

Reducción de escala temporal para series de recursos primarios

Informe técnico - Fondo sectorial de energía 173061

1 Introducción

El downscaling temporal se refiere al proceso de aumentar la resolución temporal de datos climáticos de baja resolución, como promedios mensuales o diarios, a escalas temporales más finas, como intervalos horarios o sub-horarios. Este proceso es crucial para aplicaciones que requieren datos climáticos de alta frecuencia, como modelado hidrológico, planificación de energía o análisis de eventos extremos.

El aspecto clave del downscaling temporal consiste en desarrollar relaciones empíricas entre predictores de gran escala, como salidas diarias o mensuales de modelos climáticos, y patrones temporales más finos observados en datos históricos. Estos métodos se basan en las propiedades estadísticas de datos observados de alta frecuencia para generar valores sub-diarios u horarios a partir de entradas de baja resolución.

En este informe se da cuenta de los predictores utilizados para reducir la escala temporal de variables de recursos primarios. El reporte se estructura de la siguiente manera. En la próxima sección se describen los datos utilizados en el estudio, las fuentes de datos, las variables consideradas, los períodos y la ventana geográfica utilizada para generar la información. Luego las secciones Section 3 y Section 4 presentan la metodología para obtener series de escala reducida y evaluar los predictores. La Section 5 se comentan los principales resultados de la evaluación de predictores estadísticos. El trabajo cierra con la Section 6 donde se resumen las principales conclusiones y se delinean posibles pasos para continuar en esta línea de investigación.

2 Datos utilizados en el estudio

2.1 Fuentes de datos

Los dos principales fuentes de datos que utilizamos son CMIP6 ([Eyring et al., 2016](#)), y Re-análisis v5 de ECMWF, ERA5 ([Copernicus Climate Change Service \(C3S\), 2017](#)). CMIP6 es un marco internacional coordinado para experimentos de modelado climático que proporciona simulaciones estandarizadas del sistema climático terrestre. Incluye proyecciones de estados climáticos pasados, presentes y futuros utilizando Modelos de Circulación General (GCMs). CMIP6 facilita la comparación de modelos entre instituciones y escenarios, permitiendo una evaluación robusta y cuantificación de incertidumbres en las proyecciones climáticas. Integra experimentos como ejecuciones históricas, escenarios futuros basados en Trayectorias Socioeconómicas Compartidas (SSPs) y análisis de sensibilidad. Estos conjuntos de datos se utilizan comúnmente para estudiar la variabilidad climática, evaluar sesgos de modelos e informar decisiones políticas sobre mitigación y adaptación climática. ERA5 es un conjunto de datos de reanálisis producido por el Centro Europeo de Pronósticos Meteorológicos a Plazo Medio (ECMWF). Integra datos observacionales en un modelo numérico de predicción meteorológica para generar un registro espaciotemporal consistente y de alta resolución de los estados de la atmósfera, el océano y la superficie terrestre. ERA5 proporciona datos horarios desde 1950 hasta el presente, con una resolución espacial de ~ 31 km y múltiples niveles verticales. Sus productos incluyen variables como temperatura, humedad, viento y precipitación.

Desde una perspectiva de aprendizaje automático, CMIP6 proporciona datos de alta dimensionalidad y multi-temporales adecuados para tareas de aprendizaje supervisado (por ejemplo, downscaling), aprendizaje no supervisado (por ejemplo, agrupamiento de patrones climáticos) y estudios de interpretabilidad. ERA5 sirve como un conjunto de datos de alta calidad, restringido observacionalmente, para entrenar y validar modelos, crear datos de referencia para downscaling y evaluar el rendimiento de las salidas de GCM (como las de CMIP6).

2.2 Variables utilizadas

Los datos utilizados en el trabajo están restringidos a la región geográfica delimitada como Latitud entre $[-30, -35]$ y Longitud entre $[-59, -53]$. En el caso de CMIP6 la resolución espacial depende del modelo, en tanto los datos del Reanálisis v5 de ECMWF (ERA5) tienen una resolución de $0.25^\circ \times 0.25^\circ$. Adicionalmente, se utiliza la variable de precipitaciones en la cuenca del río Uruguay relevante para Salto Grande. En todos los casos las variables son resumidas espacialmente tomando la media espacial dentro de la caja definida por las coordenadas mencionadas, de esta forma el problema de downscaling temporal se reduce a un problema univariante por variable climática.

Con respecto a los periodos de tiempo considerados. Los datos de CMIP6 se obtienen para el período (2015-2100), mientras que los datos de ERA5 se obtienen para el período (1980-2023).

En ambos casos, los datos se dividen en conjuntos de entrenamiento de 1980-2014 para ERA5, un período para evaluación correspondiente a 2015-2023 en ambas bases de datos, y un período de reducción de escala de 2024 a 2100 para los datos de CMIP6.

La Table 1 resume las variables utilizadas en el estudio, para cada variable descargada de CMIP6 se indica la variable correspondiente en ERA5 y la transformación necesaria para que ambas variables sean comparables y la frecuencia temporal en los datos CMIP6.

Finalmente, se incorporan algunas covariables externas. En particular se utiliza el mes y la duración de la luz diurna (en segundos). Adicionalmente, en el downscaling horario se incluyeron la hora, la elevación y el azimut. Para calcular las variables relacionadas con el sol, se utilizó el paquete `pvlib` (Anderson et al., 2023), fijando la posición del observador en el centro de la región geográfica utilizada.

Table 1: Correspondencia entre variables CMIP6 y ERA5

CMIP6	Descripción	ERA5	Transformación	Frecuencia
tas	Temperatura del aire a 2m de la superficie	t2m	-	Diaria
sfcWind	Velocidad del aire horizontal bidimensional cerca de la superficie	u10,v10	$\sqrt{u10^2 + v10^2}$	Diaria
pr	Suma de agua líquida y congelada por unidad de área y tiempo	tp	$tp * 1000$	Diaria
tasmax	Temperatura diaria máxima del aire a 2m de la superficie	mx2t	-	Diaria
tasmin	Temperatura diaria mínima del aire a 2m de la superficie	mn2t	-	Diaria
clt	Fracción del área horizontal ocupada por nubes	tcc	$tcc*100$	Mensual

CMIP6	Descripción	ERA5	Transformación	Frecuencia
rsdt	Radiación solar entrante	tisr	$tisr/60^2$	Mensual
rsds	Flujo radiativo de onda corta	ssrd	$ssrd/60^2$	Mensual

3 Metodología y predictores

Para describir el enfoque de downscaling temporal estadístico, se introduce Y_c como variable en la escala temporal gruesa (por ejemplo, temperatura media diaria o precipitación total diaria). Esta variable se obtiene típicamente de modelos climáticos o conjuntos de datos observacionales. Por otro lado, se denota $Y_f(t)$ como la variable correspondiente en la escala temporal más fina (por ejemplo, temperatura horaria o precipitación en el tiempo t dentro del día). Aquí, $t \in [1, T]$, donde T es el número de sub-períodos dentro del intervalo temporal grueso (por ejemplo, $T = 24$ para datos horarios en un día).

Entonces, el objetivo del downscaling temporal es estimar $Y_f(t)$ a partir de Y_c , potencialmente usando predictores adicionales. Asumimos que la relación entre Y_c e $Y_f(t)$ puede expresarse como:

$$Y_f(t) = F(Y_c, \mathbf{X}(t), \epsilon_t),$$

donde: F es una función determinística o modelo estocástico que captura la relación entre Y_c e $Y_f(t)$, $\mathbf{X}(t)$ representa predictores auxiliares disponibles en resolución temporal más fina (por ejemplo, patrones diurnos, radiación solar o velocidad del viento), y ϵ_t es un componente estocástico que representa variabilidad residual o ruido.

La Figure 1 presenta una visión general del proceso de reducción de escala. Los datos de ERA5 son transformados y agrupados para obtener las variables Y_c y $Y_f(t)$ necesarias para entrenar el modelo de downscaling. Luego, se ajusta el modelo F usando técnicas de aprendizaje automático. Finalmente, el modelo entrenado se utiliza para predecir $Y_f(t)$ a partir de datos de CMIP6 Y_c . Este proceso se repite para cada variable climática de interés, aplicando el mismo marco general.

Se utilizan diferentes modelos estadísticos de aprendizaje automático F , a los que llamaremos **predictores** para reducir la confusión con los modelos de simulación de las variables primarias. Esta terminología destaca que el objetivo es obtener predicciones, las cuales evaluaremos mediante comparaciones con los datos observados más finos.

A modo de ejemplo se describe un predictor *ingenuo* para reducir la escala de una variable diaria a horaria. Este predictor puede construirse asumiendo una forma lineal de F , incluyendo como variables explicativas la variable original y la hora del día, y un término de error aditivo:

$$Y_f(t) = aY_c + b_h + \epsilon_t,$$

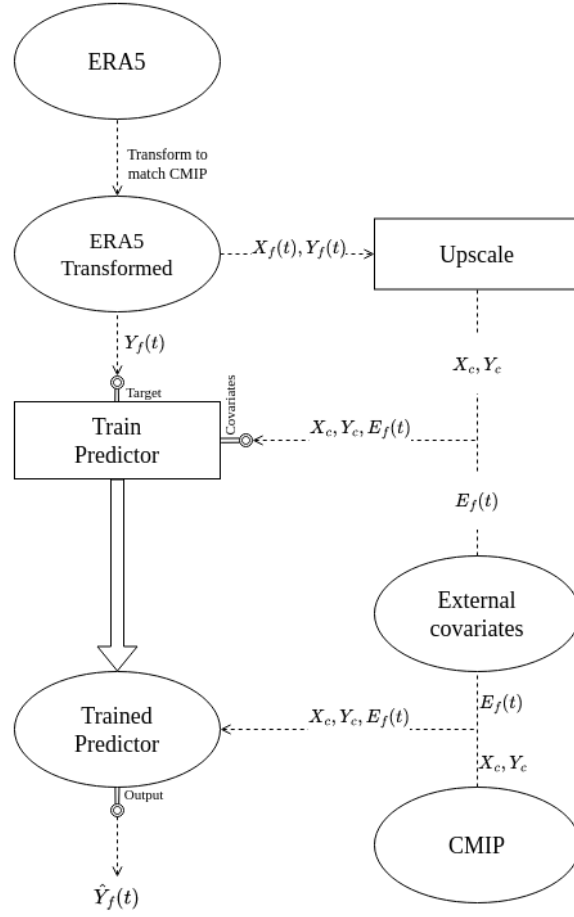


Figure 1: Visión general del proceso de downscaling aplicado para cada variable

donde $\sum_h b_h = 0$ para asegurar que la media diaria de las predicciones coincida con Y_c . Aquí, a es un coeficiente que captura la relación lineal entre la variable diaria y la variable horaria, b_h representa un efecto fijo para cada hora del día (por ejemplo, para capturar el patrón diurno típico), y ϵ_t es un término de error independiente con media cero y varianza finita. Sin dudas, el predictor recién presentado es muy simple para obtener resultados razonables. Este predictor puede ser mejorado en dos grandes aspectos.

En primer lugar, la inclusión de variables explicativas adicionales que ayuden a capturar la variabilidad horaria más allá del patrón promedio diario. En este sentido el término b_h es reemplazado por un vector de covariables $\mathbf{b}^\top \mathbf{X}(t)$ donde $\mathbf{X}(t)$ incluye todas las variables obtenidas de CMIP6, las variables externas descritas más arriba (por ejemplo, duración del día, mes del año, etc.). Además, se utilizan observaciones pasadas y futuras de las covariables. Para transformar variables de diarias a horarias, se utiliza una ventana móvil que incluye las próximas 24 horas y las 24 horas previas como covariables. Para transformaciones de mensuales a diarias, se usan los 28 días antes y después (4 semanas). El modelo lineal resultante es llamado ingenuo o naive (**nv**) en los resultados presentados.

En segundo lugar, el uso de modelos más complejos para F que puedan capturar relaciones no lineales y dependencias temporales en los datos. En particular, se utilizan estimaciones basadas en árboles de potenciación de gradiente extremo (**xgboost**), redes de Memoria a Corto y Largo Plazo (**lstm**) y Redes Neuronales Convolucionales (**cnn**). En todos los casos se utiliza un término de error aditivo para el componente estocástico.

4 Evaluación de downscaling temporal

Esta sección describe las principales estrategias propuestas para evaluar los predictores utilizados para la reducción de escala. Es importante señalar que el objetivo de las series temporales de downscaling de CMIP6 no es pronosticar el clima en cada punto de datos de la resolución más fina, sino comprender el comportamiento futuro del clima a mayor escala. Por lo tanto, muchas métricas de rendimiento tradicionales (por ejemplo, error cuadrático medio) no son útiles, ya que se basan en comparaciones de observaciones individuales. En su lugar, se necesitan métricas que reflejen aspectos distribucionales. Por ejemplo, evaluar qué tan bien se capturan la distribución empírica o los patrones de dependencia mediante la serie a escala reducida.

4.1 Estrategias de evaluación

Al momento de evaluar el downscaling, podemos considerar dos fuentes principales de error. La primera es causada por el proceso de downscaling, ya que el downscaling es la salida de un modelo estadístico, siempre habrá una diferencia entre el valor ajustado por el predictor y el valor observado en un momento dado. Una segunda fuente de error corresponde a una discrepancia entre la fuente los datos de entrenamiento y los datos a ser ajustados con los

predictores entrenados. Es decir, el producto final corresponde a las series de CMIP6 con reducción de escala mediante predictores entrenados con datos de ERA5. La señal de CMIP6 presenta sesgos respecto a la señal regional de ERA5, y adicionalmente, el escenario de cambio climático implícito utilizado para simular los datos de CMIP6 no está presente en los conjuntos de datos observados de ERA5.

El rendimiento del predictor podría evaluarse utilizando dos estrategias de comparación. Primero, una comparación clásica donde los datos de ERA5 se dividen en conjuntos de datos de entrenamiento y prueba. Luego, el rendimiento del predictor puede evaluarse haciendo downscaling de una versión agrupada de ERA5 y comparándola con los datos originales de ERA5. Los valores ajustados y observados son perfectamente comparables, y cualquier tipo de error o mala representación indica un predictor de bajo rendimiento. Nos referimos a esta estrategia como *evaluación pareada* o *evaluación de modelo perfecto*, su principal ventaja es que permite aislar la primera fuente de error y luego enfocarse en el rendimiento del predictor. Sin embargo, este no es el escenario en el que el predictor va a ser utilizado en la práctica, por lo que podría no ser una evaluación completamente verdadera del predictor.

Una segunda estrategia podría ser usar los predictores entrenados (con datos de ERA5) directamente para hacer downscaling de datos de CMIP6, y comparar la serie downscaled ajustada con los datos observados de ERA5 en el período de prueba. Nos referimos a esta estrategia como *evaluación no pareada* o *evaluación de modelo imperfecto*, su principal ventaja es evaluar los predictores de la misma manera en que se usan en el contexto práctico, pero el costo es que hay una confusión de las fuentes de error del predictor. Se podría argumentar que los sesgos en CMIP6 respecto a los conjuntos de datos de ERA5 son parte del desafío de este problema de predicción y deben ser tomados en cuenta cuando se evalúan los predictores. Por otro lado, los efectos del cambio climático pueden controlarse eligiendo un período relativamente pequeño para la prueba donde los efectos del cambio climático deberían ser despreciables.

Un conjunto de herramientas gráficas e indicadores numéricas se utiliza para comparar aspectos distribucionales de la señal observada y downscaled durante un período de prueba. En lo que sigue, se describen estas métricas enfocándose en el caso de tener una serie diaria como señal de entrada para hacer downscaling a escala horaria, modificamos ligeramente la notación introducida anteriormente para que Y_d represente la variable a escala diaria y $Y_d(h)$ a escala horaria. Vale la pena señalar que la mayoría de las métricas pueden adaptarse a otras resoluciones temporales.

4.2 Indicadores Cuantitativos

Usamos una combinación de indicadores tradicionales y desarrollados a medida para medir el rendimiento de los predictores. En total quedan 11 indicadores en la evaluación pareada y 8 en la no pareada con un indicador adicional en el caso de la variable de precipitaciones. La Table 2 resume los indicadores desarrollados a medida presentados en esta sección, indicando si se aplican en la evaluación pareada o no pareada.

Table 2: Indicadores de validación

	Pareada	No pareada
Amplitud	$\frac{\sum_{d=1}^D A_d - \hat{A}_d }{D}$	$\frac{\bar{A}}{\bar{\hat{A}}}$
Max. Diaria	$\frac{\sum_{d=1}^D mh_d - \hat{m}h_d }{D}$	$\frac{\sum_{h=0}^{23} \sum_{d=1}^D \mathbb{I}_{mh_d=h} - \mathbb{I}_{\hat{m}h_d=h} }{2D}$
Dist. Extremos	$\sum_{d d \in D*} \frac{ \mathbb{I}_{E_d=\hat{E}_d} }{ D* }$	$\sup_x F_{X_d E_d}(x) - F_{\hat{X}_d \hat{E}_d}(x) $
Cor. Signo	$\frac{\sum_{n=1}^N \mathbb{I}_{sg(X_{n+1}-X_n)=sg(\hat{X}_{n+1}-\hat{X}_n)}}{N-1}$	—
Horas lluviosas	$\sum_{d=1}^D \frac{ rh_d - \hat{r}h_d }{D}$	$\frac{\bar{r}h}{\bar{\hat{r}h}}$

Los indicadores tradicionales incluyen Error Absoluto Medio (MAE), correlación, diferencia de medias y razón de desviaciones estándar. MAE y correlación están relacionados con la *evaluación pareada* y la diferencia de medias con la *evaluación no pareada*. Además, para ambos casos, calculamos el MAE del QQplot, la Función de Autocorrelación (ACF) y el extremograma. Varios de estos indicadores sólo son aplicados en el caso de la evaluación pareada, ya que requieren una correspondencia uno a uno entre las observaciones reales y las predicciones.

Los indicadores desarrollados a medida fueron diseñados para evaluar qué tan bien el predictor puede preservar los atributos característicos a escala más fina. Por ejemplo, se analizó cómo el predictor podía recuperar la amplitud diaria. Es importante señalar que la amplitud no está disponible en los datos de CMIP6, pero se obtiene después del proceso de downscaling. Los indicadores están vinculados a alguna propiedad relevante relacionada con el problema original que no está presente en la escala original de CMIP6. Más precisamente, los indicadores están relacionados con la amplitud diaria, la hora en que ocurre el máximo diario, extremos y cómo cambia la serie.

Amplitud. La amplitud diaria se define como el rango de la variable en cada día, $A_d = \max_h \{Y_d(h)\} - \min_h \{Y_d(h)\}$. En Table 2 presenta la métrica propuesta para evaluar qué tan bien se captura la amplitud diaria, donde D Número de días, \hat{A}_d es la amplitud del día d calculada con la señal estimada downscaled, \bar{A} es la media de la amplitud de los días, y $\bar{\hat{A}}$ es la media de la amplitud estimada de los días.

Máximo diario. Se define $mh_d = \arg\max_h \{Y_d(h)\}$, la hora pico del día d . En Table 2 presenta la métrica propuesta para evaluar qué tan bien se captura la hora del máximo diario,

donde D Número de días, mh_d es la hora real del día cuando ocurrió el pico en el día d , $\hat{m}h_d$ es la hora estimada del día cuando ocurrió el pico en el día d , donde $I_{mh_d=h}$ es una función indicadora igual a 1 cuando el pico en el día d ocurrió en la hora h . Explicación análoga para $\mathbb{I}_{\hat{m}h_d=h}$

Extremos. Un valor extremo en la variable respuesta se define como un valor mayor que un umbral pre-especificado, consideramos que el valor es extremo si su valor es mayor que el valor correspondiente al cuantil 0.97 de la distribución. Nos estamos enfocando en la cola superior. Entonces, sea D^* el conjunto de días que tienen un valor extremo, y $E_d = \mathbb{I}_{d \in D^*}$, un indicador binario de si el día d tiene un valor extremo en los datos observados.

En Table 2 presenta la métrica propuesta para evaluar qué tan bien se capturan los extremos, donde \hat{E}_d es un indicador de si se predice que el día d tiene un valor extremo. La función $\mathbb{I}_{E_d=\hat{E}_d}$ es una función indicadora igual a 1 cuando ambos días tienen un valor extremo y 0 cuando uno tiene un valor extremo pero el otro no. Los valores posibles de este indicador varían entre 0 (los días con extremos en la serie real y la serie estimada nunca coinciden) y 1 (los días con extremos siempre coinciden). Valores más altos son preferibles.

En el caso de la métrica no pareada, $F_{X_d|E_d}(x)$ y $F_{\hat{X}_d|\hat{E}_d}(x)$ son las distribuciones empíricas de $X_d|E_d$ y $\hat{X}_d|\hat{E}_d$ respectivamente.

Es relevante analizar si los extremos se distribuyen entre los diferentes valores de la media diaria de la misma manera que ERA5. Por ejemplo, puede ocurrir que los extremos del valor downscaled se concentren en días con una media diaria alta, lo que significa que el predictor estaba sesgado contra hacer downscaling de extremos de corta duración (por ejemplo, períodos breves de lluvia intensa, como un aguacero intenso que dura una hora antes de detenerse). Por lo tanto, se busca comparar la distribución de $X_d|E_d$ y $\hat{X}_d|\hat{E}_d$ donde X_d es el promedio diario y \hat{X}_d es el promedio diario estimado. Para esto se utiliza el estadístico de Kolmogorov-Smirnov para dos muestras.

Correlación de signo. Este indicador mide qué tan bien el predictor captura la dirección del cambio en la serie temporal. En Table 2 presenta la métrica propuesta para evaluar qué tan bien se captura la dirección del cambio, donde N número de observaciones, sg es la función signo, X_n, X_{n+1} son los valores reales de la observación n y $n+1$, y \hat{X}_n, \hat{X}_{n+1} son los valores estimados para la observación n y $n+1$.

La función $\mathbb{I}_{sg(X_{n+1}-X_n)=sg(\hat{X}_{n+1}-\hat{X}_n)}$ es una función indicadora que es igual a 1 cuando la dirección del cambio en la serie real coincide con la dirección del cambio en la serie estimada y es igual a 0 en caso contrario. De esta forma, los valores posibles de este indicador varían entre 0 (lo que significa que siempre que la serie real aumenta, la serie estimada disminuye, o viceversa) y 1 (lo que significa que la serie estimada siempre sigue la misma dirección que la serie real). Valores más altos son preferibles.

Horas lluviosas. En el caso de la variable de precipitación es relevante analizar si el predictor puede capturar la frecuencia de horas lluviosas en un día, considerando que una hora es lluviosa si llueve más de 0.1 mm/hora. En la Table 2 presenta la métrica propuesta para evaluar qué

tan bien se capturan las horas lluviosas, donde D es el número de días, rh_d es el número de horas de lluvia en un día y \hat{rh}_d es la cantidad estimada de horas de lluvia. En el caso no pareado, \bar{rh} es la media de las horas lluviosas por día, y $\bar{\hat{rh}}$ es la media de las horas lluviosas estimadas por día.

4.3 Herramientas Gráficas

Los gráficos estadísticos son herramientas poderosas para realizar diagnósticos de modelos. En este artículo, se utiliza una combinación de siete gráficos interactivos para evaluar el rendimiento de varios métodos de downscaling. Cada gráfico se enfoca en un aspecto relevante de la distribución, la visualización general se organiza en tres niveles que representan diferentes agrupaciones de las observaciones.

Un primer nivel de gráficos se crea usando la observación directamente sin ningún esquema de agrupación. El primer gráfico compara la función de distribución acumulativa usando un gráfico de probabilidad o gráfico Q-Q, donde 100 cuantiles de la señal estimada se comparan con los cuantiles de la señal verdadera. Se utiliza una versión des-tendenciada del gráfico Q-Q donde el eje vertical se reemplaza por la diferencia de cuantil entre la señal estimada y verdadera. Los gráficos Q-Q des-tendenciados son preferibles a la versión tradicional para diagnosticar normalidad. Otra característica distribucional especialmente relevante en datos temporales es la estructura de dependencia de la serie. Dos gráficos están dedicados a medir qué tan bien se captura la estructura de dependencia mediante el modelo de downscaling. Un gráfico de autocorrelograma corresponde a graficar la correlación del valor de la serie en el presente con sus valores rezagados para diferentes rezagos. Un gráfico de barras de las autocorrelaciones de las señales de referencia verdaderas se grafica en gris mientras que la autocorrelación de cada método de downscaling se muestra como líneas en color. Luego un gráfico similar, pero con un enfoque en la estructura de dependencia extrema es un Extremograma. El extremograma mide la dependencia temporal entre eventos extremos, proporcionando información sobre la agrupación y persistencia de extremos dentro de procesos estocásticos. A diferencia de las funciones de autocorrelación tradicionales, el extremograma se enfoca específicamente en el comportamiento de la cola de la distribución, haciéndolo particularmente útil en aplicaciones que involucran eventos raros o extremos, como extremos ambientales.

Un segundo nivel de la visualización gráfica propuesta usa datos agrupados a nivel diario. Primero, se registra la hora del día donde ocurre el máximo (Max-hour). Se grafica un gráfico de barras del Max-hour de la señal verdadera y polígonos de frecuencia del Max-hour estimado con color según el predictor se grafica como una segunda capa. Este gráfico muestra qué tan bien un modelo particular captura el patrón máximo diario. Un segundo resumen diario corresponde a la amplitud o rango de la variable en cada día. El segundo gráfico en este nivel de agrupación presenta diagramas de caja de la señal verdadera con gráficos de violín coloreados como segunda capa.

El tercer nivel de la visualización gráfica corresponde a una agrupación horaria. Se grafica el perfil medio horario para evaluar qué tan bien un modelo recupera la estructura horaria a partir de datos de resolución diaria. El gráfico incluye el perfil medio horario de la señal verdadera como puntos y línea en color gris, y una segunda capa con líneas coloreadas con las medias estimadas, el gráfico usa estaciones como facetas. “

5 Resultados y Discusión

En esta sección se comentan los resultados de aplicar las estrategias de evaluación descrita más arriba. Los resultados consisten en tableros que presentan los indicadores numéricos en una tabla que destaca el mejor predictor y segundo mejor para cada métrica. Acompañado de los indicadores gráficos presentados en un gráfico interactivo.

Todos los resultados están disponibles en el repositorio https://github.com/nachalca/FSE2022_1_173061. Los tableros de la estrategia de modelo perfecto pueden obtenerse en la carpeta `reports/model_evaluation/` y los tableros de la estrategia de modelo imperfecto en la carpeta `reports/downscaling_evaluation/`.

5.1 Evaluación de modelo perfecto

El tablero contenido en el archivo `model_evaluation/pr_reduced.html` presenta los resultados para la variable precipitación. Con respecto a la distribución acumulada y la estructura de dependencia para los datos de precipitación. Todos los predictores muestran una ligera sobreestimación de los patrones de dependencia en términos de auto-correlaciones y Extremogramas con valores más altos que la señal verdadera, y una subestimación de los cuantiles más extremos. El predictor `xgboost` con una función de pérdida especial para lluvia (`xgboost_custom`) parece capturar algo mejor la precipitación extrema que el resto de los predictores. Por otro lado, Los indicadores agrupados diarios de la hora máxima de precipitación muestran un patrón relativamente uniforme (hay barras más grandes al final y al principio del día, pero esto es principalmente un efecto artificial de una secuencia de un día sin lluvia a un día lluvioso y viceversa). El patrón de media horaria es bien capturado por cada predictor excepto `xgboost_custom` que muestra una estimación sesgada hacia arriba de las medias horarias en cada estación.

El tablero contenido en el archivo `model_evaluation/tas_reduced.html` presenta los resultados para la variable temperatura. Los patrones de dependencia (correlaciones y extremos) son muy bien modelados por todos los predictores, la distribución acumulada para el predictor CNN muestra un pequeño sesgo hacia valores bajos (solo 1/2 grado) mientras que el predictor LSTM muestra un sesgo similar pero hacia valores altos de temperatura. Con respecto a los resultados de gráficos agrupados a nivel diario, todos los predictores muestran un buen desempeño capturando la hora donde ocurre el máximo, todos los polígonos de frecuencia muestran valores altos en las horas 15 y 16, el predictor CNN se concentra casi exclusivamente

en la hora 16. La distribución de amplitud diaria es bien capturada especialmente por los predictores LSTM y XgBoost.

El tablero contenido en el archivo `model_evaluation/sfcWind_reduced.html` presenta los resultados para la variable viento. Los gráficos Q-Q sugieren que los predictores XGBoost y Naive están subestimando la dispersión de la distribución ya que los cuantiles inferiores son demasiado grandes y los cuantiles superiores son demasiado bajos. Los predictores basados en redes neuronales funcionan mejor en esta métrica, y parecen capturar particularmente bien la cola superior de la distribución de viento. Con respecto a los gráficos agrupados a nivel horario, Todos los predictores recuperan el patrón de media horaria en cada estación, valores de viento más altos en las últimas horas del día y niveles más altos en la estación de Primavera.

5.2 Evaluación de modelo imperfecto

La segunda estrategia de diagnóstico consiste en comparar series de escala reducida de CMIP6 con las observadas ERA5 entre 2015 y 2023. En este caso hay 7 series de datos gruesos para cada una de las variables de respuesta (temperatura, precipitación y viento), correspondientes a una combinación de laboratorios y escenario de cambio climático. Esto resulta en que hay demasiadas métricas de rendimiento para considerar, en este informe se realizan comentarios de síntesis de resultados generales sin entrar en detalle de cada evaluación. Sin embargo, cada tablero individual se puede encontrar en la carpeta `reports/downscaling_evaluation`.

En primer lugar, los resultados sugieren que el factor de escenario no es significativo. Esto sugiere que la estrategia de diagnóstico de comparar datos CMIP6 reducidos de escala contra datos ERA5 en una ventana de ejecución corta no está contaminada de los parámetros de simulación de cambio climático. Por lo demás, los resultados varían para cada respuesta, en precipitación todos los predictores tienen un rendimiento similar, con `cnn` y `xgboost` con resultados ligeramente mejores, para temperatura el modelo lineal parece tener niveles de error más bajos y para la respuesta de viento los modelos predictores `cnn` y `lstm` muestran mejor rendimiento. En términos del efecto del laboratorio, los errores asociados con los modelos `mri_esm2_0` son mayores en precipitación y especialmente en viento.

6 Comentarios finales

Este informe presenta el trabajo relativo a los predictores para realizar reducción de escala temporal estadística de distintas variables climáticas de interés. Se presentan estrategias para la evaluación del ajuste de predictores de downscaling temporal con algunos aspectos destacables. En primer lugar, se recomienda utilizar dos tipos de contextos, llamados pareado y no-pareado, de forma conjunta para realizar una evaluación global de los resultados de predictores. Segundo, se proponen un conjunto de indicadores combinando indicadores numéricos (tradicionales y nuevos) con indicadores gráficos. El conjunto de indicadores se diseña de

forma armoniosa en tableros que permiten comparar varios predictores en distintos aspectos distribucionales relevantes de una manera integral.

Los resultados sugieren que los predictores basados en técnicas flexibles como CNN, LSTM y XGBoost son capaces de reproducir la estructura intra-diaria para proyectar series futuras horarias. Existen diferencias en algunas de las variables; por ejemplo, los predictores basados en CNN resultaron con mejor desempeño para viento, aunque esas diferencias no son muy grandes. En términos de los modelos climáticos de CMIP6, los modelos correspondientes al laboratorio `mri_esm2_0` son los únicos que muestran un desempeño algo peor que el resto.

Con base en las herramientas de evaluación presentadas aquí, se abren varias posibilidades como siguientes pasos para profundizar el estudio de los predictores para reducción de escala y su evaluación. Por un lado, aparece como necesario desarrollar indicadores de evaluación (numéricos y gráficos) específicos para cada variable de recurso primario, ya que la naturaleza de cada variable hace que los aspectos de la distribución más relevantes a tener en cuenta no sean siempre los mismos para todas ellas. En segundo lugar, una limitación importante de la aproximación presentada es que ignora los aspectos espaciales del problema. Si bien el producto final relevante para estudiar el impacto en la generación de energía es efectivamente una serie univariada por recurso primario, es posible que para implementar la reducción de escala, y sobre todo para su evaluación, sea relevante considerar aspectos espaciales. Es importante destacar que varios de los indicadores utilizados pueden generalizarse para aplicarse en dos dimensiones. Finalmente, otra extensión interesante puede ser considerar los distintos recursos primarios de manera conjunta para estudiar la distribución de extremos multivariados.

- Anderson, K., Hansen, C., Holmgren, W., Jensen, A., Mikofski, M., & Driesse, A. (2023). Pvlib python: 2023 project update. *Journal of Open Source Software*, 8(92), 5994. <https://doi.org/10.21105/joss.05994>
- Copernicus Climate Change Service (C3S). (2017). *ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate*. <https://cds.climate.copernicus.eu/>; Copernicus Climate Change Service Climate Data Store (CDS).
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5), 1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>