

TRABAJO DE FIN DE MÁSTER

# **Análisis multitemporal del retroceso glaciar mediante teledetección y modelos predictivos: El caso del glaciar de Monte Perdido (Pirineo Central, 2011-2023)**

Autor: Pablo del Olmo Díaz

Director: Raquel Montorio Lloveria

Tutor: Jesús Revuelto Benedí

Máster Universitario en Tecnologías de la información geográfica para la ordenación del territorio:  
sistemas de información geográfica y teledetección

Septiembre 2025



**Instituto  
Pirenaico  
de Ecología**



**Universidad  
Zaragoza**

**Departamento de Geografía  
y Ordenación del Territorio**



## Resumen

La pérdida de hielo glaciar se trata de un indicador claro de los efectos del cambio climático en zonas de alta montaña. Este Trabajo de Fin de Máster tiene como objetivo analizar la evolución del glaciar de Monte Perdido en el Pirineo central entre los años 2011 y 2023, a través de la integración de datos topográficos obtenidos mediante escáner de laser terrestre (TLS) y vehículos aéreos no tripulados (UAV) así como el uso de técnicas de análisis estadístico. El estudio está basado en el cálculo de cambios en el hielo glaciar a partir de la herramienta M3C2 aplicada sobre nubes de puntos de diferentes años. Posteriormente se aplicó un modelo de regresión mediante el algoritmo XGBoost, añadiendo variables topográficas y climáticas como la pendiente, el Topographic Position Index (TPI), la radiación solar potencial, el espesor de nieve y hielo y la exposición al viento. Además, se aplicaron técnicas como los valores SHAP y gráficos de dependencia parcial (PDPs) para analizar la importancia relativa de cada variable. Los resultados presentan que el TPI es la variable con mayor capacidad explicativa en la mayoría de los periodos analizados. También se identificó un posible umbral energético de radiación solar en torno a los 10.000 Wh/m<sup>2</sup> que intensifica las pérdidas, por otro lado, se identificó un comportamiento no lineal en el caso del espesor de hielo con mayores tasas de ablación en zonas entre los 10 y 15 metros. En general, el modelo mostró un buen rendimiento predictivo ( $R^2 > 0,80$ ) y un valor de RMSE ajustado.

Por tanto, la combinación de fuentes de teledetección LiDAR con modelización estadística avanzada permite hacer una caracterización de los patrones de pérdida glaciar, reproducible en otros glaciares pirenaicos altamente estudiados, así como contribuir al seguimiento de estos sistemas bajo un contexto de cambio climático especialmente agresivo con glaciares de estas características.

**Palabras clave:** Ablación glaciar, LiDAR, XGBoost

## Abstract

The loss of glacial ice is a clear indicator of the effects of climate change in high mountain areas. This Master's Thesis aims to analyse the evolution of the Monte Perdido glacier in the central Pyrenees between 2011 and 2023, through the integration of topographic data obtained using terrestrial laser scanning (TLS) and unmanned aerial vehicles (UAVs), as well as the use of statistical analysis techniques. The study is based on the calculation of changes in glacial ice using the M3C2 tool applied to point clouds from different years. A regression model was then applied using the XGBoost algorithm, adding topographic and climatic variables such as slope, Topographic Position Index (TPI), potential solar radiation, snow and ice thickness, and wind exposure. In addition, techniques such as SHAP values and partial dependence plots (PDPs) were applied to analyse the relative importance of each variable. The results show that TPI is the variable with the greatest explanatory power in most of the periods analysed. A possible solar radiation energy threshold of around 10,000 Wh/m<sup>2</sup> was also identified, which intensifies losses. On the other hand, non-linear behaviour was identified in the case of ice thickness, with higher ablation rates in areas between 10 and 15 metres. In general, the model showed good predictive performance ( $R^2 > 0.80$ ) and an adjusted RMSE value.

Therefore, the combination of LiDAR remote sensing sources with advanced statistical modelling allows for the characterisation of glacier loss patterns, reproducible in other highly studied Pyrenean glaciers, as well as contributing to the monitoring of these systems in a context of climate change that is particularly aggressive for glaciers of this type.

**Key Words:** Glacier ablation, LiDAR, XGBoost

## Índice

1. INTRODUCCIÓN.....	4
1.1 Estado de la cuestión en el entorno del glaciar de Monte Perdido .....	6
1.2 Justificación del estudio.....	7
1.3 Objetivos e hipótesis.....	8
2. ÁREA DE ESTUDIO .....	8
3. BASE DE DATOS .....	10
3.1 Láser escáner terrestre (TLS).....	10
3.2 Vehículos aéreos no tripulados (UAV/Dron) .....	11
3.3 Modelo digital del terreno. MDT02.....	11
4. METODOLOGÍA.....	12
4.1 Preprocesado y tratamiento de datos .....	13
4.2 Cálculo de cambios volumétricos con CloudCompare (M3C2 Distance) .....	13
4.2.1 Variables topográficas .....	14
4.2.2 Datos espesor nieve/espesor de hielo glaciar .....	18
4.2.3 Variables derivadas de las nubes de puntos:.....	20
4.3 Construcción del modelo estadístico XGBoost .....	20
4.3.1 Preparación, recorte y enmascarado de las variables.....	20
4.3.2 Primer análisis exploratorio: correlación de variables.....	21
4.3.3 Modelización y entrenamiento XGBoost .....	21
4.4 Evaluación de resultados. ....	22
5. RESULTADOS .....	23
5.1 Periodo TLS.....	25
5.1.1 TLS 2011-2016.....	25
5.1.2 TLS 2011-2020.....	26
5.1.3 TLS 2013-2014.....	27
5.1.3 TLS 2014-2015.....	27
5.2 Periodo UAV .....	28
5.2.1 UAV 2021-2022 .....	28
5.2.2 UAV 2023-2024 .....	29
5.2.3 UAV 2021-2024 .....	30
5.3 Comparación global.....	30
6. DISCUSIÓN .....	31
7. CONCLUSIONES.....	32
8. ÍNDICE DE ABREVIATURAS.....	33
9. ÍNDICE DE TABLAS Y FIGURAS .....	34
10. FUENTES Y BIBLIOGRAFÍA.....	34
ANEXO .....	37

## 1. INTRODUCCIÓN

El concepto de cambio global, definido por primera vez en la década de los años 70, incluye el conjunto de procesos y transformaciones que afectan al clima, los sistemas terrestres, el mundo vegetal y animal, así como a la Humanidad (Clark et al., 2001). Posteriormente, en el año 2000, se extendió el término "Antropoceno" para referirse a una nueva era geológica caracterizada por el impacto de las actividades humanas sobre el equilibrio del sistema climático (Crutzen & Stoermer., 2000). Entre los efectos más significativos de este cambio global se encuentran la pérdida de biodiversidad, el avance del deshielo de los casquetes polares y glaciares, así como la intensificación de fenómenos climáticos extremos, como sequías, grandes incendios, inundaciones y olas de calor.

El clima de la Tierra funciona como un sistema energético complejo en el que interactúan distintos subsistemas. La alteración de su equilibrio, especialmente a través del aumento de la concentración de gases de efecto invernadero de origen antropogénico, ha desencadenado un aumento sostenido de las temperaturas medias globales desde mediados del siglo XX (IPCC., 2021)), afectando de forma directa al ciclo hidrológico y a la distribución del agua en el planeta.

Uno de estos subsistemas más sensibles del climático global es la criosfera, donde el agua se almacena en forma de nieve, hielo y permafrost (IPCC., 2019). La fusión de estas reservas, además de contribuir al aumento del nivel del mar, modifica los patrones de disponibilidad de agua dulce, esenciales para la biodiversidad y el funcionamiento de ecosistemas montañosos y fluviales. La distribución del agua no es homogénea ni en el tiempo ni en el espacio, y fenómenos como el retroceso glaciar constituyen una muestra tangible de este desequilibrio provocado por el cambio global.

En un contexto de un cambio climático global acelerado, los glaciares son unos excelentes geoindicadores de lo que ocurre en el planeta y permiten conocer cómo esta evoluciona hacia un clima global cada vez más cálido. Los glaciares de los Pirineos son incluso más sensibles a estos cambios al ser los más meridionales de Europa, por lo que están desapareciendo a un ritmo elevado (Izagirre et al., 2024). Comparándolo con otras cordilleras Europeas, como los Alpes o Sierra Nevada, los glaciares el Pirineo son más sensibles a este proceso de cambio climático. Según Zemp et al. (2015) y López-Moreno et al. (2016) el retroceso en los glaciares alpinos, aunque también sea significativo, es inferior a lo observado en el Pirineo, donde las masas de hielo han sido históricamente más pequeñas y se encuentran a menor altitud que las de la cordillera alpina. Allí la línea de equilibrio glaciar se sitúa en torno a los 3000-3200 metros mientras que en los Pirineos se encuentra a una menor altitud (Chueca et al., 2005). Conviene en este punto señalar que, para ser considerado un glaciar, una masa de hielo, nieve y detritos debe tener movimiento por efecto de la gravedad. Este desplazamiento se produce cuando la masa de hielo alcanza un espesor crítico que permite la deformación del hielo, así como el desplazamiento sobre el suelo rocoso (Patterson., 1994). Por lo tanto, debe existir un flujo activo, aunque sea lento, para que la masa de hielo pueda ser considerado como glaciar. En el caso español, en Sierra Nevada, el glaciar del Veleta desapareció por completo en el siglo pasado (Gómez-Ortiz et al., 2014) y en los Picos de Europa solo se puede hablar de heleros como el del Jou Negro, que son pequeñas acumulaciones de hielo permanentes durante todo el año en lugares extremadamente sombríos (González- Trueba 2022), o como recientemente ha ocurrido en otros glaciares Pirenaicos como el glaciar de Infiernos que es actualmente considerados heleros (Revuelto et al., 2025)

En los Pirineos, desde 1984 han perdido más del 80% de superficie y el 90% de su volumen (Vidaller et al., 2021) y numerosos estudios indican que su desaparición es inminente, aproximadamente en tan solo en una o dos décadas si las condiciones climáticas actuales continúan (López-Moreno et al.,

2025). De hecho, de los 52 glaciares que existían al final de la Pequeña Edad de Hielo, solo quedan 14 masas de hielo que se pueden considerar glaciares en la actualidad (Izagirre et al., 2024, Revuelto et al., 2025). La mayoría de los glaciares han sufrido una disminución en su masa y volumen desde el final de la pequeña edad de hielo ocurrida a mediados del siglo XIX (Marshall et al., 2014). Este retroceso se ha visto incrementado especialmente en los últimos 15 años, donde eventos climáticos particularmente críticos como los sucedidos en 2022 y 2023, donde se registraron temperaturas altas y escasa acumulación de nieve a lo largo de la temporada, son factores determinantes que explican el estado actual de los glaciares pirenaicos (López-Moreno, 2025).

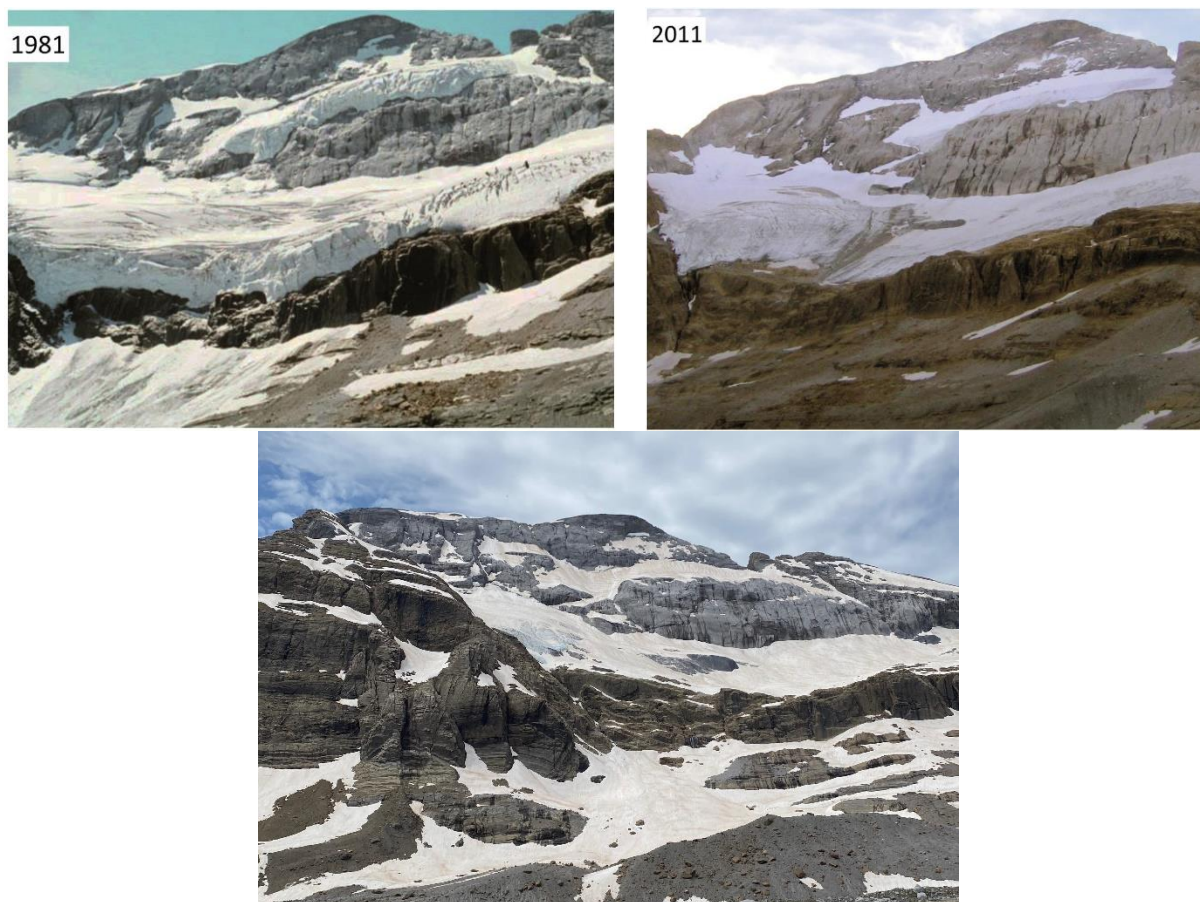


Figura 1. Fotografías del glaciar de Monte Perdido durante el verano en 1981 y 2011. (López-Moreno et al., 2016) La tercera imagen fue tomada a finales de junio de 2025. (Elaboración propia, 2025).

En concreto, el glaciar de Monte Perdido ha sido objeto de seguimiento mediante estudios desde los años ochenta, destacando investigaciones que comparan modelos digitales de elevación (DEMs) derivados de cartografía de 1981 y 1999 con datos obtenidos mediante vuelos LiDAR y campañas anuales de escaneo TLS desde 2011 (López-Moreno et al., 2016). Estas investigaciones han evidenciado tasas de pérdida de espesor de hielo de 1 metro por año en promedio entre 2011 y 2018 (López-Moreno et al., 2019) en periodos recientes, similares a las observadas en otras cordilleras europeas como los Alpes. Además, se ha comprobado que el aumento sostenido de la temperatura del aire desde la década de 1970, unido a una menor acumulación nival, ha elevado la altitud de la línea de equilibrio glaciar hasta cotas cercanas a los 2950 metros, situando al glaciar de Monte Perdido y a otros glaciares pirenaicos en una situación de desequilibrio climático crítico (López-Moreno et al., 2019)

El retroceso acelerado observado en los glaciares del Pirineo como el de Monte Perdido, el tercero con mayor superficie del Pirineo (Vidaller et al., 2021) por detrás del glaciar de Aneto y el de Ossoue,

es una consecuencia directa de las condiciones globales de la actualidad, que, según los estudios, no son compatibles con la supervivencia de los glaciares, especialmente los de cordilleras meridionales como los Pirineos (Revuelto et al., 2025) .

Por otro lado, la acelerada desaparición de ellos glaciares pirenaicos, además de ser, como ya se ha desarrollado, un claro indicador del cambio climático conlleva consecuencias directas sobre los ecosistemas que son únicos como en el contexto español y mediterráneo Esto supone un hándicap para el abastecimiento de agua o al turismo de montaña (Vidaller et al., 2021; CryoPyr-CSIC (2025).

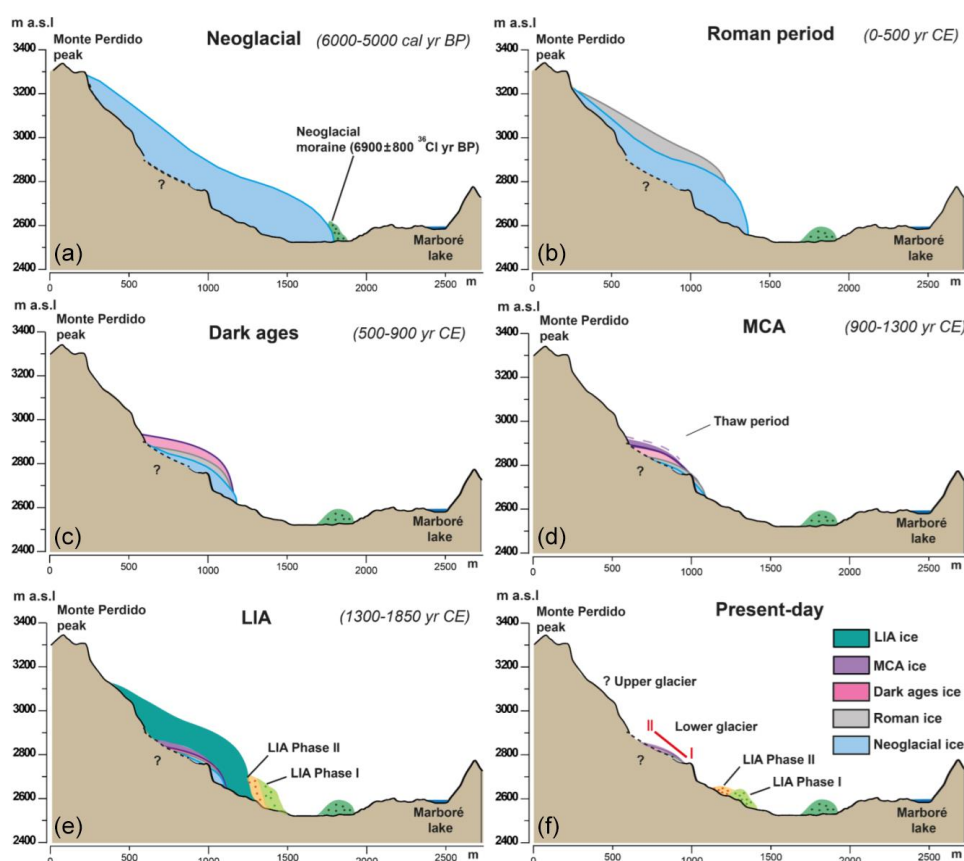


Figura 2. Esquema de la evolución del glaciar de Monte Perdido durante diferentes periodos climáticos. (Moreno et al., 2021)

Este fenómeno de pérdida continuada observada en el glaciar de Monte Perdido es especialmente alarmante ya que, según estudios, el glaciar sobrevivió en el pasado a periodos de calentamiento atmosférico como la Anomalía Climática Medieval (MCA, 900-1300 d.C) y al periodo romano (200 a.C – 500 d.C). A pesar de las temperaturas cálidas, el glaciar no sufrió un retroceso tan intenso como el registrado en los últimos años en el glaciar. Como se puede apreciar en la figura 2, se puede observar que el deshielo actual es mucho más acelerado que durante estos periodos cálidos anteriores (MCA y LIA) lo que es un indicador de la magnitud del calentamiento global de hoy en día.

### 1.1 Estado de la cuestión en el entorno del glaciar de Monte Perdido

El estudio de la evolución de los glaciares pirenaicos, y concretamente del glaciar de Monte Perdido, ha sido objeto de múltiples investigaciones en las últimas décadas, dada la situación de vulnerabilidad

extrema de estas masas de hielo ante el cambio climático. Desde los años ochenta, se han llevado a cabo diversos trabajos que han documentado la reducción de superficie y volumen del glaciar mediante técnicas tradicionales, como el análisis de fotografías aéreas, la comparación de modelos digitales de elevación (DEMs) y la interpretación de mapas topográficos. Ejemplos representativos son los trabajos de Julián y Chueca (2007) y López-Moreno et al. (2016), donde se calcularon tasas de pérdida de masa glaciar comparando datos de 1981, 1999 y principios de la década de 2010.

Tradicionalmente, los estudios se han centrado en la medición de la superficie glaciar y en la estimación de balances de masa utilizando estaciones meteorológicas situadas en el entorno del glaciar. Sin embargo, estas aproximaciones presentan limitaciones importantes: la resolución espacial es insuficiente para captar cambios locales, las estimaciones de masa son indirectas y las condiciones meteorológicas locales pueden no representar de forma precisa la dinámica del glaciar.

Con el avance tecnológico, desde 2011 se han incorporado técnicas de teledetección terrestre, como el escaneo láser terrestre (TLS) y posteriormente, del uso de vehículos aéreos no tripulados (UAV), desde 2022. Estos métodos permiten obtener nubes de puntos tridimensionales con una densidad y precisión significativamente superiores a las técnicas anteriores. En estudios más recientes como los de López-Moreno et al. (2019), Revuelto et al. (2020) o Izagirre et al. (2024), se ha demostrado que la aplicación combinada de TLS y UAV constituye una herramienta eficaz para monitorizar el estado y la evolución de glaciares pirenaicos.

Pese a estos avances, siguen existiendo limitaciones metodológicas. Muchos de los estudios realizados hasta la fecha se han centrado en analizar cambios de forma descriptiva, sin aplicar de manera sistemática técnicas de modelado estadístico que permitan interpretar las relaciones entre las variables topográficas, climáticas y la pérdida de masa glaciar. Asimismo, los estudios disponibles suelen ofrecer resultados para periodos concretos, sin una continuidad anual que permita identificar patrones interanuales de variación.

En este contexto, el presente trabajo pretende aportar una contribución diferencial, integrando series temporales continuas desde 2011 hasta 2024, aplicando técnicas de análisis estadístico avanzado para caracterizar de manera precisa el retroceso glaciar del Monte Perdido y los factores que intervienen.

## **1.2 Justificación del estudio**

La elección de este glaciar como caso de estudio se fundamenta en que la continuidad de registros topográficos desde 2011 es única en todo el Pirineo y hace que al glaciar de Monte Perdido sea uno de que cuentan con una de las mayores densidades temporales, así como espaciales de observación en Europa y en particular del Pirineo gracias a la aplicación de TLS y UAVs. Además, su situación en un entorno protegido como el Parque Nacional de Ordesa y Monte Perdido demuestra su interés como patrimonio paisajístico. Estas características convierten al Monte Perdido en un enclave ideal para evaluar con precisión el ritmo de deshielo a escala multianual, identificar los factores que más influyen en su evolución y explorar metodologías aplicables a otras regiones con masas glaciares en claro retroceso.

Este trabajo se inserta en ese marco metodológico, ya que se basa en el uso combinado de datos procedentes de TLS entre 2011 y 2022 y con datos de UAV entre 2022 y 2024 para obtener nubes de puntos tridimensionales del glaciar de Monte Perdido. A partir de estas nubes de puntos se aplica la herramienta M3C2 (Multiscale Model to Model Cloud Comparison) para calcular las diferencias de masa glaciar interanuales. Los resultados de este proceso se expresan como ganancias o pérdidas de espesor de hielo y se utilizan como variable dependiente en el posterior análisis estadístico que



incluye variables climáticas y topográficas como la pendiente, altitud, radiación solar, espesor de hielo, y espesor de nieve acumulado en distintos años etc.

El análisis posterior se lleva a cabo en el lenguaje R, mediante modelos de correlación y algoritmos de aprendizaje automático como XGBoost, también se emplean técnicas de interpretación como SHAP (SHapley Additive exPlanations) y gráficos PDP (Partial Dependence Plots) con el objetivo de comprender la importancia relativa de cada variable explicativa y la relación con el retroceso glaciar observado.

### 1.3 Objetivos e hipótesis

La principal hipótesis de este trabajo es que mediante el uso combinado de nubes de puntos tridimensionales procedentes de TLS y UAV, integrando herramientas innovadoras de análisis estadístico y aprendizaje automático, se puede obtener una mejor comprensión de los procesos que rigen la evolución de la superficie del glaciar de Monte Perdido. De esta manera se pretende identificar y cuantificar con mayor detalle los factores topográficos y climáticos que explican la variabilidad espacial de la pérdida de masa glaciar.

El objetivo general de este estudio es analizar la dinámica de retroceso del glaciar entre los años 2011 y 2024, evaluando los cambios en la superficie del hielo en relación con las variables que los condicionan, planteándose asimismo esos objetivos específicos:

1. Analizar los cambios en el hielo glaciar mediante la integración de datos LiDAR multianuales obtenidos mediante TLS y UAV
2. Caracterizar la influencia de variables topográficas y climáticas como como la pendiente, altitud, radiación, exposición al viento, curvatura del terreno, espesor de nieve y espesor de hielo en la dinámica del retroceso glaciar
3. Identificar las variables con mayor influencia y capacidad explicativa en el retroceso glaciar mediante la aplicación de un modelo estadístico y aprendizaje automático *XGBoost*
4. Interpretar los resultados a través de técnicas SHAP y gráficos PDP para entender las relaciones entre variables dependientes y la pérdida de masa glaciar en diferentes periodos de tiempo.

Con estos objetivos, se pretende describir de manera más precisa la evolución reciente del glaciar generando resultados que sirvan de referencia para su aplicación en futuras investigaciones o actuaciones de gestión medioambiental en otros glaciares con características y problemáticas similares.

## 2. ÁREA DE ESTUDIO

El glaciar de Monte Perdido se localiza dentro del parque nacional de Ordesa y Monte Perdido encuadrado en el Pirineo central a escasos kilómetros de la frontera con Francia. Este entorno fue incluido en la Lista de Patrimonio Mundial como bien mixto de tipo transfronterizo en 1997, abarcando 30.635 ha repartidas entre el lado español y el francés (zona oriental del Parc National des Pyrénées). Desde el punto de vista físico el macizo actúa como una barrera entre la vertiente atlántica, más húmeda y la mediterránea donde se presentan menores niveles de precipitación. En la vertiente española se encuentran grandes formas de relieve como los cañones de Ordesa o Pineta, mientras que en la zona francesa destacan los circos glaciares de Gavarnie o Estaubé. La zona permite observar los



procesos kársticos, fluviales, glaciares y eólicos que han generado un paisaje variado con cumbres áridas, valles tapizados de arbolado y praderas. (Ministerio de Cultura y Deporte, s.f.).

El glaciar está localizado en la ladera noreste del pico homónimo y cerca de otros picos cercanos que superan los tres mil metros como el Pico Cilindro o el Pico Marboré, también se encuentra a los pies del glaciar en la zona conocida como Balcón de Pineta, el Ibón de Marboré, a unos 2.600 metros sobre el nivel del mar. Se trata de uno de los glaciares con mayor extensión dentro de los glaciares pirenaicos y de los pocos que aún puede seguir siendo considerado como glaciar, al menos en el presente. Localizado en la ladera del macizo de Monte Perdido de 3.355m de altitud y en la actualidad se compone de tres masas diferenciadas localizadas a 2.885 y 3.110 metros respectivamente. (Izagirre et al., 2024)

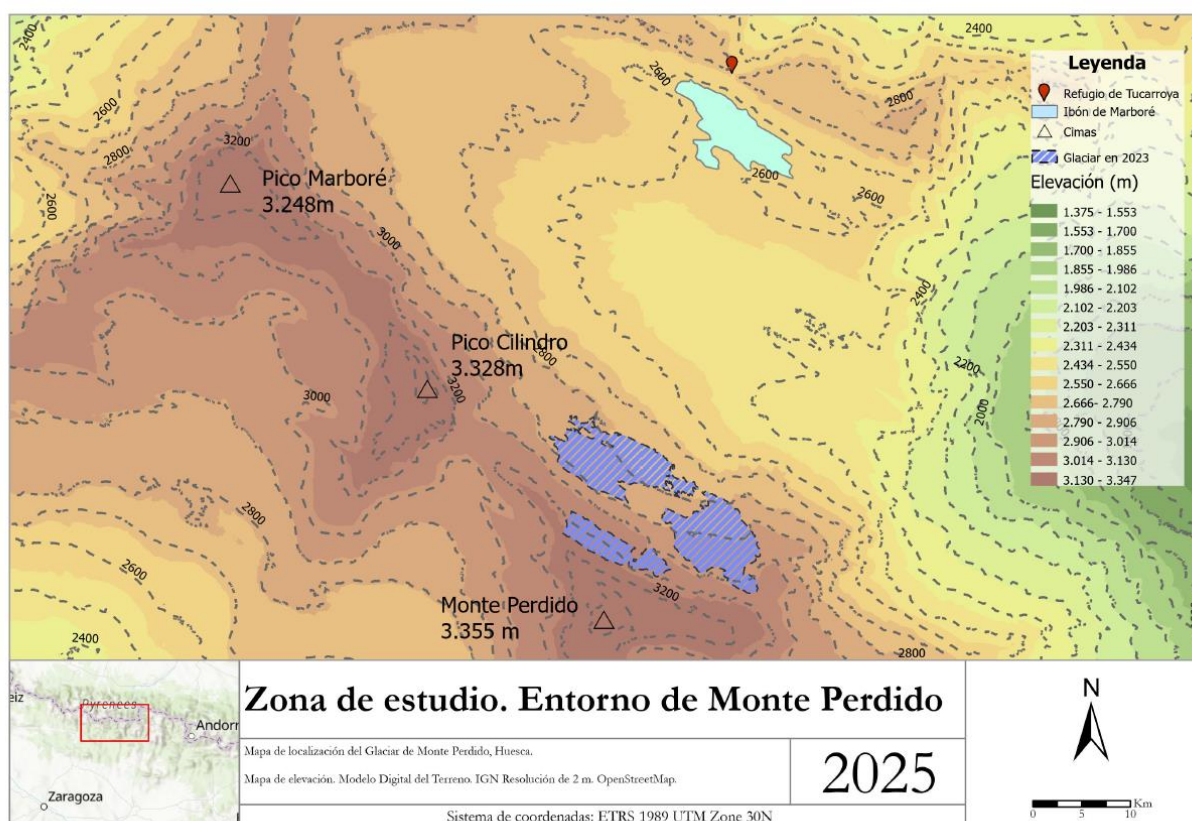


Figura 3. Zona de estudio, entorno del glaciar de Monte Perdido. Elaboración propia

Las masas situadas a menor altitud son las de mayor extensión tanto en volumen como en extensión de las dos, esto se debe a la morfología de la zona dominada por las pendientes pronunciadas y zonas de acumulación de hielo y nieve cada temporada. La clara orientación al norte-noreste del glaciar hace que cada año recibe aportes de nieve que unido a la altitud y a la limitada incidencia directa de la luz solar el glaciar continúe con presencia de hielo y cierto dinamismo hoy en día, aunque en claro retroceso.

Según investigaciones recientes llevadas a cabo en el Instituto Pirenaico de Ecología, el máximo más reciente de masa glaciar en los Pirineos se alcanzó hace unos 170 años y ha ido en una rápida y constante disminución hasta perder el 90% de la masa original. Por lo tanto, la completa desaparición de esta masa glaciares es cuestión de pocos años sobre todo tras eventos climáticos extremos críticos para los glaciares ocurridos en el 2022 y 2023 (López-Moreno et al., 2025).

### 3. BASE DE DATOS

#### 3.1 Láser escáner terrestre (TLS)

Los datos procedentes del escáner laser terrestre (TLS) comprenden el periodo entre el 2011 y el 2022 repartidos en campañas anuales de escaneo. Esta técnica ha demostrado ser altamente eficaz en lugares de alta montaña caracterizados por el difícil acceso y las condiciones climáticas extremas. Y es que la principal ventaja es que al ser un dispositivo móvil permite obtener datos con una menor dependencia de las condiciones climáticas, al contrario que ocurre con los dispositivos acoplados a aeronaves, que son muy sensibles a las condiciones de vuelo (Revuelto et al., 2021). Por lo tanto, ha sido ampliamente utilizada junto a otras técnicas geodésicas en el estudio de glaciares repartidos por el Pirineo como el glaciar del Aneto, Ossoue o Maladeta (Izagirre et al., 2024).

Esta técnica está basada en la tecnología LiDAR (Light Detection and Ranging) y permite obtener información topográfica con precisión gracias a la emisión de pulsos láseres desde un escáner colocado a una cierta distancia del glaciar. En este caso se utilizó el modelo RIEGL LPM-321 que permite conocer la distancia desde el sensor hasta la superficie del glaciar en función del tiempo que tarda en llegar el pulso laser desde que es emitido, rebotado por el hielo y recibido de vuelta.

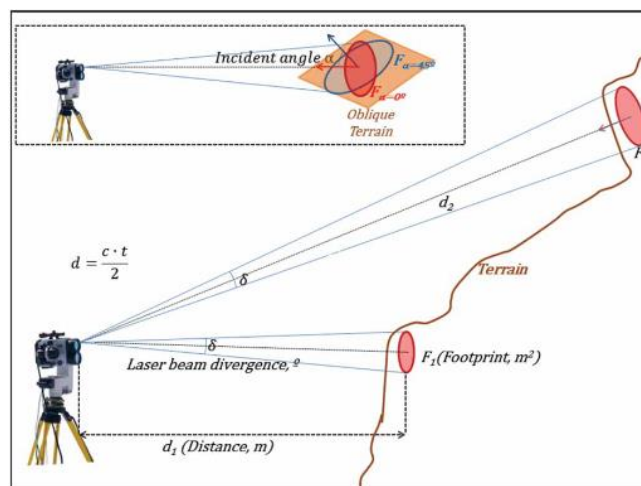


Figura 4. Esquema de funcionamiento del TLS. La zona iluminada o *footprint* depende de la distancia al objeto como del ángulo con el que el láser incide sobre la superficie del glaciar, por lo tanto, a mayor inclinación o distancia respecto al terreno, mayor será el tamaño de la huella (J. Revuelto et al. 2014).

Para una correcta toma de estos datos, el dispositivo se coloca a una distancia de alrededor de un kilómetro y medio de la masa de hielo en un punto frontal al glaciar donde se minimiza la presencia de sombras topográficas que afecten a la correcta toma de datos. El dispositivo se coloca sobre una zona lisa y estable como la roca desnuda para minimizar las vibraciones o movimientos indeseados que puedan afectar a la precisión durante el escaneo ya que es muy sensible a estos factores, pero también al viento o al frío extremo. Por ejemplo, una variación de 8' produce una desviación de 2,32 m a una distancia de 1000 m (Revuelto et al., 2014). En la figura 4 se observa como aumenta el *footprint* según la distancia y el ángulo.

Durante este proceso de toma de datos, se llevó a cabo un procedimiento conocido como registro indirecto o *target-based registration*. Consiste en el uso de objetos reflectantes ubicados de una manera estratégica alrededor del TLS. La posición de estos objetos está determinada a partir de un GPS diferencial y una estación total con alta exactitud, 0.02 m en el plano y 0.05 m en altimetría (López-Moreno et al., 2019, pp. 87–88).

El escaneo con el TLS se lleva a cabo siempre en las mismas fechas de manera periódica anual, aproximadamente a finales de septiembre o en octubre coincidiendo con el mínimo de hielo anual.

Los datos resultantes del escaneo se filtran y se georreferencian en un software especializado para que puedan ser comparables interanualmente. Para mejorar la calidad de los datos, se tienen en cuenta correcciones atmosféricas durante la toma de datos a través de ajustar la velocidad de propagación de la luz en función de variables físicas como la temperatura, presión y humedad, evitando estar expuesto a viento fuerte o insolación directa ya que podría afectar a la fiabilidad de la medida (Revuelto et al, 2014).

### **3.2 Vehículos aéreos no tripulados (UAV/Dron)**

A partir del año 2022, la toma de datos del glaciar de Monte Perdido se agregó el uso de vehículos aéreos no tripulados (UAV). Estos han sido una herramienta cada vez más utilizados para la toma de datos a gran escala gracias a su eficacia y ser una herramienta complementaria al escaneo terrestre tradicional. El uso de sensores LiDAR acoplados a diferentes plataformas móviles es cada vez más frecuente, dada su gran versatilidad para observar zonas muy extensas, su visión sintética de una amplia gama de variables de seguimiento y la rapidez de adquisición de información reduciendo costes al aumentar las superficies de rastreo (Chuvienco, 2002).

Al igual que con el TLS, estas sesiones de vuelo se llevaron a cabo en octubre coincidiendo con el mínimo glacial anual, lo que permite captar la extensión real del hielo al no existir la nieve depositada en los meses anteriores.

Concretamente en el 2022 y 2023 se realizaron trabajos en el glaciar de manera paralela con el TLS y con el UAV con el objetivo de validar los resultados entre sí, los buenos resultados permitieron que, en las campañas de 2024 y 2025 se realizaran exclusivamente con UAV. (Izagirre et al., 2024).

Una ventaja destacable de los UAV frente al TLS es la rapidez en la toma de datos. Mientras que una campaña TLS puede requerir una jornada completa de trabajo en campo, además del transporte de equipos pesados, un solo vuelo con dron, que no suele superar una hora, permite cubrir toda la superficie del glaciar con gran detalle. Además, el conjunto de UAV, baterías y tablet no supera los 5 kg de peso, lo que facilita su transporte y rápido despliegue sobre el terreno del dron, en este caso, se trabajó con un dron multirrotor DJI Matrice 300RTK.. Esta eficiencia operativa ha sido clave en su adopción como método principal a partir del 2023 por parte del equipo del Instituto Pirenaico de Ecología.

Los datos obtenidos con el UAV se georreferencian mediante GPS-RTK instalado en el propio aparato garantizando una elevada precisión, que cuentan con un error de geolocalización inferior a los 10cm en altimetría y 5 cm en planimetría.

A partir de estas dos fuentes principales de datos, las campañas de escaneo mediante el TLS y las adquisiciones mediante drones se generan las cartografías de pérdidas de espesor de hielo. Esta variable se obtuvo mediante la herramienta m3c2 (Multiscale Model to Model Cloud Comparison) que permitió conocer las variaciones de espesor y volumen de hielo. Esta variable se utilizó posteriormente como variable dependiente en los análisis de correlación y análisis estadístico.

### **3.3 Modelo digital del terreno. MDT02**

El Instituto Geográfico Nacional ofrece a través del centro de descargas del CNIG diferentes modelos digitales a diferente escala y nivel de procesamiento.

El Modelo Digital de Terreno empleado en este trabajo corresponde a un MDT de la segunda cobertura LiDAR del PNOA con resolución de 2 metros, esta no es la versión más actualizada y no es tan precisa como la tercera cobertura, sin embargo, el motivo por el cual se decidió usar esta cobertura fue que la tercera cobertura no está completamente terminada, y la tesela que corresponde al entorno del glaciar de Monte Perdido no estaba registrada a fecha de realización de este trabajo.

Además de las nubes de puntos generadas gracias a TLS y UAV, para este estudio se ha utilizado el MDT como fuente base para la obtención de variables topográficas que serán las variables independientes en el análisis que se desarrollará más adelante. Estas variables esenciales fueron la pendiente, la radiación potencial, índice de posición curvatura del terreno (TPI) y el índice de viento (Sx) (Winstral et al., 2002). Estas variables permiten caracterizar la dinámica glaciar y se han integrado con el conjunto de datos para el modelado posterior. Asimismo, también se añadieron al estudio dos variables adicionales, el espesor de nieve, calculado con el TLS y el espesor de hielo, calculado con GPR (López Moreno et al., 2019)

#### 4. METODOLOGÍA

Con la base datos descrita en el apartado anterior se aplicaron los procesamientos descritos a continuación: Este trabajo se desarrolló bajo el marco de los proyectos de investigación relacionados con la dinámica glaciar llevados a cabo por el Instituto Pirenaico de Ecología (IPE, CSIC) así como con la ayuda de los investigadores de dicha institución. Gran parte de la información y datos básicos para la realización de este trabajo fueron suministrados por este Instituto. La metodología aplicada en términos generales que se detallarán a lo largo de este apartado se estructuró en diferentes fases.

En primer lugar, se comenzó con la recopilación y preparación de diferentes fuentes de datos, tras esto, se generaron y se seleccionaron las variables suficientemente representativas del relieve y de la estructura del glaciar, que posteriormente se analizaron mediante técnicas estadísticas y de modelización. Con este proceso, que se presenta esquematizado en la figura 5, permitió obtener una visión detallada y consistente del comportamiento del glaciar en los diferentes periodos analizados.

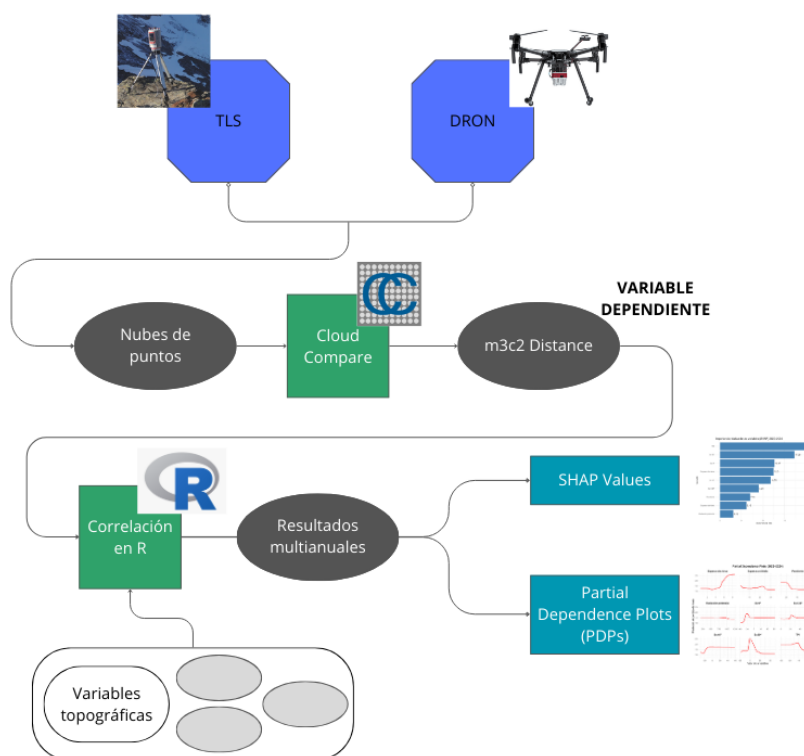


Figura 5. Esquema metodológico del trabajo. Elaboración propia.

En la figura 5 se muestra un esquema del proceso metodológico llevado a cabo en el trabajo, en primer lugar, se obtuvieron las nubes de puntos del glaciar en diferentes años a partir de los registros LiDAR llevados a cabo por el TLS y vuelos de dron. Posteriormente, se utilizó el programa Cloud Compare para generar a variable de cambio de masa a través de la herramienta m3c2 Distance, luego, se inició en el lenguaje R un modelo estadístico XGBoost con variables topográficas y climáticas obtenidas de manera externa. Los resultados fueron generados para diferentes periodos tiempo y para cada uno de ellos, métricas como el RMSE o el coeficiente de determinación ( $R^2$ ). Finalmente, se aplicaron técnicas de interpretación de modelos mediante SHAP y gráficos PDP para identificar las variables más influyentes en el retroceso glaciar. Esto no solo permitió conocer el peso de cada variable topográfica sino también visualizar la manera en la que se comporta relación con la pérdida de masas glaciar.

#### **4.1 Preprocesado y tratamiento de datos**

La información inicial fue recopilada a partir de las campañas de TLS, UAV y el Modelo Digital del Terreno (MDT). Cabe destacar que una parte de las nubes de puntos generadas mediante TLS y UAV ya fue entregada preprocesada, con las correcciones geométricas y de georreferenciación previamente aplicadas por el propio IPE. Por tanto, en este trabajo el preprocesado se centró en la comprobación de la alineación espacial de dichas nubes, la integración con el resto de las fuentes de datos y el ajuste de resolución, buscando la coherencia espacial de las distintas capas de información para el posterior análisis estadístico y caracterización precisa de la evolución glaciar. Los diversos sistemas de coordenadas y proyecciones se adaptaron al sistema ETRS89 / UTM zona 30 adecuado para la zona de estudio. Las nubes de puntos finales estaban registradas en formato LAS con la asignación de color RGB a cada punto.

#### **4.2 Cálculo de cambios volumétricos con CloudCompare (M3C2 Distance)**

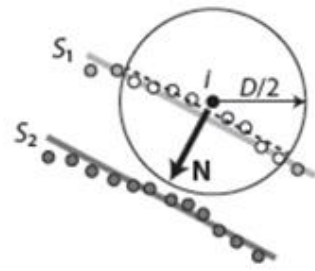
El cálculo de cambios de elevación del glaciar se llevó a cabo utilizando el software Cloud Compare. Este programa está especializado en el tratamiento de nubes de puntos, como puntos fuertes, es código abierto con lo que su uso es libre y es capaz de procesar grandes volúmenes de datos 3D (Lague et al., 2013).

Para obtener la variable de cambio de elevación en el glaciar de Monte Perdido se aplicó la técnica "Multiscale Model to Model Cloud Comparison" (M3C2) para la comparación precisa de nubes de puntos tridimensionales. Este método logra superar las limitaciones que se presentan en otros métodos tradicionales como la diferencia de modelos digitales de elevación (DoD), distancia de nube a malla (C2M) o la comparación directa nube a nube (C2C) (Lague et al, 2013).

A partir del trabajo realizado por Lague et al 2013 se puede describir el funcionamiento de esta herramienta. En primer lugar, se realiza el cálculo de la distancia perpendicular desde el punto hasta la superficie. Consiste en estimar los vectores normales en cada punto (core point), usando una aproximación mediante el ajuste de un plano a los puntos vecinos dentro de un radio previamente definido. La orientación del vector se establece respecto a los puntos externos definidos, resultando una medida que permite evaluar las distancias entre dos superficies.



Step 1 (M3C2): Calculation of normal  $\mathbf{N}$  at a scale  $D$  around the core point  $l$  in cloud  $S_1$ .



Step 2 (M3C2): Average distance between the two cloud measured along  $\mathbf{N}$  between mean points  $l_1$  and  $l_2$ , derived at a scale  $d$ .

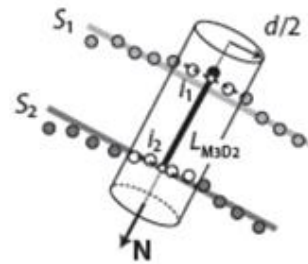


Figura 6. Representación del funcionamiento del algoritmo M3C2. Fase 1: cálculo de las normales locales. Fase 2: cálculo de la distancia entre nubes mediante una ventana cilíndrica. James et al., 2017.

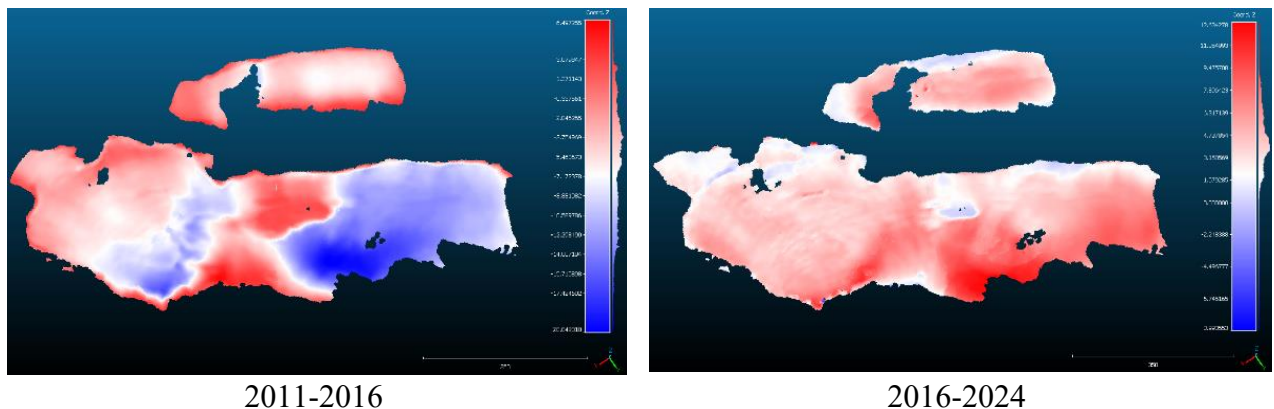


Figura 7. Ejemplos de cálculos M3C2 calculados en Cloud Compare.

#### 4.2.1 Variables topográficas

En este apartado se describen las variables topográficas a partir de las que se ha tratado de explicar las pérdidas de espesor del glaciar y que han sido empleadas como variables explicativas en el modelo estadístico. Las variables topográficas se obtuvieron en el software ArcGIS en su versión PRO, aprovechando la gran capacidad de gestionar archivos de carácter geográfico en un mismo entorno.

- **Pendiente (Slope):** Esta variable obtenida a partir del MDT del entorno de Monte Perdido. Para ello se utilizó la herramienta *Slope* en ArcGis pro, obteniendo la pendiente en grados para cada píxel del DEM. La pendiente influye directamente en la dinámica de acumulación y fusión de nieve, y condiciona además el movimiento potencial del hielo, al afectar a la velocidad y dirección del flujo glaciar (Cuffey & Paterson, 2010). Por ello, resulta esencial incluirla como variable explicativa para determinar la distribución espacial de la pérdida de masa glaciar.
- **Radiación potencial solar potencial (Solar Radiation):** Variable calculada para un periodo de tiempo desde el 1 de enero hasta el 31 de diciembre del 2025 gracias a la herramienta en ArcGIS Pro de *Area Solar Radiation*. La variable computa la radiación solar potencial que llega a una superficie durante un periodo de tiempo determinado, la unidad es vatios hora por metro cuadrado ( $\text{Wh/m}^2$ ). El cálculo se realizó a partir del modelo digital de elevaciones (MDE) estableciendo una ventana de cielo uniforme y considerando la latitud del glaciar

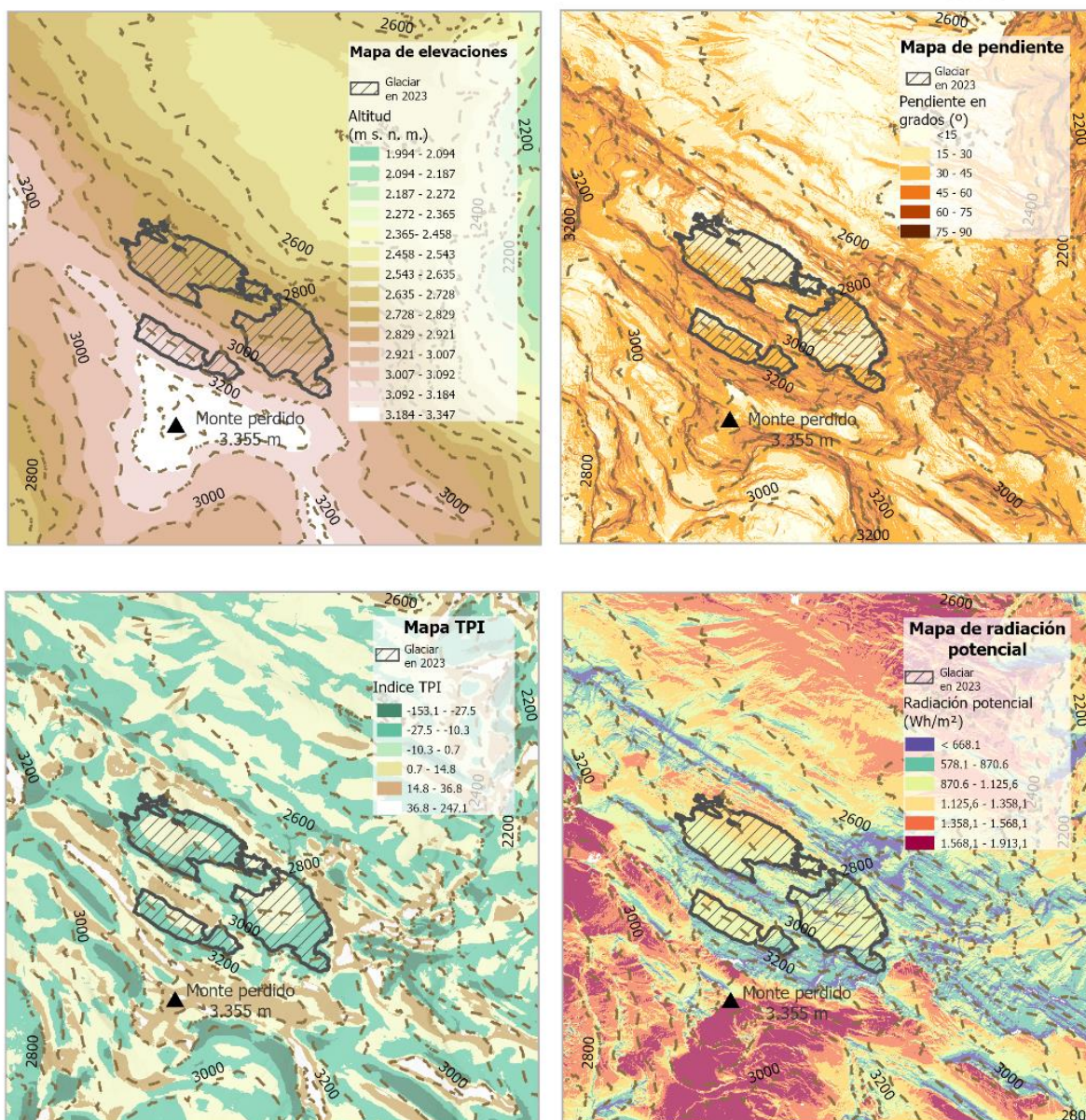
Monte Perdido (42°39' N). Se realizó promediando datos diarios acumulados, de modo que la variable muestra el balance de energía teórica anual por unidad de superficie. Los parámetros relativos a resolución espacial, tamaño de ventana y otros ajustes se mantuvieron en sus valores por defecto, según las recomendaciones de ESRI para entornos montañosos. Esta variable se ha incluido en el trabajo debido a su influencia en el balance energético del glaciar. Son varios los estudios que han demostrado que esta variable es fundamental en los procesos de fusión de nieve y hielo (Hock, 2005)

- **TPI (*Topographic Position Index*):** El TPI se calculó con ArcGis Pro con el objetivo de caracterizar la morfología de la superficie glaciar mostrando las zonas convexas, planas o cóncavas lo que influyen en los procesos de acumulación y posterior fusión de nieve y hielo. Este índice refleja la diferencia de altitud entre un píxel y el promedio de altitud de las celdas vecinas dentro de un entorno definido (De Reu et al., 2013). Para este trabajo se utilizó una ventana circular de 50m, suficiente para la zona de estudio. Los valores de TPI cercanos a 0 corresponden con zonas planas mientras que los valores positivos representan lugares elevados y convexas, y los negativos muestran las concavidades del terreno. Añadir el TPI al modelo estadístico se basa en la capacidad de esta variable para explicar patrones de distribución espacial de la nieve (Revuelto et al., 2014).
- **Wind Slope Parameter(Sx):** Esta variable permite caracterizar la exposición del terreno frente al viento. Es muy relevante para comprender la distribución de la nieve en zonas de montaña. Como se puede ver en la figura 8, este parámetro tiene una alta capacidad predictiva para la distribución espacial de la nieve en cuenca alpinas y subalpinas, además esta relación es consistente en diferentes escalas temporales e interanuales (Winstral et al., 2013). El parámetro se calculó a partir del programa MAXUS creado por Adam Winstral, necesita ser ejecutado en el Command Prompt de Windows a partir del archivo del modelo digital de elevación utilizado previamente, pero en esta ocasión en formato ASCII (.asc) con una distancia de búsqueda de 1000 metros. Se calculó en las direcciones de 0°, 45°, 90° y 135° (con ventanas de 45° para cada uno de ellos). Con estos 4 ángulos se logra cubrir de manera efectiva toda la rosa de los vientos con intervalos de 45° ya que se obtienen de manera inversa la relación potencial con sus ángulos complementarios (180°, 225°, 270° y 315°). Los valores positivos indican zonas apantalladas del viento y los valores negativos aquellos que están expuestos (figura 8).



## VARIALES TOPOGRÁFICAS EN EL GLACIAR DE MONTE PERDIDO

Comparativa cartográfica de elevación, pendiente, Índice TPI y radiación potencial



Mapas calculados a partir del Modelo Digital del Terreno (MDT) de 2 m del Instituto Geográfico Nacional (IGN), obtenido en abril 2024.  
Realizado en ArcGIS Pro utilizando la herramienta aspect, TPI Index y Area Solar Radiation (Enero - Junio)  
Proyección: ETRS89 / UTM Zona 30N

Fuente de datos: IGN — MDT 2 m

Realizado por Pablo del Olmo Díaz

0 0.2 0.4 0.8  
Kilometers



Figura 8. Representación de variables topográficas incluidas en el modelo estadístico. Elaboración propia.



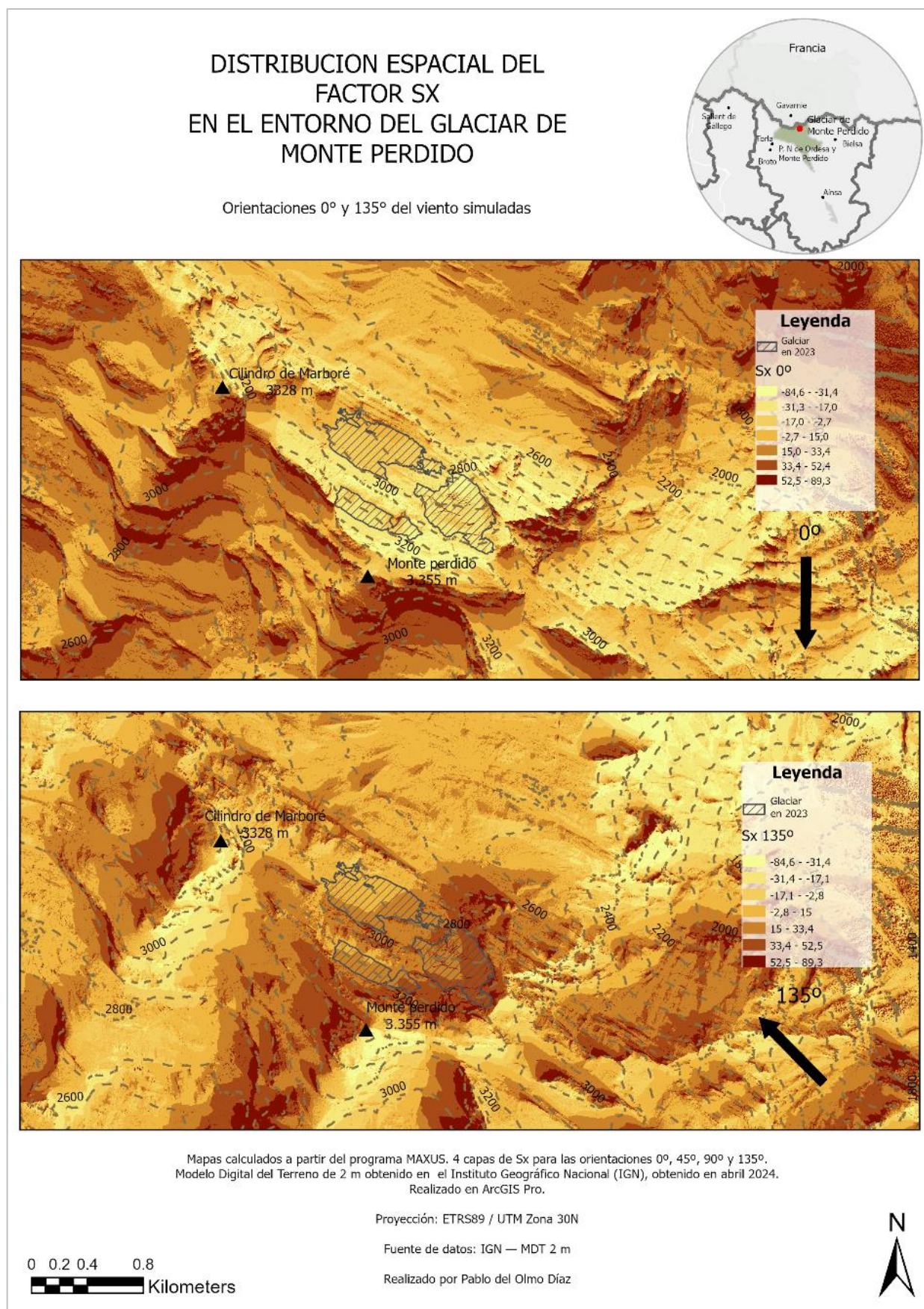


Figura 9. Representación del índice Sx para la orientación 0° y 135°. Elaboración propia. Las flechas indican la dirección del viento dominante en cada caso.

Otras variables topográficas calculadas, pero no añadidas al script final debido al bajo peso explicativo en los resultados finales, fueron la orientación y el Índice de Humedad Topográfica

(TWI). La orientación indica la orientación de una determinada ladera, fue calculada en ArcGIS Pro a partir de MDT, se calcula en grados respecto al norte geográfico siguiendo el sentido horario con valores entre 0° y 360°. Por parte del TWI, este es un indicador que combina la pendiente y TPI estimando las condiciones de una zona de terreno a retener la humedad (Beven et al., 1979).

#### 4.2.2 Datos espesor nieve/espesor de hielo glaciario

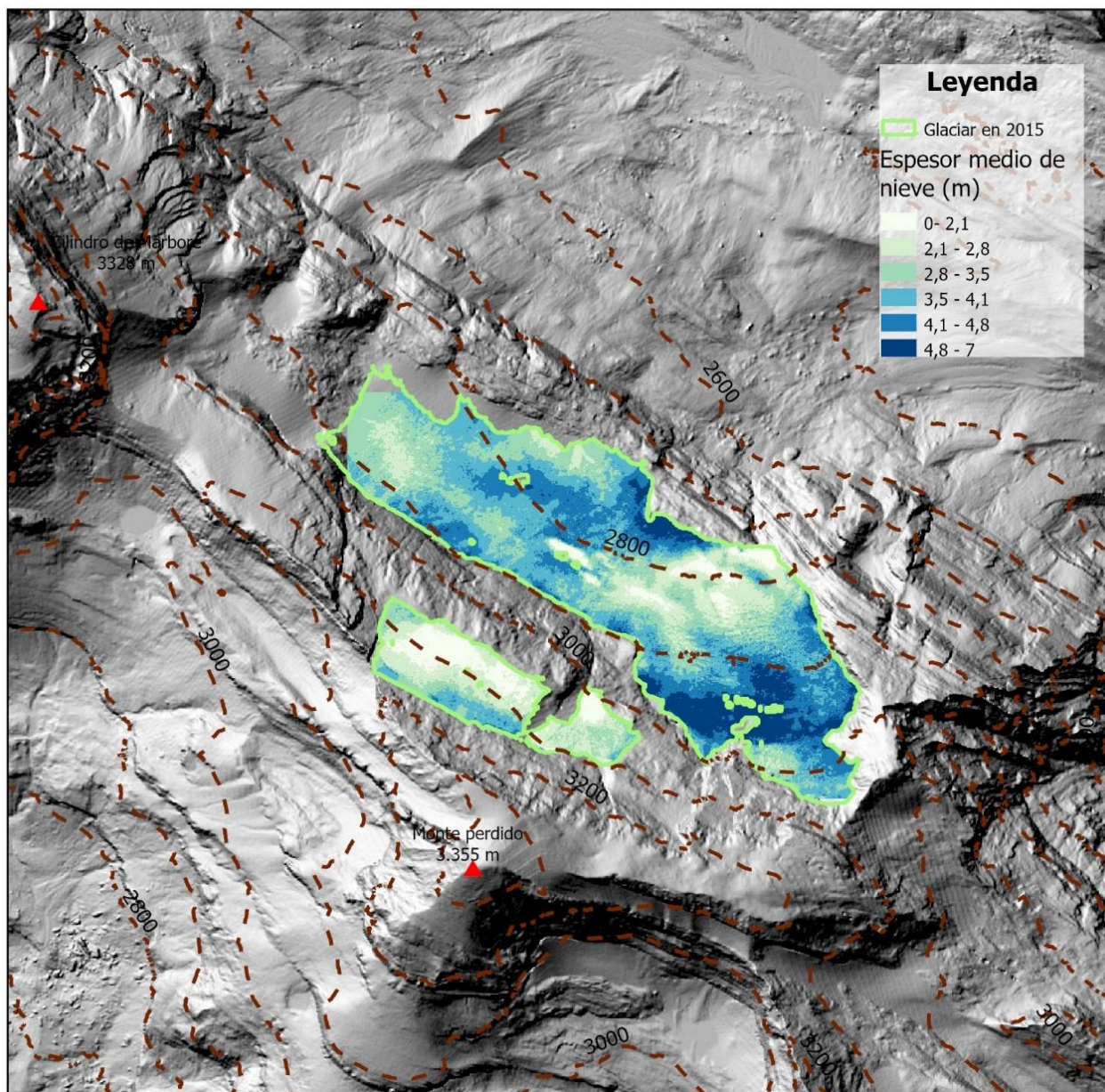
En este apartado se indican las variables que nos son puramente topográficas y que tiene cierta relación con la climatología y la topografía. El espesor de nieve observado al final del periodo de acumulación de nieve (abril -mayo) y el espesor de hielo. Estas variables fueron seleccionadas por su influencia directa en la distribución y pérdida de la masa glaciario. Ambas fueron obtenidas a partir de los estudios realizados por el Instituto Pirenaico de Ecología en campañas pasadas, ambas en formato ráster.

- **Espesor de nieve:** Los datos de espesor de nieve fueron proporcionados por el IPE y fueron derivados de campañas previas realizadas mediante TLS, correspondientes a los años 2014, 2015 y 2017, al final de la primavera, cuando se estima que el espesor de nieve es máximo, así como el valor medio de estos años. Estos datos fueron utilizados dependiendo del periodo de año analizado, por ejemplo, para intervalos previos al 2014 se utilizó el dato de espesor nival de 2014 y para datos posteriores al 2017, (2023-24 o 2022-23), se aplicó el más reciente, es decir 2017. Sin embargo, para periodos más dilatados como 5 o 10 años se utilizó la capa del promedio de la nieve.
- **Espesor de hielo:** Obtenido también a partir de campañas de medición del IPE con un GPR (López-Moreno et al., 2019). En este estudio se utilizaron registros en formato ráster de 2016, obtenido de la interpolación de la campaña de GPR y en 2023, sustrayendo la diferencia de espesor entre 2016 y 2023 al espesor interpolado en 2016.



## ESPESOR DE NIEVE MEDIO REGISTRADO EN EL PERIODO 2014-2017

Espesor medio de nieve de los años 2014 , 2015 y 2017.  
Glaciar de Monte Perdido



Mapa de espesor medio de nieve en metros para el periodo 2014-2017.  
Datos obtenido con TLS (Terrestrial Laser Scanning). Realizado en ArcGIS Pro

Proyección: ETRS89 / UTM Zona 30N

Fuente de datos: Instituto Pirenaico de Ecología

Realizado por Pablo del Olmo Díaz

0 0.2 0.4 0.8  
Kilometers



Figura 10. Distribución espacial del espesor de nieve promedio registrado entre 2014 y 2017.  
Elaboración propia.

#### 4.2.3 Variables derivadas de las nubes de puntos:

En este apartado se mencionan una serie de variables obtenidas en el software Cloud Compare a partir de las nubes de puntos. El software permite generar variables para caracterizar la superficie representada tales como: *Omnivariance* (para detectar zonas de mayor irregularidad superficial), *Sphericity*, *Planarity* (ambas para identificar superficies uniformes frente a las irregulares) y *Verticality* (para identificar escarpes o grietas donde se pueda producir una mayor fusión) (Weinmann et al., 2017).

Estas variables se calcularon con el objetivo de ser incluidas en el modelado estadísticos. Sin embargo, tras comprobar el peso de estas variables en el modelo se observó que estas variables no presentaban una relación significativa con la pérdida de masa glaciar, además tampoco mejoraban la capacidad predictiva del modelo XGBoost. Por este motivo, no fueron incluidas en la versión final del modelo estadístico.

### 4.3 Construcción del modelo estadístico XGBoost

El análisis estadístico y predictivo de la pérdida de masas glaciar de Monte Perdido se elaboró mediante un modelo de regresión utilizando el algoritmo XGBoost (Chen & Guestrin, 2016) Este proceso se ha creado integralmente en el lenguaje R, en este entono se han cargado las variables topográficas y se ha procesado a su análisis, como variable dependiente se toma el dato de balance de masa glaciar (M3C2) obtenido previamente como ya se ha indicado y se relaciona con las variables independientes topográficas y climáticas mencionadas en el apartado anterior.

#### 4.3.1 Preparación, recorte y enmascarado de las variables

Al comienzo de este script se realiza la carga de las librerías que posteriormente serán utilizadas para el tratamiento de los datos geospaciales, modelado estadístico y visualización, estas librerías son *terra*, *sf*, *dplyr*, *xgboost* y *iml*.

El código completo que recoge este proceso, así como la preparación y alineación de las variables en R, se incluye en el Anexo.

Después, se carga la pérdida de masa glaciar (M3C2) para un periodo de año en concreto que será la variable dependiente en el modelo. Luego se procede con las variables topográficas (Pendiente, TPI, Radiación Potencial y viento Sx) y climáticas (espesor de nieve y hielo) que serán las variables independientes.

Todas las capas de reprojectaron al sistema de referencia ETRS89/UTM zona 30N con código EPSG:25830 habitual en España. En el script esto se lleva a cabo mediante la función *crs()* del paquete *terra*, lo que permite el sistema referencial original de cada ráster como para asignar el sistema adecuado en concordancia con el resto de las capas. También, para asegurar que las capas coincidan en resolución espacial, extensión se utilizaron las funciones *res()*, *dim()* y *ext()*.

El área de referencia para el análisis fue variando en función del periodo analizado, el recorte del glaciar se importó en formato shapefile, en algunos casos fue proporcionado por el IPE y en otros fue calculado a partir de los datos de espesor de hielo en ArcGis PRO. En el script se le asigna la

proyección EPSG:25830 y se rasterizan las variables a una resolución de 2 metros. Luego se le aplica la función *SpatVector* lo que permite usar esta capa como recorte con el resto de las capas ráster.

Todas las capas se recortaron y enmascararon utilizando el límite glaciario adecuado a cada periodo analizado, de modo que el análisis estadístico se circunscribe exclusivamente al espacio reconocido como glaciario vigente en el periodo de análisis, para ello se aplicaron las funciones *crop()* y *mask()*.

#### 4.3.2 Primer análisis exploratorio: correlación de variables

En primer lugar, para poder manejar las variables de manera correcta se creó un *stack* (*stack\_vars*) con las variables ya recortadas, se les asignó un nombre reconocible mediante la función *names()*. Luego se extrajeron los valores por píxel mediante *as.data.frame()*, generando un *dataframe* en el que cada fila está correspondida a un píxel y cada columna a una de las variables dependientes.

Para valorar la relación entre la pérdida glaciaria y el resto de las variables se realizó un análisis de correlación de Pearson. Con el objetivo de conseguir un análisis más robusto se realizó un muestreo aleatorio mediante la configuración de un bucle de 10.000 iteraciones, lo que logra una mayor estabilidad y precisión y en cada iteración se extrajo una muestra aleatoria de 100 píxeles, generándose las correlaciones para cada muestra. Esto permite estimar la distribución de los coeficientes de correlación evitando de esta manera la dependencia de una única muestra. Posteriormente se visualizaron las correlaciones mediante matrices *corrplot* y diagramas de dispersión *ggplot2*.

#### 4.3.3 Modelización y entrenamiento XGBoost

El algoritmo XGBoost o *eXtreme Gradient Boosting*) es un método de aprendizaje basado en árboles de decisión que mejora la capacidad predictiva de un modelo a través de *boosting* por gradiente, es decir, el modelo se entrena múltiples veces y cada nuevo modelo intenta corregir los errores que pueda haber en el anterior. La principal ventaja radica en la capacidad que presenta para manejar relaciones complejas y en su eficacia a nivel computacional, lo que lo convierte en un algoritmo útil para trabajar con conjuntos de datos grandes o con múltiples variables predictoras (Chen & Guestrin, 2016).

En el script se aplicó XGBoost al *dataframe* ya creado previamente, pero con el objetivo de evaluar correctamente la capacidad del modelo y evitar el sobreajuste, la muestra original se dividió de manera aleatoria en dos conjuntos, el primero de entrenamiento con el 70% de los datos mientras que el segundo se trata del conjunto de validación con el 30% restante (James et al., 2021). Tras esto, mediante *set.seed()* se crea un generador de números pseudoaleatorios, esta función lo que garantiza es que cada vez que se ejecute el código se obtengan los mismos resultados, lo que permite que el código sea reproducible.

Una vez ya se definieron los conjuntos de datos de entrenamiento y validación, se procedió a la configuración y entrenamiento del modelo XGBoost. Este modelo se ejecutó mediante la función *xgb.train()* del paquete *xgboost*, en este punto se definieron los parámetros empleados;

- *objective = "reg:squarederror"*. Especifica que se está realizando una regresión

- $\eta = 0.03$ . Se trata de la tasa de aprendizaje, regulando las correcciones en cada iteración. Los valores recomendados están entre 0.02 (mejor aprendizaje, pero se necesitan mayores recursos computacionales) y 0.03 (Chen & Guestrin, 2016).
- $\max\_depth = 5$ . Asigna la profundidad máxima de los árboles de decisión. Valores recomendados entre 3 y 10 (Brownlee, J. 2016).
- $\text{subsample} = 0.7$  y  $\text{colsample\_bytree} = 0.7$ . En cada árbol se utilizará el 70% de las variables disponibles, lo que introduce aleatoriedad y reduce el riesgo de sobreajuste.
- $\text{min\_child\_weight} = 10$ ,  $\text{gamma} = 1$ ,  $\text{lambda} = 1$ : otros parámetros con valores asignados por defecto.
- $\text{eval\_metric} = "rmse"$ : indica que la métrica de evaluación aplicada fue el error cuadrático medio

Tras configurar los parámetros del modelo, se creó una *watchlist* que monitoriza la evolución del modelo tanto en el de entrenamiento como con el de validación. Para evitar el sobreajuste se creó  $\text{early\_stopping\_rounds} = 10$  lo que hace que el entrenamiento se detenga si durante 10 iteraciones el RMSE no mejora.

La implementación detallada de lo expresado en este epígrafe se puede consultar en el Anexo

#### 4.4 Evaluación de resultados.

Para conocer la calidad del ajuste y su capacidad predictiva, se evaluó el modelo una vez ya se había entrenado el modelo. Para ello se generaron las siguientes métricas, las principales son el error cuadrático medio (1) y el coeficiente de determinación (2), también se calcularon otras métricas como el error absoluto medio (MAE), se trata del cálculo de la diferencia entre los valores reales y los predichos por el modelo, el error porcentual absoluto medio (MAPE) que muestra el error como porcentaje del valor observado, y el Nash-Sutcliffe Efficiency (NSE) que mide la eficiencia predictiva del modelo si se compara con el de una predicción basada en la media observada.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

Además de las métricas numéricas, se aplicaron otras técnicas más avanzadas de interpretación del modelo.

En primer lugar, se evaluó la contribución individual de cada variable predictora en el modelo, para ello se utilizaron los valores SHAP (*Shapley Additive Explanations*). Esta técnica proporciona una interpretación completa de cómo se comporta el modelo creado, permitiendo identificar cuáles son las variables (topográficas o climáticas) que tienen mayor capacidad explicativa. Además, los gráficos SHAP permiten mostrar tanto la magnitud como la dirección del efecto de cada variable explicativa sobre el modelo (Molnar, C., 2022). La aplicación de esta técnica se ha asentado como una de las más robustas por su capacidad de interpretación de modelos complejos (Lundberg & Lee, 2017).

Con el objetivo de mejorar la interpretación de los resultados, en segundo lugar, se generaron gráficos de dependencia local o PDP mediante el paquete en *r* *iml*. En términos generales los gráficos PDP muestran cómo cambia la predicción del modelo al modificar una de las variables explicativas dejando las demás constantes (Molnar, C., 2022). Esta técnica permite identificar, por ejemplo, si una variable como la pendiente o el espesor de hielo aumentan, se produce, o no, una mayor pérdida de



masa glaciar de una manera progresiva o por el contrario existe un punto donde a partir del cual se intensifican estas pérdidas. Los gráficos PDP permiten detectar de manera visual la tendencia o cambios bruscos en el modelo en función de cada variable topográfica o climática.

## **5. RESULTADOS**

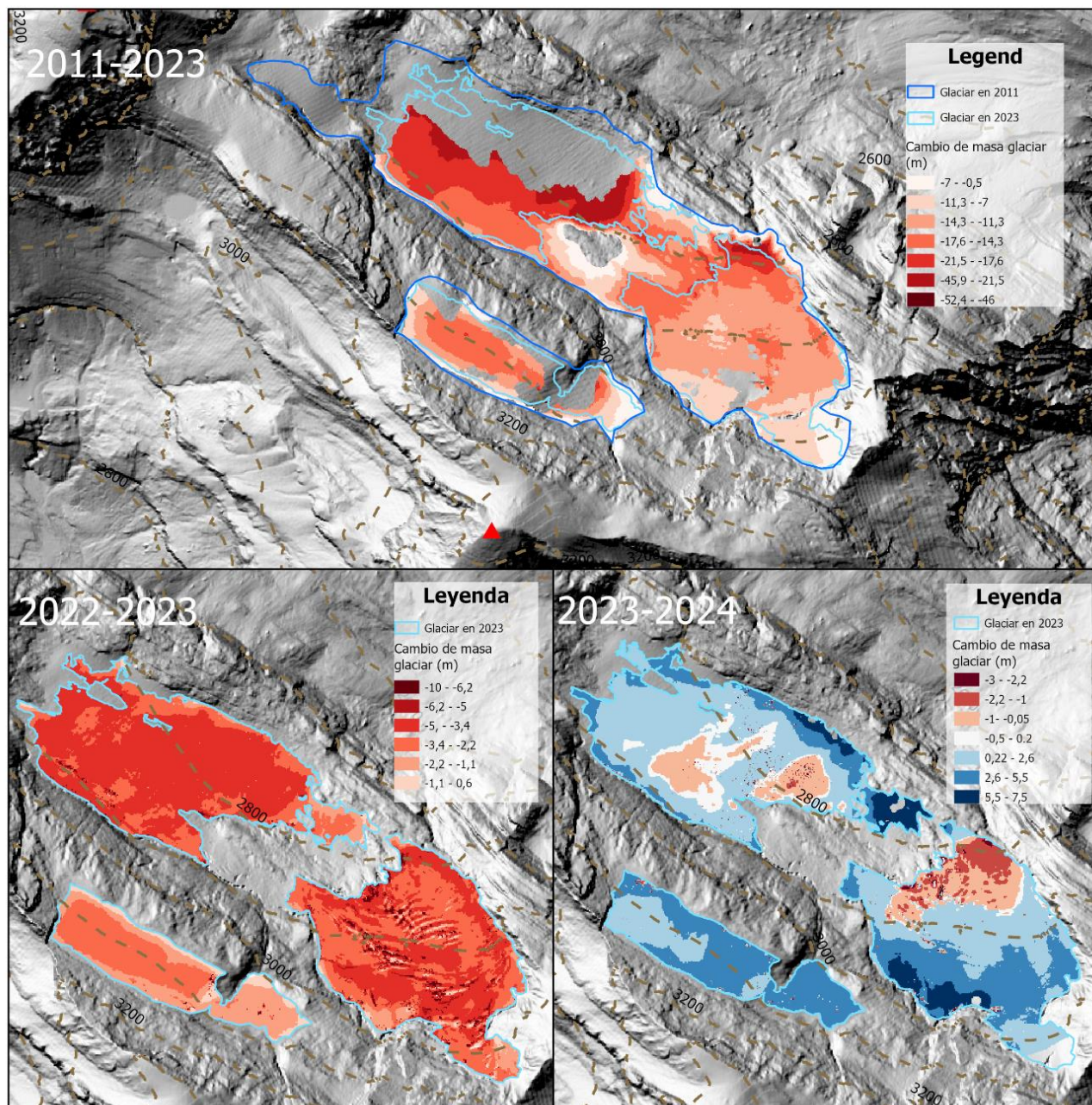
En este capítulo se muestran los resultados obtenidos del análisis del glaciar de Monte Perdido de manera estructurada en diferentes periodos temporales y técnicas de adquisición de datos.

En el primer periodo se incluyen las comparaciones temporales medidas mediante el TLS para los intervalos 2011-2016, 2011-2020, 2016-2024, así como los resultados para años específicos con condiciones extremas de mucha o poca nieve, 2013-2014 y 2014-2015.

El segundo periodo es el relativo a las medidas obtenidas con UAV. Se obtuvieron resultados para el 2021-2022, 2022-2023 caracterizados por la poca nieve acumulada en ese periodo, 2023-2024 periodo con mucha nieve acumulada y un análisis de un periodo más amplio 2021-2024.

## VARIACIÓN EN LA MASA GLACIAR

Variación en la masa del glaciar del Monte Perdido en los periodos:  
2011-2023, 2022-2023 y 2023-2024



Variación de masa glaciar, comparación anual  
Datos previos al 2022 obtenidos TLS (Terrestrial Laser Scanning) y con Dron a partir del 2022,  
luego procesados en Cloud Compare (m3c2 distance). Realizado en ArcGIS Pro

Proyección: ETRS89 / UTM Zona 30N

Fuente de datos: Instituto Pirenaico de Ecología

Realizado por Pablo del Olmo Díaz

0 0.2 0.4 0.8  
Kilometers



FIGURA 11. Evolución de la masa de hielo en diferentes periodos. Elaboración propia a partir de datos IPE.

Por último, se analiza el periodo más extenso disponible, desde 2011 hasta 2024 para proporcionar una idea general del cambio glaciar en toda la ventana temporal analizada.

En primer lugar, se calculan unos estadísticos básicos para conocer la pérdida media de hielo registrado, así como la variabilidad espacial a través de la desviación estándar, también se añaden los valores máximos y mínimos registrados en cada intervalo (tabla 1).

	Periodo	Media	Desviación std	Mediana	p5 (mínimo)	p95 (máximo)
TLS	2011-2016	-4.70	2.28	-4.36	-8.75	-1.44
	2011-2020	-7.31	4.09	-7.13	-14.08	-1.41
TLS/UAV	2016-2024	-6.47	5.38	-5.78	-16	1.83
TLS	2013-2014	-0.04	0.72	0.02	-1.62	0.99
	2014-2015	-2.09	0.74	-2.05	-3.2	-1.02
UAV	2021-2022	-3.97	0.78	-3.86	-5.5	-2.91
	2022-2023	-1.18	1.69	-1.46	-4.24	0.15
	2023-2024	1.31	1.49	0.96	-0.34	4.28
	2021-2024	-0.61	3.08	0.88	-6.25	2.51
TLS/UAV	2011-2024	-7.04	8.29	-5.88	-22.38	2.10
TLS/UAV	2011-2023	-13.62	4.79	-13.68	-21.49	-4.79

Tabla 1. Resumen estadístico de la pérdida de hielo (m) para cada periodo analizado. Se incluyen los percentiles 5 y 95 para evitar los *outliers*.

Con el objetivo de evaluar los resultados del modelo XGBoost obtenidos en los diferentes intervalos temporales analizados se presentan a continuación (tabla 2) las estadísticas resultantes para cada uno de ellos. Con estas métricas se puede comparar la calidad predictiva del modelo, principalmente destacando el error cuadrático medio (RMSE) y el coeficiente de determinación ( $R^2$ ).

	Periodo	RMSE	$R^2$	MAE	MAPE	NSE
TLS	2011-2016	0.80	0.86	0.53	25.4	0.86
	2011-2020	1.21	0.89	0.74	29.41	0.89
TLS/UAV	2016-2024	1.12	0.96	0.70	10.58	0.96
TLS	2013-2014	0.23	0.88	0.16	25.4	0.88
	2014-2015	0.52	0.70	0.32	21.2	0.70
UAV	2021-2022	0.61	0.83	0.36	7.82	0.83
	2023-2024	0.80	0.83	0.47	198.31	0.85
	2021-2024	0.75	0.91	0.45	44.01	0.91
TLS/UAV	2011-2024	1.12	0.96	0.70	10.58	0.96
TLS/UAV	2011-2023	2.38	0.85	1.28	7.19	0.85

Tabla 2. Métricas estadísticas del modelo XGBoost para los distintos periodos analizados.

## 5.1 Periodo TLS

### 5.1.1 TLS 2011-2016

Durante el periodo 2011-2016 se registró una pérdida media de 4.7 metros de espesor de hielo. El modelo mostró un RMSE de 0.82 y  $R^2$  de 0.87, logrando un buen ajuste entre los valores predichos y observados.

El análisis SHAP muestra que el TPI es la variable más influyente en el modelo con un 19,4%, seguido de la radiación potencial (15,8%), el espesor de nieve (13,7%) y el espesor de hielo (12,4%), mientras que la variable con menor influencia en el modelo fue la pendiente con el 4.82% (figura 12).

Los gráficos de dependencia parcial (PDP) permitieron identificar como la pérdida de hielo glaciar era superior en zonas con valores intermedios de TPI, es decir, zonas de ladera entre crestas y depresiones. En cuanto a la radiación potencial, el grafico PDP muestra un incremento de las pérdidas cuando se alcanzan los valores superiores a 1000 Wh/m<sup>2</sup>. Por otro lado, los valores más altos de pérdida se localizaron en espesores de hielo intermedios de entre 10 y 15 metros, mientras que en zonas con poco espesor menores de 5 metros y en zonas de hielo grueso, más de 25 metros, la pérdida fue menor. Además, el espesor de nieve tiene un efecto protector, disminuyendo de manera significativa el retroceso glaciar a medida que aumentaba el espesor de nieve registrado.

En el caso de la pendiente, y los índices de viento (Sx) no se identifican patrones que puedan ser consistentes, por lo que su contribución pueda ser dependiente de otros factores topográficos locales.

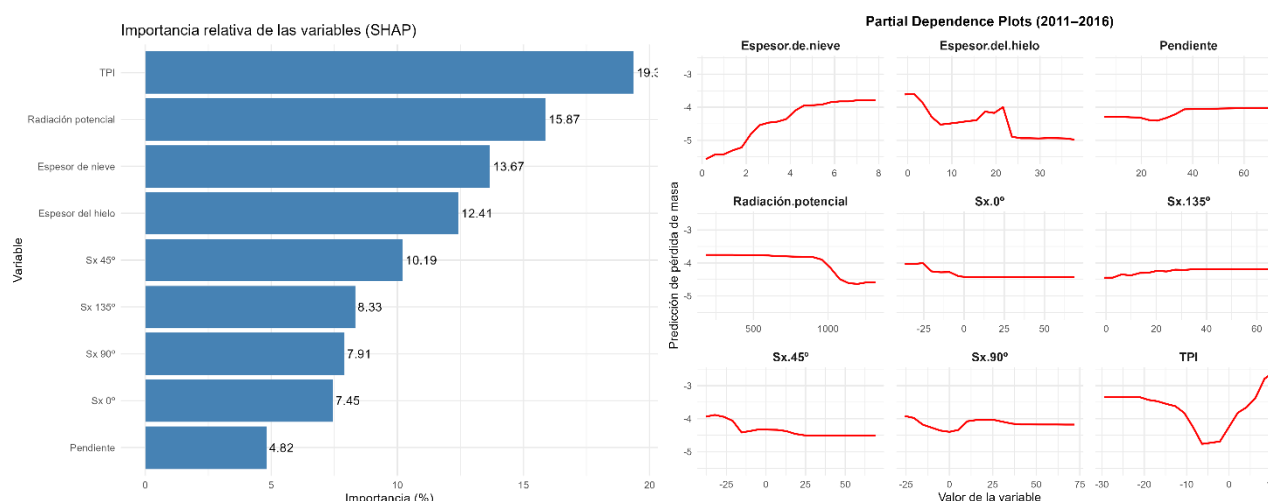


Figura 12. Gráficos SHAP y PDP del modelo XGBoost para el periodo 2011–2016. Elaboración propia

### 5.1.2 TLS 2011-2020

Durante este periodo el modelo mostro un buen rendimiento con un RMSE de 1.21 y R<sup>2</sup> de 0.89. El análisis SHAP mostro que el TPI fue de nuevo la variable con más peso en el modelo con una importancia notable del TPI con un 26.3%, seguido del espesor de hielo (17.95%) y espesor de nieve (16.38%) (figura 13).

De nuevo se observó que las mayores pérdidas se registraron en las zonas de transición entre las zonas deprimidas y las elevadas, según el grafico TPI. Por otro lado, en el caso del espesor de hielo, las mayores pérdidas se reconocen en espesores de entre 10 y 15 metros, mientras que a partir de los 20m se estabilizan. Por parte del espesor de nieve, la relación es inversa, un menor espesor de nieve supone unos niveles superiores de pérdida, lo que podría indicar que la nieve funciona como una capa protectora del hielo. Respecto a la radiación, se puede apreciar un ligero aumento de las pérdidas justo antes de llegar a los 1000 Wh/m<sup>2</sup>, aunque de una manera más atenuada que en el periodo anterior (figura 13).



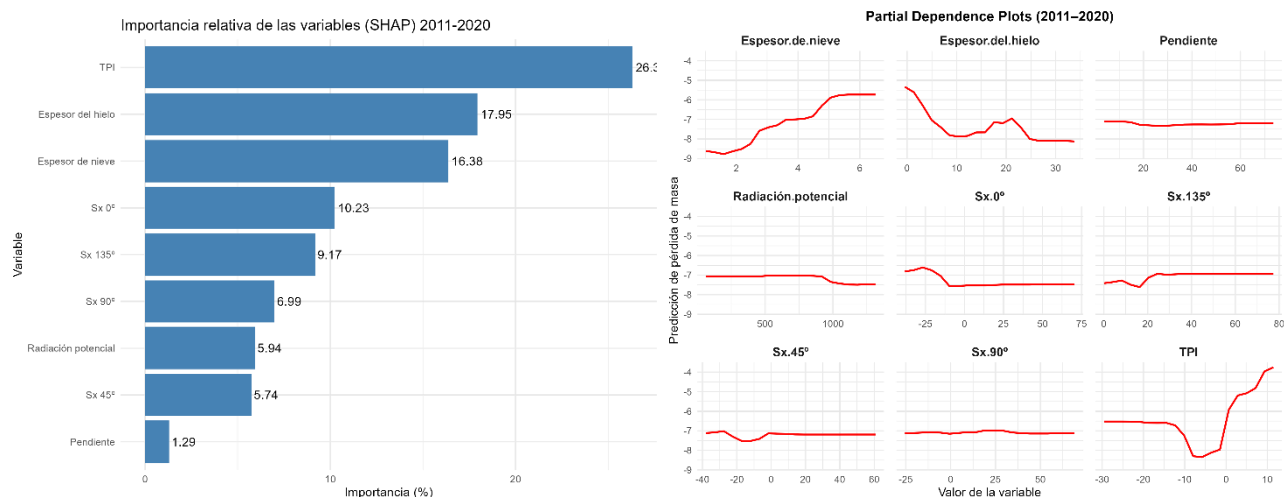


Figura 13. Gráficos SHAP y PDP del modelo XGBoost para el periodo 2011–2020. Elaboración propia

### 5.1.3 TLS 2013-2014

El periodo 2013-2014, marcado por unos niveles de acumulación nival superiores a lo habitual, presentó un comportamiento diferente a los periodos anteriormente mostrados. En este caso, el espesor de nieve fue la variable más influyente según el análisis SHAP (figura 14) presentando un 29 %, seguido del TPI (18.3%) y el espesor de hielo (11.88%), mientras que las variables de viento Sx, radiación y pendiente tuvieron una influencia menor.

Los gráficos PDPs (figura 14) reflejan una pérdida glaciara más elevada cuando la acumulación nival es menor, y una estabilización en las pérdidas cuando la acumulación de nieve supera los 4-5 metros. En cambio, el comportamiento del TPI es diferente a los anteriores periodos ya que no se identifica de forma clara una pérdida máxima en valores cercanos a 0 como si ocurría en los otros periodos. Por parte del espesor de hielo, las pérdidas se mantienen estables hasta los 25 metros, pero en mayores profundidades de hielo las pérdidas de acentúan.

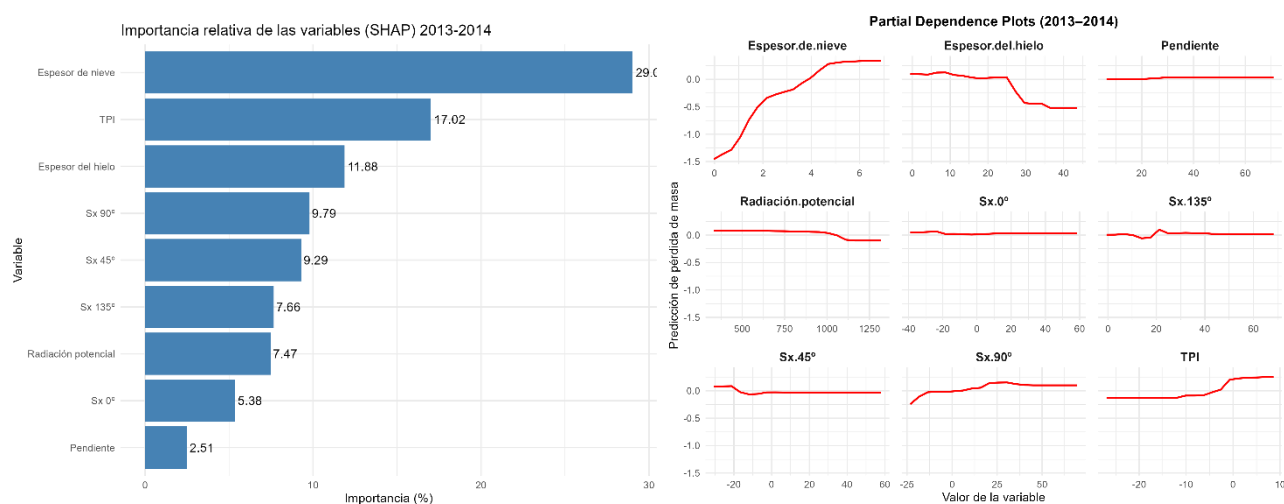


Figura 14. Gráficos SHAP y PDP del modelo XGBoost para el periodo 2013–2014. Elaboración propia

### 5.1.3 TLS 2014-2015

Este periodo está marcado por un menor nivel de acumulación nival, lo que permite conocer el comportamiento del glaciar en un contexto de escasez de nieve. El modelo ofreció unos resultados con menor  $R^2 = 0.71$ , aunque se considera adecuado.

En este caso, el TPI fue la variable con más peso en los resultados (figura 15) con un resultado de 21.5 %, seguido de los índices de viento Sx, que en este caso adquirieron mayor peso que en otros periodos. El espesor de hielo, el espesor de nieve y la radiación potencial aportaron contribuciones intermedias.

Los gráficos PDP reflejan la relación inversa entre el espesor de nieve y la pérdida de hielo glaciar, esta curva presenta menor pendiente que en el caso del periodo anterior. El TPI muestra una trayectoria ascendente constante, a mayor valor de TPI, supone mayor pérdida de masa glaciar, por lo que se registra en este periodo mayores pérdidas en zonas de depresión o vaguadas y menores pérdidas en zonas de cresta.

En el caso de los índices de viento Sx, en este periodo si presentan tendencias más claras que en otros periodos, por ejemplo, el Sx 90°, se observa (figura 15) que en valores comprendidos entre 0° y 25° la pérdida de hielo es superior que en otras direcciones por lo que orientaciones cercanas al este estarían asociadas con una mayor pérdida.

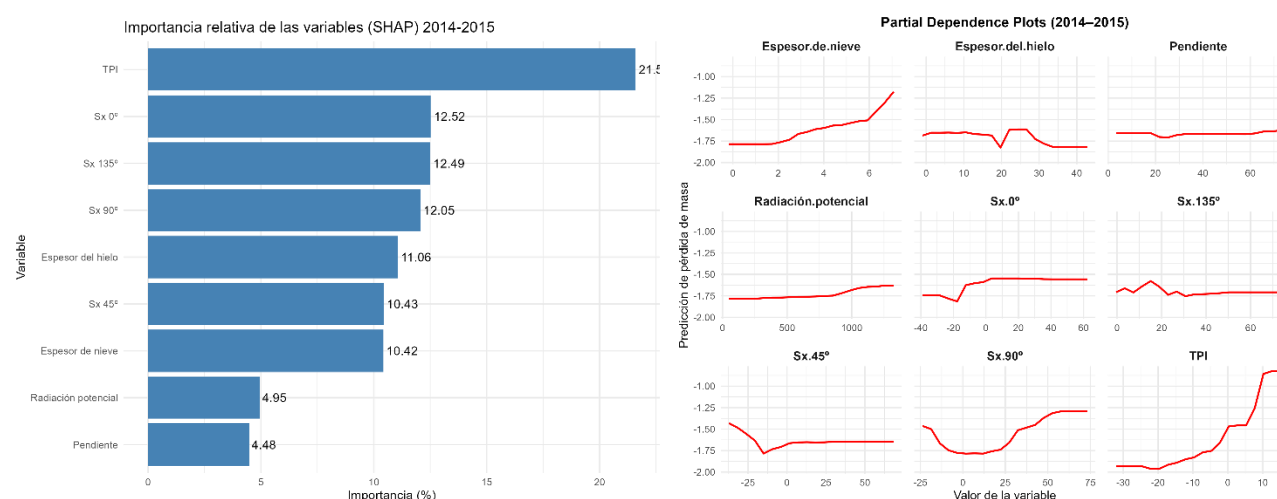


Figura 15. Gráficos SHAP y PDP del modelo XGBoost para el periodo 2014–2015. Elaboración propia.

## 5.2 Periodo UAV

### 5.2.1 UAV 2021-2022

Este periodo obtenido mediante técnicas de vuelo de UAV, esta caracterizado por la baja acumulación nival. A pesar de que la fuente de datos sea diferente a la de periodos anteriores, el modelo obtiene buenos resultados (RMSE de 0.61 y un  $R^2$  de 0.83).

El análisis SHAP (figura 14) revela que la radiación potencial es la variable con mayor peso predictivo con el 17.1% seguida de los valores Sx 135°, Sx 0° y TPI, mientras que el espesor de nieve y el espesor de hielo pierden peso en este modelo.

En el gráfico PDPs (figura 16) se observa como las pérdidas del glaciar se reducen al aumentar la capa de hielo glaciar, mientras que la capa de nieve muestra una ligera tendencia que indica que la masa de hielo disminuye al aumentar el espesor de nieve registrado en este rango. Por parte de la radiación potencial, esta se mantiene neutra hasta alcanzar valores superiores a los 1000 Wh/m².

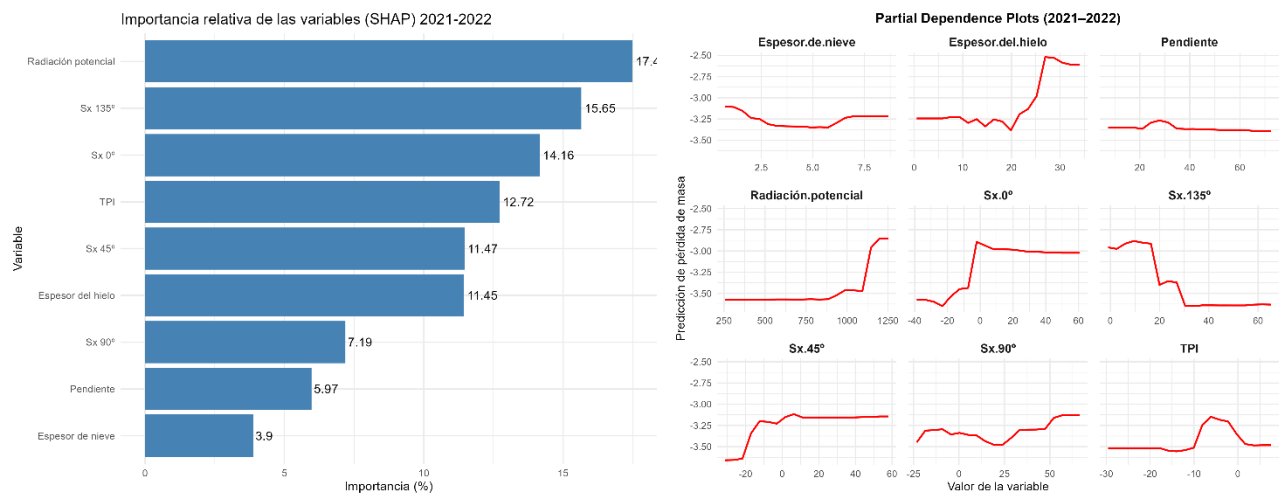


Figura 16. Gráficos SHAP y PDP del modelo XGBoost para el periodo 2021–2022. Elaboración propia

## 5.2.2 UAV 2023-2024

En este periodo caracterizado por una alta acumulación nival el modelo presentó un correcto rendimiento con  $R^2$  de 0.85 y un RMSE de 0.81 demostrando una adecuada capacidad predictiva.

El gráfico SHAP (figura 17) muestra que la variable TPI es el factor más determinante con el 20.6% seguido de los índices de viento 90° y 0°. Los gráficos de dependencia parcial, en este caso, muestran que las mayores pérdidas se recogen en los niveles comprendidos entre -10 y +10 lo que se corresponde con las zonas de transición entre vaguadas y crestas. En cuanto al espesor de nieve, se observa una relación inversa clara sobre todo a partir de los 5m, donde a mayor espesor de nieve se observan mayores pérdidas, esto es algo que es contradictorio con lo observado en la mayoría de los otros periodos. Por lo tanto, en este periodo las mayores pérdidas se dieron en zonas de mayor acumulación. Por parte de la variable Sx 90° se observa claramente como las pérdidas máximas se dan cuando el valor es 0, por lo que las zonas más expuestas al viento del este contribuyen a una mayor ablación, al menos en este periodo.

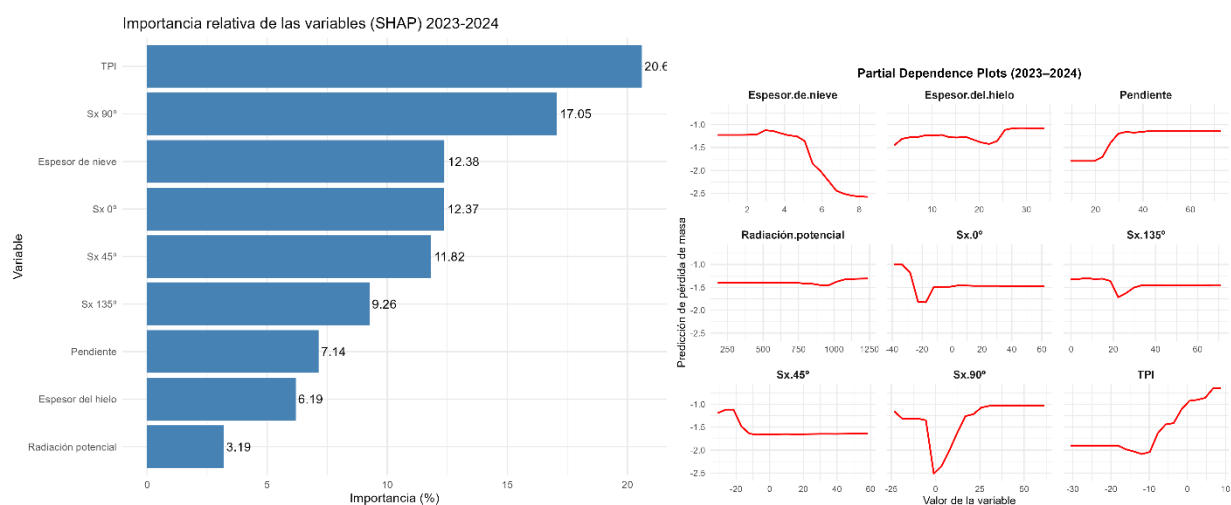


Figura 17. Gráficos SHAP y PDP del modelo XGBoost para el periodo 2023–2024. Elaboración propia



### 5.2.3 UAV 2021-2024

A continuación, se muestran los resultados en el periodo que comprende el uso de UAV para el seguimiento de la evolución del glaciar. El modelo volvió a presentar un buen rendimiento predictivo con RMSE de 0.75 y  $R^2 = 0.91$ .

En la figura 16 se muestra el gráfico SHAP que muestra que la variable TPI es la más relevante del análisis con el 17.3% seguido por los índices de viento con pesos comprendidos entre 15% y 10%. Variables como el espesor de nieve, hielo y pendiente tiene una importancia menor siendo la variable con menos importancia en el modelo la radiación potencial solar. En el gráfico PDP (figura 18), mostrado a la derecha del gráfico 16, se observa como en el TPI las mayores pérdidas del glaciar se dan en zonas con valores negativos en TPI, especialmente por debajo de -10, correspondiendo con zonas de vaguada, sin embargo, en este periodo en zonas de cresta se observa una menor tasa de pérdida de masa glaciar. Por parte del espesor de nieve, las mayores pérdidas se dan en zonas de poco espesor de nieve, lo que podría indicar el factor protector de la nieve. Por parte de la pendiente las pérdidas se acentúan en zonas de pendiente mayor a 30° lo que podría indicar que en pendientes grandes el glaciar no logra retener la masa de hielo.

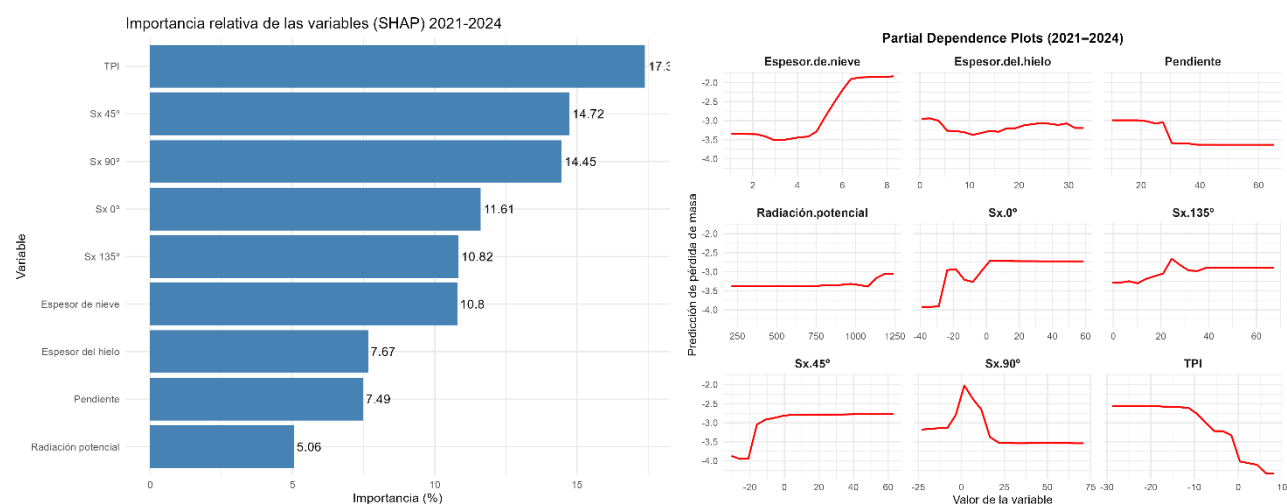


Figura 18. Gráficos SHAP y PDP del modelo XGBoost para el periodo 2021–2024. Elaboración propia

### 5.3 Comparación global

En el análisis del periodo completo, desde 2011 hasta 2023, se obtuvo mediante la comparación de datos de TLS (2011) y UAV (2023), lo que generó un modelo de cambio glaciar continuo. El modelo presentó un comportamiento aceptable con un  $R^2$  de 0.85 y un RMSE de 2.38, es cierto que el valor del error es superior a lo registrado en otros periodos más reducidos.

El análisis SHAP (figura 19) indica que la variable con más peso en este periodo es el índice de viento con dirección 45° con un 19% seguido por el espesor de hielo, el TPI y el índice Sx 135°. La radiación potencial y el espesor de nieve ocuparon lugares más bajos en la jerarquía de importancia relativa, siendo la pendiente de nuevo la variable con menor peso.

Los gráficos de dependencia parcial muestran que en zonas donde el espesor de hielo es bajo, las pérdidas son mayores, por lo que las pérdidas de masa glaciar son menores según aumentan los niveles de hielo acumulado. Por parte del TPI, se observa que cuando esta variable tiene valores de entre -10

y 0, las pérdidas son más acentuadas, lo que asocia a zonas de depresión o vaguadas. En cuanto al espesor de nieve, no se observa una clara tendencia, sin embargo, si se puede apreciar como las pérdidas son menores a medida que aumenta el espesor de nieve. En relación con los índices de viento, destaca la importancia de la dirección 45°, muestra un claro aumento de la pérdida en la dirección que coincide con el glaciar lo que puede suponer que la incidencia directa del viento en el glaciar suponga un aumento de la fusión o causar una menor acumulación de nieve.

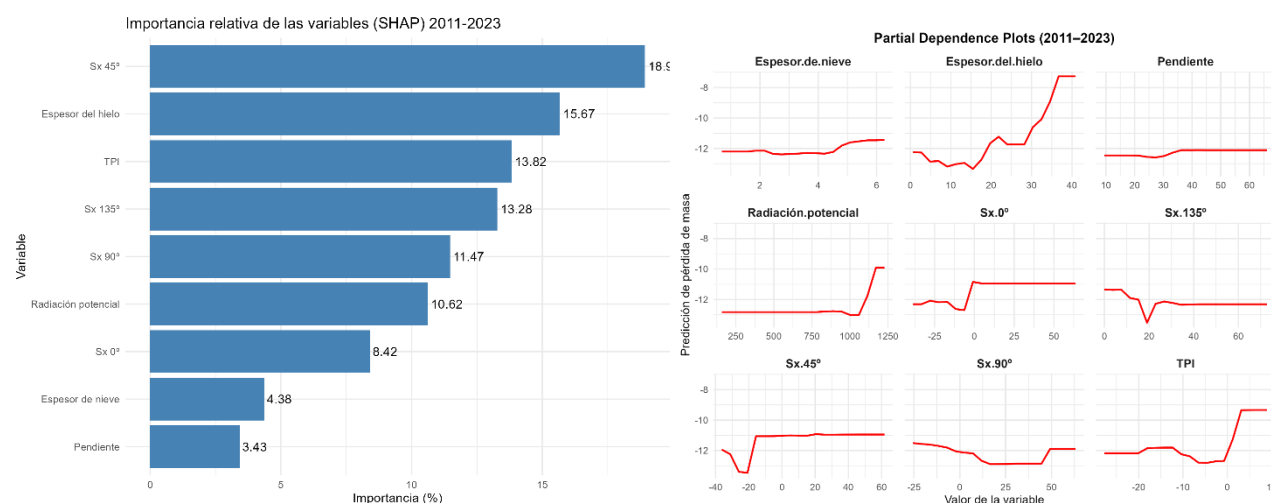


Figura 19. Gráficos SHAP y PDP del modelo XGBoost para el periodo 2011–2023. Elaboración propia

## 6. DISCUSIÓN

Gracias al análisis multitemporal mediante datos TLS y UAV y la posterior modelización estadística con XGBoost, se han obtenido resultados que permiten identificar algunos condicionantes de la pérdida de hielo en el glaciar en el periodo 2011-2023.

Uno de los aspectos relevantes observados en los resultados es que a partir de los gráficos PDPs, se puede identificar un posible umbral energético en la variable de radiación potencial solar. De hecho, en varios periodos analizados, especialmente en registros tomados con TLS como en 2011-2016, se aprecia un incremento de las tasas de pérdidas glaciar cuando los valores de radiación solar potencial supera los 10.000 Wh/m<sup>2</sup>. Este resultado coincide con lo señalado por Revuelto et al. (2020), donde se indica que la radiación solar es uno de los principales controladores de la ablación en los glaciares pirenaicos, condicionando la respuesta diferencial de las superficies de hielo en función de la orientación y el sombreado topográfico. Pero, por el contrario, en los datos obtenidos con UAV, la tendencia es diferente, donde valores elevados de radiación no provocan una pérdida mayor, lo que puede sugerir que a escalas cortas temporales influyen también otros factores, lo que va en línea con lo indicado por Vidaller et al. (2021) al analizar la sensibilidad espacial de la fusión en microcuencas nivales.

Por otra parte, el Índice de Posición Topográfica (TPI) se sitúa como el factor más determinante en la pérdida de hielo en la mayoría de los periodos analizados, lo que muestra la importancia de la morfología del terreno en la dinámica de acumulación y fusión de la nieve y hielo glaciar. Los resultados muestran mayores tasas de pérdida en zonas intermedias ubicadas entre zonas elevadas de escarpes y zonas deprimidas en las laderas, donde es probable que durante los meses de invierno estas zonas sean las que más acumulación nival reciban y posteriormente sufran los procesos de fusión de una manera más intensa debido a la tendencia general de pérdida de hielo glaciar. Esto ha sido ya señalado en trabajos contrastados sobre glaciares en los Pirineos y Alpes donde la topografía ejerce

un papel controlador en la persistencia de la nieve (López-Moreno et al., 2016; Grunewald & Scheithauer, 2010).

Otro factor importante es el espesor de nieve acumulado, esta variable funciona de manera general como un factor que protege el hielo glaciar, actuando como una capa que logra reducir la radiación y las condiciones térmicas adversas que provocan la fusión del hielo. Si embargo, esta relación no se presenta de manera consistente en todos los periodos analizados. Por ejemplo, en periodos cortos marcados por altos niveles de acumulación nival como 2023-2024, el comportamiento observado tiene un comportamiento diferente, identificándose mayores pérdidas en zonas con espesores elevados de nieve. Esto podría deberse a que episodios excepcionales como el mencionado, donde una elevada acumulación ocurre tras un periodo de baja acumulación nival y por ende una alta fusión glaciar, lo que modifica la dinámica habitual que tiene lugar en periodos más prolongados.

La exposición al viento analizada mediante el factor  $S_x$ , muestra también un papel relevante en este análisis glaciar. Cuando el viento incide de manera directa frente al glaciar, es decir el valor de  $S_x$  es cercano a 0, se observan mayores pérdidas, esto es especialmente destacable en el periodo global y en el periodo 2023-2024. Esto podría justificarse porque cuando el viento incide de manera frontal, redistribuye la nieve acumulada y reduce la capacidad protectora de la nieve lo que incrementa la fusión (Damseaux et al. 2025)

De manera general, el modelo estadístico elaborado ha sido capaz de identificar los patrones entre las variables topográficas, climáticas y la pérdida glaciar. No obstante, algunos periodos tuvieron resultados estadísticos de baja calidad por lo que fueron desestimados como 2022-2023, dando prioridad aquellos con una robustez adecuada. De la misma manera, algunas variables fueron desestimadas de ser añadidas en el modelo, ya que tras revisar los resultados de los gráficos SHAP. Las variables de índice de humedad topográfica (TWI), orientación y variables derivadas del programa Cloud Compare (Omnivariance, Sphericity, Planarity y Verticality) presentaron de manera sistemática valores bajos en el gráfico SHAP, por lo que se desestimaron del modelo final por su bajo peso y su posible redundancia con las demás variables. Cabe recordar que aparece en el Anexo de este trabajo el procedimiento desde el procesado de las variables hasta la interpretación con las técnicas SHAP y PDP.

Finalmente es adecuado mencionar algunas limitaciones presentadas a lo largo del estudio, por ejemplo, el dato espesor de nieve y hielo estaban disponibles solo en ciertos años, concretamente 2014, 2015 y 2017 para el caso del espesor de nieve y 2016 y 2023 en el caso de espesor de hielo. Por lo tanto, en algunos casos se realizaron interpolaciones lo que puede añadir incertidumbre.

## 7. CONCLUSIONES

Aunque ya existan numerosos estudios previos que trabajan a partir de los datos procedentes de la monitorización de los glaciares pirenaicos, especialmente el glaciar de Monte Perdido, con TLS y más recientemente con UAV, son pocos los trabajos que se han realizado combinando ambos tipos de datos con un enfoque de aprendizaje automático como XGBoost para analizar conjuntamente los condicionantes topográficos y climáticos de la ablación glaciar. Por tanto, este trabajo se convierte en una aportación novedosa al campo del análisis del glaciar de Monte Perdido al combinar la teledetección LiDAR con el modelado estadístico avanzado.

Este Trabajo de Fin de Máster demuestra que la combinación de nubes de puntos multitemporales obtenidas mediante el escáner láser terrestre (TLS) y vehículos aéreos no tripulados (UAV), unida al cálculo de cambios con M3C2 y el posterior análisis estadístico con XGBoost, es una metodología

válida para describir la pérdida de hielo glaciar de Monte Perdido entre el 2011 y el 2023. Los resultados muestran en primer lugar que en las series TLS la ablación se dispara cuando la radiación potencial solar rebasa los 10.000 Wh/m<sup>2</sup>, mientras que los registros con UAV no se aprecia tal comportamiento, lo que es un indicador de la influencia que provoca la fuente de datos en el resultado final. Asimismo, el factor más determinante en la mayoría de los casos es el TPI, indicando que en las zonas de transición entre crestas y vaguadas se concentran las mayores pérdidas quizá al combinar el hecho de que son lugares que reciben altas acumulaciones de nieve y una fuerte fusión en verano. Otro factor observado es el factor protector de la nieve sobre el hielo glaciar donde normalmente a mayor espesor de nieve acumulado, menor es el nivel de deshielo glaciar. Además, en algunos periodos se identifica un rango crítico de espesor de hielo en torno a los 10-15 metros donde la ablación alcanza su máximo y en cuando al viento, el viento que incide frontalmente al glaciar favorece la redistribución de la nieve y el aumento de la fusión del hielo. Concretamente, en el año 2023-2024, caracterizado por una acumulación excepcional tras un año de fuerte ablación, se registraron pérdidas notables incluso donde el espesor nival era alto, lo que muestra el complejo que es caracterizar estos procesos.

En cuanto al modelo XGBoost creado, ha demostrado un buen comportamiento con valores de coeficientes de determinación por encima del 0,80 y errores RMSE ajustados. Los periodos con métricas pobres se descartaron para evitar conclusiones equivocadas y se eliminaron del modelo variables correlacionadas entre sí y con baja capacidad explicativa con el fin de optimizar el modelo y reducir variables redundantes. Cabe destacar también las limitaciones en la disponibilidad de datos en los espesores de nieve y hielo añadidos al modelo, así como posibles errores en la alineación entre las nubes de puntos TLS/UAV o incertidumbres procedentes de la interpolación de espesores de nieve y hielo

Este estudio ha demostrado una herramienta capaz de analizar a partir de la monitorización continua del glaciar de Monte Perdido la relación de las variables climáticas y topográficas en el proceso de pérdida de hielo. Con ello, se contribuye a una mejor comprensión del retroceso glaciar en entornos sensibles como los Pirineos y se sientan las bases para el desarrollo de estrategias más eficaces de seguimiento, predicción y adaptación frente al cambio climático en zonas de alta montaña.

Futuras líneas de investigación pueden centrarse en incorporar nuevas variables ya sean climáticas, topográficas o de otro tipo, refinar el modelo estadístico con mayor profundidad de análisis, aumentar la frecuencia en la extracción de variables como la nieve o el hielo o aplicar este modelo a otros glaciares de similares características. En conjunto, el estudio confirma que la dinámica del glaciar de Monte Perdido está gobernada por una interacción compleja de factores topoclimáticos (morfología del terreno, radiación, nieve, viento y espesor de hielo) y demuestra la utilidad de enfoques para anticipar su evolución en un escenario de cambio climático acelerado.

## **8. ÍNDICE DE ABREVIATURAS**

PDP. Partial Dependence Plots

MDT. Modelo Digital del Terreno

UAV. Vehículo Aéreo no Tripulado / Unmanned Aerial Vehicle

TLS. Terrestrial Laser Scanner / Escáner láser terrestre

TPI. Topographic Position Index

## 9. ÍNDICE DE TABLAS Y FIGURAS

Figuras	Página
Figura 1 Fotografías del glaciar de Monte Perdido en diferentes fechas	5
Figura 2. Esquema de la evolución del glaciar en diferentes periodos históricos	6
Figura 3. Zona de estudio	9
Figura 4. Esquema de funcionamiento del TLS	10
Figura 5. Esquema metodológico del trabajo	12
Figura 6. Representación del funcionamiento del algoritmo M3C2	14
Figura 7. Ejemplos de cálculos M3C2	14
Figura 8. Representación de variables topográficas incluidas en el modelo estadístico.	16
Figura 9. Representación del índice Sx para la orientación 0° y 135°	17
Figura 10. Distribución espacial del espesor de nieve promedio registrado entre 2014 y 2017	19
Figura 11. Evolución de la masa de hielo en diferentes periodos	24
Figura 12. Gráficos SHAP y PDP del modelo XGBoost para el periodo 2011–2016.	26
Figura 13. Gráficos SHAP y PDP del modelo XGBoost para el periodo 2011–2020	27
Figura 14. Gráficos SHAP y PDP del modelo XGBoost para el periodo 2013–2014.	27
Figura 15. Gráficos SHAP y PDP del modelo XGBoost para el periodo 2014–2015.	28
Figura 16. Gráficos SHAP y PDP del modelo XGBoost para el periodo 2021–2022	29
Figura 17. Gráficos SHAP y PDP del modelo XGBoost para el periodo 2023–2024.	29
Figura 18. Gráficos SHAP y PDP del modelo XGBoost para el periodo 2021–2024.	30
Figura 19. Gráficos SHAP y PDP del modelo XGBoost para el periodo 2011–2023.	31

## 10. FUENTES Y BIBLIOGRAFÍA

Beven, K. J., & Kirkby, M. J. (1979). *A physically based, variable contributing area model of basin hydrology.* *Hydrological Sciences Bulletin*, 24(1), 43–69.  
<https://doi.org/10.1080/02626667909491834>

Brownlee, J. (2016). *XGBoost with Python: Gradient boosted trees with XGBoost and scikit-learn.* *Machine Learning Mastery.*

Centro Nacional de Información Geográfica (CNIG). (2021). *Modelo Digital del Terreno MDT02.* Instituto Geográfico Nacional de España.  
<https://centrodedescargas.cnig.es/CentroDescargas/buscadorCatalogo.do>



- Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. <https://doi.org/10.1145/2939672.2939785>
- Chueca, J., Julián, A., & Saz, M. A. (2005). *The retreat of Pyrenean glaciers from the Little Ice Age: A new approach using digital photogrammetry*. *Geografiska Annaler: Series A, Physical Geography*, 87(3), 201–212. <https://doi.org/10.1111/j.0435-3676.2005.00254.x>
- Clark, W. C., Crutzen, P. J., & Schellnhuber, H. J. (2001). *Science for global sustainability: Toward a new paradigm*. *Nature*, 419, 809–810.
- Crutzen, P. J., & Stoermer, E. F. (2000). *The ‘Anthropocene’*. *Global Change Newsletter*, 41, 17–18.
- CryoPyr-CSIC. (2025). *CryoPyr en el Día Internacional de los Glaciares en la UNESCO: La desaparición de los glaciares del Pirineo es cuestión de pocos años*. Instituto Pirenaico de Ecología – CSIC. <https://cryopyr.csic.es/2025/03/24/cryopyr-en-el-dia-internacional-de-los-glaciares-en-la-unesco-la-desaparicion-de-los-glaciares-del-Pirineo-es-cuestion-de-pocos-anos/>
- Cuffey, K. M., & Paterson, W. S. B. (2010). *The Physics of Glaciers* (4th ed.). Academic Press.
- Damseaux, A., et al. (2025). *Impact of snow thermal conductivity schemes on pan-Arctic permafrost dynamics in the Community Land Model version 5.0*. *The Cryosphere*. <https://doi.org/10.5194/tc-19-1539-2025>
- De Reu, J., Bourgeois, J., Bats, M., Zwertvaegher, A., Gelorini, V., De Smedt, P., & Crombé, P. (2013). *Application of the topographic position index to heterogeneous landscapes*. *Geomorphology*, 186, 39–49. <https://doi.org/10.1016/j.geomorph.2012.12.015>
- Esri. (2024). *Area Solar Radiation (Spatial Analyst)*. En *ArcGIS Pro 3.3 tool reference*. Environmental Systems Research Institute. Recuperado de <https://pro.arcgis.com/es/pro-app/3.3/tool-reference/spatial-analyst/area-solar-radiation.htm>
- Grunewald, K., & Scheithauer, J. (2010). *Europe’s southernmost glaciers: response and adaptation to climate change*. *Journal of Glaciology*, 56(195), 129–142. <https://doi.org/10.3189/002214310791190947>
- Hock, R. (2005). *Glacier melt: A review of processes and their modelling*. *Progress in Physical Geography*, <https://doi.org/10.1191/0309133305pp453ra>
- IPCC. (2019). *Special Report on the Ocean and Cryosphere in a Changing Climate (SROCC)* [H.-O. Pörtner, D.C. Roberts, V. Masson-Delmotte, et al. (Eds.)]. Intergovernmental Panel on Climate Change. <https://www.ipcc.ch/srocc/>
- IPCC. (2021). *Climate Change 2021: The Physical Science Basis*. Cambridge University Press.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning* (2nd ed.). Springer. 10.1007/978-1-0716-1418-1
- James, M. (2017). *Precision maps and 3-D uncertainty-based topographic change detection with structure-from-motion photogrammetry*.
- Julián, A., & Chueca, J. *Cambios recientes en el Glaciar de Monte Perdido: evolución de la superficie glaciar y del volumen de hielo entre 1981 y 1999*. *Revista de Geografía*, 41(2), 155–173. Universidad de Zaragoza.
- Lundberg, S.. M., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*. *Advances in Neural Information Processing Systems*, 30.

<https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>

López-Moreno, J. I., Alonso-González, E., Monserrat, O., Del Río, L. M., Otero, J., Lapazaran, J., Luzi, G., Dematteis, N., Serreta, A., Rico, I., Serrano-Cañadas, E., Bartolomé, M., Moreno, A., Buisan, S., & Revuelto, J. (2019). Ground-based remote-sensing techniques for diagnosis of the current state and recent evolution of the Monte Perdido Glacier, Spanish Pyrenees. *Journal of Glaciology*, 65(249), 85–100. <https://doi.org/10.1017/jog.2018.96>

López-Moreno, J. I., Revuelto, J., Izagirre, E., Alonso-González, E., Vidaller, I., & Bonsoms, J. (2025). No hope for Pyrenean glaciers. *Annals of Glaciology*. Advance online publication. <https://doi.org/10.1017/aog.2025.10015>.

Marshall, S. (2014): *Glacier retreat crosses a line*, *Science*, 342, 872, doi: 10.1371/science.1258584

Ministerio de Cultura y Deporte. (s.f.). Pirineos – Monte Perdido. Gobierno de España. <https://www.cultura.gob.es/cultura/areas/patrimonio/mc/patrimoniomundial/bienes-declarados/por-ano-de-inscripcion/1997/monteperdido.html>

Molnar, C.. (2022). *Interpretable Machine Learning* (2nd ed.). <https://christophm.github.io/interpretable-ml-book/>

Paterson, W.. S. B. (1994). *The Physics of Glaciers* (3rd ed.). Pergamon Press.

Revuelto, J., Izagirre, E., Rico, I., Rio, L., Serrano, E., Vidaller, I., Rojas-Heredia, F., & López-Moreno, J. I. (2025). *The last years of Infiernos Glacier and its transition to a new paraglacial stage*. *Journal of Glaciology*, 71, e35, 1–13. <https://doi.org/10.1017/jog.2025.22>

Revuelto, J., López-Moreno, J. I., Azorin-Molina, C., & Vicente-Serrano, S. M. (2014). *Topographic control of snowpack distribution in a small catchment in the central Spanish Pyrenees: intra- and inter-annual persistence*. *The Cryosphere*, 8(6), 1989–2006. <https://doi.org/10.5194/tc-8-1989-2014>

Revuelto, J., López-Moreno, J. I., & Alonso-González, E. (2021). *Light and shadow in mapping alpine snowpack with unmanned aerial vehicles in the absence of ground control points*. *Water Resources Research*, 57(6), e2020WR028980. <https://doi.org/10.1029/2020WR028980>

Rico, I., Izaguirre, E., Serrano, E. y López-Moreno, J. I. (2017). «Current glacier area in the Pyrenees: an updated assessment 2016». *Pirineos*, 172: e029. Disponible en: <https://Pirineos.revistas.csic.es/index.php/Pirineos/article/view/281/354>

Vidaller, I., López-Moreno, J. I., Revuelto, J., Azorin-Molina, C., Alonso-González, E., Rico, I., & Serrano, E. (2021). *Toward an ice-free mountain range: Demise of Pyrenean glaciers during 2011–2020*. *Geophysical Research Letters*, 48(4), e2020GL090397. <https://doi.org/10.1029/2020GL090397>

Weinmann, M., Jutzi, B., Hinz, S., & Mallet, C. (2017). *Geometric features and their relevance for 3D point cloud classification*. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-1/W1, 157–164. <https://doi.org/10.5194/isprs-annals-IV-1-W1-157-2017>

Winstral, A., Elder, K., & Davis, R.. E. (2002). *Spatial snow modeling of wind-redistributed snow using terrain-based parameters*. *Journal of Hydrometeorology*. [https://journals.ametsoc.org/view/journals/hydr/3/5/1525-7541\\_2002\\_003\\_0524\\_ssmowr\\_2\\_0\\_co\\_2.xml](https://journals.ametsoc.org/view/journals/hydr/3/5/1525-7541_2002_003_0524_ssmowr_2_0_co_2.xml)

Winstral, A., Marks, D., & Gurney, R.. (2013). *Simulating wind-affected snow accumulations at catchment to basin scales*. *Advances in Water Resources*, 55, 64–79. <https://doi.org/10.1016/j.advwatres.2012.08.011>

## ANEXO

A continuación, se muestra el script utilizado para generar y evaluar el modelo.

1. Carga de las variables topográficas y climáticas en formato ráster
2. Adición del contorno del glaciar
3. Recorte y enmascarado de las capas al tamaño del glaciar y asignación el CRS
4. Construcción de un *stack* con las variables y extracción de los valores píxel a una tabla
5. Presentación de una primera matriz de correlación
6. Análisis de la robustez de las correlaciones a través de un muestreo aleatorio
7. Entrenamiento mediante XGBoost (separación 70/30, *early stop*, RMSE/R<sup>2</sup>) y obtención de la importancia de las variables SHAP.
8. Generación de Partial Dependence Plots (PDP)

```
# TFM Monte Perdido
# =====

install.packages("patchwork")
install.packages("corrplot")

# Cargar paquetes
library(terra)
library(sf)
library(dplyr)
library(corrplot)
library(ggplot2)
library(patchwork)
library(raster)
library(tidyr)

##

# 0) Limpiar todo en memoria
rm(list=ls())

gc()
```

```

# 1. rásters

m3c2 <- rast("tu_ruta") # pérdida de masa glaciar (archivo 2011-2016)

pend      <- rast("tu_ruta")      # pendiente
twi       <- rast("tu_ruta")      # índice de humedad
expos     <- rast("tu_ruta")      # exposición
tpi       <- rast("tu_ruta")
rad_pot   <- rast("tu_ruta")

wind_0    <- rast("tu_ruta")
wind_45   <- rast("D:/Desktop/TFM/sx_programa/maxus_45.asc.tif")
wind_90   <- rast("D:/Desktop/TFM/sx_programa/maxus_90.asc.tif")
wind_135  <- rast("D:/Desktop/TFM/sx_programa/maxus_135.asc.tif")

espesor_2016 <- rast("tu_ruta ")
nieve_2017   <- rast("tu_ruta ")

# Dar CRS a la capa de masa (si no lo tiene o para forzarlo a EPSG:25830)
crs(m3c2) <- "EPSG:25830"
crs(m3c2)

# 2. shp del glaciar
glaciar_2016 <- st_read("tu_ruta.shp")
glaciar_2016 <- st_transform(glaciar_2016, crs(m3c2)) # asegurar CRS compatible
glaciar_vect <- vect(glaciar_2016)                  # convertir a SpatVector para
terra
crs(glaciar_vect)

# Aplico CRS del m3c2 al resto de variables (igualando CRS)
crs(pend)      <- crs(glaciar_vect)
crs(twi)       <- crs(glaciar_vect)
crs(expos)     <- crs(glaciar_vect)
crs(tpi)       <- crs(glaciar_vect)
crs(rad_pot)   <- crs(glaciar_vect)

crs(wind_0)    <- crs(glaciar_vect)

```

```

crs(wind_45) <- crs(glaciar_vect)
crs(wind_90) <- crs(glaciar_vect)
crs(wind_135) <- crs(glaciar_vect)

crs(espesor_2016) <- crs(glaciar_vect)
crs(nieve_2017) <- crs(glaciar_vect)

plot(wind_0, main = "SX 0")

# 3. Recortar y enmascarar todos los rásters con glaciar
m3c2_2016_2024_mask <- mask(crop(m3c2, glaciar_vect), glaciar_vect)
pend_mask <- mask(crop(pend, glaciar_vect), glaciar_vect)
twi_mask <- mask(crop(twi, glaciar_vect), glaciar_vect)
expos_mask <- mask(crop(expos, glaciar_vect), glaciar_vect)
tpi_mask <- mask(crop(tpi, glaciar_vect), glaciar_vect)
rad_pot_mask <- mask(crop(rad_pot, glaciar_vect), glaciar_vect)

wind_0_mask <- mask(crop(wind_0, glaciar_vect), glaciar_vect)
wind_45_mask <- mask(crop(wind_45, glaciar_vect), glaciar_vect)
wind_90_mask <- mask(crop(wind_90, glaciar_vect), glaciar_vect)
wind_135_mask <- mask(crop(wind_135, glaciar_vect), glaciar_vect)

espesor_mask <- mask(crop(espesor_2016, glaciar_vect), glaciar_vect)
nieve_mask <- mask(crop(nieve_2017, glaciar_vect), glaciar_vect)

# Rellenar los valores NA de la capa de masa con 0
m3c2_2016_2024_mask[is.na(m3c2_mask)] <- 0

# Resample. reajusta el resto de los rásters para que tengan la misma resolución,
extensión y alineación
pend_mask <- resample(pend_mask, m3c2_mask)
twi_mask <- resample(twi_mask, m3c2_mask)
expos_mask <- resample(expos_mask, m3c2_mask)
tpi_mask <- resample(tpi_mask, m3c2_mask)
rad_pot_mask <- resample(rad_pot_mask, m3c2_mask)

```



```

wind_0_mask <- resample(wind_0_mask, m3c2_mask)
wind_45_mask <- resample(wind_45_mask, m3c2_mask)
wind_90_mask <- resample(wind_90_mask, m3c2_mask)
wind_135_mask <- resample(wind_135_mask, m3c2_mask)

```

```

espesor_mask <- resample(espesor_mask, m3c2_mask)
nieve_mask <- resample(nieve_mask, m3c2_mask)

```

```
# 4. Crear stack de variables
```

```

stack_vars <- c(
  m3c2_mask,
  pend_mask, twi_mask, expos_mask, tpi_mask, rad_pot_mask,
  wind_0_mask, wind_45_mask, wind_90_mask, wind_135_mask,
  espesor_mask, nieve_mask
)

```

```
# 5. Asignar nombres a las variables
```

```

names(stack_vars) <- c(
  "Masa", "pendiente", "TWI", "exposicion", "TPI", "Radiacion",
  "Wind_0", "Wind_45", "Wind_90", "Wind_135", "Espesor_2016", "Nieve"
)

```

```
# 6. Extraer valores por píxel dentro del glaciar de 2011
```

```
df <- as.data.frame(stack_vars, na.rm = TRUE)
```

```
# 7. Calcular matriz de correlación
```

```

cor_matrix_2016_2024 <- cor(df, method = "pearson")
print(cor_matrix_2016_2024)

```

```
# 8. Visualizar la matriz
```

```
corrplot(cor_matrix_2016_2024, method = "circle", type = "upper", tl.col = "black")
```

```
# 9. Ejemplo de gráfico de dispersión: Radiación vs pérdida de masa
```

```

ggplot(df, aes(x = Radiacion, y = Masa)) +
  geom_point(alpha = 0.3) +

```

```

geom_smooth(method = "lm", col = "blue") +
labs(title = "Correlación entre Radiación y pérdida de masa glaciar",
      x = "Radiación", y = "Pérdida de masa (m)")

names(df)

# Extraer solo las correlaciones con "Masa"
corr_con_masa <- cor_matrix_2016_2024["Masa", ]
corr_con_masa <- corr_con_masa[names(corr_con_masa) != "Masa"]

# Convertir a data.frame para ggplot
df_corr <- data.frame(
  Variable = names(corr_con_masa),
  Correlacion = as.numeric(corr_con_masa)
)

# Ordenar por magnitud de correlación
df_corr <- df_corr %>% arrange(desc(abs(Correlacion)))

# barplot
ggplot(df_corr, aes(x = reorder(Variable, Correlacion), y = Correlacion, fill =
Correlacion)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  scale_fill_gradient2(low = "red", mid = "white", high = "blue", midpoint = 0) +
  labs(title = "Correlación de variables con pérdida de masa glaciar",
      x = "Variable", y = "Correlación con Masa") +
  theme_minimal()

# Modelo de regresión múltiple
modelo <- lm(Masa ~ ., data = df)
summary(modelo)

# ANÁLISIS DE CORRELACIÓN CON MUESTREO ALEATORIO

set.seed(123) # para reproducibilidad

```

```

# Número de iteraciones y tamaño de muestra
n_iter <- 1000
n_muestra <- 100

# Variables explicativas (excepto 'Masa')
vars <- names(df)[names(df) != "Masa"]

# Inicializar matriz de resultados
cor_resultados <- matrix(NA, nrow = n_iter, ncol = length(vars))
colnames(cor_resultados) <- vars

# Bucle de iteraciones
for (i in 1:n_iter) {
  muestra <- df[sample(1:nrow(df), n_muestra), ]
  cor_i <- sapply(vars, function(v) cor(muestra$Masa, muestra[[v]], method =
"pearson"))
  cor_resultados[i, ] <- cor_i
}

# Convertir a data.frame
cor_df <- as.data.frame(cor_resultados)

cor_long <- pivot_longer(cor_df, cols = everything(),
                        names_to = "Variable", values_to = "Correlacion")

ggplot(cor_long, aes(x = reorder(Variable, Correlacion, FUN = median), y =
Correlacion)) +
  geom_boxplot(fill = "steelblue", alpha = 0.6) +
  geom_hline(yintercept = 0, linetype = "dashed") +
  coord_flip() +
  labs(title = "Distribución de correlaciones por muestreo aleatorio (n = 1000)",
       x = "Variable explicativa", y = "Correlación con Masa") +
  theme_minimal()

# Calcular medias de correlaciones
cor_mean <- colMeans(cor_df)

```

```

# Convertir a data.frame y ordenar de mayor a menor
cor_mean_df <- data.frame(
  Variable = names(cor_mean),
  Correlacion = round(as.numeric(cor_mean), 3)
)

cor_mean_df <- cor_mean_df[order(-cor_mean_df$Correlacion), ]

# Mostrar tabla ordenada
print(cor_mean_df)

rasters <- list(
  m3c2_2016_2024_mask, pend_mask, twi_mask, expos_mask,
  tpi_mask, rad_pot_mask, wind_0_mask, wind_45_mask, wind_90_mask, wind_135_mask,
  espesor_mask, nieve_mask
)

# Comprobar dimensiones, resolución, CRS, extent
identical_dimensions <- all(sapply(rasters, function(x) all(dim(x) ==
dim(rasters[[1]]))))

identical_resolution <- all(sapply(rasters, function(x) all(res(x) ==
res(rasters[[1]]))))

identical_extent <- all(sapply(rasters, function(x) ext(x) == ext(rasters[[1]])))
identical_crs <- all(sapply(rasters, function(x) crs(x) == crs(rasters[[1]])))

all_ok <- identical_dimensions && identical_resolution && identical_extent &&
identical_crs
print(paste("¿Todos los rásters están correctamente alineados?", all_ok))

### GMS ###

#install.packages("xgboost")
library(xgboost)
library(ggplot2)

# Asegurarse de que 'df' tiene la variable dependiente "Masa"

```

```

dependent_var <- df$Masa
predictors <- df[, setdiff(names(df), "Masa")]

# Convertir a matriz para xgboost
pixel_matrix <- as.matrix(predictors)

# División en entrenamiento y validación (70/30)
set.seed(123)
train_index <- sample(1:nrow(df), 0.7 * nrow(df))
train_data <- pixel_matrix[train_index, ]
train_labels <- dependent_var[train_index]
val_data <- pixel_matrix[-train_index, ]
val_labels <- dependent_var[-train_index]

# Convert to DMatrix for xgboost
dtrain <- xgb.DMatrix(data = train_data, label = train_labels)
dval <- xgb.DMatrix(data = val_data, label = val_labels)

##### Gradient Boosted Model #####
# Parámetros del modelo
params <- list(
  objective = "reg:squarederror", # Regression task
  eta = 0.03, # Learning rate - (tu comentario decía 0.02 si quieres baja aquí)
  max_depth = 5, # Maximum depth of trees
  subsample = 0.7, # Fraction of data to use per iteration
  colsample_bytree = 0.7, # Fraction of features to consider per tree
  min_child_weight = 10, # Minimum sum of instance weight (hessian) needed in a child
  gamma = 1, # Minimum loss reduction required to make a split
  lambda = 1, # L2 regularization
  alpha = 0, # L1 regularization
  eval_metric = "rmse" # Evaluation metric
)

# Create a watchlist to track train and validation performance
watchlist <- list(train = dtrain, val = dval)
xgb_model <- xgb.train(

```



```

params = params,
data = dtrain,
nrounds = 1000, # Maximum number of boosting rounds
watchlist = watchlist,
early_stopping_rounds = 10, # Stop if no improvement for 10 rounds
verbose = 1 # Print progress
)

# Print model summary
print(xgb_model)

##### Make predictions on the validation data #####
# Evaluar el modelo
preds <- predict(xgb_model, newdata = dval)
rmse <- sqrt(mean((val_labels - preds)^2))
cat("Validation RMSE:", round(rmse, 3), "\n")

r2 <- 1 - sum((val_labels - preds)^2) / sum((val_labels - mean(val_labels))^2)
cat("Validation R²:", round(r2, 3), "\n")

### Get feature importance #####
importance_matrix <- xgb.importance(model = xgb_model)

# Print the importance matrix
print(importance_matrix)
# xgb.plot.importance(importance_matrix) # opcional

# ==== SHAP: importancia individual ====
shap_values <- predict(xgb_model, dval, predcontrib = TRUE)

shap_df <- as.data.frame(shap_values[, -ncol(shap_values)])
mean_shap <- colMeans(abs(shap_df))
shap_importance <- data.frame(
  Feature = names(mean_shap),
  Importance = (mean_shap / sum(mean_shap)) * 100
)

```

```

# Gráfico final SHAP

p_shap <- ggplot(shap_importance, aes(x = reorder(Feature, Importance), y =
Importance)) +

  geom_bar(stat = "identity", fill = "steelblue") +

  coord_flip() +

  labs(title = "Importancia relativa de las variables (SHAP)",

       x = "Variable", y = "Importancia (%)") +

  theme_minimal() +

  geom_text(aes(label = round(Importance, 2)), hjust = -0.1)

ggsave("ruta_output",

       plot = p_shap, width = 8, height = 6, dpi = 300)

## evaluate model

evaluate_model <- function(actual, predicted) {

  # Remove any potential NA pairs

  valid_idx <- complete.cases(actual, predicted)

  actual <- actual[valid_idx]

  predicted <- predicted[valid_idx]

  # Residuals

  residuals <- actual - predicted

  # Metrics

  rmse <- sqrt(mean(residuals^2))

  mae <- mean(abs(residuals))

  bias <- mean(predicted - actual)

  # R-squared

  ss_res <- sum((actual - predicted)^2)

  ss_tot <- sum((actual - mean(actual))^2)

  r_squared <- 1 - ss_res / ss_tot

  # MAPE (avoid division by 0)

  mape <- mean(abs((actual - predicted) / ifelse(actual == 0, NA, actual)), na.rm =
TRUE) * 100

```

```

# Nash-Sutcliffe Efficiency
nse <- 1 - sum((actual - predicted)^2) / sum((actual - mean(actual))^2)

# Return as a named list or data frame
data.frame(
  RMSE = rmse,
  MAE = mae,
  R2 = r_squared,
  Bias = bias,
  MAPE_percent = mape,
  NSE = nse
)
}

# After predicting your validation set
preds <- predict(xgb_model, newdata = dval)
resultados <- evaluate_model(val_labels, preds)
resultados$periodo <- "2011-2016 sin CC"
resultados$nieve <- "Nieve 2014"

# Ruta completa del archivo de salida
ruta_resultados <- "ruta_output_resultados"

# Guardar resultados como CSV (añadir si ya existe)
write.table(resultados,
  file = ruta_resultados,
  sep = ";", row.names = FALSE,
  col.names = !file.exists(ruta_resultados),
  append = TRUE)

# BLOQUE ENTRENAMIENTO MODELO + PDP

library(iml)
library(dplyr)
library(xgboost) # ya cargado arriba

```

```

# 1. Asegurar nombres compatibles y consistentes
df <- df %>% rename_with(make.names)

# 2. Preparar datos
dependent_var <- df$Masa
predictors <- df[, setdiff(names(df), "Masa")]
pixel_matrix <- as.matrix(predictors)
colnames(pixel_matrix) <- colnames(predictors)

# 3. Entrenar modelo desde cero (rápido para PDP)
set.seed(123)
dtrain_all <- xgb.DMatrix(data = pixel_matrix, label = dependent_var)
xgb_model_pdp <- xgboost::xgb.train(
  params = list(
    objective = "reg:squarederror",
    eta = 0.03,
    max_depth = 5,
    subsample = 0.7,
    colsample_bytree = 0.7,
    min_child_weight = 10,
    gamma = 1,
    lambda = 1,
    alpha = 0,
    eval_metric = "rmse"
  ),
  data = dtrain_all,
  nrounds = 100,
  verbose = 0
)

# 4. Lista de variables para PDP
variables_pdp <- c(
  "pendiente", "TWI", "exposicion", "TPI", "Radiacion",
  "Wind_0", "Wind_45", "Wind_90", "Wind_135", "Espesor_2016", "Nieve"
)

```

```
# 5. Crear muestra y objeto predictor
```

```
set.seed(42)
```

```
sample_data <- df[sample(nrow(df), min(1000, nrow(df))), ]
```

```
predictor <- iml::Predictor$new(
```

```
  model = xgb_model_pdp,
```

```
  data = sample_data[, setdiff(names(sample_data), "Masa")],
```

```
  y = sample_data$Masa,
```

```
  predict.fun = function(model, newdata) {
```

```
    colnames(newdata) <- colnames(predictors)
```

```
    predict(model, as.matrix(newdata))
```

```
  }
```

```
)
```

```
# 6. Generar PDPs
```

```
pdp_data <- data.frame()
```

```
variables_validas <- c()
```

```
for (feature_name in variables_pdp) {
```

```
  cat("Probando variable:", feature_name, "\n")
```

```
  pdp_result <- FeatureEffect$new(predictor, feature = feature_name, method = "pdp")
```

```
  if (!is.null(pdp_result$results) && nrow(pdp_result$results) > 0) {
```

```
    df_temp <- pdp_result$results %>%
```

```
    rename(X = all_of(feature_name),
```

```
      Predicted = setdiff(names(pdp_result$results), feature_name)[1]) %>%
```

```
    mutate(Feature = feature_name)
```

```
    if (nrow(df_temp) > 0) {
```

```
      pdp_data <- bind_rows(pdp_data, df_temp)
```

```
      variables_validas <- c(variables_validas, feature_name)
```

```
    } else {
```

```
      cat("> Sin datos tras renombrado para:", feature_name, "\n")
```

```
    }
```

```
  } else {
```

```
    cat("> Sin resultados para:", feature_name, "\n")
```

```
  }
```



```
}
```

```
# Diagnóstico
```

```
cat("Variables con datos válidos:", paste(variables_validas, collapse = ", "), "\n")
```

```
cat("Total de puntos en pdp_data:", nrow(pdp_data), "\n")
```

```
# 7. Graficar y guardar solo si hay datos
```

```
if (nrow(pdp_data) > 0) {
```

```
  p <- ggplot(pdp_data, aes(x = X, y = Predicted)) +
```

```
    geom_line(color = "red", size = 1) +
```

```
    facet_wrap(~ Feature, scales = "free_x", nrow = 3) +
```

```
    theme_minimal(base_size = 14) +
```

```
    labs(
```

```
      title = "Partial Dependence Plots (2011-2016, sin CC)",
```

```
      x = "Valor de la variable",
```

```
      y = "Predicción de pérdida de masa"
```

```
    ) +
```

```
    theme(
```

```
      strip.text = element_text(size = 14, face = "bold"),
```

```
      plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
```

```
      panel.spacing = unit(1, "lines")
```

```
    ) +
```

```
    scale_y_continuous(expand = expansion(mult = c(0.05, 0.05)))
```

```
  ggsave(
```

```
    filename = "tu_ruta",
```

```
    plot = p,
```

```
    width = 10,
```

```
    height = 8,
```

```
    dpi = 300
```

```
  )
```

```
  plot(p)
```

```
} else {
```

```
  cat("\n⚠ No se generó ningún gráfico: no hay datos en pdp_data.\n")
```

```
}
```

