Moreno-Montes et al 2026 Reviewer Comments

Title: Decadal predictions of wind, solar and compound power indicators to support the European renewable energy sector

Author(s): Sara Moreno-Montes et al.

MS No.: egusphere-2026-1205

MS type: Research article

General Comments

This manuscript evaluates the ability of a multi-model hindcast of decadal predictions to forecast energy indices across Europe for forecast years 1-3. The authors report some skill for forecasts of solar PV potential, while more limited skill is present for wind capacity factors. They also construct a compound energy drought indicator and identify seasonally dependent prediction skill. While the approach presented here is novel, there are several issues with the methodological approach and interpretation of the results which may delay publication until further evaluation has been conducted.

We sincerely thank the reviewer for the thorough and constructive evaluation of our study, which will certainly improve the quality and readability of the manuscript. We appreciate the detailed comments provided, which have helped us identify several important methodological and interpretative aspects that required clarification or further analysis. Point-to-point responses to each comment (in blue) are presented below.

The decision to regrid the decadal predictions (~1 degree resolution) to the higher resolution of ERA5 (0.25 degree) is questionable here, as this gives the impression that decadal predictions have the ability to capture small scale variability in specific regions. This makes it more difficult to interpret the skill maps presented, as much of the significant positive skill is demonstrated in scattered, small scale areas. It would be more appropriate to regrid ERA5 and the decadal prediction systems to the resolution of the coarsest model used (e.g., 1.25 degrees) to provide a fair comparison of skill on the appropriate spatial scales of the decadal predictions. Especially as much of the discussion is focused on slight changes in skill over small regions, a fairer comparison, like the coarse regridding suggested, would help to clarify signal from noise. This would help the interpretation of the findings, as the areas which do show significant skill should be visible as larger regions. This would further demonstrate the robustness of the findings presented here.

We fully acknowledge that statistical downscaling has well-known limitations, including its reliance on empirical relationships, assumptions of stationarity, and the fact that it does not increase the intrinsic predictability of the system (Maraun et al., 2010; Maraun & Widmann, 2018; Hewitson et al., 2014). Despite these limitations, statistical downscaling techniques are widely used in the context of climate services, where the objective is to provide climate information at spatial scales relevant for end users and sectoral applications (e.g., energy, water management), rather than to enhance dynamical skill (Maraun et al., 2010; WMO, 2011; Soares et al., 2017). In this framework, higher-resolution information is often required to derive impact-relevant indicators.

In our case, bringing all datasets to a common spatial grid is necessary to ensure consistency between predictors and the reference dataset (ERA5), which is used for both calibration and evaluation. While degrading ERA5 to the coarser model resolution would be an alternative, this choice is not neutral, as it may smooth part of the local variability contained in the observational-based reference dataset. We therefore opted to preserve the ERA5 grid.

We agree that spatial features at scales smaller than the native resolution of the prediction systems should be interpreted with caution, as they may arise from the downscaling procedure or sampling variability rather than robust predictive skill. In the revised manuscript, we will explicitly state this limitation and adopt a more cautious and consistent interpretation of dispersed and weak skill patterns. The next text will be added: "Small-scale and spatially isolated areas of significant skill are interpreted with caution, as they may arise from the downscaling procedure or sampling variability and do not necessarily indicate robust predictive signals." (Page 6, line 172)

To minimise the risk of introducing artefacts, we adopted a simple approach based on interpolation combined with quantile mapping. This choice is supported by intercomparison studies showing that more complex downscaling methods do not systematically outperform simpler approaches in terms of skill (e.g., Manzanas et al., 2018; Gutiérrez et al., 2019; Moreno-Montes et al., 2026). Quantile mapping corrects systematic biases while preserving the non-Gaussian characteristics of wind distributions, which are essential for the indicators analysed here.

We also acknowledge the reviewer's suggestion that performing the analysis at a coarser resolution would provide a useful benchmark. While the current framework is designed to support impact-oriented applications, we will assess the sensitivity of the results at a coarser grid (e.g., the coarsest model resolution) to evaluate the robustness of the findings. These additional analyses are currently being performed.

The density of figures and subplots described in the results section makes the findings of the paper hard to follow. For example, on page 7 the results of Figure 1 (10 panels) are described with reference to Figure S5 (15 panels) in the supplementary information. In order to make the findings of this paper more impactful, it might be appropriate to condense the findings described in the results section and reduce references to the supplementary material to only where necessary. This would help the reader to better understand the main findings of the paper and improve the clarity of the results.

We thank the reviewer for this helpful comment. We agree that the current density of figures and cross-references to the supplementary material may difficult the readability of the results section. In the revised manuscript, we will streamline the presentation by condensing the main findings, reducing the number of panels where possible, and limiting references to supplementary figures to cases where they are strictly necessary. In particular, we will combine Figure 1 with Figure S7, Figure 3 with Figure S10, and Figure 5 with Figure S14. This will improve the clarity and accessibility of the results.

The authors report that the main contributing factor to the number of energy drought days (NED) during the winter (DJF) is the number of effective wind generation days (wind-Neff). However, in the supplementary material (Figure S13b, i) they demonstrate that solar-Neff values are missing across most of Europe (aside from southern Iberia). This means that the number of ineffective solar days component of the NED index is effectively constant. The authors repeatedly claim that the areas of NED skill during DJF primarily align with wind-Neff. While this is correct, this is also misleading, as for most of the area shown in Figure 5b and 5g, only the number of ineffective wind days contributes, so the NED shown is no longer a compound index. For clarity, the authors should consider presenting the results differently for DJF in Figure 5 or rephrasing the findings to be clear about what is shown.

We agree that during DJF the solar component of the NED index shows very limited variability over most of Europe, as solar-Neff is effectively zero in large regions. As a consequence, the variability of NED in this season is primarily controlled by the wind component. We will clarify this point in the manuscript to avoid misinterpretation.

However, NED remains a compound indicator by definition, as it is based on the joint occurrence of ineffective wind and solar conditions. In winter, the lack of effective solar production means that this condition is always fulfilled, so NED effectively reflects the occurrence of ineffective wind days. This does not imply a change in the definition of the index, but rather a seasonal regime in which one component does not contribute to variability. We will revise the text to make this interpretation explicit.

Overall, this paper applies a novel methodology to demonstrate some notable skill for energy-sector metrics over Europe. However, methodological choices currently limit the robustness of the results presented. Until these issues are addressed, principally that of fair model comparison (coarse resolution), publication should be delayed. To allow time for these points to be addressed, I recommend that publication to ESD is declined until these issues can be addressed.

Specific Comments

- Pg3, lines 65-66: “In this study, five different decadal forecast systems contributing to the Decadal Climate Prediction Project (DCPP; Boer et al., 2016) of the Coupled Model Intercomparison Project Phase 6 (CMIP6; Eyring et al., 2016), are used.”
    - Technically, only three different decadal forecast systems are used, as 3 of the 5 forecast system are different initialisations of EC-Earth3. Although the ensemble size is detailed in the SM, it would be useful to mention this here. Has the drift of each of the models been corrected here, as in Appendix E of Boer et al., 2016 (10.5194/gmd-9-3751-2016)?
- We agree that the description of the forecast systems may be ambiguous and that the ensemble composition should be more clearly stated in the main text. While the five forecast systems correspond to distinct DCPP contributions, three of them are based on the EC-Earth3 model but differ in their initialization strategies and experimental configurations. We will revise the manuscript to clarify this point and explicitly state the ensemble composition.
    - Page 3, line 66: “These forecast systems are the EC-Earth3 (with three different initialisation strategies), IPSL-CM6A-LR and MPI-ESM1.2-HR.”

- We thank the reviewer for raising this point. In the revised manuscript, model drift will be addressed by applying a mean bias correction prior to calibration. Specifically, model outputs will be adjusted to the observed climatology by removing the model mean and replacing it with the corresponding observed mean, both computed over a common reference period defined by the forecast configuration. Given that the analysis uses start dates from 1960 to 2016 and forecast years 1–3, the common period across all lead times spans from 1963 to 2017. In addition, to ensure consistency with the cross-validation framework used in the calibration, the climatological means will be computed excluding the target year being calibrated. That is, for a given year, the reference climatology is estimated using all other years within the 1963–2017 period, except the year under consideration. This approach will reduce the impact of model drift while preserving the temporal variability of the simulations and avoids introducing artificial skill. The EQM calibration will then be applied to these adjusted time series.
    - Page 4, line 97: "Prior to calibration, model outputs are adjusted to the observed climatology by removing the model mean and replacing it with the corresponding observed mean. These climatologies are computed over a common reference period (1963–2017), defined by the availability of all forecast years (1–3) across the selected start dates (1960–2016). To ensure consistency with the cross-validation framework, the climatological means are estimated excluding the target year being calibrated. This step reduces the impact of model drift while preserving the temporal variability of the simulations (Boer et al. 2016). The EQM calibration is then applied to the adjusted time series."

- Pg 3, lines 67-68: "the indicators are computed for the three corresponding CMIP6 non-initialized historical forcing simulation models."
    - It would also be useful to include the number of members here.
- The number of members will be included in the main text:
    - Page 3, line 67
    - "the indicators are computed for the three corresponding CMIP6 non-initialized historical forcing simulation models (7 members of EC-Earth, 11 members of IPSL-CM6A-LR and 2 members of MPI-ESM1.2-HR)"

- Pg 3, line 83: "To ensure consistency across datasets, variables from all model simulations are first interpolated to the ERA5 grid (0.25°)."
    - Can you explain why you chose to regrid the model simulations to a higher resolution for comparison, rather than regridding the ERA5 data to a constant model resolution (e.g., all models and ERA5 data compared at 2.5 x 2.5 degrees)? No new information is gained from the models from regridding to the higher resolution of ERA5. It would be useful to know how much the results of this study would change, if all analysis was done with models and ERA5 at a 1.25 degree (or 2.5 degree) resolution, to ensure a fair comparison.
- Renewable energy users currently rely on reanalysis products for their applications; therefore, it is essential that decadal predictions are provided at the same spatial resolution. This represents a fundamental requirement for the provision of climate services, as these users often lack the computational capacity and technical

resources needed to further downscale or increase the spatial resolution of decadal prediction products themselves. We acknowledge that this downscaling procedure does not add new information from the models and may give the impression of resolving finer-scale variability than is intrinsically represented. As discussed above, we will clarify this point in the manuscript and ensure a more cautious interpretation of small-scale skill patterns.

- Page 3, line 83: "To provide information at a spatial scale relevant for impact-oriented applications, a statistical downscaling approach is applied to all model simulations, bringing them to the ERA5 grid (0.25°). This downscaling consists of a combination of spatial interpolation and calibration using Empirical Quantile Mapping (EQM), described below. A simple method is adopted, as previous studies have shown that more complex downscaling techniques often provide limited added value in terms of predictive skill compared to simpler approaches (e.g., Manzanas et al., 2018; Gutiérrez et al., 2019; Moreno-Montes et al., 2026), while minimising the risk of introducing artificial skill. After interpolation and before calibration, wind-specific…"

In addition, we are assessing the sensitivity of the results to spatial resolution by repeating the main analysis at a coarser grid (e.g., the coarsest model resolution) in order to provide a consistent benchmark.

- Pg 4, lines 90-91: "In addition, ERA5 wind speeds are adjusted using the Global Wind Atlas (GWA) to better represent long-term mean conditions at hub height."
  - I am not clear here as to whether the 100m wind speeds are used from ERA5 or the 10m wind speeds passed through the power law transform. It would be useful to be clearer with this. In the SM, it appears that 10m wind speeds were used from ERA5 and then extrapolated using the power law up to 100m. How would the results change if 100m wind speeds from ERA5 were used instead?
- We confirm that 10 m wind speeds from ERA5 are used and extrapolated to hub height using a power law formulation, ensuring methodological consistency with the model outputs, which do not provide native wind speeds at 100 m. Applying the same vertical extrapolation framework to both observations and model simulations avoids introducing inconsistencies associated with comparing differently derived wind products.

  Using native ERA5 100 m wind speeds would nevertheless represent an alternative observational reference. Native ERA5 100 m wind speeds are diagnostically derived within the IFS boundary-layer scheme using stability- and roughness-dependent formulations, rather than being directly simulated variables (Peña and Gryning, 2008). Consequently, ERA5 native 100 m winds and the power-law-derived winds represent different hub-height formulations. Following the reviewer's suggestion, we compared the 1961–2019 climatologies obtained from ERA5 10 m winds extrapolated to 100 m against native ERA5 100 m winds (Figure A1). The figure shows the difference between the power-law-derived and native ERA5 100 m climatologies. Results reveal spatially coherent differences, with predominantly negative values over land and positive values over ocean regions, indicating that the power-law approach tends to underestimate mean hub-height winds over land and overqwsestimate them over the ocean relative to ERA5 100 m winds. Consequently, using native ERA5 100 m winds would modify the mean state of the wind-based indicators and associated capacity factors. Although differences in the mean state are evident, the impact on predictive skill is expected to be more limited, since the evaluation primarily focuses on temporal variability and relative skill within a consistent ERA5–DCPP framework.

The purpose of the methodology is therefore not to reproduce native ERA5 100 m winds as accurately as possible, but to ensure that ERA5 and the decadal prediction systems are treated consistently within the same framework. Since the DCPP systems only provide near-surface winds, applying the same extrapolation approach to both ERA5 and model outputs allows the evaluation to focus on the predictive skill of the calibrated DCPP framework itself, rather than on differences arising from inconsistent hub-height wind formulations. We will clarify this methodological choice and discuss the associated uncertainty in the revised manuscript.

- Pg 3, line 88: “The same extrapolation framework is applied to both ERA5 and the prediction systems to ensure methodological consistency, since the forecast systems do not provide native wind speeds at 100 m.”

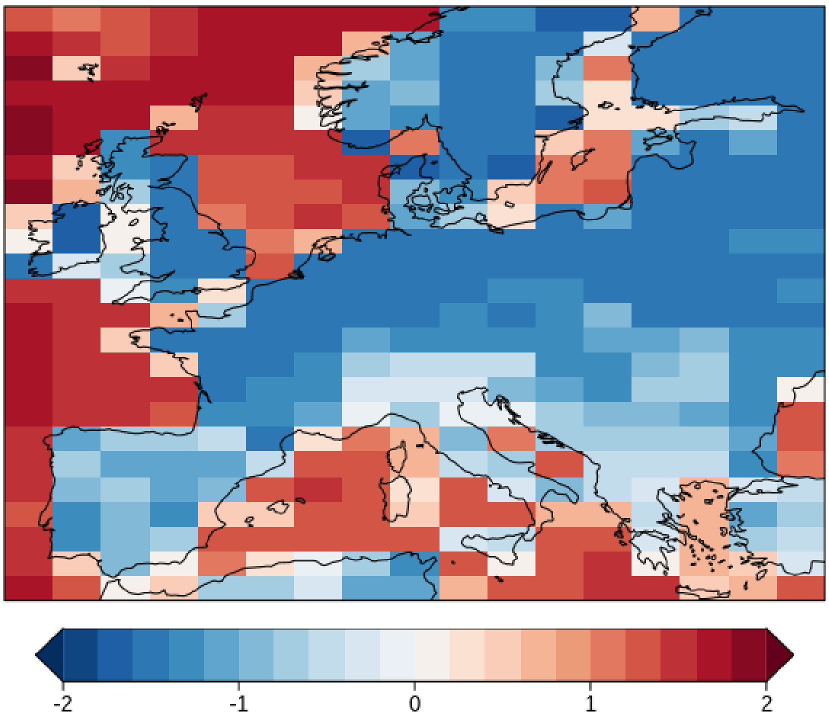

Figure A1. Difference between the 1961–2019 climatological mean 100 m wind speeds derived from ERA5 10 m winds using the power-law extrapolation and native ERA5 100 m wind speeds.

- Pg 4, lines 91-92: “Figure S1 shows the climatological ratio between ERA5 and GWA mean 6-hWIND over 1961–2019, which is applied as a pointwise correction to the ERA5 data.”

- With this pointwise bias correction of ERA5 to the global wind atlas, is this an additive (correcting the mean) or a multiplicative (correcting the spread) bias correction? It would be useful to have greater detail on how exactly this works.

- The adjustment of ERA5 wind speeds using the GWA is applied as a multiplicative correction factor, computed as the ratio between GWA and ERA5 mean wind speeds. This approach adjusts the mean state while preserving temporal variability, and we will clarify this in the manuscript with the next sentence:
  - Page 4, line 90: "Figure S1 shows the climatological ratio between ERA5 and GWA mean 6-hWIND over 1961–2019, which is applied as a multiplicative pointwise correction factor (GWA/ERA5) to adjust the mean wind speed while preserving the temporal variability of the ERA5 data."

- Pg 4, lines 95-96: "EQM maps the quantiles (percentiles) of the modelled data to the observed ones without assuming any underlying distribution and then it corrects the model biases."
  - With the Empirical Quantile Mapping, it would be useful to clarify the moments of the distribution that this method acts to correct. This is currently performed for each forecast year and member, but how would it change the results if the EQM was performed across all ensemble members for each of the models in turn? With the signal to noise issue (Scaife and Smith, 2018, 10.1038/s41612-018-0038-4), we cannot assume that the individual members can be regarded as equivalent realisations of climate variability as in the observations. Will the EQM method applied to each member in turn artificially inflate the variance? How is the signal to noise issue accounted for in this study?

- Empirical Quantile Mapping (EQM) corrects the distribution of the variable by adjusting empirical quantiles at regular intervals (here, percentiles from the 1st to the 99th, one by one). This approach allows the method to correct not only the mean and variance, but also higher-order moments such as skewness. We will clarify this in the manuscript:
  - Page 4, line 95: "By adjusting empirical quantiles at 1% intervals (from the 1st to the 99th percentile), EQM corrects the distribution of the variable, including the mean, variance, and higher-order moments."

  Regarding the calibration strategy, we acknowledge that an alternative approach would be to estimate the EQM transfer function from the pooled distribution of all ensemble members for each model. In this study, the member-by-member strategy has been used because it preserves a more comparable sample size between observations and model data during calibration. To increase the robustness of the estimated empirical distributions while maintaining this consistency, the calibration was performed using a moving temporal window around each target day.

  This approach is also consistent with the findings of Roberts and Leutbecher (2025), who first analysed the impact of pooled-member and member-by-member approaches on anomaly estimation, and subsequently evaluated their effect on ensemble calibration and reliability diagnostics. They showed that member-by-member anomaly estimation preserves statistically consistent spread-error relationships, whereas pooled-member approaches can introduce biases in ensemble reliability diagnostics. Although their framework differs from the present study, as it focuses on anomaly estimation and simpler calibration approaches rather than EQM-based full-distribution calibration, they similarly found that differences between pooled-member and member-by-member approaches become small when sufficient calibration data are available.

However, we acknowledge that ensemble members cannot be considered independent realisations of the observed climate and that signal-to-noise limitations are inherent to decadal prediction systems (Scaife and Smith, 2018). To assess the potential relevance of this issue in the calibrated system, we additionally evaluated the Ratio of Predictable Components (RPC; Figure A2), computed as the ratio between the predictable signal estimated from the ensemble mean and that represented by the ensemble members. In this framework, RPC values significantly larger than 1 indicate that the ensemble mean is more strongly correlated with observations than expected from the ensemble-member signal, a behaviour commonly associated with signal-to-noise issues in decadal prediction systems. Statistical significance was assessed using a 5-year block-bootstrap procedure in which start dates were resampled with replacement and the RPC was recalculated for each resampled dataset to estimate confidence intervals. Confidence intervals were estimated from 1000 bootstrap resamples.

RPC was analysed only in regions with ACC > 0, since negative ACC values do not allow a meaningful interpretation of RPC. The results show that there are not generally significant RPC greater than 1 across all indicators and seasons, although some localized and statistically significant regions remain, particularly for the NED indicator during JJA. Overall, the results suggest that signal-to-noise inconsistencies may affect specific regions and seasons, but do not dominate the behaviour of the indicators over the analysed domain.

We will clarify in the revised manuscript that future studies could further investigate the signal-to-noise characteristics of the calibrated system.

“Furthermore, as decadal prediction systems are known to be affected by signal-to-noise problems (Scaife and Smith, 2018), future studies could further investigate the impact of signal-to-noise characteristics of the calibrated ensemble or apply correction methods such as the NAO-matching proposed by Smith et al. (2020).”

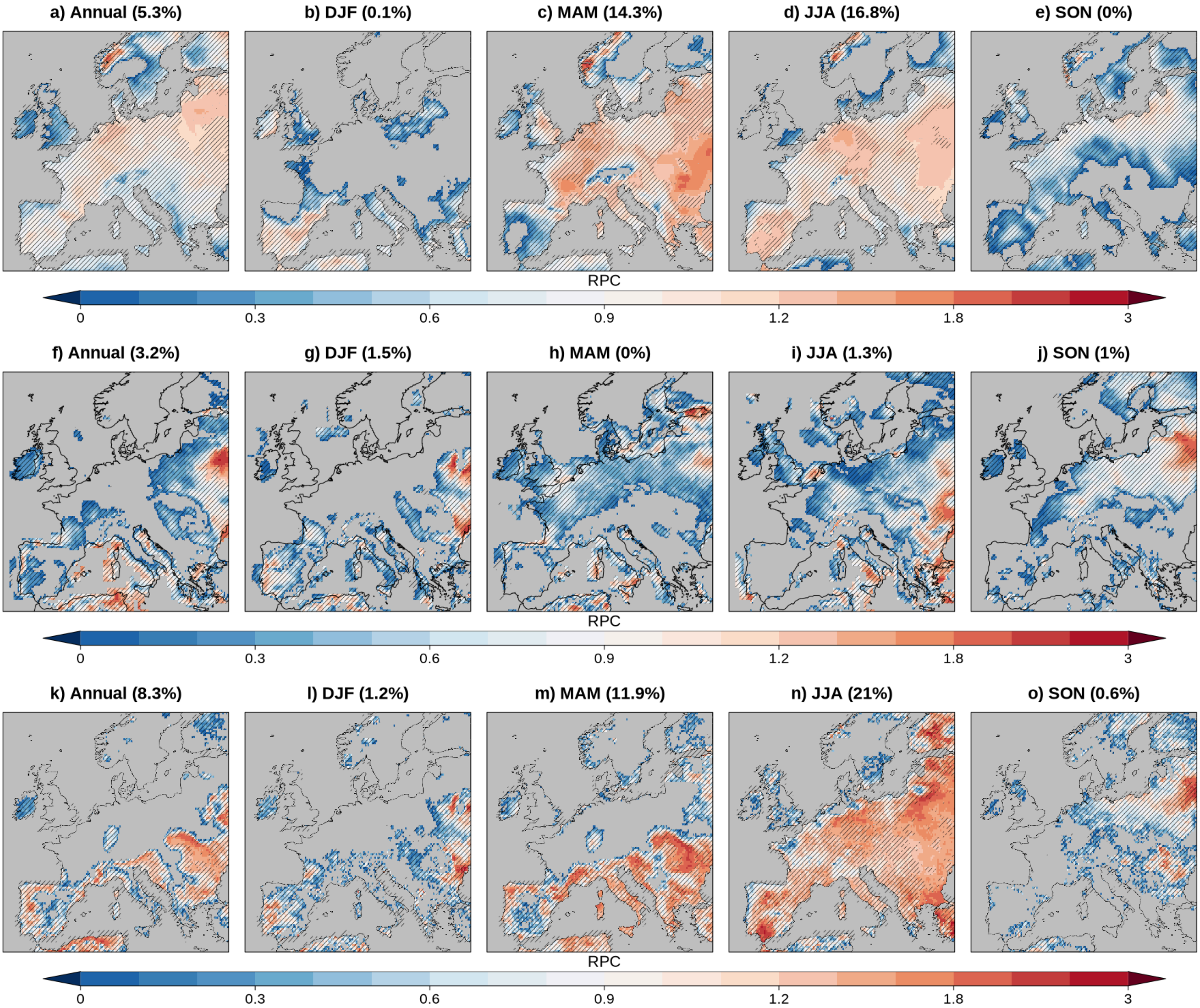

Figure A2. Ratio of Predictable Components (RPC) for PVpot (top row), WCF (middle row), and NED (bottom row) for Annual, DJF, MAM, JJA, and SON averages over forecast years 1–3. RPC was computed only in regions with positive ACC. Hatched areas indicate regions where RPC is not significantly different from 1 based on a bootstrap test using resampled start dates. Percentages in panel titles indicate the fraction of the analysed area with statistically significant RPC values.

- Pg 4 lines 111-114: “PVpot is computed as a dimensionless metric that combines RSDS, TAS and SFCWIND to represent the performance of photovoltaic cells under ambient environment. PV power generation at a specific location is determined by multiplying PVpot with the nominal installed capacity of that location.”
    - How do you determine the nominal installed capacities at each location here?
- Thank you for the comment. PVpot is a dimensionless indicator representing relative photovoltaic performance under given meteorological conditions and does not depend on prescribed installed capacities. Installed capacity would only be required to translate this indicator into absolute power generation for specific applications.

- Pg 7, Figure 1:
    - It would help the reader to have a title for the colorbar here (e.g., ACC). Also, it may help to have ylabels on the far left plots in each row, identifying “Full field” and “Residual” correlation skill. This goes for Figures 3, 5, and similar plots in the supplementary material as well (e.g., S5 onwards).
- The colorbar title and labels will be added to the correspondent figures.

- Pg 7 lines 182-183: “This reduction is mainly due to northern Europe and parts of Iberia, where MAM skill is generally non-significant.”
    - Presumably due to the lack of skill in northern Europe here?
- The sentence will be changed to:
    - “This reduction is associated with the lack of significant skill in northern Europe and parts of Iberia during MAM.”

- Pg 8 lines 210-211: “On the other hand, DCPP skill (Figure 1b) is lower than HIST skill (Figure S7b) in areas such as southern Iberia or the Alps”
    - To my eye, the DCPP (Figure 1b) shows no skill over the Alps and marginal skill over eastern Spain. This marginal skill over Spain increases slightly in HIST, but the skill over the alps in Figure S7b appears too small to be significant. It might be helpful to remove the statement about the Alps here for clarity, therefore.
- The statement about the Alps will be removed and sentence simplified to:
    - “On the other hand, DCPP skill (Figure 1b) is lower than HIST skill (Figure S7b) in areas such as southern Iberia.”

- Pg 9, Figure 2:
    - Can you explain how the trend of 0.00 in Figure 2b DJF Scandinavia can be statistically significant?
- Thank you for pointing this out. In the previous version, trend significance was assessed using the standard linear regression p-value, which did not account for temporal autocorrelation. We have now revised the trend significance test and applied a modified Mann–Kendall test following Hamed and Rao (1998), which corrects the variance of the test in the presence of serial dependence.

  With this revised approach, the DJF trend for Scandinavia in Figure 2b is no longer statistically significant. The significance of several other weak trends has also been reduced. We will update Figure 2, Figure 4, Figure 6 and Figure S8 accordingly and now provide the revised heatmaps, in which the trend values and significance markers reflect the autocorrelation-corrected test.

    - Page 6, line 160: “Trends are estimated from the annual time series using a linear fit, while statistical significance is assessed independently using a modified Mann-Kendall test that accounts for temporal autocorrelation following Hamed and Rao (1998).”

  As an example, the updated Figure 2 will be:

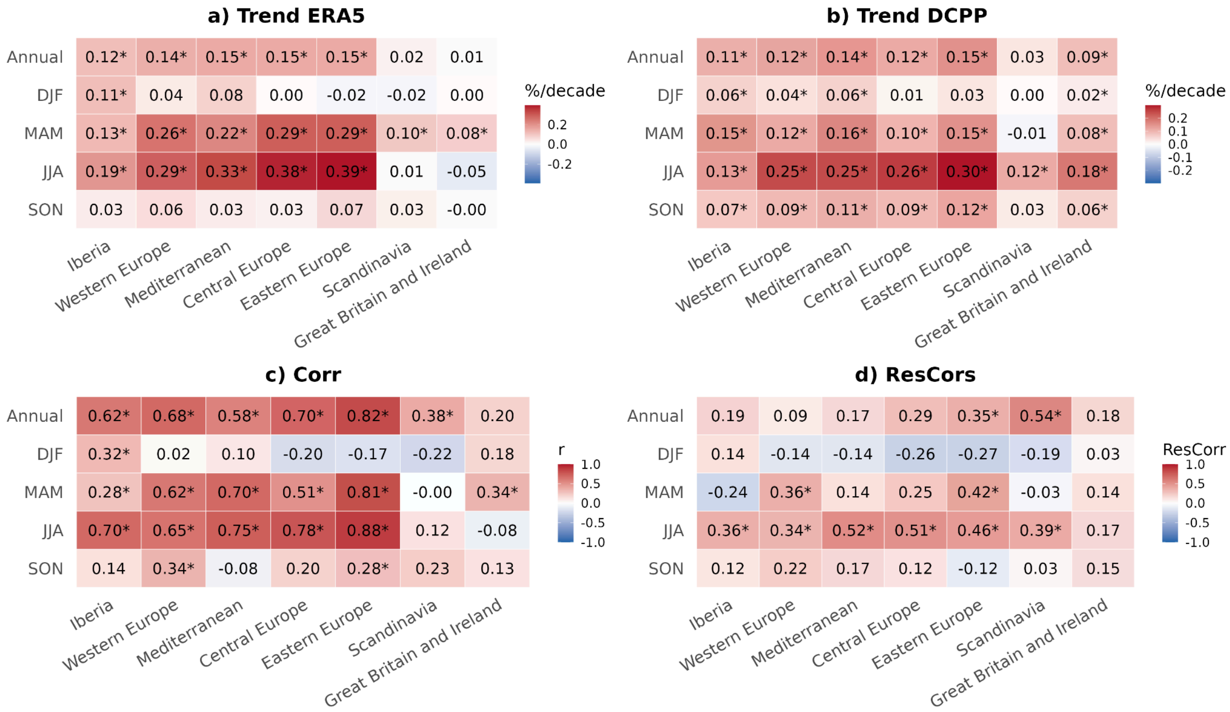

Figure 2. PVpot trends derived from ERA5 for years 1961-2019 with a three years rolling mean (a) and from the DCPP for the start dates 1960-2016 for over forecast years 1–3 (b), the correlation between ERA5 and DCPP (c), and the ResCorr between DCPP and ERA5 respect HIST (d) shown for the annual mean and for each season across the seven european sub-regions (Figure S4). Statistically significant trends and correlations at the 95% confidence level are marked with an asterisk. Trends significance are assessed using a modified Mann–Kendall test accounting for temporal autocorrelation.

- Figures 2, 4, 6:
    - In all of these figures, it would be helpful to make sure the colorbars for trends (a and b) and correlations (c and d) are the same in each figure. It makes it more difficult to compare the trends particularly when the colorbar scales are slightly different.
- Thank you for the suggestion. The colorbars will be corrected in the revised version ranging all of them from -1 to 1.

- Pg 10, lines 248-249: “The ResCorr isolates the added value of initialization beyond the historical simulations, providing insight into the role of internally generated variability.”
    - The residual correlations show the added value of initialisation, but do not necessarily isolate the internal variability. Initialisation can improve the response to external forcing when initialisation corrects the model response to external factors (Smith et al., 2019, https://doi.org/10.1038/s41612-019-0071-y). Consider rephrasing this sentence accordingly.

- We agree that residual correlation does not strictly isolate internally generated variability, as initialization may also improve the response to external forcing. We will rephrase this point to provide a more accurate interpretation.
    - Page 10, line 248-249: "The ResCorr isolates the added value of initialization beyond the historical simulations, providing insight into climate variations not captured by the historical simulations."

- Figures 3-6:
    - It would be useful to include the captions here, rather than refer back to the previous Figure through "…Same as Figure X". This would make it easier for the reader to interpret the findings. As mentioned above for previous figure, it would be very useful to include captions for the colorbar (e.g., ACC) and potentially ylabels identifying what is shown in each row.
- The complete captions and labels will be included in the revised manuscript.

- Page 12, lines 293-294: "For the annual mean and DJF (Figure 3f–g), initialization reduces WCF skill over Scandinavia and northern Europe, although HIST skill in these regions is already non-significant"
    - To my eye, there is no significant positive skill over Scandinavia and Northern Europe in either the DCPP or HIST results for WCF in Figure 3 or Figure S10. Suggesting that "initialisation reduced WCF skill over…" could be misleading here, as, although this statement is correct looking at Figure 3f-3g, it suggests the existence of predictable skill for WCF over Scandinavia and Northern Europe in the first place. I find it difficult to interpret the significance of initialisation reducing skill where there is none in the first place. Consider rephrasing or removing.
- We agree that this statement may be misleading in regions where skill is not statistically significant. We will therefore remove this sentence to improve clarity.

- Pg 12, line 295: "The skill reduction may be related to the high interannual variability of WCF in these regions (Figure S9f–g),"
    - In Figure S9, I find it very difficult to interpret whether the interannual standard deviation over Scandinavia and Northern Europe is higher than in other regions. To my eye, the magnitudes of interannual variability are similar in Scandinavia and Northern Europe to the rest of Europe, with the one exception to this being the Baltic region. I am not clear how the skill reduction here (i.e., less negative skill in HIST to more negative skill in DCPP) relates to interannual variability. Consider rephrasing or drop.
- We agree that this interpretation is not sufficiently supported by the figure and may be unclear. We will therefore remove this statement to improve clarity.

- Page 12, lines 302-304: "During JJA (Figure 3i), initialization increases skill over parts of eastern Europe and central Germany, while reducing it over Iberia, where HIST skill is already low (Figure S10d), resulting in a higher fraction of significant area for the DCPP than for the HIST"
    - In this case, the areas of significance over central Germany are very small so could be a result of noise. Additionally, as mentioned in the point above, I am not clear on the significance of negative skill (i.e., no skill) becoming more negative over Iberia. The implications of this to me are unclear.

- We agree that changes in regions where skill is non-significant, such as over Iberia, are not meaningful in terms of predictive performance, and we will remove this part of the sentence. We also acknowledge that the small areas of significant skill over central Germany may be influenced by sampling variability. We will clarify this point in the revised manuscript to ensure a more cautious interpretation.
    - Page 12, lines 302-304: "During JJA (Figure 3i), initialization increases skill over parts of eastern Europe, while small areas of significant skill over central Germany should be interpreted with caution, as they may be influenced by sampling variability."

- Page 12, lines 305-308: "Overall, model initialization affects WCF skill mainly in winter and summer: it leads to a marked reduction of skill in DJF, while it enhances skill in JJA."
    - I am not fully convinced on the evidence provided to justify this statement. 'Marked reduction of skill in DJF, as mentioned above, suggests that there was some predictable skill present in the first place, which there is not in Figure 3b. Similarly, the enhancement of skill in JJA only appears to occur over Northern Europe. I would consider rephrasing this statement to be more precise.
- We agree that the original statement was too strong and could be misleading, particularly in suggesting a reduction of skill in regions where skill is not statistically significant. We will therefore revise this sentence to provide a more precise interpretation. In particular, we will distinguish between the reduction of significant skill over southern Europe in DJF and the more localized improvements in JJA over eastern Europe, avoiding general statements at the seasonal scale.
    - Page 12, line 305: "Overall, initialization has a limited impact on WCF skill, with only localized improvements in JJA, mainly over eastern Europe."

- Pg 12-13, lines 321-323: "In addition, ResCorr is significantly positive in Central Europe in MAM, even though the correlation is not significant and trend agreement is weak, indicating the increase of skill due to the initialization in this case."
    - Could the significance in this case potentially be by chance? As with 95% significance we would expect 1 in 20 samples to show significant correlations by chance. This result is not reflected in the spatial plot in Figure 3h.
- We agree that, although the spatially averaged ResCorr over Central Europe is significantly positive, the signal is spatially limited and not consistently reflected in the corresponding map. This suggests that the regional significance may be influenced by localized areas of positive values and should therefore be interpreted with caution. We will revise the text accordingly to avoid overinterpretation.
    - Page 12-13, line 321: "In addition, ResCorr is significantly positive in Central Europe in MAM; however, this signal is driven by localized positive values (Figure 3h) and may be influenced by sampling variability."

- Pg 13, lines 323-324: "Other regions show non-significant but positive ResCorr values, suggesting local improvements due to initialization."
    - I am not clear as to whether insignificant residual correlation values of 0.1-0.2 suggest local improvements due to initialisation. As these are such low values, they could be the result of noise. Additionally, where there is no correlation skill shown in the region in Figure 4c, skill does not improve, rather

negative skill becomes less negative. Perhaps it would be more appropriate to focus the discussion on findings where there is both positive skill in Figure 4c and positive residual correlation skill in Figure 4d.

- We agree that non-significant and low positive ResCorr values do not necessarily indicate meaningful improvements in predictive skill. In particular, when correlation skill remains negative, a positive ResCorr may simply reflect a reduction in negative skill rather than a transition to useful predictive performance. We will revise the text to clarify this point and to focus the discussion on cases where both correlation skill and ResCorr indicate consistent improvements.
    - Page 13, lines 323-324: "Other regions show weak and non-significant positive ResCorr values. Where correlations are already positive, these may indicate localized improvements due to initialization, as in parts of western Europe during MAM. In regions with weak or negative correlations, however, they mainly reflect a reduction in negative skill rather than a transition to meaningful predictive skill."

- Pg 13, lines 330-331: "in other cases providing additional skill independently of it."
    - I am not clear on where initialisation provides additional skill independent of agreement in the trends. Is this referring to the example of MAM Central Europe in Figure 4d?
- We agree that the previous wording was unclear and may have suggested an interpretation not fully supported by the results. This statement will be revised to avoid implying additional skill independent of trend agreement and to focus on cases where both correlation skill and ResCorr are consistently positive.
    - Page 13, line 328: "Taken together, these results show that initialization can enhance skill in some regions, mainly where both correlation skill and ResCorr are consistently positive."

- Pg 13, lines 332-333: "Using climate projections, Pryor et al. (2005) have identified slight decreases in wind speed and wind energy density over northern Europe."
    - Are CMIP3 class models a fair comparison for the results shown here? Have there been other, more recent, studies that support the findings of this research?
- We agree that Pryor et al. (2005), based on CMIP3 models, is not directly comparable with the CMIP6-based simulations used in the present study. We will therefore revise this section to include more recent CMIP6-based analyses. In particular, Miao et al. (2023) reported historical decreases in surface wind speed across Europe during 1979–2014, identifying Europe as the Northern Hemisphere region with the strongest declining trends in several CMIP6 models, including the ones analysed in the present work. We will use this reference in the manuscript.
    - Page 13, line 332: "Using CMIP6 historical simulations, Miao et al. (2023) identified negative historical wind speed trends over Europe across a broad CMIP6 ensemble, including the three CMIP6 models analysed in the present study."

- Pg 13-14, lines 336-338: "Observational analyses by Vautard et al. (2010) and Torralba et al. (2017a) have revealed spatially heterogeneous near-surface wind trends across Europe, with predominantly negative signals in several regions."

    - What level of agreement is there between the observed regional trends shown here and in the studies referenced? Do the areas of trend increase/decrease shown in Figure 4c line up with previous analysis in this area?
- The spatial patterns of wind speed trends shown in Figure 4c are in partial agreement with previous studies. In particular, the negative trends over central and eastern Europe and the weak or heterogeneous signals over southern regions are broadly consistent with both Torralba et al. (2017) and Vautard et al. (2010).

  However, some discrepancies are found. In winter, differences are present with respect to Torralba et al. (2017), particularly over northern Europe and the British Isles, where our results show positive trends while their analysis indicates neutral or negative trends. Differences are also present with Vautard et al. (2010) over northern Europe and the British Isles, where negative trends are reported in their ERA-Interim analysis but positive ones are found here. These differences are likely related to the use of different datasets, time periods and methodological approaches. We will clarify this comparison in the revised manuscript.

    - Page 13, line 336: "Observational and reanalysis-based studies (e.g. Vautard et al., 2010; Torralba et al., 2017a) have reported spatially heterogeneous near-surface wind trends across Europe. Our results are broadly consistent in showing negative trends over central and eastern Europe and weak signals over southern regions. However, differences remain over northern Europe and the British Isles, where these previous studies reported neutral or negative trends while positive trends are found here, likely due to differences in datasets, periods and methodologies."

- Pg 14, lines 360-361: "In our case, the prior calibration of all climate variables against ERA5 ensures consistent magnitudes…"
    - Consistent magnitudes of what? Magnitudes of variability?
- We agree that the wording was unclear. We will revise the text to explicitly state that the calibration aligns the statistical distributions of the variables with ERA5, including both mean and variability, which supports the use of absolute thresholds.
    - Page 14, line 360: "In our case, the prior calibration of all climate variables against ERA5 reduces systematic biases relative to the observational reference, including biases in mean values and variability. This helps avoid artificial shifts in the frequency of threshold exceedances caused by biases in the raw model data."

- Pg 14-15, lines 370-373: "By contrast, both-Neff exhibits more spatially fragmented skill, with isolated regions of significant values. In solar-Neff and both-Neff, some areas appear as missing values because correlations cannot be computed when the indicator remains equal to zero for all years, and solar energy droughts are expectable when the amount of daily sun hours is low (Figure S13b, e)."
    - I am not clear on the presentation of NED for DJF in Figure 5. Figure S13b, g, and i, suggest that for much of Europe, the solar indicator is equal to zero for all years, due to the low amount of daily sunshine hours. In this case, most of the area in Figure 5b and 5g is solely dependent on the number of ineffective days for the wind resource, as the number of ineffective days for solar is

constant (i.e., all winter days are ineffective for solar in the greyed out region in Figure S13). Therefore, this would suggest that most of the area in Figure 5b and 5g does not show the combined NED index, as only one of the variables (e.g., wind speed) varies. It would be useful to clarify this and potentially modify the figure to show only the area where the NED index incorporates both solar and wind speed, as this may be misleading to the reader otherwise.

- As noted above, we agree that during DJF the solar component of the NED index shows very limited variability over most of Europe, as solar-Neff is effectively zero in large regions. As a consequence, the variability of NED in this season is primarily driven by the wind component. However, NED remains a compound indicator by definition, as it is based on the joint occurrence of ineffective wind and solar conditions. In winter, the lack of effective solar production implies that this condition is always fulfilled in a big part of the region, so NED effectively reflects the occurrence of ineffective wind days. This represents a seasonal regime of the compound indicator rather than a change in its definition. A similar, though more limited, situation occurs in some northern regions during SON. We will revise the manuscript to clarify this interpretation and avoid potential misinterpretation of the DJF results.
    - Page 14, line 366:
    - "Figure 5a–e shows the ACC between the DCPP and ERA5 of NED, while Figure S13 presents the corresponding ACC for solar-Neff (Figure S13a–e), wind-Neff (Figure S13f–j), and both-Neff (Figure S13k–o). The highest fraction of significant ACC values occurs in JJA (Figure 5d), when most of eastern Europe and parts of central Europe and Iberia exhibit significant skill. The annual mean (Figure 5a) and MAM (Figure 5c) also show significant values over parts of eastern Europe and Iberia. In DJF, solar-Neff remains equal to zero over most of Europe (Figure S13b), indicating that effective solar days are absent across large areas. Consequently, NED variability and skill in these regions are largely controlled by wind conditions (Figure 5b), except over southern Iberia where both components contribute to the index variability. A similar but more limited situation occurs over northern regions during SON. Outside these areas, skill in SON is generally weak and confined to parts of eastern Europe and the Baltics.

      The skill patterns of solar-Neff (Figure S13a-e) closely resemble those of solar PVpot (Figure 1a-e), while wind-Neff (Figure S13f-j) shows patterns similar to those of WCF (Figure 3a-e), as expected given that all three indicators reflect energy-production efficiency. By contrast, both-Neff (Figure S13k-o) exhibits more spatially fragmented skill.

      The spatial distribution of NED skill reflects the seasonally varying contribution of solar- and wind-related conditions. For the annual mean (Figure 5a), significant NED skill over parts of eastern Europe and Iberia broadly coincides with regions where both solar-Neff (Figure S13a) and wind-Neff (Figure S13f) show positive skill. Over parts of southeastern Europe, including the Carpathian region and Greece, localized NED skill is more consistent with solar-Neff than with wind-Neff. In MAM (Figure 5c), NED skill over southern Europe generally follows the solar-Neff signal, although

NED remains non-significant in several regions where solar-Neff shows skill. During JJA (Figure 5d), the widespread NED skill closely resembles the solar-Neff pattern, with lower values over parts of western Europe. Finally, in SON (Figure 5e), NED skill is weak and only locally significant. Its spatial pattern is more consistent with wind-Neff (Figure S13j), although some contribution from solar-Neff (Figure S13e) is also apparent."

- Pg 15, lines 376-378: "By contrast, over parts of the Alps and southeastern region, NED exhibits significant skill despite wind-Neff being non-significant, indicating a dominant contribution from solar-Neff."
  - In Figure 5a (annual NED correlations), I cannot clearly see a significant area of positive skill over the Alps. While there is some skill over eastern Europe, both solar-Neff and wind-Neff appear to show some significant positive skill over this region in Figure S13, although the magnitude of the positive skill is lower for wind-Neff.
- We agree that the reference to the Alps is not clearly supported by the figure and will be removed. Regarding the regional description, we acknowledge that the term "southeastern Europe" was too imprecise. We intended to refer to areas spanning eastern and southeastern Europe, including the Carpathian region and Greece, where NED shows localized areas of significant skill. In these regions, the signal is more consistent with solar-Neff than with wind-Neff, which generally shows weaker or non-significant values. We will revise the text to clarify this point.
  - Page 15, line 376: "Over parts of southeastern Europe, including the Carpathian region and Greece, localized NED skill is more consistent with solar-Neff than with wind-Neff."

- Pg 15: lines 378-379: "In DJF (Figure 5b), NED skill over Iberia and eastern Europe aligns mainly with wind-Neff skill (Figure S13g)."
  - As with the point described above for lines 370-373, I think this statement could be misleading, as only wind-neff skill varies over eastern Europe and much of Iberia. The solar-neff is effectively constant. In this case, surely the NED skill over eastern Europe can only align with wind-neff skill, as solar-neff is constant?
- We agree that, during DJF, NED variability over large parts of Europe is effectively controlled by wind conditions because solar-Neff remains equal to zero in many regions. This point will be clarified earlier in the revised paragraph describing Figure 5 and Figure S13, avoiding a separate interpretation of DJF NED skill solely in terms of wind-Neff correspondence.

- Pg 15, lines 381-383: "Finally, in SON (Figure 5e), NED skill is confined to parts of Poland and the Baltic region, closely matching the wind-Neff skill pattern (Figure S13j) and in points of southeastern Europe, which are skillful at solar-Neff (Figure S13e).
  - Much of the NED skill shown in Figure 5e is non-significant here. Additionally, both solar-Neff and wind-Neff show skill over Poland, suggesting that both sources contribute, rather than only the wind-Neff mentioned in this statement here. Additionally, the area of significant skill in southeastern Europe is small, suggesting that it could be related to noise.

- We agree that the SON pattern was overinterpreted in the original text. NED skill in SON is weak and mostly non-significant, so any attribution to a single component should be made cautiously. The revised text now states that the SON NED pattern is more consistent with wind-Neff, while acknowledging that solar-Neff may also contribute in some regions. We also avoid interpreting the small isolated areas over southeastern Europe as robust skill.
  - Page 15, line 381: “Finally, in SON (Figure 5e), NED skill is weak and only locally significant. Its spatial pattern is more consistent with wind-Neff (Figure S13j), although some contribution from solar-Neff (Figure S13e) is also apparent.”

- Page 15, Lines 385-386: “During periods of high solar availability (JJA and, in the southern region, MAM), NED skill is largely driven by solar-Neff, whereas in low-radiation seasons (DJF and SON) wind-Neff becomes the main modulator.”
  - As with the comments above, I think this conclusion may be misleading for the DJF result, as, for much of the European area shown, only the number of ineffective wind days contributes, as the number of ineffective solar days is effectively constant. For much of the region in DJF therefore, wind-(in-)Neff is the only modulator, so it is misleading to suggest that it is the main modulator. This point is not clarified in enough detail in the main text. Additionally, the suggestion that wind-Neff is the main modulator for SON may also be misleading, as this is only true for the limited area of skill over southeastern Europe, while for Poland/Baltic area, both the wind-Neff and solar-Neff appear to contribute.
- We agree that the original statement was overly general given the spatial variability and complexity of the results. Following the reviewer’s suggestions, we consider that a generalised conclusion on the seasonal modulation of NED based on a single dominant energy source is not sufficiently supported. We will therefore remove this paragraph in the revised manuscript to avoid overinterpretation and keep the discussion focused on the regionally consistent and robust features of the results.

- Page 16, lines 393-395: “In MAM (Figure 5h), initialization enhances skill over parts of southern and eastern Europe while reducing it over Iberia, leading to a modest improvement of DCPP skill relative to HIST (Figure 5c vs Figure S14c).”
  - Much of the ResCorr skill in Figure 5h is marginal and insignificant, suggesting that these areas could be due to chance. The difference in significant area of ~4% is relatively small. Therefore, the improvements from initialisation are marginal and the DCPP skill is approximately the same as the HIST (i.e., initialisation does not significantly degrade the skill).
- We agree that, in MAM, the ResCorr values are generally weak and spatially heterogeneous. This suggests that the apparent changes associated with initialization are likely influenced by sampling variability rather than reflecting a robust signal. We also acknowledge that the difference in the fraction of significant area between DCPP and HIST is small, and therefore does not support a clear improvement in skill due to initialization. We will revise the text accordingly to provide a more cautious interpretation.

    - Page 16, line 393: “In MAM (Figure 5h), ResCorr values are generally weak, indicating that the impact of initialization is limited, and that DCPP skill remains broadly comparable to HIST (Figure 5c vs Figure S14c).”

- Page 16: lines 398-400: “For this indicator, initialization generally increases predictability under conditions of higher sunshine duration (JJA and MAM) and reduces it under the opposite conditions (DJF and SON).”
    - Once again, any findings related to solar-Neff driving DJF results are misleading, as only the number of inefficient wind days contributes across most of Europe. The only season where initialisation appears to increase skill via ResCorr in Figure 5 is JJA, with MAM, DJF, and SON all showing few areas with any ResCorr skill. I would suggest this sentence is rephrased for clarity, as the MAM ResCorr result is not equivalent to the JJA ResCorr result.
- We agree that the original statement was too general and could be misleading. The clearest and most consistent impact of initialization is found in JJA, where DCPP shows higher skill than HIST. In the other seasons, including MAM, DJF and SON, the ResCorr signal is weak, spatially limited, or not robust, and does not support a consistent interpretation of either improvement or degradation of skill. We will revise the text accordingly.
    - Page 16, line 396: “In SON (Figure 5j), ResCorr negative values often occur in regions where skill is already non-significant, indicating a limited impact of initialization. Overall, a clear impact of initialization is mainly found in JJA, where DCPP shows higher skill than HIST. In the other seasons, the effect of initialization is limited or not robust, with weak and spatially heterogeneous signals that do not support a consistent improvement or degradation of skill.”

- Page 16, lines 403-404: “In JJA, Iberia, Western Europe, Mediterranean, Central and Eastern Europe display significant negative trends in both datasets, which results in significantly positive correlations (Figure 6c).”
    - It would be useful here to understand the relative contributions of trends in the number of ineffective wind and solar days, in absolute terms. I suspect that the trends in solar radiation contribute more strongly than the trends in wind speed. It would be useful to include greater detail on this here.
- We agree that analysing the relative contributions of wind and solar components provides important insight into the drivers of NED trends. To address this, we will include a new analysis of trends in solar-Neff and wind-Neff (Figure A3), which allows a more direct physical interpretation of the results. This analysis shows that in JJA the strongest NED trends are primarily associated with changes in solar-Neff, while wind-Neff exhibits weaker and less coherent signals. The manuscript will be revised accordingly.

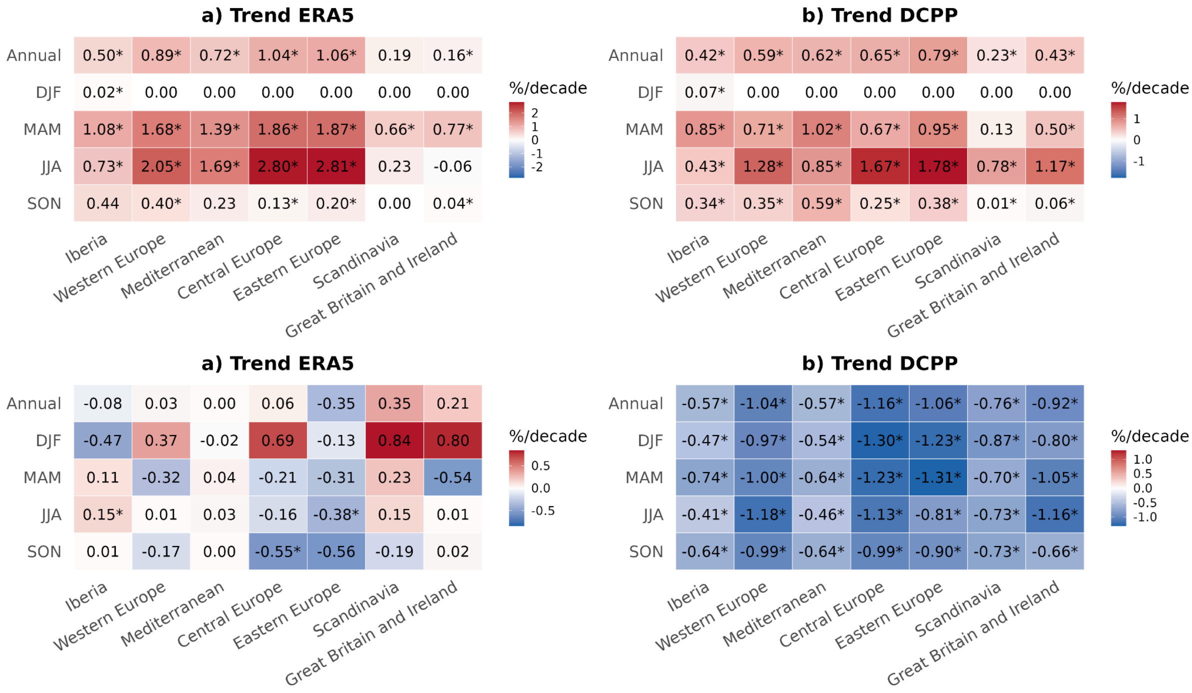

Figure A3. Regional trends for ERA5 (first column) and DCPP (second column) for solar-Neff (a-b) and wind-Neff (c-d). Statistically significant trends using a modified Mann–Kendall test accounting for temporal autocorrelation at the 95% confidence level are marked with an asterisk.

- ○ Page 16, line 417: “To better understand the drivers of NED variability and trends, solar-Neff and wind-Neff trends are analysed separately (Figure SX). The results show a clear seasonal dependence. In JJA, the strong negative NED trends in both ERA5 and DCPP are primarily associated with marked increases in solar-Neff (Figure SXa-b), while wind-Neff changes are weaker (Figure SXc-d). In MAM, solar-Neff also drives NED trends in ERA5, although with smaller magnitude, whereas in DCPP this signal is less clear due to strong negative trends in wind-Neff. In DJF, solar-Neff shows negligible variability over most regions, and NED trends are therefore directly anticorrelated with wind-Neff. In SON, both components exhibit mixed signals in ERA5, while in DCPP negative wind-Neff trends tend to dominate, leading to positive NED trends.”

- Page 16: lines 418-419: “NED variability is largely controlled by the energy source that dominates effective production in each season: wind in DJF and SON, and solar in JJA and, secondarily, in MAM.”
  - ○ See the above points on the number of ineffective wind days in DJF being the only contributor to NED here.
- We agree that the original statement was too general, particularly for DJF. In this season, solar-Neff shows very limited variability over most of Europe, so NED variability is directly anticorrelated with wind conditions. Following the revision described above, the interpretation of NED variability has been reformulated based on the direct analysis of solar-Neff and wind-Neff trends (Figure SX), providing a more precise and seasonally consistent description.

- Figures 2, 4, 6
  - It would be useful to have subplot headings here, to convey the information more clearly to the reader, these could be: “a) ERA5 trends”, “b) DCPP trends”, “c) Corr”, “d) ResCorr”.
- We will add the subtitles to the corresponding figures.

- Page 17, lines 429-430: “Additionally, DCPP trends show seasonal patterns that broadly resemble the future evolution of hybrid energy droughts projected by (Kapica et al., 2024), particularly in DJF and SON.”
  - It would be useful to have further clarification on how these resemble the results of Kapica et al., 2024, particularly regarding the similarities and differences of the NED metric used here compared to the hybrid metric from Kapica.
- We agree that the comparison with Kapica et al. (2024) required further clarification. A better qualitative agreement is found in MAM, where both studies show a north-south contrast (decreasing drought conditions in southern Europe and increasing conditions in northern regions), and to a lesser extent in JJA. In DJF, some regional agreement is also found, particularly over northern and eastern Europe, where both studies indicate increasing energy drought conditions. In addition, discrepancies remain in southern regions such as Iberia and France, where the two studies show opposite trends. In SON, the agreement is weak, with Kapica et al. (2024) indicating decreasing drought conditions while our results show a general increase.

  We will revise the manuscript accordingly and clarify that Kapica et al. (2024) use a percentile-based definition of droughts, whereas the NED metric employed here is based on fixed, physically motivated thresholds. The comparison is therefore limited to qualitative spatial patterns.

  The following sentences will be added:

  - Page 17, line 429: “Additionally, DCPP trends show some qualitative similarities with the future evolution of energy droughts projected by Kapica et al. (2024), particularly in MAM and over parts of northern and eastern Europe in DJF. However, Kapica et al. (2024) use a percentile-based definition of droughts, whereas the NED metric employed here is based on fixed physically motivated thresholds. The comparison is therefore limited to broad spatial patterns, and important regional differences remain, particularly in SON.”

- Page 18, lines 445-446: “This strong seasonal dependence indicates that PVpot predictability is highest during periods of higher and persistent solar radiation, while such predictability is more limited under weaker and more variable radiative conditions.”
  - To what extent is the highest predictability during the summer driven by the fact that the trends are strongest during the summer (Figure 2)?
- We agree that the higher predictability observed in summer may be highly related to the stronger and more spatially coherent trends shown in Figure 2, which can enhance correlation skill. However, the original statement linked this behaviour to

more persistent solar radiation, which was not explicitly assessed in this study. To avoid introducing an unsupported interpretation, we will remove this sentence in the revised manuscript.

- Page 19: lines 455-457: "Significant skill is consistently found over parts of eastern Europe across seasons, while other regions show skill only seasonally: central Europe in JJA, southern Europe in DJF, and northern Europe in MAM and SON."
    - I am not convinced that significant skill is found over southern Europe in DJF (Figure 3b). While there is some skill over eastern Europe, the remaining areas of significant positive skill are small and spatially heterogenous.
- We agree that the areas of significant skill over southern Europe in DJF are limited, spatially fragmented, and not particularly robust. We will revise the text to reflect this more accurately.
    - Page 18, line 455: "Significant skill is consistently found over parts of eastern Europe across seasons, while other regions show weaker and more seasonally dependent skill, including central Europe in JJA and northern Europe in MAM and SON."

- Page 18, lines 459-460: "Initialization tends to enhance skill in summer and reduce it in winter, while its impact is generally small in the transition seasons."
    - Most of the ResCorr in Figure 3i is over Eastern Europe, suggesting that the enhancement of skill in summer is more regionally constrained than suggested in this statement. The reduction of skill during the winter is mostly negative skill becoming more negative, suggesting that there wasn't skill to reduce in the first place. Consider rephrasing.
- We agree that the original statement was too general. In JJA, the enhancement of skill is mainly confined to eastern Europe, while in DJF the negative ResCorr values largely reflect already low or negative skill becoming more negative rather than a meaningful reduction of predictive skill. We will revise the text accordingly.
    - Page 18, line 458: "The impact of initialization is regionally dependent and generally limited. The clearest skill enhancements are found in JJA over parts of eastern Europe, while in other seasons the impact of initialization is weaker and spatially heterogeneous."

- Page 19, lines 466-469: "Seasonal differences in NED skill reflect the dominant energy source. During high-radiation periods (JJA and partly MAM), NED predictability is mainly driven by solar-Neff, whereas during low-radiation seasons (DJF and SON) wind-Neff becomes the main contributor."
    - See the above points on the number of inefficient wind days being the only contributing factor to skill during DJF. Consider rephrasing and/or changing narrative above.
- We agree that the original statement was too general, particularly for DJF. We will revise the text to provide a more precise and regionally consistent interpretation of the relative contribution of wind and solar components across seasons.
    - Page 19, line 466: "Seasonal differences in NED skill reflect the relative contribution of wind and solar components. In JJA, NED predictability is mainly driven by solar-Neff. In DJF, variability over much of Europe is largely driven by wind conditions, while in MAM and SON the relative contribution of both components varies regionally."

- Page 19, line 480: “However, ERA5 has been identified as the best reanalysis for wind (Ramon et al., 2019)”
    - It would be useful for the reader to be more specific about the particular characteristics of ERA5 which make is preferable to other reanalysis products here.
- We agree that the original statement was too general. Ramon et al. (2019) show that ERA5 outperforms other reanalyses primarily in terms of correlation with observations and representation of temporal variability, with statistically significant improvements over a multi-reanalysis mean in a substantial fraction of locations. However, differences among reanalyses remain large for mean wind and long-term trends. We will revise the text to clarify these aspects.
    - Page 19, line 480: “However, ERA5 has been identified as one of the most reliable reanalysis datasets for near-surface wind variability, particularly in terms of correlation with observations and representation of temporal variability at turbine-relevant heights (Ramon et al., 2019).”

- Page 19, lines 492-494: “In addition, hybrid approaches that combine percentile-based metrics with physically motivated thresholds may help balance climatological comparability and operational relevance when assessing compound renewable-energy risks.”
    - How can percentile-based metrics be combined with physical thresholds in a hybrid approach? Does this involve selecting the percentiles based on physical thresholds or vice versa?
- We agree that the original statement was unclear. To avoid introducing ambiguity, we will simplify this part of the discussion.

Technical corrections

- Pg 8 Line 216: “Figure 2 summarizes the regional behaviour of PVpot long-term trends across the seven European sub-regions (Figure S2)”
    - Figure S2 should be Figure S4
- Pg 12, line 301: “and the DCPP (Figure 3c) has higher skill values than HIST (Figure S10h)”
    - Figure S10h should be Figure S10c
- Pg 14, line 353: “…almost all seasons closely resembles the patterns found for the WCF standard deviation (Figure S8f–j),”
    - Figure S8f-j should be Figure S9f-j
- Pg 15, line 373: “and solar energy droughts are expectable when…”
    - Replace ‘expectable’ with ‘to be expected’ or ‘expected’
- All these technical corrections will be changed in the revised version of the manuscript

Gutiérrez, J. M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., ... & Pagé, C. (2019). An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross‐validation experiment. International journal of climatology, 39(9), 3750-3785.

Hamed, K. H., & Rao, A. R. (1998). A modified Mann-Kendall trend test for autocorrelated data. Journal of hydrology, 204(1-4), 182-196.

Hewitson, B. C., Daron, J., Crane, R. G., Zermoglio, M. F., & Jack, C. (2014). Interrogating empirical-statistical downscaling. Climatic change, 122(4), 539-554.

Manzanas, R., Gutiérrez, J. M., Fernández, J., Van Meijgaard, E., Calmanti, S., Magariño, M. E., ... & Herrera, S. (2018). Dynamical and statistical downscaling of seasonal temperature forecasts in Europe: Added value for user applications. Climate Services, 9, 44-56.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., ... & Thiele‐Eich, I. (2010). Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. Reviews of geophysics, 48(3).

Maraun, D., & Widmann, M. (2018). Statistical downscaling and bias correction for climate research. Cambridge University Press.

Moreno-Montes, S., Delgado-Torres, C., Duzenli, E., Pérez-Zanón, N., Marcos-Matamoros, R., & Soret, A. (2026). Comparative analysis of statistical downscaling methods for multi-model decadal climate predictions over Western Europe. Climate Services, 42, 100639.

Peña, A., & Gryning, S. E. (2008). Charnock's roughness length model and non-dimensional wind profiles over the sea. Boundary-layer meteorology, 128(2), 191-203.

Roberts, C. D., & Leutbecher, M. (2025). Unbiased calculation, evaluation, and calibration of ensemble forecast anomalies. Quarterly Journal of the Royal Meteorological Society, 151(771), e4993.

Soares, P. M., Cardoso, R. M., Lima, D. C., & Miranda, P. M. (2017). Future precipitation in Portugal: high-resolution projections using WRF model and EURO-CORDEX multi-model ensembles. Climate Dynamics, 49(7), 2503-2530.

WMO (World Meteorological Organization). (2011). Climate Knowledge for Action: A Global Framework for Climate Services–Empowering the Most Vulnerable, Report of the High-Level Taskforce for the Global Framework for Climate Services.