

SOCIEDAD CHILENA DE INGENIERÍA HIDRÁULICA
XXIV CONGRESO CHILENO DE INGENIERÍA HIDRÁULICA

**EVALUACIÓN DE MÉTODOS ESTADÍSTICOS PARA PRONÓSTICOS DE
VOLUMEN DE DESHIELO EN CUENCAS DEL MAULE**

ALEJANDRA ISAMIT FAURE¹
CONSTANZA VÁSQUEZ DIAZ²
NICOLÁS VÁSQUEZ PLACENCIA³
XIMENA VARGAS MESA⁴
MIGUEL LAGOS ZÚÑIGA⁵

RESUMEN

El pronóstico de volúmenes de deshielo es un insumo necesario para diversos usuarios de agua que usan este recurso en época estival. Este problema se resuelve usualmente mediante enfoques de regresión lineal múltiple y presenta una complicación para cuencas en donde la componente nival no es tan dominante. En este estudio se evalúa el desempeño histórico del pronóstico en las cuencas de Maule en Armerillo (MArm) y Achibueno en la Recova (AchR) aplicando tres métodos estadísticos: Regresión Lineal Múltiple (MLR), Principal Components Regression (PCR) y Parcial Least Squares Regression (PLSR). Los predictores son la precipitación acumulada, el caudal medio mensual, el equivalente en agua de nieve y la cobertura nival fraccional. El modelo que presenta un mejor desempeño corresponde al PLSR, sin embargo, al evaluar su desempeño en distintos tipos de años, se aprecia una sobrestimación del volumen en años normales y subestimación en años húmedos, siendo esta más significativa en la cuenca de MArm. En ambas cuencas la adición de información del mes de septiembre mejora el modelo en mayor parte que al agregar también información del mes de octubre, por lo que se puede pronosticar el volumen de deshielo de la temporada de primavera-verano a fines de septiembre. El error volumétrico promedio encontrado es de 8.4% y 11% para Marm y AchR respectivamente, lo que permite utilizar esta metodología con propósitos operacionales.

¹Estudiante Pregrado, Departamento de Ingeniería Civil, Universidad de Chile – maria.isamit@ug.uchile.cl

²Estudiante Pregrado, Departamento de Ingeniería Civil, Universidad de Chile – constanza.vasquezd@gmail.com

³ Profesional Docente, Depto. Ing. Civil e Ingeniero Civil en Arcadis. nicolas.vasquez.pl@uchile.cl

⁴ Profesor Asociado, Departamento de Ingeniería Civil, Universidad de Chile – xvargas@uchile.cl

⁵ Profesional Docente, Depto. Ing. Civil e Investigador Asociado AMTC, U. de Chile – mlagosz@uchile.cl

1. INTRODUCCIÓN

Para el análisis de los recursos disponibles para el riego por parte de los agricultores y para controlar el funcionamiento de centrales hidroeléctricas es de suma importancia hacer una gestión eficiente del recurso hídrico. La demanda de recursos hídricos en Chile ocurre, principalmente, en la temporada de primavera-verano (septiembre a marzo).

Diversos modelos dinámicos y estadísticos de variables hidrológicas, influyentes en la estimación de recursos hídricos, han sido generados. Por ejemplo, la Dirección General de Aguas (DGA) genera pronósticos anuales del volumen de deshielo y caudales medios mensuales, a partir de modelos de regresión lineal múltiple, para la temporada de riego del periodo primavera-verano, con el propósito de estimar la situación hidrológica general del país, específicamente en lo relativo a las zonas de riego de las principales cuencas del Norte Chico, zona central y sur (desde la cuenca del río Copiapó hasta la cuenca del río Ñuble) (DGA, 2018). Sin embargo, los pronósticos realizados pueden llegar a sobreestimar el volumen de deshielo hasta un 60% en años secos, por lo que nace la necesidad de analizar nuevos modelos de carácter estadísticos que consideren otros parámetros y/o técnicas de modelación.

La precipitación acumulada en un periodo de tiempo y el caudal medio mensual han sido propuestos como predictores para realizar pronósticos. Lagos (2015) al considerar la fracción de área cubierta de nieve o fSCA como predictor del volumen de deshielo disminuye el error de los modelos hasta un 4%. Por otra parte, Mendoza (2014) propone utilizar el máximo valor del equivalente en agua de nieve o SWE_{max} como predictor en modelos de predicción de volumen de deshielo.

Este trabajo pretende responder a las incógnitas que surgen al generar un modelo que prediga los volúmenes de deshielo tales como, ¿Mejoran las estimaciones al utilizar diferentes modelos?, ¿Con cuánto tiempo de anticipación es posible realizar un buen pronóstico? y ¿Cómo es el desempeño de un pronóstico según el tipo de año (seco, normal o húmedo)?

Para contestar las preguntas se definen los objetivos de este estudio, los cuales corresponden a generar diferentes modelos de predicción de volumen de deshielo a partir de diferentes técnicas como Multiple Linear Regression (MLR), Principal Components Regression (PCR) y Partial Least Squares Regression (PLSR). Luego, a partir de la selección del mejor modelo comparando los valores del Error Cuadrático Medio (MSE) se determina el tiempo de adelanto y la capacidad de predicción de años secos, normales y húmedos.

2. ZONA DE ESTUDIO

Para resolver las preguntas planteadas se analizan dos cuencas pertenecientes a la Región del Maule: Maule en Armerillo y Achibueno en la Recova (Figura 1). Se utilizan aquellas estaciones hidro-meteorológicas pertenecientes a la DGA cuyos datos se encuentran disponibles en línea, para de este modo, permitir un uso operacional de la herramienta. En la Tabla 1 se muestra un resumen de la información de las estaciones y las variables utilizadas para realizar el pronóstico.

Para el caso de Maule en Armerillo se utilizó una serie reconstruida en régimen natural y en el caso de Achibueno, se considera esta cuenca en régimen prácticamente natural.

Tabla 1. Estaciones hidro-meteorológicas consideradas en este estudio. Pp: Precipitación, T: Temperatura, EA: Equivalente en agua de nieve (SWE).

Código Figura	Estación	Código BNA	UTM N	UTM E	Variable utilizada
M1	Río Maule en Armerillo	07321002-K	6046528	308442	Q y Pp
M2	Río Achibueno en la Recova	07354002-K	6013026	279757	Q
M3	Río Ancoa en el Morro	07355002-5	6023788	292742	Pp
P1	Lo Aguirre	07301000-4	6018203	358258	EA Máximo

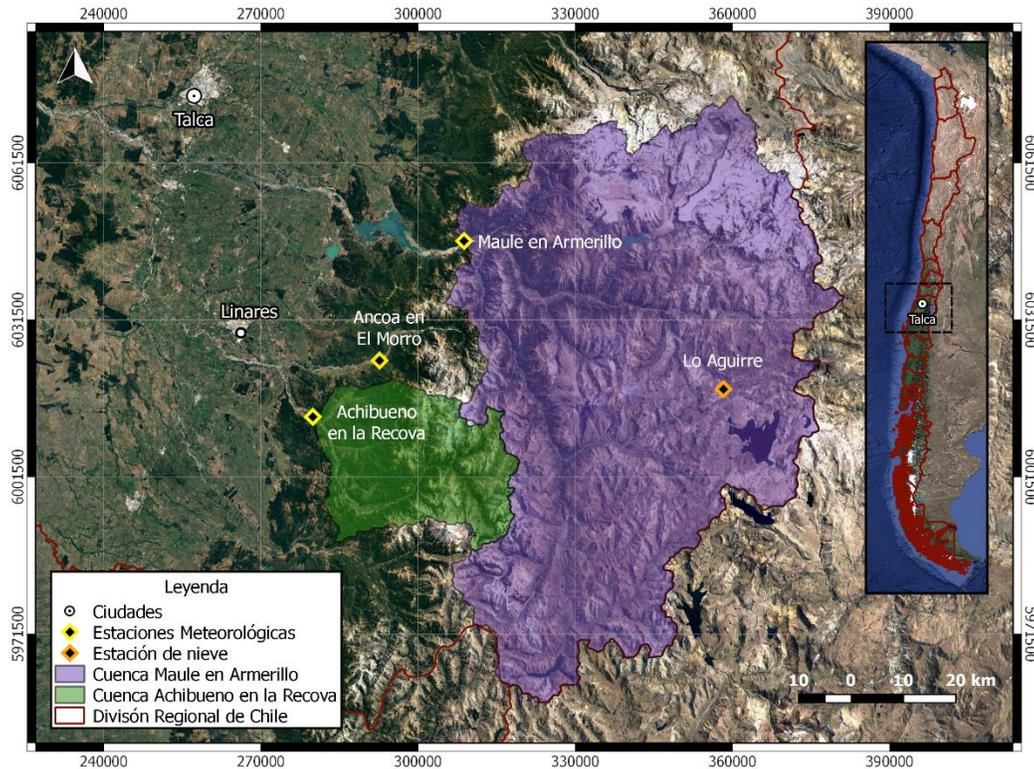


Figura 1. Ubicación de las cuencas de estudio y estaciones meteorológicas utilizadas para la generación del pronóstico. WGS84 zona 19S.

3. METODOLOGÍA

La metodología empleada en este estudio consiste en utilizar métodos estadísticos que busquen explicar un volumen de deshielo (Y), el cual se entiende como la acumulación de caudal entre los meses de septiembre y marzo, teniendo en consideración ciertos predictores que pueden ser observados en invierno (X). Tanto los predictores empleados como las técnicas estadísticas se describen en las siguientes secciones:

Predictores utilizados

Cada modelo parte del siguiente set potencial de predictores obtenidos para el periodo entre los años 2001 y 2018:

- Caudal medio de abril hasta agosto (QmAMJJA)
- Caudal medio de junio hasta agosto (QmJJA)

- Precipitación acumulada de abril hasta agosto (PpAMJJA)
- Precipitación acumulada de junio hasta agosto (PpJJA)
- Equivalente en agua de nieve máximo de la temporada (SWE_{max})
- Cobertura nival fraccional máxima en el mes de agosto (fSCA_{max_A})
- Cobertura nival fraccional media en el mes de agosto (fSCA_{med_A})
- Cobertura nival fraccional del 31 de agosto (fSCA_31A)
- Cobertura nival fraccional media en los meses de junio hasta agosto (fSCA_{med_JJA})

La obtención de la cobertura nival fraccional se realiza utilizando el producto Snow Cover de MODIS (MOD10 y MYD10), debido a que dichas imágenes se encuentran afectadas por la cobertura de nubes, se implementa una corrección de nubes propuesta por Lagos et al. (2015) que consiste en una adaptación de la metodología propuesta por Gafurov y Bárdossy (2009), presenta un resumen del procesamiento en la figura 2. Debido a que sólo se tienen imágenes MODIS desde el 2001, el análisis se realiza para el periodo (2001/02-2018/19).

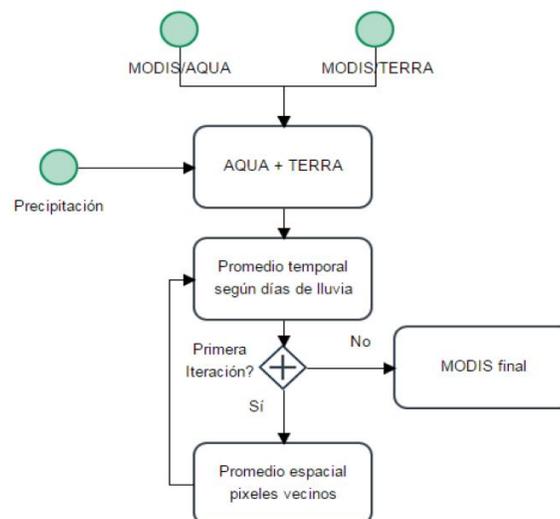


Figura 2. Procedimiento de corrección de subes de los satélites AQUA y TERRA. (Fuente: Lagos et al, 2015).

Metodologías para pronóstico volumétrico

Los métodos estadísticos usados en este estudio corresponden a Regresión lineal múltiple, Regresión de Componentes Principales y Partial Least Square Regression, las que se describen a continuación. Para los tres métodos estadísticos fue llevada a cabo una validación cruzada dejando un valor fuera (LOOCV, por sus siglas en inglés).

Regresión lineal múltiple (MLR)

Este método consiste en ajustar predictores x_i y una variable de interés a predecir y (en este caso el volumen de deshielo) mediante una función de regresión lineal, como se muestra en la Ecuación 1, donde β_i corresponden a los parámetros ajustados para cada predictor y ϵ corresponde al error del modelo.

$$y = \beta_0 + \sum_i \beta_i \cdot x_i + \epsilon \quad (1)$$

Para definir que predictores generan la mejor regresión lineal múltiple se utiliza el método forward de selección de predictores (Wilks, 2011). Este consiste en añadir un predictor a la vez a la regresión y calcular el coeficiente R^2 , seleccionando como predictor a aquel que maximiza el incremento de R^2 , repitiendo el proceso hasta que el R^2 no mejore en un 0,5%.

Principal Components Regression (PCR)

Es una técnica de regresión basada en el análisis de los componentes principales (PCA), Para todo el set de predictores potenciales (X) se calculan los componentes principales y estos componentes se relacionan mediante regresión lineal múltiple con los volúmenes de deshielo de cada cuenca. La ventaja de usar esta técnica es que al usar una transformación ortogonal de posibles variables correlacionadas, es posible reducir el número de predictores iniciales, obteniendo un conjunto de predictores no correlacionados entre sí.

Para definir los componentes principales se utiliza la totalidad de los predictores propuestos y se identifica el número de componentes necesarios para explicar al menos el 80% de la varianza del volumen con el menor error cuadrático medio.

Partial Least Square Regression (PLSR)

Al igual que el PCR es un método estadístico que se basa en el análisis de componentes principales, pero en vez de encontrar hiperplanos de máxima varianza entre una variable de respuesta y los predictores, encuentra un modelo de regresión lineal al proyectar las variables predichas y observadas a un nuevo espacio.

Luego, el modelo se ejecuta con todos los predictores propuestos y se identifica el número de componentes principales necesarias de la misma forma que se escoge para PCR.

Evaluación y selección de modelos de pronóstico

Para comparar los resultados de los diferentes modelos y definir el mejor se calcula el *Root Mean Square Error* o RMSE. Luego aquel modelo con el menor valor de RMSE se considera el mejor.

Selección de los años del tipo seco, normal y húmedo

Para determinar que años del periodo de tiempo estudiado (2001-2018) son de carácter seco, normal y húmedo se calcula el índice normalizado de precipitación o SPI (Mckee et al, 1993), que se basa en la probabilidad de precipitación para cualquier escala temporal. Teniendo en cuenta la precipitación observada, la probabilidad de precipitación se transforma en un índice que define las distintas intensidades de la sequía. En la tabla 2 se presentan los valores del índice normalizado de precipitación.

Tabla 2. Valores del índice normalizado de precipitación. (Fuente: OMM 2012)

2,0 y más	Extremadamente húmedo
1,5 a 1,99	Muy húmedo
1,0 a 1,49	Moderadamente húmedo
-0,99 a 0,99	Normal o aproximadamente normal
-1,0 a -1,49	Moderadamente seco
-1,5 a -1,99	Severamente seco
-2 y menos	Extremadamente seco

Utilizando la serie de precipitaciones de la estación Maule en Armerillo para el periodo de tiempo entre los años 2001 al 2018, se define el SPI mensual para una escala de tiempo de 12 meses para cada año. Luego, aquel año con SPI promedio anual mayor a 2 se considera el año húmedo, aquel con promedio menor 0,99 y mayor a -0,99 el año normal y aquel menor a -2 el año seco.

Tiempo de adelanto

El pronóstico de volumen de deshielo es realizado originalmente a fines del mes de agosto, por lo que se desea evaluar cómo funcionan las actualizaciones del pronóstico a fine de septiembre y a fines de octubre, incorporando los predictores de la **Tabla 3**.

Tabla 3 Predictores añadidos para evaluar tiempo de adelanto.

Predictores añadidos pronóstico Oct-Mar	Predictores añadidos pronóstico Nov-Mar
- Caudal medio de septiembre (QmS)	- Predictores añadidos en Septiembre
- Precipitación acumulada en septiembre (PpS)	- Caudal medio de octubre (QmO)
- Cobertura nival fraccional media en septiembre (fSCAmed_S)	- Precipitación acumulada en octubre (PpO)
	- Cobertura nival fraccional media en el mes de octubre (fSCAmed_O)

4. RESULTADOS

Características finales de cada modelo

Utilizando el método *forward selection* (Wilks, 2010), se obtuvo que el modelo de regresión lineal múltiple para cada cuenca queda definido por los predictores expuestos en la Tabla 3.

Tabla 4. Predictores seleccionado para MLR por método forward.

Cuenca Maule en Armerillo	Cuenca Achibueno en la Recova
PpAMJJA	PpJJA
QmJJA	fSCAmed_JJA
mSWE	fSCAmax_A
fSCA_31A	PpAMJJA
fSCAmed_JJA	QmJJA
	fSCA_31A
	mSWE

Para los modelos PCR y PLSR el número de componentes definidos para cada cuenca se presentan en la **Tabla 5**.

Tabla 5. Número de componentes para cada método utilizado.

Modelo	Maule en Armerillo	Achibueno en la Recova
PCR	5	5
PLSR	4	6

A partir de los resultados se decide crear un modelo de tipo MLR con cuatro predictores, seleccionando aquellos que tengan mayor correlación con el volumen de deshielo para cada categoría (precipitación, caudal y fSCA). Los predictores utilizados fueron PpAMJJA, fSCAmed_JJA, mSWE y QmJJA.

Elección del modelo

Según los resultados del RMSE para cada método estadístico, el mejor modelo para Maule en Armerillo es el PLSR (**Tabla 6**), mientras que para Achibueno en la Recova es el MLR Forward y en segundo lugar el PLSR, por lo que para trabajar con un modelo general en la zona de estudio se escogió PLSR.

Tabla 6. Errores máximos obtenidos para las metodologías utilizadas en las cuencas Maule en Armerillo y Achibueno en la Recova.

Modelo	RMSE en Maule en Armerillo (Mill. m ³)	RMSE en Achibueno en la Recova (Mill. m ³)
PCR	565	63
PLSR	497	55
MLR Forward	503	52
MLR Propuesto	584	95

En las Figuras 3 y 4 se presenta la predicción del volumen de deshielo para el periodo entre los años 2001-2018 para cada cuenca y modelo, como si se estuviera realizando un pronóstico en cada una de esas temporadas.

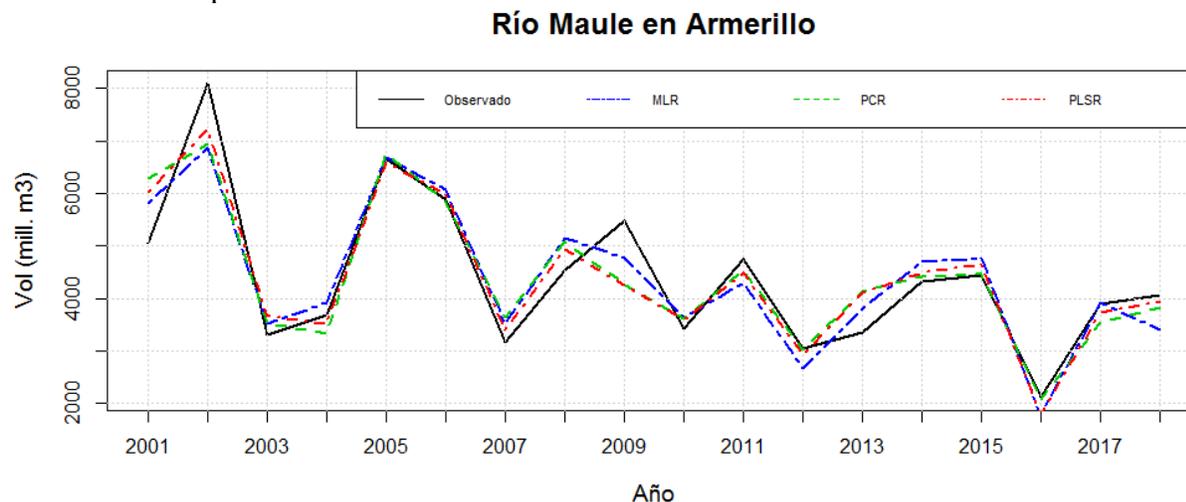


Figura 3. Volúmenes de deshielo obtenidos para Maule en Armerillo obtenidos mediante distintas metodologías.

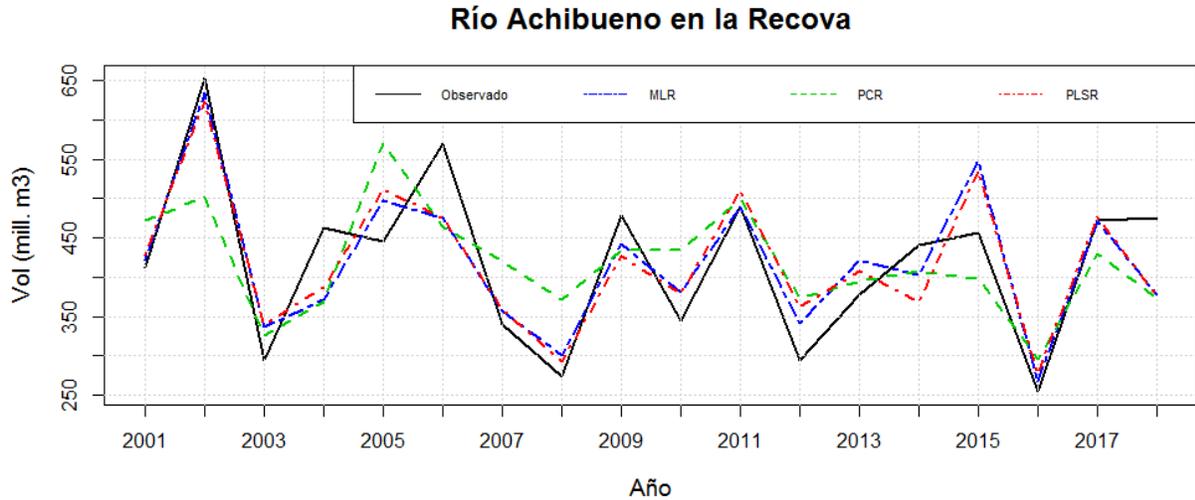


Figura 4. Volúmenes de deshielo obtenidos para Achibueno en la Recova obtenidos mediante distintas metodologías.

Tiempo de adelanto de la predicción

Realizando dos modelos, uno con la información de septiembre y otro con la información de septiembre y octubre se obtiene el volumen de deshielo para ambas cuencas. Los resultados de la incorporación de esta información se presentan de forma resumida en la Figura 5. Es posible apreciar que el valor de la mediana del error se acerca en los distintos tiempos de pronóstico al cero, sin embargo, la dispersión se incrementa.

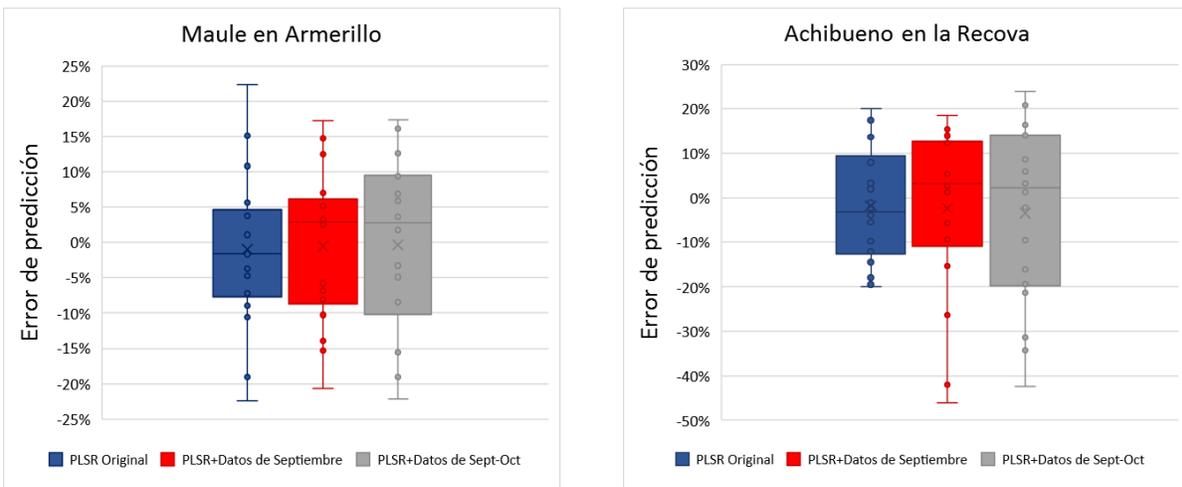


Figura 5. Errores de predicción con distintos tiempos de adelanto para Maule en Armerillo y Achibueno en la Recova. Cada punto representa el error de predicción de un año.

Utilizando los errores en valor absoluto, se obtiene que el error promedio del pronóstico de Agosto realizado para Maule en Armerillo es un 8%, mientras que para Achibueno en la Recova es un 11%. Para ambas cuencas este error porcentual promedio aumenta en una unidad para los pronósticos realizados en los meses siguientes.

Evaluación del modelo según el tipo de año

Utilizando la metodología SPI se obtiene que del periodo de tiempo entre el año 2001-2018 el año seco corresponde al año 2016, el normal al año 2009 y el húmedo al año 2006.

Para analizar el desempeño del modelo PLSR en los años seco, normal y húmedo se realizan tres ajustes PLSR sin considerar el año que desea pronosticarse. En las tablas 6 y 7 se resumen los resultados en ambas cuencas.

Tabla 6. Resumen de volúmenes obtenidos con PLSR y el error de predicción en cada tipo de año en Maule en Armerillo.

Año	Volumen Observado (Mill. m3)	Volumen PLSR (Mill. m3)	Error predicción (%)
Seco	2117	1796	-15
Húmedo	5882	5967	1
Normal	5479	4618	-16

Tabla 7. Resumen de volúmenes obtenidos con el modelo y el error de predicción en cada tipo de año en Achibueno en la Recova.

Año	Volumen Observado (Mill. m3)	Volumen PLSR (Mill. m3)	Error predicción (%)
Seco	254	268	6
Húmedo	570	467	-18
Normal	478	440	-8

5. DISCUSIÓN

Se observa de los resultados del método *forward selection*, que en la cuenca Maule-Armerillo se necesitan de menos predictores que en Achibueno en la Recova para ajustar el modelo de regresión lineal múltiple (5 predictores Maule v/s 7 predictores Achibueno). Analizando los resultados de los modelos PLSR y PCR se vuelve a observar que en Achibueno en la Recova se necesitan más componentes para generar los ajustes que en Maule en Armerillo.

Se observa del cálculo del tiempo de adelanto, que agregar información de septiembre y octubre entrega resultados similares que solo utilizar información hasta agosto. Adicionalmente, se observa que en ambos casos el volumen de deshielo es subestimado en los años húmedos. Para ambas cuencas existe un aumento de la varianza del error al incorporar la información de los meses siguientes.

Evaluando el desempeño del modelo en los años seco, normal y húmedo se observa que el funcionamiento de la metodología es diferente en ambas cuencas, en ambos casos se subestima el volumen del año normal, mientras que los años seco y húmedo son subestimados en una cuenca y acertados (<5%) en la otra.

El error promedio de predicción de los pronósticos de volumen de deshielo realizados por la DGA en la cuenca Maule en Armerillo durante los años 2009-2019 es de 11.2%, sobreestimando en un 42% el volumen de deshielo de la temporada 2016-2017, que dado su SPI se caracteriza como una temporada severamente seca según la Tabla 2. Al comparar estos resultados con los obtenidos con el modelo se observa que el error promedio de predicción es menor que el de la DGA, siendo equivalente a 8.4%. Adicionalmente, para la temporada 2016-2017 el modelo propuesto subestima en un 15% el volumen de deshielo, por lo que se observan mejoras en los pronósticos al utilizarlo.

6. CONCLUSIONES

El mejor modelo para Maule en Armerillo es el PLSR, mientras que para Achibueno en la Recova es el MLR Forward, sin embargo, el modelo PLSR funciona de manera muy similar por lo que para realizar un análisis de toda la zona de estudio se considera el PLSR como el mejor modelo.

Sobre el desempeño del modelo en la predicción de los años seco, normal y húmedo se concluye que ocurre una subestimación del año normal, posiblemente por el efecto que genera el periodo en estudio (2001-2018), el cual ha sido caracterizado como una mega sequía (Garreaud et al, 2017). Por otra parte, el año húmedo es subestimado en Achibueno en la Recova, es posible que esto se deba a que este año aporta información considerable considerando los pocos años húmedos del periodo. Esto podría ser corregido aumentando el periodo de tiempo estudiado para tener un mayor número de años del tipo húmedo que aporten información al modelo.

Al evaluar los pronósticos para distintos tiempos de inicialización, se observa que para ambos casos la información de octubre no mejora el modelo y lo empeora. Sin embargo, la varianza dependiendo de la cuenca aumenta o disminuye al agregar información de los meses de septiembre y octubre. Una posible justificación para este empeoramiento es la elección de predictores incorporados en la actualización del pronóstico (caudal medio mensual, precipitación acumulada y área cubierta por nieve).

Dada la naturaleza de las cuencas nival y pluvio-nival analizadas, se concluye que estos métodos estadísticos son capaces de generar pronósticos de volúmenes robustos, los que pueden ser implementados de forma operacional por las distintas asociaciones de usuarios de agua que lo requieran, puesto que entregan en promedio errores inferiores a los entregados por DGA.

REFERENCIAS

División Hidrológica. (2018). *Pronóstico de caudales de deshielo. Temporada de riego 2018-2019*. Santiago: Dirección General de Aguas.

Gafurov, A., & Bárdossy, A. (2009). Cloud removal methodology from MODIS snow cover product. *Hydrology and Earth System Sciences*, 13(7), 1361-1373.

Garreaud, R. D., Alvarez-Garretón, C., Barichivich, J., Boisier, J. P., Christie, D., Galleguillos, M., ... & Zambrano-Bigiarini, M. (2017). The 2010-2015 megadrought in central Chile: impacts on regional hydroclimate and vegetation. *Hydrology & Earth System Sciences*, 21(12).

Lagos, M.; Cornwell, E.; McPhee, J. y Vargas, X. La Cobertura Nival Fraccional Como Predictor De Volumen Estacional De Deshielo. XXII Congreso Chileno De Ingeniería Hidráulica. Pontificia Universidad Católica De Chile. 2015.

Mendoza, P. A., B. Rajagopalan, M. P. Clark, G. Cortes, and J. McPhee (2014), A robust multimodel framework for ensemble seasonal hydroclimatic forecasts, *Water Resour. Res.*, 50, 6030–6052, doi:10.1002/2014WR015426.

Organización Meteorológica Mundial. (2012). *índice normalizado de precipitación. Guía del usuario*. Ginebra: OMM.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). Academic press.