

PISAC



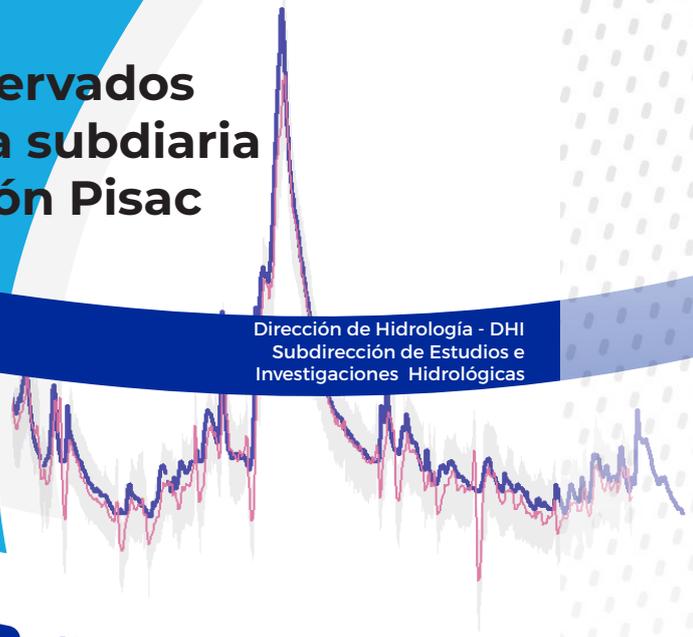
PERÚ

Ministerio  
del Ambiente



# Asimilación de caudales observados en la modelación hidrológica subdiaria del río Vilcanota en la estación Pisac

Dirección de Hidrología - DHI  
Subdirección de Estudios e  
Investigaciones Hidrológicas



Diciembre  
2022

**Citar:**

Llauca, H. (2023). Asimilación de caudales observados en la modelación hidrológica subdiaria del río Vilcanota en la estación Pisac. Servicio Nacional de Meteorología e Hidrología del Perú-SENAMHI, (2022). Estudio Final. <https://hdl.handle.net/20.500.12542/2613>

# Asimilación de caudales observados en la modelación hidrológica subdiaria del río Vilcanota en la estación Pisac





## DIRECTORIO

PhD. Guillermo Antonio Baigorria Paz  
PRESIDENTE EJECUTIVO DEL SENAMHI

Ing. Oscar Gustavo Felipe Obando  
DIRECTOR DE HIDROLOGÍA DEL SENAMHI

PhD. Waldo Sven Lavado Casimiro  
SUBDIRECTOR DE ESTUDIOS E INVESTIGACIONES HIDROLÓGICAS DEL SENAMHI

REVISOR DEL ESTUDIO  
PhD. Waldo Sven Lavado Casimiro  
Ing. Oscar Gustavo Felipe Obando

AUTOR DEL ESTUDIO  
**MSc. Ing. Harold Llauca Soto**  
Especialista en Hidrología  
Subdirección de Estudios e Investigaciones Hidrológicas (SEH)

Diciembre  
2022

## TABLA DE CONTENIDO

	<i>Pág.</i>
<b>1. INTRODUCCIÓN.....</b>	<b>8</b>
<b>2. ENFOQUE SECUENCIAL DE ASIMILACIÓN DE DATOS.....</b>	<b>9</b>
2.1. Asimilación de datos .....	9
2.2. Filtro de Kalman de Conjuntos (EnKF) .....	10
2.3. Filtro de Partículas (PF) .....	10
<b>3. MATERIALES Y MÉTODOS .....</b>	<b>11</b>
3.1. Cuenca y datos .....	11
3.2. Modelación hidrológica .....	13
3.2.1. Modelo precipitación-escorrentía GR4H .....	13
3.2.2. Calibración del modelo .....	14
3.3. Evaluación de la incertidumbre .....	15
3.3.1. Incertidumbre en las forzantes del modelo .....	15
3.3.2. Incertidumbre en las variables de estado del modelo .....	15
3.3.3. Incertidumbre en las observaciones de caudal .....	15
3.4. Diseño de experimentos de asimilación.....	17
3.5. Evaluación de pronósticos subdiarios.....	18
<b>4. RESULTADOS Y DISCUSIÓN .....</b>	<b>19</b>
4.1. Calibración del modelo y estimación de la incertidumbre .....	19
4.2. Beneficios de la asimilación de caudales observados .....	22
<b>5. CONCLUSIONES Y RECOMENDACIONES.....</b>	<b>25</b>
<b>6. REFERENCIAS BIBLIOGRÁFICAS.....</b>	<b>26</b>
<b>ANEXOS .....</b>	<b>28</b>

## ÍNDICE DE FIGURAS

	<i>Pág.</i>
<b>Figura N° 1.</b> Mapa de ubicación de la cuenca del río Vilcanota en la estación Pisac. ....	11
<b>Figura N°2.</b> Variación mensual (1981-2016) de la precipitación media, caudal medio, y temperatura media del aire en la cuenca del río Vilcanota en Pisac. ....	12
<b>Figura N° 3.</b> Estructura del modelo GR4H (Li et al., 2015). ....	14
<b>Figura N° 4.</b> Calibración y validación del modelo GR4H en la cuenca Vilcanota en Pisac, empleando datos de precipitación horaria de IMERG-E' y GSMaP-NRT'.	13
<b>Figura N° 5.</b> Simulación de caudales horarios en la estación Pisac para el periodo lluvioso 2022 (febrero-marzo); empleando perturbaciones de precipitación horaria de IMERG-E' y GSMaP-NRT' y el Filtro de Kalman de Conjuntos (EnKF) y Filtro de Partículas (FP). ....	19
<b>Figura N° 6.</b> Simulación del contenido de humedad del suelo en la cuenca Vilcanota en Pisac para el periodo lluvioso 2022 (febrero-marzo); empleando perturbaciones de precipitación horaria de IMERG-E' y GSMaP-NRT' y el Filtro de Kalman de Conjuntos (EnKF) y Filtro de Partículas (FP). ....	21
<b>Figura N° 7.</b> Evaluación del desempeño (NSE, BIAS, MRMSE, CRPS) de pronósticos de caudales horarios en la estación Pisac para el periodo lluvioso 2022 (febrero-marzo), en tiempos de pronóstico de 1, 6, 12, 24 y 48 horas. ....	23
<b>Figura N° 8.</b> Pronóstico de caudales horarios en la estación Pisac para el periodo lluvioso 2022 (febrero-marzo), en tiempos de pronóstico de 1, 6, 12, 24 y 48 horas; empleando perturbaciones de datos de precipitación horaria de GSMaP-NRT' y el Filtro de Partículas (PF). ....	24

## ÍNDICE DE TABLAS

	<i>Pág.</i>
<b>Tabla N° 1.</b> Estaciones pluviométricas e hidrométricas con registro de horario en el periodo 2017-2022 en el dominio de estudio.....	13
<b>Tabla N° 2.</b> Productos de precipitación estimada por satélite en tiempo cuasi real.....	13
<b>Tabla N° 3.</b> Estadísticos de desempeño del modelo GR4H en la cuenca Vilcanota en Pisac. .....	15
<b>Tabla N° 4.</b> Estadísticos de evaluación de pronóstico hidrológico.....	18
<b>Tabla N° 5.</b> Estadísticos de desempeño del modelo GR4H en la cuenca Vilcanota en Pisac. .....	20

## 1. INTRODUCCIÓN

La gestión del riesgo de desastres, en materia de riesgos hidrometeorológicos, involucra el conocimiento y entendimiento de la dinámica de los procesos hidrológicos en una cuenca, con especial énfasis en la relación precipitación-escorrentía. En ese sentido, es importante contar con sistemas de monitoreo y pronóstico de caudales, en tiempo cuasi-real y a escala subdiaria, que faciliten la vigilancia operacional de caudales en una cuenca (Alvarez-Garretón et al., 2015; Klemeš, 1986), con la finalidad de proporcionar información oportuna a los tomadores de decisiones para la adopción de medidas de mitigación de los riesgos ante inundaciones. Herramientas como la modelación hidrológica permiten evaluar los cambios en la respuesta hidrológica de una cuenca, frente a perturbaciones en sus forzantes (Devia et al., 2015). Asimismo, la modelación a nivel operacional requiere de la implementación de un *framework* que facilite al usuario la actualización de los estados del modelo en cada paso de tiempo, y la evaluación en tiempo cuasi-real de las salidas de este (Noh et al., 2013).

Llauca et al. (2021a) desarrolló un modelo horario semidistribuido con fines de pronóstico de caudales para la cuenca del río Vilcanota en Intihuantana km 105. Para ello se empleó el modelo conceptual GR4H y se utilizaron diversos productos de precipitación estimada por satélite, combinados con información observada en estaciones pluviométricas automáticas, como datos de entrada en la modelación hidrológica a escala horaria. No obstante, la reducción de incertidumbre en la previsión hidrológica dificulta aún la vigilancia de crecidas producto de tormentas de corta duración, y más aún, la previsión de crecidas instantáneas (Li et al., 2015). En ese sentido, la aplicación de técnicas de asimilación de datos es una herramienta muy utilizada para mejorar las predicciones hidrológicas de flujos y estados en la modelación hidrológica basado en la incorporación de observaciones de caudales, humedad de suelo, evapotranspiración, entre otras variables disponibles (Nayak et al., 2021; Noh et al., 2013; Xiong et al., 2019).

Por ello, en el presente estudio se evalúan dos esquemas de asimilación de datos del tipo secuencial integrados al modelo GR4H en la cuenca Vilcanota en Pisac. Para su aplicación se perturbaron dos productos de precipitación estimada por satélite y corregidos con datos de estaciones automáticas. Finalmente, este estudio plantea responder las preguntas científicas:

- ¿Es factible incorporar esquemas de asimilación de datos secuenciales en un modelo hidrológico conceptual en una cuenca andina con escasa información?
- ¿Pueden mejorarse los pronósticos hidrológicos subdiarios en la cuenca Vilcanota en Pisac al asimilar datos de caudales observados en la cuenca?

## 2. ENFOQUE SECUENCIAL DE ASIMILACIÓN DE DATOS

### 2.1. Asimilación de datos

Los datos hidrometeorológicos observados (e.g. precipitación, temperatura del aire, caudal) pueden ser utilizados como forzantes del modelo (variables de entrada). No obstante, estos también pueden ser combinados con el modelo y utilizados en un esquema de actualización de las variables de estado en base a los errores que tienen las salidas del modelo con respecto a las observaciones (Vrugt et al., 2013). Esta última aplicación es conocida como Asimilación de Datos (DA). De acuerdo con (Clark et al., 2008), la idea central de la DA es cuantificar simultáneamente errores en el modelo hidrológico y observaciones; para luego actualizar los estados del modelo de tal manera que se combine óptimamente las simulaciones y observaciones.

De acuerdo con Mendoza et al. (2012), el éxito que pueda tener cualquier técnica de asimilación está fuertemente ligado a las fuentes de error involucradas. Por una parte, están los errores de las observaciones asimiladas por un modelo, los que pueden ser de tipo instrumental (directamente asociado a la instrumentación utilizada), de muestreo (problemas en cuanto a escala espacial de medición, interpolación o extrapolación inadecuada, etc.) y de transformación desde el espacio de observación al espacio de estados del modelo. Por otra parte, se encuentran los errores del modelo, que pueden ser de tres tipos: estructurales, paramétricos y asociados a los datos. Los errores estructurales son aquéllos derivados de los supuestos y simplificaciones utilizados en la formulación, que inevitablemente reproducirá de manera imperfecta el comportamiento del sistema. Los errores paramétricos están ligados a la representación espacio – temporal de propiedades heterogéneas del sistema real (Liu and Gupta, 2007).

Muchas técnicas de asimilación de datos pueden diferenciarse por el costo computacional y numérico para su aplicación. En la actualidad se han desarrollado diversas técnicas de asimilación de datos observados de caudales en un modelo hidrológico con fines de pronóstico (Sun et al., 2016). Por ejemplo, estos pueden seguir un enfoque variacional o secuencial.

En los métodos variacionales se asume usualmente un modelo de error constante en el tiempo, y se centra en identificar un conjunto de estados del modelo que minimicen una función de costo definida por la diferencia entre estados y observaciones. De esta manera se plantea un típico problema de minimización, en donde se aplica una versión linealizada del modelo hidrológico para calcular la gradiente de la función de costo. Una de las ventajas de los métodos variacionales es que no requieren predecir la matriz de covarianza de errores y ni de una formulación del espacio – estado como en los métodos secuenciales.

Por otro lado, los métodos secuenciales, conocidos también como técnicas de filtrado, actualizan el sistema siempre que existan observaciones disponibles ( $Y_t$ ). La actualización secuencial de los estados del modelo en el tiempo  $t(X_t)$  es definido como:

$$X_t = M[X_{(t-1)}, \theta, U_t, \Omega_t];$$

donde  $M$  es el modelo que depende de las entradas ( $U_t$ ), los parámetros ( $\theta$ ), y los errores del modelo ( $\Omega_t$ ). Los métodos secuenciales son más utilizados en aplicaciones en tiempo real, como es el caso de modelos hidrológicos operacionales con fines de pronóstico de caudales.

De acuerdo a (Piazzini et al., 2021), las técnicas de DA intentan mejorar la calidad de las predicciones incorporando la incertidumbre asociada a las mediciones (observaciones) y predicciones del modelo (simulaciones). Uno de sus principales retos es estimar cómo afectan los errores a las observaciones y estados del modelo. Esto último es importante para combinar óptimamente el estado del sistema estimado *a priori* (e.g. el estado base,  $X_t^b$ ) con las observaciones, para obtener un estado actualizado o de análisis ( $X_t^a$ ).

En la actualidad, las técnicas de DA más usadas en hidrología son el Filtro de Kalman de Conjuntos (EnKF), el Filtro de Partículas (PF), entre otros. A continuación, se describe de manera simplificada estos algoritmos de asimilación.

## 2.2. Filtro de Kalman de Conjuntos (EnKF)

El EnKF (Evensen, 2009) está basado en el Filtro de Kalman (KF). El KF es un método secuencial de asimilación de datos que fue diseñado para aplicaciones en sistemas lineales y procesos de medición con estadísticas de error gaussianas (Xiong et al., 2019). Debido a ello, el KF es notoriamente inestable si las no linealidades son fuertes (Clark et al., 2008). Sin embargo, la integración de esta ecuación no es factible computacionalmente para sistemas ambientales a gran escala. Para superar estas limitaciones, el EnKF utiliza un conjunto de trayectorias finitas, en el cual, las estimaciones del error del modelo son producto de asumir que la media del conjunto de trayectorias es “verdadera”, y calculando la varianza de las diferencias de cada miembro del conjunto y la media (Clark et al., 2008). Además, el EnKF utiliza un modelo no lineal para propagar los estados de conjunto, y evita las linealizaciones que hacen que el KF sea propenso a fallar.

## 2.3. Filtro de Partículas (PF)

El PF se diferencia de otros métodos de DA debido a que los estados del modelo no son “actualizados”. En lugar de ello, se asigna una probabilidad (peso) a cada miembro del conjunto basado en la diferencia entre cada miembro del conjunto y observaciones, y el error relativo entre el modelo y observaciones (Noh et al., 2013). Luego, la distribución de probabilidad

de las predicciones del modelo puede calcularse como una combinación ponderada de los miembros del conjunto. Por ello, una de las ventajas del PF respecto al EnKF es que no asumen una distribución gaussiana de los errores.

### 3. MATERIALES Y MÉTODOS

#### 3.1. Cuenca y datos

La cuenca del río Vilcanota se ubica al sureste del territorio peruano, en la región de Cusco y la vertiente hidrográfica del Atlántico (Figura N° 1). Esta cuenca es recurrentemente afectada por severas inundaciones debido al incremento del nivel del río Vilcanota producto de lluvias de gran intensidad. Recientes estudios han mostrado la gran necesidad de contar con simulaciones de caudales en la cuenca para mitigar los efectos de estos peligros hidrológicos y efectos del cambio climático (Llauca et al., 2021b; Saavedra et al., 2021). La cuenca del Vilcanota en la estación hidrológica Pisac tiene una extensión aproximada de 6900 km<sup>2</sup>, y presenta un relieve accidentado característico de los Andes Tropicales. En este trabajo se definió un dominio de estudio entre las latitudes 12.9°S y 14.8°S, y longitudes 70.6°W y 72.7°W; para el procesamiento de datos grillados de precipitación horaria estimada por satélite.

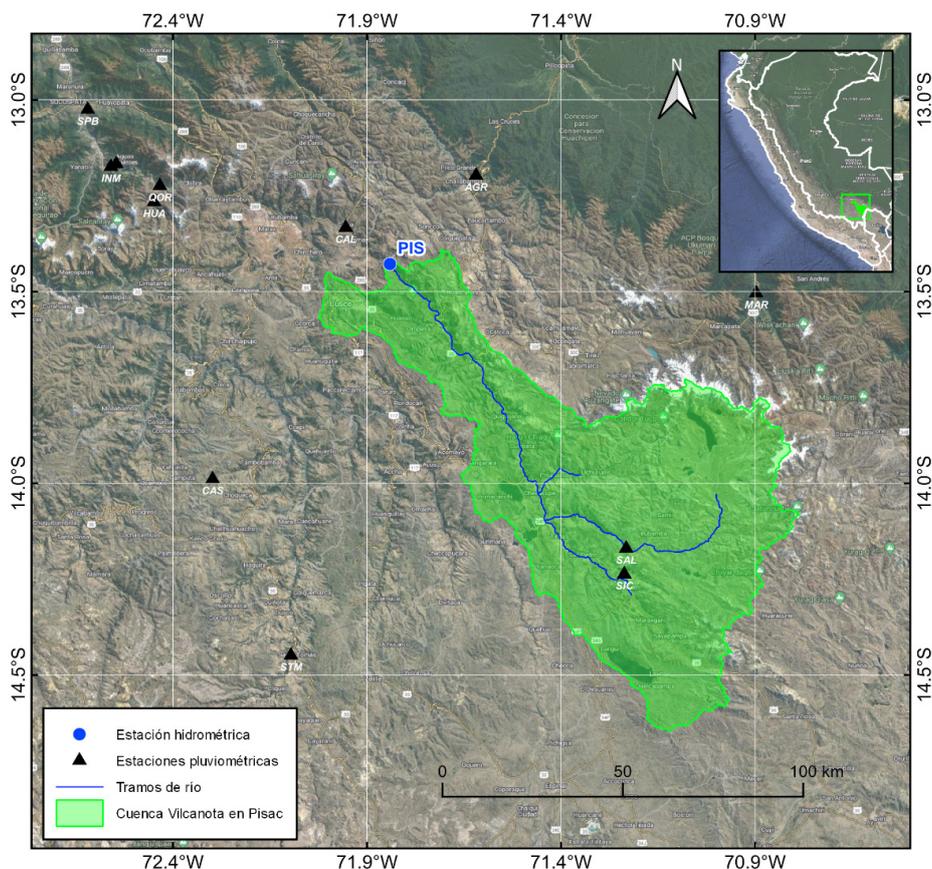
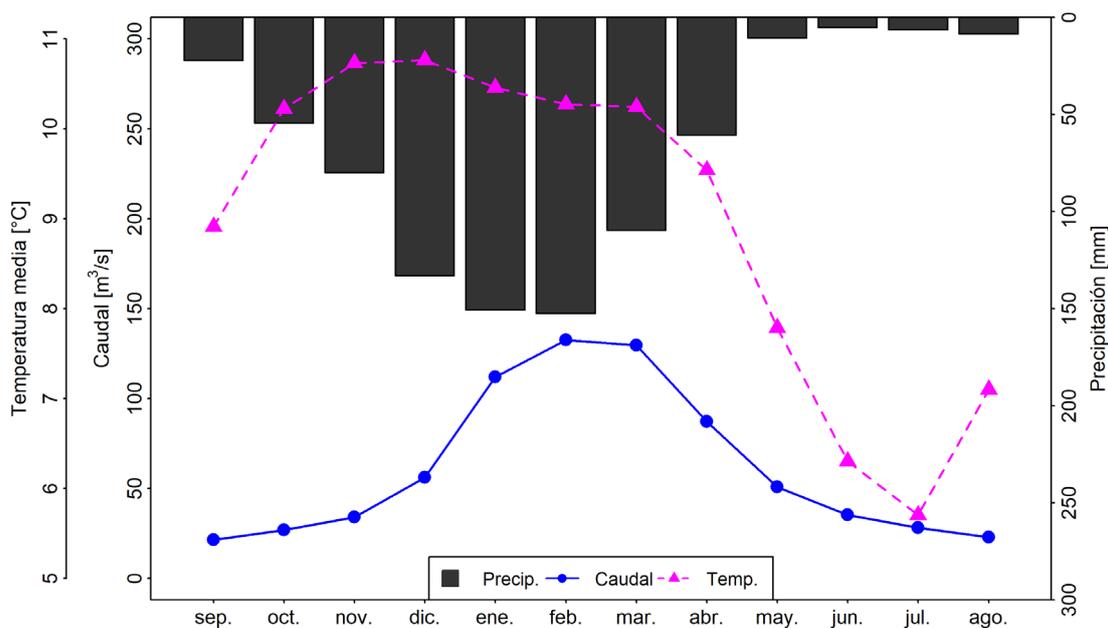


Figura N° 1. Mapa de ubicación de la cuenca del río Vilcanota en la estación Pisac.

La cuenca Vilcanota en Pisac presenta un hidroclimatología marcada (Figura N° 2) durante el año hidrológico (setiembre - agosto), con variaciones de la precipitación mensual con distribución unimodal y un periodo de máximas tasas de precipitación entre diciembre a marzo, alcanzando hasta los 150 mm/mes. La temperatura media del aire oscila entre los 6 °C a 11 °C, con un periodo cálido presente de octubre a abril y un periodo frío entre mayo y setiembre. Los caudales a la salida de la cuenca oscilan entre los 20 m³/s y 140 m³/s. El periodo de máximos caudales se presenta entre enero y abril, con valores extremos en febrero y marzo; mientras que el periodo de estiaje ocurre entre mayo y diciembre.



**Figura N° 2.** Variación mensual (1981-2016) de la precipitación media, caudal medio, y temperatura media del aire en la cuenca del río Vilcanota en Pisac.

Este estudio se recopiló información horaria de 12 estaciones pluviométricas automáticas y 1 estación hidrológica, para el periodo del 01 de enero de 2017 al 31 de julio de 2022. La Tabla N° 1 muestra el detalle de las estaciones hidrometeorológicas seleccionadas, mientras que su distribución espacial sobre el dominio de estudio se presenta en la Figura N° 1.

**Tabla N° 1.** Estaciones pluviométricas e hidrométricas con registro de horario en el periodo 2017-2022 en el dominio de estudio.

Nombre	Abrev.	Departamento	Provincia	Distrito	Latitud [°]	Longitud [°]	Elev. [msnm]
Acjanaco Gore	AGR	Cusco	Paucartambo	Kosnipata	-13.197	-71.619	3466
Calca	CAL	Cusco	Calca	Calca	-13.333	-71.955	2921
Casaccancha	CAS	Apurímac	Cotabambas	Tambobamba	-13.989	-72.298	4033
Huayllabamba	HUA	Cusco	Urubamba	Machupicchu	-13.266	-72.449	2976
Intihuatana M	INM	Cusco	Urubamba	Machupicchu	-13.174	-72.56	1778
Machupicchu	MAC	Cusco	Urubamba	Machupicchu	-13.167	-72.546	2399
Marcapata Gore	MAR	Cusco	Quispicanchi	Marcapata	-13.504	-70.897	1792
Qorihuayrachina	QOR	Cusco	Urubamba	Machupicchu	-13.224	-72.434	2517
Salcca	SAL	Cusco	Canchis	Combapata	-14.17	-71.232	3920
San Pablo	SPB	Cusco	La Convención	Huayopata	-13.025	-72.62	1228
Santo Tomas	STM	Cusco	Chumbivilcas	Santo Tomas	-14.449	-72.096	3665
Sicuani	SIC	Cusco	Canchis	Sicuani	-14.237	-71.237	3534
Pisac*	PIS	Cusco	Calca	Pisac	-13.428	-71.841	2791

\*Estación del tipo hidrométrica (observaciones de caudal).

**Tabla N° 2.** Productos de precipitación estimada por satélite en tiempo cuasi real.

Producto	Versión	Nombre corto	Institución	Resolution	Latencia (UTC-5)
Integrated Multi-satellite Retrievals for GPM	Early V06B	IMERG-E	NASA	0.1° x 0.1°	5 hr
Global Satellite Mapping of Precipitation	Near Real Time V06	GSMaP-NRT	JAXXA	0.1° x 0.1°	5 hr

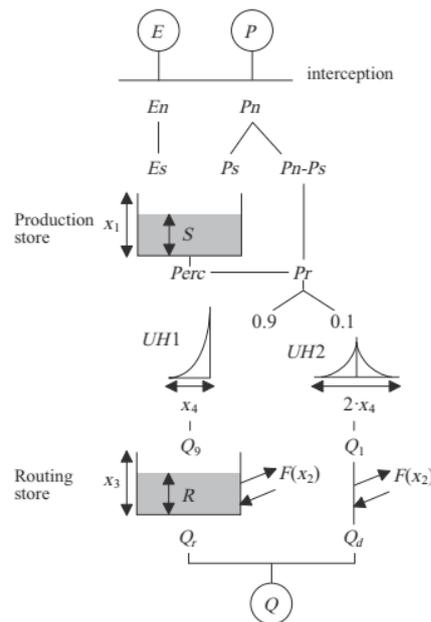
Con la finalidad de complementar el registro temporal y espacial de la precipitación horaria sobre la cuenca se seleccionaron dos productos de precipitación estimada por satélite con menor tiempo latencia, es decir, menor tiempo desde la captura de información por parte del satélite hasta estar disponible para la descarga del usuario. La Tabla N° 2 resume las principales características de los productos seleccionadas en este estudio.

### 3.2. Modelación hidrológica

#### 3.2.1. Modelo precipitación-escorrentía GR4H

El modelo GR4H (Le Moine, 2008) es la adaptación horaria del modelo GR4J (Perrin et al., 2003) de cuatro parámetros (X1, X2, X3 y X4). En este modelo, la precipitación y evapotranspiración potencial a escala horaria son consideradas para determinar la precipitación (Pn) y evapotranspiración (En) neta. Parte de Pn se pierde en el almacenamiento en el reservorio suelo S (Ps); mientras que lo restante conforma la

precipitación efectiva (siendo:  $P_t = P_n - P_s$ ). El contenido de humedad en el reservorio del suelo decae por percolación (Perc). Posteriormente,  $P_t$  es ruteada a la salida de la cuenca de la siguiente manera: 10% es transitado vía un solo hidrograma unitario, y el 90% restante es ruteada a través de un hidrograma unitario y un reservorio no lineal (R). Adicionalmente, una función de pérdidas (F) es aplicada a ambos flujos para representar el intercambio subterráneo (pérdida o ganancia en el sistema). La estructura del modelo GR4H se muestra en la Figura N° 3.



**Figura N° 3.** Estructura del modelo GR4H (Li et al., 2015).

Siguiendo la misma metodología empleada en (Llauca et al., 2021a), se calculó la precipitación media areal de la cuenca para el periodo 2017-2022 a partir de los productos IMERG-E' y GSMaP-NRT' (con corrección de sesgo). La evapotranspiración potencial de la cuenca se estimó a partir de la desagregación horaria de la climatología diaria de la evapotranspiración calculada el método de (Hargreaves and Samani, 1985) y datos de temperatura del aire del producto PISCOtemp.

### 3.2.2. Calibración del modelo

En modelo GR4H fue calibrado y validado empleando los datos de caudales horarios observados en la estación Pisac. Se tomó el periodo del 01 enero de 2017 00:00 horas al 31 agosto de 2020 23:00 horas para la calibración, y el periodo de 01 de setiembre de 2020 00:00 horas al 31 de julio de 2022 23:00 horas para la validación del modelo. Los primeros 1200 primeros valores del periodo de calibración se consideraron como periodo de calentamiento (warm-up) con la finalidad de independizar las salidas de las condiciones iniciales del modelo.

Los parámetros del modelo fueron calibrados empleando el algoritmo de optimización global Shuffled Complex Evolutionary (SCE-UA) (Duan et al., 1993). Siguiendo la metodología presentada en (Li et al., 2015), se utilizó una función objetivo ( $F_{obj}$ ) que es el promedio aritmético de la métrica de eficiencia de Nash Sutcliffe (Nash and Sutcliffe, 1970) en espacio logarítmico ( $\logNSE$ ), la eficiencia de Nash Sutcliffe con transformación Box-Cox ( $BoxNSE$ ), la eficiencia de Kling-Gupta ( $KGE$ ) (Gupta et al., 2009) y el sesgo ( $BIAS$ ). La  $F_{obj}$  empleada se muestra a continuación; asimismo, las ecuaciones para el cálculo de los estadísticos de desempeño de resumen en la Tabla N° 3.

$$F_{obj} = 0.25 * F_{\logNSE} + 0.25 * F_{BoxNSE} + 0.25 * F_{KGE} + 0.25 * F_{BIAS}$$

**Tabla N° 3.** Estadísticos de desempeño del modelo GR4H en la cuenca Vilcanota en Pisac.

Estadístico	Ecuación	Unid.	Min, Max, Óptimo	Énfasis
Eficiencia de Nash-Sutcliffe en espacio logarítmico ( $\logNSE$ )	$\logNSE = 1 - \frac{\sum_{t=1}^n (\log(Qsim_t) - \log(Qobs_t))^2}{\sum_{t=1}^n (\log(Qobs_t) - \overline{\log(Qobs)})^2}$	-	$-\infty, 1, 1$	Sensible a flujos bajos
Eficiencia de Nash-Sutcliffe en con transformación Box-Cox ( $BoxNSE$ )	$BoxNSE = 1 - \frac{\sum_{t=1}^n (Qsim'_t - Qobs'_t)^2}{\sum_{t=1}^n (Qobs'_t - \overline{Qobs'})^2}$ $Q' = \frac{(Q + 1)^\gamma - 1}{\gamma}$	-	$-\infty, 1, 1$	Sensible a flujos medios
Eficiencia de Kling-Gupta ( $KGE$ )	$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2}$ $r = \frac{\sum_{t=1}^n [(Qsim_t - \overline{Qsim})(Qobs_t - \overline{Qobs})]}{\sqrt{\sum_{t=1}^n (Qsim_t - \overline{Qsim})^2} \sqrt{\sum_{t=1}^n (Qobs_t - \overline{Qobs})^2}}$ $\alpha = \frac{\sigma_{sim}}{\sigma_{obs}}; \quad \beta = \frac{\mu_{sim}}{\mu_{obs}}$	-	$-\infty, 1, 1$	Sensible a la varianza y flujos altos.
Sesgo ( $BIAS$ )	$BIAS = \left[ \max \left( \frac{\overline{Qsim}}{\overline{Qobs}}, \frac{\overline{Qobs}}{\overline{Qsim}} \right) - 1 \right]^2$	%	$0, +\infty, 0$	Tendencia promedio de los datos simulados

### 3.3. Evaluación de la incertidumbre

Para aplicar las técnicas de asimilación de datos en modelos hidrológicos es importante cuantificar la incertidumbre en las componentes del modelo hidrológico (Liu and Gupta, 2007). En ese sentido, a continuación, se presenta un breve resumen de la modelación de los errores propagados en el modelo GR4H, desde las entradas hasta las salidas, siguiendo la misma metodología propuesta por Clark et al. (2008) y replicada por Piazza et al. (2021) en los modelos GR4J a través del paquete airGRDataAsim (Piazza and Delaigue, 2021).

### 3.3.1. Incertidumbre en las forzantes del modelo

El conjunto de previsiones meteorológicas se generó perturbando las series de tiempo medias areales y un multiplicador de ruido estocástico ( $\phi_p$ ) aplicado a cada paso de tiempo horario, acorde a la metodología propuesta por Clark et al. (2008). Se generaron perturbaciones de ambas forzantes meteorológicas (precipitación y evapotranspiración) para 50 miembros de conjunto, considerando un error fraccional de 0.65 y una decorrelación de 1 hora para la precipitación y 2 horas para la evapotranspiración.

$$p' = p\phi_p$$

$$\phi_p = (1 - \varepsilon_p) + 2u_p \varepsilon_p;$$

donde  $u_p$  es un número aleatorio uniforme, puesto que  $\phi_p$  es una realización de una distribución uniforme con límites de  $1 - \varepsilon_p$  a  $1 + \varepsilon_p$ .

### 3.3.2. Incertidumbre en las variables de estado del modelo

Si bien una de las mayores fuentes de incertidumbre en la modelación hidrológica, es importante incorporar la incertidumbre en los estados del modelo para representar la incertidumbre en el sistema. De acuerdo con Moradkhani et al. (2005), es posible perturbar las variables de estado aplicando un ruido gaussiano (con distribución normal). La varianza del ruido es asumida igual a la varianza de las variables de estado en análisis. No obstante, el valor de la varianza es restringida a ciertos umbrales para evitar inestabilidades.

### 3.3.3. Incertidumbre en las observaciones de caudal

Es conocido que existen incertidumbres en las mediciones de caudales, asociado a la construcción de las curvas altura-gasto. En ese sentido, en este trabajo también se considera la modelación del error en los caudales observados. El ruido en los caudales observados se generó a partir de una distribución normal con media 0 y varianza ( $\sigma_{obs}^2$ ), parametrizado como una función de los caudales observados ( $Q_{obs}$ ), tal como se muestra en la siguiente ecuación:

$$\sigma_{obs}^2 = (\varepsilon_{obs} Q_{obs})^2$$

Similar a Piazzzi et al. (2021), se consideró un valor de  $\varepsilon_{obs}$  de 0.1 y un umbral mínimo de la varianza igual al cuantil 10 ( $Q_{10}$ ).

### 3.4. Diseño de experimentos de asimilación

A continuación, se resumen los experimentos realizados en el presente trabajo, para aplicar los algoritmos secuenciales de asimilación de datos:

- a. **Modelo Open Loop (OL):** Corresponde a la corrida inicial del modelo GR4H (con parámetros previamente calibrados y validados) con el conjunto de forzantes meteorológicas perturbadas (50 miembros), y sin aplicar técnicas de asimilación de datos.
  - ✓ IMERG-E'+OL: Corresponde al modelo GR4H ejecutado con las 50 perturbaciones del producto IMERG-E con corrección de sesgo (IMERG-E').
  - ✓ GSMaP-NRT+OL: Corresponde al modelo GR4H ejecutado con las 50 perturbaciones del producto GSMaP-NRT con corrección de sesgo (GSMaP-NRT').
  
- b. **Filtro de Kalman de Conjuntos (EnKF):** Corresponde a la corrida del modelo GR4H (con parámetros previamente calibrados y validados) con el conjunto de forzantes meteorológicas perturbadas (50 miembros) y el algoritmo EnKF para asimilar los caudales observados horarios.
  - ✓ IMERG-E'+EnKF: Corresponde al modelo GR4H ejecutado con las 50 perturbaciones del producto IMERG-E con corrección de sesgo (IMERG-E') y aplicando el algoritmo EnKF.
  - ✓ GSMaP-NRT+EnKF: Corresponde al modelo GR4H ejecutado con las 50 perturbaciones del producto GSMaP-NRT con corrección de sesgo (GSMaP-NRT') y aplicando el algoritmo EnKF.
  
- c. **Filtro de Partículas (PF):** Corresponde a la corrida del modelo GR4H (con parámetros previamente calibrados y validados) con el conjunto de forzantes meteorológicas perturbadas (50 miembros) y el algoritmo PF para asimilar los caudales observados horarios.
  - ✓ IMERG-E'+PF: Corresponde al modelo GR4H ejecutado con las 50 perturbaciones del producto IMERG-E con corrección de sesgo (IMERG-E') y aplicando el algoritmo PF.
  - ✓ GSMaP-NRT+PF: Corresponde al modelo GR4H ejecutado con las 50 perturbaciones del producto GSMaP-NRT con corrección de sesgo (GSMaP-NRT') y aplicando el algoritmo PF.

### 3.5. Evaluación de pronósticos subsidiarios

Para evaluar el desempeño del conjunto de pronósticos de cada experimento de asimilación sobre distintos horizontes de pronóstico se emplearon métricas de evaluación detalladas en la Tabla N° 4. En base a la estacionalidad de los caudales máximos en la cuenca (Figura N° 2), se seleccionaron los meses más lluviosos del 2022 (febrero y enero) para evaluar el desempeño del conjunto de pronósticos para horizontes de pronóstico de 1, 6, 12, 24 y 48 horas; tomando como línea base los estados actualizados (*a posteriori*) del modelo GR4H (S, R, HU1 y HU2) luego de asimilar los datos observados de caudales en la estación Pisac y aplicar los algoritmos EnKF y PF. Es decir, para el periodo del 01 de febrero de 2022 al 31 de marzo de 2022 se realizaron pronósticos con horizonte de hasta 48 horas en cada paso de tiempo del periodo mencionado.

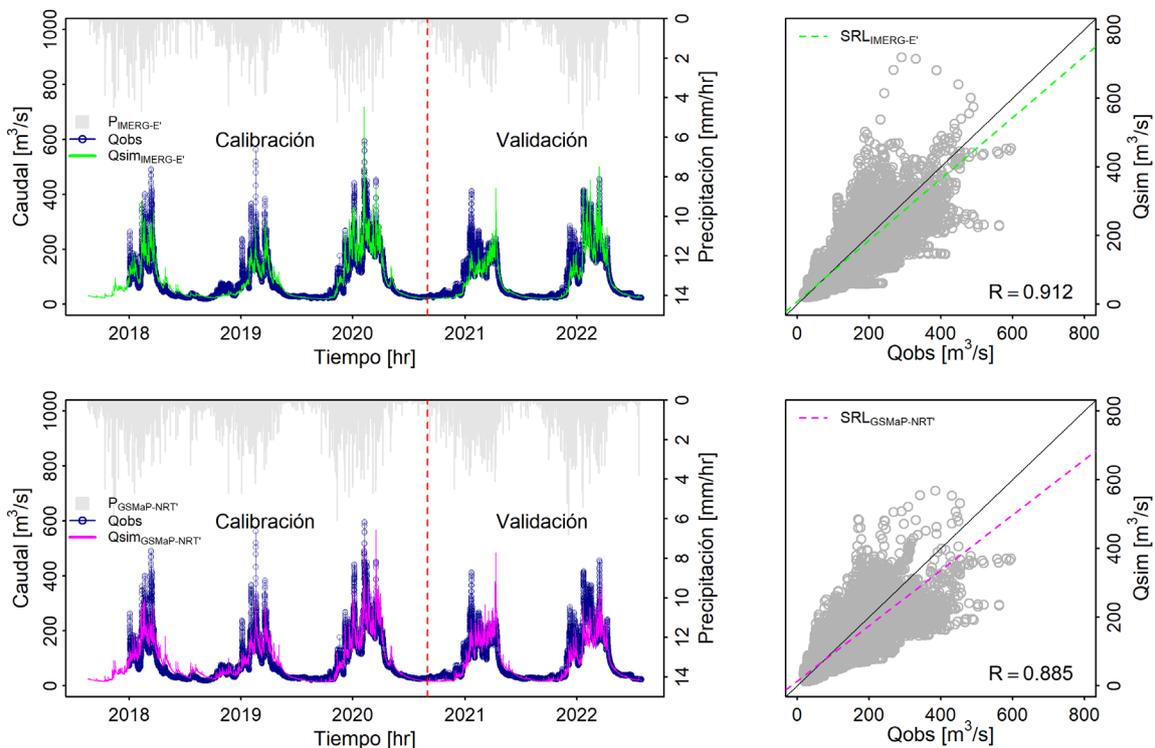
**Tabla N° 4.** Estadísticos de evaluación de pronóstico hidrológico.

Estadístico	Ecuación	Unid.
Eficiencia de Nash-Sutcliffe (NSE)	$NSE = 1 - \frac{\sum_{t=1}^n (Q_{fcst_t} - Q_{obs_t})^2}{\sum_{t=1}^n (Q_{obs_t} - \bar{Q}_{obs})^2}$	-
Sesgo de la media del conjunto (BIAS)	$BIAS = \frac{1}{n} \sum_{t=1}^n (Q_{fcst_t} - Q_{obs_t})$	[m³/s]
Media del error cuadrático medio del conjunto (MRMSE)	$MRMSE = \frac{1}{n} \sum_{t=1}^n \sqrt{\frac{1}{F} \sum_{t=1}^F (Q^j_{fcst_t} - Q_{obs_t})^2}$	[m³/s]
Puntuaciones de rango de probabilidad continuo (CRPS)	$CRPS = \frac{1}{n} \sum_{t=1}^n \int_{-\infty}^{\infty} (F_{fcst_t}(Q) - F_{obs_t}(Q))^2 dQ$ $F_{obs_t}(Q) = \begin{cases} 0 & (Q < Q_{obs_t}) \\ 1 & (Q \geq Q_{obs_t}) \end{cases}$	[m³/s]

## 4. RESULTADOS Y DISCUSIÓN

### 4.1. Calibración del modelo y estimación de la incertidumbre

Las series de caudales observados y simulados durante el periodo de calibración y validación del modelo GR4H en la cuenca Vilcanota en Pisac se muestra en la Figura N° 4 (paneles izquierdos). La serie de tiempo azul representa los datos observados y líneas verde (IMERG-E') y magenta (GSMaP-NRT') corresponden a los valores simulados empleando datos de precipitación grillados a escala areal (barras grises). De ello se observa que las series simuladas representan adecuadamente la variabilidad temporal de los caudales en la cuenca, logrando simular los periodos de avenidas y estiaje durante la calibración y validación. Los paneles derechos de la Figura N° 4 muestra los diagramas de dispersión del caudal horario observado y simulado en Pisac. Se observa que existe relación lineal entre la variable observada y simulada, con altos coeficientes de correlación ( $R_{\text{IMERG-E}'}=0.912$  y  $R_{\text{GSMaP-NRT}'}=0.885$ ). Las pendientes de los modelos de regresión lineal simple (SRL) respecto a la línea de identidad 1:1 (línea negra) dan cuenta de una ligera subestimación de los caudales horarios por sobre los 100 m<sup>3</sup>/s aproximadamente; siendo esta subestimación ligeramente mayor en el modelo forzado con GSMaP-NRT'.



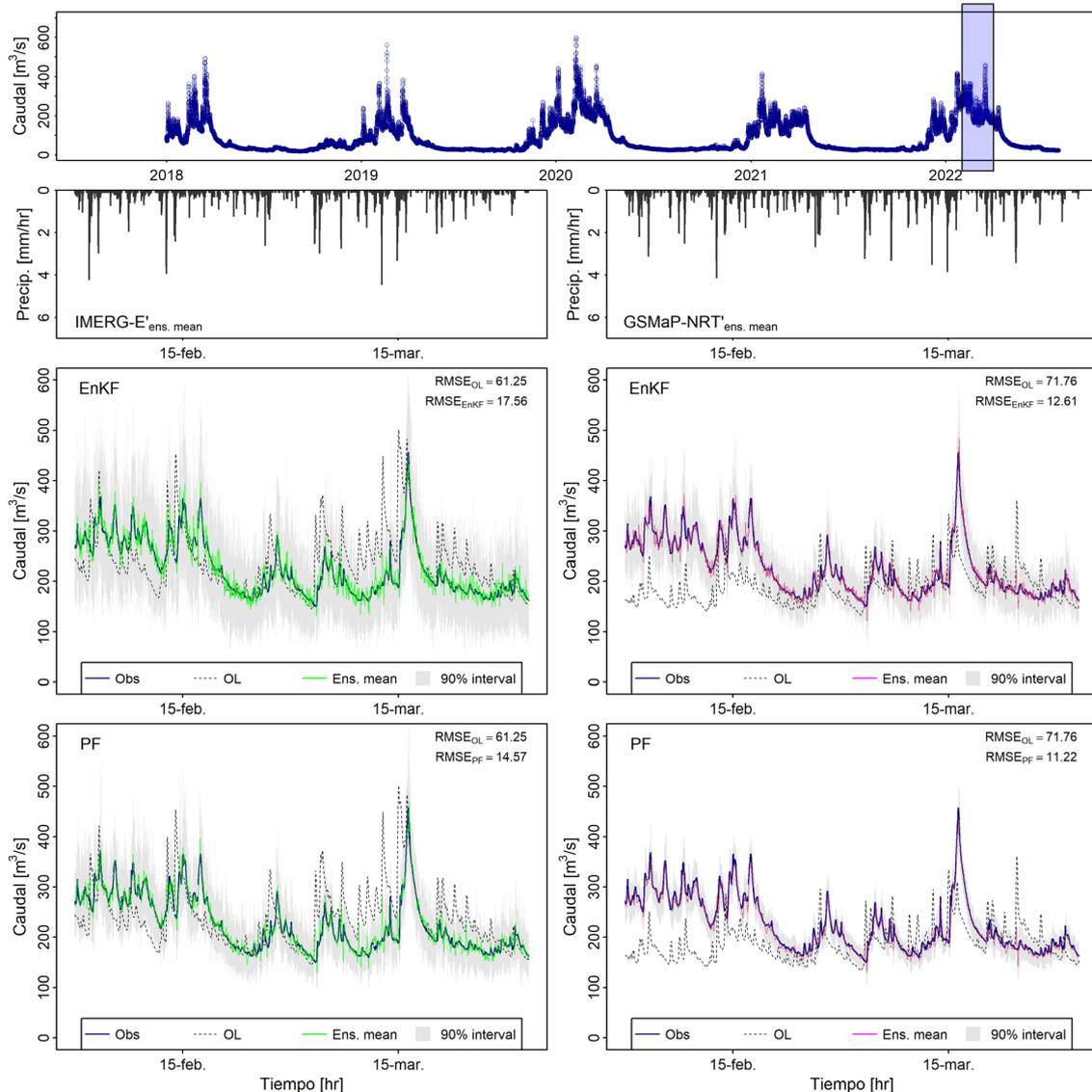
**Figura N° 4.** Calibración y validación del modelo GR4H en la cuenca Vilcanota en Pisac, empleando datos de precipitación horaria de IMERG-E' y GSMaP-NRT'.

El resumen de los estadísticos de desempeño obtenidos para la calibración y validación del modelo se presentan en la Tabla N° 5. En términos del logNSE, BoxNSE y KGE el desempeño del modelo en los flujos bajos, medios y altos (mayor a 0.780), respectivamente, es ligeramente mayor empleando datos IMERG-E' en comparación a GSMaP-NRT'. Asimismo, estos indicadores de eficiencia son mayores durante la calibración y decaen en la validación. En términos del BIAS, se evidencia que existe una muy ligera subestimación de las simulaciones (valores positivos cercanos a cero).

**Tabla N° 5.** Estadísticos de desempeño del modelo GR4H en la cuenca Vilcanota en Pisac.

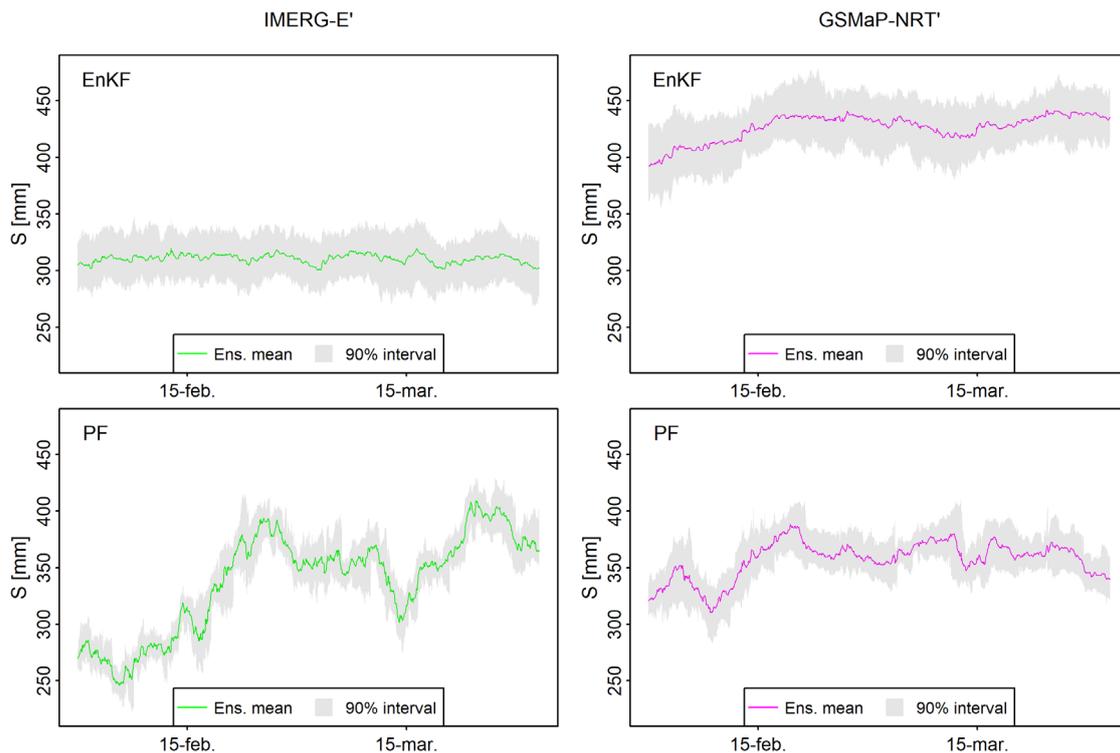
Métrica	Calibración		Validación	
	IMERG-E'	GSMaP'	IMERG-E'	GSMaP'
<b>logNSE</b>	0.875	0.792	0.878	0.786
<b>BoxNSE</b>	0.883	0.831	0.878	0.819
<b>KGE</b>	0.912	0.871	0.869	0.789
<b>BIAS</b>	0.003	0.003	0.016	0.029

La Figura N° 5 muestra las series de caudales horarios observados (línea azul) y simulados del modelo Open Loop (OL, línea negra segmentada), valor promedio (de verde y magenta) e intervalos de confianza al 90% (gris) del conjunto de simulaciones (EnKF y PF) forzadas con perturbaciones de los IMERG-E' y GSMaP-NRT' (barras grises oscuras) para el periodo lluvioso del 01 de febrero al 31 de marzo de 2022 (caja del panel superior de la Figura N° 5). Los resultados muestran el mejor ajuste de las simulaciones a las observaciones luego de aplicar la asimilación de datos con IMERG-E'+EnKF, IMERG-E'+PF, GSMaP-NRT'+EnKF y GSMaP-NRT'+PF a diferencia de las simulaciones del OL. Esto queda evidenciado en la reducción de la Raíz del Error Cuadrático Medio (RMSE) los experimentos de asimilación respecto a la simulación OL (por ejemplo, de 71.76 m<sup>3</sup>/s a 12.61 m<sup>3</sup>/s en el experimento GSMaP-NRT'+EnKF). También se observa que el menor RMSE (11.22 m<sup>3</sup>/s) se obtuvo en el experimento GSMaP-NRT'+PF; mientras que el mayor RMSE (17.56 m<sup>3</sup>/s) se obtuvo en el experimento IMERG-E'+EnKF. Finalmente, se observa que la banda de incertidumbre al 90 % de intervalo de confianza es menor en los experimentos con la forzante GSMaP-NRT' en comparación a IMERG-E', para un conjunto compuesto por 50 miembros.



**Figura N° 5.** Simulación de caudales horarios en la estación Pisac para el periodo lluvioso 2022 (febrero-marzo); empleando perturbaciones de precipitación horaria de IMERG-E' y GSMaP-NRT' y el Filtro de Kalman de Conjuntos (EnKF) y Filtro de Partículas (FP).

La incertidumbre en las estimaciones del contenido de humedad en el suelo se muestra en la Figura N° 6. Las series de tiempo de color verde y magenta representan la media del conjunto de simulaciones de la variable de estado  $S$  (50 miembros) para los cuatro experimentos de asimilación; mientras que las bandas grises representan las bandas de incertidumbre para un intervalo de confianza del 90%. Se observa que existe alta incertidumbre asociada a la representación del contenido de agua en el suelo a través del reservorio de producción del modelo GR4H. Las simulaciones con el algoritmo PF presentan más variabilidad en los valores de  $S$  durante el periodo lluvioso 2022 (de 230 mm a 410 mm en IMERG-E'+PF y de 230 mm a 408 mm en GSMaP-NRT'+PF). Asimismo, se evidencia que el rango de variabilidad de  $S$  estaría sujeta al experimento de asimilación utilizado.

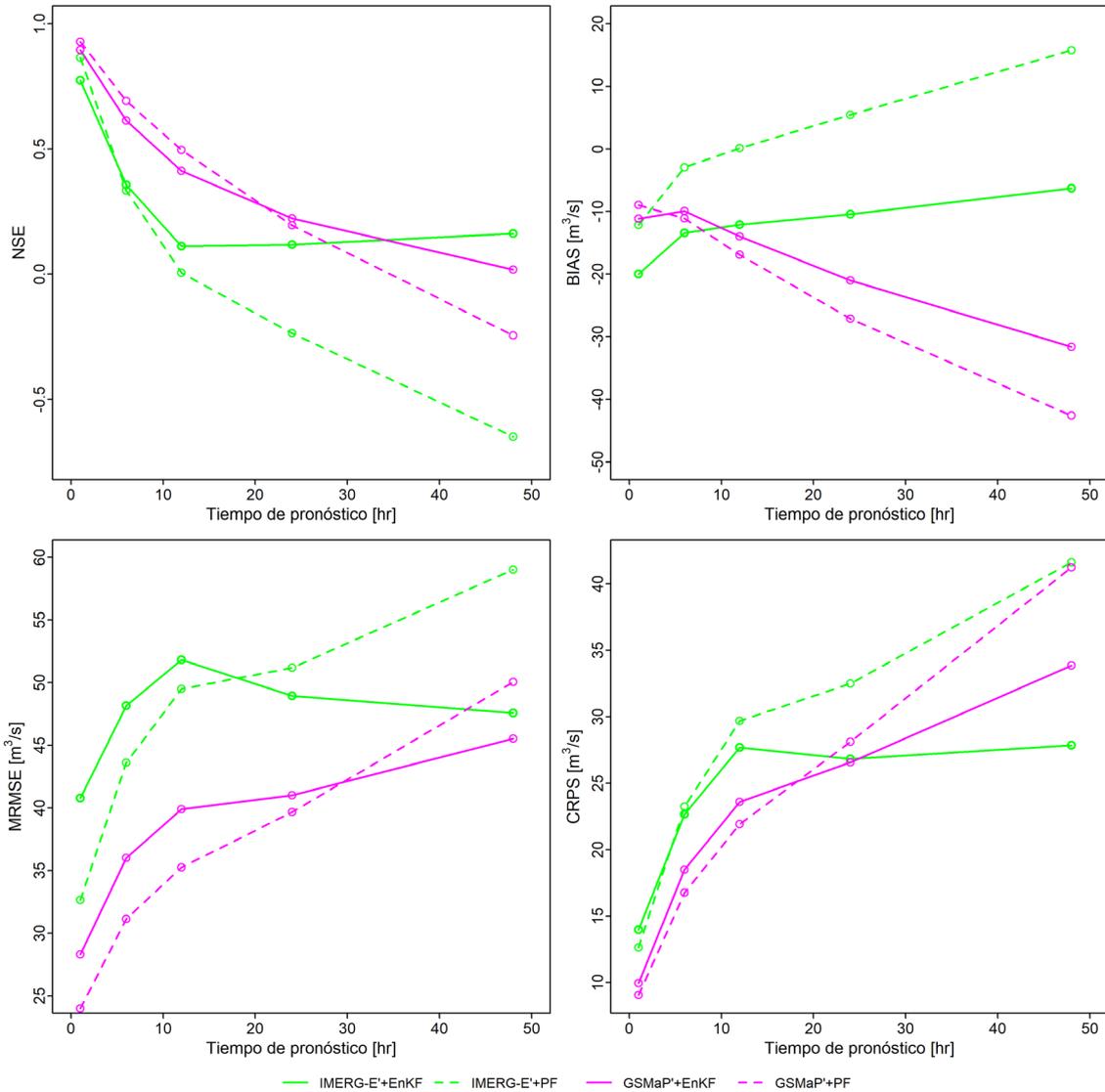


**Figura N° 6.** Simulación del contenido de humedad del suelo en la cuenca Vilcanota en Pisac para el periodo lluvioso 2022 (febrero-marzo); empleando perturbaciones de precipitación horaria de IMERG-E' y GSMaP-NRT' y el Filtro de Kalman de Conjuntos (EnKF) y Filtro de Partículas (FP).

#### 4.2. Beneficios de la asimilación de caudales observados

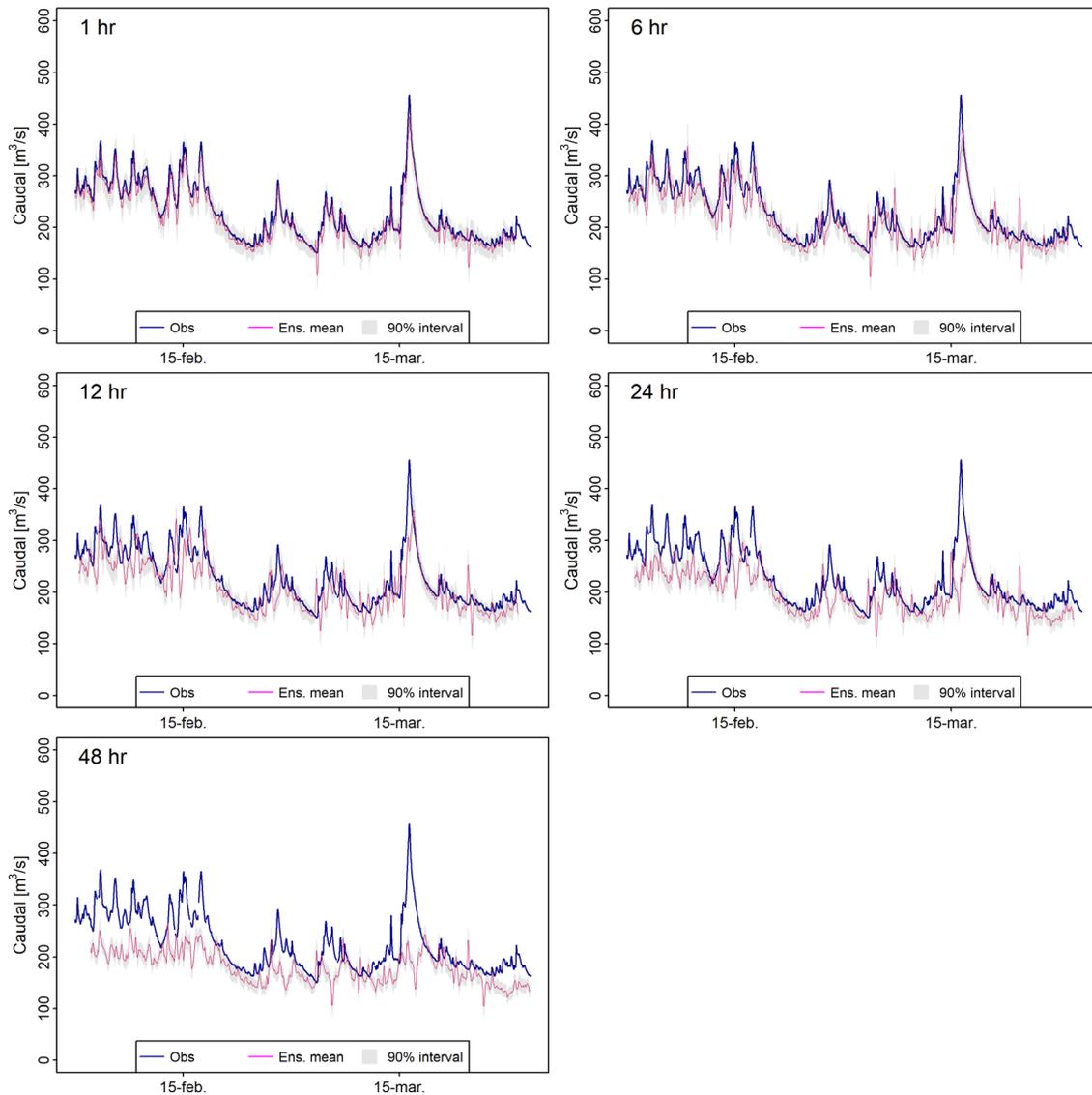
El desempeño de los pronósticos de caudales horarios para horizontes de 1, 6, 12, 24 y 48 horas se muestra en la Figura N° 7. Las métricas de NSE, BIAS, MRMSE y CRPS evidencian el decaimiento del desempeño de los pronósticos conforme incrementa el horizonte de pronóstico. En términos del NSE, la eficiencia de los cuatro experimentos de asimilación decae notablemente ( $NSE \leq 0.5$ ) a partir de las 6 horas de pronóstico; siendo los experimentos GSMaP-NRT'+EnKF y GSMaP-NRT'+PF aquellos con mejor ajuste. En términos del BIAS, los experimentos con GSMaP' muestran un incremento del sesgo positivo (tendencia a la sobreestimación de caudales horarios) luego de las 6 horas de pronóstico, mientras que los experimentos IMERG-E' muestran un incremento del sesgo negativo (tendencia a la subestimación de caudales horarios). Hacia las 48 horas de pronóstico el BIAS tiene un rango amplio de variación entre los  $-45 \text{ m}^3/\text{s}$  y  $15 \text{ m}^3/\text{s}$ . El desempeño de los pronósticos en términos del MRMSE empeora conforme el horizonte de pronóstico. Se observa que hacia 1 hora de pronóstico los experimentos valores muy diferentes entre  $22 \text{ m}^3/\text{s}$  (GSMaP-NRT'+PF) a  $40.5 \text{ m}^3/\text{s}$  (IMERG-E'+EnKF) y se incrementa conforme se alcanza las 48 horas de pronóstico. También se observa también que, en el experimento IMERG-E'+EnKF, el MRMSE llega a su máximo valor a las 12 horas de pronóstico y decae hacia las 48 horas. Finalmente, el

CRPS indica que el ajuste de las funciones de distribución acumulada (CDF) de los caudales pronosticados respecto a las observaciones empeora mientras incrementa el horizonte de pronóstico.



**Figura N° 7.** Evaluación del desempeño (NSE, BIAS, MRMSE, CRPS) de pronósticos de caudales horarios en la estación Pisac para el periodo lluvioso 2022 (febrero-marzo), en tiempos de pronóstico de 1, 6, 12, 24 y 48 horas.

La Figura N° 8 muestra la simulación de los caudales horarios pronosticados en el periodo de lluvias de febrero y marzo 2022 para horizontes de pronóstico de 1, 6, 12, 24 y 48 horas. Los resultados muestran el decaimiento de la representación de las observaciones de caudales horarios conforme incrementa el horizonte de pronóstico. Gráficamente se observa que las series simuladas coinciden con las simuladas hasta un horizonte aproximado de 12 horas para luego notarse subestimaciones notables hacia las 24 y 48 horas de pronóstico.



**Figura N° 8.** Pronóstico de caudales horarios en la estación Pisac para el periodo lluvioso 2022 (febrero-marzo), en tiempos de pronóstico de 1, 6, 12, 24 y 48 horas; empleando perturbaciones de datos de precipitación horaria de GSMaP-NRT' y el Filtro de Partículas (PF).

## 5. CONCLUSIONES Y RECOMENDACIONES

En el presente trabajo se explora la aplicación de dos técnicas de asimilación de datos secuenciales (el Filtro de Kalman de Conjuntos, EnKF; y el Filtro de Partículas, PF) para actualizar los estados del modelo hidrológico GR4H en la cuenca Vilcanota en Pisac. Para simular los caudales subdiarios en la cuenca se utilizaron las forzantes de precipitación horaria con corrección de sesgo de los productos IMERG-E' y GSMaP-NRT'. Asimismo, como punto de partida, se ejecutó el modelo Open Loop (OL) perturbando las forzantes de precipitación y evapotranspiración potencial, y generando un conjunto con 50 miembros. Las principales conclusiones del presente trabajo se mencionan a continuación:

- Se demostró la factibilidad de incorporar datos observados de caudales horarios en un esquema secuencial de asimilación de datos aplicando los algoritmos del EnKF y PF en un modelo hidrológico agregado, perturbando las forzantes meteorológicas del modelo para generar conjunto de datos de entrada con 50 miembros. La aplicación del EnKF y PF en un modelo conceptual agregado requirió de cuatro horas de procesamiento por cada experimento (en un periodo a paso horario desde enero 2017 a julio 2022); por lo que su implementación en un esquema de modelación semidistribuida y de base física debería implementarse en un clúster de alto rendimiento. No obstante, su implementación operativa (actualización de estados a un tiempo de pronóstico de 48 horas) requeriría de mucho menos tiempo de cómputo.
- Las simulaciones para los meses lluviosos del 2022 (enero y febrero) en la cuenca Vilcanota en Pisac muestran la reducción de los errores en las salidas del modelo para las simulaciones con asimilación de datos respecto a la ejecución sin asimilación (OL). Se evidencia también buenas métricas de desempeño en las simulaciones de caudales horarios con hasta tiempos de pronóstico de 6 horas. Luego de ello, el desempeño de las predicciones decae significativamente hacia las 48 horas.
- En futuros trabajos, se recomienda incorporar experimentos de asimilación de datos en un esquema de modelación semidistribuida del modelo GR4H, y evaluar las diferencias con la implementación en un esquema agregado. Por otro lado, se considera explorar la asimilación de datos de sensoramiento remoto de humedad del suelo para reducir la incertidumbre en las estimaciones en esta variable de estado.

## 6. REFERENCIAS BIBLIOGRÁFICAS

- Alvarez-Garretón, 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. *Hydrology and Earth System Sciences* 19, 1659–1676. <https://doi.org/10.5194/hess-19-1659-2015>
- 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>
- Devia, G.K., Ganasri, B.P., Dwarakish, G.S., 2015. A Review on Hydrological Models. *Aquatic Procedia* 4, 1001–1007. <https://doi.org/10.1016/j.aqpro.2015.02.126>
- Duan, Q.Y., Gupta, V.K., Sorooshian, S., 1993. Shuffled complex evolution approach for effective and efficient global minimization. *J. Optim. Theory Appl.* 76, 501–521. <https://doi.org/10.1007/BF00939380>
- Evensen, G., 2009. *Data Assimilation*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-03711-5>
- 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. <https://doi.org/10.1016/j.jhydrol.2009.08.003>
- Hargreaves, G.H., Samani, Z.A., 1985. Reference crop evapotranspiration from ambient air temperature. *Chicago. Applied Engineering in Agriculture* 1, 96–99. <https://doi.org/10.13031/2013.26773>
- Klemeš, V., 1986. Operational testing of hydrological simulation models. *Hydrological Sciences Journal* 31, 13–24. <https://doi.org/10.1080/02626668609491024>
- Le Moine, N., 2008. Le bassin versant de surface vu par le souterrain : une voie d'amélioration des performances et du réalisme des modèles pluie-débit ?
- Li, Y., Ryu, D., Western, A.W., Wang, Q.J., 2015. Assimilation of stream discharge for flood forecasting: Updating a semidistributed model with an integrated data assimilation scheme. *Water Resour. Res.* 51, 3238–3258. <https://doi.org/10.1002/2014wr016667>
- Liu, Y., Gupta, H.V., 2007. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resour. Res.* 43, 160. <https://doi.org/10.1029/2006WR005756>
- Llauca, H., Lavado-Casimiro, W., León, K., Jimenez, J., Traverso, K., Rau, P., 2021a. Assessing Near Real-Time Satellite Precipitation Products for Flood Simulations at Sub-Daily Scales in a Sparsely Gauged Watershed in Peruvian Andes. *Remote Sensing* 13, 826. <https://doi.org/10.3390/rs13040826>
- Llauca, H., Lavado-Casimiro, W., Montesinos, C., Santini, W., Rau, P., 2021b. PISCO\_HyM\_GR2M: A Model of Monthly Water Balance in Peru (1981–2020). *Water*. <https://doi.org/10.3390/w13081048>

- 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>
- Moradkhani, H., Hsu, K.-L., Gupta, H., Sorooshian, S., 2005. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resour. Res.* 41. <https://doi.org/10.1029/2004wr003604>
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology.* [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Nayak, A.K., Biswal, B., Sudheer, K.P., 2021. Role of hydrological model structure in the assimilation of soil moisture for streamflow prediction. *J. Hydrol.* 598, 126465. <https://doi.org/10.1016/j.jhydrol.2021.126465>
- Noh, S.J., Tachikawa, Y., Shiiba, M., Kim, S., 2013. Sequential data assimilation for streamflow forecasting using a distributed hydrologic model: Particle filtering and ensemble Kalman filtering. *IAHS-AISH Publication 357*, 341–349.
- Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. *J. Hydrol.* 279, 275–289. [https://doi.org/10.1016/S0022-1694\(03\)00225-7](https://doi.org/10.1016/S0022-1694(03)00225-7)
- Piazzì, G., Delaigue, O., 2021. Ensemble-based data assimilation with GR hydrological models (v. 0.1.3).
- Piazzì, G., Thirel, G., Perrin, C., Delaigue, O., 2021. Sequential data assimilation for streamflow forecasting: Assessing the sensitivity to uncertainties and updated variables of a conceptual hydrological model at basin scale. *Water Resour. Res.* 57. <https://doi.org/10.1029/2020wr028390>
- Saavedra, D., Mendoza, P.A., Addor, N., Llauca, H., Vargas, X., 2021. A multi-objective approach to select hydrological models and constrain structural uncertainties for climate impact assessments. *Hydrol. Process.* <https://doi.org/10.1002/hyp.14446>
- Sun, L., Seidou, O., Nistor, I., Liu, K., 2016. Review of the Kalman-type hydrological data assimilation. *Hydrol. Sci. J.* 61, 2348–2366. <https://doi.org/10.1080/02626667.2015.1127376>
- Vrugt, J.A., ter Braak, C.J.F., Diks, C.G.H., Schoups, G., 2013. Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications. *Adv. Water Resour.* 51, 457–478. <https://doi.org/10.1016/j.advwatres.2012.04.002>
- Xiong, M., Liu, P., Cheng, L., Deng, C., Gui, Z., Zhang, X., Liu, Y., 2019. Identifying time-varying hydrological model parameters to improve simulation efficiency by the ensemble Kalman filter: A joint assimilation of streamflow and actual evapotranspiration. *J. Hydrol.* 568, 758–768. <https://doi.org/10.1016/j.jhydrol.2018.11.038>

## ANEXOS