Pronóstico de volúmenes y caudales estaciones para cuencas de Chile Central

Metodología

Una de las nociones básicas que debemos entender es la definición de un año hidrológico. El año hidrológico generalmente comienza en otoño y termina en verano del año siguiente. Esta traslación del año calendario tiene la intención de considerar todas las posibles fuentes de acumulación y derretimiento dentro de un periodo anual. Para el caso específico de Chile, la convención es que el año hidrológico comienza el 1 de abril de cada año y termina el 31 de marzo del año siguiente.

El uso de los pronósticos de caudales suele ser más intensivo después de septiembre, sin embargo, dependiendo de la industria, por ejemplo, en la agricultura hay decisiones del tipo de cultivo o la extensión de las plantaciones que consideran la cantidad de agua disponible durante el periodo de crecimiento.

El pronóstico de caudales funciona a partir de múltiples entradas o predictores que son agregadas temporalmente y/o espacialmente para cada una de las cuencas en estudio.

Por ejemplo, uno de los predictores podría ser la precipitación mensual (variable), acumulada para el año hidrológico en estudio (agregación temporal), esta precipitación proviene de un modelo de reanálisis meteorológico grillado que debe ser promediado para obtener un solo valor para la superficie de la cuenca (agregación espacial).

El pronóstico se basa en la relación que existe entre los varios predictores y el predicando o variable objetivo. Para nuestro caso, el predicando corresponde al volumen estacional de septiembre a marzo. El volumen se distribuye ocupando los coeficientes de Pardé (caudal medio mensual normalizado por el caudal medio de la temporada) de los años más cercanos basado en la distancia euclidiana de los predictores.

Predictores

Modelo estadístico

El modelo estadístico consiste en una regresión lineal múltiple de los diferentes predictores.

Los predictores utilizados se normalizan utilizando el método centrar/escalar que consiste en restar a cada individuo de un predictor el promedio de la muestra (centrar) y luego dividir por la desviación estándar (escalar). Tanto el promedio como la desviación se calculan ocupando los años de la muestra de entrenamiento que corresponden a los años hidrológicos 1980 a 2020. Existen otras alternativas de normalización como minmax (restar el mínimo y dividir por el máximo), sin embargo, el tipo de normalización no suele ser influyente en el resultado final de la regresión.

La utilización de regresiones lineales requiere cumplir supuestos. Para verificar la normalidad del predicando (volumen) se aplicó el test de Shapiro-Wilks (SW). Este test tiene como hipótesis nula que la variable viene de una distribución normal. Por lo tanto, si el p-value es menor al nivel alfa (generalmente 0.05) la hipótesis nula se rechaza y por lo tanto la distribución no sería normal. Cabe mencionar que este tipo de metodología no identifica la distribución sino que aprueba o rechaza la hipótesis de normalidad. En caso de ser rechazada la hipótesis, si la variable en el histograma se encuentra sesgada a la derecha o la izquierda, la muestra se puede transformar aplicando alguna transformación de orden de magnitud, por ejemplo, logaritmo natural o raíz cuadrada. Para los volúmenes se observaron principalmente volúmenes sesgados (skew) a la derecha, es decir, una mayor frecuencia de caudales bajos comparado con medios y altos. Por lo que se aplicó, si la hipótesis nula se rechazaba, una transformación logarítmica y nuevamente se realizó el test de Shapiro-Wilks a la nueva variable. La variable (volumen o logaritmo del volumen) que pase el test de Shapiro-Wilks se utiliza como predicando en la regresión lineal múltiple.

La regresión lineal múltiple se basa en la optimización computacional de los