Dear Professor Joaquim G. Pinto,

We would like to sincerely thank you for the opportunity to revise our manuscript and for coordinating a careful and constructive review process. The referees provided an extensive and detailed set of comments, and we have undertaken a substantial effort to address each of them thoroughly. The revised version of the manuscript now includes significant clarifications, additional analyses, and an expanded methodological context that have, in our view, considerably strengthened the contribution.

In particular, we have rewritten and expanded Section 2.2, added new experiments related to the emulation of the FWI Fuel Moisture Codes, incorporated additional diagnostics regarding the role of precipitation, and enhanced the interpretability discussion through dedicated XAI analyses. These additions required substantial work, including new computations and figure generation, but we believe they materially improve the manuscript and more clearly situate its scientific scope.

While some reviewer requests extended beyond the intended focus of the study -such as the analysis of several severe wildfire episodes-, we have carefully clarified the boundaries of the manuscript and have complemented the text with concise results that illustrate the physical consistency of the emulators, without deviating from the core scientific objective. We trust that these revisions demonstrate our commitment to addressing the reviewer's concerns in a constructive and rigorous manner.

We hope that the revised manuscript meets the journal's standards and look forward to your consideration.

Sincerely,

The authors

Review of Deep Learning Emulation of Multivariate Climate Indices: A Case Study of the Fire Weather Index in the Iberian Peninsula. Rev2 RC3

We sincerely thank the reviewer for taking the time to review our manuscript and for the constructive comments. We believe their feedback has improved the clarity of the manuscript and overall quality of this work.

The authors investigate the use of deep learning models to emulate the original FWI over the Iberian Peninsula using daily mean inputs. The manuscript has already undergone a first round of review by two reviewers, whose comments I have carefully read and with which I fully agree. A point-by-point reply was provided by the authors, and the comments suggested by the reviewers have now been incorporated into the new version of the manuscript. This was done systematically by the authors and the revised version of the manuscript is an improvement over the previous one.
The manuscript is generally well written and the methodology is sound, although some aspects are not entirely clear. In my opinion, the manuscript can be accepted after minor revision mentioned below.

Main comment: The manuscript clearly lacks a more substantial and critical discussion of the results, and the conclusions section remains overly generic, failing to provide meaningful insights or synthesis. In its current form, the section does not adequately reflect the relevance or implications of the study. My suggestion for the conclusions section is: (1) a very short paragraph stating the problem and how you addressed it in the paper; (2) a point-by-point summary of the main results; (3) a discussion of your results in the context of previous studies, including those conducted in similar regions or using comparable approaches; (4) the limitations of the study; and (5) how your findings could be applied in future research. Some of this information is already present in this section, but in the current form it is difficult to identify and extract the most relevant points, which weakens the overall clarity and impact.

We thank the reviewer for this constructive and insightful comment. We agree that the previous conclusions section was overly generic and did not sufficiently synthesise the relevance and implications of our findings. Following your recommendations, we have fully rewritten and reorganised the Conclusions section to provide a clearer and more informative synthesis of the study.
The revised section now (i) briefly restates the problem and our approach; (ii) summarises the main results in a concise, point-by-point structure; (iii) discusses our findings in the context of previous work on FWI emulation, proxy-based approaches and related studies in Iberia and similar regions; (iv) explicitly outlines the main limitations of our framework; and (v) highlights avenues for future research and potential applications, **including an explicit mention to the on-going climate service development in which the emulation is intended to be implemented, of practical relevance <https://pti-clima.csic.es/>**. In the last point, we also include a brief mention of some preliminary results on FWI component emulation, suggested by another reviewer. We believe that this restructured conclusion improves the clarity, impact and scientific value of the manuscript, and more adequately reflects the contribution of our work.

Therefore, the *Conclusions* section in the new version of the manuscript is the following:

In this study we addressed the challenge of emulating the reference Canadian Fire Weather Index (FWI), which requires instantaneous noon meteorological inputs that are rarely available in observational databases or climate model output. We proposed a deep-learning framework capable of reconstructing the reference FWI from commonly available daily variables and evaluated its performance over the Iberian Peninsula.

Our results show that (i) all emulators reproduce the main spatial and temporal characteristics of the reference FWI, including high-percentile behaviour (critical for extreme fire danger detection); (ii) they consistently outperform the ``traditional'' Proxy FWI –as introduced by Bedia et al 2014 a) for regional climate projections– , particularly in the detection of extreme fire-weather events; (iii) daily mean predictors are sufficient to recover the reference FWI signal with high fidelity; (iv) precipitation can be omitted with limited accuracy loss during the JJAS fire season; and (v) the U-Net and DeepESD architectures offer the best overall balance across metrics. These findings highlight the potential of DL-based emulators to provide accurate fire-weather information when sub-daily inputs are unavailable.

Placed in the context of previous work, our results address a well known problem with FWI reconstruction from model outputs (Herrera et al., 2013), that can also be circumvented through the use of FWI-derived indicators (e.g.: Matteo et al., 2025), posing a challenge in future climate impact studies and introducing additional uncertainty in the resulting projections. Our study extends earlier studies that relied on proxy variables to approximate the noon-LST FWI signal (Bedia et al., 2014a), demonstrating that DL methods can recover the nonlinear relationships between daily predictors and instantaneous fire-weather conditions more effectively than proxy formulations. This improvement is particularly relevant for Mediterranean-type climates, where strong diurnal cycles amplify mismatches between daily means and noon conditions, and for climate-change applications relying on coarse temporal resolution model output.

The study does, however, have limitations. The emulators were trained solely on ERA5-Land, which is not an observational ground truth; also, for simplicity, the analysis is restricted to the generic JJAS fire season (although the simulations are performed for the entire period), even though the fire season period varies across ecoregions within the Iberian Peninsula (see e.g. Bedia et al., 2014b); and we do not investigate individual wildfire events or fuel-specific fire behaviour, which lie outside the intended scope of this study. In addition, the emulators are sensitive to the spatiotemporal characteristics of the training dataset and may require adaptation in environments with different climatic regimes.

Our preliminary experiments on the individual fuel-moisture codes (FFMC, DMC, DC) show that FFMC and DMC can be emulated accurately from daily inputs, while the DC exhibits a stronger dependence on precipitation due to its long-term moisture memory. As this study focuses on the final FWI index, which is most widely used in climate and impact applications, we leave a full component-level assessment to a companion work. Ongoing research is extending this framework to other regions and seasons, incorporating additional environmental predictors, and analysing the physical behaviour of the individual FWI components in greater depth.

Minor comments

1) If I understand correctly, the difference between P1 and P2 lies solely in the inclusion of the directional components of wind. The authors mention several times that this distinction allows them to assess the impact of using the wind speed magnitude versus the wind speed components, which could be relevant for determining fire danger. However, the improvements appear to be minimal, or even worse when incorporating wind direction. Can the authors comment on this?

We thank the referee for this insightful comment. Indeed, the distinction between P1 and P2 lies solely in the inclusion of the zonal and meridional wind components in P2, while P1 only includes wind speed magnitude. As the results indicate, the inclusion of wind components does not lead to an improvement in the emulation skill of the FWI and, in some metrics, slightly degrades performance.

We interpret this result as evidence that, for the purpose of emulating the FWI itself, the wind direction information does not play a key additional role beyond what is already captured by wind speed magnitude. Since the FWI formulation primarily depends on wind speed rather than direction, the magnitude appears to contain most of the relevant information required to reproduce the index. Consequently, the deep learning model is likely able to extract all necessary wind-related signals from wind speed alone.

Moreover, adding the wind components introduces information that is partly redundant with wind speed and potentially noisier in the context of FWI emulation. This redundancy may hinder the learning process and explain the slight degradation observed in P2, despite the model's internal feature selection capabilities. This suggests that, even for deep learning approaches, a careful screening of predictors remains beneficial. This predictor screening was conducted during the present study, as the inclusion of variables from the 850 hPa level in addition to the existing surface predictors did not provide any additional information to the deep learning model. Furthermore, incorporating upper-level variables had a detrimental effect on model performance, leading to worse results compared to using surface predictors alone.

We remark, however, that this finding does not imply that wind direction is unimportant for fire danger assessment more broadly. Wind direction is crucial for fire spread, behavior, and risk management applications. Our result simply indicates that, when the objective is the statistical emulation of the FWI index, the additional directional wind information does not provide added value and may even be counterproductive.

We have added a brief explanation in the new manuscript version at Section 3.4, indicated in L447-L450 of the manuscript: "*This suggests that, for emulating the FWI index itself, wind speed magnitude already captures the relevant wind-related information, and the addition of directional components introduces redundant or noisy predictors that can hamper model performance, despite the internal feature selection of deep learning methods. Consequently, a prior predictor screening may still be necessary to ensure optimal results.*"

2) Another potentially misleading aspect is the use of the original FWI variables in its computation, compared with the predictors used in the emulation models. It would strengthen the manuscript if the authors explicitly mentioned this distinction, either within the main text or at least in the conclusions section.

Before discussing the performance of the DL models across the validation indices, we first highlight how daily aggregation of the input data affects model performance (L310-L326). Figure D1 in Appendix D presents the results from the Dense, DeepESD, and U-Net models trained using 12:00 UTC input variables (temperature, 24-hour accumulated precipitation, relative humidity, and wind speed) to compute the FWI. This sensitivity experiment evaluates the models' ability to learn the transfer function defining the FWI using inputs at 12:00 UTC, consistent with the temporal resolution of the reference index. This analysis complements our other model configurations by isolating the effect of temporal aggregation. Specifically, we compare performance when models are trained with instantaneous inputs (i.e., values at 12:00 UTC and 24-hour precipitation) versus daily mean inputs. This framework allows us to separate the intrinsic biases of each model from the additional error introduced by using daily-aggregated predictors. Figure D1 shows that models trained with instantaneous inputs exhibit lower bias and improved accuracy while maintaining a similar spatial error pattern. This confirms that part of the error observed in experiment P0 (Figure 3) stems from the mismatch in temporal resolution between the predictors and the reference FWI.

3) The manuscript does not include any time series comparisons. For example, for a single grid point the authors could show the reference FWI, P0 versus the emulation models (P1 and P2) for a particular extreme FWI spell in the test data (maybe a particular in 2021 where several wild fires occurred in Iberia). Including such a comparison would help the reader assess how well the models capture the temporal evolution at a local scale and would strengthen the evaluation.

We thank the referee for this comment. The figure attached shows monthly, spatially aggregated time series by climatological region for the reference FWI (FWI Ref), the Proxy FWI, and the deep learning based estimates over the 2018–2021 test period. The Proxy FWI displays noticeable deviations from FWI Ref, including biases in magnitude and a damped seasonal variability. In contrast, the deep learning models consistently improve agreement with FWI Ref, more accurately reproducing both the timing and amplitude of the seasonal cycle as well as short-term variability. These results demonstrate that the deep learning approaches effectively correct the limitations of the Proxy FWI and provide a closer approximation to the reference FWI across different climatological regions. To complete our study with this time series comparisons, we have decided to add this figure to the manuscript as Figure A2. This new figure is located at Appendix A, where climatological regions are described. Therefore, in this appendix section we complement the presentation of the climatological regions with this time series comparison

highlighting the performance of each deep learning model and the FWI proxy in the different regions.

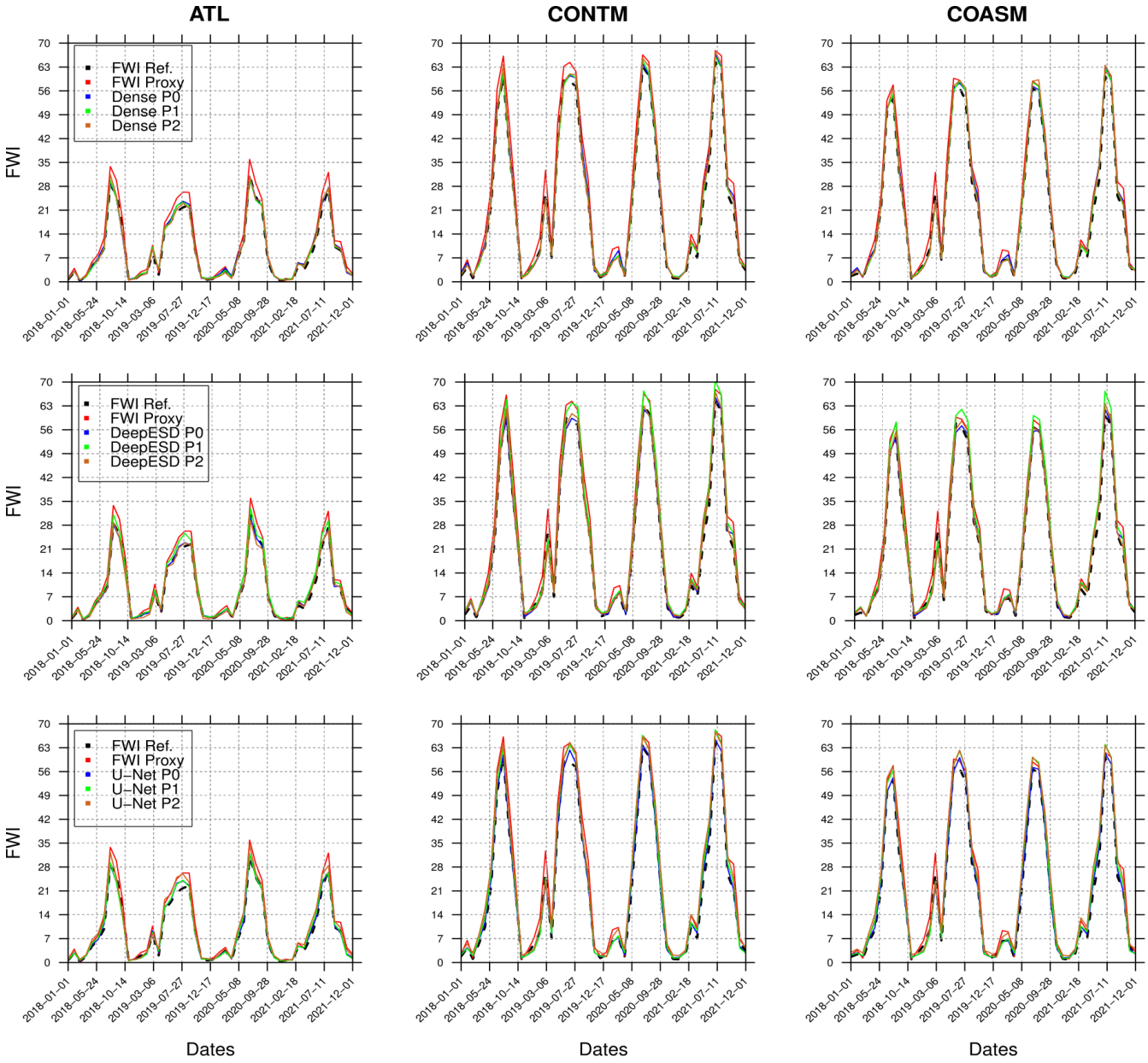

Figure 1: Monthly, spatially aggregated time series by climatological region for the reference FWI (FWI Ref), the Proxy FWI, and the deep learning based estimates over the 2018–2021 test period. The first, second and third row show the results for the Dense, DeepESD and U-Net models respectively.

References

Bedia, J., Herrera, S., Camia, A., Moreno, J. M., and Gutierrez, J. M.: Forest Fire Danger Projections in the Mediterranean using ENSEMBLES Regional Climate Change Scenarios, Climatic Change, 122, 185–199, https://doi.org/10.1007/s10584-013-1005-z, 2014a

Bedia, J., Herrera, S., and Gutierrez, J. M.: Assessing the predictability of fire occurrence and area burned across phytoclimatic regions in Spain, Natural Hazards and Earth System Science, 14, 53–66, https://doi.org/10.5194/nhess-14-53-2014, 2014b.

Herrera, S., Bedia, J., Gutierrez, J. M., Fernandez, J., and Moreno, J. M.: On the projection of future fire danger conditions with various instantaneous/mean-daily data sources, Climatic Change, 118, 827–840, https://doi.org/10.1007/s10584-012-0667-2, 2013.

Matteo, A., Garnés-Morales, G., Moreno, A., Andreia, R., Azorín-Molina, C., Bedia, J., Giuseppe, F. D., Dunn, R. J. H., Herrera, S., Provenzale, A., Quilcaille, Y., Vázquez, M. A. T., and Turco, M.: Challenges in assessing Fire Weather changes in a warming climate, https://doi.org/10.48550/arXiv.2503.01818, arXiv:2503.01818 [physics], 2025

Review of Deep Learning Emulation of Multivariate Climate Indices: A Case Study of the Fire Weather Index in the Iberian Peninsula. Rev2 RC4

Dear Professor C. Gouveia,

We would like to sincerely thank you for the time and care devoted to reviewing our manuscript and for providing such detailed and thoughtful comments. We greatly appreciate the opportunity to clarify the scope of our work and to further strengthen the manuscript in response to your observations. Several of your remarks have helped us refine both the methodological description and the contextual framing of the study, and we have implemented substantial revisions accordingly.

We would like to clarify that the primary scientific objective of this study is the estimation of the FWI for climatic applications, particularly in the context of climate simulations. The computation of FWI requires instantaneous meteorological variables at local noon standard time. However, such sub-daily information is often unavailable in climate data repositories, especially in climate projection datasets. Consequently, the commonly available data do not provide the optimal temporal representation required for a standard FWI calculation. Within this context, the aim of the study is not to validate FWI against observations, but rather to assess whether deep learning models can reproduce a reference FWI more accurately than common proxy approaches using daily mean information to compute the index.

We have also clarified why certain expectations related to operational wildfire forecasting and fire management fall outside the intended scope of this work. The practical value of the proposed emulators lies primarily in their capacity to generate rapid, spatially consistent fire-weather information suitable for climate services, climate-change impact assessments, and large-scale evaluations of natural hazards. This was the intended meaning of the term “operational” in the previous version, which has now been revised to avoid ambiguity. These applications differ fundamentally from short-range, high-resolution meteorological forecasting and detailed fuel or terrain characterization required for real-time firefighting operations. By making this distinction explicit, we aim to more clearly position our contribution within the domain of climate-related natural hazard assessment rather than tactical wildfire suppression.

At the same time, we would like to note that the present manuscript focuses primarily on the emulation of the FWI itself, as it remains the most widely used indicator in climate and impact studies (as noted in the introduction of the paper). Nevertheless, in response to your comments, in the revised manuscript we have included a concise presentation of selected results on the emulation of the FWI sub-indices to demonstrate the internal consistency of our approach at the component level. A comprehensive and dedicated analysis is beyond the main scope of this paper.

In summary, the manuscript has been expanded and revised in three main directions:

(i) a clearer discussion of the role of the FWI components and their relationship to the main index has been incorporated and some new results on FWI components have been included (see Sec. 3.6).

(ii) additional experiments have been included to assess predictor sets with and without precipitation, supported by a more comprehensive evaluation under varying fire-weather conditions; and

(iii) the section on model interpretability has been substantially enhanced through the integration of detailed explainability analyses, providing stronger physical insight into the behaviour of the emulators.
We trust that the revisions presented here address your main concerns and contribute to a clearer and more robust manuscript. A detailed, point-by-point response follows below. The authors' responses are indicated in blue font and interwoven with the referee's comments.

Célia Gouveia, cmgouveia@fc.ul.pt

This manuscript aims explore how deep learning (DL) models can replicate the reference Fire Weather Index (FWI) across the Iberian Peninsula, using only daily input variables from ERA5-Land data. I have several criticisms related with the applicability of the proposed proxy that only emulates FWI, disregarding the other important sub-indices used for operational measures (e.g., FFMC, DC, ISI). In particular, the exclusion of precipitation that as a crucial role for DC was not evaluated during the weather transition regimes. Another important issue is the strong focus of the present work on Deep Learning methods and in Fire Hazard itself. These three issues are key points that should be improved if the authors envisage, as they stated, the use of the present proxy to help spatial planning and management, and adaptation decision-making. Thus, I believe the paper should not be accepted as it is and requires the major revisions listed below before been considered for publication in Natural Hazards and Earth System Sciences journal.

Major suggestions/comments:

FWI Emulation without considering the FWI sub-indices
FWI is included in the Canadian Forest Fire Weather Index System, that is a standard scientific system used to assess wildland fire danger. Specific daily weather observations are used to calculate six numeric components that track fuel moisture and potential fire behavior. In particular, the system is divided into three fuel moisture codes and three fire behavior indices. Higher values across all components indicate greater fire danger. Agencies use these sub-indices to guide daily management decisions, such as pre-positioning firefighting resources and predicting ignition potential. In particular, the Fine Fuel Moisture Code (FFMC) and the Initial Spread Index (ISI) are critical indicators of ease of ignition and flammability rate of spread immediately after ignition. On the other hand, the Drought Code (DC) tracks long-term, seasonal drought effects and is used to predict difficulties in extinguish practices. As the authors claim that they aim to show that "DL emulators can enhance the accessibility and applicability of FWI in impact studies, facilitating a more efficient and widespread use in climate impact studies and decision-making", I do not see how this can be done using only the FWI index. The use of the sub indices is crucial in order to translate the 'indice value' in proxies of weather and fuel

characteristics that are key factors for wildfire spread and propagation. In my opinion, as wildfire expert with experience in early warning advice and considering the importance of wildfire as a natural hazard where the fire forecast is crucial, I strongly suggest to compute proxy for the sub-indices. Otherwise I suppose the exercise done in this paper is more related with AI studies, than to wildfires risk assessment.

We thank the reviewer for this detailed comment and for highlighting the scientific relevance of the individual components of the Canadian Forest Fire Weather Index System. We fully agree that the six sub-indices (FFMC, DMC, DC, ISI, BUI, and FWI) play distinct roles in operational wildfire danger assessment, and that agencies often rely on these components to support short-range decision-making and tactical resource allocation.
At the same time, we would like to clarify that the objective of the present study is not to reproduce the full operational structure of the CFFDRS, nor to provide a forecasting tool for real-time wildfire suppression activities. As now emphasized in the revised manuscript, our goal is specifically to emulate the reference FWI, which is the most widely used indicator in climate-scale applications, impact assessments, and large-area evaluations of natural hazards. These domains, aligned with the scope of Natural Hazards and Earth System Sciences, differ fundamentally from the operational context in which agencies use the sub-indices to guide tactical wildfire management. Operational fire brigades typically rely on short-lead numerical weather prediction, higher-resolution meteorological fields, and localized fuel or terrain data that fall outside the scope of this work.
We fully acknowledge, however, that a deeper understanding of the consistency of the emulators at the sub-index level is scientifically valuable. For this reason, and although it lies beyond the intended scope and size of the present manuscript, we have undertaken an extended analysis of the individual FWI components as part of a broader research effort and included in the revised manuscript a selection of results that illustrate the ability of our deep learning framework to emulate not only the aggregated FWI but also its intermediate sub-indices. These results demonstrate strong internal coherence with the physical structure of the system and support the robustness of the approach. More details on this are provided as an annex to this response for further information.
We hope that this clarification removes any potential misunderstanding regarding the ambition and scope of the study. The manuscript now explicitly states that the purpose of the proposed emulators is to enhance the accessibility of the FWI for climate-service applications and climate-change impact studies (contexts in which the aggregated FWI is the standard metric) rather than to replace the sub-indices used in operational fire-management decision support.

Exclusion of precipitation for FWI emulation

Precipitation directly affects the three primary moisture codes, though each requires a different "effective" amount to overcome surface tension and penetrate the fuel layer. Because fine fuels have a short time lag (~16 hours), a small amount of rain can cause a rapid drop in the FFMC, significantly lowering immediate ignition risk. On the other hand, as the "long-term memory" of

the system, the DC is only affected by significant or sustained rainfall that can penetrate deep, compact organic layers. This means that even during severe events, light may temporarily lower surface ignition risk (FFMC) but fails to penetrate deeper fuel layers. However prolonged periods of rain that saturate both fine fuels and deep fuels can meaningfully reduce the risk during established severe wildfire events. On the other hand, in case of out of high fire season precipitation has an important role that in terms of preparation and mitigation techniques are crucial. In fact, the FFMC and DC values are extremely important to plan management practices and prescribe fires.
Therefore, I strongly suggest the author to evaluate the performance of emulated FWI in cases of rapid weather change regime during the fire season and check if the change in FWI values due to precipitation events are realistic. A set of events shoud be analysed and results presented; these wildfires demonstrate the impact of weather patterns on fire management in the Iberian Peninsula, as precipitation can significantly aid in firefighting efforts:
1.Portugal (2017)
The October 2017 extreme fires in Portugal and adjacent NW Spain were followed by widespread precipitation on 17 October, which reduced fire danger and helped to control most ongoing fires;
2.Portugal (2022)
Fires in Central Portugal: Significant wildfires occurred in July 2022, and heavy rains in late August helped extinguish ongoing blazes.
3. Spain (2021)
Sierra de la Culebra Fire: This fire burned for several weeks in June and July 2021. Rainfall in mid-July contributed to controlling the fire.
4. Spain (2020)
Navarra Wildfires: Wildfires in Navarra during July 2020 saw significant rainfall in early August that helped extinguish some active fires.
5. Spain (2019)
Gran Canaria Fires: Major wildfires occurred in August 2019, with rainfall in early September aiding firefighting efforts.
6. Portugal (2019)
Fires in the Algarve: Several wildfires in the Algarve region in July 2019 were followed by precipitation in August, helping to control the situation.
7. Spain (2018)
Castilla y León Fires: In July 2018, wildfires were prevalent, and rainfall in August assisted in extinguishing some blazes

The reviewer rightly highlights the important role of precipitation in modulating the different moisture codes of the FWI system and its influence on fire danger under rapidly changing weather regimes. We fully acknowledge that rainfall can trigger sharp reductions in fine fuel moisture (FFMC) and, when persistent, can also affect the deeper layers represented by DMC and DC, with clear operational implications for fire suppression and fuel management. While a detailed event-by-event analysis of historical wildfire episodes lies beyond the scope of the present study (which focuses on

evaluating the performance of deep learning emulators trained on daily predictors) we have incorporated several complementary analyses that directly address the reviewer's concerns in a systematic way:

- First, we evaluate the emulators' ability to reproduce each individual component of the FWI, allowing us to assess the sensitivity of different fuel moisture codes to meteorological drivers, including precipitation.
- Second, we apply the eXplainable AI (XAI) methods used in the manuscript to quantify the relative importance of each predictor across different fire-danger conditions, allowing us to examine how the model internally responds to precipitation signals and how the deep learning architecture compensates for the absence of explicit precipitation information when producing fire-danger estimates.

Importantly, the exclusion of precipitation in one of the predictor sets is not intended to disregard its physical relevance, but rather to explore a limiting scenario in which a variable that is typically among the least reliably represented in climate models and reanalyses is deliberately omitted. This allows us to assess the robustness of the emulation framework when precipitation is uncertain, unavailable, or unsuitable for direct use, conditions that frequently arise in climate projection contexts.

Moreover, some of the physical consistency analyses suggested by the reviewer, including the detailed assessment of FWI components and their response to rapid weather changes, are part of a subsequent manuscript **currently under preparation**, where we investigate the physical coherence of the emulated components in greater depth. We mention these efforts here to acknowledge the validity and relevance of the reviewer's concerns, even though such analyses exceed the methodological focus of the present contribution.

Together, these elements provide a rigorous and transparent framework that addresses the reviewer's suggestions while maintaining the scope of the current study.

Focus on DL methods

The present work is mainly focus on DL methods to emulate climate indices in general and FWI in particular. In my opinion this is an important topic, but more appropriate to a AI journal. Many DL models are opaque, making it difficult for researchers and decision-makers to understand the reasoning behind a specific prediction. This is especially true when some case studies are not presented to allow a better understanding of the natural hazard event. In natural hazard science, this lack of "physical explainability" can lead to a lack of trust from government agencies and emergency responders. Therefore, if the authors aim to propose an index that show potential use for government agencies and emergency responders in natural hazard domains, a strong effort to provide weather and physical explanations of the obtained index in selected cases should be provided, in this particular case the ability to characterize fuel moisture content and fire spread and intensity of firefront.

We appreciate the reviewer's thoughtful comments regarding the emphasis on deep learning (DL) methods and the importance of physical interpretability in natural hazard applications. We

fully agree that the operational use of fire-danger indicators by government agencies and emergency responders requires transparent methods whose behaviour can be understood in physically meaningful terms. While the primary objective of the present work is to assess the feasibility and performance of DL models for emulating the Fire Weather Index (FWI) using daily-scale predictors, we have taken several steps to ensure that our approach remains physically grounded and not merely algorithmic.
First, beyond reporting error metrics, we explicitly apply **eXplainable AI (XAI) techniques** to quantify the relative importance of each meteorological predictor under different fire-danger regimes. This analysis provides insight into the internal functioning of the DL emulators and allows us to verify their physical consistency, including their response to key drivers such as temperature, relative humidity, wind, and precipitation. Second, we evaluate the emulators' skill in reproducing the individual components of the FWI system, which offers additional evidence that the DL models are capturing physically meaningful relationships related to fuel-moisture dynamics. These two elements help mitigate the "black-box" nature of DL methods and strengthen the physical interpretability of the results.
We also acknowledge the reviewer's concern regarding the absence of detailed case studies illustrating specific wildfire events. Such event-focused analyses (particularly those involving the evolution of fuel-moisture codes, fire spread behaviour, and abrupt changes induced by precipitation) are indeed highly valuable for end-users. For this reason, a more exhaustive assessment of the physical coherence of the emulated FWI components and their behaviour during rapid weather transitions constitutes the core focus of a second manuscript currently in preparation, where we analyse selected episodes and explore the DL-based emulation framework in greater depth. We mention this ongoing work here to emphasise that we consider the reviewer's concerns entirely valid and are actively addressing them, while maintaining the methodological scope of the present contribution.

Overall, our intention in this manuscript is not to replace physically based indices with opaque ML tools, but to demonstrate that DL-based emulators can provide accurate and physically interpretable approximations of established fire-danger indices when only daily predictor information is available, an increasingly relevant scenario in both climate applications and operational settings.

General

Page 4. Detail more about each one of the 5 sub-indices and FWI, the role of each meteorological variable on them and how they are related with FWI

We thank the reviewer for this useful suggestion. In response, Section 2.2 has been substantially revised to include a clearer and more detailed description of all FWI sub-indices (FFMC, DMC, DC, ISI and BUI), their respective physical meaning, and the role of the

meteorological variables driving each component. The revised text now explains how the different moisture codes respond to temperature, relative humidity, wind speed, and precipitation at distinct timescales, and how these components are sequentially combined to produce the final FWI. These additions improve the clarity and physical grounding of the manuscript, fully addressing the reviewer's request while providing a coherent link to the new Section 3.6 devoted to fuel-moisture-code emulation.

Page 8: increase readability of Figure 1

Thank you for your comment. The figure has been revised to improve readability. The layout of the different architectures has been adjusted to better utilize the available space within the figure.

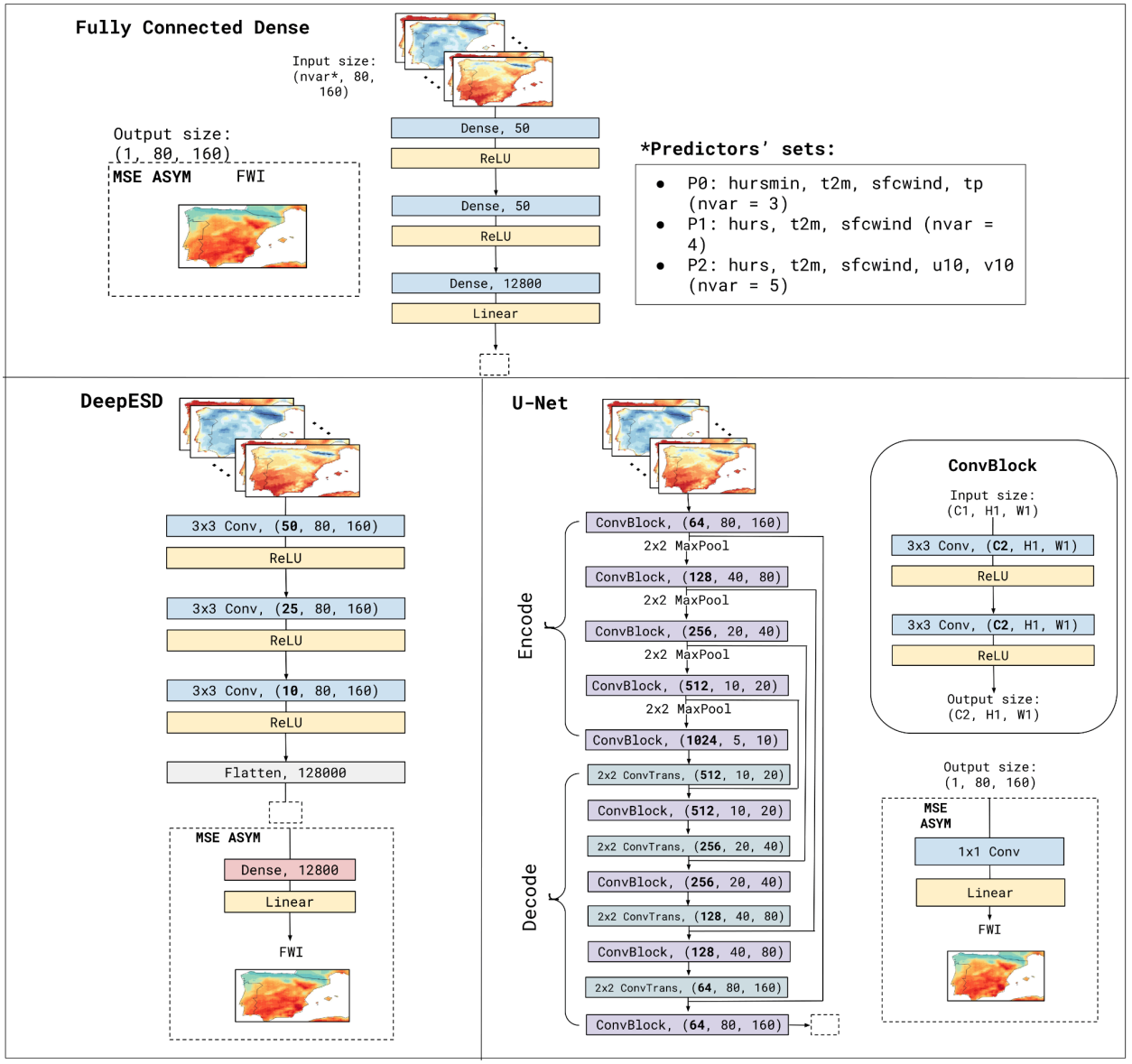

Figure 1: Schematic representation of the DL architectures used in this study, showcasing the different types of layers and their corresponding dimensions. The figure also highlights the distinct sets of predictors used as training inputs.

Table 3: provide the spatial mean of FWI for the different classes or alternatively the Iberian maps of the FWI values for each class

Thanks for your comment. We have added the spatial mean value for each category in Table 3. The new version is shown below:

Level	FWI Percentile Range	Spatial Mean Value
Low	Below 40th percentile	17.57
Moderate	40th – 65th percentile	42.27
High	65th – 85th percentile	55.67
Very High	85th – 95th percentile	67.02
Extreme	Above 95th percentile	80.70

Table 1. FWI Classes According to AEMET (Based on Percentiles). The right column shows the FWI reference spatial mean per category during the fire season. This Table corresponds to Table 3 of the manuscript.

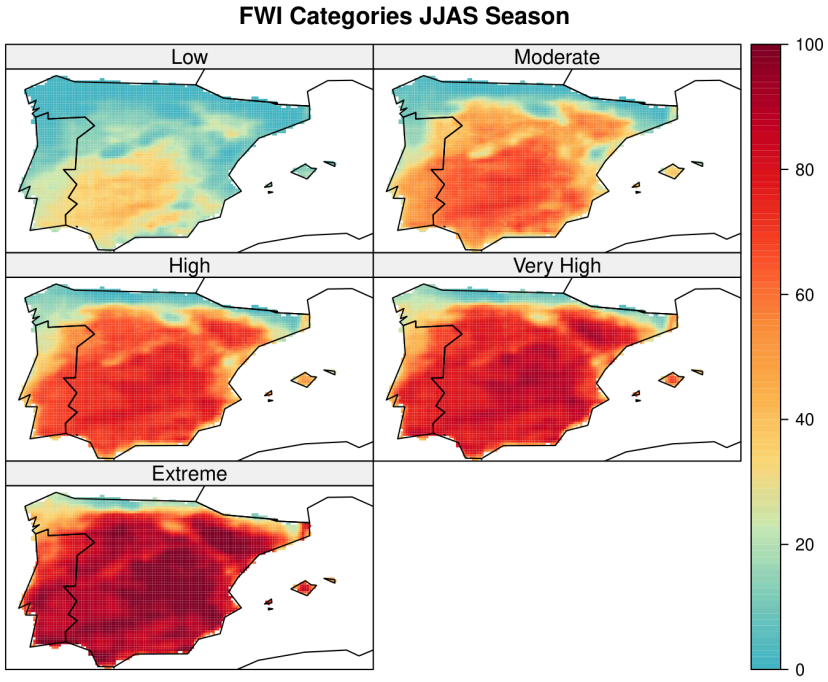

Figure 2: FWI climatologies by fire danger level during the fire season (June to September).

Figure 2: Figure FWI 12UTC is missing for FreqFWI95

It is not explicitly shown in Figure 2 because, by construction, each grid point in the reference data records exactly 5% of days above its local 95th percentile threshold during the season. In

other words, the value in each grid across the spatial map is 0.05. It is explained in the manuscript. This is indicated in L269-L271 of the manuscript.

Figure 2: High bias (overestimation) and MAE over areas with higher fire frequency. Comment on the robustness of the ProxyFWI

Thanks for pointing to this critical issue. It is actually one of the main points in favor of using deep-learning emulators that can consistently learn the “true” FWI behaviour. The relatively high bias in the frequency of FWI95 events obtained with the Proxy FWI reflects the structural limitations inherent to replacing instantaneous noon-LST meteorological variables with daily aggregated predictors, rather than a lack of robustness of the proxy formulation itself. In regions where fire danger exhibits a strong diurnal cycle, extreme FWI values typically occur around midday when temperature peaks and relative humidity reaches its minimum. By construction, the Proxy FWI uses daily mean temperature and wind, daily minimum relative humidity, and 24-h accumulated precipitation. These aggregated predictors effectively smooth out the sharp sub-daily variations that drive threshold exceedances, leading to either under- or over-counting FWI95 events depending on local climatological conditions.
This issue becomes particularly visible in the frequency metric because it is more sensitive to small shifts around the percentile threshold than continuous quantities such as mean FWI or FWI95 climatology, which the Proxy reproduces more faithfully. Indeed, Figure 2 shows that the spatial gradients of mean FWI, FWI95 and Max Spell95 are well captured by the Proxy FWI. The high relative bias in FWI95 frequency therefore reflects intrinsic differences between the temporal resolution of the predictors (daily aggregates) and the target metric (instantaneous noon-LST extremes), rather than instability or inconsistency of the Proxy FWI. These limitations are well-documented and expected when projecting FWI from daily-scale climate data, which is why the Proxy FWI is used here primarily as a simple, physically consistent baseline against which the added value of the deep-learning emulators can be demonstrated.
Importantly, the deep-learning emulators substantially mitigate this limitation of the Proxy FWI. By learning the nonlinear relationships between daily aggregated predictors and the instantaneous noon-LST FWI, the emulators are able to recover the timing and magnitude of extreme fire-weather events more accurately than the Proxy formulation. This results in markedly lower biases in FWI95 frequency and improved reproduction of high-danger spell characteristics. These gains highlight the value of the proposed emulation framework in contexts where only daily meteorological data are available, providing a closer approximation to the reference FWI than the traditional Proxy approach. We have revised Section 3.1 accordingly (L284-296).

Figure 6: Explain the low role of wind for extreme cases must be explained

The relatively low saliency of wind speed for extreme fire danger categories reflects both the characteristics of the training data and the role of wind within the Fire Weather Index at very

high danger levels. Extreme fire danger conditions in Iberia are predominantly associated with persistent heat, very low relative humidity, and prolonged fuel drying, rather than with short-lived wind peaks. Once fuels are critically dry, incremental increases in wind speed have a reduced impact on the overall fire danger classification compared to thermodynamic controls. This leads the model to rely more strongly on temperature and relative humidity when discriminating between extreme and very extreme categories.

In addition, wind exhibits higher temporal variability and weaker spatial coherence than the other predictors, particularly when using daily aggregated inputs. As a result, its contribution to stable, large-scale extreme fire danger patterns is smaller, which is consistently reflected in the learned saliency.

We have clarified this interpretation in the revised manuscript to emphasize that low wind saliency in extreme categories does not imply physical irrelevance, but rather a secondary role relative to fuel dryness under extreme fire weather conditions, indicated in L387-395 of the manuscript.

Minor:

The acronym ASYM is not defined in the text

The acronym is defined in the Section 2.4.4, it is indicated in L196 of the manuscript: “the mean squared error (MSE), which serves as a standard loss function, and an asymmetric loss function (ASYM)..”

The acronym XAI is not defined in the text

The acronym is defined in the Introduction section, this is indicated in L63 of the manuscript: “ It also introduces new challenges related to model interpretability, requiring methods of eXplainable Artificial Intelligence (XAI)...”

References

Lawson, B.D., Armitage, O.B., 2008. Weather Guide for the Canadian Forest Fire Danger Rating System.

ANNEX

FUEL MOISTURE CODE EMULATION AND PHYSICAL ATTRIBUTION THROUGH EXPLAINABILITY METHODS

We further include preliminary analyses for the reviewer demonstrating how precipitation, when used as part of the predictor set, influences the emulation of the various fuel-moisture codes of the Canadian FWI system. For conciseness, in this analysis we focus on two sets of predictor sets, namely the predictor sets P0 (with precipitation) and P1 (without precipitation), as displayed in Table 1 of the manuscript.

DROUGHT CODE EMULATION

The Drought Code (DC) represents the long-term moisture content of deep, compact organic layers and is the slowest-responding component of the Canadian FWI System (Lawson & Armitage, 2008). Its evolution reflects the cumulative effects of sustained drying driven by temperature and atmospheric moisture deficits, with a r**esponse time of several weeks to months**. Only substantial or prolonged precipitation events can effectively reduce the DC, as deep organic horizons exhibit strong resistance to infiltration; light or short-lived rainfall typically has no measurable impact. Because of this long memory, the DC is a key determinant of the seasonal buildup of deep-layer dryness, influencing the potential for persistent smouldering combustion and high-intensity wildfire behaviour under extended drought conditions.
In the context of our DL-based emulation framework, **the DC provides a particularly stringent test of whether the model can learn the slow-accumulating**, integrative nature of deep-fuel moisture without relying explicitly on sub-daily meteorological information. Since effective precipitation events capable of reducing the DC are comparatively rare, precipitation enters the emulation problem as a high-impact but low-frequency predictor. This makes the DC especially relevant for evaluating whether the DL models can internalize the distinction between negligible and meaningful rainfall inputs, and whether they can approximate the long-memory behaviour of the code even when trained solely on daily aggregates. By comparing emulator skill with and without precipitation included in the predictor set (and by analysing predictor relevance via XAI methods) we assess not only the model's capacity to reproduce the DC itself, but also the degree to which the DL architecture can compensate for missing or poorly represented precipitation information. This is particularly important for climate-change applications, where precipitation is often one of the least reliable variables in climate model output, and where the DC's long-term integrative nature makes it sensitive to such uncertainties.

Figure 3 presents the reference Drought Code (DC) climatology for the fire season (JJAS) corresponding to the test period (2018-2021), together with the U-Net emulations obtained using the P0 (with precipitation) and P1 (without precipitation) predictor sets and their associated climatological biases. The reference DC field (top row) shows the expected large-scale north-south gradient across the Iberian Peninsula, with markedly higher seasonal dryness in the central and southern regions and substantially lower values in the cooler, moisture-rich northern areas.

The U-Net emulations (middle row) closely reproduce this spatial structure for both predictor sets, with only minor differences between P0 and P1. This similarity is physically consistent with the behaviour of the DC during the fire season, a period characterized by minimal and infrequent precipitation across most of the region. Because deep-layer organic horizons require substantial or prolonged rainfall to register an effective moisture increase, and such events are rare during JJAS, the DC is only weakly influenced by the day-to-day precipitation variability captured in the daily predictor sets. As a result, including precipitation (P0) provides only marginal additional information relative to P1.

The bias fields (bottom row) reflect this behaviour: both predictor sets show comparable spatial bias patterns, with slight underestimation in the wetter northern areas and small overestimation in the drier central and southern regions. The close agreement between P0 and P1 demonstrates that the DL model can emulate the long-memory behaviour of the DC even without explicit precipitation input, largely because, during the fire season, precipitation contributes minimally to the DC signal compared to temperature and atmospheric moisture deficits. This reinforces the interpretation that DC emulation is primarily driven by the slow-varying predictors, while precipitation, though physically meaningful, exerts a limited operational influence during JJAS.

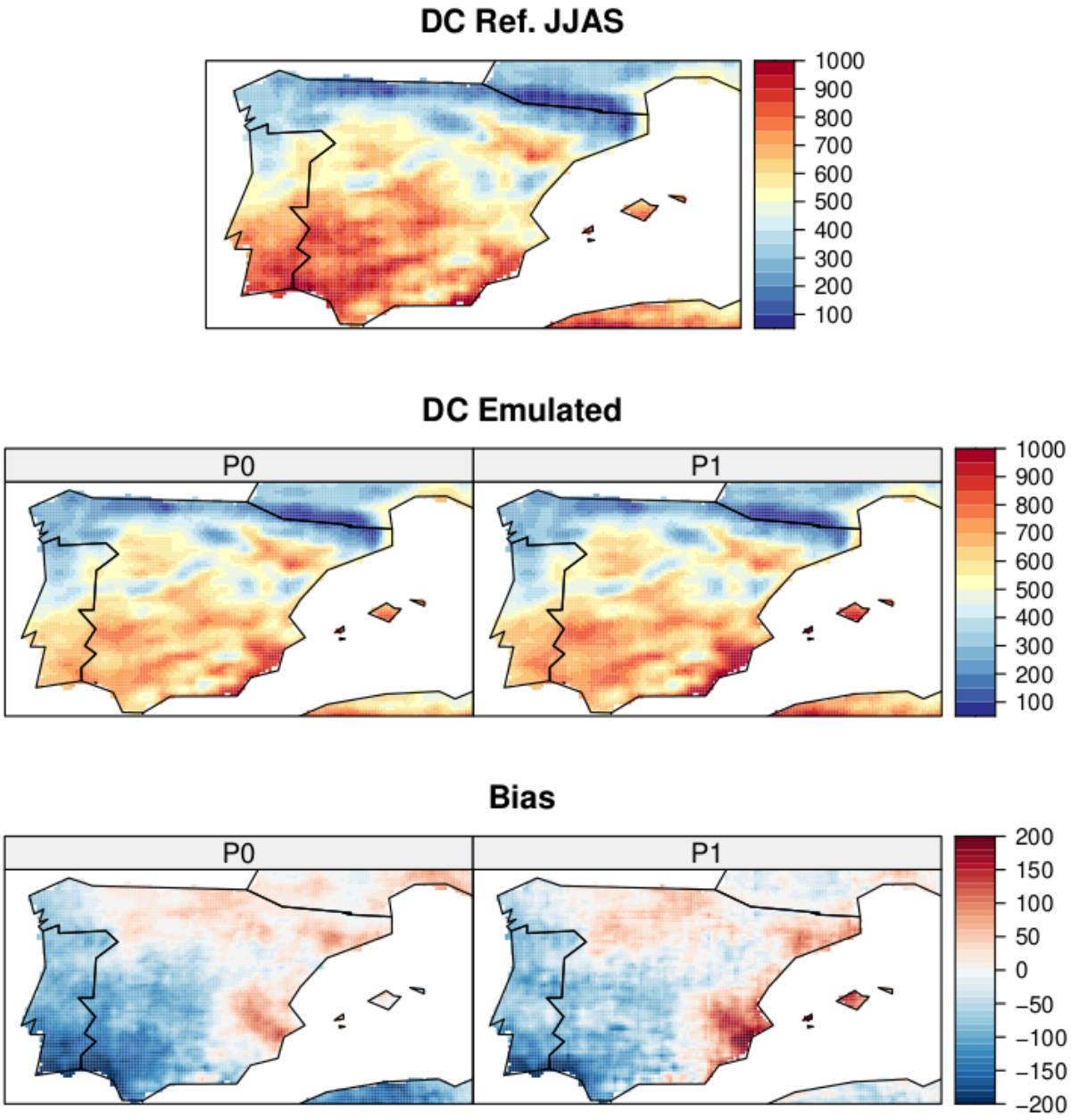

Figure 3: Climatologies of the reference and emulated Drought Code (DC) and the corresponding prediction biases for the fire season. The first row shows the reference DC climatology for the fire season (JJAS) during the test period 2018-2021. The second row presents the U-Net model climatologies: the left panel corresponds to the P0 predictor set (including precipitation), and the right panel corresponds to the P1 predictor set. The third row displays the climatological biases of the U-Net predictions relative to the reference.

Figure 4 shows the spatially aggregated daily Drought Code (DC) time series for the test period (2018-2021), comparing the reference values (black) with the U-Net emulations using the P0 (with precipitation, red) and P1 (without precipitation, blue) predictor sets across the three

climatological regions defined in the Appendix A (ATL, CONTM, COASM). In all regions, the emulated series reproduce the seasonal cycle and interannual variability of the reference DC remarkably well, including both the build-up of deep-layer drought during the fire seasons and the subsequent relaxation phases.

In the Atlantic region (ATL, top row, the wettest part of Iberia during JJAS) the reference DC exhibits lower peaks and more frequent short-term drops associated with precipitation events. Both emulators track these features closely, with P0 occasionally following sharp decreases slightly more closely than P1 when rainfall temporarily reduces DC values. However, the overall differences between P0 and P1 remain small, indicating that even in this comparatively wetter region the long-term DC evolution is dominated by the cumulative drying signal rather than by day-to-day precipitation variability.

In the Continental Mediterranean (CONTM) and Coastal Mediterranean (COASM) regions (middle and bottom rows), where fire-season precipitation is rarer and deep-layer fuels remain generally drier throughout the entire JJAS season, both P0 and P1 closely follow the reference DC throughout the period, including the magnitude and timing of the seasonal maxima. The two emulated curves are almost indistinguishable at this aggregated regional scale, reinforcing the notion that, during the fire season, DC behaviour is primarily controlled by sustained atmospheric moisture deficits and temperature, while precipitation acts as a high-impact but low-frequency modifier. The close agreement between P0 and P1 in all three regions confirms that the U-Net architecture is able to emulate the long-memory dynamics of the DC with rather good fidelity, and that the absence of explicit precipitation information in P1 does not substantially degrade performance, even in the wetter ATL region.

Although both emulators (P0 and P1) reproduce the large-scale temporal evolution of the DC remarkably well across all regions, the time-series comparison reveals that the highest DC peaks tend to be slightly underestimated relative to the reference. This behaviour is physically consistent with the nature of the DC itself: as a long-memory index driven by cumulative drying, its sharpest peaks occur following extended periods of warm, dry conditions. These rapid rises are difficult to emulate with daily-scale predictors alone, and the smoothing effect inherent to the deep-learning architectures tends to moderate the most pronounced maxima. Importantly, this underestimation appears in both P0 and P1, indicating that even when precipitation is explicitly included, the model's limitations stem from the intrinsic difficulty of reconstructing fine-scale temporal accelerations in a slowly evolving variable using only daily aggregated meteorological inputs.

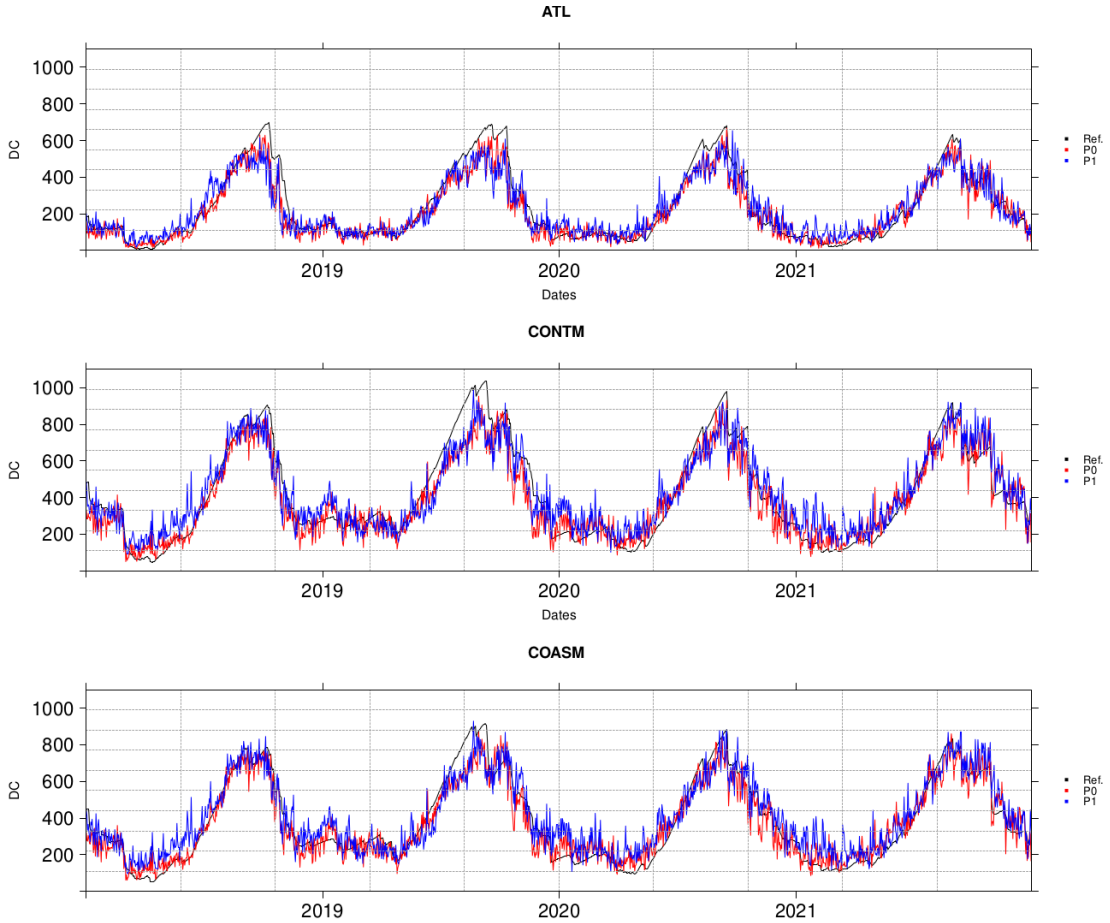

Figure 4. Spatially aggregated daily DC time series for the test period (2018–2021), shown for the reference (black) and the emulated simulations (P0 in red; P1 in blue) across the different climatological regions presented in the Appendix.

The saliency analysis for the Drought Code (DC) emulation in JJAS, aggregated by climatological region and fire-danger category, further clarifies the relative influence of the predictors and the limited role of precipitation. When precipitation is included in the predictor set (**Fig. 5**, P0, top panel), temperature and relative humidity systematically emerge as the dominant drivers across danger classes and regions ATL and COASM, with saliency values that clearly exceed those of precipitation, with wind generally playing a secondary role, except in CONTM where it is more relevant. Precipitation shows only modest contributions at low and medium danger levels and becomes almost negligible for high, very high and extreme DC conditions, consistent with the physical formulation of the DC, which only responds to substantial or prolonged rainfall capable of penetrating deep organic layers.
When precipitation is removed from the predictor set (**Fig. 6**, P1, bottom panel), the saliency patterns remain qualitatively very similar: temperature and relative humidity still account for most of the explanatory power, with wind retaining a secondary role and no evidence of compensating spurious dependencies on other variables. The stability of these patterns across

regions (including the wetter ATL sector) and across fire-danger categories indicates that the emulator is primarily relying on slow-varying atmospheric moisture deficits and thermal conditions to reproduce the long-memory behaviour of the DC. Taken together with the climatological maps and regional time series, these saliency results support the conclusion that precipitation, while physically relevant in principle, plays a minor role in DC emulation during the fire season, and that the deep-learning model behaves in a manner consistent with the known physics of the Drought Code. The systematic overestimation of DC during winter by the P1 predictor set (excluding precip) depicted in the time series of **Fig. 4** can be attributable to the higher importance of precipitation to emulate the index for lower percentiles, as depicted in saliency plots (**Fig. 3**), particularly over the COASM region.

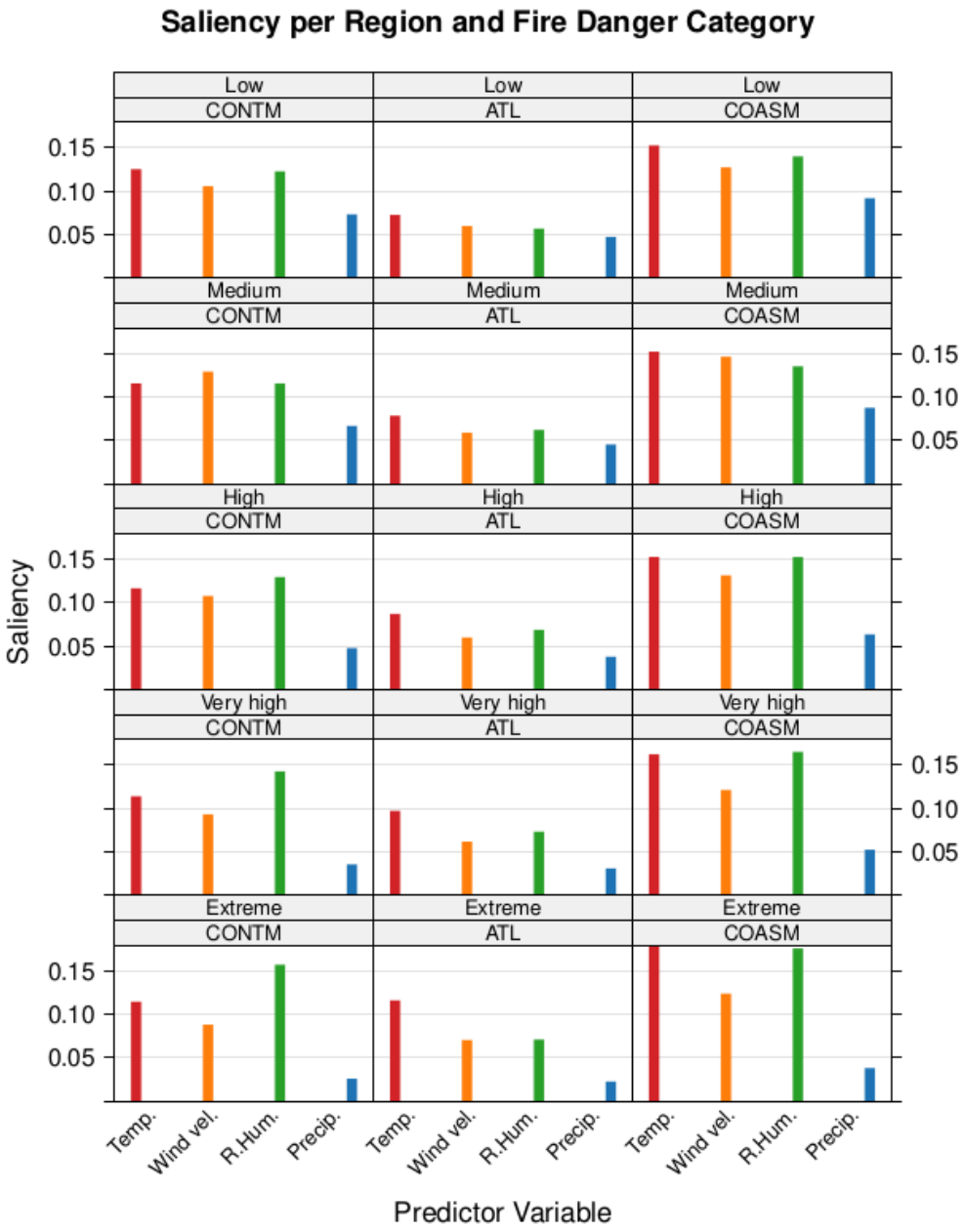

Figure 5: DC aggregated saliency of predictor variables by climatic region and fire danger category for the U-Net model considering P0 predictors set. Columns represent the climatic regions defined in Figure A1 of the manuscript, while rows correspond to the fire danger categories described in Section 3.2 of the manuscript . The displayed saliency values are derived from saliency maps, normalized at each grid point. Results for the JJAS season, test period (2018-2021).

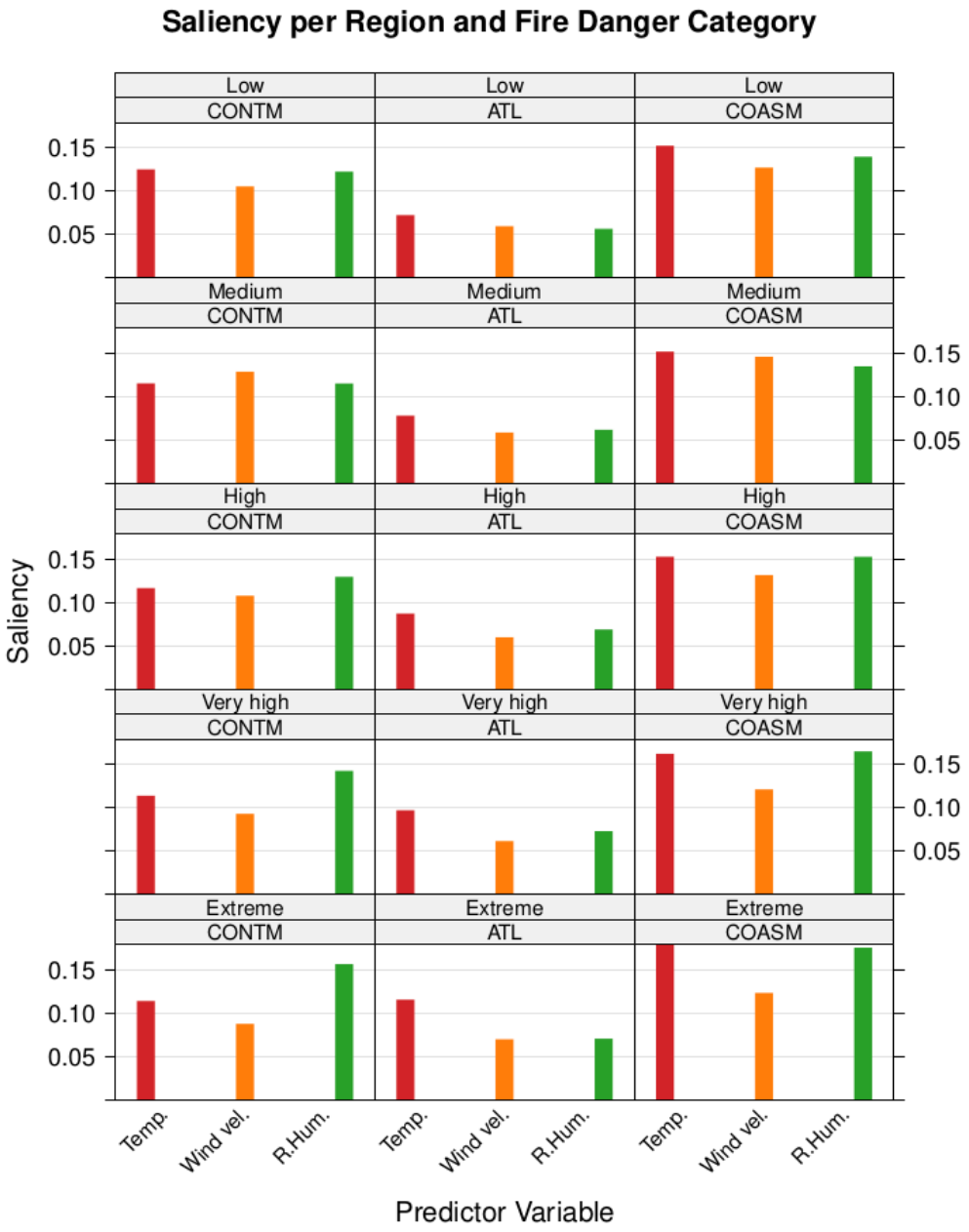

Figure 6: DC aggregated saliency of predictor variables by climatic region and fire danger category for the U-Net model considering P1 predictors set. Columns represent the climatic regions defined in Figure A1 of the manuscript, while rows correspond to the fire danger categories described in Section 3.2 of the manuscript . The displayed saliency values are derived from saliency maps, normalized at each grid point. Results for the JJAS season, test period (2018-2021).

DUFF MOISTURE CODE EMULATION

The Duff Moisture Code (DMC) represents the moisture content of loosely compacted organic layers in the upper duff, responding on a medium-term timescale of several days to weeks (Lawson & Armitage, 2008). Its evolution reflects the balance between atmospheric drying—primarily driven by temperature and relative humidity—and the infiltration of precipitation that exceeds the minimum effective rainfall required to penetrate beyond the surface litter. Because of this intermediate response time, the DMC plays a central role in determining seasonal fire danger: it governs the availability of sub-surface fuels that contribute to sustained fire spread once ignition occurs.
In the context of our DL-based FWI emulation, the DMC represents an especially informative component for assessing the role of precipitation as a predictor. Unlike the FFMC, which responds sharply to light rainfall, or the DC, which only reacts to substantial and persistent precipitation, the DMC occupies a regime where rainfall is important but selective: only events exceeding the effective threshold meaningfully reduce its value. This makes the DMC a useful diagnostic for understanding whether DL emulators can internalize the non-linear, threshold-dependent nature of rainfall infiltration when trained solely on daily-scale variables. By comparing predictor sets with and without precipitation, and by analysing predictor relevance through XAI methods, we evaluate not only whether the models can reproduce the DMC itself, but also how they compensate for the absence of precipitation information, a key question for climate-change applications where precipitation is often poorly represented or inconsistently available.

Figure 7 shows that the U-Net emulator accurately captures the main spatial patterns of the Duff Moisture Code (DMC) during the fire season. The reference climatology (top row) exhibits the expected north-south gradient across the Iberian Peninsula, with lower DMC values over the wetter Atlantic-influenced regions and higher values in the drier interior and southern sectors. The emulated fields (middle row) reproduce these large-scale features well for both predictor sets: P0, which includes precipitation, and P1, which excludes it. Differences between the two emulations are modest, reflecting the intermediate sensitivity of the DMC to rainfall: only precipitation events exceeding the effective infiltration threshold significantly reduce DMC values, and such events are relatively infrequent during JJAS.
The bias maps (bottom row) indicate that both predictor sets produce similar and spatially coherent discrepancies. P0 tends to slightly underestimate the DMC, particularly in areas where effective rainfall events are more likely to occur, while P1 shows small positive biases in the driest regions and local underestimation in the northern fringe. Overall, both emulators provide consistent results, confirming that the DMC during the fire season is primarily governed by temperature and atmospheric moisture deficits, with precipitation playing a secondary (though not negligible) role at daily scales.

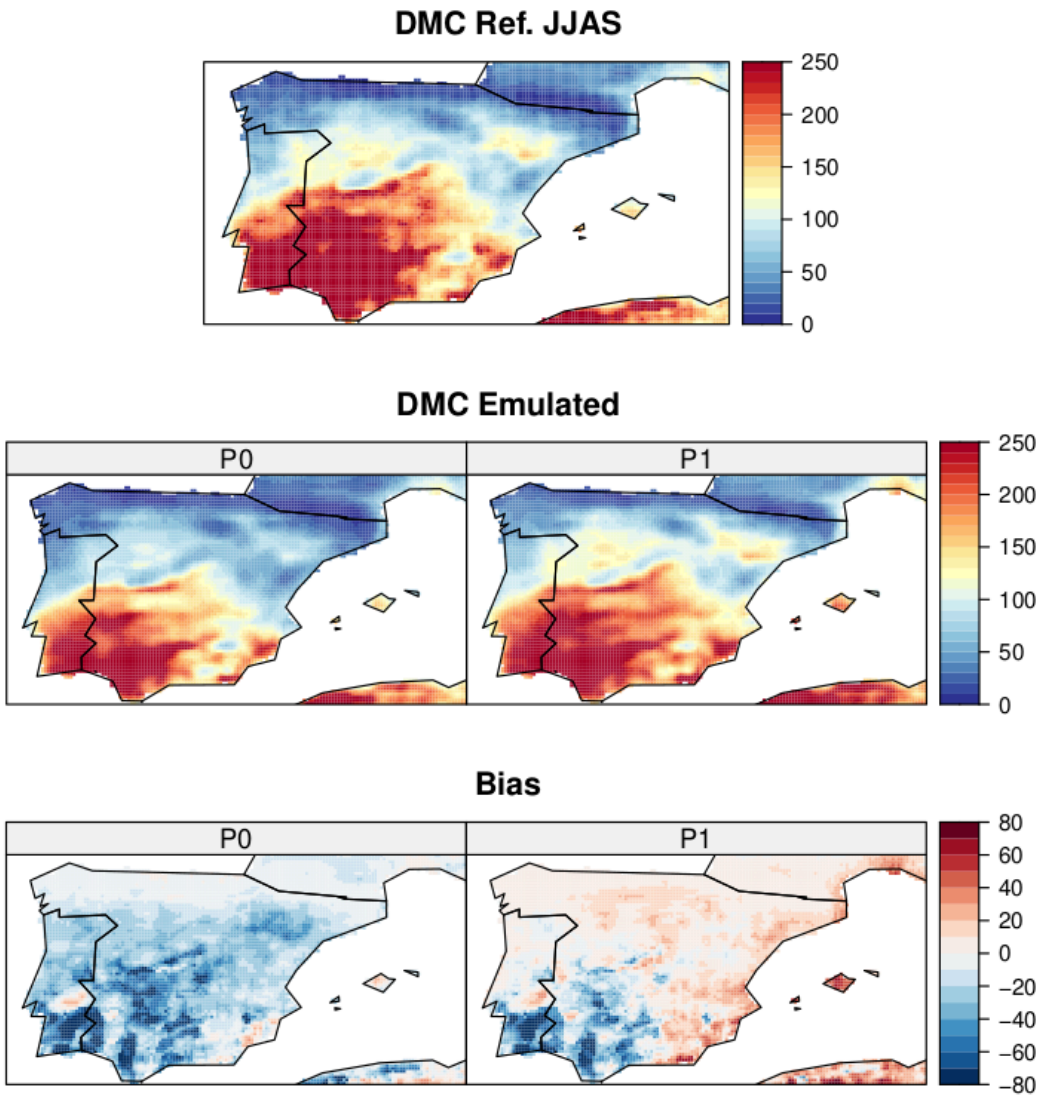

Figure 7: Climatologies of the reference and emulated Duff Moisture Code (DMC) and the corresponding prediction biases for the fire season and test period (2018-2021). The first row shows the reference DMC climatology. The second row presents the U-Net model climatologies: the left panel corresponds to the P0 predictor set (including precipitation), and the right panel corresponds to the P1 predictor set. The third row displays the climatological biases of the U-Net predictions relative to the reference.

The saliency analyses for the Duff Moisture Code (DMC) reveal that the emulator's performance is largely driven by temperature and relative humidity, with precipitation contributing only marginally during the fire season (**Fig. 8**). In fact, the DMC emulation performs slightly better when precipitation is excluded (P1, **Fig. 7**), particularly in the core Mediterranean regions, where rainfall events during JJAS are rare and often insufficient to exceed the effective

infiltration threshold required to influence the DMC. The aggregated saliency maps corroborate this behaviour: when precipitation is included (P0, **Fig. 8**), its contribution remains consistently low across all regions and fire-danger categories, while temperature and relative humidity dominate the predictor importance. When precipitation is removed (P1, **Fig. 9**), the overall saliency structure remains nearly unchanged, with the model shifting reliance toward the same two primary drivers rather than compensating via spurious dependencies on other variables. This stability across regions (including the wetter ATL area) and across all fire-danger levels supports the physical interpretation that (similarly to DC), DMC variability during JJAS is governed mainly by medium-term drying processes controlled by atmospheric moisture deficits, with precipitation acting as a secondary and irregular contributor.

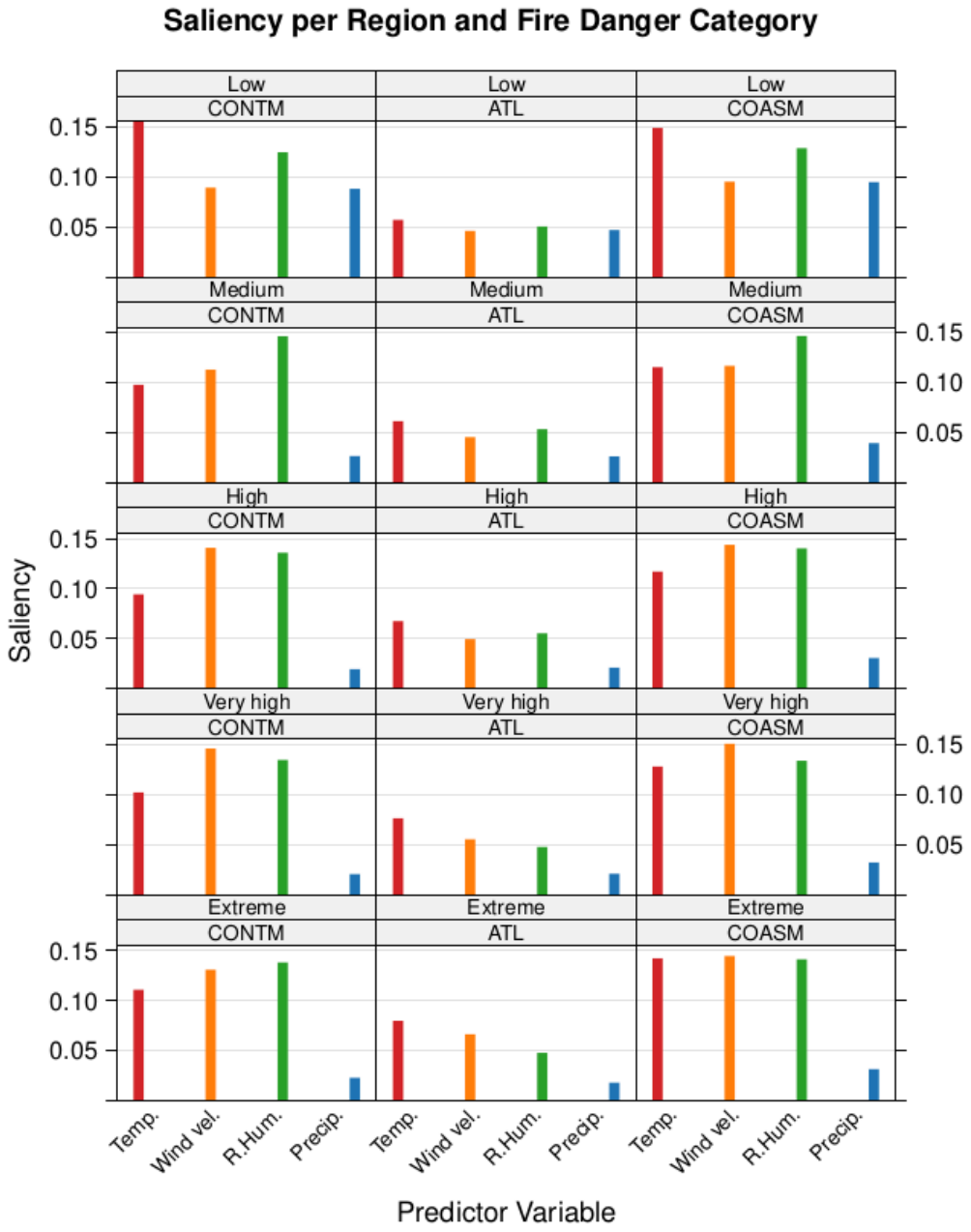

Figure 8: DMC aggregated saliency of predictor variables by climatic region and fire danger category for the U-Net model considering P0 predictors set. Columns represent the climatic regions defined in Figure A1 of the manuscript, while rows correspond to the fire danger categories described in Section 3.2 of the manuscript . The displayed saliency values are derived from saliency maps, normalized at each grid point. Results for the JJAS season, test period (2018-2021).

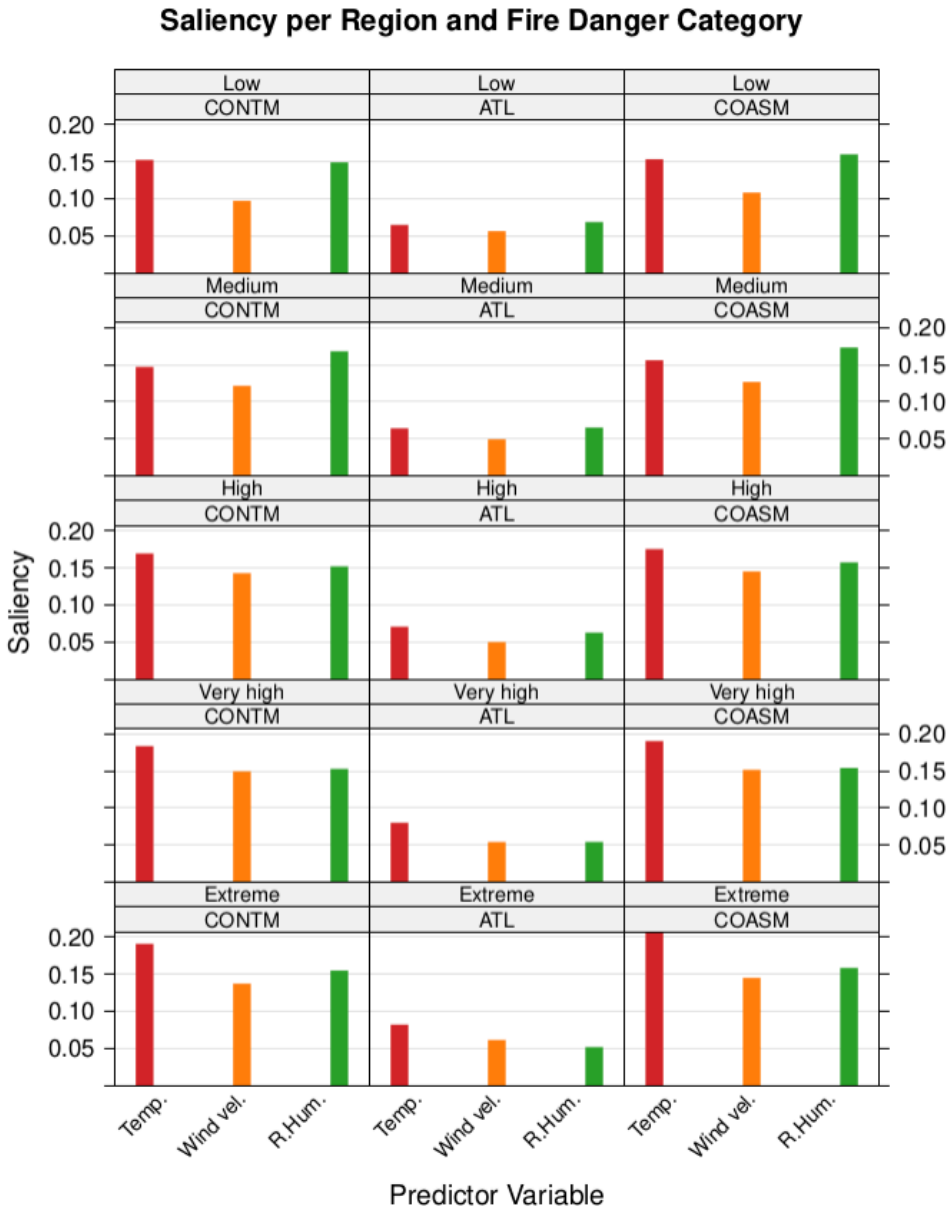

Figure 9: DMC aggregated saliency of predictor variables by climatic region and fire danger category for the U-Net model considering P1 predictors set. Columns represent the climatic regions defined in Figure A1 of the manuscript, while rows correspond to the fire danger categories described in Section 3.2 of the manuscript . The displayed saliency values are derived from saliency maps, normalized at each grid point. Results for the JJAS season, test period (2018-2021).

FINE FUEL MOISTURE CODE EMULATION

The Fine Fuel Moisture Code (FFMC) captures the moisture content of surface litter and fine fuels, responding on a very short timescale (typically a few hours to one day) making it the

fastest-varying component of the Canadian FWI System (Lawson & Armitage, 2008). Its evolution is strongly controlled by near-surface meteorological conditions, especially temperature, relative humidity, wind speed, and the presence of even small amounts of precipitation. Because fine fuels exhibit low water-holding capacity and rapid evaporation, the FFMC is highly sensitive to light rainfall events that may have little or no impact on deeper organic layers. This strong and immediate reactivity makes the FFMC a key indicator of ignition likelihood and the rapid onset or cessation of fire danger conditions.
Within the framework of our DL-based emulation approach, the FFMC offers a unique opportunity to evaluate whether daily-scale predictors contain sufficient information to model a process that is inherently sub-daily and highly responsive to short-duration precipitation. Whereas the DMC and DC require rainfall above certain effective thresholds to register changes, the FFMC reacts sharply to even minimal precipitation inputs, making the role of rainfall as a predictor particularly pronounced for this component. Consequently, examining emulator performance with and without precipitation in the predictor set provides insight into the ability of the models to approximate fine-fuel moisture dynamics when high-frequency meteorological information is unavailable. Through XAI-based analyses of predictor relevance, we further assess to what extent the deep learning architecture can either exploit precipitation signals when present or compensate for their absence by leveraging correlated meteorological variables. This evaluation is especially relevant for operational and climate-change contexts, where precipitation is often uncertain or poorly resolved, yet surface fuels—and therefore the FFMC—remain critical determinants of day-to-day fire danger.

Figure 10 shows that the U-Net emulator accurately reproduces the Fine Fuel Moisture Code (FFMC) climatology during the fire season. The reference field (top row) displays very high FFMC values across most of the Iberian Peninsula, with slightly lower values confined to the northern Atlantic-influenced region—reflecting the dominance of dry surface-fuel conditions during JJAS.
The emulated climatologies (middle row) closely match the spatial structure of the reference, both when precipitation is included (P0) and when it is excluded (P1). The emulator captures the extent of high FFMC values across the interior and southern sectors, as well as the north–south gradient associated with moisture availability. Differences between P0 and P1 are minor, consistent with the fact that FFMC responds strongly to very light rainfall, and such events are relatively infrequent during the fire season.
The bias maps (bottom row) show small and spatially coherent discrepancies. P0 tends to slightly underestimate FFMC, particularly in northern coastal areas, whereas P1 exhibits a weak positive bias across central and southern Iberia. These patterns indicate that the exclusion of precipitation has a measurable but limited effect on the FFMC emulation at seasonal scales. Overall, the emulator performs well under both predictor configurations, with temperature, relative humidity and wind speed acting as the dominant controls during JJAS, while precipitation—though more relevant for FFMC than for deeper moisture codes—contributes relatively little due to its low seasonal occurrence.

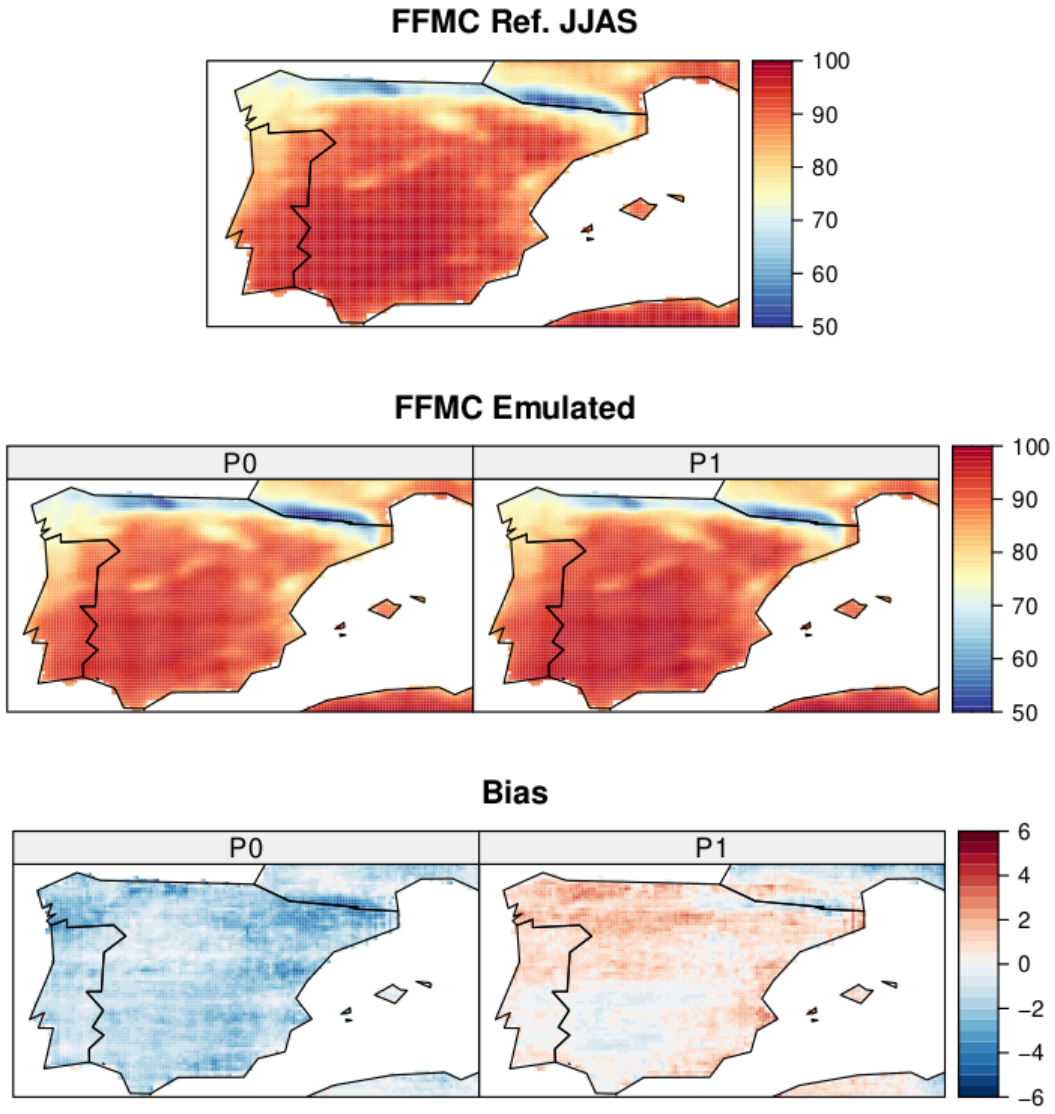

Figure 10: Climatologies of the reference and emulated Fine Fuel Moisture Code (FFMC) and the corresponding prediction biases for the fire season, test period (2018-2021). The first row shows the reference FFMC climatology. The second row presents the U-Net model climatologies: the left panel corresponds to the P0 predictor set (including precipitation), and the right panel corresponds to the P1 predictor set. The third row displays the climatological biases of the U-Net predictions relative to the reference.

The regional FFMC time series (**Figure 11**) show that both emulators –P0 (with precipitation) and P1 (without precipitation)-- successfully reproduce the high-frequency variability and rapid fluctuations that characterise fine-fuel moisture during the fire season. Across all three climatic regions (ATL, CONTM, COASM), the emulated curves closely track the reference, capturing

both the timing and magnitude of day-to-day drops associated with humidity or rainfall events, as well as the rapid recoveries driven by drying.
Differences between P0 and P1 remain small at this aggregated scale, reflecting the limited number of effective precipitation events during JJAS and the dominant influence of temperature, relative humidity, and wind on FFMC evolution. P0 tends to follow the sharpest moisture-increase events slightly more closely (consistent with FFMC's sensitivity to light rainfall) whereas P1 occasionally exhibits a small positive offset during sustained dry periods. Overall, however, both predictor configurations provide a faithful representation of the regional FFMC dynamics, capturing the fast response and strong temporal variability that distinguish this surface-fuel index.

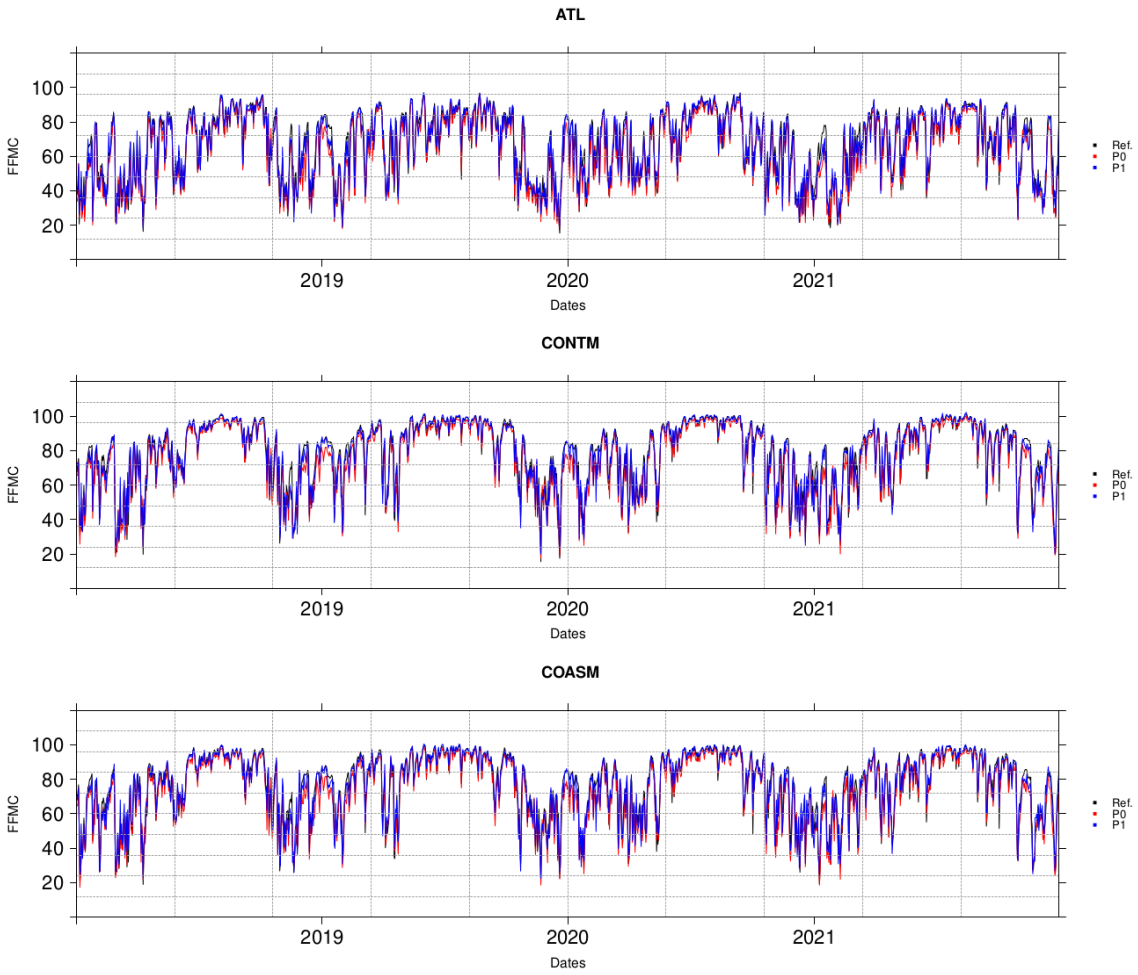

Figure 11: Spatially aggregated daily FFMC time series for the test period (2018–2021), shown for the reference (black) and the emulated simulations (P0 in red; P1 in blue) across the different climatological regions presented in the Appendix.

The saliency analysis for the FFMC emulation confirms that the U-Net model relies on physically consistent drivers across regions and fire-danger categories. When precipitation is included in the predictor set (P0, **Fig. 12**), temperature and relative humidity clearly dominate the saliency

patterns in all three climatic regions, with wind speed playing a secondary but non-negligible role, especially at higher danger levels. Precipitation exhibits only modest contributions and mainly at low and medium fire-danger categories, reflecting the fact that, although FFMC is sensitive to light rainfall, effective JJAS precipitation events are relatively infrequent over Iberia. When precipitation is excluded (P1, **Fig. 13**), the overall saliency structure remains very similar: temperature, relative humidity and wind speed retain their leading roles, with slightly increased saliencies that compensate for the absence of rainfall, rather than revealing any shift toward unintuitive or spurious predictors. This stability of predictor importance between P0 and P1 indicates that the emulator captures FFMC variability primarily through temperature, humidity and wind (consistent with the physical formulation of the index) and that precipitation, while relevant at event scale, exerts only a limited influence on the seasonal emulation performance during the fire season.

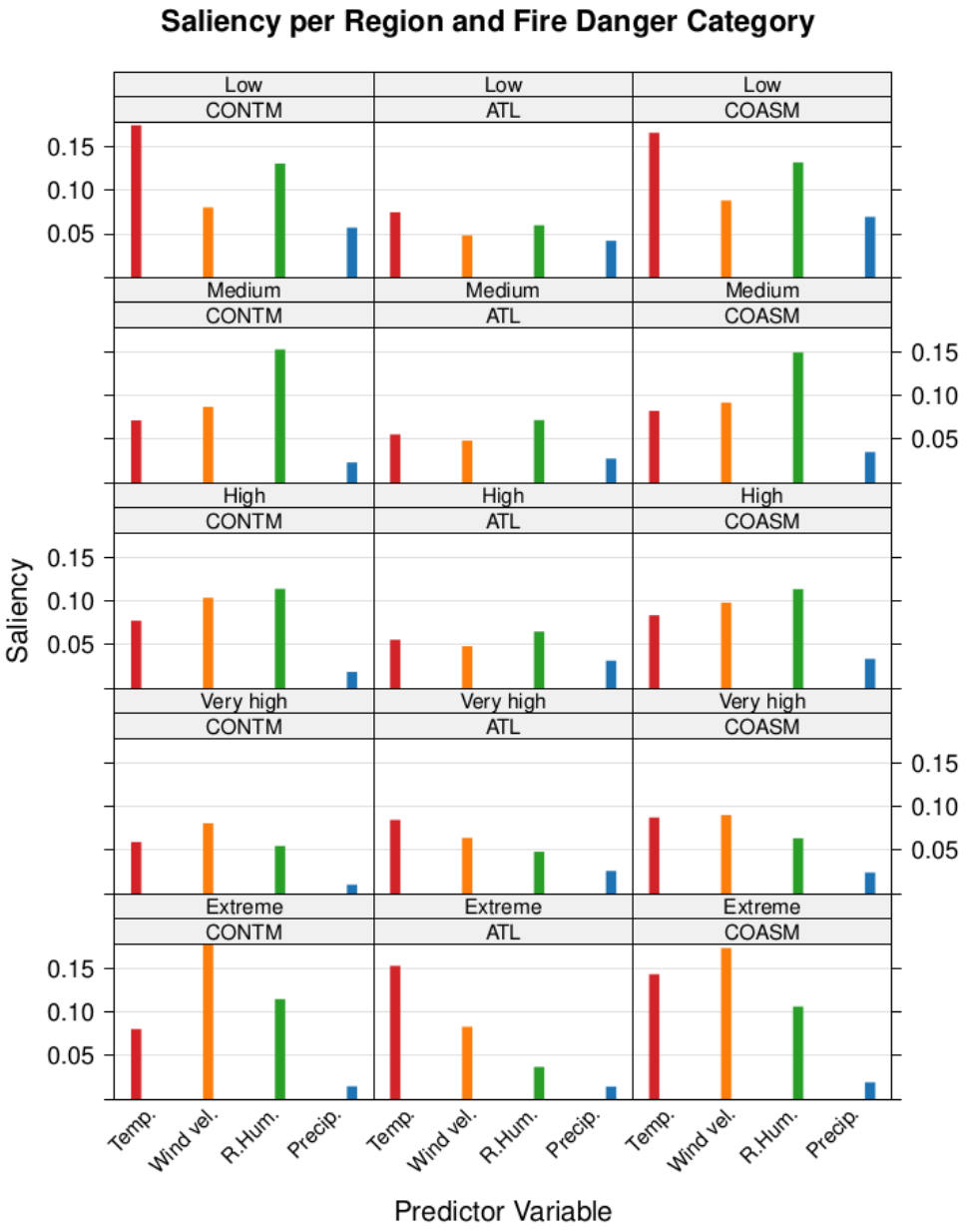

Figure 12: FFMC aggregated saliency of predictor variables by climatic region and fire danger category for the U-Net model considering P0 predictors set. Columns represent the climatic regions defined in Figure A1 of the manuscript, while rows correspond to the fire danger categories described in Section 3.2 of the manuscript . The displayed saliency values are derived from saliency maps, normalized at each grid point. Results for the JJAS season, test period (2018-2021).

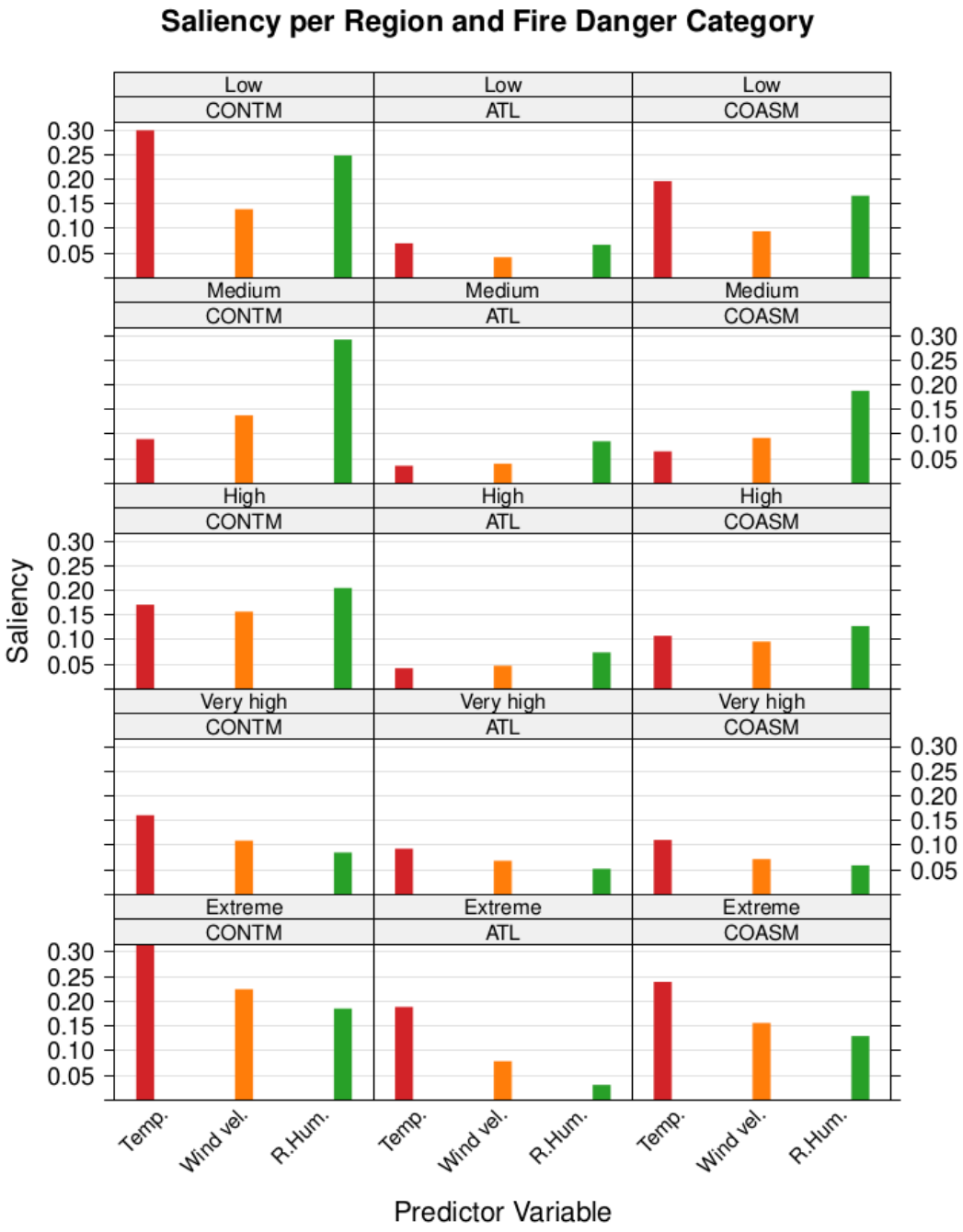

Figure 13: FFMC aggregated saliency of predictor variables by climatic region and fire danger category for the U-Net model considering P1 predictors set. Columns represent the climatic regions defined in Figure A1 of the manuscript, while rows correspond to the fire danger categories described in Section 3.2 of the manuscript . The displayed saliency values are derived from saliency maps, normalized at each grid point. Results for the JJAS season, test period (2018-2021).

CONCLUSION ON FUEL MOISTURE CODE EMULATION EXPERIMENTS

In order to summarize all the above results, we have included in the revised manuscript a new subsection entitled **"3.6 Fuel Moisture Code Sub-indices Assessment"**, which synthesizes and formalizes the analyses presented above, providing a concise summary of the evaluation of the DL framework when applied to the individual fuel moisture components. Within this subsection, a new table is provided summarizing the performance of the DL emulators for the selected FWI sub-indices, based on a set of validation metrics already employed throughout the manuscript.

Code	Predictors sets	MAE	MAE FWI95	Time series correlation	Rel. Bias Freq. FWI95
DC	P0	105.62	227.51	0.72	0.81
	P1	118.17	225.74	0.61	0.82
DMC	P0	50.63	173.32	0.75	0.93
	P1	48.23	139.81	0.85	0.77
FFMC	P0	3.58	2.91	0.98	0.67
	P1	3.36	2.16	0.97	0.52

Table 2. Validation results of DL emulators for different predictors sets and FMC sub-indices; DC, DMC and FFMC. This Table corresponds with Table 5 of the new version of the manuscript.