California Wintertime Precipitation Bias in Regional and Global Climate Models

Peter Caldwell *

Lawrence Livermore National Lab, Livermore, CA

May 20, 2010

* *Corresponding author address:* Peter Caldwell, L-103, Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, CA 94566. E-mail: caldwell19@llnl.gov.

Abstract

In this paper, wintertime precipitation from a variety of observational datasets, regional climate models (RCMs), and general circulation models (GCMs) is averaged over the state of California (CA) and compared. Several averaging methodologies are considered and all are found to give similar values when model grid spacing is less than 3°. This suggests that CA is a reasonable size for regional intercomparisons using modern GCMs.

Results show that reanalysis-forced RCMs tend to significantly overpredict CA precipitation. This appears to be due mainly to overprediction of extreme events; RCM precipitation frequency is generally underpredicted. Overprediction is also reflected in wintertime precipitation variability, which tends to be too high for RCMs on both daily and interannual scales.

Wintertime precipitation in most (but not all) GCMs is underestimated. This is in contrast to previous studies based on global blended gauge/satellite observations which are shown here to underestimate precipitation relative to higher-resolution gauge-only datasets. Several GCMs provide reasonable daily precipitation distributions, a trait which doesn't seem tied to model resolution. GCM daily and interannual variability is generally underpredicted.

1 Introduction

In recent years, the focus of climate science has shifted from proving/disproving the existence of global warming to providing guidance for climate change adaptation planning (Shukla et al., 2009). This new role is more challenging because climate impacts vary from region to region and depend not just on the sign but also on the magnitude of future change. General circulation models (GCMs) are our best tools for forecasting future climate, but vary in the amount and geographical distribution of their predicted changes. In the face of this uncertainty, model intercomparisons provide a critical sense of the the range of possibilities confronting us.

A key problem with GCMs is that their grid spacing is typically measured in hundreds of km, which is too coarse to capture regional features (such as lakes or mountains) that may play a central role in determining how climate change affects our day-to-day lives. This is a particular problem for precipitation (Pr), which depends strongly on local topography. In order to obtain information at the needed scales, GCM output is often downscaled to finer resolution. This can be done through the use of statistical relationships between GCM-scale and fine-scale climate variables or by running a high-resolution regional climate model (RCM) forced at the boundaries by GCM data. Both of these techniques have drawbacks. Statistical downscaling methods can only be trained on current climate data, so it is unclear whether the relationships underlying any particular method will continue to hold in a different climate. RCM predictions are uncertain because their boundary condition treatment and physics are complex and imperfect.

Because RCMs present both benefits and drawbacks relative to GCMs, it is useful to assess the value they add. There are two ways that RCMs are expected to add value. First, RCMs provide information on scales too small to be resolved by GCMs. The validity of this benefit is irrefutable, and by itself justifies the use of RCMs by researchers interested in local climate. The second expectation is that RCMs are more accurate because they better resolve physical processes and the local terrain. This means that even when averaged to GCM scale, downscaling should in theory yield better results. The physical processes controlling Pr are in particular expected to improve with resolution because Pr depends heavily on topography (which becomes more realistic at higher resolution) and because a larger fraction of precipitation is explicitly resolved at higher resolution, reducing dependence on the (more empirical) convective parameterization. Expectation of improved orographic Pr simulation at higher resolution is pervasive in the literature (e.g. Tibaldi et al., 1990; Leung and Ghan, 1995; Brankovic and Gregory, 2001; Rauscher et al., 2009).

There are already many papers showing that dynamical downscaling adds value, but most have focused on metrics that reward RCMs for having output at higher resolution (e.g. by comparing against high-resolution or point measurements or by praising RCM maps for their fine spatial structure). Since these studies convolve the two types of "added value" noted above, they fail to show whether downscaling actually improves large-scale accuracy. GCM-scale improvement can be gleaned from studies that consider regional averages. For example, Christensen et al. (1998) find RCM Pr bias over Scandinavia to be worse than its forcing GCM, but suggest that this could be due to problems with the observations. Leung et al. (2003) find downscaling to increase precipitation (Pr) bias over the Columbia River basin, but to decrease error over the Sacramento/San Joaquin area. In Duffy et al. (2006), Pr from 4 RCMs averaged over the Western United States fail to improve upon the results from their driving GCMs. Seth et al. (2007) found their RCM to have trouble reproducing the annual cycle of Pr over 4 South American subregions, adding little value except in Northeast Brazil. Jacob et al. (2007) compute Pr bias for 13 different RCMs over 8 European subregions (as part of the PRUDENCE project); downscaling reduces bias in just over half of their cases. Sylla et al. (2009) show mixed benefits from downscaling over 8 African subregions.

None of the aforementioned studies focus on the value added at the GCM scale and most make no explicit mention of the differences between results from downscaling versus from the driving GCM. Caldwell et al. (2009) (hereafter C09) computed regional averages for a particular RCM/GCM combination over California (CA) with a focus on GCM-scale improvements; they conclude that their regional model had generally worse Pr bias than its forcing GCM. This study examines whether the C09 result is typical for GCM/RCM pairings over CA.

Lack of improvement in RCMs could come from several sources. For example, specification of lateral boundary conditions for limited area models is still imperfect (e.g. Staniforth, 1997). Additionally, model performance may actually not be improved by increased resolution as commonly expected. This has been investigated in a variety of previous studies (using both RCMs and GCMs) and is nicely summarized in Rauscher et al. (2009) (hereafter R09). Briefly, many studies show higher resolution to improve Pr simulation (e.g. Colle et al., 2000; Mass et al., 2002; Duffy et al., 2003; Iorio et al., 2004; Gao et al., 2006; Rojas, 2006). Other research (e.g. Pope and Stratton, 2002; Leung and Qian, 2003) find little improvement or even degradation in Pr at higher resolution. Results seem to be regionally and seasonally dependent. Duffy et al. (2003) find greatest improvement during fall and winter, which they attribute to the fact that convective (parameterized) Pr is less important during these seasons. R09, however, find no improvement with resolution in winter, with some in summer. Discrepancy between these studies could be due to diminishing returns at higher resolution (as noted by Colle et al., 2000; Mass et al., 2002): R09 compares 25 km and 50 km RCM simulations, while Duffy et al. (2003) compare GCM runs at 55 km, 75 km, and 310 km. This study adds to the discussion by examining whether resolution is the leading indicator of Pr bias across a variety of models.

Another motivation for this work was the realization that most climate models (particularly RCMs) overpredict wintertime Pr over the W coast of the US (C09 and references therein; Leung et al., 2003; Coquard et al., 2004; Phillips and Gleckler, 2006). This seems to also be the case for other coastal regions (e.g. R09; Christensen et al., 1998), but doesn't hold for inland mountain regions (Rasmussen, 2009). There is some evidence that overprediction increases at higher resolution (e.g. Colle et al., 2000; Mass et al., 2002; Leung and Qian, 2003). A limited number of studies suggest that this effect is due to sensitivity of physical parameterizations rather than increased sharpness of topography (Giorgi and Marinucci, 1996; Han and Roads, 2004; Gao et al., 2006).

In this study, we compare CA-average wintertime Pr as simulated by a large number of RCMs and GCMs in order to assess the consistency of RCM overprediction and to get a better sense of the benefits of resolution and downscaling. We focus on CA because its huge irrigated-agriculture industry, large population, and subtropical position place great demands on its water resources. Additionally, CA is interesting because downscaling is expected to add the most value in regions like CA which have complex topography, yet the above studies suggest that this expectation may not be borne out. We focus on wintertime precipitation because this is when CA gets the bulk of its water supply. Statewide averages are used because CA is large enough to be resolved by current-generation GCMs but small enough to be meaningful as a climatic unit. We evaluate model ability to capture the observed CA-average Pr statistics rather than ability to reproduce temporal or spatial anomaly patterns because:

1. The temporal evolution of our GCM runs are only constrained by SST and sea ice distributions, so can't be expected to match any particular pattern of temporal evolution,

2. Focus on scales resolved by all models precludes spatial anomaly evaluation on smaller scales and focus on CA precludes analysis on larger scales.

Evaluation of model response to climate forcing would be a better test of ability to predict climate change, but (as typical for climate studies) appropriate forcing response data is not available. Even though some of the issues uncovered in this study can be masked by bias correction, their analysis is useful because bias is the physical manifestation of errors in model physics, which means that a model with a bad mean state is unlikely to simulate climate change realistically. Additionally, the nonlinearity of atmospheric processes means that even a perfect model would get the wrong climate response if its initial state was inaccurate.

Experimental design and datasets used are explained in the next two sections. This is followed by the results section (broken into subsections dealing with mean bias, probability distributions, and variability) and followed up by conclusions.

2 Methodology

As noted in the introduction, evaluating whether downscaling actually improves upon resolved-scale GCM results requires comparison at a scale resolved by both the RCM and its driving model. We use regional averaging because it meets this requirement, allows for quick and easy comparison of data on differing grids, and reduces model noise.

There are also drawbacks associated with regional averaging. One downside to this approach is that it hides information on sub-regional spatial scales. An insidious example of this was found in C09, where GCM performance was found to best that of an RCM partially because GCM bias spread over a larger area which fell partly outside of the study area.

Analysis of the CA average is also potentially complicated by the fact that the factors controlling northern and southern CA climate are somewhat different. To the extent that GCM data can be trusted on smaller scales, it appears that GCMs and RCMs have similar dry biases in southern CA and differ mainly in performance in the north and central part of the state (not shown). For this reason and because southern CA precipitation is only a small contributor to the statewide total (e.g. C09), the results shown here can be thought of as dominated by contributions from the northern and central portion of the state.

Another challenge is deciding how to actually do the averaging. For grid cells contained entirely within the averaging region, this is straightforward. However, even at 50 km spacing only about 60% of the grid cells touching CA would fall into this category (Table 1). Thus it is clear that the utility of regional averaging depends on our ability to properly treat cells straddling the regional boundary. Proper treatment of boundary cells, however, is a philosophical issue in the sense that the averaging strategy of choice will depend on what information is assumed to be carried by model grid cells. In this study, the absence of an optimal averaging strategy is handled by applying several reasonable methodologies and using inter-method agreement as a measure of averaging uncertainty.

One approach is simply to compute the cell-area weighted average of all cells whose centers lie within the state. An illustration of this method (hereafter the simple method) is provided in Fig. 1. The simple method is attractive because it is easy to implement, but suffers from the flaw that a minute shift in cell position may determine the inclusion/exclusion of a cell. An approach that avoids this sensitivity is to weight boundary cells by the fraction of their area which is contained in CA. Computing fractional areas is challenging, however, particularly for a region as complicated as CA. A good approximation to this method that is much easier to implement is to regrid the data to very fine resolution, then to apply the simple method described above to the fine-scale data. If the regridding method conserves area averages, the resulting CA average differs only from fractional weighting through error induced by applying the simple method to the fine-resolution grid (which approaches zero as the fine-resolution grid spacing becomes small). We implement such a technique (hereafter the conservative method) using the regridding scheme of Jones (1999) and mapping all data to the uniform 1/4th degree grid used by the National Oceanographic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) Unified observations described later.

Conservative regridding is appropriate if model data is assumed to be uniformly distributed within each grid cell, but may give misleading results if the field of interest in actuality varies smoothly in space. In this case, a method that takes relationships between neighboring cells into account may be more appropriate. Bilinear interpolation is a simple method that does this. This technique is also of interest because it is much easier to implement than conservative regridding and is therefore more likely to be used. Including this method in our study allows us to test whether the complexity of conservative regridding is warranted.

3 Data

For this study, regional model data is taken from North American Regional Climate Change Assessment Program (NARCCAP) Experiment 1 output which is publicly available at http://narccap.ucar.edu/. This data consists of 6 hrly output for 1981-2004 from 6 different RCMs forced by sea surface and lateral boundary conditions supplied by the National Center for Environmental Prediction (NCEP) Reanalysis II (Kanamitsu et al., 2002). For GCM data, we use Atmospheric Model Intercomparison Project (AMIP) experiment data from the Coupled Model Intercomparison (CMIP3) archive, which is publicly available at http://www-pcmdi.llnl.gov/ipcc/about_ipcc.php. We use AMIP data because it is more directly comparable to observations, but recognize that these runs neglect the bias induced by ocean coupling. We use data from all 13 models which supply Pr from at least one AMIP realization at monthly resolution. In order to increase the range of resolutions explored, we also include the lone 50 km resolution GCM (hereafter GFDL Hi) included in the 1st NARCCAP experiment. Model details are included in Table 1.

Model performance is assessed by comparison against gridded observations. Unfortunately, precipitation observations are relatively uncertain (Nijssen et al., 2001; Groisman et al., 1996; Xie and Arkin, 1995). In an attempt to identify observational uncertainty, we include observational data from 6 different sources in this study. These include University of Washington (UW; Hamlet and Lettenmaier, 2005), NOAA CPC Unified (Unified; Higgins et al., 2000), Climatic Research Unit (CRU) version 2.1 (Mitchell and

Jones, 2005), University of Delaware version 1.02 (UDel), Global Precipitation Climatology Project (GPCP) version 2 (Gruber and Levizzani, 2008), and CPC Merged Analysis of Precipitation (CMAP; Xie and Arkin, 1997). References for each of these datasets are given in Table 2. It is worth noting that the UW dataset is scaled to match the Parameter-elevation Regressions on Independent Slopes Model (PRISM) model (Daly et al., 1994) in long-term mean; because of this, including PRISM in this study would be redundant. Since PRISM adjusts Pr based on topographic factors such as elevation, aspect, and slope, the UW data can be considered to have a more sophisticated treatment of mountainous terrain (which should result in higher Pr than predicted by simple interpolation). The raw data for many of these products overlap; differences between datasets could reflect variations in interpolation method as much as differences in raw data sources. The UW and Unified datasets use data from the National Climatic Data Center (NCDC) Cooperative Observer gauge data which has a station density of around 7000 daily reports over the US. Unified also includes CPC Cooperative stations and Higgins et al. (1996) data, which add a significant number of additional stations. CRU uses its own gauge dataset that contains around 8,300 monthly measurements worldwide. UDel combines Global Historical Climate Network and Legates and Willmott (1990) data for a total global station density of over 20,000. GPCP and CMAP both use Global Precipitation Climatology Center (GPCC) gauge data (6500-7000 stations globally) in combination with data from a variety of satellite platforms.

UW, Unified, CRU, UDel, and CMAP data do not include corrections for gauge undercatch, while GPCP correct following Legates (1987). This is an important consideration because high-elevation precipitation during winter is generally in the form of snow, which is underpredicted by gauges because it tends to to flow around sensors. This means that the Pr observations used here are likely to be underestimates. To our knowledge, no estimates of CA-area wintertime gauge bias exist, though Fig. 8 of Adam and Lettenmaier (2003) puts December-February zonal-average total undercatch error over land in CA latitudes at 15-25%.

Our comparison focuses on the period 1981 through 1998 because this is the only period for which data is available from all sources. We consider winter to consist of November through March (NDJFM) because this is the period of significant CA precipitation (C09 Fig. 3). CA averages for all observational datasets are created following the same methods as used for the models.

4 Results

4.1 Mean Precipitation

NDJFM Pr averaged over 1981-1998 is presented for each of the datasets in Fig. 2. In order to depict statistical significance graphically, values are given as bias relative to Unified (which has NDJFM 1981-1998 Pr of 3.0 mm day^{-1}). Errorbars are 95% confidence intervals computed using a 2-tailed t-test applied to the (annual-resolution) timeseries of the difference between model and Unified data. A dataset is statistically different from Unified if its confidence interval doesn't include the x-axis. While precipitation itself isn't appropriate for a t-test because it is zero-bounded and therefore non-normal, bias does not suffer from this problem and does follow a normal distribution (not shown). Each year is taken to be an independent sample because the 1-lag autocorrelation between years is less than 0.3 (generally quite a bit less) for all models while the threshold for statistical significance for 18 years of data is 0.4 (Zar, 1999).

UW, UDel, and CRU have small mean bias (Fig. 2; using them instead of Unified would have little effect on our results. GPCP and CMAP, on the other hand, yield substantially lower Pr estimates. This could be due to the GPCC gauge data used by both projects. This dataset contains fewer stations than used by CRU and many fewer stations than Unified UDel, and UW. This would cause a low bias if the omitted stations were predominantly in mountainous terrain, where climatological precipitation tends to be higher. The use of satellite data could also cause bias: Gruber and Levizzani (2008) note that passive microwave estimates sometimes fail to capture orographic enhancement, and that this error propagates into the GPCP (and presumably CMAP) final products. Because of these shortcomings, it seems likely that GPCP and CMAP estimates of CA Pr are too low.

The size of each confidence interval is related to the correlation between the dataset tested and Unified; UW, CRU, and UDel datasets have small intervals because they track Unified very well. The fact that these observational datasets are statistically different from Unified illustrates the important impact of differing approaches to selecting and processing station data. RCMs tend to have smaller confidence intervals than GCMs because they are forced by reanalysis, which ties them more strongly to the current climate. GCMs with multiple realizations are an exception to this rule. For these models, realizations are considered to be independent and statistics are computed on the time series of bias concatenated over all ensemble members. This approach is reasonable because the average pairwise correlation between realizations for a given model is less than 0.072 for all models except FGOALS. FGOALS runs are correlated at 0.34; its uncertainty is probably underestimated here. Low correlation between ensemble members (which implies that SST has little effect on simulated CA Pr) was also found in Phillips (2006).

One potential concern with this study is that the sampling period is relatively short and SST forcing leaves the GCMs only weakly constrained, so results may reflect natural variability more than model climatology. This is addressed by plotting individual ensemble-member values as black dots in Fig. 2. It is clear that in all cases the natural variability within an individual model is much smaller than the differences seen between models.

Each color in Fig. 2 indicates a different averaging technique. For UW, CRU, and UDel, only masked averaging was used because their native grids are already of comparable resolution to Unified. For grid spacing less than 3°, averaging technique does not have a strong impact on our conclusions. Note that this does not mean that models are actually resolving CA topography correctly, just that little error is induced by averaging. Because averaging technique does not make a difference, the rest of this study uses conservative regridding. The 3 coarsest models are omitted from further discussion because they are not adequately resolved.

An interesting result (which echoes the findings of C09 and other papers noted in Section 1) is that all RCMs except HadRM3 and CRCM significantly overpredict wintertime Pr. As noted earlier, observational undercatch error likely exaggerates (but is not wholly responsible for) the apparent wet bias. The source of bias is not immediately obvious and we leave its identification for future work. Consistency between RCMs is important because it suggests that the cause is fundamental to the dynamical downscaling approach rather than arising from the details of a particular code. It is also worth noting that spectral nudging used by CRCM and RSM does not seem to have a systematic effect on RCM bias - CRCM bias is smallest and RSM bias is among the largest.

GCMs, on the other hand, generally underpredict Pr (though some overpredict and a few have larger bias than any RCM). This result contradicts the findings of Coquard et al. (2004), who found all Coupled Model Intercomparison Project (CMIP) phase 2 models to overpredict wintertime Pr by more than 50% and Phillips and Gleckler (2006) who found substantial overprediction of west coast January Pr in the CMIP3 models. Difference between our study and theirs is seen here to result at least partly from use of CMAP and GPCP data as truth in the previous studies. Salathe et al. (2007) also found annual average Pr from CMIP3 models to be generally overpredicted using an earlier NCEP reanalysis as validation. Fig. 2 shows that reanalysis Pr is not necessarily a good surrogate for reality. Another possible reason for differences between our results and those of previous studies is that we use AMIP simulations, while previous work focused on coupled ocean-atmosphere GCMs. Lower Pr in AMIP runs is perhaps unsurprising since west-coast SSTs are consistently overpredicted in the CMIP3 archive (Solomon et al., 2007), which should cause excessive upstream evaporation and resultingly enhanced on-shore moisture flux.

Another interesting feature of Fig. 2 is that GCM bias does not seem to be related to model resolution. This suggests that insufficient resolution is not the leading source of model bias, implying that better parameterizations - not simply increased resolution - are required to improve climate predictions. It should, however, be noted that most of the GCMs considered here are too coarse to resolve CA's mountains. It is possible that resolution is important, but must be finer than some threshold to make a difference. In this context, it is interesting that the 50 km GFDL Hi model behaves very similarly to the RCMs. This suggests that perhaps resolution, not lateral boundary forcing, is responsible for elevated Pr in regional models.

Another key finding of this study is that RCM bias does not appear to be systematically smaller than that from GCMs. As noted in the introduction, this does not imply that downscaling is useless (since high-resolution output is itself very valuable), but it does suggest that the "upscale benefit" from more realistically simulating processes and terrain does not seem to be realized in CA. Identifying why increased resolution doesn't translate to better simulation would be a big step forward for regional climate modeling.

4.2 Precipitation Distributions

Fig. 3 shows the CA-average Pr exceedance probability distribution for each dataset available at daily resolution. Good agreement between UW and Unified suggests a close understanding of the true distribution, though it should be remembered that both datasets are based on very similar raw data and both are subject to the same systematic biases (such as undercatch).

It is interesting that all RCMs except CRCM overpredict heavy (>20 mm day^{-1}) Pr and 4 of 6 RCMs underpredict Pr frequency (days with Pr>0.1 mm day^{-1}). It is also worth noting that low mean Pr in HadRM3 appears to result from partial compensation between underprediction of weak events and overprediction of strong events, while CRCM does relatively well in the mean because it doesn't overpredict strong events, although it strongly overpredicts Pr frequency.

A problem with using CA averages to evaluate extreme Pr is that overprediction could be due to exaggerated storm spatial extent rather than overpredicted local intensity. Similarly, CA-average frequency bias could be driven by errors in the spatial distribution of rain rather than its frequency of occurrence. To clarify the source of bias, we plot in Fig. 4 the fraction of RCMs and GCMs overpredicting frequency or 99th percentile Pr as a function of location. In order to keep our analysis resolution-independent, these graphics were created by comparing each model against Unified data conservatively regridded to that model's grid. Composite maps were then created by conservatively regridding each model's bias map to the Unified grid and counting the number of models with positive bias for each resulting cell. GFDL Hi was omitted from this analysis due to technical problems. This graphic shows that almost all RCMs over-predict the magnitude of 99th percentile Pr events over most of CA, confirming our impression from Fig. 3 that overprediction of extreme wet events is the source of the bias found in Sect. 4.1. RCM Pr frequency is generally underpredicted; as in C09, compensation between frequency and intensity errors acts to reduce RCM mean-state Pr bias. Interestingly, RCMs seem to underpredict heavy Pr in the southwestern portion of the state. Replacing Unified data with UW observations reduces the areal extent but not the existence of this underpredicted region[1]. Reasons for this difference are unclear, but could be due to differences in topography or to greater tropical influence at lower latitudes. Underprediction of Southern CA mean Pr was also noted in Sect. 2.

Fig. 4 shows GCMs to be much less consistent in their biases than RCMs. This is also seen in Fig. 3, which shows that some GCMs yield very realistic probability distributions while others behave quite poorly. In general, model resolution does not appear to be a good predictor of GCM skill. The GFDL Hi model, however, again looks similar to the RCMs. This suggests that resolution rather than lateral boundary forcing may be responsible for RCM bias, and that the difference between 50 and 100 km resolution may be important even if resolution differences between coarser models don't appear to be.

Our GCM results contrast with previous studies, where strong rainfall was found to

[1]Using UW instead of Unified data has no qualitative effect on our RCM conclusions. We chose Unified data because lack of UW data just off the coast caused problems when regridding observations to coarse GCMs.

be too infrequent and light rainfall too prevalent (even using the same model versions considered here). Such a bias pattern is expected if high-resolution observations are compared against lower-resolution models, since averaging to coarser resolution tends to dilute maxima and minima. This dilution probably explains GISS ER bias in Dai (2006) since GISS ER at 4° x 5° resolution is compared to observations at 2.5° x 2.5° resolution, but fails to explain bias in the 3 other models considered in Dai, which have resolutions similar to the observations. Further, Sun et al. (2006) found little difference between the frequency and intensity[2] of light (1-10 mm day^{-1}) Pr at station, 1°, and 3° resolutions, and found the majority of models to overpredict light-Pr statistics even when compared to the 3° observations. Resolution was found to matter more for heavy (>10 mm day^{-1}) Pr intensity, but even compared to the 3° data, many models underpredicted heavy Pr. Results from Sun are not, however, directly comparable with the current study because Sun focused on global maps of June-August Pr and used color scales tuned to pick up global maxima/minima rather than midlatitude detail. Still, it seems clear that factors other than model/observation resolution discrepancies are playing a role in the difference between our results and those of previous studies. As noted earlier, one reasonable and testable hypothesis is that AMIP models are better than coupled ocean-atmosphere models at reproducing Pr statistics.

[2]intensity is defined as the average magnitude of rain events.

4.3 Variability

Temporal variability in the models is investigated in Fig. 5. This graphic shows the standard deviation of NDJFM-averaged Pr and (where available) the standard deviation of wintertime-only daily Pr. We focus on the variability of the CA average rather than the average of CA variability because the latter measure would be resolution dependent. Using only rainy days to compute the daily variance increases the HadRM3 value to 6 mm day^{-1} but otherwise does not significantly impact the results.

Observational estimates are again consistent, suggesting that we can say with some confidence whether model variability is too high or too low. HadRM3 and CRCM results look relatively good at both daily and annual timescales but the remaining 4 RCMs overpredict variance. This is perhaps not surprising since the models which overpredict variance also overpredict climatological Pr and especially the frequency of high rain rates. GCMs, on the other hand, generally underpredict variance at both daily and interannual timescales with higher resolution offering no improvement. This is consistent with the findings of Dai (2006). Daily variability in the GFDL Hi model is similar to the RCMs (as might be expected from the previous results shown here), but interannual variance is underpredicted, similar to the other GCMs.

5 Conclusions

In this study, we evaluate the effect of resolution on CA wintertime Pr as simulated by a variety of regional and global models. We note that resolution is expected to add value through more accurate spatial distribution and through more realistic physical representation of terrain and physical processes. We focus on this second benefit by evaluating averages over the state of CA. We find that the CA average is well resolved by all models with grid spacing finer than 3° in the sense that the CA mean for these models is essentially independent of averaging method. This does not mean that GCMs are able to resolve the terrain and processes important to CA regional climate, though we find little evidence that adding these details through finer resolution improves model performance at the CA-average scale. The fact that improved resolution doesn't translate to improved simulation is a key finding of this study. This result is somewhat surprising because Pr is strongly affected by topography, so increased the realism of mountain terrain should provide a huge advantage to high-resolution models.

Understanding and removing the source of bias at high resolution is critical for accurate regional climate prediction. While identifying the source of this bias is beyond the scope of this paper, we do offer some clues. Consistency between RCMs suggests that the source of bias is fundamental rather than tied to the particulars of a certain code. Further, wet bias seems to be associated with strong Pr events, while Pr frequency is generally underpredicted. The fact that the 50 km GFDL Hi GCM behaves similarly to the RCMs hints that resolution - not boundary forcing - is responsible for Pr bias. These features suggest that detailed analysis of a series of extreme-precipitation case studies at various resolutions would be a useful avenue of research.

Another important finding of this work is that GPCP, CMAP, and NCEP II show a dry bias in CA wintertime mean-Pr relative to the rest of the observational datasets. Previous studies concluding that GCMs overpredict Pr along the west coast of the U.S. were based on comparison against these datasets; based on the more accurate UW, Unified, CRU, and UDel datasets, the GCMs considered here actually tend to underpredict CA-mean precipitation. Additionally, we find no evidence of overpredicted rainfall frequency or underpredicted heavy precipitation in our global simulations (in contrast to previous work), though we do note that these simulations do underestimate daily and interannual Pr variability (which is consistent with previous work). Our differing conclusions may stem partially from careful use of resolution-independent metrics. This is unlikely to provide a complete explanation, however, and we hypothesize that use of specified-SST runs is also playing a role by removing warm SST biases offshore and hence reducing moisture flux into CA.

Finally, we note that model bias and intermodel agreement both provide a sense of the uncertainty inherent in Pr prediction from climate models. Based on the results presented here, we conclude that significant caution should be taken in interpreting model results for Pr.

Acknowledgments

We gratefully acknowledge NOAA's Earth Systems Research Laboratory Physical Science Division, the Surface Water Modeling group at UW, CRU at the University of East Anglia, and the GPCP and CMAP teams for providing observational data. In particular, Unified, CMAP, and UDel data were provided by NOAA/OAR/ESRL PSD in Boulder, Colorado, USA from their Web site at http://www.cdc.noaa.gov/. We would also like to thank the international modeling groups that provided their GCM data for analysis and the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for collecting and archiving that data. Thanks also go to the NARCCAP modeling groups and the NARCCAP program, which is funded by the National Science Foundation (NSF), the U.S. Department of Energy (DoE), NOAA, and the U.S. Environmental Protection Agency (EPA) Office of Research and Development. We also wish to thank Celine Bonfils, Karl Taylor, Ben Santer, and Dave Bader for useful discussions. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 as part of its Laboratory Directed Research and Development Program.

References

Adam, J. C. and D. P. Lettenmaier, 2003: Adjustment of global gridded precipitation for systematic bias. *J. Geophys. Res.*, **108**, 1–14, doi:10.1029/2002JD002499.

Brankovic, C. and D. Gregory, 2001: Impact of horizontal resolution on seasonal integrations. *Climate Dyn.*, **18**, 123–143.

Caldwell, P. M., H.-N. S. Chin, D. C. Bader, and G. Bala, 2009: Evaluation of a WRF-based dynamical downscaling simulation over California. *Climatic Change*, **95**, 499–521.

Christensen, O. B., J. H. Christensen, B. Machenhauer, and M. Boltzet, 1998: Very high-resolution regional climate simulations over Scandinavia - present climate. *J. Clim.*, **11**, 3204–3229.

Colle, B., C. F. Mass, and K. J. Westrick, 2000: MM5 precipitation verification over the Pacific Northwest during the 1997-99 cool seasons. *Wea. Forecasting*, **15**, 730–744.

Coquard, J., P. Duffy, K. Taylor, and J. Iorio, 2004: Present and future surface climate in the Western USA as simulated by 15 global climate models. *Climate Dyn.*, **23**, 455–472.

Dai, A., 2006: Precipitation characteristics in eighteen coupled climate models. *J. Climate*, **19**, 4605–4630.

Daly, C., R. Neilson, and D. Phillips, 1994: A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteor.*, **33**, 140–158.

Duffy, P. B., R. W. Arritt, J. Coquard, W. Gutowski, J. Han, J. Iorio, J. Kim, L.-R. Leung, J. Roads, and E. Zeledon, 2006: Simulations of present and future climates in the Western United States with four nested regional climate models. *J. Clim.*, **19**, 873–895.

Duffy, P. B., B. Govindasamy, J. P. Iorio, J. Milanovich, K. R. Sperber, K. E. Taylor, M. F. Wehner, and S. L. Thompson, 2003: High-resolution simulations of global climate, part 1: present climate. *Clim. Dyn.*, **21**, 371–390.

Gao, X., Y. Xu, Z. Zhao, J. S. Pal, and F. Giorgi, 2006: On the role of resolution and topography in the simulation of East Asia precipitation. *Theor. Appl. Climatol.*, **86**, 173–185.

Giorgi, F. and M. R. Marinucci, 1996: An investigation of the sensitivity of simulated precipitation to model resolution and its implications for climate studies. *Mon. Wea. Rev.*, **124**, 148–166.

Groisman, P., D. Easterling, R. Quayle, V. Golubev, A. Krenke, and A. Mikhailov, 1996: Reducing biases in estimates of precipitation over the United States: Phase 3 adjustments. *J. Geophys. Res.*, **101**, 7185–7196.

Gruber, A. and V. Levizzani, 2008: Assessment of global precipitation products. Technical Report WRCP-128, WMO/TD-NO. 1430, World Climate Research Program.

Hamlet, A. F. and D. P. Lettenmaier, 2005: Production of temporally consistent gridded precipitation and temperature fields for the continental United States. *J. Hydrometeor.*, **6**, 330–336.

Han, J. and J. Roads, 2004: U.S. climate sensitivity simulated with the NCEP regional spectral model. *Clim. Change*, **62**, 115–154, doi:10.1023/B:CLIM.0000013675.66917.15.

Higgins, R., J. E. Janowiak, and Y. Yao, 1996: A gridded hourly precipitation database for the United States. Atlas no 1, National Center for Environmental Prediction/Climate Prediction Center, Camp Springs, MD 20746, USA.

Higgins, R., W. Shi, E. Yarosh, and R. Joyce, 2000: Improved United States precipitation quality control system and analysis. Atlas no 7, National Center for Environmental Prediction/Climate Prediction Center, Camp Springs, MD 20746, USA.

Iorio, J. P., P. B. Duffy, B. Govindasamy, S. L. Thompson, M. Khairoutdinov, and D. Randall, 2004: Effects of model resolution and subgrid-scale physics on the simulation of precipitation in the continental United States. *Clim. Dyn.*, **23**, 243–258.

Jacob, D., L. Bärring, O. B. Christensen, J. H. Christensen, M. de Castro1, M. Déqué, F. Giorgi, S. Hagemann, M. Hirschi, R. Jones, E. Kjellström, G. Lenderink, B. Rockel, E. Sánchez, C. Schär, S. I. Seneviratne, S. Somot, A. van Ulden, and B. van den Hurk, 2007: An inter-comparison of regional climate models for Europe: model performance in present-day climate. *Clim. Change*, **81**, 31–52.

Jones, P. W., 1999: First- and second-order conservative remapping schemes in spherical coordinates. *Mon. Wea. Rev.*, **127**, 2204–2210.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP-DEO AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, 1631–1643.

Legates, D. R., 1987: A climatology of global precipitation. *Pub. in Climat.*, **40**, 85 pp.

Legates, D. R. and C. J. Willmott, 1990: Mean seasonal and spatial variability in gauge-corrected, global precipitation. *Int J. Climatol.*, **10**, 111–127.

Leung, L. R. and S. Ghan, 1995: A subgrid parameterization of orographic precipitation. *Theor. Appl. Climatol.*, **52**, 95–118.

Leung, L. R. and Y. Qian, 2003: The sensitivity of precipitation and snowpack simulations to model resolution via nesting in regions of complex terrain. *J. Hydromet.*, **4**, 1025–1043.

Leung, L. R., Y. Qian, J. Han, and J. O. Roads, 2003: Intercomparison of global reanalysis and regional simulations of cold season water budgets in the Western United States. *J. Hydromet.*, **4**, 1067–1087.

Mass, C., , D. Ovens, K. Westrick, and B. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? *Bull. Amer. Meteor. Soc.*, **83**, 407–430.

Mitchell, T. D. and P. D. Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. of Climatology*, **25**, 693–712.

Nijssen, B., G. M. O'Donnell, D. P. Lettenmaier, D. Lohmann, and E. F. Wood, 2001: Predicting the discharge of global rivers. *J. Climate*, **14**, 3307–3323.

Phillips, T. J., 2006: Reproducibility of seasonal land surface climate. *J. of Hydromet.*, **7**, 114–136.

Phillips, T. J. and P. J. Gleckler, 2006: Evaluation of continental precipitation in 20th century climate simulations: the utility of multimodel statistics. *Water Resources Research*, **42**, W03202, doi:10.1029/2005WR004313.

Pope, V. D. and R. A. Stratton, 2002: The processes governing horizontal resolution sensitivity in a climate model. *Clim. Dyn.*, **19**, 211–236.

Rasmussen, R., 2009: High resolution simulation of seasonal snowfall over Colorado and some impacts of climate change, submitted, J. Clim.

Rauscher, S. A., E. Coppola, C. Piani, and F. Giorgi, 2009: Resolution effects on regional climate model simulations of seasonal precipitation over europe, in press, Clim. Dyn. doi:10.1007/s00382-009-0607-7.

Rojas, M., 2006: Multiply nested regional climate simulation for southern South America: sensitivity to model resolution. *Mon. Weath. Rev.*, **134**, 2208–2223.

Salathe, E., P. W. Mote, and M. W. Wiley, 2007: Review of scenario selection and downscaling methods for the assessment of climate change impacts on hydrology in the United States Pacific Northwest. *Int. J. Climatol.*, **27**, 1611–1621.

Seth, A., S. A. Rauscher, S. J. Camargo, J.-H. Qian, and J. S. Pal, 2007: RegCM3 regional climatologies for South America using reanalysis and ECHAM global model driving fields. *Clim. Dyn.*, **28**, 461–480.

Shukla, J., R. Hagedorn, B. Hoskins, J. Kinter, J. Marotzke, M. Miller, T. Palmer, and J. Slingo, 2009: Revolution in climate prediction is both necessary and possible. A declaration at the World Modeling Summit for Climate Prediction. *Bull. Amer. Meteor. Soc.*, **90**, 175–178.

Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller, eds., 2007: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996 pp.

Staniforth, A., 1997: Regional modeling: A theoretical discussion. *Meteorol. Atmos. Phys.*, **63**, 15–29.

Sun, Y., S. Solomon, A. Dai, and R. W. Portman, 2006: How often does it rain. *J. Climate*, **19**, 916–934.

Sylla, M. B., E. Coppola, L. Mariotti, F. Giorgi, P. M. Ruti, A. Dell'Aquila, and X. Bi, 2009: Multiyear simulation of the African climate using a regional climate model (RegCM3) with the high resolution ERA-interim reanalysis, in press, Clim. Dyn. doi:10.1007/s00382-009-0613-9.

Tibaldi, S., T. N. Palmer, C. Brankovic, and U. Cubasch, 1990: Extended-range predictions with ECMWF models: Influence of horizontal resolution on systematic error and forecast skill. *Q.J.R. Meteorol. Soc*, **116**, 835–866.

Xie, P. and P. A. Arkin, 1995: An intercomparison of gauge observations and satellite estimates of monthly precipitation. *J. Appl. Meteor.*, **34**, 1143–1160.

—, 1997: Global precipitation: a 17-year montly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539–2558.

Zar, J. H., 1999: *Biostatistical Analysis*. Prentice Hall, Upper Saddle River, NJ, 4th edition.

List of Figures

Table 1: Data details. Acronyms used below: UQAM = Université du Québec à Montréal, UC = University of California, NW = Northwest, CCSM = Center for Climate System Research, NIES = National Institute for Environmental Studies, FRCGC = Frontier Research Center for Global Change, NCAR = National Center for Atmospheric Research, GFDL = Geophysical Fluid Dynamics Laboratory, IAP = Institute of Atmospheric Physics, MRI = Meteorological Research Institute, IPSL = Institut Pierre Simon Laplace, NASA = National Aeronautics and Space Administration, GISS = Goddard Institute for Space Studies, INM = Institute for Numerical Mathematics.

Type	Name	Center (Country)	Res (deg)	% boundry cells
RCMs	CRCM	Ouranos/UQAM (Canada)	0.36x0.45	37.7
	RSM	UC San Diego/Scripps (USA)	0.36x0.48	37.8
	HadRM3	Hadley Centre (UK)	0.42x0.52	41.0
	RegCM3	UC Santa Cruz (USA)	0.42x0.54	42.3
	WRF	Pacific NW National Lab (USA)	0.44x0.56	42.4
	MM5	Iowa State University (USA)	0.44x0.56	42.4
GCMs	GFDL Hi	GFDL (USA)	0.50x0.63	41.6
	MIROC Hi	CCSM/NIES/FRCGC (Japan)	1.12x1.13	72.0
	CCSM	NCAR (USA)	1.40x1.40	77.1
	HADGEM	Met Office (UK)	1.25x1.88	79.4
	BCC	Beijing Climate Center (China)	1.87x1.88	90.9
	ECHAM5	Max Plank Institute (Germany)	1.87x1.88	90.9
	GFDL	GFDL (USA)	2.02x2.50	93.8
	CNRM	Meteo France (France)	2.79x2.81	100
	FGOALS	IAP (China)	2.79x2.81	100
	MIROC Med	CCSM/NIES/FRCGC (Japan)	2.79x2.81	100
	MRI	MRI (Japan)	2.79x2.81	100
	IPSL	IPSL (France)	2.53x3.75	100
	GISS	NASA/GISS (USA)	4.00x5.00	100
	INM	INM (Russia)	4.00x5.00	100

Table 2: Observational datasets. WCRP = World Climate Research Program.

Name	Center (Country)	Res (deg)	Reference
UW	University of Washington (USA)	0.13x0.13	www.hydro.washington.edu/Lettenmaier/ Data/gridded/index_hamlet.html
Unified	NOAA (USA)	0.25x0.25	www.cdc.noaa.gov/cdc/data.unified.html
CRU	Climatic Research Unit (UK)	0.50x0.50	www.cru.uea.ac.uk/ timm/grid/ CRU_TS_2_1.html
UDel	University of Delaware (USA)	0.50x0.50	www.cdc.noaa.gov/data/gridded/ data.UDel_AirT_Precip.html
CMAP	Climate Prediction Center (USA)	2.50x2.50	www.cdc.noaa.gov/data/gridded/ data.cmap.html
GPCP	WCRP (international)	2.50x2.50	www.gewex.org/gpcp.html

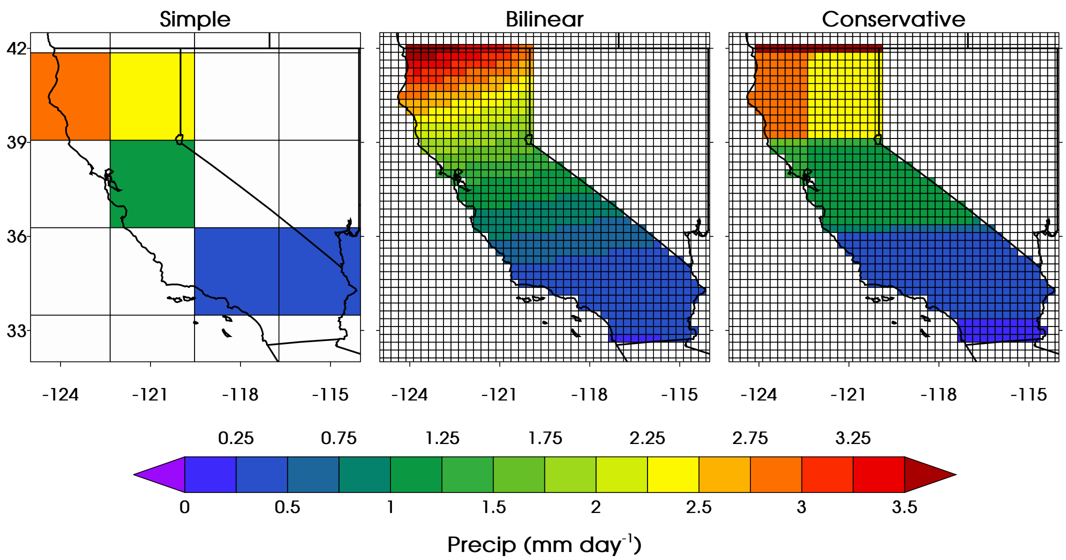

Figure 1: Illustration of our gridding methodologies using the CNRM model as an example. For each method, the CA average is the area-weighted average of all colored cells.

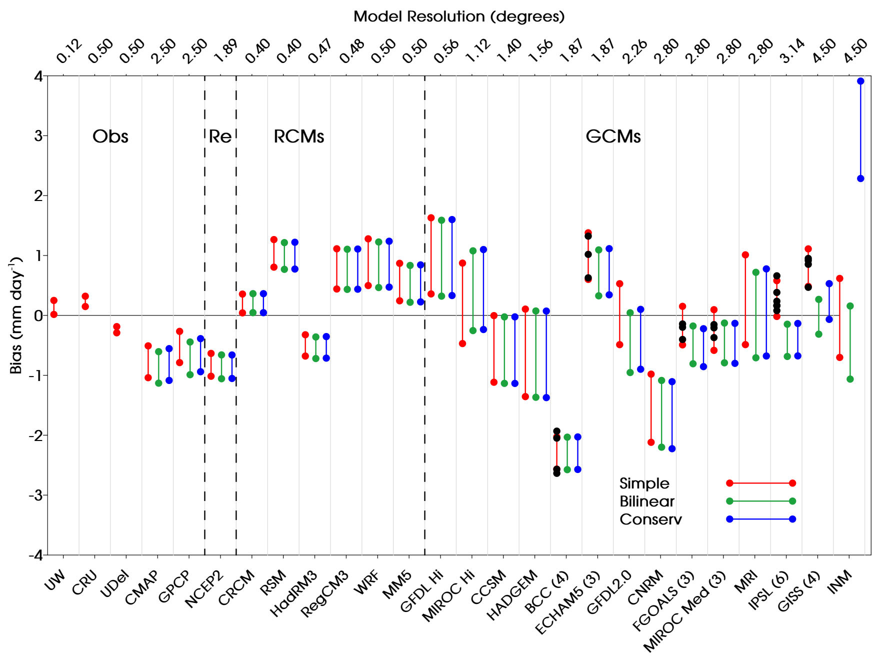

Figure 2: 95% confidence intervals for bias in NDJFM 1981-1998 precipitation averaged over CA using the methods from Fig. 1. In this graphic, Unified data is used as “truth” and datasets are separated by type (Obs=Observations, Re=Reanalysis, RCMs, GCMs). Within each type, datasets are arranged from lowest to highest resolution (resolutions are indicated at top). Averages over ensemble members are made where possible; in these cases number of members is indicated after model name and individual ensemble-member values are presented as black dots.

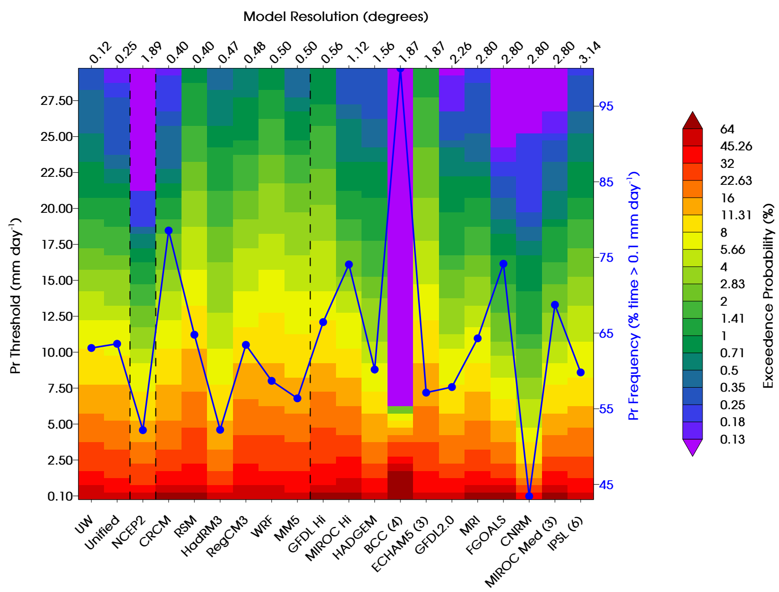

Figure 3: Cumulative probability distributions for conservatively-averaged daily Pr, separated by data type (note logarithmically-scaled color axis). Frequency of precipitation >0.1 mm day^{-1} is overplotted in blue with scale on right.

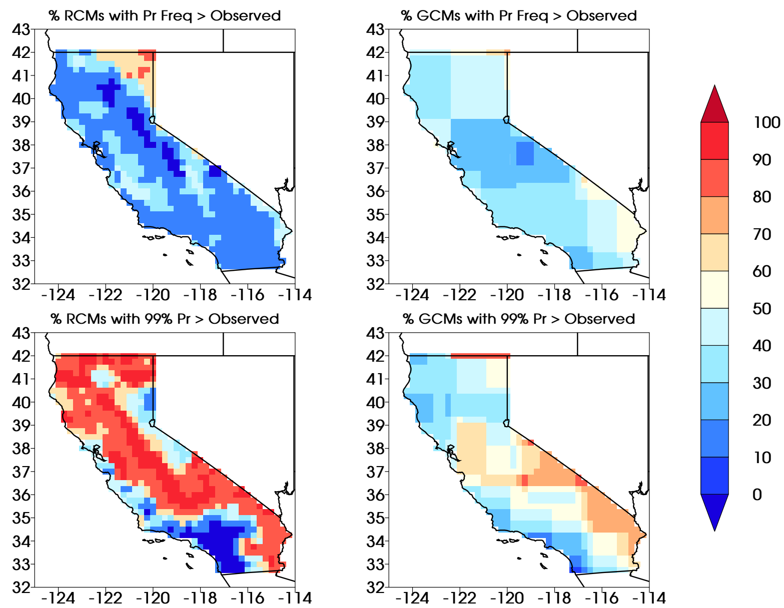

Figure 4: Top panels: percentage of models overpredicting Pr frequency (defined as Pr>0.1 mm day^{-1}). Bottom panels: 99th percentile Pr. Unified data is taken as truth. RCMs are compared in the left panels and GCMs in the right panels. See text for details.

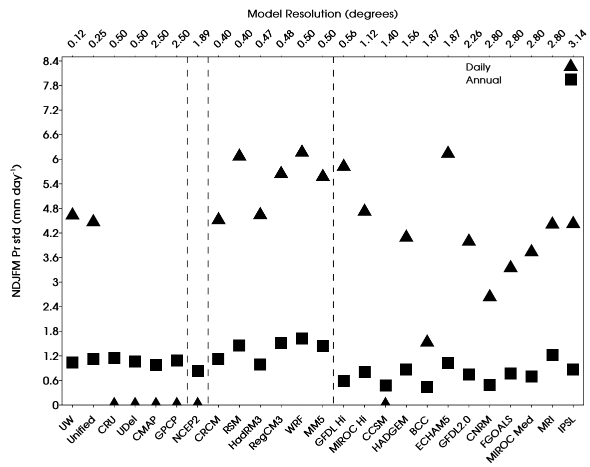

Figure 5: Pr standard deviation for NDJFM-averaged data (squares) and daily data within NDJFM (triangles). Daily values for models lacking daily-resolution data are mapped to zero for reference.