The study presents the SCREAM/E3SM model run at global-scale to 100 m horizontal resolution over the San Francisco Bay Area in the US for two selected weather forcing periods.

Overall, the manuscript is written in much detail which is appreciated but feels lengthy. I strongly suggest moving the technical sections of the manuscript to either supplemental or appendix. That space could be used to discuss more about the capabilities of the model. The use of varying grid resolution for refining over regions of interest is quite challenging and is important as authors noted. Demonstrating that the model can deliver measurable skill gains relative to the baseline run at 3.25 km while remaining numerically stable is an advancement for the modeling community. To my knowledge, this is the first LES-scale within a global model framework.

Having said that, some scientific and technical issues need further treatment before the manuscript is publishable, especially the generalization of the model performance based on two case studies. I understand the computational cost associated with each case, however, to accept the SCREAM model for routine use in the LES community, it would be better to have either a small ensemble, or another synoptically driven flow (perhaps Diablo winds!) simulated, or a different region to answer few questions such as - How does the model perform for mid or high-latitudes? - What would be the sensitivity to the initial and lateral boundary conditions? I therefore recommend a major revision.

Thank you for your insightful, balanced comments, and for recognizing the significance of this work! We would like to emphasize that the primary contribution of this study is to demonstrate and document the feasibility of running E3SM SCREAM at 100 m resolution – as you noted, this represents the first such attempt within a global model framework. Prior to performing the simulations, we were unsure whether such a setup would even be viable. Thus, during the early design phase of the project (especially while applying for computing resources), it was not feasible to plan for a broader set of simulations after a successful proof-of-concept, since success was far from guaranteed.

Frankly, the simulations and analyses presented in the current manuscript have already exhausted (actually exceeded) the resources approved. Much of the allocated computing time was consumed during the trial-and-error phase, including months of development, test runs, and debugging prior to launching the full simulations. The additional 100 m simulations you suggested are exciting and we deeply appreciate your recognition of their potential, however, they go far beyond what this work is intended to cover.

We would like to take this opportunity to clarify the scope and contribution of this work:

First, this is a proof-of-concept study, focused on demonstrating feasibility rather than conducting a scaled-up, systematic evaluation of model performance for routine use. The latter, in our view, should be carried out by larger collaborative teams with expertise across various scientific processes and evaluation metrics.

Second, beyond demonstrating feasibility, this study identifies key limitations and opportunities for improvement in the current toolchains. For instance, we were unable to use 100 m topography consistent with the model resolution because the existing topography tool runs only in serial interpolation from the 250 m source DEM to model grids triggered out-of-memory issues on all available machines. A practical solution would involve developing an MPI interpolation tool and simplifying the mapping algorithm. Such development needs fall into several categories. Until they are implemented, additional simulations would remain prohibitively expensive. Moreover, proceeding without fixing these known bottlenecks is impractical. Clear improvements exist that could greatly reduce resource demands; otherwise, new simulations would again require 10× the reasonable cost which is currently not feasible. As far as we know, these toolchain developments are already part of broader community plans.

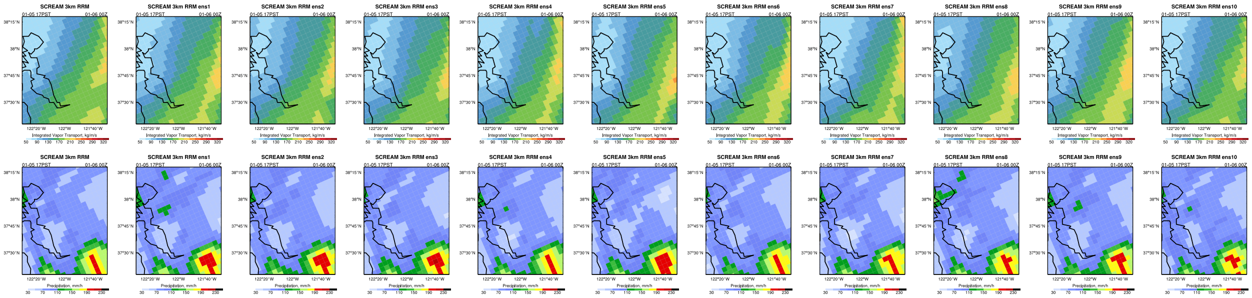
Third, because this work serves as a first demonstration of feasibility and a documentation, the methodology itself is a central contribution. A key mission of this paper is to record every step as clear as possible, so that interested users can replicate the process. Given that many readers may not be interested in highly technical details, we have reorganized the Methods section: detailed content has been moved into deeper-level subsections, and summarizing sentences were added at the start of the Methods.

All in all, we hope this clarifies the role and scope of the study, the practical challenges of designing/conducting new 100 m simulations, and the broader community efforts already in progress. We are equally eager to see more E3SM SCREAM simulations at 100 m resolution and more in-depth analyses in the future, which would undoubtedly expand the model's scientific impact. Once current toolchain limitations are resolved, additional simulations will become much more feasible – especially with the support of GPU resources.

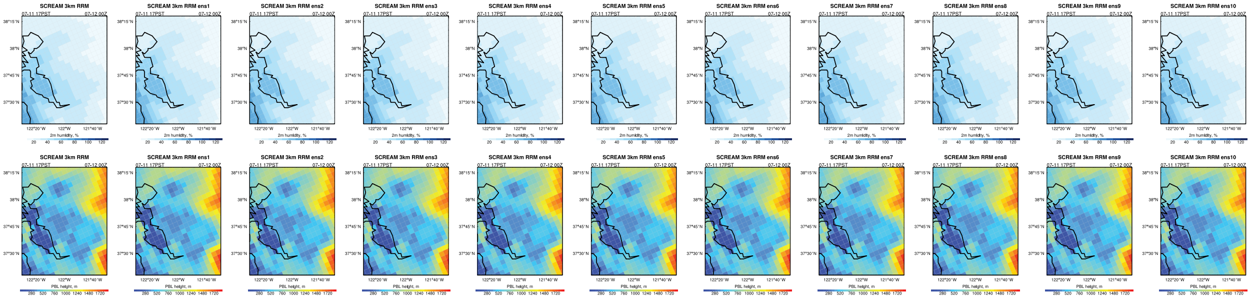
We sincerely thank you again for your careful review and constructive suggestions! In response, we have incorporated three major updates to the analysis: 1) two sets of ensemble simulations for CA-3km for each case, 2) energy spectra analysis, and 3) a comparison of SGS TKE, fluxes, and variances in CA-3km vs. BA-100m. Following your specific comments, we have updated five figures, added four figures and two tables in the main text.

=============

To quantify the sensitivity to small perturbations in the initial conditions, we conducted two small ensembles (10 members) for the CA-3km simulations of both cases. The ensembles were generated by applying random perturbations to the initial temperature profiles at all grid points. Fig. R2.1 and Fig. R2.2 shows the domain-averaged vertically integrated moisture transport and accumulated precipitation at the final timestep for the Storm2008 case, and 2 m relative humidity and online-diagnosed boundary layer height for the Stratocumulus2023 case. Apart from small uncertainty in the location of the precipitation maximum in Storm2008, the moisture transport and overall precipitation patterns are highly robust. Differences in the Stratocumulus2023 case are almost unable to distinguish visually.

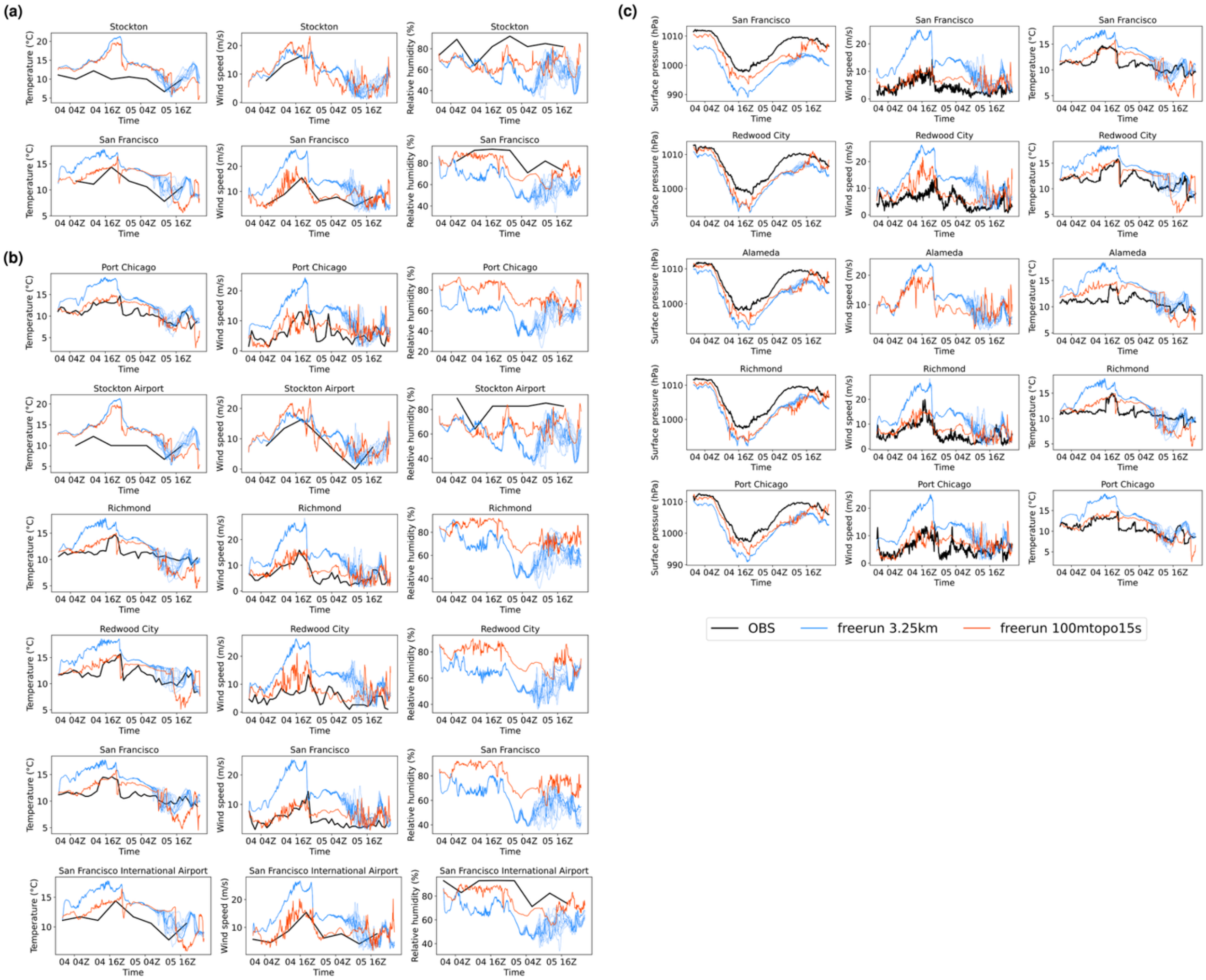


R2.1 vertically integrated vapor transport (top) and accumulated precipitation (bottom) at the final timestep of the Storm2008 event, for the single-realization control simulation (left column) and ensembles 1–10 (2nd to 11th columns).

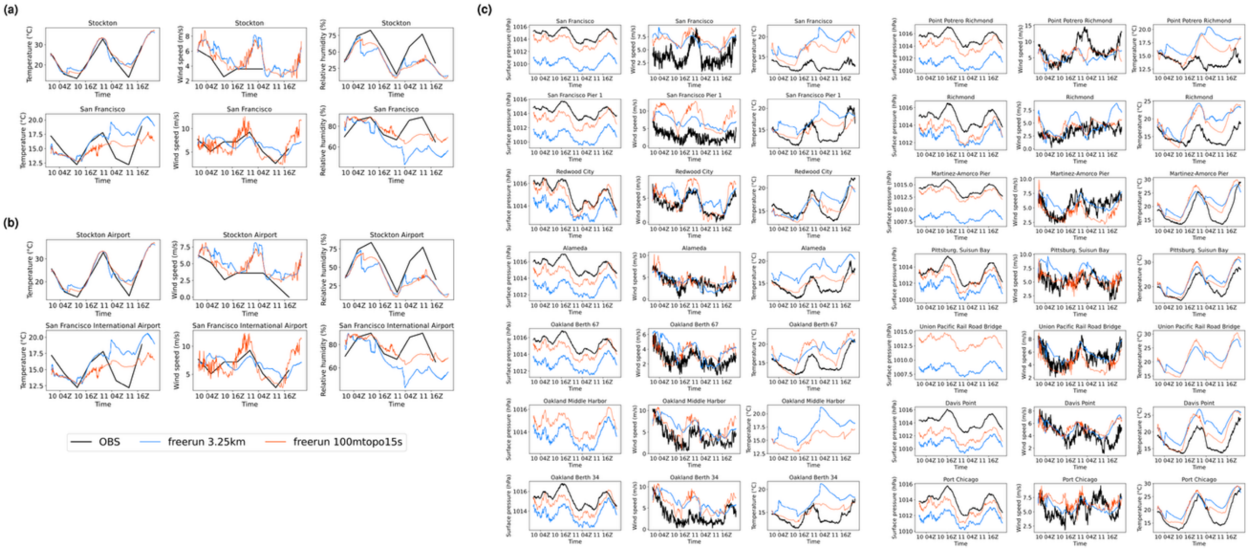


R2.2 Same as R2.1 but for 2 m relative humidity (top) and boundary layer height (bottom) for the Stratocumulus2023 event.

Fig. R2.3 and Fig. R2.4 show timeseries of in situ timeseries evaluation for all ensemble members over the simulation period. Thick lines represent the control runs, and thin lines represent ensemble members. Note that these plots differ slightly from their counterparts in the main text: here, to better show high-frequency variability, we did not resample the raw 5 min instantaneous model outputs to match the obs, nor did we compute variability based on the model outputs within each observational window (as was done in the Meteomanz and ISD plots in the updated manuscript). Instead, we directly plot the 5 min timeseries in Fig. R2.3 and Fig. R2.4.



R2.3 (a) Time series at each station for the Storm2008 event, with black, blue, and orangered lines representing Meteomanz observations, the 3.25 km California SCREAM-RRM simulation, and the 100 m Bay Area SCREAM-RRM simulation, respectively. (b) and (c) are the same as (a), but for ISD and Tides and Currents observations. For the 3.25 km simulations, the thick line indicates the single-realization control run, and the thin lines represent ensemble members 1–10.



R2.4 Same as R2.3 but for the Stratocumulus2023 event.

For Storm2008, ensemble spread in wind speed and relative humidity begins to emerge after the 34th simulation hour, with temperature showing moderate spread and surface pressure showing very little. Nevertheless, the ensemble spread does not alter the first-order comparisons. For Stratocumulus2023, ensemble spread remains negligible throughout the two simulation days.

The metrics plots in the main text have been updated to include standard deviation bars for CA-3km, representing ensemble spread. The relatively large bias and RMSE in CA-3km are maintained, reinforcing that resolution sensitivity played a dominant role, rather than random variability:

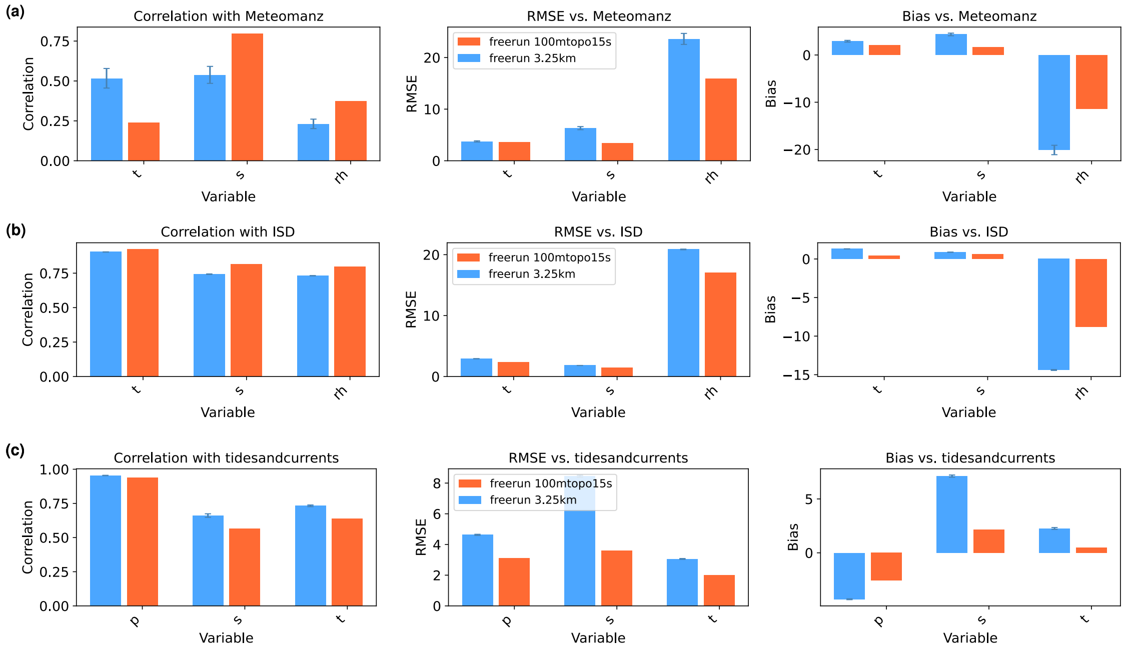


Figure 10. Skill scores for the Storm2008 event are shown for near-surface temperature (t), wind speed (s), relative humidity (rh), and surface pressure (p). These are compared against observations from (a) Meteomanz, (b) ISD, and (c) Tides and Currents, and presented as three overall metrics: Pearson correlation coefficient (left), root-mean-square error (RMSE, middle), and bias (right). The blue and orangered bars indicate simulation results from the 3.25 km California SCREAM-RRM and the 100 m Bay Area SCREAM-RRM, respectively. The ensemble spread in the 3.25 km simulation is represented by standard deviation bars.

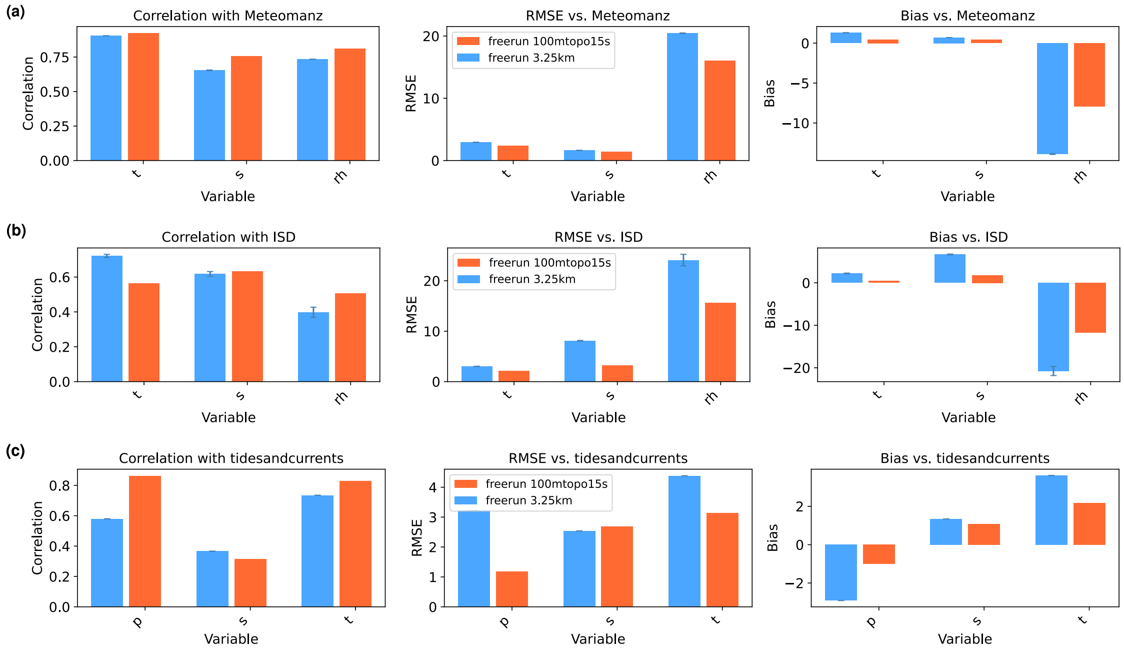


Figure 14. Same as Fig. 10 but for the Stratocumulus2023 event.

We have added the following descriptions to the Methods and Results sections of the manuscript:

> “\subsubsection{Sensitivity to initial condition} → in Methods
>
> In addition to the single-realization control runs, we conducted small ensembles (10 members each) for both events in the 3.25 km California RRM to quantify sensitivity to small perturbations in the initial conditions. Ensembles were generated by adding random perturbations to the initial temperature profiles across all grid points. Due to computational resource constraints, we were unable to perform ensemble simulations for BA-100m. Ensemble spread is represented by standard deviation bars in the 3.25 km California RRM metrics.
>
> \subsubsection{Storm2008} → in Results
>
> Ten ensemble members were run for CA-3km to assess sensitivity to initial condition perturbations. In Fig.~\ref{metricsStorm2008}, the vertical bars on CA-3km values represent the standard deviation across ensemble members. The relatively large bias and RMSE remain, highlighting the key role of resolution sensitivity rather than random variability. Ensemble spread appears after hour 34 of the simulation, most prominently in wind speed and relative humidity, but does not alter the first-order comparisons with observations or BA-100m. Spatially, except for small uncertainty in the location of the precipitation maximum, the moisture transport and precipitation patterns are highly robust (not shown).
>
> \subsubsection{Stratocumulus2023} → in Results
>
> The CA-3km ensemble shows virtually no ensemble spread throughout the two-day simulation (Fig.~\ref{metricsStratocumulus2023}).”

Specific comments:

1. Sec 2.1: For a horizontal resolution of 100 m, the 30 m near surface layer thickness seems high. Are you sure you are resolving the surface shear with such an aspect ratio?

Comparison with IGRA sounding observations shows that near-surface wind shear is generally well captured in the 100 m simulations, except for a slight underestimation in Storm2008 on 01-05 23PST and Stratocumulus2023 on 07-11 05PDT. Mirocha et al. (2010) found that an aspect ratio between 2 and 4 yielded the best agreement with similarity solutions in their boundary-layer LES simulations using WRF. The near-surface aspect ratio in SCREAM falls within this range.

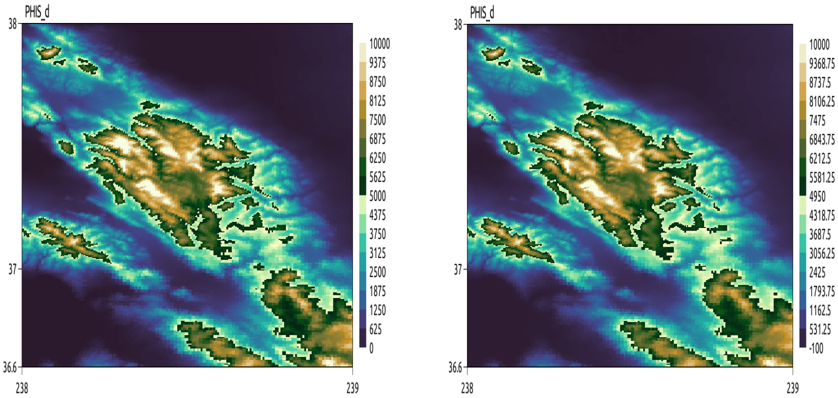
We have added this discussion to the IGRA results section.

References:

1. Mirocha, J. D., Lundquist, J. K., & Kosović, B. (2010). Implementation of a nonlinear subfilter turbulence stress model for large-eddy simulation in the Advanced Research WRF model. Monthly Weather Review, 138(11), 4212-4228.

2. Sec 2.3: The topography generation was discussed in much detail, however, no information was provided whether the extra smoothing has any effect on the terrain induced flows.

We were unable to compare the simulation “effects” between 6× and 12× smoothing because the run with 6× smoothing crashed.. We note that 12× smoothing is the recommended reference value in the E3SM v2 Topography Toolchain workflow. The figure below compares the topography after 6× (left) and 12× (right) smoothing, showing only minor visual differences:



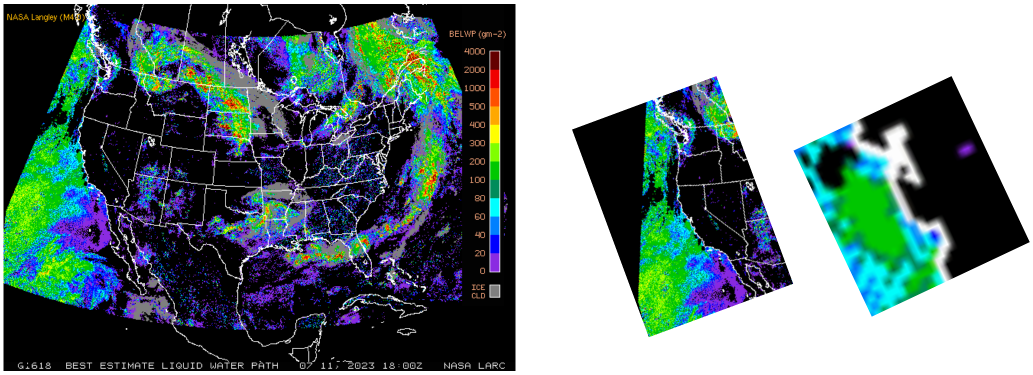
R2.5 BA-100m topography generated using 6 (left) and 12 (right) smoothing iterations.

3. Sec 2.4: What is the justification for increasing the hyper viscosity coefficient?

Looking back, we tested the sensitivity of several dycore parameters in the early setup tests, and the hyperviscosity coefficient was one of the commonly adjusted ones: raising it helps mitigate numerical instability. However, in retrospect, it appears that topography played a dominant role than this parameter. However, the hyperviscosity coefficient was not reverted to its default value when the simulations were eventually run.

4. Figure 5: Replace the GOES LWP map with a zoomed-in version so that it is easy to qualitatively compare against the model plots.

Since only the GOES images were available (not the raw data), we relied on a screenshot for a rough qualitative comparison. As you can see, after zooming in, the coarse resolution of GOES leaves very few pixels within the domain of interest:



R2.6 GOES-East/West Merged CONUS LWP Best Estimate, shown at three spatial scales: full North America (left), California (middle), and the Bay Area (right).

5. Even though the observations are sparse, the selection of the Storm2008 case warrants a quick comparison between the model simulated precipitation amounts and the observed ones. Capturing the extreme precip amounts has been a challenge for many fine resolution models. Showing that SCREAM at 100m does a reasonable job would make the discussion more robust.

During our initial evaluation, we did search for in situ precipitation records, but found that all available sources (Meteomanz, ISD, tidesandcurrents) either lacked precipitation observations entirely or reported maximum rates no higher than 3.2 mm/h, with most time steps missing data altogether. However, Storm2008 brought record-breaking extreme rainfall to the region (https://en.wikipedia.org/wiki/January_2008_North_American_storm_complex).

We appreciate your thoughtful and balanced comment! And we would like to reiterate that this study is not intended as a comprehensive evaluation of the model. Such an effort would require the involvement of experts across multiple areas. As noted in the discussion, we acknowledge that more complete information prior to the simulation might have guided us toward selecting an event with denser observational coverage. These are valuable lessons and experiences that will inform future works.

6. Figure 6: Some station names are overlapped, making it difficult to identify them.

Thank you for pointing this out! We have adjusted the figure to improve the readability of station names.

7. Figures 10 and 11 could be converted to tables.

Did you mean the original Fig. 10 and Fig. 14? We are quite fond of the detailed view in Fig. 11, as it provides information that summary metrics alone cannot convey (such as the high-frequency variability in observed winds). For the summary metrics figures, we've actually received suggestions in other works to convert tables into plots for clearer visual interpretation.

In response to your comment, we've added corresponding tables for the summary metrics while retaining the original plots, so readers can refer to whichever format they prefer:

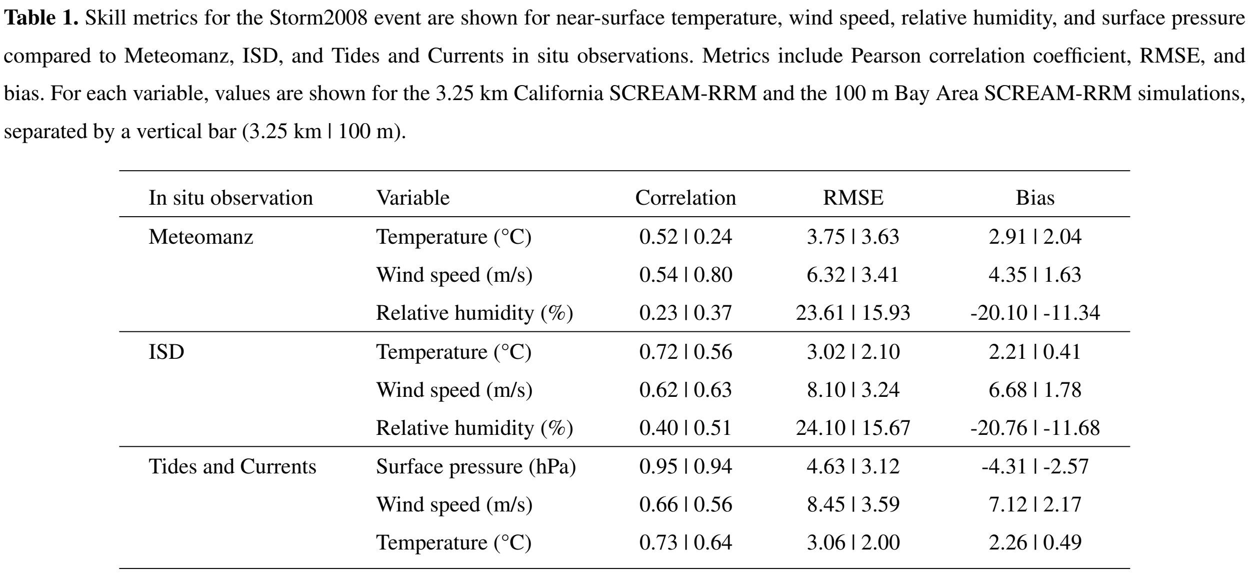


Table 1. Skill metrics for the Storm2008 event are shown for near-surface temperature, wind speed, relative humidity, and surface pressure compared to Meteomanz, ISD, and Tides and Currents in situ observations. Metrics include Pearson correlation coefficient, RMSE, and bias. For each variable, values are shown for the 3.25 km California SCREAM-RRM and the 100 m Bay Area SCREAM-RRM simulations, separated by a vertical bar (3.25 km | 100 m).

In situ observation	Variable	Correlation	RMSE	Bias
Meteomanz	Temperature (°C)	0.52 | 0.24	3.75 | 3.63	2.91 | 2.04
	Wind speed (m/s)	0.54 | 0.80	6.32 | 3.41	4.35 | 1.63
	Relative humidity (%)	0.23 | 0.37	23.61 | 15.93	-20.10 | -11.34
ISD	Temperature (°C)	0.72 | 0.56	3.02 | 2.10	2.21 | 0.41
	Wind speed (m/s)	0.62 | 0.63	8.10 | 3.24	6.68 | 1.78
	Relative humidity (%)	0.40 | 0.51	24.10 | 15.67	-20.76 | -11.68
Tides and Currents	Surface pressure (hPa)	0.95 | 0.94	4.63 | 3.12	-4.31 | -2.57
	Wind speed (m/s)	0.66 | 0.56	8.45 | 3.59	7.12 | 2.17
	Temperature (°C)	0.73 | 0.64	3.06 | 2.00	2.26 | 0.49

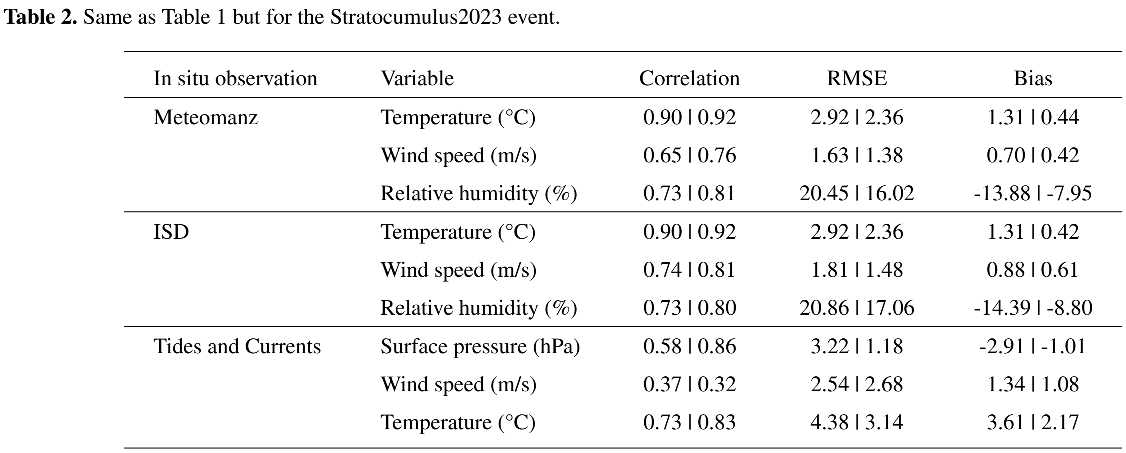


Table 2. Same as Table 1 but for the Stratocumulus2023 event.

In situ observation	Variable	Correlation	RMSE	Bias
Meteomanz	Temperature (°C)	0.90 | 0.92	2.92 | 2.36	1.31 | 0.44
	Wind speed (m/s)	0.65 | 0.76	1.63 | 1.38	0.70 | 0.42
	Relative humidity (%)	0.73 | 0.81	20.45 | 16.02	-13.88 | -7.95
ISD	Temperature (°C)	0.90 | 0.92	2.92 | 2.36	1.31 | 0.42
	Wind speed (m/s)	0.74 | 0.81	1.81 | 1.48	0.88 | 0.61
	Relative humidity (%)	0.73 | 0.80	20.86 | 17.06	-14.39 | -8.80
Tides and Currents	Surface pressure (hPa)	0.58 | 0.86	3.22 | 1.18	-2.91 | -1.01
	Wind speed (m/s)	0.37 | 0.32	2.54 | 2.68	1.34 | 1.08
	Temperature (°C)	0.73 | 0.83	4.38 | 3.14	3.61 | 2.17

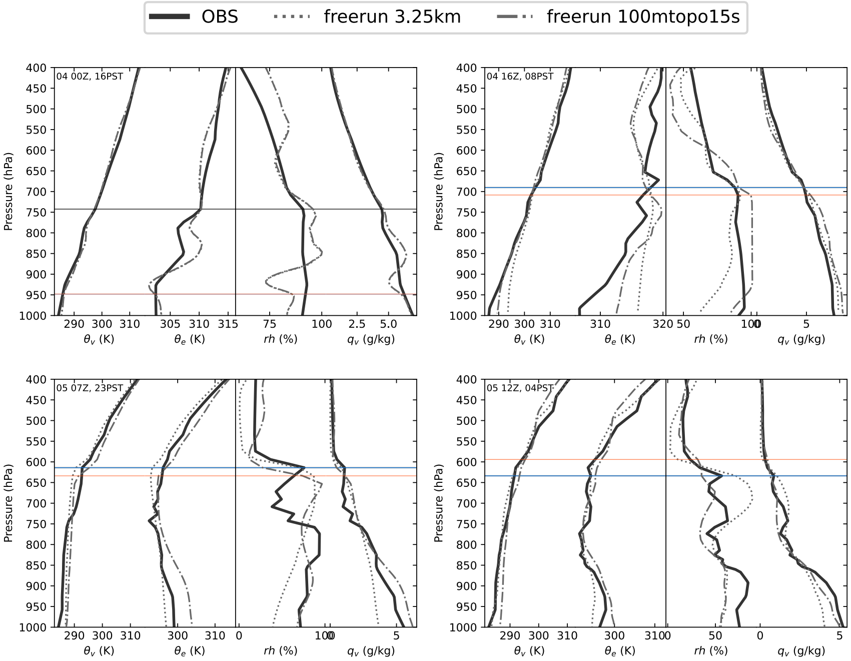
8. Figure 13 shows a significant difference in the mixed layer profiles at Oakland. Especially at 23PST on 2008-01-05, the dew point difference got a lot better between the 3.25 km and 100 m runs. This is a considerable difference and needs a comment or two in the manuscript. Also, it would be better to compare the boundary layer height estimations against the soundings to see if the LES is capturing all the turbulent motions.

Thank you for pointing this out! We have now explicitly referenced this figure in the sentence discussing the improved dew point in the 100 m run.

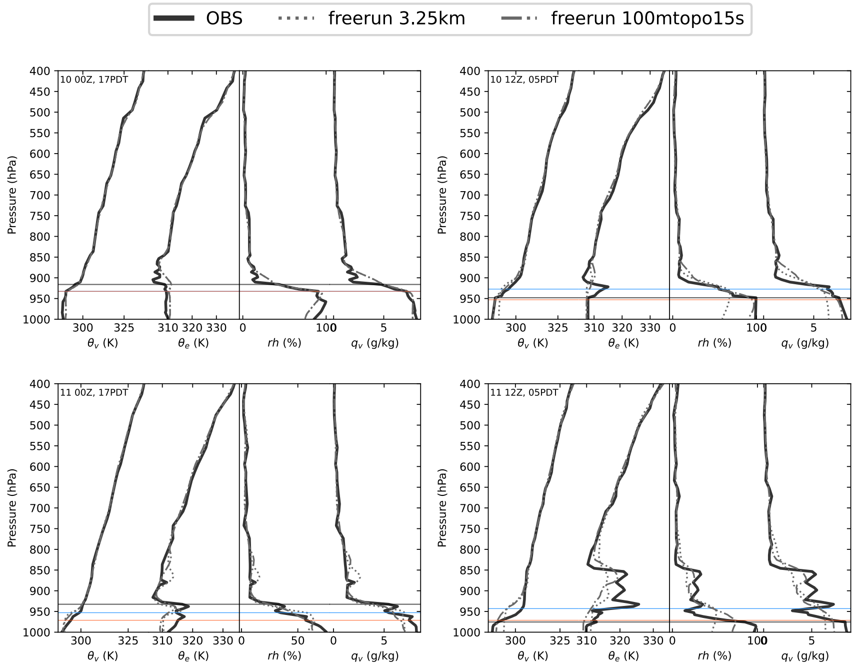
Also thanks for the nice suggestion about PBLH! While SCREAM includes an online diagnostic PBLH based on the Richardson number in SHOC, this method relies on surface friction velocity, which is not available from observations. We attempted to infer observational PBLH using the same approach by assuming a friction velocity, but the resulting estimates corresponded poorly with subjectively identified PBLH from the sounding plots.

Therefore, we adopted a simple RH-gradient method with the following steps: 1) Compute the forward finite difference of RH with respect to pressure (dRH/dp). 2) Identify RH inflection points that meet the following criteria: a) dRH/dp is positive at the current level, b) dRH/dp is larger at the current level than at the next lower level, c) dRH/dp becomes negative at the next lower level or the ratio between current and the next lower levels exceeds 2. 3) Among all detected RH inflection points, the one with the steepest gradient is taken as the PBL top.

Fig. R2.7 and Fig. R2.8 show the PBLH estimates from IGRA (black), CA-3km (blue), and BA-100m (orangered), respectively. Note that All profiles and sounding data were interpolated to the reference pressure levels before analysis. In the Storm2008 case, except at the initial time, the PBLH from CA-3km is visually indistinguishable from the IGRA estimate. The PBLH in the BA-100m simulation is generally lower. In the Stratocumulus2023 case, except at the initial time and at 07-11 00Z, the PBLH from BA-100m is closer to IGRA, whereas CA-3km tends to overestimate the stratocumulus boundary layer height.



R2.7 Vertical profiles of virtual potential temperature, equivalent potential temperature, relative humidity, and specific humidity at Oakland International Airport for the Storm2008 event. IGRA, the 3.25 km California RRM, and the 100 m Bay Area RRM are shown as thick solid, dotted, and dotted-dashed lines, respectively, with their corresponding estimated PBL heights indicated by black, blue, and orangered lines.



R2.8 Same as R2.7 but for the Stratocumulus2023 case.

A brief discussion of the PBLH comparison has been added to the main text:

> “\subsubsection{Storm2008} → in Results
>
> Estimated planetary boundary layer height (PBLH) at Oakland International Airport using a RH-gradient method agrees closely between CA-3km and IGRA, with differences that are visually indistinguishable. The BA-100m simulation generally underestimates PBLH.
>
> \subsubsection{Stratocumulus2023} → in Results
>
> We note that the estimated PBLH using a RH-gradient method from BA-100m aligns better with IGRA, while CA-3km tends to overestimate the stratocumulus boundary layer height.”

9. Lack of turbulence characterization in the study. Given how much of the manuscript relies on the LES capability, the turbulence metrics such as velocity power spectra, resolved to sub-grid turbulent kinetic energy, flux-gradient relations are expected. LES models are often judged by these metrics, and providing at least one such metric

would convince readers about the SCREAM model LES capabilities. Authors have hinted that improved performance could be linked to capturing turbulent mixing (L525) but did not provide any supporting evidence.

Thank you and Reviewer #1 for this valuable suggestion! In response, we have added two analyses: 1) SGS turbulent kinetic energy, fluxes, and 2) KE and w spectra. These additional analyses provide a more explicit characterization of turbulence and help support the LES-scale capabilities of SCREAM.

This content has been added into two subsections in the main text:

> "\subsection{Sub-grid-scale flux}
>
> Building on the resolution sensitivity study of DP-SCREAM in \citet{Bogenschutz2023}, SCREAM exhibits characteristics of a scale aware model. As horizontal resolution increases, the partitioning between SGS and resolved turbulence diminishes. This scale awareness, inherent to the SHOC parameterization \citep{Bogenschutz_Krueger2013}, enables SCREAM to operate effectively at 100 m resolution without the need for parameter tuning. Specifically, Fig. 9 and Fig. 16 in \citet{Bogenschutz2023} show that in the marine stratocumulus case, maritime shallow cumulus case, and mixed-phase Arctic stratocumulus case, as the horizontal grid spacing Δx decreases, the contribution of SGS moisture flux becomes smaller while the resolved flux becomes increasingly dominant. In the 100 m DP-SCREAM simulations, above 0.2 km, the proportion of SGS moisture flux is negligible, and the resolved flux is nearly unity.
>
> In our simulations, although we did not output high-frequency resolved moisture flux or TKE, Fig.~\ref{SHOCtimeplevStorm2008} and Fig.~\ref{SHOCtimeplevStratocumulus2023} shows significant reductions in SGS TKE, moisture flux/variance, \emph{w} variance, and the third moment of \emph{w} in the BA-100m simulations compared to the CA-3km simulations. For clarity, the plotted ranges in these figures differ between the two resolutions, with the CA-3km values being much larger.

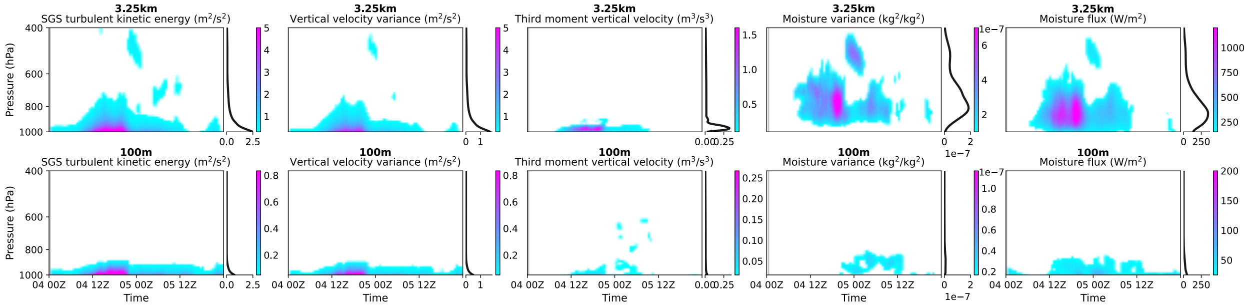


Figure 18. Simulated Sub-grid-scale (SGS) variables in the 3.25 km California RRM (top) and 100 m Bay Area RRM (bottom) during the Storm2008 event. From left to right: TKE, vertical velocity variance, third-moment vertical velocity, moisture variance, and moisture flux. Each panel consists of a time–evolution shading plot on the left and a vertical profile averaged over the simulation period on the right. For clarity, the colorbars differ between the 3.25 km and 100 m simulations, with the former being much larger.

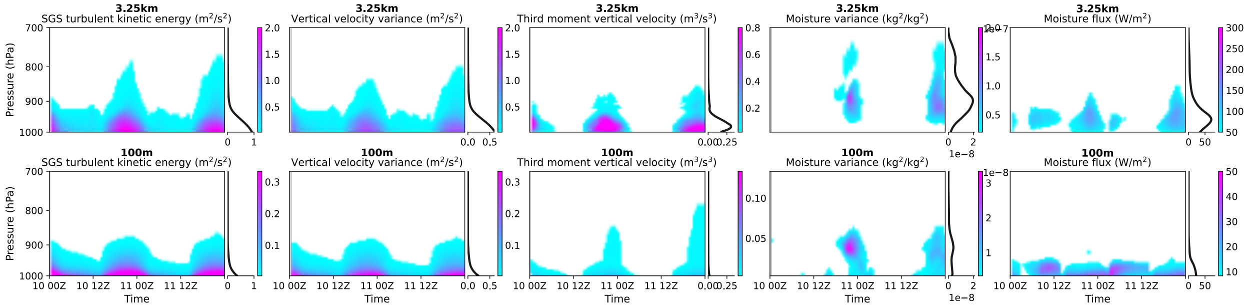


Figure 19. Same as Fig. 18 but for the Stratocumulus2023 event.

At 100 m resolution, simulations are close to, but largely below, the turbulence gray zone, where the grid spacing becomes comparable to the dominant eddy scale. The gray zone typically spans the transition from mesoscale models, which rely on ensemble-averaged vertical fluxes, to LES, where most turbulent motions are explicitly resolved and subgrid closures play only a minor dissipative role \citep{Wyngaard2004}. In this transitional regime, subgrid transport is best treated with 3D turbulence schemes that represent the full stress tensor \citep[e.g.,][]{Wyngaard2004,Chow2019,Honnert2020}, whereas SHOC currently parameterizes only vertical mixing. Thus, at coarser resolutions such as 800 m, a 3D implementation of SHOC would likely be beneficial. At 100 m, near the lower edge of the gray zone, the need for 3D turbulence remains uncertain and case-dependent, pending the implementation and testing of such a scheme in SCREAM."

"\subsection{Energy spectra} → in Methods

For global models, spherical harmonics are a natural method for spectral decomposition, but they are not suitable for limited-area regional models and RRMs. For regional outputs, Discrete Fourier Transforms (DFT) and Discrete Cosine Transforms (DCT) are commonly used. DFT requires detrending or windowing, while detrending can artificially remove large-scale gradients, and windowing can distort spectra for already periodic fields \citep{Errico1985, Denis2002}. DCT mirrors the field to ensure symmetry before applying a Fourier transform, and is reliable for fields with spectral slopes between –4 and 1. DCT was originally developed for digital image, audio, and video compression (e.g., JPEG), but it has also been used to diagnose energy spectra in numerical simulations \citep[e.g.,][]{Denis2002, Selz2019, Prein2022}.

We used the scipy.fft package (\url{https://docs.scipy.org/doc/scipy/tutorial/fft.html}, last access: 17 September 2025) for Discrete Cosine Transforms. Since we only output high-frequency relative vorticity, divergence, and \emph{w} profiles, we computed rotational and divergent KE spectra as well as \emph{w} spectra at every level using 10 min instantaneous outputs. The raw outputs were on the dynamical GLL grid; they were horizontally interpolated using the NCO-native first-order conservative algorithm to 0.03° (3.25 km California RRM) and 0.001° (100 m Bay Area RRM) over the small domain of the 100 m mesh (237.5E–238.5E, 37.3N–38.3N), and vertically interpolated to SCREAM's 128 reference pressure levels using NCO's default method. The two-dimensional spectra were projected onto the zonal and meridional directions, then averaged in time (from the 6th simulation hour to the end of the simulation) and in the vertical (within 100 hPa around 200, 500, and 850 hPa, respectively).

\subsection{Energy spectra} → in Results

Energy spectra provide a benchmark for evaluating the transition from large-scale quasi-2D motions to small-scale 3D turbulence. The canonical k^{-3} slope at synoptic scales and $k^{-5/3}$ slope at mesoscale wavelengths \citep{Nastrom_Gage1985} are widely used to assess effective resolution and numerical diffusion in atmospheric models \citep[e.g.,][]{Skamarock2004, Jablonowski_Williamson2011, Caldwell2021}. Numerous studies have examined KE spectra in global and regional models \citep[e.g.,][]{Bierdel2012, Skamarock2014, Durran2017, Menchaca_Durran2019, Prein2022, ZhangY2022, Khairoutdinov2022, Silvestri2024}, with some studies have emphasized the rotational and divergent components \citep[e.g.,][]{Hamilton2008, BlažicaN2013, Selz2019}. Spectra of vertical velocity (\emph{w}) are also informative, as they emphasize divergent motions and typically peak at mesoscale wavelengths

\citep{Bryan2003, Schumann2019}. Figures~\ref{spectraStorm2008}–\ref{spectraStratocumulus2023} show KE and \emph{w} spectra for the Storm2008 and Stratocumulus2023 cases.

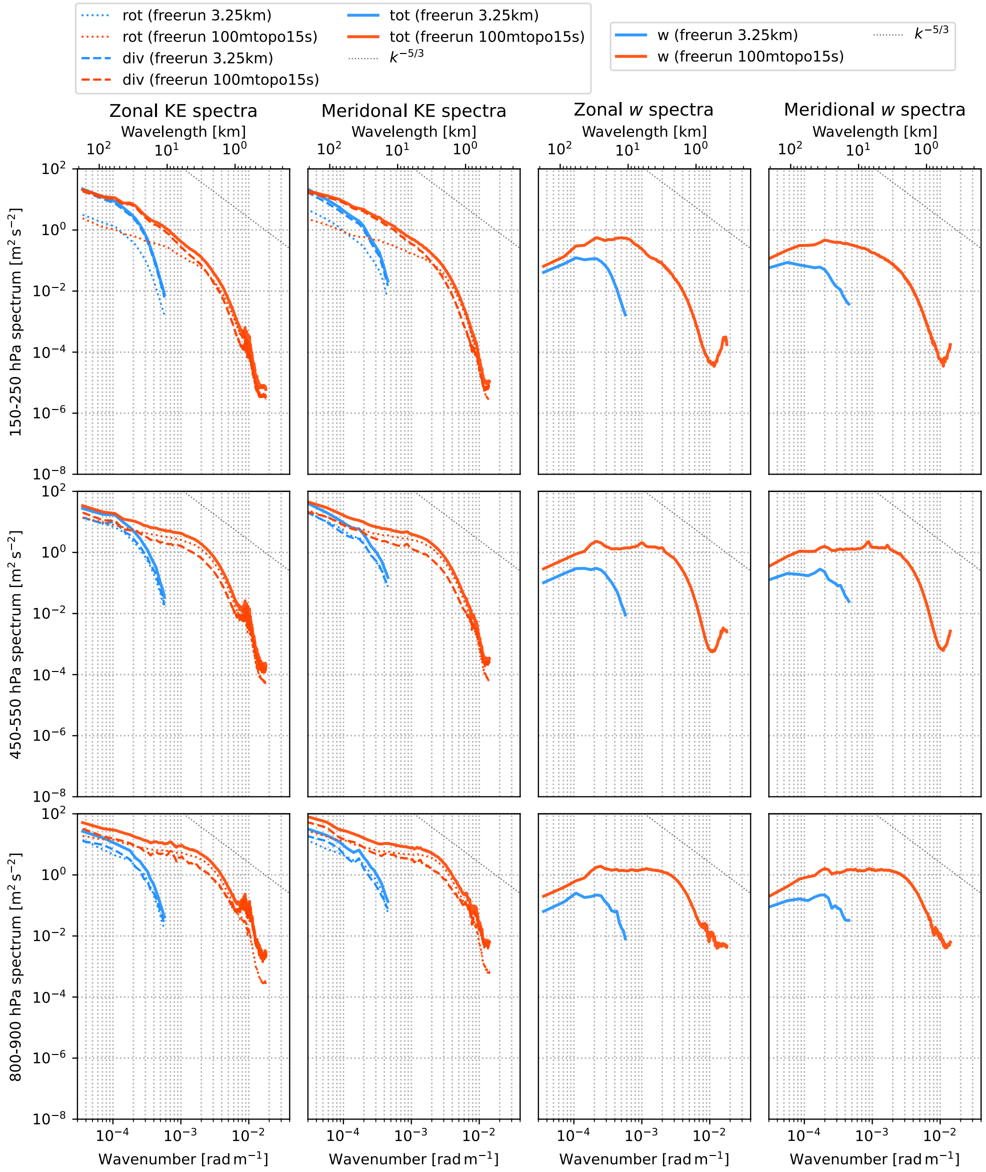


Figure 20. Energy spectra simulated by the 3.25 km California RRM (blue) and the 100 m Bay Area RRM (orangered) for the Storm2008 case. From left to right: zonal kinetic energy (KE), meridional KE, zonal vertical velocity (w), and meridional w spectra. From top to bottom: averages centered at 200 hPa, 500 hPa, and 850 hPa, each using a 100 hPa vertical window. In the KE spectra, the total, rotational, and divergent components are shown as thick solid, thin dotted, and thin dashed lines, respectively. A reference $k^{-5/3}$ slope line is shown in the top-right corner.

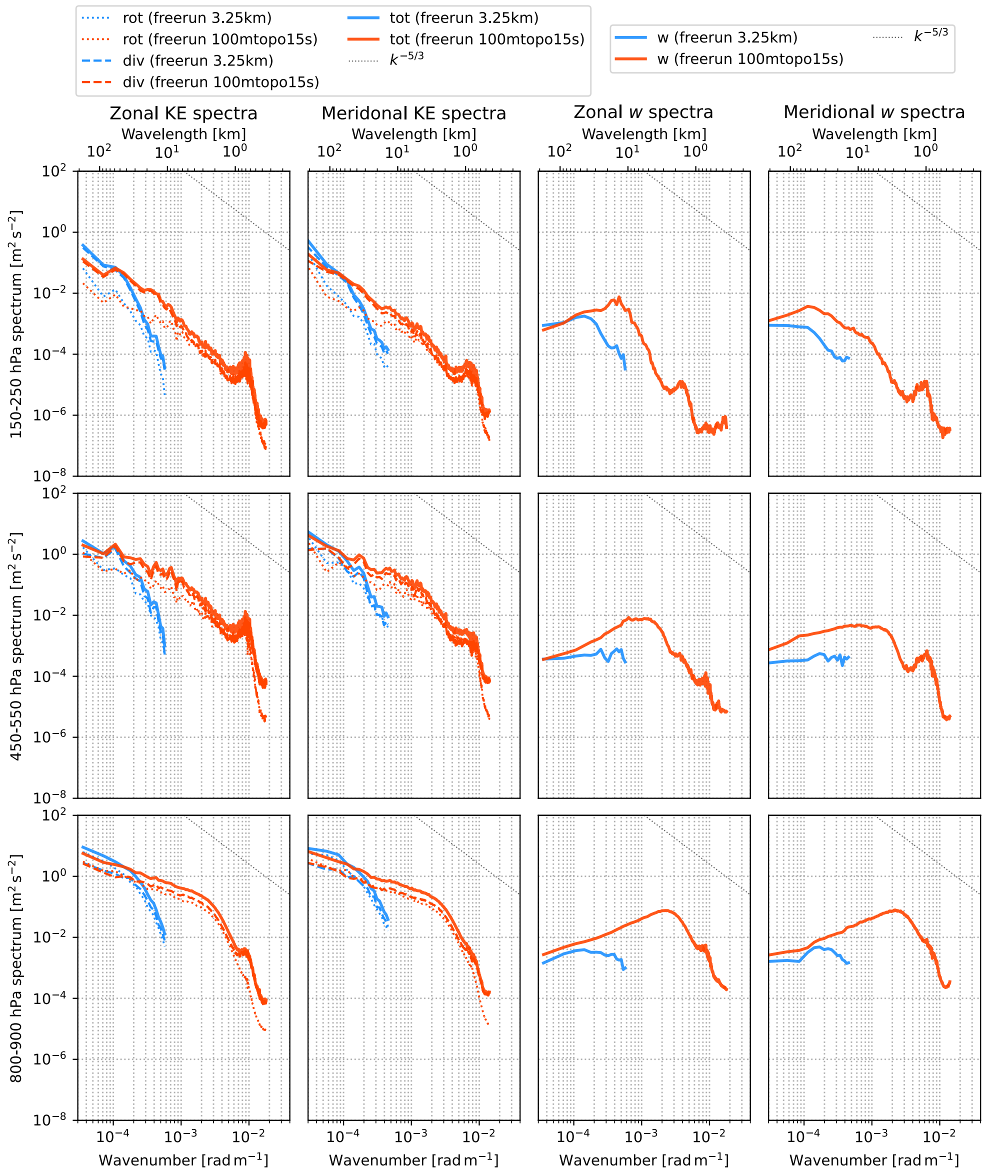


Figure 21. Same as Fig. 20 but for the Stratocumulus2023 event.

First, the CA-3km simulations roll off sooner than the global spectra in \citet{Caldwell2021}; however, we caution that the two differ in important ways (DCT over a region vs. spherical harmonics globally; two two-day events vs. 40-day statistics). For BA-100m, the effective resolution in Stratocumulus2023 is shorter than in Storm2008 if the roll-off standard is applied. However, within the mesoscale regime (10 km–100 m), the Storm2008 KE spectra flatten relative to $k^{-5/3}$ before steepening again, which corresponds to the earlier roll-off. The mesoscale flattening of the KE spectra in this case may reflect a genuine accumulation of mesoscale energy, given that this was a record-breaking extreme event. In both cases, it is robust that BA-100m contains much more small-scale energy than CA-3km.

A notable feature in both events is a sharp KE increase between 1 km and 500 m (approximately) in BA-100m. We suspect this results from the blocky mountain effect, caused by the mismatch between the 800 m cubed-sphere topography (the highest-resolution global dataset available) and the model's 100 m grid spacing. Toolchain memory limits prevented higher-resolution topography, so the effective topographic resolution lags behind the model Δx, producing unnaturally flat peaks and steep slopes. These slopes can generate excess high-wavenumber energy, consistent with the abrupt KE rise below 1 km. An alternative is contamination by inflow of coarser-resolution energy from the surrounding 800 m mesh via lateral boundaries. However, this is inconsistent with: 1) the effect being stronger in Stratocumulus2023 than in Storm2008, whereas boundary advection should amplify it in the latter; and 2) the 800 m mesh itself having an effective resolution of \textasciitilde4.8 km, with KE decaying rapidly beyond that, making a rise near 500 m hard to explain. To confirm the blocky mountain hypothesis, sensitivity tests with 100 m cubed-sphere topography are needed. To rule out lateral boundary effects, larger RRM domains need be tested \citep[cf.][]{Bogenschutz2024}. In either case, toolchain memory upgrades are essential.

The \emph{w} spectra exhibit a mesoscale peak, consistent with \citet{Bryan2003, Schumann2019}, with a cutoff near 1–3 km in BA-100m. Below 1 km, the ratio of \emph{w} to KE spectra approaches unity in Storm2008 but remains below one in Stratocumulus2023, except in the lower troposphere. The temporal evolution further shows rapid development of small- and mesoscale \emph{w} spectra, which are well developed within the first simulation hour (not shown).

Overall, KE spectra vary substantially across these two events, reflecting the influence of different forcing mechanisms and dynamical regimes. This aligns with \citet{Menchaca_Durran2019} and \citet{Selz2019}, and does not support a universal mesoscale KE spectrum."

10. As noted in the manuscript, despite improvements, the model consistently underestimates the surface pressure at the majority of stations. Where does this bias come from? Is it the model dynamics or the initial conditions? Authors should include some discussion on this.

Thanks for asking this! Since all simulations were initialized using ERA5 data (with both atmospheric and land surface conditions consistently initialized), our study is unable to isolate or quantify the contribution of initial conditions to the surface pressure bias. However, one potential factor related to initialization is the surface adjustment procedure. Because ERA5 has much coarser topography than SCREAM, failing to adjust surface pressure according to the high-resolution model topography would introduce spurious low or high pressure and hydrostatic imbalance at the model surface, which in turn triggers strong artificial gravity wave oscillations during spin-up. To mitigate this, the surface pressure adjustment was applied based on the difference in elevation between the source (ERA5) and target (SCREAM) topographies. This procedure uses the hydrostatic equation by assuming a dry adiabatic lapse rate (Trenberth, 1993). However, this assumption may introduce bias relative to the true surface pressure.

To our knowledge, some groups have started using O(10)-km reanalysis data blended with km-scale assimilated forecasts (e.g., 3 km High-Resolution Rapid Refresh forecast) to initialize km-scale simulations. These groups might have examined whether initializing with ERA5 vs. ERA5 + HRRR leads to meaningful differences. Replacing ERA5 with an alternative reanalysis product is also a possibility, although it would require careful coordination of atmospheric and land surface spin-up under consistent forcing. An alternative view is that, although the model spectrum tends to lack sufficient mesoscale kinetic energy during initialization, the mesoscale spectrum can rapidly develop (Skamarock, 2004). In this case, providing only large-scale kinetic energy in the initial conditions may not lead to significant loss. Whether model dynamics alone could lead to surface pressure errors of 1–2 hPa remains unclear.

We have added a brief discussion in the Results section:

> “Despite improvements, the model consistently underestimates surface pressure at the majority of stations. It remains unclear whether this bias is related to model dynamics. One possible factor tied to the initial conditions is the surface adjustment procedure, which assumes a dry adiabatic lapse rate. Some groups have begun blending reanalysis data at O(10)-km resolution with \emph{k}-scale analysis products to initialize \emph{k}-scale models. An alternative view suggests that, while mesoscale kinetic energy is initially lacking, it can rapidly spin up \citep{Skamarock2004}, so supplying only large-scale energy may not result in significant degradation.”