Response to RC1

This is a well-written and structured manuscript, easy to read and follow meanwhile without losing the rigour of scientific work. It deals with the evaluation of atmospheric models for simulating solar and wind power in California, with multiple layers of investigation, i.e., evaluating the impact of using different atmospheric models, spatial resolutions, and solar and wind power generation models. Answering these questions would help users to understand the impact of different factors on solar and wind power calculations, and to decide on the best solution given the available resources. This is particularly important in the context of the current global transition toward a carbon-neutral energy system. I believe this manuscript could be further improved by addressing the points listed below:

Thank you for your careful reading of the manuscript and for your positive comments! In response to your questions and suggestions, we provide detailed replies to each point below. Following your recommendations regarding figure formatting, we have updated the original Figs. 3-7, 9, and 11-16. In addition, because the EIA data for 2024 have recently become available, we updated all figures related to EIA and observations (reanalysis), except for Figs. 1, 7, and 14. The additional year of observations increases the sample size for EIA plant-level analyses, although the climatological comparison results remain almost unchanged.

1. The necessity or irreplaceability of the work needs to be sharpened. What makes it so important to evaluate solar and wind power generation derived from the atmospheric models (or meteorological data sets) you mentioned in California? What are their special things compared with other alternatives, or can they represent the others?

Thank you for raising this question, which helps clarify the contribution of this work!

First, regardless of the accuracy of datasets, meteorological data that can be used *directly* to derive wind and solar power generation are limited. Energy modeling generally requires at least hourly meteorological variables critical to wind and solar generation, which already excludes most existing climate simulations. Wind power requires winds at the turbine hub height, while solar power requires the total surface downwelling shortwave radiation and either the direct or diffuse components. This further restricts the available datasets, including widely used reanalysis products such as ERA5 and MERRA2, which typically provide only near-surface wind speed and total surface downwelling shortwave radiation. When direct datasets are unavailable, key meteorological fields are commonly reconstructed using temporal interpolation, vertical extrapolation based on boundary-layer wind profile assumptions, and radiation decomposition. Given our ability to perform high-frequency, km-scale simulations directly, generating physically consistent fields through explicit modeling is therefore a natural choice.

Second, although the importance of km-scale simulations has been emphasized in many studies, most existing work has focused on Europe. In the United States, and particularly in California with its complex topography, relevant studies are very limited. To our knowledge, PLUSWIND's evaluation of HRRR represents one of the few studies addressing such applications, and it evaluates only wind power. Moreover, HRRR is based on WRF, a limited-area regional model, similar to most existing km-scale modeling studies. Consequently, it remains unclear whether many of the conclusions drawn from such studies are transferable to other modeling frameworks, particularly global climate models. To our knowledge, global climate models have rarely treated energy applications as a core focus, whereas one of the development goals of E3SM is to better support energy-relevant applications. Expanding the model sample is therefore valuable in itself; in our case, the impact of using different models is non-negligible.

In this context, the evaluation of the E3SM 25kmNARRM by Lee et al. 2025 represents the first assessment of E3SM for energy applications, and one of the primary motivations of this study is to examine whether the poor performance identified at 25km resolution over California can be substantially improved in km-scale simulations. While this paper focuses on present-day conditions, the simulations are part of paired experiments designed to contrast present and future conditions, This provides additional benefit, given that HRRR and other widely used models are restricted to historical periods.

We have added these points to the Introduction section.

References

Millstein, D., Jeong, S., Ancell, A., & Wiser, R. (2023). A database of hourly wind speed and modeled generation for US wind plants based on three meteorological models. *Scientific Data*, *10*(1), 883.

Lee, H. H., Arthur, R. S., Golaz, J. C., Signorotti, M. V., Wert, J. L., & Watson, J. P. (2025). Using multiple high-resolution datasets to benchmark the energy exascale earth system model (E3SM) for renewable resource assessment. *AIP Advances*, *15*(10).

2. Adding information about the status quo of renewable energy installation in California and how much they cover the local electricity demand may help readers further understand the importance of your work.

Thanks for this helpful suggestion! We have added the following sentences to the Introduction:

> “According to the California Energy Commission (CEC), California currently has approximately 87.8 GW of installed electric generation capacity distributed across more than 1,600 power plants statewide in 2023 (\url{https://www.energy.ca.gov/data-reports/energy-almanac/california-electricity-data/2023-total-system-electric-generation}, last access: February 20, 2026). Natural gas remains the dominant technology, accounting for about 39.7 GW (45\%) of total nameplate capacity. Renewable energy resources contribute approximately 32.9 GW (37.5\%) of installed capacity, including about 20.9 GW of solar PV capacity (24\%) and 6.3 GW of wind capacity (7\%). In terms of electricity generation rather than installed capacity, solar and wind energy supplied approximately 19.2\% and 6.5\%, respectively, of California’s in-state electricity generation, together accounting for about one quarter of annual energy supply.”

3. The limitations of this study should be uncovered and discussed in the manuscript.

Previously, the limitations were discussed within the Results and Conclusions. Based on your suggestion, we have added a separate paragraph in the Conclusions to summarize and discuss the limitations of this study.

> “Several limitations of this study are worth noting. First, although the EIA monthly data represent the best available ``observations'' they are not automatically collected, and the quality-control procedures used to derive monthly total generation are unclear, and imputation introduces an additional concern. As a result, we cannot disentangle the impacts of data processing or operational losses, such as outages and curtailment, nor can we fully avoid the effects of imputation, which would otherwise substantially reduce the effective sample size and damp the spread of interannual variability. Second, none of the E3SM/SCREAM simulations conducted here are weather hindcasts, i.e., the simulated timeseries do not correspond one-to-one with historical observations. Thus, our evaluation is limited to climatological statistics rather than specific time series or historical events. Finally, we did not conduct an extensive sensitivity analysis and instead chose to follow the default assumptions of

PySAM, as detailed plant-level optimization of the energy model is not the primary focus of this work."

4. The conclusions you have made based on the results in the work, are they applicable to other regions? Or, are they applicable to other models?

Thanks for asking this! The conclusion that km-scale simulations perform substantially better than O(10 km) simulations for wind power generation is likely applicable to other regions with complex terrain, as topography is one of the most dominant lower boundary conditions controlling winds. For PV power generation, sensitivity to different meteorological models may also be broadly applicable, since the representation of clouds (especially low clouds) remains one of the largest sources of uncertainty in atmospheric models. The sensitivity of PV generation to axis tracking arises from the geometric design of PV panels and thus should be robust. Whether the overestimation found in PySAM that increases with the number of tracking axes applies to other power generation models remain an open question.

The conclusion that PV power generation is relatively insensitive to horizontal resolution may, however, only apply to regions where mesoscale cloud-precipitating processes are not prevailing. Although California contains extensive mountain ranges, precipitation is primarily associated with large-scale systems (atmospheric rivers) and is relatively infrequent overall, particularly during summer when solar power is at the peak. Local orographically forced precipitation and mesoscale convective systems are largely absent.

We have added these discussions to the corresponding parts of the Conclusions.

5. Are monthly CF observations really able to identify the difference of the high- and coarse-resolution atmospheric modeling outputs by evaluating their derived power results? The fluctuations in renewables, especially in wind, have been largely filtered out at the monthly scale, for which high-resolution models are good at solving while coarse-resolution models are not. Would not evaluating the direct solar and wind output variables from atmospheric models against observations from, e.g., synoptic stations and weather masts, make more sense for this purpose? Since hourly or finer time scale CF are not reachable as mentioned in the manuscript.

Yes, monthly CFs may smooth out sub-hourly or hourly variability, and EIA data limitations prevent us from directly assessing the impact of this smoothing. As you noted, most available meteorological station observations are near-surface (10-m). Previous studies have shown that interannual variability in wind power generation is only weakly correlated with observed near-surface wind speed (Millstein et al. 2022), motivating PLUSWIND to evaluate HRRR using EIA generation data rather than surface wind observations. Wind speeds at turbine hub height must therefore be extrapolated from near-surface winds, typically using power-law assumptions, whose exponents depend on atmospheric stability and surface roughness and perform poorly under stable conditions (Frank et al. 2020). This is particularly relevant in California, where wind power generation tends to be stronger at night, when atmospheric stability is higher and associated uncertainties are larger.

Tall-tower wind observations would provide a more direct benchmark for hub-height winds, but to our knowledge no such publicly available datasets exist for California.

We have added this discussion to the subsection “Uncertainty in EIA monthly data and interpretation of discrepancies”.

References

Frank, C. W., Pospichal, B., Wahl, S., Keller, J. D., Hense, A., & Crewell, S. (2020). The added value of high resolution regional reanalyses for wind power applications. *Renewable Energy*, *148*, 1094-1109.

Millstein, D., Bolinger, M., & Wiser, R. (2022). What can surface wind observations tell us about interannual variation in wind energy output?. *Wind Energy*, *25*(6), 1142-1150.

6. Please justify in this manuscript why this work chose to use grid-cell (figure 3, 4, 5, 8; 11, 12) based evaluation, instead of evaluating over every plant (figure 9; 15) all the time?

Thanks for asking this! This choice was made because our primary goal is model evaluation at the model-resolution (i.e., grid-cell) scale, rather than plant-level applications. This approach has several advantages:

1) It avoids artificially reducing the spread of the modeled climatological seasonal cycle. In plant-level evaluations, multiple plants located within the same model grid cell would be represented identically. This would substantially reduce the apparent monthly climatological spread and make modeled variability appear artificially smaller than that from observations.

2) Aggregating to the model grid ensures consistency between plant-level evaluations and assessments of spatial patterns across California, as both are conducted on the same grid. This facilitates a coherent transition between statewide spatial perspectives and local perspectives.

We have added this explanation to the subsection “EIA monthly CF data”:

> “We adopted a gridcell-based evaluation framework instead of plant-level evaluation for the following reasons: 1) It avoids artificially reducing the spread of the modeled climatological seasonal cycle. In plant-level evaluations, multiple plants located within the same model grid cell would be represented identically. This would substantially reduce the apparent monthly climatological spread and make modeled variability appear artificially smaller than that from observations. 2) Aggregating to the model grid ensures consistency between local evaluations and assessments of spatial patterns across California, as both are conducted on the same grid. In the seasonal-cycle analysis, we applied a weighted average based on total nameplate capacity within each grid cell, so that the EIA mean line remains consistent across comparisons at different model resolutions.”

7. The study indicated in line 99 that the only difference between SCREAM-3kmCARRM and SCREAM-800mCARRM is the horizontal resolution, and in line 100-101 it also indicated that the physical parameterization is sensitive to horizontal resolution. Then, why do you think using the same suit of physical parameterization in these two experiments is fair and reasonable?

We apologize for the ambiguity caused by the original wording. Lines 100–101 were referring to other GCMs, in which physical parameterizations may depend strongly on horizontal resolution; this statement does not largely apply to SCREAM. In SCREAM, the turbulence parameterization employs a scale-aware mixing length, and the model has been shown to exhibit good scale awareness over resolutions ranging from approximately 100 m to 5 km (a larger fraction of turbulent kinetic energy is explicitly resolved as resolution increases, while the representation of clouds and thermodynamics does not change substantially with resolution) (Bogenschutz et al. 2023). Such properties require deliberate model design and are often absent in models that are tuned for a specific resolution regime. As a result, turbulence schemes in other GCMs that lack this scale-aware property may artificially exaggerate the contribution of subgrid transport at 800 m resolution.

On the other hand, we acknowledge that horizontal turbulent mixing should ideally be enabled in the 800 m simulations. This option was not yet implemented in the model version used here, which may have partly contributed to the smaller-than-expected sensitivity. This limitation is related to the turbulence gray-zone problem (e.g., Wyngaard, 2004; Chow et al., 2019; Honnert et al., 2020)

We have removed the original sentence and revised the discussion in the conclusion:

> "Regarding the lack of large sensitivity between 3.25 km and 800 m SCREAM-RRMs, this likely depends on SCREAM's turbulence scheme and may not necessarily generalize to other models. SCREAM's SHOC turbulence scheme is scale-aware and scale-insensitive, meaning that a larger fraction of turbulent kinetic energy is explicitly resolved as resolution increases, while the representation of clouds and thermodynamics does not change substantially with resolution \citep{Bogenschutz2023}. Turbulence schemes lacking this property, which are generally used in coarser GCMs, may artificially exaggerate the contribution of subgrid transport at 800 m resolution.
>
> On the other hand, the muted sensitivity may also be attributed in part to model errors associated with the turbulence gray zone \citep[e.g.,][]{Wyngaard2004,Chow2019,Honnert2020}; see also discussions in \citet{Zhang2025100m}. Specifically, horizontal turbulent mixing (neglected in most PBL schemes including SHOC) may become non-negligible at 800 m, especially in complex terrain where three-dimensional effects matter \citep[e.g.,][]{Juliano2022,Arthur2025b}."

8. Since this work recognized the effect of using different loss schemes in the generation estimate models, it should disclose the details of the wake scheme used in the PySAM/SAM, even though it is the default one. A brief intro as you did to PLUSWIND in lines 263-264 would do the job for PySAM/SAM. A great solution would also be adding words justifying why a certain wake scheme is chosen to be used in the workflow.

Thank you for raising this point! In PySAM.Windpower, we use the WindPowerNone model with the default parameter set. The wake model employed is the simple wake model ('wind_farm_wake_model': 0.0), together with an external wake loss of 1.1% ('wake_ext_loss': 1.1). Documentation for the simple wake model is available at https://samrepo.nrelcloud.org/help/wind_power.html. This model computes distances between downwind and crosswind turbines based on their relative positions within a wind farm and accounts for the effects of power coefficient, thrust coefficient, and turbulence intensity on wind speed. The default parameter set specifies relative x and y coordinates for 32 turbines within a generic wind farm. As noted in our response to the first question, because the primary focus of this study is model evaluation and the complexity of meteorological datasets, we chose to follow the default settings in PySAM to maintain focus.

We have added this description along with more details of PV model parameters in the subsubsection “Coupling with PySAM to estimate wind and solar CFs”:

> “The default parameter set is used for \emph{PySAM.Windpower.default("WindPowerNone")}, including the use of a generic turbine power curve without plant- or manufacturer-specific corrections, which neglects variations in turbine physical characteristics such as rotor diameter and rated capacity, and the use of fixed 90 m hub-height wind inputs without turbine-specific height correction. The default wake calculation includes the simple wake model (\url{https://samrepo.nrelcloud.org/help/wind_power.html}, last accessed: February 20, 2026), together with an external wake loss of 1.1\%. The simple wake model computes distances between downwind and crosswind turbines based on their relative positions within a wind farm and accounts for the effects of power coefficient, thrust coefficient, and turbulence intensity on wind speed. The default parameter set specifies relative x and y coordinates for 32 turbines within a generic wind farm. Because the primary focus of this study is model evaluation and the complexity of meteorological datasets, we chose to follow the default settings in PySAM to maintain focus.
>
> The \emph{PySAM.Pvsamv1.default("FlatPlatePVNone")} configuration is a detailed PV model that includes separate representations of the module, inverter, and cell temperature. The default use of nominal operating cell temperature (NOCT) model \citep{SAMPV2018} relies on effective irradiance transmitted to the cell, air temperature, and near-surface wind speed. In addition to the default fixed-axis tracking configuration, single- and dual-axis tracking configurations are enabled by setting \emph{SystemDesign.subarray1_track_mode}. The default option None is used for subarray shading, meaning no self-shading or backtracking is applied to avoid row shading.”

Now follows some technical comments:

1. Make sure every acronym mentioned in the abstract has its full name explained there (e.g., SCREAM-RRM is a bit surprised to me)

Thank you for pointing this out! For SCREAM–RRM, we have now provided the full names when the acronyms first appear in the abstract, e.g., SCREAM (Simple Cloud-Resolving E3SM Atmosphere Model) and RRM (Regionally Refined Model).

2. To be precise with the used language, solar and wind power are not renewable technologies, wind turbines and solar PV systems are. Solar and wind power are renewable energy sources. Please reformulate line 23 accordingly, and check throughout the manuscript to get rid of similar errors.

Thank you for the correction! We have revised the sentence to use the wording "renewable technologies like wind turbines and solar photovoltaic (PV) systems," and we have checked the manuscript to ensure consistent terminology wherever "technologies" are mentioned.

3. In line 97 the study used "water years", which is a field-specific terminology. It is suggested to explain its meaning at its first occurrence (line 97, while its first explanation now is in line 172) or use a more common term like weather years.

The reference to "water years" originally in L97 has been removed. We have also revised the question-framing portion of the Introduction, as we found that the original questions were overly broad and introduced excessive new information and technical details. We now focus on stating the purpose of each comparison, while deferring detailed descriptions to the Methods section.

4. The headline index might be wrong in section 2.1 and 2.2. Now it appears nothing presented under the headline 2.1, which should not be the case.

Section 2.2 was originally intended to be a first-level subsubsection under Section 2.1 (i.e., 2.1.1). This indexing issue has now been corrected.

5. Please check the entire Methods section, many acronyms mentioned without introducing its full name at the 1st occurrence, such as SVD, MPAS, HICCUP

Except for HICCUP, which does not have a full name, we have added the full names for SVD and MPAS at their first occurrence: singular value decomposition and the Model for Prediction Across Scales.

6. Line 216, repetitive to line 114-115, please reformulate and avoid using the exact same sentence

We have revised the final sentence in the opening paragraph of the Methods section to read:

“Modeled CFs are computed by coupling meteorological outputs from SCREAM-CARRMs, E3SM-25kmNARRM, and HRRR with PySAM, and evaluated against monthly plant-level CFs from EIA as well as across the spatial distribution of California grid points.”

7. Line 274, repetitive word used “preprocessing”

The second “preprocessing” has been deleted.

8. Please add references to underpin your perspectives in line 315-316

We have added references to (e.g., Millstein et al., 2022, 2023) at the end of this sentence.

9. Figure 3: suggest to rearrange the legend into multiple columns (like 3) to save space, and to add index for subplots like a), b), c), which applies to other figures as well in this manuscript

Thank you for the suggestion! We have rearranged the legend of Fig. 3 into three columns and added subplot indices a), b), c), etc. for Figs. 3, 4-5, 7-8, and 11-12, 14-15 in the new version.

10. Figure 4: suggest to keep legend only one time for one row since they are the same, the same applies to Figure 5 and other figures from solar

We have simplified the legends in Figs. 4-5 and 11-12 by keeping the legend only in the first column.

11. Line 423: I don't see an underestimation in SoCal from HRRR, at least not from the bold line. Also it seems to me that NorCal has the least discrepancy present instead of SoCal.

You are correct, SoCal and NorCal were mistakenly reversed here.. This has been corrected.

12. Figure 6: to save space, suggest to use a shared colorbar, put repetitive names as row name and column name, and show latitudes and longitudes only at the first column and the last row, the same applies to other figures from solar, otherwise some numbers are hard to read at the moment in these figures

Following your suggestion, we revised Figs. 7-8 and 14-15 (the figure order has been slightly adjusted, with the diurnal cycle figures moved earlier) by using one colorbar per column and by removing repeated center strings (variable + season names), which are now included in the colorbar titles. Because the NCL panel layout reduces the size of panels containing axis labels, displaying latitude and longitude only on selected rows or columns would produce inconsistent subplot sizes. We thus retained latitude and longitude labels in all California maps to keep panel sizes consistent.

13. Figure 7: texts are hard to recognize, save space by reducing repetitive information as suggested before, try to use a landscape layout or split them into two panels to improve the visualization

As noted in the previous comment, we revised this figure by simplifying the colorbars and removing the center strings.

14. Figure 14: why not showing the "difference" subplots you have in Figure 7?

For figures intended to display and preserve the original high-resolution information, we did not include difference plots, because difference maps between simulations at different resolutions tend to be very noisy, especially at higher resolution. Therefore, we did not include difference maps for the California regional plots in current Figs. 8 and 15. In contrast, for the global circulation plots we included difference maps because the data were first regridded to the ERA5 0.25° grid, which is appropriate for examining large-scale circulation. Nevertheless, we provide here the difference maps (relative to ERA5) for 90-m U/V winds, surface downwelling solar flux (FSDS), and 2-m temperature, with all data remapped to the 3kmCARRM physical grid:

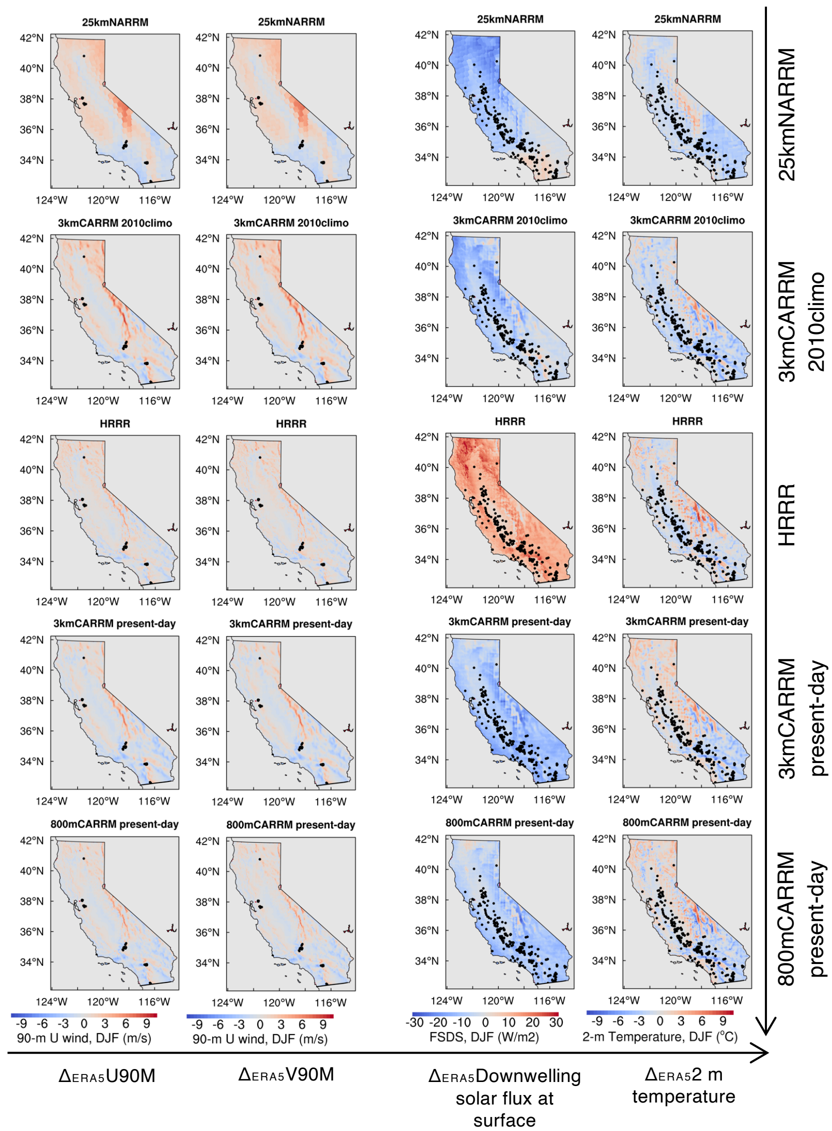

Fig. R1. Direct meteorological drivers of wind and solar CFs in winter: difference relative to ERA5. From left to right: total downwelling solar flux at surface (FSDS), 2 m temperature, 90 m zonal winds (U90M) and 90 m meridional winds (V90M). All datasets were first remapped to the 3kmCARRM physical grid.

15. Figure 9 and 15: try to adjust the legend so that it would not overlap the plot

The legends have been adjusted to avoid overlapping with the plots. In addition, these figures have been updated to show mean absolute bias relative to EIA in the plant-aggregated climatology, consistent with the seasonal cycle figures.

Response to RC2

This paper presents a detailed evaluation of wind and solar energy generation estimates derived from E3SM SCREAM regionally refined simulations over California, with comparisons against HRRR, E3SM-25km NARRM, and EIA-reported capacity factors. The topic is timely and relevant, particularly given the growing importance of high-resolution meteorological modeling for renewable energy assessment. The modeling framework and the scope of the analysis are potentially valuable. However, in its current form, the manuscript requires substantial revision before it can be considered for publication. The primary issues concern clarity, precision, and organization of the writing, as well as conceptual ambiguities in the framing of the research questions and interpretation of results. Detailed comments are provided below.

Thank you for your careful reading of the manuscript and for your supportive comments! In response to your questions and suggestions, we provide detailed replies to each point below. Following your suggestions regarding wording, structure, and presentation, we have revised the manuscript accordingly, particularly in the Introduction and Results sections. In addition, because the EIA data for 2024 have recently become available, we updated all figures related to EIA and observations (reanalysis), except for Figs. 1, 7, and 14. The additional year of observations increases the sample size for EIA plant-level analyses, although the climatological comparison results remain almost unchanged.

Major comments

1. Several abbreviations are introduced without clear or consistent definition, particularly in the abstract.

(1) The abbreviation RRM is used inconsistently. The manuscript refers to “regional mesh refinement” in the context of CARRM and “regionally refined model” in the context of NARRM. A consistent naming convention should be adopted and clearly introduced at first use.

(2) SCREAM is introduced in a convoluted manner in the abstract (“derived from the US Department of Energy’s Simple Cloud-Resolving Energy Exascale Earth System Model (E3SM) Atmosphere Model (SCREAM)”), which makes it unnecessarily difficult to identify what SCREAM stands for. The full name should be stated clearly and directly at first mention.

(3) ERA5 and MERRA-2 are repeatedly referenced without explicitly stating that they are reanalysis datasets, nor is it specified which meteorological variables from these datasets are used. In multiple places, the manuscript refers to the reanalysis data as “meteorological models” or “meteorological model inputs,” which is inaccurate and potentially misleading.

Thank you for pointing out these semantic ambiguities!

1) In the literature, RRM can refer both to the Regionally Refined *Mesh* and to a Regionally Refined *Model*. Strictly speaking, RRM is a technical approach for modifying grid-related configurations within a given host model, and it does not exist independently of an underlying model. However, in the literature it is often simplified to refer to a specific model version that applies a regionally refined mesh. We have chosen to consistently retain the use of Regionally Refined Model throughout the manuscript.

2) By directly referencing E3SM, we have revised the description of SCREAM in the abstract to its more commonly used full name “Simple Cloud-Resolving E3SM Atmosphere Model (SCREAM)”.

3) When ERA5 is first introduced, its full name naturally indicates that it is a reanalysis product, whereas the full name of MERRA2 does not explicitly include this information. We have therefore added explicit statements at first mention clarifying that both are reanalysis datasets “...compared to the coarse-resolution reanalyses such as ERA5 and MERRA2”. We do not calculate CFs using ERA5 or MERRA2 in this study but cite their use in the existing literature. In the subsubsection “PLUSWIND-derived CFs” we have added which variables are used in PLUSWIND: “They use hub-height wind speeds from HRRR, model-level wind speeds from ERA5, and near-surface wind speeds from MERRA2 to estimate wind power generation”.

We apologize for the misleading word choice. The phrase meteorological model inputs appeared once in the original text: “This suggests that the effect of modeling assumptions between PLUSWIND and PySAM is larger than the effect of internal corrections within PLUSWIND, but smaller than the effect of varying meteorological model inputs”. Here, “inputs” refers to meteorological datasets (i.e., outputs from atmospheric models) used as inputs to the generation model (e.g., PySAM). To avoid ambiguity, we have revised this to “varying meteorological model datasets”. In this paper, unless otherwise specified, each meteorological model corresponds one-to-one with its associated dataset; e.g., HRRR, ERA5, and MERRA2 are distinct models, each paired with the meteorological dataset generated by that model.

2. The research questions listed in the introduction are not well separated conceptually and often mix scientific questions with methodological choices.

(1) First goal (generation modeling assumptions): This appears primarily to be a methodological sensitivity analysis, yet it is framed as a main scientific objective. It is not clear, until later sections, why this comparison is essential for the broader goals of the paper. If this is a key component, the motivation and what the “generation modeling assumptions” represent should be stated more clearly in the introduction.

(2) Second goal (meteorological models): The phrase “what do meteorological models represent” is vague. It is unclear whether this refers to differences in simulated meteorological fields or differences in model physics. The authors should be more precise in their wording. In addition, the comparison between SCREAM-3kmCARRM forced by a 2005–2014 climatology and HRRR simulations from 2018–2022 is not a fair comparison. While the manuscript argues that monthly climatology justifies this approach, interannual variability during 2018–2022 may still strongly affect the results, and this issue is not adequately addressed. There is also an inconsistency in how the “2010climo” period is defined: it is described as 2005–2014 in the text, whereas Table 1 lists 2010–2019. This discrepancy should be clarified and made consistent.

Thank you for raising this point! These two questions are indeed related. In the original structure, generation modeling assumptions were listed as the first item, and meteorological models as the second, but both are components of the first core question we aim to address: Is km-scale E3SM (SCREAM) combined with PySAM an effective framework? The second core question is: Can km-scale E3SM simulations substantially reduce the pronounced wind power biases identified in 25 km E3SM simulations over California, and how important is horizontal resolution for wind and solar energy simulations?

For the first question, assessing “effectiveness” requires multiple reference benchmarks. The benchmark closest to applications is the EIA monthly data; however, discrepancies between simulated and EIA CFs arise from multiple sources of uncertainty: (1) biases in SCREAM-CARRMs simulation data, (2) PySAM energy modeling, and (3) EIA data collection and quality control. We therefore include a second benchmark, HRRR, another km-scale meteorological dataset validated in the wind energy literature. By combining HRRR with the same PySAM settings applied to SCREAM-CARRMs, uncertainties related to horizontal resolution and energy modeling are removed, isolating differences arising solely from the meteorological model (SCREAM-CARRMs vs. HRRR). A third benchmark is the PLUSWIND-derived HRRR wind CFs, which, when compared with HRRR + PySAM, removes meteorological uncertainty (both use HRRR) and isolates differences only due to the generation model (PySAM vs. PLUSWIND).

To the specific questions:

1) The motivation for introducing generation modeling assumptions in the first goal is as follows. We begin by comparing HRRR + PLUSWIND vs. HRRR + PySAM to first assess how large the variability arising from the use of different generation models can be. This is to validate our use of PySAM by comparing against a well-established dataset. While this comparison resembles a sensitivity analysis (as it involves two different wind power models), the focus of our study is the energy evaluation of SCREAM-CARRMs and the effects of grid resolution. Thus, except for axis-tracking, we adopt the default PySAM settings rather than doing a detailed sensitivity analysis.

2) In the second goal, “meteorological models” refers specifically to different meteorological models at comparable horizontal resolution, namely SCREAM-3kmCARRM vs. HRRR. In a climatological context, we treat meteorological models (including physics) and the resulting simulation datasets as closely linked, as differences among simulated fields are primarily driven by model differences. We fully agree with the concern regarding mismatched simulation periods; this is an unavoidable challenge when comparing climatological simulations with reanalysis data. We also recognize that the original question framing was too broad. As discussed above, HRRR serves as one of several benchmarks for evaluating SCREAM-3kmCARRM, analogous to the role of PLUSWIND in the first question. Both questions arise because EIA-based evaluation alone contains multiple sources of uncertainty, motivating additional comparisons to help gauge the relative importance of different uncertainty sources (horizontal resolution, meteorological model, and generation model).

The 2010climo simulation is forced by present-day observational climatology of the 2010s, with SST and sea ice prescribed from the climatological annual cycle averaged over 2005–2014. The nominal simulation period spans 2010–2019. However, for climate-type simulations, nominal years can be misleading, particularly when other datasets correspond to actual calendar years. Therefore, we have revised Table 1 to report simulation durations rather than nominal simulation years for all E3SM simulations.

=======

Following your suggestion, we have reorganized and clarified the questions listed in the introduction:

“By estimating CFs using PySAM and comparing all modeled results with EIA monthly data, we construct multiple angles of comparison. Our high-level study objectives are to access: 1) whether km-scale SCREAM-CARRM climate data combined with PySAM constitutes an effective energy modeling framework, and 2) whether it can substantially reduce the pronounced wind energy biases identified in 25 km E3SM simulations over California (i.e., how important is horizontal resolution is for wind and solar energy simulations?).

EIA provides the benchmark most directly relevant to applications; however, discrepancies relative to EIA reflect combined uncertainties associated with SCREAM-CARRM datasets, PySAM energy modeling, and EIA data quality. We therefore introduce additional evaluation targets to approximately assess the relative importance of these uncertainties. Specifically, we first assess whether PySAM is applied appropriately, then compare against HRRR at comparable resolution to evaluate whether SCREAM-3kmCARRM has similar performance, and finally examine the role of horizontal resolution. Accordingly, we address the following three questions:

\begin{enumerate}

\item \emph{Does the use of different energy models affect wind CFs? How do different axis-tracking methods affect solar CFs?}

This question focuses on validating our use of PySAM in assessing whether the SCREAM-CARRM + PySAM framework is effective. For wind, we compare HRRR-derived CFs from PLUSWIND with those estimated using PySAM, with both evaluated against EIA monthly CFs. We also reproduce the PLUSWIND comparisons among ERA5, MERRA2, and HRRR, which helps assess the relative importance of horizontal resolution and generation modeling. For solar, we evaluate the impacts of fixed, single-axis, and dual-axis tracking assumptions in PySAM.

\item \emph{How different are wind and solar CFs between SCREAM-3kmCARRM and HRRR at comparable horizontal resolution?}

Using the same PySAM configuration, we assess the performance of SCREAM-CARRM by comparison with HRRR. Although these two meteorological models differ substantially in model structure, physical parameterizations, and simulation periods (HRRR forecasts over 2018–2022 vs. SCREAM-3kmCARRM forced by observed climatology over 2005–2014), they share a very similar horizontal resolution at the kilometer scale. We therefore ask whether, despite these differences, comparable horizontal resolution can dominate and lead to statistically similar wind and solar CFs.

\item \emph{How much does horizontal resolution affect wind and solar energy calculations?}

We first compare CFs computed with PySAM from the E3SM-25kmNARRM and 3.25 km SCREAM-CARRM simulations to assess whether km-scale resolution substantially reduces the wind generation biases identified at 25 km. Sensitivity to further resolution refinement is then assessed by comparing 3.25 km and 800 m SCREAM-CARRM simulations, which share identical configurations (driven by CMIP6 forcings and evaluated for the present day) and differ only in horizontal resolution.

\end{enumerate}"

3. The manuscript states that SCREAM-RRM requires nudging because it lacks a scale-aware deep convection scheme. However, nudging does not directly address the absence of scale-aware convection. The authors should clarify the actual purpose of nudging in these simulations, for example whether it is primarily used to constrain large-scale circulation. The choice of a 2-day relaxation timescale also appears relatively weak and should be justified. In addition, the explanation of how "reduced nudging strength near the surface and model top" allows "free-running conditions over California" while maintaining large-scale constraints elsewhere is not sufficiently clear.

Yes, nudging is used primarily to constrain the large-scale circulation. It does not directly compensate for the absence of a scale-aware deep convection scheme; rather, it helps maintain realistic large-scale thermodynamic and dynamical conditions, which in turn limits the unrealistic buildup of instability in the absence of a convective parameterization and ensures consistent large-scale forcing into the refined domain.

Regarding the nudging details, we apologize for the lack of clarity in the original description. The statement "this allows free-running conditions over California while nudging the coarser-resolution domain" corresponds to "using the E3SM regional configuration with a tanh-based smooth windowing function to constrain the horizontal domain." Because SCREAM is a global model, a spatial weighting function is used to vary nudging strength across the domain; in E3SM, this is implemented via a tanh-based smooth windowing function. Vertically, nudging strength is also defined via tanh-based smooth window, with users specifying topmost and bottommost model levels and transition gradients (deltas) in the user namelist. We adopted the default configuration, in which nudging spans all vertical levels with delta values of 0.1 at both the top and bottom, resulting in weaker nudging near the surface and at the top and full-strength nudging elsewhere. The two-day relaxation timescale follows previous SCREAM-3kmCARRM climate simulations (Zhang et al., 2024). In earlier experiments we tested shorter timescales and found that a 2-day relaxation produced the most consistent precipitation patterns between RRM and the nudging data.

To clarify, we have revised the nudging description accordingly:

> “Nudging data are interpolated onto the RRM physics grid using NCO’s linear-in-log-pressure vertical interpolation and TempestRemap’s high-order horizontal interpolation. U, V, T, and Q nudging with a 2-day relaxation timescale is applied following \citep{Zhang2024}. Regional nudging in E3SM employs a tanh-based smooth windowing function that is zero within the refined region and transitions toward one in the surrounding coarse-resolution domain \citep[e.g.,][]{Tang2019}, allowing free-running conditions over California and applying nudging in the coarser-resolution domain. Vertically, the default tanh-based smooth window is applied, which reduces the effective nudging strength near the surface and at the top and full strength in between.”

4. Section 3.1 emphasizes substantial uncertainties in the EIA monthly data and explicitly states that model–EIA differences should be interpreted as “discrepancies” rather than biases. Given this, the scientific value of using EIA data as the primary reference for evaluating model performance is unclear. The authors should clarify how conclusions about model skill can be drawn when the reference dataset itself is highly uncertain, and whether the PLUSWIND dataset might provide a more reliable benchmark.

Thank you for raising this issue! As discussed in our response to earlier comments, this is precisely why we supplement EIA-based evaluation with additional datasets. These include comparisons between HRRR + PLUSWIND vs. HRRR + PySAM to confirm our correct use of PySAM, as well as comparisons between HRRR + PySAM and SCREAM-3kmCARRM + PySAM.

Despite its limitations, EIA remains the best publicly available dataset of observed generation and one of the most widely used public energy datasets in the U.S. (\url{https://catalystcoop-pudl.readthedocs.io/en/latest/index.html}, last accessed: February 22, 2026). HRRR CFs from PLUSWIND closely resemble our HRRR + PySAM results and have therefore been used as a benchmark for wind CFs, although PLUSWIND does not include solar. Both HRRR + PLUSWIND and HRRR + PySAM are subject to uncertainties associated with HRRR model biases, particularly for solar CFs. Indeed, HRRR solar simulations deviate more from reference datasets (EIA for CFs and ERA5 for surface downwelling shortwave radiation) than SCREAM-3kmCARRM, suggesting caution in treating HRRR solar results as closer to reality. Unfortunately, more suitable observational datasets are either unavailable or not publicly accessible. Our conclusions therefore emphasize SCREAM's potential applicability in regions with similar climates, based on the available comparisons.

We have expanded the discussion of potential alternative datasets in the section "Uncertainty in EIA monthly data and interpretation of discrepancies" and added the following interpretation:

> "Given these limitations, the use of monthly EIA data in this study reflects a trade-off between observational fidelity and data availability. We therefore caution against attributing discrepancies between modeled generation and EIA monthly CFs solely to deficiencies in either the atmospheric simulations or the power generation modeling. Instead, these discrepancies should be interpreted in light of the combined uncertainties inherent in both the observational data and the meteorological–generation modeling framework. This is why we also include comparisons with PLUSWIND and HRRR to help assess the relative importance of different uncertainty sources."

5. Section 3.2 ("Evaluation group design and key questions") is vague and reads more like informal notes than formal academic writing. This section should be rewritten to clearly and concisely describe the evaluation framework.

Thank you for this suggestion! We have rewritten this section as follows:

> "\subsection{Evaluation design and structure}
>
> The Results section is organized to address the three guiding questions introduced in the Introduction through a set of controlled comparisons. All CFs are calculated using a consistent PySAM configuration unless otherwise noted. Because discrepancies relative to EIA reflect combined uncertainties from meteorological datasets, power-generation modeling, and EIA data quality, additional benchmarks are introduced to help contextualize the relative importance of these uncertainty sources.
>
> For both wind and solar energy, the analysis follows a common structure in general:

\begin{enumerate}

\item \emph{Generation model dependence (Q1).}

For wind, we compare CFs derived from HRRR using PLUSWIND and PySAM to assess whether generation model choices materially affect CF estimates. For solar PV, we quantify the sensitivity of modeled CFs to axis-tracking assumptions within PySAM. These tests serve as a validation of PySAM before inter-comparing meteorological datasets.

\item \emph{Meteorological dataset dependence at comparable resolution (Q2).}

We compare SCREAM-3kmCARRM 2010climo against HRRR, both of which provide km-scale meteorological fields to PySAM. This comparison quantifies the magnitude of CF differences attributable to the meteorological model when horizontal resolution is not the primary differentiator.

\item \emph{Horizontal resolution sensitivity (Q3).}

We compare present-day SCREAM-CARRM simulations at 3.25 km with the E3SM-25kmNARRM simulation to assess whether km-scale configurations substantially reduce the pronounced wind CF phase errors identified at 25 km resolution over California. Sensitivity to further refinement of resolution is then examined by comparing the 3.25 km and 800 m CARRM simulations.

\end{enumerate}

Within each comparison group, results are presented separately for wind and solar. For each energy type, we first show multi-year monthly CF climatology aggregated by plant clusters, followed by diagnostics that help interpret CF differences, including diurnal cycles of simulated CFs, statewide spatial patterns of CFs and key meteorological driver variables used by PySAM. Each energy-type section concludes with a summary plot to synthesize the main findings across meteorological datasets."

6. The statement that "each following subsection presents results by energy type (wind or solar)" is misleading, as the first evaluation (generation modeling assumptions) applies only to wind. A consistent subsection addressing generation modeling assumptions for solar is missing.

Following your earlier comments, we have clarified this issue by explicitly distinguishing between wind and solar. For wind, the question is whether the use of different energy models affect wind CFs; for solar, the question is how do different axis-tracking methods affect biases in modeled solar CFs. We have placed the axis-tracking results in a dedicated subsubsection.

The content of this subsection has been updated accordingly:

> "\subsubsection{How do different axis-tracking methods affect solar CFs?}
>
> Figure~\ref{SolarMon3kmCARRMF2010vsHRRR} shows the solar CFs in California as estimated by PySAM, using SCREAM-3kmCARRM 2010climo and HRRR simulations. Compared to wind, the seasonal cycle of solar CFs is much simpler, with a clear summer peak and winter minimum. Even for dual-axis systems, despite having only three samples, the seasonal pattern remains smooth. This likely reflects the relatively stable solar resource in California's sunny climate; in cloudier regions with higher solar variability, such smoothness may not hold.
>
> Across the three tracker classes, the overestimation of simulated solar CFs relative to EIA records increases with tracker complexity. For fixed-axis arrays, both SCREAM-3kmCARRM 2010climo and HRRR reproduce the observations well: on average, 3kmCARRM deviates by less than 3 pp in every month. For single-axis systems, the overestimation in both models remains below 6 pp, while for dual-axis systems it stays below 13 pp.
>
> This monotonic increase suggests that the tracker algorithms implemented in PySAM are somewhat idealized. Solar tracking systems are designed to increase energy production by maintaining the panel orientation close to perpendicular to the incoming solar radiation. In principle, dual-axis trackers should therefore achieve higher efficiency than single-axis systems. However, this pattern is not evident in the EIA generation records. One possible reason is that dual-axis systems are mechanically more complex, and operational constraints associated with the control and driving mechanisms may offset part of the theoretical energy gain. In addition, the benefit of multi-axis tracking may be limited in hot climates, where the increased incident irradiance can raise module temperatures and reduce PV output due to thermal losses \citep{Hammoumi2022}.
>
> In the SAM PV model, tracker orientation follows geometric tracking algorithms that determine panel tilt and azimuth from solar position and array geometry:

> for single-axis systems, the panel azimuth tracks the solar azimuth while the tilt remains fixed relative to the rotation axis, whereas for dual-axis systems both tilt and azimuth follow the solar zenith and azimuth angles, respectively \citep{SAMPV2018}. In practice, mechanical constraints, tracking control strategies, and stow conditions can lead to slower or less precise tracking, reducing the realized energy gains relative to these idealized simulations."

In addition, we revised the grouping logic for tracking types in the Methods section as follows:

> "For PV plants, we group samples by both model grid point and axis-tracking type. For example, if a grid point G contains three single-axis plants and one fixed-axis plant, two groups are formed: G_1 (single-axis) and G_0 (fixed). The seasonal cycles and evaluation metrics are then computed by clustering according to axis-tracking type. This approach preserves the plant-level tracking type information reported in the EIA dataset, and the total nameplate capacity within each cluster does not vary with model resolution. As will be shown in the results section, the axis-tracking type has a substantial influence on PV power generation."

7. Sections 3.3.2–3.3.6 are unevenly structured. Some subsections emphasize resolution effects even though they are nominally focused on meteorological model differences, while others contain only a single paragraph. The organization of these subsections should be improved to better align with the stated evaluation goals.

Section 3.3.2 compares HRRR + PySAM with SCREAM-3kmCARRM 2010climo + PySAM. Although the two simulations have similar horizontal resolutions, the models themselves differ in many aspects. As noted above, the inference regarding which source of uncertainty is more influential emerges only when this comparison is considered together with the subsequent analysis of horizontal resolution sensitivity and the contrasting results between wind and solar CFs. To clarify this point, we moved the discussion related to horizontal resolution from this section to the following subsection.

In addition, we added a subsection discussing the diurnal cycle of solar generation and moved the corresponding figure forward to immediately follow the EIA monthly evaluation, because the diurnal-cycle analysis is also based on EIA plant sampling (although no hourly EIA data are available here).

As explained in our responses to the previous two comments, we have revised the descriptions introducing each subsection accordingly. Because these revisions involve extensive changes throughout the results section, please refer to the tracked-changes version for the detailed modifications.

8. In Figure 4 and subsequent figures, the EIA curves appear different between the top and bottom panels. The reason for this difference is unclear and should be explained.

Thanks for asking this! This choice was made because our primary goal is model evaluation at the model-resolution (i.e., grid-cell) scale, rather than plant-level applications. This approach has several advantages:

1) It avoids artificially reducing the spread of the modeled climatological seasonal cycle. In plant-level evaluations, multiple plants located within the same model grid cell would be represented identically. This would substantially reduce the apparent monthly climatological spread and make modeled variability appear artificially smaller than that from observations.

2) Aggregating to the model grid ensures consistency between local evaluations and assessments of spatial patterns across California, as both are conducted on the same grid.

In addition, we identified a minor issue in the previous version of the code used for the seasonal cycle plots. Small differences in the EIA curves after aggregation to different model grids arose from an incorrect application of nameplate capacity weights during gridcell averaging. This has been corrected in the revised manuscript, and the resulting changes are minor and do not affect the conclusions.

We have added this explanation to the subsection “EIA monthly CF data”:

> “We adopted a gridcell-based evaluation framework instead of plant-level evaluation for the following reasons: 1) It avoids artificially reducing the spread of the modeled climatological seasonal cycle. In plant-level evaluations, multiple plants located within the same model grid cell would be represented identically. This would substantially reduce the apparent monthly climatological spread and make modeled variability appear artificially smaller than that from observations. 2) Aggregating to the model grid ensures consistency between local evaluations and assessments of spatial patterns across California, as both are conducted on the same grid. In the seasonal-cycle analysis, we applied a weighted average

based on total nameplate capacity within each grid cell, so that the EIA mean line remains consistent across comparisons at different model resolutions."

9. In some regions (e.g., Kern County and Southern California), the seasonal cycles between datasets do not appear as similar as claimed in the text, for example the statement that "both models captured the July peak well." These interpretations should be revisited and aligned more closely with what is shown in the figures.

We apologize that the discussion of SoCal and NorCal was inadvertently reversed in several places! The statement in question actually refers to NorCal. The relevant text has been corrected.

10. Lines 475–485 discuss deficiencies in the simulation of large-scale circulation, but then abruptly conclude that large-scale circulation plays a relatively minor role. The logical connection between these statements is unclear and requires clarification.

The initial discussion examines large-scale circulation as a commonly used diagnostic to understand the high winter wind speeds in the 25 km simulation. However, the ridge pattern in SCREAM-3kmCARRM 2010climo closely resembles that in 25kmNARRM. If biases in large-scale circulation were the dominant cause of the winter wind bias, the 3 km simulation would be expected to exhibit similarly strong biases, which it does not. This indicates that large-scale circulation is unlikely to be the primary driver of the high winter wind bias in 25kmNARRM. This may not be surprising for California, where complex topography implies a stronger influence of local, small-scale processes on wind patterns. Nevertheless, examining circulation differences remains a useful diagnostic to support this inference.

11. Section 3.4.1 aims to assess the impact of meteorological models on solar energy, but the results primarily describe model behavior without clearly answering this question.

Thank you for this reminder! We have refined the question to "How different are solar CFs between SCREAM-3kmCARRM and HRRR at comparable horizontal resolution?" Given the many differences between the two models, our evaluation quantifies the magnitude of differences in simulated solar CFs but cannot isolate specific causal mechanisms. Nevertheless, we examined key meteorological drivers of solar CFs (surface downwelling shortwave radiation) in ERA5 over the same period as HRRR (2018–2022) and found spatial patterns very similar to the 2013–2024 mean (Fig. R2). This suggests that HRRR solar CFs averaged over 2018–2022 would likely be comparable to those averaged over the EIA record period, implying that the observed differences primarily reflect meteorological model differences. Clouds, which dominate surface shortwave radiation, remain one of the largest sources of uncertainty in atmospheric modeling as they depend on multiple interacting physical processes.

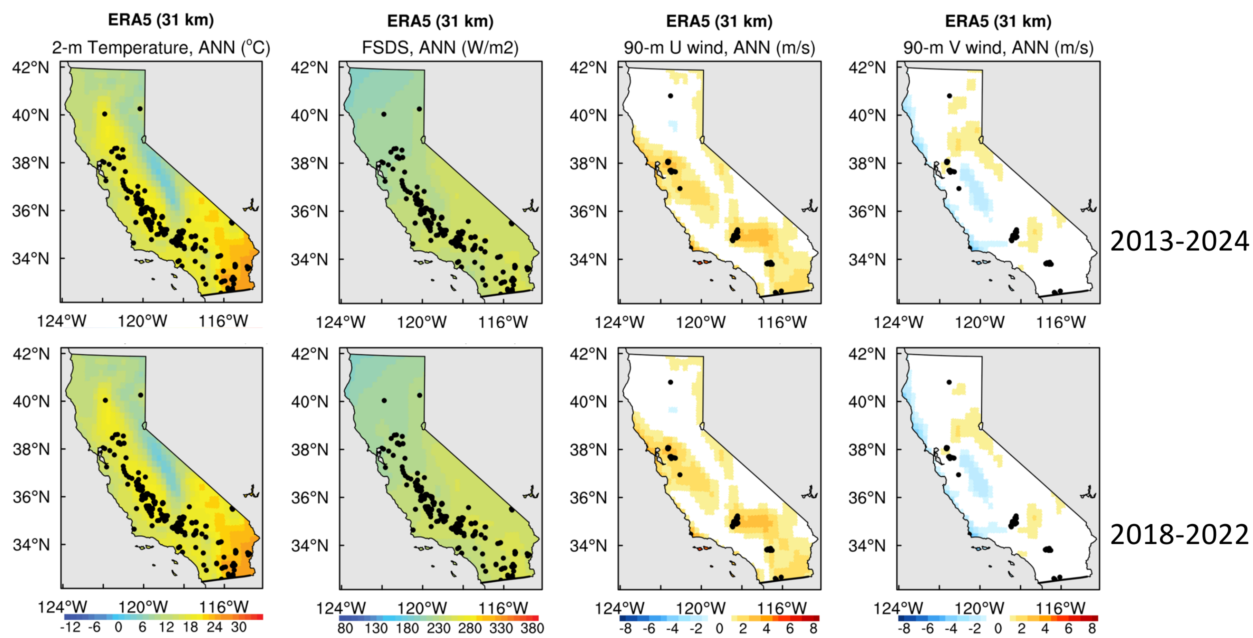

Fig. R2. Direct meteorological drivers of annual-mean solar and wind CFs. From left to right: total downwelling solar flux at surface (FSDS), 2 m temperature, 90 m zonal winds (U90M) and 90 m meridional winds (V90M).

We also noted that this section previously conflated axis-tracking sensitivity with meteorological model differences. Axis-tracking has now been moved to a separate subsection, and the above discussion has been added here.

12. Section 3.4.3 is titled “large-scale circulation,” yet large-scale circulation is not discussed in the current version of this section. In addition, the conclusions discuss generation modeling assumptions for both wind and solar energy, whereas the introduction and the corresponding research question frame this issue almost exclusively in terms of wind (“How much do generation modeling assumptions impact the wind energy?”). This inconsistency should be addressed.

Thank you for pointing out these issues! We have deleted large-scale circulation in the section title. And the introduction now clearly distinguishes generation modeling questions for wind and solar.

Minor comments

1. Introduce full names before abbreviations for SHOC, P3, RTE+RRTMGP, SPA, and MPAS.

Except for MPAS, we previously placed the full names of these terms in parentheses (followed by the citation, separated by ";"), because the abbreviations are often used more widely than their full names in practice. In this revision, we have reversed the order and added the full name of MPAS.

2. Section title 2.2.2: "Souce" or "Source"?

Corrected.

3. Line 176: replace "simulations" with "this type of experiment simulates".

Corrected.

4. Line 185: provide the full name for MPAS.

Added: the Model for Prediction Across Scales (MPAS).

5. Line 221: "eia-processor" → "EIA-processor".

Modified.

6. Lines 225–230: rewrite in a more formal, non-conversational style.

Thanks for this suggestion! We revised this paragraph to use a more formal and descriptive style to describe the filtering criteria:

> “The EIA data first underwent the following plant-level quality control steps:
>
> \begin{enumerate}
>
> \item Only plants with a nameplate capacity of at least 2.5 MW were included.
>
> \item Only plants located in the state of California were included.
>
> \item For solar generation, only plants classified as ``Solar Photovoltaic'' were retained (accounting for 97.3\% of all solar facilities), while those categorized as ``Solar Thermal without Energy Storage'' were excluded. In this study, the term \emph{solar} refers specifically to \emph{PV (photovoltaic)} power.
>
> \item For solar PV plants, only those with a reported array configuration in the EIA dataset were retained, as indicated by at least one of the fields ``Single-Axis Tracking?'', ``Dual-Axis Tracking?'', or ``Fixed Tilt?''.
>
> \item Plants were excluded if CF was below 5\% for more than nine months or if more than 12 months of data were missing.
>
> \end{enumerate}”

7. Line 245: clarify what “independent-plant ratio > 15%” means and how it is computed.

We have added one sentence here:

> “The independent-plant ratio is defined as the fraction of plants in a given year whose monthly CF time series are not perfectly correlated with those of other plants (identified using a Pearson correlation distance threshold of 10^{-6}). “

8. Line 262: clarify what “models” refers to.

Added:

> "Air density correction for all models (ERA5, MERRA2, HRRR) was based on MERRA2 air density, interpolated to hourly resolution at hub height."

9. Line 274: "preprocessing" is duplicated.

The second "preprocessing" has been deleted.

10. Line 428: "suggests" is repeated.

The second "suggests" has been deleted.

11. Figure 7 caption needs revision, as large-scale circulation is one of the drivers influencing wind CF, rather than a separate category from the direct drivers.

We have deleted "large-scale circulation" from the summary sentence in Figure 7 caption.