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:

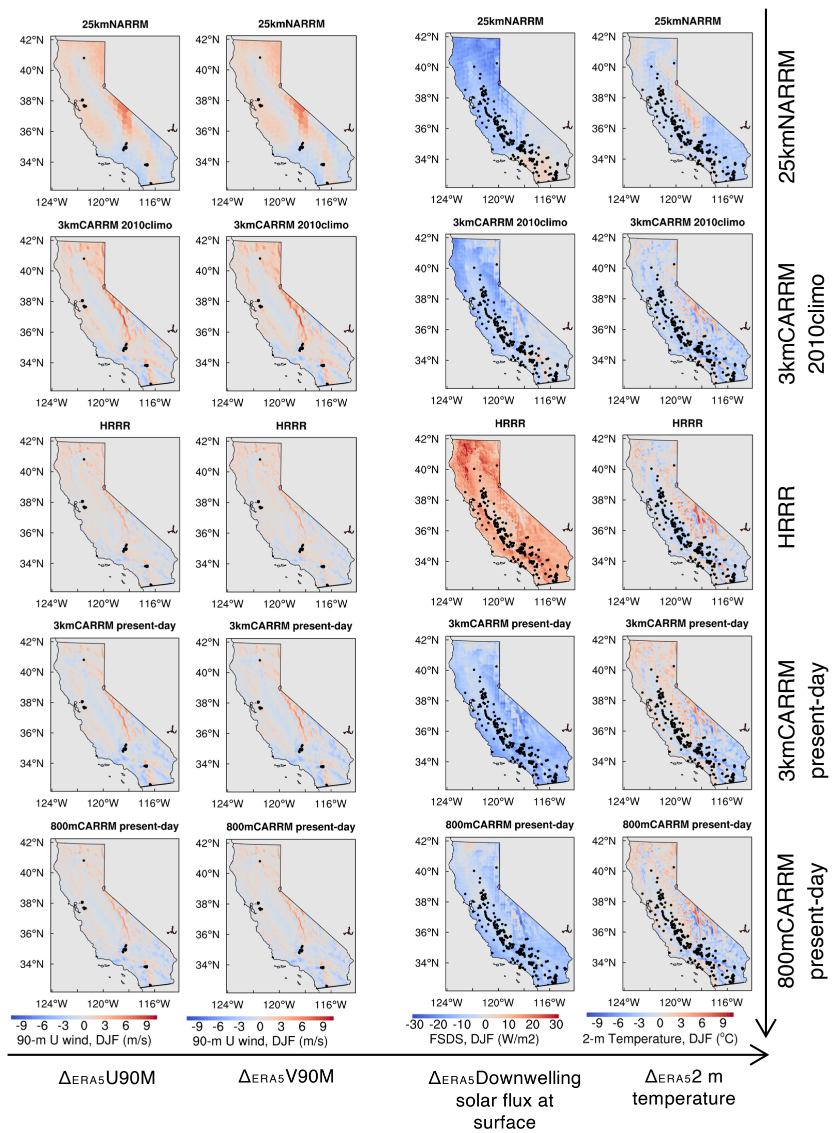

Fig. R1. Direct meteorological drivers of wind and solar CFs in winter: difference relative to ERA5. From left to right: total downwelling solar flux at surface (FSDS), 2 m temperature, 90 m zonal winds (U90M) and 90 m meridional winds (V90M). All datasets were first remapped to the 3kmCARRM physical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