

In ‘A study of irrigation impacts on California’s climate with the variable-resolution CESM,’ Huang and Ullrich use a variable-resolution version of the CESM to test the hypothesis that irrigation in the Central Valley (CV) of California is an important component of the region’s surface energy budget that must be parameterized in climate models in order to properly simulate temperature statistics. They perform three experiments: a control simulation with no irrigation parameterization and two simulations with differing degrees of parameterized irrigation. They show that the two simulations with parameterized irrigation have temperature statistics that appear to match better with observations than the control simulation. They also explore differences in heat stress and humidity and show evidence that parameterized irrigation reduces the former and increases the latter.

This study builds on a number of previous modeling studies that have explored the importance of irrigation in controlling the climate of the CV region in the following ways: (1) it employs high resolution over the CV region, (2) it uses a different irrigation parameterization, (3) it uses a variable resolution atmospheric model (rather than the low-resolution global and moderate-resolution limited-area models that have been previously used), and (4) it explores the effect of irrigation on temperature statistics, including extreme heat. Importantly, this study verifies (albeit with a not-entirely independent model) the findings of previous studies that irrigation generally lowers temperatures in the CV region. These are important contributions to the literature and warrant eventual publication.

The paper has been prepared for submission to Geophysical Research Letters, which generally publishes papers that meet one or more of the following criteria:

- High impact innovative results with broad geophysical implications at the forefront of one or several AGU disciplines.
- Results with immediate impact on the research of others and requiring rapid publication.
- Instrument or methods manuscript introducing an innovative technique that makes new science advance possible, with immediate applications to AGU disciplines.

Broadly speaking, I would characterize the contributions noted above as incremental—but important—advances in our understanding of the role of irrigation on the CV region, and these contributions are mainly relevant to persons who use climate models to investigate climate processes and change in the CV and other similarly irrigated regions. Given this, I do not think that GRL is the appropriate forum for this manuscript. I would instead recommend that this be submitted to JGR-Atmospheres, J. Climate, or Climate Dynamics. I believe that submission to any of these journals could be done with minimal modification to the manuscript.

Aside from issues of journal scope, there are a few two issues that detract from the quality of the manuscript: (1) there is a null hypothesis (that differences among the simulations are purely random and not due to irrigation parameterization) that the authors do not attempt to falsify, and (2) the authors do not adequately place the new results from their study in the context of previous literature. A thorough description of these two issues follows:

Major issues

Falsification of the null hypothesis

The central hypothesis of the manuscript (though not explicitly stated as such) is that irrigation in the Central Valley (CV) of California is an important component of the region's surface energy budget that must be parameterized in climate models in order to properly simulate temperature statistics. This is tested by comparing simulations with parameterized irrigation and without.

The null hypothesis for this experimental design is that differences among the simulations is due to random variation. The authors show a number of difference plots, but only in one instance to the authors refer to statistical significance (presumably relative to the null hypothesis that I have just described).

Throughout the manuscript, the authors should attempt to falsify this null hypothesis, they should report how they do so, and they should report their confidence that the reported differences can not be explained by the null hypothesis (e.g., report statistical significance or, even better, confidence intervals).

This would best be done by repeating an ensemble of these simulations, reporting the ensemble difference, and reporting the degree to which this difference is significantly different from 0. However, given that this is computationally expensive, this could be done in two stages. First, the authors should show differences between the IRG simulations; based on visual inspection of plots in the existing manuscript, I believe that these differences will be quite small relative to differences between the IRG simulations and the control. This would establish that intrinsic variability (plus some differences due to the different level of irrigation) is small relative to the effect of irrigation, and that significance can then be established by simple metrics like the Student's T-test.

Assuming that the authors go this route, then the following should be done:

1. Figure 3 should include difference plots for IRG-IRG(0.5) to show the level of intrinsic variability
2. Figure 3 should include difference plots for NRG-IRG
3. The NRG-IRG should include hatching to denote statistical significance as calculated by a simple T-test [but please use the spatial autocorrela-

tion adjustment for significance described by Ventura et al. (2004, doi: 10.1175/3199.1)]

4. Table 2 should include a row for the NRG-IRG difference, and it should include an indication of significance, as calculated using a simple T-test (or confidence intervals)
5. Please use a Kolmogorov-Smirnov test for the PDFs shown in Figure 5 to assess whether the NRG and IRG simulations are significantly different.

Discussion of manuscript context

The authors do an excellent job in the Introduction section of describing the existing relevant literature around irrigation and climate in the CV region. However, in the discussion section, the authors do not revisit these papers to place their new findings in the context of this previous literature. Therefore, it is difficult to glean from this manuscript how exactly this manuscript has advanced our knowledge of how irrigation impacts climate in the CV region (and possibly other regions). In the second paragraph of this review, I believe I have succinctly summarized the new contributions from this manuscript; however, I had to infer these unique contributions partly from combining the author's results with my own understanding of the literature. I should have been able to pull extract these contributions explicitly from the discussion section.

The discussion section would also benefit from an expanded discussion of the implications of this study: i.e., it should explicitly describe 'why are these new results important, and what do they mean for future research'?

Minor issues

Stratocumulus

I think that the authors have missed a great opportunity in this manuscript: to re-evaluate the claim of Lo et al. (2013, doi:10.1002/jgrd.50516) that irrigation in the CV region affects off-shore atmospheric stability and the presence of stratocumulus clouds. There are two key reasons why the Lo et al. (2013) study may have presented faulty results. First, they use an incredibly low-resolution atmospheric model, and the results that they present are well within the model's diffusion stencil; therefore the 'remote' effect on atmospheric stability could be purely due to horizontal (and presumably not realistic) diffusion. Second, they use the CAM3 physics parameterization, which (a) uses a boundary layer parameterization that is known to be inappropriate for the boundary layer conditions that occur in the offshore region of California, and (b) parameterizes stratocumulus clouds simply as being proportional to lower tropospheric stability.

The simulation in the current manuscript uses much higher resolution in the CA region, and it uses the CAM5 physics package, which has a parameterization that is explicitly designed for the marine boundary layer conditions like those off the shore of California. It would be quite simple for the authors to simply replicate some of the analyses done by Lo et al. (2013): i.e., create difference plots of low cloud cover, cloud LWP, etc. I think there is a strong possibility that the authors would come to a different conclusion than that of Lo et al.

Precipitation tails

In Figure 4c, the upper tails of the whisker boxes for both IRG simulations are about 2x higher than the control and observations, but this is not mentioned or explored in the manuscript. This should at least be mentioned, and it would be best if this were explored: what season is this intense precipitation happening? Nominally, the highest precipitation is in the winter, but that is when irrigation is lowest; so does this mean that irrigation in the NRG simulations is triggering summertime convection?

Absurdly high correlation values

In Table 1, the correlation coefficients between the simulation and the observations is absurdly high: 0.999. I think that this indicates that the way in which the correlation coefficient is calculated is not appropriate. The caption and the manuscript text are a bit vague about how this is calculated, but the caption for the table seems to imply that the authors did the following to calculate these coefficients: (1) averaged Tmax within the CV region for the model and observations, (2) averaged these values over the JJA months for each year, and (3) calculated the correlation coefficient from these averaged values. This would be a reasonable way to evaluate the correlation, and it would reflect the degree to which the simulations reproduce the interannual variability of the observations. However, a coefficient of 0.999 implies that the model is doing a perfect job of capturing interannual variability, which I strongly doubt is the case.

Specific issues

line 116: ‘mostly described’; the authors should qualify what is meant by ‘mostly’ (if the following paragraph does so, it should be indicated in this sentence that differences with Sacks et al. are described in the following paragraph)

line 117 (and Eqn 1): this parameterization is designed to approximate human behavior, but this is not noted in the manuscript. It might be useful to note this and to justify how/why (in simple terms) this parameterization (and $\alpha=0.7$) is a reasonable approximation of what humans aim to do when irrigating. One

way of describing it might be that when we choose to irrigate, we irrigate enough such that the plants aren't water stressed, but not so much that the soil is completely saturated.

lines 129-132: why do you spin up the NRG simulation in this way? This should be described briefly in the manuscript.

line 130: 'was obtained': from where? The source should be cited.

line 131: 'realistic': could you be more specific here? Do you mean to say that the 3 min dataset provides irrigation coverage that is more appropriate for the 28km grid?

line 133: why do you use the term 'non-arctic grass' here? It seems slightly odd to be so specific (I recognize that this might be CLM jargon). I think it would be appropriate to just say 'grass' here.

line 137: 'in response to Mediterranean summers': it might be useful to again iterate that this is humans responding to Mediterranean summer conditions; we choose to irrigate when it is hot and dry.

line 139: I did not see any supplementary figures?

line 154+3 (note that line numbers drop briefly after line 154): '155 degrees of freedom': do you mean grid points and/or CLM columns here? I would avoid 'degrees of freedom' since it ambiguously implies that you might have adjusted for spatial autocorrelation.

line 212: 'significant': do you mean statistically significant here? If not, I would choose another word. If so, you should specify how you are evaluating statistical significance.

line 216: Comparing->Compared

lines 267-268: 'statistically significant': you should report how you evaluate statistical significance (what test, what are the criteria for rejecting the null hypothesis, etc., so that a reader could reproduce your results and your assessment of significance).

Figure 3: Which variable does the last 3-panel set of differences refer to, T_{min} or T_{avg}? Also, why is there not a third row of difference plots corresponding to the other variable?

Figure 3: it might be useful to come up with a way of visually emphasizing the differences in the CV region. This could be done, for example, by using NCL to overlay a field that is transparent where irrigation fraction (Fig 2) is high and semi-opaque where irrigation fraction is low.