la columna de suelo para la estimación de la humedad.

En términos porcentuales, la porosidad puede ser estimada a través de la densidad real del suelo y la densidad aparente, como se muestra en la ecuación (B.2).

$$\text{porosidad} = 1 - \frac{\text{densidad aparente del suelo}}{\text{densidad de partículas de suelo}} \quad (\text{B.2})$$

Tanto la densidad de las partículas de suelo (i.e., densidad real) como aparente se deben definir como variables de entrada al configurar el modelo VIC. Así, la variable definida en VIC como “*soil density*”, corresponde a la densidad de las partículas de suelo, mientras que la variable “*bulk density*”, que depende de la materia orgánica, textura (arena, limo y arcilla), densidad de los minerales presentes y nivel de compactación del suelo, corresponde a la densidad aparente del mismo. Considerando que el presente estudio parte de la base de lo realizado en el proyecto Actualización del Balance Hídrico Nacional (DGA, 2019b, 2019a, 2018, 2017), se mantienen los valores utilizados en dicho trabajo. Allí, los valores permanecen constantes para todos los píxeles y equivalentes a

1.569 kg/m³ y 2.600 kg/m³ para *bulk* y *soil density* respectivamente. La decisión metodológica adoptada en los estudios de la DGA supone, implícitamente, que la columna de suelo es homogénea, y que la diferencia en los procesos hidrológicos asociado a cada capa se puede caracterizar, o compensar, a través de la profundidad de estas, sus propiedades de conductividad hidráulica saturada y los parámetros de la curva de infiltración variable.

B.2 Perturbación de cobertura de nieve y equivalente de nieve en agua iniciales

A diferencia del caso de la perturbación de la humedad de suelo por capa, en el caso de fSCA y SWE iniciales, sólo se perturba el SWE y el ensemble es fSCA es el producto de la perturbación antes indicada. A su vez, la perturbación sólo se aplica en la temporada de acumulación (i.e., abril a septiembre) para no interferir en el proceso de derretimiento simulado por VIC (i.e., sólo se modifica el valor con el cual comienza el derretimiento) e inducir inconsistencias numéricas en el balance energético.

B.3 Estimación de tasa de recarga

Siguiendo el enfoque propuesto en el proyecto de Actualización del Balance Hídrico Nacional (DGA, 2019a, 2019b, 2018, 2017), para estimar la recarga a partir de las simulaciones generadas con VIC, se supone que la capa más profunda (i.e., capa de suelo 3), corresponde a un acuífero. De este modo, a partir del balance de masa, la variación de humedad de suelo en dicha capa (i.e., $\Delta SM3$) en un intervalo de tiempo Δt , estará dada por la ecuación (B.3), donde tanto la recarga (flujo de entrada al volumen de control) como el flujo base (salida del volumen de control) se miden en mm/día:

$$\frac{\Delta SM3}{\Delta t} = Recarga - Flujo base \quad (B.3)$$

Tomando en cuenta que tanto SM3 como el flujo base son variables modeladas con VIC, se puede despejar la incógnita de la ecuación (B.3), es decir, la recarga, según lo cual se obtiene la expresión (B.4).

$$Recarga = \frac{\Delta SM3}{\Delta t} + Flujo base \quad (B.4)$$

Asimismo, la variación de humedad de suelo en la tercera capa (i.e., SM3) puede escribirse en su forma discreta, como se muestra en (B.5). Cabe mencionar que, para $t = 1$, se supone que no existe recarga en el sistema, por lo que la ecuación (B.4) se comienza a aplicar para $t > 1$.

$$\frac{\Delta SM3}{\Delta t} = \frac{SM3(t) - SM3(t - 1)}{\Delta t} \quad (B.5)$$

A modo de ejemplo, en la Figura B.2 se grafican los resultados obtenidos para la estimación de la recarga según la ecuación (B.4), aplicada a los casos sin y con asimilación, según la configuración C1.E1 para la cuenca 5722001.

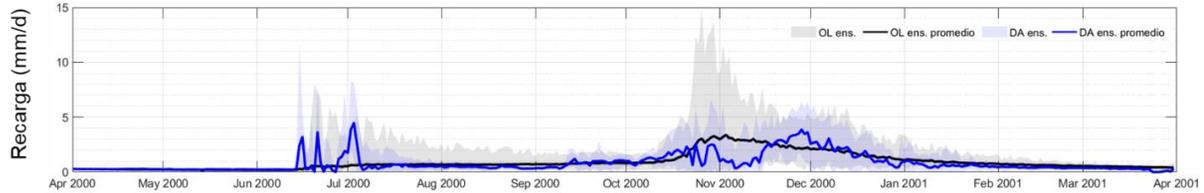


Figura B.2: Ejemplo de estimación de recarga en el periodo 2000/2001 para la cuenca Arrayán en la Montosa según los casos sin asimilación y con asimilación para la configuración C1.E1.

B.4 Formato y estructura de los archivos de estado en VIC (*state files*)

En esta tesis se perturbaron, además de las forzantes del modelo, los estados iniciales asociados a cada paso de tiempo (i.e., secuencialmente). Para llevar a cabo lo anterior, se modificó el archivo de estados que genera VIC para la fecha indicada por el usuario en la configuración del modelo (i.e., archivo global). El archivo corresponde a un documento de texto (".txt") en el cual se ordenan por filas, que dan cuenta del tipo de cobertura vegetal de la celda y banda de elevación (en caso de existir), y columnas la información asociada a los estados del modelo. En la Figura B.3 se muestra un archivo de estados de VIC para un píxel de la cuenca Estero Arrayán en la Montosa, indicando algunas componentes de interés. Cabe mencionar que, el archivo de estados contiene todos los píxeles que son ejecutados desde un mismo archivo de simulación VIC. Para mayor información respecto al archivo de estados y otras componentes de VIC, se refiere al lector al sitio web <https://vic.readthedocs.io/en/master/>.

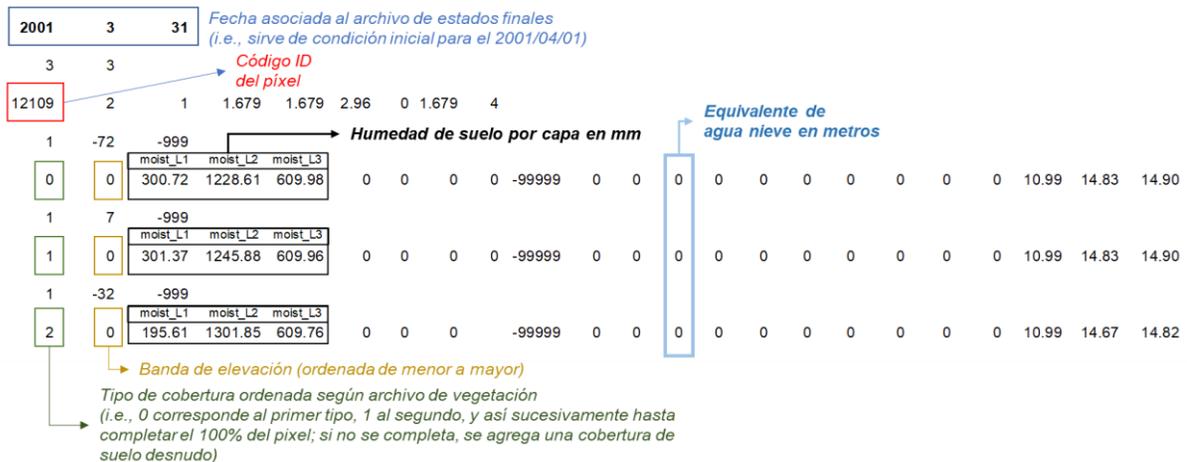


Figura B.3: Ejemplo de archivo de estados VIC para el final del día 31 de marzo de 2001, por lo que estos valores sirven como condición inicial para el día siguiente.

En resumen, para inicializar el algoritmo de asimilación de datos se corre en modo calentamiento o warm-up hasta el día anterior a la inicialización y se guarda el estado final. Por ejemplo, si se desea inicializar el 01-04-2000, se corre en modo calentamiento hasta el 31-03-2000 y se guarda el estado final del sistema. Así, se utiliza dicho estado como condición inicial de simulación, y se perturban los estados deseados (en el caso de estudio, humedad de suelo por capa y SWE). Cabe recordar que, en cada paso de tiempo, **se debe guardar el archivo de estados completo** (i.e., todo el arreglo), pues grandes variaciones entre los términos que componen los estados pueden generar problemas de convergencia y, con ello, que el modelo no termine las simulaciones.

B.5 Propuestas de mejora en la calibración de errores: MAP

Una variable desconocida “ θ ” se puede trabajar como una variable aleatoria, suponiendo cierto conocimiento inicial sobre su distribución (i.e., distribución a priori). A partir del teorema de Bayes, si se considera Y una variable aleatoria observada, la distribución a posteriori, $P(\theta|Y)$, se puede expresar a partir de la ecuación B.6. No obstante, para resolver dicha ecuación se requiere conocer $P(Y)$, lo cual se puede obtener, con un alto costo computacional, aplicando la ley de probabilidades totales.

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)} \quad (\text{B.6})$$

Si se supone que $P(\theta|Y)$ contiene toda la información sobre la variable aleatoria desconocida θ , se puede utilizar dicha distribución para encontrar un punto o intervalo estimativo de θ . Dicho método corresponde a la “Estimación máxima a posteriori” (MAP por sus siglas en inglés), y busca encontrar un θ tal que se maximice la expresión B.6. Así, aplicado al caso de estudio, B.6 se puede reescribir como B.7.

$$P(\theta|Q_{obs}, u, \theta) \propto L(\theta) = P(\theta) \cdot \prod_{t=1}^N P(Q_{obs,t}|\theta, u, \theta) \quad (\text{B.7})$$

En términos prácticos, al igual que, se propone implementar el método MAP como en Li et al.(2014) según el algoritmo de optimización presentado en la Tabla B.2. Cabe mencionar que si la variable aleatoria θ se distribuye uniformemente en un intervalo finito, la estimación MAP resulta equivalente a la estimación de máxima verosimilitud (MLE).

Tabla B.1: Esquema del algoritmo empleado para implementar computacionalmente MAP.

Paso 1	: Una muestra de errores es seleccionada desde el espacio de errores θ definido.
Para $t = 1, \dots, N$ realizar desde el Paso 2 a 4 (ciclo iterativo de simulaciones con el modelo hidrológico VIC).	
Paso 2	: Para el paso de tiempo t , se generan <i>ensembles</i> a partir de los errores definidos en el Paso 1 para cada variable de interés. <ul style="list-style-type: none"> • Forzando el modelo con cada miembro del ensemble definido en el Paso 2, se generan simulaciones con el modelo hidrológico para obtener la variable observada simulada (e.g., escorrentía) y la variable modelada (e.g., humedad de suelo).
Paso 3	: <ul style="list-style-type: none"> • A partir del error de observación del Paso 1 se perturba la variable observada simulada. • Se ajusta una distribución normal $N(\mu_{pred}, \sigma_{pred}^2)$ a la variable observada perturbada para determinar μ_{pred} y σ_{pred}^2 para el ensemble estimado.
Paso 4	: Dada la densidad de probabilidad de obtener, por ejemplo, el caudal registrado en el control fluviométrico (i.e., observación) a partir de $N(\mu_{pred}, \sigma_{pred}^2)$ se obtiene $P(Q_{obs,t} \theta, u, \theta)$.
Término del ciclo iterativo de simulaciones con el modelo hidrológico VIC.	
Paso 5	: <ul style="list-style-type: none"> • La densidad de probabilidad de los errores se calcula como el producto de la densidad de probabilidad de los errores de cada variable. • Se supone una distribución Gaussiana del error de observación con media ϵ_q y desviación estandar $10\% \cdot \epsilon_q$. • Como no se tiene información a priori del resto de los errores, se supone una distribución a priori no informativa.
Paso 6	: Los resultados derivados del ciclo iterativo (Paso 2 a 4) son ponderados con los resultados del Paso 5, según lo cual se calcula la probabilidad a posteriori de los errores. Se retorna al paso 1 hasta la convergencia del algoritmo de optimización.

ANEXO C: COMPARACIÓN DE CONFIGURACIONES ADOPTADAS PARA LA ASIMILACIÓN DE DATOS

En esta sección, se incluyen figuras y tablas que permiten apoyar los hallazgos reportados en el cuerpo principal de la tesis y que, por temas prácticos, han sido ubicadas en el presente anexo divididas por sección. A continuación, se realiza una breve descripción de las figuras presentadas en cada sección del presente anexo.

- **Sección C.1:** Se presenta la razón entre la raíz del error cuadrático medio (RMSE) promedio de los ensembles generados con y sin asimilación (i.e. $RMSE_{DA}/RMSE_{OL}$). A diferencia de la Figura 5.9, las figuras presentadas muestran el valor agregado de asimilar datos respecto al caso OL según las simulaciones en los subperiodos abril-septiembre y octubre-marzo para el año hidrológico 2000/01. A su vez, se presentan las diferencias en el RMSE de los casos con y sin asimilación a través de tablas.
- **Sección C.2:** Se presentan los impactos de la asimilación de datos, respecto al caso de no asimilación, en las estimaciones de flujos y estados hidrológicos de los cuales no se tienen observaciones. Las figuras presentadas ilustran los resultados obtenidos para el periodo 1) abril-septiembre y 2) octubre-marzo del año hidrológico 2000/01.

Comentarios generales:

- La Figura C.2 muestran que los problemas del caso de asimilación C2.E2 radican en el mal desempeño del modelo durante el periodo octubre-marzo. Asimismo, entre las cuencas 4311001 a 5722001, se observa un NRMSE en general mayor al resto de las cuencas del dominio.
- De las Figuras C.1 y C.2, y las Tablas C.2 y C.3, se desprende que, en términos estacionales, el efecto de la asimilación sobre la escorrentía es mayor en el periodo octubre-marzo en cuencas de régimen nival ubicadas en la zona B y C del dominio de estudio (i.e., 4311001 a 5722001), y abril-septiembre en el resto de las cuencas. Dichas diferencias se pueden atribuir a la importancia relativa que tiene la reducción de la dispersión de las simulaciones de SWE respecto a las observaciones de referencia en la respuesta hidrológica de cada una de las cuencas.
- De la Figura C.3 y C.4 se desprende que, en general, las mayores diferencias porcentuales entre el caso OL y DA se generan en el periodo abril a septiembre, asociadas principalmente a flujo base y recarga, lo cual se verifica en todas las cuencas. Sólo la cuenca 7330001 presenta diferencias importantes (hasta 400% respecto a OL) en el periodo octubre-marzo, las cuales corresponden principalmente a diferencias en el flujo base y recarga.

C.1 Valor agregado de la asimilación de datos en la estimación de escorrentía media diaria, fSCA y SWE.

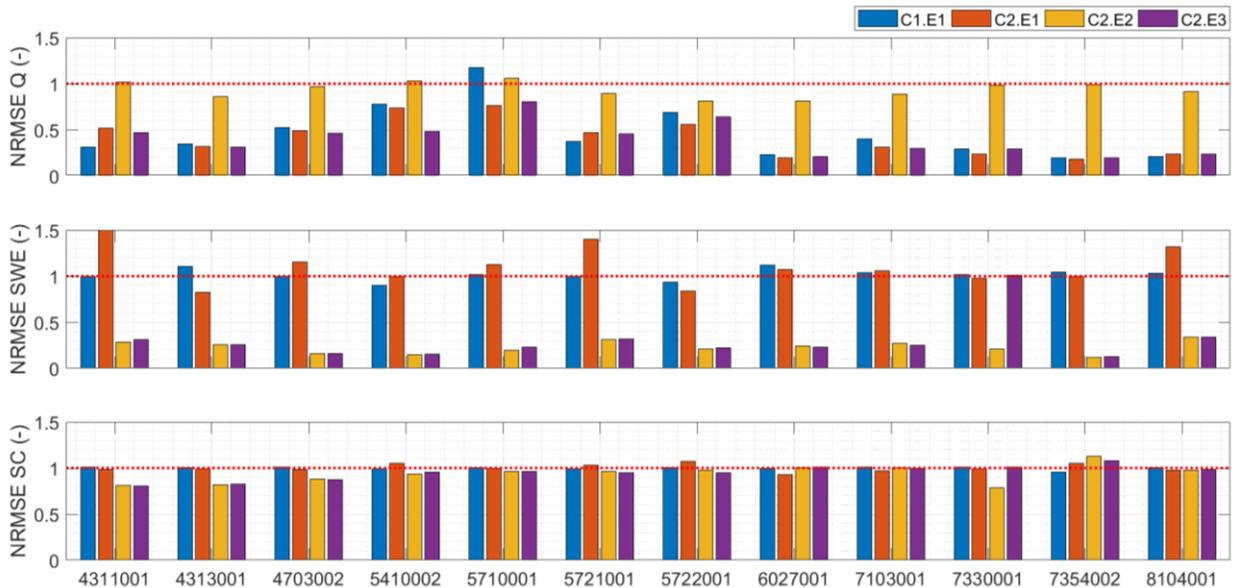


Figura C.1: Evaluación del valor agregado de asimilar datos respecto a generar un conjunto de simulaciones sin asimilación aplicada a la estimación de escorrentía (Q), equivalente de nieve en agua (SWE) y cobertura de nieve (SC) en el periodo abril-septiembre del año hidrológico 2000/01. La línea roja segmentada representa el valor en que los resultados sin asimilación y con asimilación son equivalentes. El valor máximo en el eje Y se ha truncado en 1.5 con fines comparativos.

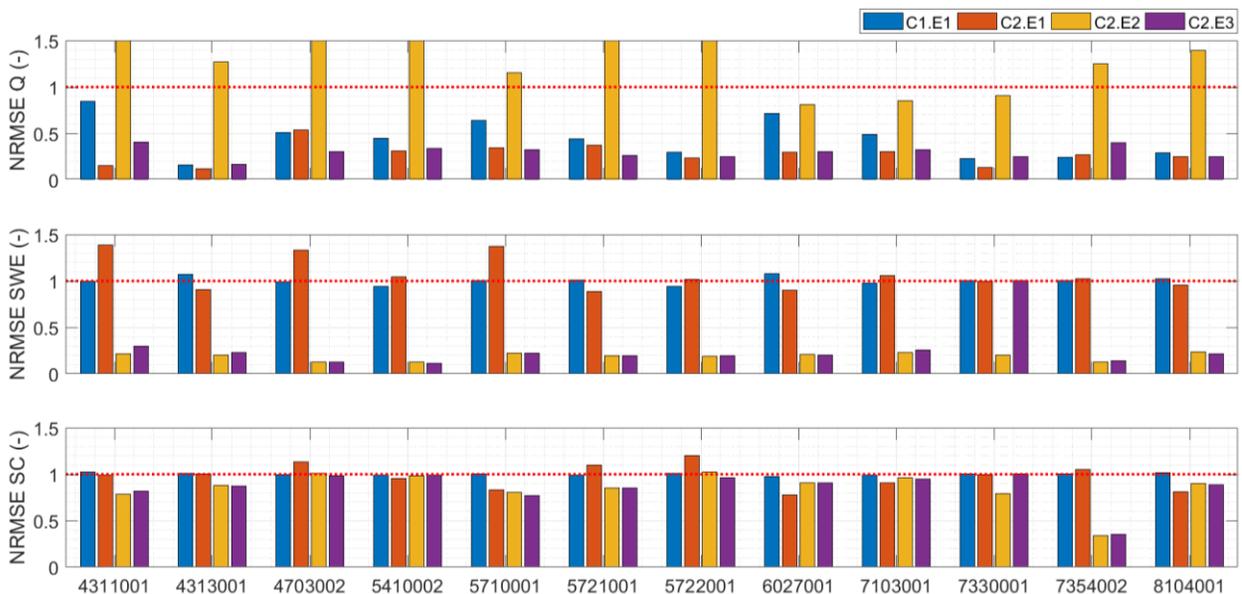


Figura C.2: Evaluación del valor agregado de asimilar datos respecto a generar un conjunto de simulaciones sin asimilación aplicada a la estimación de escorrentía (Q), equivalente de nieve en agua (SWE) y cobertura de nieve (SC) en el periodo octubre-marzo del año hidrológico 2000/01. La línea roja segmentada representa el valor en que los resultados sin asimilación y con asimilación son equivalentes. El valor máximo en el eje Y se ha truncado en 1.5 con fines comparativos.

Tabla C.1: Diferencia entre la raíz del error cuadrático medio (RMSE) promedio de los ensemble simulados para el año hidrológico 2000/01 con asimilación de datos (DA) según las diferentes configuraciones propuestas y el caso sin asimilación (OL) asociado. El valor en cada celda se obtiene de la operación $RMSE_{DA}-RMSE_{OL}$ dividido por un término de referencia, el cual corresponde a la escorrentía media anual para Q, y SWE y fSCA máximo durante la temporada para SWE y fSCA respectivamente.

BNA	C1.E1			C2.E1			C2.E2			C2.E3		
	ΔQ (mm/d)	ΔSWE (mm)	ΔSC (%)	ΔQ (mm/d)	ΔSWE (mm)	ΔSC (%)	ΔQ (mm/d)	ΔSWE (mm)	ΔSC (%)	ΔQ (mm/d)	ΔSWE (mm)	ΔSC (%)
4311001	-11%	0%	0,6%	-36%	19%	-0,8%	48%	-15%	-11,4%	-28%	-14%	-11,4%
4313001	-34%	1%	0,2%	-51%	-3%	-0,5%	13%	-14%	-9,9%	-48%	-14%	-9,7%
4703002	-7%	0%	0,3%	-8%	6%	0,6%	12%	-22%	-3,5%	-13%	-22%	-4,0%
5410002	-7%	-2%	-0,4%	-9%	0%	1,0%	12%	-26%	-1,9%	-10%	-26%	-1,2%
5710001	-1%	0%	0,1%	-8%	5%	-1,0%	2%	-16%	-1,8%	-8%	-15%	-2,0%
5721001	-9%	0%	-0,3%	-11%	2%	1,3%	15%	-17%	-1,7%	-13%	-17%	-2,1%
5722001	-8%	-1%	0,1%	-12%	-2%	2,6%	124%	-17%	-0,5%	-11%	-17%	-1,6%
6027001	-9%	1%	-0,4%	-11%	0%	-4,8%	-3%	-15%	-0,3%	-11%	-15%	0,0%
7103001	-5%	0%	0,3%	-6%	1%	-1,7%	-1%	-12%	-0,1%	-7%	-12%	-0,6%
7330001	-8%	1%	0,4%	-8%	-1%	-0,3%	0%	-35%	-10,3%	-8%	-35%	0,4%
7354002	-13%	1%	-1,3%	-13%	0%	1,6%	0%	-32%	-1,3%	-13%	-32%	-2,5%
8104001	-12%	0%	0,5%	-11%	2%	-6,3%	0%	-9%	-4,5%	-11%	-9%	-4,1%
PROM.	-10%	0%	0%	-15%	3%	-1%	19%	-19%	-4%	-15%	-19%	-3%
DESV.	8%	1%	1%	14%	6%	3%	36%	8%	4%	12%	8%	4%

Tabla C.2: Diferencia entre la raíz del error cuadrático medio (RMSE) promedio de los ensemble simulados para periodo abril-septiembre del año hidrológico 2000/01 con asimilación de datos (DA) según las diferentes configuraciones propuestas y el caso sin asimilación (OL) asociado. El valor en cada celda se obtiene de la operación $RMSE_{DA}-RMSE_{OL}$ dividido por un término de referencia, el cual corresponde a la escorrentía media anual para Q, y SWE y fSCA máximo durante la temporada para SWE y fSCA respectivamente.

BNA	C1.E1			C2.E1			C2.E2			C2.E3		
	ΔQ (mm/d)	ΔSWE (mm)	ΔSC (%)	ΔQ (mm/d)	ΔSWE (mm)	ΔSC (%)	ΔQ (mm/d)	ΔSWE (mm)	ΔSC (%)	ΔQ (mm/d)	ΔSWE (mm)	ΔSC (%)
4311001	-16%	0%	0,6%	-11%	28%	-1,0%	0%	-18%	-13,0%	-12%	-17%	-13,3%
4313001	-12%	1%	0,2%	-14%	-4%	-0,7%	-3%	-18%	-11,9%	-14%	-18%	-11,5%
4703002	-2%	0%	0,5%	-2%	4%	-0,7%	0%	-24%	-4,9%	-2%	-24%	-5,3%
5410002	-2%	-3%	-0,5%	-2%	0%	2,2%	0%	-26%	-2,8%	-4%	-26%	-1,7%
5710001	1%	0%	0,1%	-2%	2%	-0,1%	0%	-16%	-1,1%	-2%	-15%	-1,1%
5721001	-3%	0%	-0,4%	-2%	7%	1,1%	0%	-12%	-1,4%	-2%	-12%	-2,0%
5722001	-4%	-1%	0,1%	-7%	-4%	2,6%	-3%	-20%	-0,8%	-5%	-20%	-1,7%
6027001	-14%	1%	-0,4%	-16%	2%	-4,9%	-4%	-17%	0,4%	-15%	-17%	0,9%
7103001	-8%	1%	0,6%	-9%	1%	-1,7%	-1%	-12%	0,3%	-9%	-13%	-0,3%
7330001	-10%	1%	0,6%	-11%	-1%	-0,4%	0%	-48%	-12,2%	-10%	-48%	0,6%
7354002	-18%	1%	-2,1%	-18%	0%	2,2%	0%	-32%	5,4%	-18%	-31%	3,6%
8104001	-17%	0%	0,2%	-14%	4%	-3,1%	-2%	-9%	-3,1%	-14%	-9%	-1,9%
PROM.	-9%	0%	0%	-9%	3%	0%	-1%	-21%	-4%	-9%	-21%	-3%
DESV.	7%	1%	1%	6%	9%	2%	1%	11%	6%	6%	11%	5%

Tabla C.3: Diferencia entre la raíz del error cuadrático medio (RMSE) promedio de los ensemble simulados para periodo octubre-marzo del año hidrológico 2000/01 con asimilación de datos (DA) según las diferentes configuraciones propuestas y el caso sin asimilación (OL) asociado. El valor en cada celda se obtiene de la operación $RMSE_{DA}-RMSE_{OL}$ dividido por un término de referencia, el cual corresponde a la escorrentía media anual para Q, y SWE y fSCA máximo durante la temporada para SWE y fSCA respectivamente.

BNA	C1.E1			C2.E1			C2.E2			C2.E3		
	ΔQ (mm/d)	ΔSWE (mm)	ΔSC (%)	ΔQ (mm/d)	ΔSWE (mm)	ΔSC (%)	ΔQ (mm/d)	ΔSWE (mm)	ΔSC (%)	ΔQ (mm/d)	ΔSWE (mm)	ΔSC (%)
4311001	-7%	0%	0,7%	-50%	5%	-0,4%	72%	-11%	-7,1%	-35%	-10%	-5,9%
4313001	-46%	1%	0,3%	-69%	-1%	0,2%	21%	-9%	-4,2%	-65%	-9%	-4,4%
4703002	-9%	0%	0,0%	-12%	7%	3,4%	17%	-20%	0,2%	-17%	-20%	-0,5%
5410002	-11%	-2%	-0,3%	-15%	1%	-0,9%	18%	-26%	-0,4%	-14%	-26%	-0,2%
5710001	-5%	0%	0,1%	-12%	7%	-2,9%	3%	-15%	-3,4%	-13%	-15%	-4,0%
5721001	-12%	0%	-0,1%	-15%	-3%	1,9%	22%	-20%	-2,8%	-18%	-20%	-2,8%
5722001	-12%	-1%	0,2%	-15%	0%	3,5%	181%	-12%	0,4%	-15%	-12%	-0,9%
6027001	-1%	1%	-0,7%	-3%	-1%	-6,1%	-1%	-12%	-2,5%	-3%	-12%	-2,6%
7103001	-2%	0%	-0,3%	-2%	1%	-2,2%	-1%	-11%	-0,9%	-2%	-11%	-1,1%
7330001	-4%	0%	0,0%	-4%	0%	-0,1%	0%	-20%	-6,1%	-3%	-20%	0,0%
7354002	-3%	0%	0,0%	-3%	1%	1,3%	1%	-33%	-17,0%	-2%	-32%	-16,8%
8104001	-4%	0%	1,2%	-5%	-1%	-13,8%	3%	-9%	-7,4%	-5%	-9%	-8,5%
PROM.	-10%	0%	0%	-17%	1%	-1%	28%	-16%	-4%	-16%	-16%	-4%
DESV.	12%	1%	0%	21%	3%	5%	52%	7%	5%	18%	7%	5%

C.2 Impactos estacionales de DA respecto a OL en flujos y estados VIC

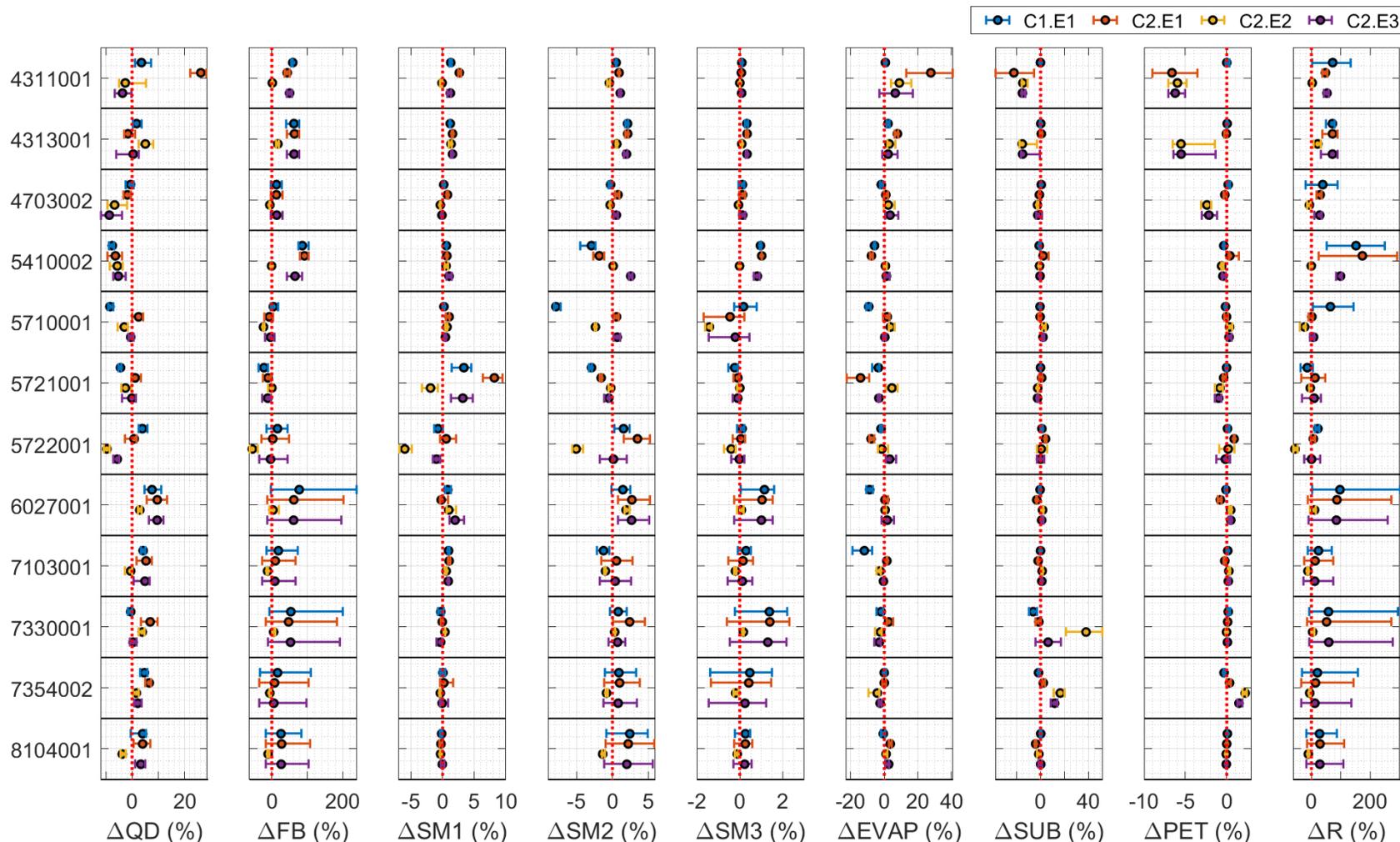


Figura C.3: Diferencias porcentuales en flujos y estados derivados de las simulaciones con VIC según las configuraciones de asimilación de datos y los casos sin asimilación asociado a cada una. La línea roja segmentada representa el caso en que no hay diferencias entre las simulaciones con y sin asimilación de datos. La extensión de las barras representa el percentil del 25% y 75% asociado a las diferencias estimadas para cada ensemble evaluado a nivel diario en el periodo abril-septiembre del año hidrológico 2000/01. El punto al centro de cada una de las barras representa la mediana del ensemble.

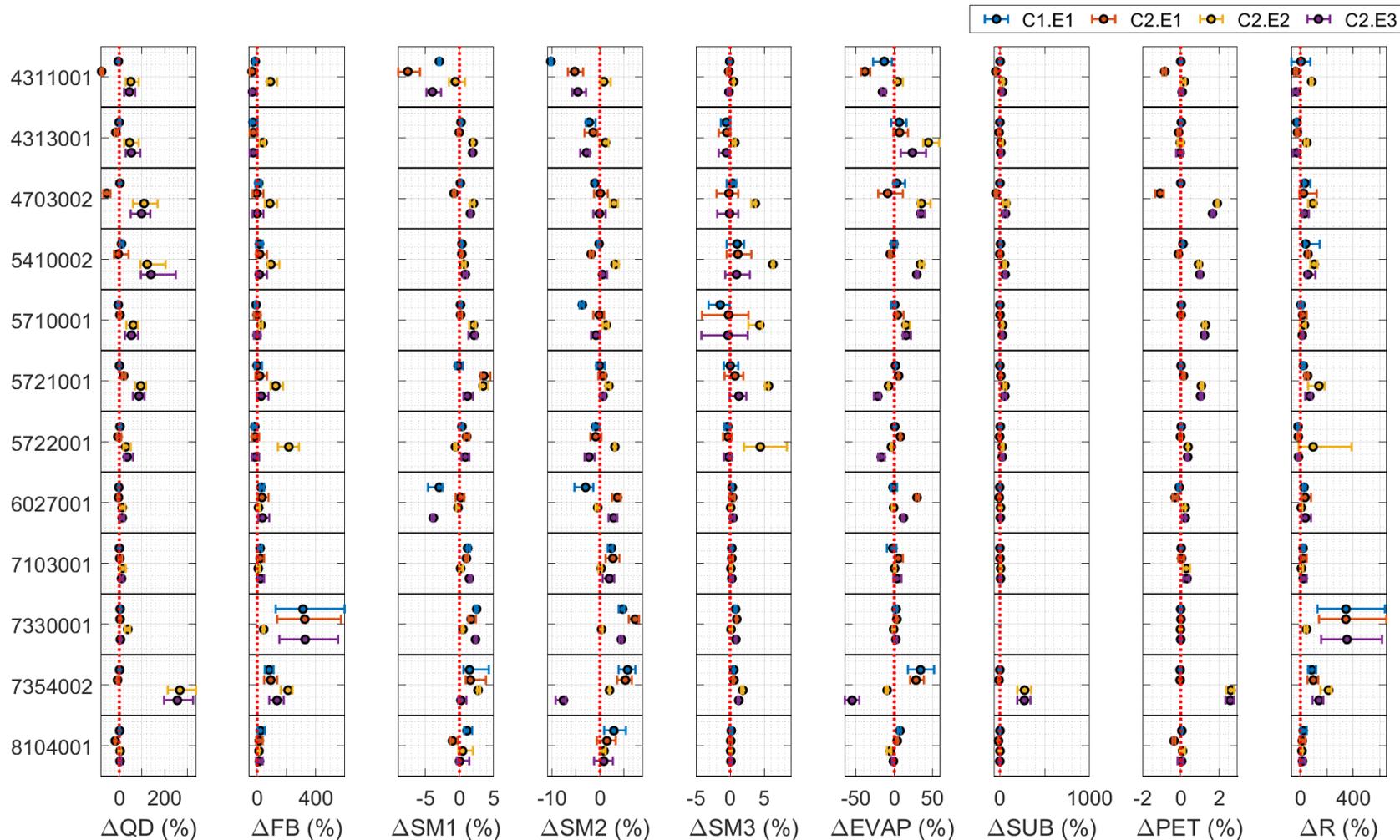


Figura C.4: Diferencias porcentuales en flujos y estados derivados de las simulaciones con VIC según las configuraciones de asimilación de datos y los casos sin asimilación asociado a cada una. La línea roja segmentada representa el caso en que no hay diferencias entre las simulaciones con y sin asimilación de datos. La extensión de las barras representa el percentil del 25% y 75% asociado a las diferencias estimadas para cada ensemble evaluado a nivel diario en el periodo octubre-marzo del año hidrológico 2000/01. El punto al centro de cada una de las barras representa la mediana del ensemble.