a) Objetivo del hito:

El objetivo del hito es generar un pronóstico hibrido para 45 cuencas de Chile Central, con diferentes tiempos de emisión a largo del año mediante un proceso automatizado.

b) Variables críticas estudiadas y rangos alcanzados:

Las variables críticas incluyen los índices climáticos mensuales estándar (SOI, ONI, PDO, NIÑO1,2,3,4), los datos horarios de reanálisis meteorológicos ERA5 de precipitación y temperatura, y las variables de estado de almacenamiento del modelo hidrológico como el manto nival, 2 estanques de agua subsuperficial y humedad del suelo. Como las bases de datos son de servicios externos (NOAA, ECMWF), los formatos y permisos de acceso pueden cambiar con el tiempo. El servicio de ERA5 tienen un retraso de aproximadamente 6 días respecto al presente, lo que debe considerarse en la ejecución.

Se probó el sistema retrospectivamente entre los años 1981 y 2019, incluyendo emisiones entre el 1 de mayo y el 1 de marzo (11 meses de emisión). Las métricas de evaluación nos indican que los primeros meses del año hidrológico (abril, mayo, junio) el pronóstico con los predictores actuales tiene coeficientes de correlaciones menores a 0.5 y el error probabilístico (CRPSS) medio es mayor 0.5 desde 1 de agosto. Por lo tanto, se puede ejecutar el pronóstico desde el 1 de agosto con mayor seguridad y precisión, los resultados en julio son variados y suelen ser más precisos en años húmedos (probabilidad de excedencia 25%).

c) Breve descripción de los experimentos, tratamientos y/o pruebas realizadas:

Se realizaron varias etapas de procesamiento y análisis de datos. Primero, se descargaron índices climáticos y datos meteorológicos. Estos datos se preprocesaron, escalando a nivel de cuenca y corrigiendo el sesgo (error sistemático) ocupando Shaake-Shuffle y DQM. Luego, se ejecuta un modelo hidrológico conceptual TUW utilizando datos diarios corregidos de precipitación y temperatura entre 1981 y el presente. A partir de estos datos, se ejecuta un modelo estadístico con predictores independientes descartando variables correlacionadas y seleccionando la mejor combinación de predictores de acuerdo con el criterio de Akaike (que minimiza el número de predictores con menor error). Finalmente, se utiliza el modelo de similitud KNN para distribuir el volumen en caudales medios mensuales dada la forma (caudal normalizado por el volumen) de años similares y se exportaran 2 archivos con los resultados para el año objetivo en formato csv. Este formato ha sido probado exitosamente para ser importado y mostrados en la página web [www.pronosticoscaudales.cl](http://www.pronosticoscaudales.cl) sección administración.

Para el preproceso meteorológico (agrupación mensual, promedio espacial, formateo) se ocupa el software CDO (Climate Data Operators) y para el resto de los procesos se ocupan códigos R (versión superior a 4.0). CDO se ejecuta internamente en R ocupando la línea de comando.

La ejecución de los procesos ha sido probada en estructuras Linux (Ubuntu) y Unix (MacOS), se han adaptado los códigos para que funcionen adecuadamente en ambos sistemas. En Windows la instalación de CDO es más inestable y puede variar dependiendo de la versión, desde Windows 10 se incluye una versión de Linux dentro del sistema por lo que se podría instalar ocupando la línea de comando (no probado).

La transferencia de los archivos a la página web no se encuentra automatizada dado que depende de la infraestructura del receptor (DGA). Por ejemplo, se puede correr el algoritmo con un cronometro en un servidor local que también contiene la página web o se pueden llevar todos los archivos a la nube y ejecutar remotamente desde internet. En general, se recomienda revisar los archivos antes ser publicados en la plataforma web para una mayor confiabilidad.

d) Principales resultados logrados y conclusiones:

Se logró la automatización del proceso de generación de pronósticos híbridos para 45 cuencas entre latitudes 27 ˚ S y 37 ˚ S. Dado los resultados de errores determinísticos y estadísticos se recomienda la ejecución cada mes comenzando desde agosto y finalizando en febrero, es decir 7 meses de emisión. La metodología permitió manejar un gran volumen de datos y los códigos permiten dar flexibilidad en la ejecución para mejorar o reparar cada proceso.

Anexo: Automatización del algoritmo del pronóstico

La automatización corresponde a la ejecución no supervisada de los varios procesos para generar el pronóstico.

Los procesos automatizados corresponden a:

1. Descarga de índices climáticos: se utilizan índices mensuales estándar descargados de los servicios del NOAA.

2. Adquisición de datos meteorológicos: Descarga de las bases de datos de los reanálisis meteorológicos ERA5 para Chile Central actualizados al presente desde la plataforma de ECMWF.

3. Preproceso meteorológico: escalamiento a nivel de cuenca, agregación diaria y corrección del sesgo (error sistemático) a partir de CR2MET.

4. Ejecución de modelo hidrológico: ocupando las series de tiempo históricas de precipitación y temperatura se ejecuta el modelo TUW, consiguiendo 4 variables de estado de los almacenamientos (manto nival, 2 estanques de agua subsuperficial y humedad del suelo).

5. Ejecución del modelo estadístico: se ocupan los almacenamientos hidrológicos e índices climáticos para generar predictores independientes en una regresión lineal múltiple para cada cuenca y fecha de emisión. La variable objetivo es el volumen estacional acumulada de septiembre a marzo, para emisiones antes de octubre.

6. Distribución de caudales: se aplica el modelo de similitud KNN para encontrar los 6 años más similares a los predictores actuales. Se distribuye el volumen estaciones ocupando coeficientes de Pardé de los años similares.

7. Exportación de datos: los datos se agrupan en dos archivos csv (pronóstico volumen y pronóstico caudales mensuales) que contienen los resultados para las 45 cuencas estudiadas y se da formato para importarlos en la plataforma web.

El algoritmo del pronóstico se encuentra guardado actualmente en un repositorio privado en Github.com. Dentro de Github se puede automatizar la ejecución periódica de algunos procesos, pero debido a que se requiere tanto R como CDO, se recomienda instalar inicialmente de manera local con un programador temporal. Una solución más definitiva puede ser ejecutar todos los códigos en la nube en servicios como AWS (Amazon) o Azure (Microsoft), donde se puede tener una maquina virtual que utiliza Linux/UNIX.

Para ejecutar el algoritmo, se recomienda descargar después del 6 de cada mes para obtener los datos completos del mes anterior. Por ejemplo, para la emisión del 1 de mayo del 2023 se debe esperar hasta el 6 de mayo para ejecutar el algoritmo con todos los datos del mes de abril debido al tiempo de retraso de las bases de datos externas (ECMWF y NOAA).

El código central se llama “CORRER\_PRONOSTICO.R” y se puede correr tanto manualmente ocupando Rstudio o por línea de comando ocupando Rscript. Para la primera ejecución se requiere input de credenciales del usuario de descarga por parte del usuario.

A screenshot of a computer

Description automatically generated with medium confidence

Figura 1. Lista de archivos para correr el algoritmo del pronóstico.

La descarga de la carpeta a un computador local se puede hacer manual con la opción Code > Download ZIP.

A screenshot of a computer

Description automatically generated

Figura 2. Botón de descarga del repositorio

El código “CORRER\_PRONOSTICO.R” ejecuta una serie de funciones para asegurar que existan las dependencias de paquetes y librerías necesarias. La descarga de librerías ha funcionado satisfactoriamente en UNIX/Linux, algunos paquetes pueden requerir instalación manual para la primera ejecución dependiendo del computador y sistema operativo. El algoritmo no ha sido probado en Windows.

El software R (versión probada 4.2.2) se puede conseguir de los repositorios oficiales del CRAN (<https://cran.r-project.org/bin/>) y Rstudio (versión 2023.03.1+446) que corresponde a la interfaz gráfica más utilizada con R se puede descargar desde Posit (<https://posit.co/products/open-source/rstudio/>).

Antes de ejecutar R, se debe instalar el software CDO (Climate Data Operators) que se puede conseguir usando los pasos de la pagina oficial (<https://code.mpimet.mpg.de/projects/cdo>). La versión probada en nuestros computadores es v2.2.0 (última disponible al 31 mayo 2023).

También se requiere acceso a una cuenta de Copernicus (gratuita) para la descarga de datos, la que se puede conseguir en <https://cds.climate.copernicus.eu/user/login>. La página principal del algoritmo (función descargar\_era5) pide el usuario (CDS\_user) y clave (CDS\_key). En algunos computadores (Linux) también se pide instalar un archivo keyring.

La ejecución del pronóstico implica una serie de pasos o sub-funciones de dependencias que se muestran en la siguiente figura 3. En general, cada función tiene parámetros para modificar una parte del algoritmo, en “CORRER\_PRONOSTICOS.R” se ha configurado los parámetros principales para una persona con entendimiento básico de R.

A picture containing diagram, line, origami, pattern

Description automatically generated

Figura 3. Dependencias directas de funciones del modelo de pronósticos operativo.

Las carpetas del repositorio se dividen en varias secciones:

* La carpeta “base” contiene los archivos más globales del algoritmo en todos sus procesos y puede entenderse como un nuevo paquete o un conjunto de clases.
* La carpeta “utils” contiene códigos derivados de la “base” y que se ocupan para un uso específico, como correr una cuenca, si se quiere estudiar los resultados finales o exportar solo una parte de los resultados o sub-procesos.
* La carpeta “data\_input” contiene las múltiples entradas del modelo estadístico y las entradas del modelo hidrológico. Temporalmente también es donde se guardan los archivos descargados de ERA5 y índices climáticos.
* La carpeta “data\_output” contiene los resultados de los pronósticos. Se pueden encontrar las figuras de pronósticos y comparaciones por cuenca, los archivos en formato .csv para subir a la pagina web ([www.pronosticocaudales.cl](http://www.pronosticocaudales.cl)), los resultados detallados de cada proceso de ejecución (data de entrenamiento, pronósticos volúmenes, pronósticos caudales, métricas de evaluación, modelo de regresión, predictores, etc.)
* La carpeta “reportes” contiene algunos reportes realizar para la ejecución de proyecto.

El algoritmo completo para una fecha de emisión (ejecutando “CORRER\_PRONOSTICO.R”), cuando todas las librerías se encuentras instaladas, se demoró lo que muestra la tabla 1.

Tabla 1. Especificaciones de los computadores utilizados para probar el algoritmo de pronósticos.

|  |  |  |  |
| --- | --- | --- | --- |
| Sistema Operativo | Ram | Procesador | Tiempo ejecución |
| MacOS 13.3.1 | 16 GB | Apple M1 pro | 6 minutos |
| Linux Ubuntu | 32 GB | Intel | 15 minutos |

El tiempo de ejecución puede variar tanto por el rendimiento del procesador como eventos externos. Por ejemplo, la base de datos de ERA5 puede tener una cola de espera de hasta 1 hora, generalmente la descarga de archivos es casi instantánea, aunque se debe considerar que la descarga podría tomar más tiempo.

Para ejecutar el algoritmo el 6 de cada mes, en ambiente UNIX se puede ejecutar el comando “crontab -e” donde en el archivo abierto se puede pegar el código “0 0 6 \* \* Rscript ruta\_al\_archivo /CORRER\_PRONOSTICOS.R”. En Windows se podría utilizar “Task Scheduler”.

El script principal “CORRER\_PRONOSTICO.R” se adjunta a continuación.

A screenshot of a computer

Description automatically generated with medium confidence