White Paper - Downscaling Using Dynamical and Statistical Methods

This white paper, describes development of downscaling and hydrologic modeling techniques applied to California weather, climate and hydrologic phenomena. ~ Submitted on behalf of Project Team

Additional submitted attachment is included below.

Downscaling Using Dynamical and Statistical Methods

White Paper

Dan Cayan, David Pierce, Laurel DeHaan, Janin Guzman-Morales, Alexander Gershunov, Rachel Clemesha, Andrew Martin, Paul Ullrich, Lele Shu, Hoori Ajami, Adam Schreiner-McGraw

Table of Contents

1. Introduction

Downscaled weather, climate and hydrology data from multiple climate scenarios are needed to translate global climate model (GCM) climate projections to spatial and temporal details that are relevant to decision makers—to understand historical variation and future projected changes and variation of temperature, precipitation, and other variables over California's complex terrain, from coast to interior.

Users must contend with a growing accumulation of global climate models—multiple models, most providing simulations for multiple future scenarios. This collection of models contains range of variability within and across models, and thus the uncertainty that arises from different model constructs, from different scenarios of future climate drivers (e.g. emissions, aerosols, land use, land cover, etc.), and from various forms of climate variability. Additionally, many GCMs provide an ensemble of model simulations for a given scenario, which are useful to disentangle the role of unpredictable and chaotic natural climate variations from trends produced by anthropogenic greenhouse gases and aerosols. To investigate the range of regional climate changes and impacts that are contained in the growing set of climate projections requires regional downscaling techniques that are on one hand sophisticated enough to represent complex regional climate structure, including extreme events, and on the other hand are sufficiently efficient so that the computational load is not prohibitive.

Ongoing research funded by CEC under the EPIC program (EPC-16-063) aims to develop downscaling techniques, exploiting hybrid dynamical-statistical elements, to provide improved resolution of California's highly varying (space and time) meteorological conditions including winds and coastal cloudiness and how they are affected by aerosols. Also included in this research is exploratory modeling of selected hydrological systems including surface and ground water systems.

This white paper, a report on work to-date undertaken under project EPC-16-063, will describe development of downscaling and hydrologic modeling techniques applied to California weather, climate and hydrologic phenomena. The following elements are included: statistical and dynamical downscaling of meteorological variables; hybrid techniques using both statistical and dynamical methods to obtain better downscaling results than would otherwise be possible, hydrological modeling including surface and ground water components, and ground water modeling of selected aquifers. Some complementary elements are also presented, which inform or stem from these downscaling and hydrologic modeling approaches. These include observed datasets, and an assessment of variables of interest to California decision makers that can be expected from dynamical and statistical downscaling. Finally, we report on "lessons learned", including opportunities and challenges.

2. Statistical Downscaling

Statistical downscaling exploits statistical relationships between local climate variables (e.g., precipitation, surface air temperature) and regional or large-scale predictors (e.g. general circulation modeled precipitation and temperature, or atmospheric pressure fields). By applying these statistical relationships to the output of global climate model simulations, more spatially-resolved local climate characteristics can be inferred. Further capabilities of the Localized Constructed Analogues (LOCA) technique are being developed under the present project to provide improved spatial and temporal resolution and new vector wind downscaling of California's highly varying (space and time) meteorological conditions. Furthermore, for example, California's investor-owned utilities (IOUs) often use only relatively few meteorological stations to characterize temperature variability in their service areas and would like the simulations of the historical period to be as close as possible (in the statistical sense) to the historical data for these stations. The need for fidelity with historical observations—including statistical properties of means, overall variation and properties of extreme events—must be considered in developing downscaled climate data.

Statistical downscaling is computationally efficient, but relies upon variability and patterns from a selected period of time, existing codes that have been developed for a small subset of climate variables, and an extensive training dataset. Statistical techniques operate under the assumption that some statistical properties of the training dataset are valid for the period (often future period) that is being downscaled. Which exact properties are assumed unchanging in the future depends on the method, and is usually referred to as the "stationarity assumption" of the method. The training datasets are usually a collection or rendition of historical observations, but could also be formed from dynamical model output from either a historical or a future period, an approach that is particularly useful when direct observations are lacking. Some well-used historical training datasets (see Section 6) have continuous gridded observations (or samples) and broad spatial coverage. These datasets have had monthly, daily and in some cases sub-daily time sampling.

2.1 Localized Constructed Analogs (LOCA) Downscaling

Localized constructed analogs (LOCA) is a statistical downscaling technique (Pierce et al. 2014; Pierce and Cayan 2015) that uses past history to add improved fine-scale detail to global climate models. David Pierce et al. at Scripps Institution of Oceanography, UCSD has developed and used LOCA to downscale 32 global climate models from the CMIP5 archive at a 1/16th degree spatial resolution (http://loca.ucsd.edu/). LOCA downscaling was implemented over various spatial domains, contingent on available historical training data for given variables of interest, ranging from the California region to most of North America from central Mexico through Southern Canada. For the California Fourth Climate Change Assessment generation of LOCA, the historical period was 1950-2005, along with two future scenarios: RCP 4.5 and RCP 8.5 over the period 2006-2100 (although some models stop in 2099). Variables produced for the California Fourth Climate Change Assessment were daily values of minimum and maximum temperature, precipitation, humidity, wind speed, and downward short wave radiation.

The LOCA method is a statistical scheme that produces downscaled estimates suitable for hydrological simulations using a multi-scale spatial matching scheme to pick appropriate analog

days from observations. First, a pool of candidate observed analog days is chosen by matching the model field to be downscaled to observed days over the region that is positively correlated with the point being downscaled, which leads to a natural independence of the downscaling results to the extent of the domain being downscaled. Then the one candidate analog day that best matches in the local area around the grid cell being downscaled is the single analog day used there. Most grid cells are downscaled using only the single locally selected analog day, but locations whose neighboring cells identify a different analog day use a weighted combination of the center and adjacent analog days to reduce edge discontinuities. By contrast, existing constructed analog methods typically use a weighted average of the same 30 analog days for the entire domain. By greatly reducing this averaging across analog cases, LOCA produces better estimates of extreme days, constructs a more realistic depiction of the spatial coherence of the downscaled field, and reduces the problem of producing too many light-precipitation days. The methods, algorithms, and validation of LOCA are available in a series of publications (Pierce et al. 2014, Pierce et al. 2015, Pierce and Cayan 2015, Pierce et al. 2018). Interested readers should consult those works for details of how the method works, which will not be repeated here.

In the EPC-16-063 research presented here, increased spatial and temporal resolution of LOCA downscaling is developed and tested, as an advance on previous work. Hourly temporal resolution applications are included, applied to temperature, humidity and winds. Increased spatial resolution, to 3km resolution from the former 6km resolution is developed. Downscaling of vector winds, not just wind speed, is also achieved. As described below, the introduction of fine scale dynamical model reanalysis results provides vital training datasets that enable these developments, resulting from the EPC-16-063 research effort.

Although the LOCA method is more computationally expensive than existing constructed analog techniques, LOCA is still practical for downscaling numerous climate model simulations with limited computational resources. Because LOCA is not resource limited, LOCA downscaling can afford to be conducted over many global climate models (GCMs), including outliers. Because of its high cost, WRF dynamical downscaling should be limited to the minimum expected to faithfully reproduce the range of expected climate variability. Analysis shows that 6 downscaled members should be considered a minimum to downscale for regional modeling purposes (Pierce et al. 2009), ideally selected from different GCMs to reduce the multi-model averaged error. Beyond this minimum, we consider that 10 downscaled members will create an ensemble whose mean and standard deviation may be expected to reasonably estimate the full suite of projections, if the GCM projections are chosen carefully.

2.2 LOCA: Sub-Daily Resolution – Hourly Temperature

Some applications require sub-daily information, such as hourly temperatures for energy applications, hourly winds for wildfire analysis, and hourly solar insolation for photovoltaic analyses. Up to now the California assessments have been focused on daily data, but in this recent work, conducted under EPC-16-063, we have been pushing the analysis to the sub-daily time resolution to address these issues.

There are two distinct approaches to sub-daily downscaling addressed here. For some historical reanalyses, such as MERRA-2 (Molod et al. 2015; Hinkelman 2019) and ERA5 (Hersbach et al.

2019), hourly data is already available, albeit on a coarse spatial grid (~30-50 km). Likewise, some meteorological station observations are available at hourly resolution. When available, this hourly data can be made use of directly. For example, in the historical hourly wind downscaling, each hour of the day is downscaled using the hourly reanalyses and the hourly training data (described in more detail below). In other situations, primarily future model projections, the climate models saved only daily data, so that observed hourly data has to be combined with the daily model projections to produce projections of hourly temperature. This methodology that will now be described in the context of future hourly temperature projections at selected meteorological stations of relevance to California energy applications.

Under the aforementioned ongoing CEC grant (EPC-16-063), the first thrust at sub-daily downscaling was hourly time resolution temperature, which enables investigation of energy-related temperature extremes. There are existing techniques to disaggregate daily Tmin and Tmax to hourly values, but they have important limitations. For example, a common approach is to fit a climatologically determined diurnal cycle to the Tmin and Tmax values and then take the hourly values from this fitted curve. (For some applications, such as agricultural degree days, a triangle is often fitted instead of a sinusoid.) The drawback of this approach is that it discards important aspects of sub-daily variability. I.e., two days can have very different progressions of hourly temperature even though they have the same Tmin and Tmax. Our development here was to devise a better method of disaggregating daily Tmin and Tmax to hourly values that retains realistic sub-daily variability, which can be of importance to energy industry stakeholders.

To retain realistic sub-daily variability, we disaggregated future model projections of daily Tmin and Tmax to an hourly time step at 29 meteorological stations using an analog day matching technique similar to what the LOCA downscaling approach uses. Besides having application to CEC and other energy utility concerns, this exercise provided opportunity to devise and test an hourly disaggregation technique that translates to other applications, e.g. hourly wind and humidity downscaling, which is described below in Section 4.1.

The approach to generating hourly temperatures given that day's Tmin and Tmax is as follows. First we construct the 3-day sequence of model Tmin and Tmax from the day before, the day of, and the day after the model day being disaggregated to hourly values. We then identify the best matching observed 3-day sequence of Tmin and Tmax in the training data set, subject to the constraint that the central matching analog day must be in a +/- 45 day-of-year window around the day-of-year being disaggregated. There are two reasons for using a 3 day sequence of Tmin-Tmax rather than matching only the day being disaggregated. 1) Matching on 6 values provides a better constraint than matching on 2 values. 2) The point of the exercise is to generate days with realistic transitions from one day to the next, which is better addressed by matching on the days before and after the day being disaggregated. A weighted RMSE was used to evaluate the quality of the match between the model Tmin/Tmax series and the historical observations. The weights are 1 for the central day, and 0.5 for the preceding and subsequent days. This emphasizes the match in the day being disaggregated, while still taking into consideration information from the previous and subsequent days.

We compared our analog day approach to the more traditional method of using climatological diurnal cycles as described previously, which we call the "climatological curve" approach. The

climatological diurnal cycles for this more traditional method were chosen based on station, month, and quartile of the diurnal temperature range for the day being disaggregated. Our intent is to examine whether the analog day approach captures observed sub-daily variability that the traditional climatological curve approach discards.

Example hourly sequences produced by the analog day method and the climatological curve method are shown in Figure 1A. By construction, the climatological curve approach (dotted red line) generates a smooth solution with very little variability in the afternoon and nighttime hours, which is unlike what is seen in the observations (solid black line). The analog matching approach (broken blue line) does capture realistic variability at all hours of the day.

Figure 1: Examples of hourly temperature disaggregation at Bakersfield, CA.
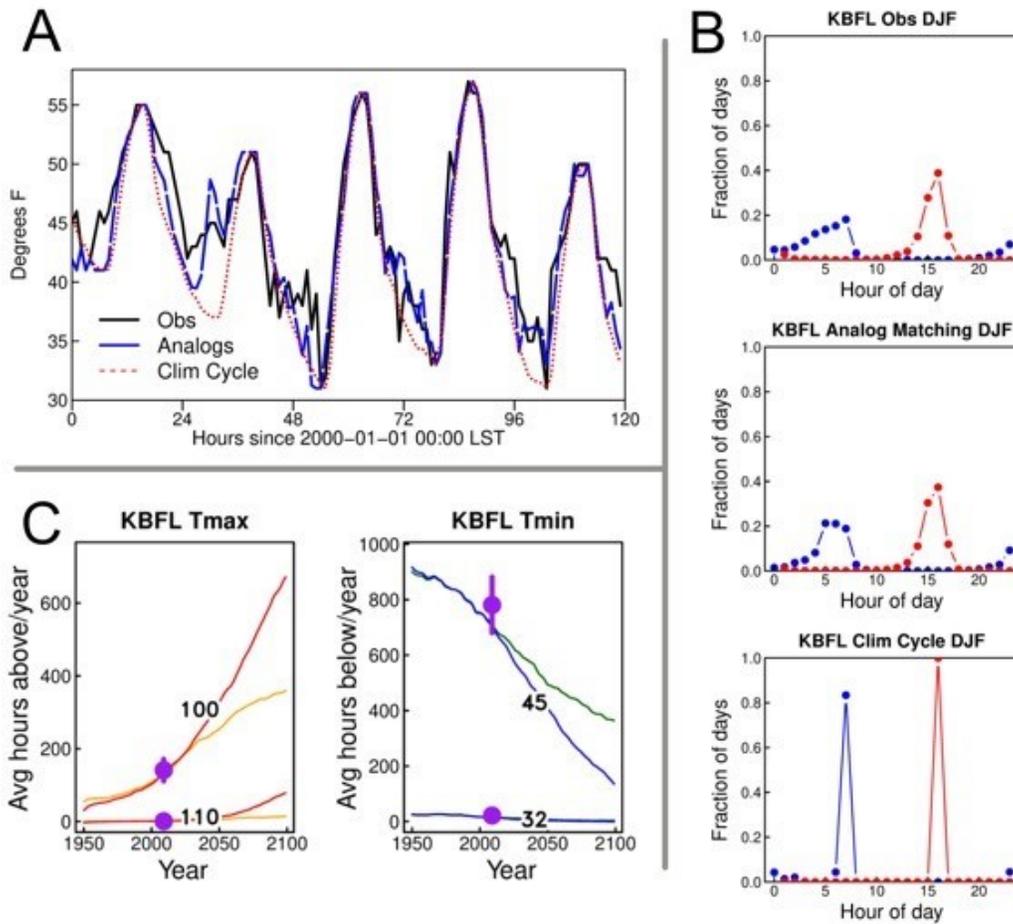
Time series of the original observed hourly temperature values (black), the analog day disaggregation (blue), and the disaggregation based on fitting climatological diurnal cycles (dotted red). B: Histograms of the fraction of time that each hour is the warmest (red) or coldest (blue) hour of the day, in winter (Dec-Jan-Feb), for the observations (top), analog day disaggregation (middle), and climatological diurnal cycle (bottom) methods. C: Multi-model ensemble average projected change in the number of hours either above (red and orange lines) or below (blue and green lines) the indicated temperature threshold in degrees F. The red and blue lines are for RCP 8.5; the orange and green lines are for RCP 4.5. The purple dot and whisker show the observed mean value and 95% confidence interval from 2000-2018.

A comparison between the original hourly observations and cross-validated disaggregations from the analog matching technique and climatological curve technique is shown in Figure 1B. As expected from the way the methods are constructed, an analog approach captures the observed variability extremely well. Histograms of the fraction of time that each hour is the warmest (red) or coldest (blue) hour of the day, in winter (Dec-Jan-Feb), for the observations (top), analog day disaggregation (middle), and climatological diurnal cycle (bottom) methods. In comparison to the analogue approach, the climatological curve approach severely underestimates variability on time scales shorter than a day, reaching a deficit in variance spectral power of nearly two orders of magnitude at the highest frequencies. Based on these and other analyses presented in the full CEC report on the hourly temperature downscaling (Pierce and Cayan, 2019), we conclude that the analog day matching approach to disaggregating daily Tmin/Tmax to hourly values does a much better job of capturing sub-daily variability than does the more traditional approaches. Application of the analogue method to projected climate scenarios is shown in Figure 1C, which illustrates the multi-model. Multi-model ensemble average projected change in the number of hours either above (red and orange lines) or below (blue and green lines) the indicated temperature threshold in degrees F. The red and blue lines are for RCP 8.5; the orange and green lines are for RCP 4.5. The purple dot and whisker show the observed mean value and 95% confidence interval from 2000-2018. These plots clearly indicate the simulated rise in frequency of "hot" hours, the decrease in the frequency of "cold" hours, and the marked divergence, in about 2040, of the upward (hot hour) trajectory or downward (cold hour) trajectory of the RCP8.5 scenario relative to that of the RCP4.5 scenario. Overall, the hourly disaggregation technique developed under the EPC-16-063 program achieves its goals of generating hourly future projected temperature values that match observations over the historical period, correctly replicate the global climate model projected trends, and preserve realistic variability on sub-daily time scales.

3. Dynamical Reanalysis

3.1 Global and Regional Reanalyses—Drivers of Fine Scale Models

Historical reanalysis using a dynamical model offers a way to develop a more complete, dynamically consistent set of observed records needed to investigate mechanisms controlling various forms of variability, including extreme events. Global and regional reanalyses results, which capture the larger scale atmospheric environment, are used as input to finer scale (downscaled) dynamical and statistical models. Recent reanalyses (e.g. ERA5 global (hourly, ~31km spatial resolution, 1979--present) from the European Center for Medium Range Forecasting, MERRA-2 global (hourly, $0.5° \times 0.625°$ grid, 1980-present) from the NASA Goddard Space Flight Center) and NARR (Mesinger et al. 2006) regional over CONUS (3-hourly, 32km, 1979-present) from NOAA NCEP are available over the relatively modern satellite era. Although spatial resolution provided by the two global reanalyses has become finer than was available just a few years ago, the 30-50km gridding is still too coarse to capture important structure over California's complex coast and interior landscape. In this research we conducted dynamical downscaling experiments to replicate historical climate conditions aimed at simulating coastal low cloud (CLC) including marine stratocumulus clouds (MSc), stratus and fog, along and over the California coast, along with wind and humidity over the California region.

3.2 Dynamical Model Downscaling

Dynamical downscaling involves running high-resolution climate models on regional domain (usually a limited sub-domain). Regional dynamical models typically employ low-resolution climate model output as boundary conditions. These models use physical principles to reproduce local climates, but are computationally intensive. Often, regional dynamical models are customized to better replicate certain physical variables (e.g. coastal stratus clouds) by tailoring model structure and physical parameterizations. Like all methods, dynamical model output contains biases, which often are more than minor incremental offsets. Spatial resolution can be refined by nesting successively smaller domains inside larger coarser domain simulations. Regional dynamical models are generally more computationally expensive than statistical methods, but produce a full set of atmospheric variables that are dynamically consistent and can be sampled over high temporal resolution. Dynamical models generate a high volume dataset, they are usually confined to a limited set of climate model scenarios and sometimes a restricted period of time.

Regional dynamical downscaling resulting from a "one-size-fits-all" model configuration may have limitations. Although the Weather Research and Forecasting (WRF) mesoscale numerical weather model (Skamarock et al. 2008) has become a commonly deployed dynamical downscaling tool, each WRF application must use a tailored configuration to reproduce target processes with the best accuracy possible. There are well-documented WRF configurations for hurricanes, severe convection, fire weather, air pollution meteorology, wind resource forecasting, and atmospheric rivers. Each application requires trade-offs in model parameters that reduce biases in the targeted variables but can increase biases in others. In this case, we conduct dynamical downscaling experiments to replicate historical climate conditions using a configuration of WRF that is aimed at simulating marine stratocumulus clouds (MSc) along and over the California coast. We call this model "WRF-CA-CLC", where the acronym CLC stands for "Coastal Low Clouds". We conducted experiments to test WRF-CA-CLC configurations and their sub-grid physical process models, including different boundary layer configurations (mixing/advection, convection, and model vertical resolution), and how they affect cloudiness and, secondarily, winds. These experiments and the configuration parameters tested can be found in Table 1. Key findings that emerged are described in Sections 3.2 - 3.4.

3.3 Inland Penetration of Marine Stratocumulus in WRF

MSc cover large areas of the North Pacific Ocean offshore CA during the summer months, but the extent of their inland migration is both not-well-characterized and critical for understanding energy supply and demand. Figure 2 shows the intra-seasonal extent of albedo by GOES satellite observations and WRF-CA-CLC v0 (see Table 1) simulations. Albedo is a measure of the reflection of sunlight by the combined effects of the atmosphere and the earth's surface. For the region and season studied here, albedo is driven primarily by the presence of bright MSc over dark surfaces. Figure 2 shows the mean coverage of clouds, but also displays the daily variability of cloud coverage extent through the coefficient of variability. While WRF-CA-CLC is able to simulate inland penetrating clouds, this analysis indicates fine grain improvements to target in further model development. Such improvements may be critical for understanding fluctuating patterns in energy demand and anticipating solar energy resources on a regional scale (this

analysis was performed by Dr. Martin's undergraduate intern Alexis Harris, who participated in the Portland State University (PSU) Center for Climate and Aerosol Research (CCAR) REU program during June – Aug, 2019; see http://web.pdx.edu/~anmarti2/Talks.html).

**Figure 2. Mean daily summer season (May – Sept) albedo (%)**
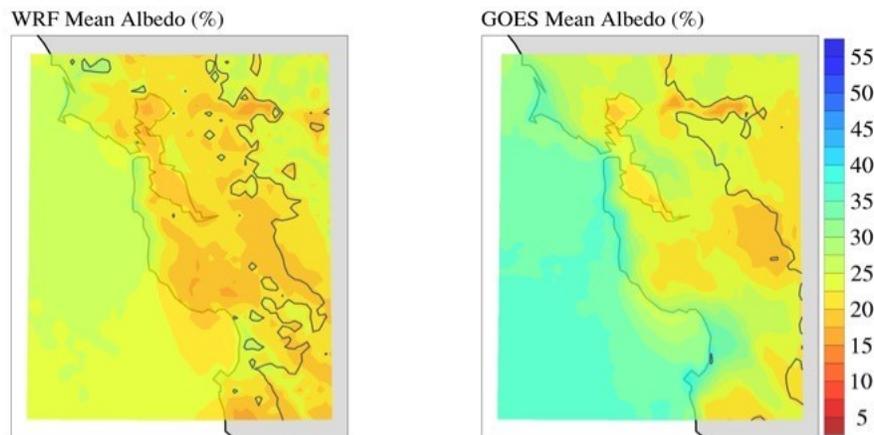
From Left) WRF-CA-MSc v0 and Right) GOES. Black line indicates where coefficient of variability (σ/μ) exceeds 1.0.

3.4   Intra-seasonal Variability of Saturation Simulated in the Coastal Boundary Layer

The mechanisms responsible for moistening the subtropical marine boundary layer during extensive MSc episodes are well-documented (Wood 2012; Clemesha et al. 2017). Less well-documented is the vertically resolved intra-seasonal variability in boundary layer, moisture and clouds. Figure 3 shows an analysis emerging from the active period created by comparing WRF-CA-CLC simulations to daily balloon-borne soundings, collected by National Weather Service observers and obtained from the University of Wyoming at http://weather.uwyo.edu/upperair/sounding.html. The analysis period was the typical MSc season, May – Sept, 2010. Both simulations and sounding observations are valid for cloudy mornings near 00 UTC and drawn from three CA locations near the coast: Oakland, Miramar, and Vandenburg Air Force Base. Three configurations of WRF-CA-CLC from Table 1 are shown, to investigate effects of different model boundary layer parameterizations on clouds. The center panel in Figure 3 shows that all WRF-CA-CLC versions overestimate the daily variability of saturation in the lowest atmospheric layers but underestimate saturation near the surface at pressures below 950 hPa. Crucially, these model disparities occur near the average top of the cloud layer, meaning all configurations regularly produce much too dry boundary layers below the marine stratocumulus cloud deck. The far-right panel shows that in these critical layers with pressure greater than 950 hPa, The total water mixing ratio is biased low in all versions of the model. In the far-left panel, the cloud water mixing ratio variability shows that model configurations with the Total Energy – Mass Flux (TEMF) boundary layer often produce much more cloud in lower atmospheric layers, despite the low bias in total water content.

Figure 3: Profiles of Observed and WRF-CA-CLC Simulated Cloud Properties
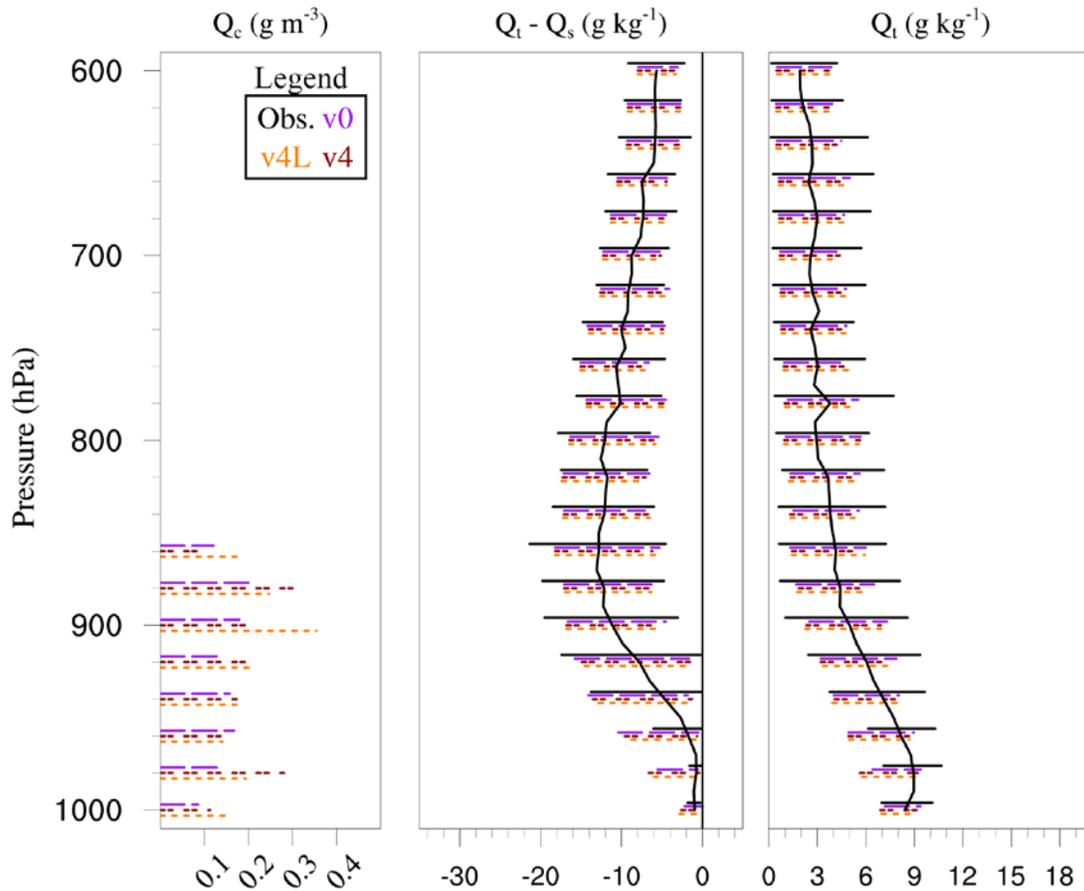
Morning (12 UTC) lower tropospheric profiles during summer 2010 simulations of Left) cloud liquid water (g m^-3) from WRF-CA-CLC versions v0, v4, v4L for all coastal national weather service sounding sites (NKX, VBG, OAK) where GOES albedo at 15 UTC indicated cloudy conditions. Horizontal lines drawn between lower/upper 10% values. Center) As in left panel, except profiles of saturation deficit: $Q_t - Q_s$ (g kg^{-1}). Observed values calculated from NWS sounding temperature and dewpoint. Bold black line indicates observational mean profile. Thin black vertical line indicates saturation. Where the quantity approaches saturation, there is water available to condense clouds. Right) As in center panel, but for total water mixing ratio (Q_t - g kg^{-1}). A threshold albedo equal 0.3 was used to detect cloudy conditions. Approximately 220 observations is maximum number in any vertical layer.

3.5 Simulated MSc are Limited by Boundary Layer Turbulent Mass Fluxes

Figure 4 shows the accuracy of MSc presence by two configurations of WRF-CA-CLC by comparison to GOES satellite observations of albedo. In the mean, v4, a version of WRF-CA-CLC including the TEMF boundary layer model, is far more accurate than v0, the WRF-CA-CLC baseline version. This is true for all configurations using the TEMF boundary layer model compared to all other versions. The TEMF model allows more vigorous development of turbulent mass fluxes, especially near boundary layer top where the other boundary layer models tested do

9

not allow prognostic treatment of turbulent mass fluxes. Combined with the above result, this indicates that MSc in WRF-CA-CLC simulations are limited by turbulent mass fluxes in marginal saturation environments. This result, along with the intra-seasonal variability result above, are being developed for a peer-reviewed manuscript (Martin et al., *in-prep.*).

Figure 4: July 2010 mean albedo at 15Z (7 PST) from WRF v0, and WRF v4 vs GOES observations.
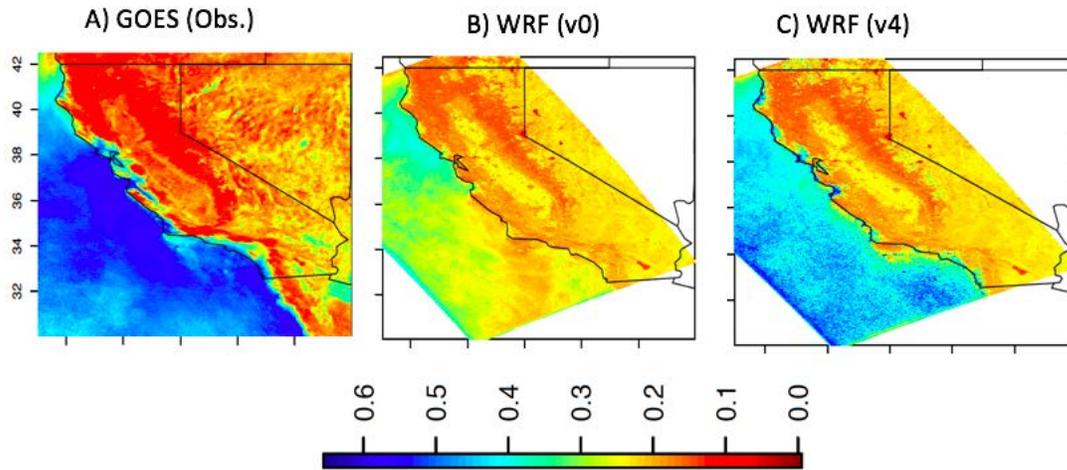
These different parameterizations had significant influence on marine cloudiness, and ultimately the model version (v4 in our tests, Table 1) was selected on the basis of producing greater marine cloudiness than the others, which were strongly biased towards unrealistically clear conditions.

Table 1: WRF-CA-CLC Sensitivity tests.

Perturbation Test name	BL Variables	Mix/Advect	Shallow Convection
v0 (Baseline)	TKE	Off	Off
v1	TKE	Off	On
v2	TKE	On	Off
v3	TKE	On	On
v4	TEMF	Off	Off
v5	TEMF	On	Off

Columns refer to parameterized processes a,b,c where the prognostic BL Variables turbulent kinetic energy (TKE) and total energy and mass flux (TEMF) are used, Mixing and advection (Mix/Advect) of cloud hydrometeors in the boundary layer, and shallow convection and turned on or off. Green highlighting shows tests that were performed for May – Nov, 2010.

Turning to the characteristics of modeled near-surface winds, the variants of the model produced relatively small differences in winds, as compared to those observed from a relatively dense network of weather stations deployed over San Diego County by San Diego Gas and Electric (examples shown below in Figures 10 and 11). And, modeled winds can be adjusted via bias correction, but how well this performs and the degree to which this can be applied uniformly over

a broad region are still open questions, largely owing to temporally-limited, spatially spotty, and poor quality wind observations.

Additionally, a crucial factor in the degree of fidelity of the dynamical downscaled historical results is the choice of the large scale model that is used as boundary conditions for the fine scale WRF simulation. A comparison between downwelling solar radiation from different atmospheric reanalyses (Figure 5) makes this point—the ERA5 reanalysis produces patterns of marine and coastal cloud that are far superior to those from the NARR and MERRA-2 Reanalyses, as gaged by their comparison to observed cloud albedo from GOES data.

Figure 5: Daily averaged surface downwelling solar radiation (W/m²) over the West coast of the U.S. from (left to right) GOES satellite observations, MERRA2 reanalysis, ERA5 reanalysis, and the NARR reanalysis.
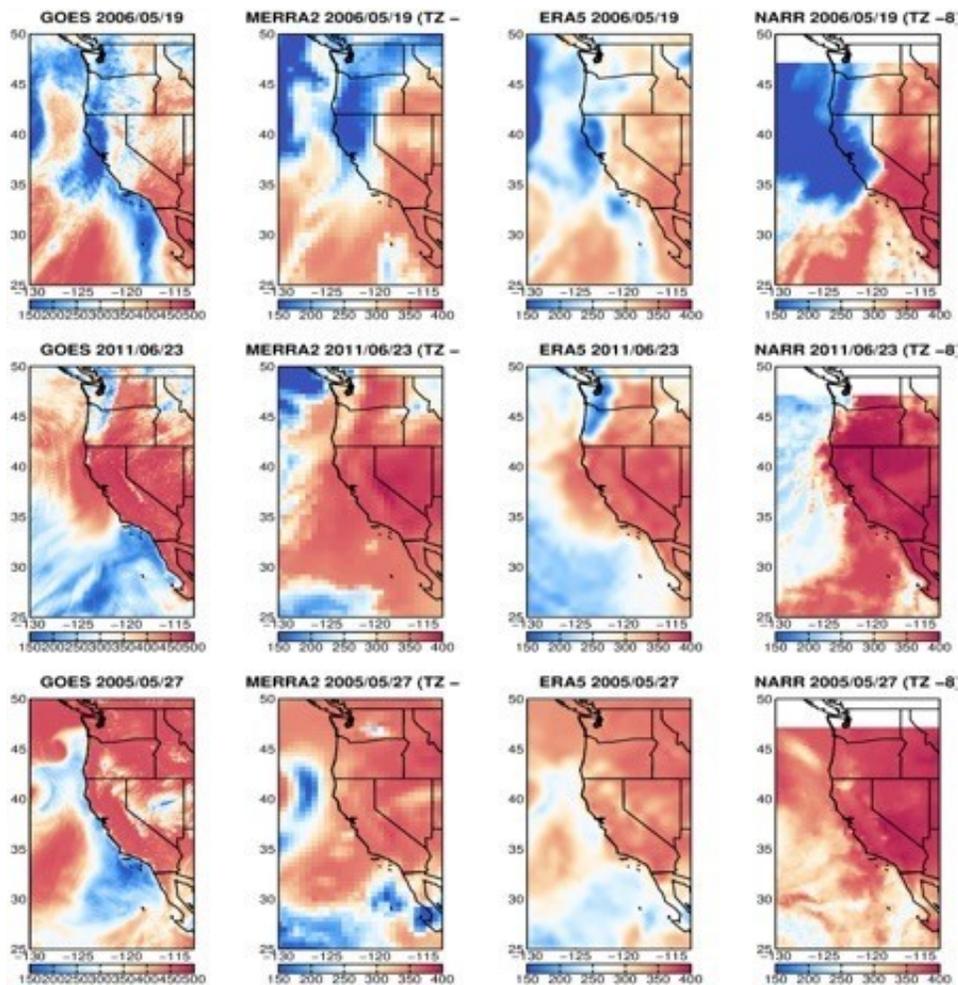
Each row shows one day, as per the panel title. Days are selected based upon having a strong land/sea contrast in values along the coast of Southern California. So, for instance, blue areas indicate heavy marine stratus clouds (low surface solar radiation), red colors indicate clear conditions (high surface solar radiation).

Daily averaged surface downwelling solar radiation (W/m^2) over the West coast of the U.S. from (left to right) GOES satellite observations, MERRA2 reanalysis, ERA5 reanalysis, and the NARR reanalysis. Each row shows one day, as per the panel title. Days are selected based upon having a strong land/sea contrast in values along the coast of Southern California. So, for instance, blue areas indicate heavy marine stratus clouds (low surface solar radiation), red colors indicate clear conditions (high surface solar radiation). Here we note that ERA5 is a very new (~mid-2019) dataset, and dynamical downscaling using ERA5 as WRF boundary conditions is still nascent, both to the community and to this project. Future dynamical downscaling of MSc over coastal regions of CA may benefit from using ERA5 as boundary conditions, rather than NARR (current). Further tests to verify this are proceeding.

4. Hybrid Dynamical-Statistical Downscaling

Hybrid downscaling exploits the ability and output of dynamical downscaling in capturing fine scale atmospheric features, combining a dynamical model's high spatial and temporal resolution datasets with the bias correction and computational savings of a statistical model. This technique can be applied to downscaling multiple GCMs and also to 'backcast' historical climate using a global atmospheric model reanalysis. Importantly, the efficiency of the hybrid technique allows us to downscale multiple emission scenarios and multiple time periods. For example, the ~40-yr, 3 km spatial resolution vector wind dataset constructed here via hybrid downscaling techniques is generated using a 15-yr high resolution WRF model run as training data. Simply extending the WRF model run to 40 years rather than using the hybrid approach would have been prohibitively expensive.

4.1 Application of Hybrid Downscaling to Hourly Wind and Humidity

Following the temperature downscaling described above in section 2.2, the next cut at sub-daily resolution downscaling is vector winds and humidity at hourly resolution. Previously, wind speed, but not wind direction was downscaled using the LOCA technique. Vector wind downscaling required an extension of LOCA downscaling to consider vector fields along with an auxiliary variable, sea level pressure (SLP), used because of SLP's key role in representing regional circulation patterns and in generating pressure gradients that drive wind. Specifically, U, V, and SLP are jointly downscaled by LOCA, then the wind speed from the downscaled U and V fields is computed and biased corrected to the training data's wind speed field. This approach helps preserve the correct relationships between U and V in the downscaled data while still having wind speed distributions that agree with the training data.

Figure 6: Climatology of annual mean wind speed from several reanalyses observational datasets indicates the broad range of magnitudes, somewhat disparate spatial patterns exists between different products.
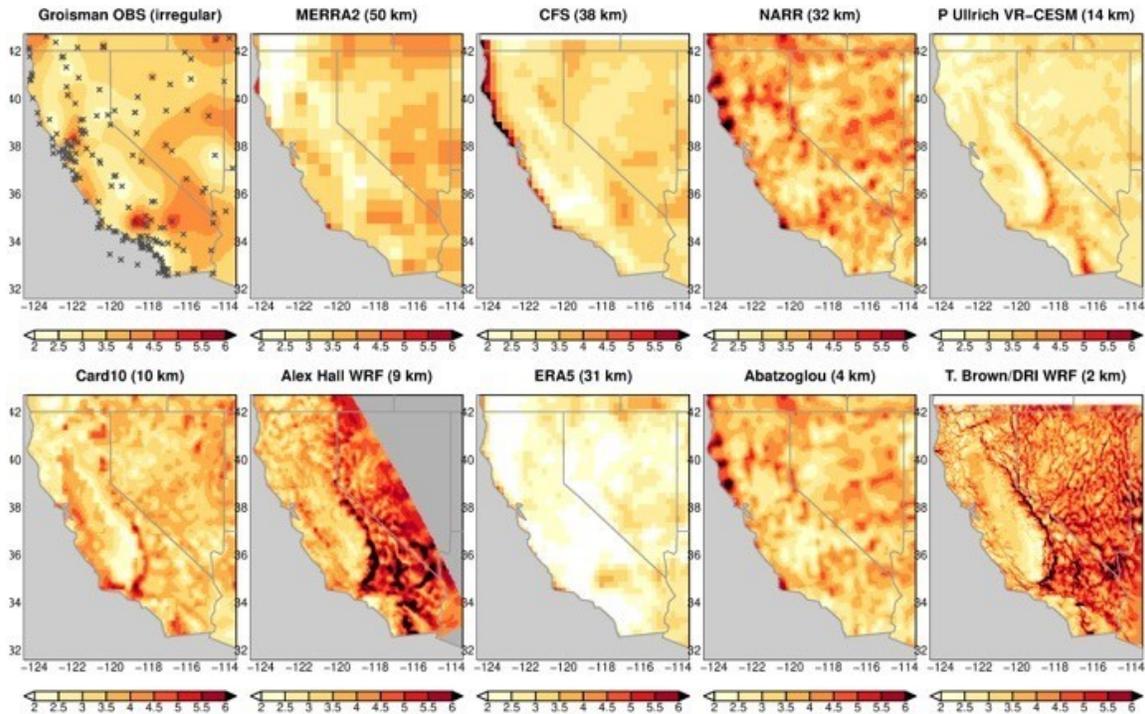
As with the temperature downscaling described above, the LOCA technique is extended to hourly time resolution by virtue of the availability of hourly training data from the dynamical model and global reanalysis being downscaled. Surface and near-surface winds are only sparsely measured, and those locations where winds are measured usually provide only limited time histories which are often contaminated with spurious measurement errors, which reinforces the need for regional modeled wind products (Guzman-Morales et al. 2016). Our understanding of wind structure over California and surrounding regions is muddied by lack of quality surface observations, and differences that are presented by different model results. For example, the climatology of annual mean wind speed from several reanalyses and observational datasets (Figure 6) indicates the broad range of magnitudes and disparate spatial patterns exists between different products. Here, for a training dataset for the LOCA statistical downscaling scheme, we employ a regional high spatial resolution wind record from a fine scale atmospheric model, making this a hybrid downscaling approach. Here we employ the numerical model wind dataset from a WRF regional atmospheric model run by Tim Brown and colleagues at the Desert Research Institute (DRI) which has strong correspondence to topographic features (Figure 6, lower rhs) and is being used by the wildfire prevention and management community (Brown et. al 2016; Sapsis et al.. 2016). The DRI-WRF product is analyzed onto a 3km spatial grid, and provides hourly temporal resolution. The DRI-WRF training data was bias corrected by the DRI-WRF providers to available RAWS observations by computing errors between the observed and model winds at grid cells that contain RAWS stations, then interpolating the errors between station locations using an inverse distance weighting approach. Errors are computed separately for each season and hour of day to account for the

seasonal and diurnal variability of wind and its possible misrepresentation in the WRF model output.

Figure 7: LOCA downscaled winds and relative humidity for three Santa Ana events.
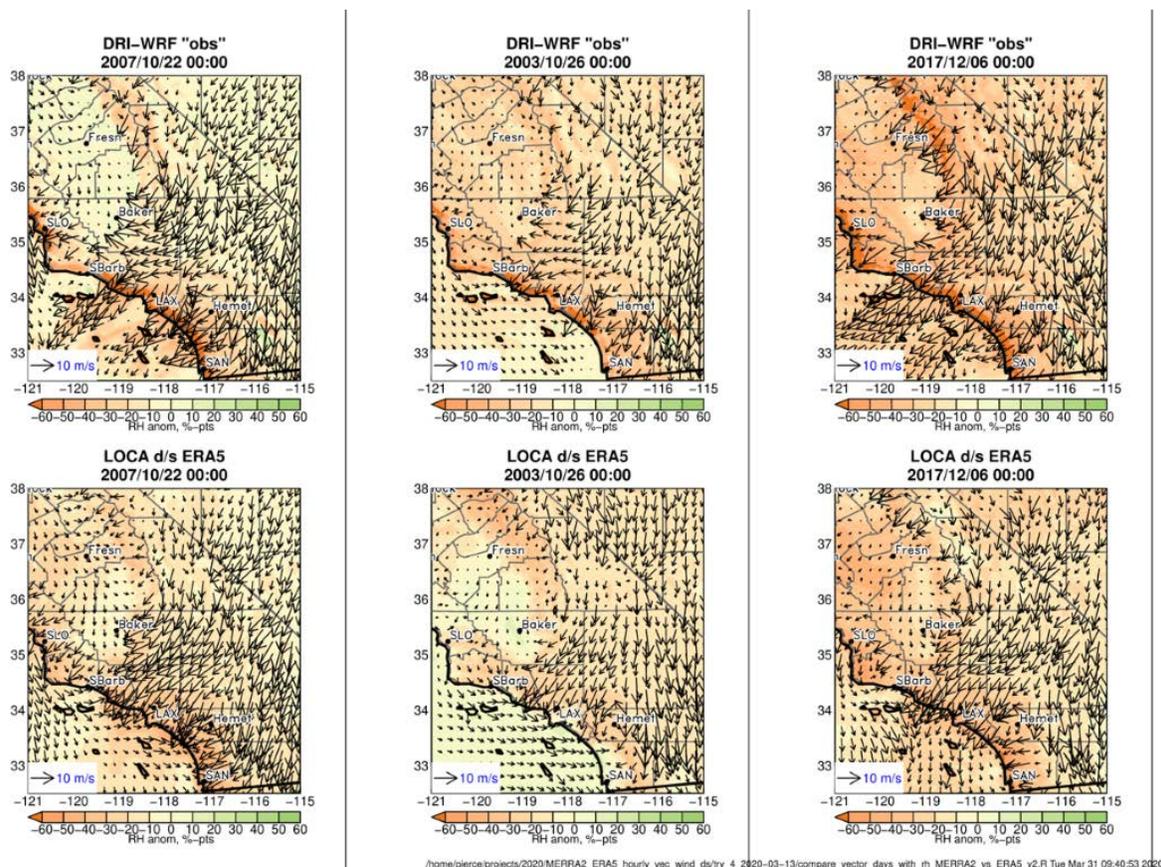
A comparison of winds and relative humidity (RH) in individual strong Santa Ana events in 10/22/2007, 10/26/2003 and 12/06/2017 at 0Z universal time produced by LOCA downscaled ERA5 reanalysis with that from DRI-WRF. Downscaled results determined from portions of training dataset that do not include these cases. Colors show RH anomaly (%) from long term average, while arrows show the vector wind speed (m/s) and direction.

In downscaling the vector winds with LOCA (Figures 7-9) it was found that a full multivariate downscaling process including vector winds, sea level pressure, and relative humidity simultaneously in the analog selection process did not represent the relative humidity fields as well as when relative humidity was downscaled individually. Adding more predictor fields to the LOCA process makes it progressively harder for closely matching analog days to be found, so there is a tradeoff between how many predictor fields are selected and the quality of the final downscaled field. We therefore changed the process so that SLP and vector wind are downscaled together (SLP being retained because of the important role of pressure gradients in determining the wind field), but relative humidity is downscaled separately. Note that the fields are all still connected through the physics of the driving GCM (both MERRA-2 and ERA5 were downscaled over their available time period, beginning 1980 and 1979, respectively). The temporal

14

consistency of different downscaled variables is vital to capturing phenomena that inflict compound impacts.

Figure 8: A comparison of seasonal wind speeds in the training data (marked observations here) and LOCA downscaled data set.
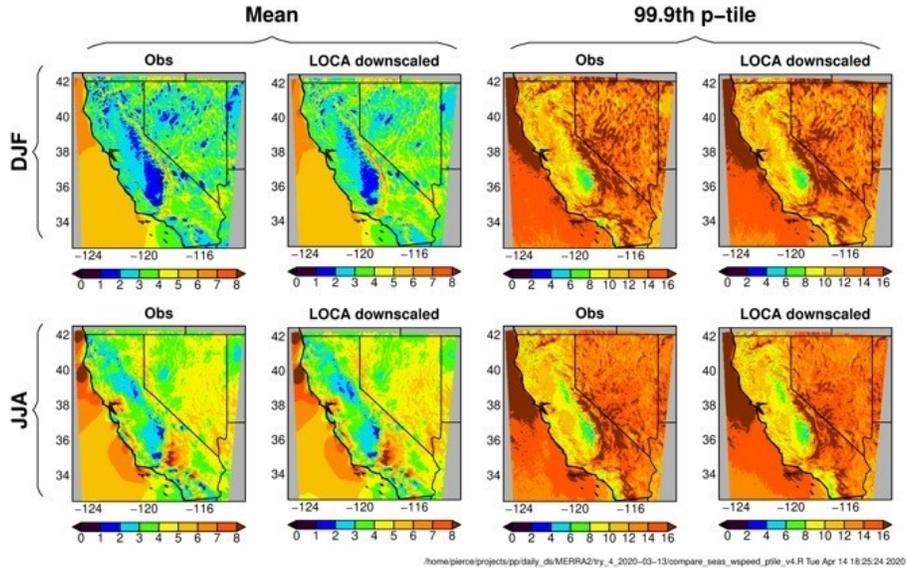
Top row is winter (Dec-Jan-Feb), lower row is summer (Jun-Jul-Aug). In each row, the left two columns show the mean wind speed (m/s), while the right two columns show extreme (99.9th percentile) values. All analyzed data is hourly taken at 4 PM local standard time.

Figure 9: A comparison of an individual strong Delta Breeze event and Sundowner wind, LOCA vs. WRF training data
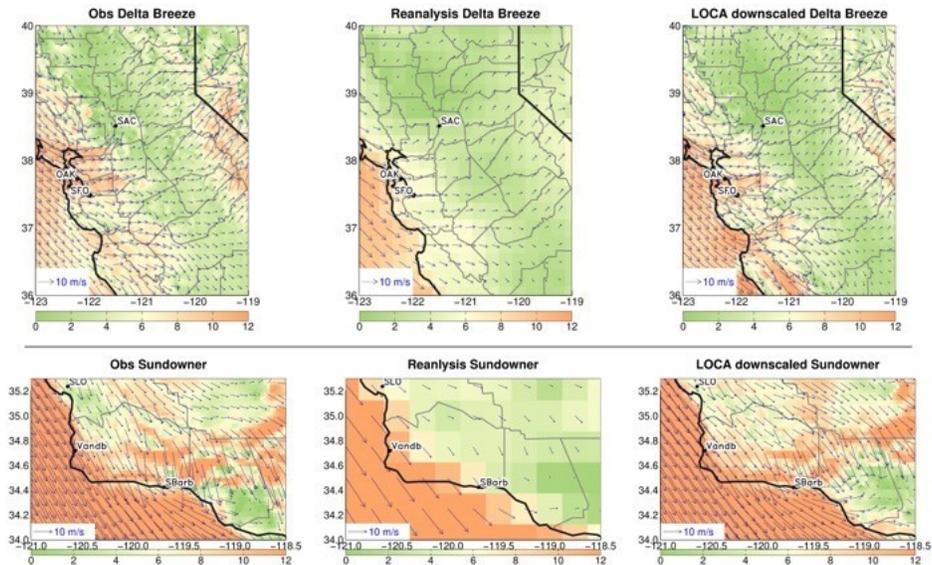
A comparison of an individual strong Delta Breeze event (top row, 2017/06/11 at 1 PM local standard time) and Sundowner wind event (bottom row, 2009/04/15 at 9 PM local standard time) in the DRI-WRF training data (marked Observations here), from the ERA5 reanalysis, and the ERA5 reanalysis downscaled with LOCA. Downscaled results determined from portions of training dataset that do not include these cases. Colors show wind speed in m/s, while arrows show the vector wind speed and direction.

4.2 Bias Correction

Even state-of-the-art high resolution simulations, such as the WRF runs that will be completed as part of this effort, have significant biases in the simulated output fields. This will be addressed using bias correction, which combines information from the WRF simulations with historically observed data to reduce the biases in the model output, resulting in a more realistic final data set.

A necessity in bias correcting winds and other variables is the availability of observational data over a time period that has sufficient representation of variability from daily extremes to interannual fluctuations. An observational dataset, becoming increasingly useful, is the weather data collected by public utilities; Figures 10 and 11 compare dynamical modeled winds from WRF-CA-CLC v4 against observed winds from the San Diego Gas and Electric (SDG&E) weather stations in the San Diego County region, illustrating relatively good agreement over a longer period and also over a shorter period during a Santa Ana wind event. Systematic bias in the modeled winds appears to exist. Biases are generally positive (modeled winds exceed observed SDGE winds), and these overestimated speeds are largest in desert regions east of the local coastal mountains (Laguna Range) and to a lesser extent over the coastal regions. General agreement by month is best over the crest and on the western slopes of the coastal topography (Figure 10). For the Santa Ana wind event considered at a mountain crest location (Julian), the overestimation bias is strongest in the early afternoon and evening as well as in the longer temporal persistence of the modeled Santa Ana wind event (Figure 11). This being a short sample, bias assessment and correction will need to be carried out over many more events. Results from the short modeling experiment carried out here are sufficient to indicate that modeled winds are not exceedingly sensitive to the details of the WRF physics package chosen, certainly much less so than is the case for coastal stratus clouds.

Figure 10: Wind speed mean difference (WRF v4-SDGE obs) for Aug – Nov 2010.
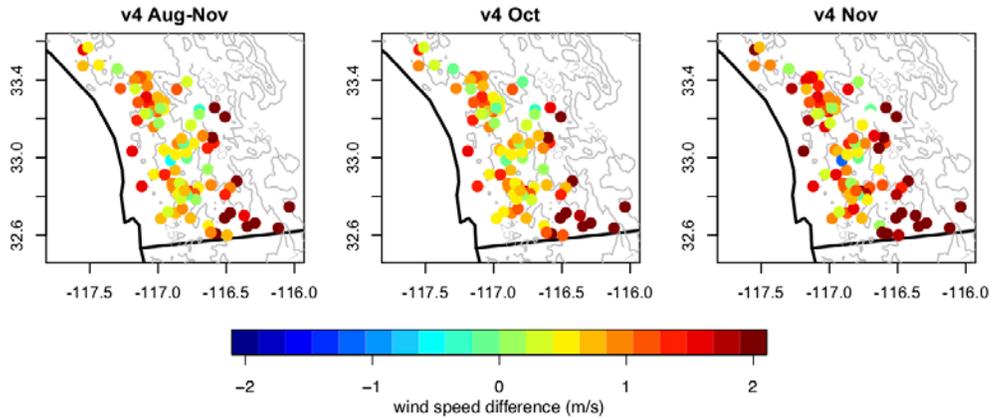
Wind speed mean difference (WRF-v4-SDGE obs) for Aug – Nov 2010. Differences are shown for the complete period (most left column) and for Oct and Nov (subsequent two columns). SDGE 10min-resolution observations are averaged to 3hrs to match WRF output resolution.

Figure 11: WRFv4 simulation of Santa Ana wind that occurred on 13 November 2010.
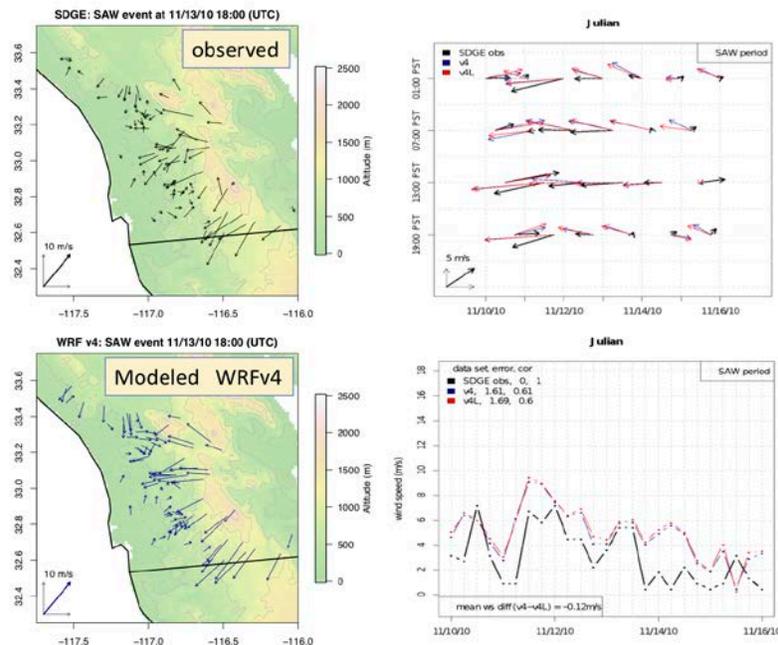
The 3-hr wind speed from WRFv4 tends to exceed that observed by SDGE weather stations. WRF simulated wind directions quite closely reproduce the reversed East-to-Wind that occurs during Santa Ana events.

In both dynamical and statistical models, the ideal bias correction procedure would jointly adjust multiple variables using a full multi-variate approach. E.g. a model may not produce uniform biases in T and P in simulating warm and cool precipitation events. The existing LOCA bias correction scheme (Pierce et al. 2015) addresses this using conditional bias correction; for example, the bias correction of temperature is conditional upon the presence or absence of precipitation so that the differing impacts of snow and rain are better captured in the final result. Although under the ongoing EPC-16-063 project, a multi-variate bias correction technique has been developed, this come at a significant price in much more computational demand, by about a factor of 30. There is no clear path forward in regards to reducing the large time demands of the multi-variate bias correction. Since it is an iterative process, we must accept that there may be no way around this, and we will have to evaluate whether the 30-fold increase in time needed to do this form of multivariate bias correction is worthwhile and acceptable given our need to process all of California for multiple models, ensemble members, and scenarios. This might largely be determined by how many models, ensemble members, and emissions scenarios are desired in forthcoming CMIP6 downscaling effort.

5. State of the Art Hydrologic Modeling Approaches

Modelers, policymakers, and stakeholders have an ongoing and growing need for high-resolution and detailed information about hydrological flows and the temporal-spatial distribution of water in a watershed. This need reflects the growing importance of coupling research with detailed long-term predictions and projections for ecological systems and the environment, agricultural development, and food security under future climate change. Assessments of the effects of climate variability and climate change in regions such as California also requires information on soil moisture and groundwater fluctuations, which are related to streamflow and reservoir management.

5.1 SHUD, the *Simulator for Hydrologic Unstructured Domain*

The Simulator for Hydrologic Unstructured Domains (SHUD - pronounced "SHOULD"; https://www.geosci-model-dev-discuss.net/gmd-2019-354/) is a multi-process, multi-scale integrated hydrologic model using the semi-discrete Finite Volume Method. The SHUD modeling system is useful for rapid, reproducible, and automated hydrologic modeling. As a descendant of the Penn State Integrated Hydrologic Model (PIHM), SHUD builds upon over 15 years of past hydrologic modeling experience. SHUD is a robust, integrated modeling system with the potential for providing scientists with new insights into surface and groundwater hydrology. It is expected that SHUD will enable a deeper understanding of water in California, particularly through future efforts targeting integrating scientific principles with water management infrastructure.

In the present study, SHUD has been configured and employed for modeling of surface and groundwater in several domains of various sizes: in the Wagon Creek and the Catchments Attributes and Meteorology for Large-sample Studies (CAMELS) watersheds, over the Sacramento River watershed from the highlands to the delta, and over the entire state of California. With its ability to work across scales, SHUD enables hydrology to be studied in many possible contexts, such as the behavior of water in regions of rough topography, in light of water resource

and stormwater management, and in conjunction with related fields such as limnology, agriculture, geochemistry, geomorphology, water quality, and ecology, climatic and land-use change. Whereas many existing hydrologic modeling systems are unable to deal with California's rough topography and significant topographic variation, SHUD has been designed with this relatively extreme regime in mind. In general, SHUD is a valuable scientific tool for modeling and understanding hydrologic action and response.

Figure 12: The basic calculation schemes in SHUD model.
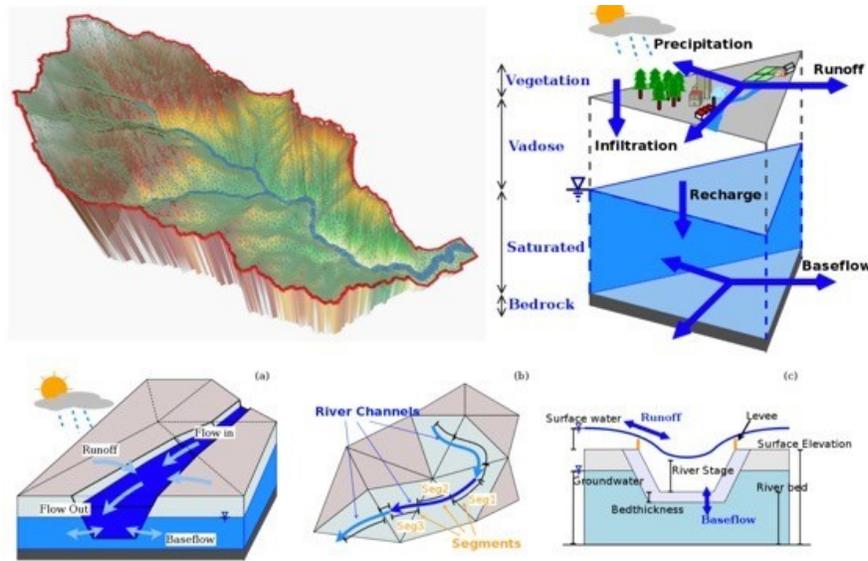
SHUD integrates overland flow, snow accumulation/melting, evapotranspiration, subsurface and groundwater flow, and river routing, while using a robust and realistic strategy for capturing the physical processes in a watershed. SHUD incorporates one-dimensional unsaturated flow, two-dimensional groundwater flow, and river channels connected with hillslopes via overland flow and baseflow (Figure 12). SHUD design is based on a concise representation of hydrodynamics of a watershed and river basin, which allows for interactions among major physical processes operating simultaneously, but with the flexibility to add or drop constitutive relations between states and processes depending on the objectives of the numerical experiment. SHUD is a distributed hydrological model in which the domain is discretized using an unstructured triangular irregular network (e.g., Delaunay triangles) generated with constraints (geometric and parametric). A local prismatic control volume is formed by the vertical projection of the Delaunay triangles forming each layer of the model. Given a set of constraints (river network, watershed boundary, elevation, and hydraulic properties), an "optimized mesh" is generated. The "optimized mesh" enables hydrologic processes on the unstructured mesh to be calculated efficiently, stably and rationally (Farthing and Ogden, 2017; Vanderstraeten and Keunings, 1995; Kumar et al., 2009). River volume cells are also prismatic, with trapezoidal or rectangular cross-section, and maintain a topological relationship with the Delaunay triangles.

Figure 13: The hydrograph in Cache Creek (simulation versus observation) in the calibration (2001-07-01 to 2003-06-30) and validation periods (2003-07-01 to 2007-06-30).
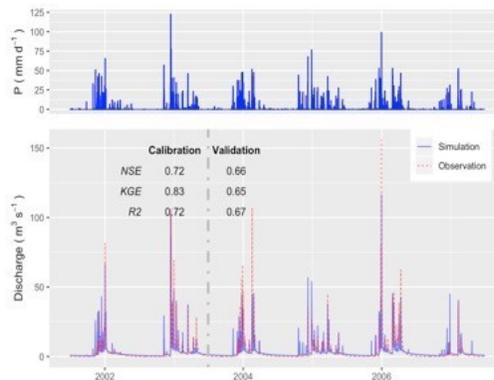
Figure 14: SHUD Model domain built for Central Valley. The number of cells is ~12500 and equivalent resolution is about 12 km² (3.5 * 3.5km).
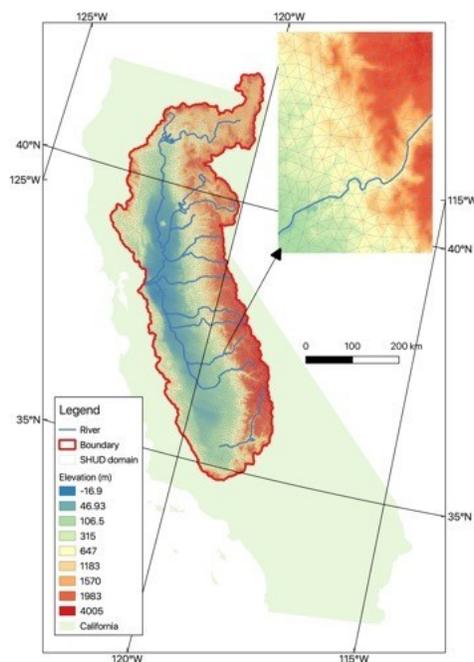
Performance of SHUD in simulating historical variability in particular sub-basins was good, as judged against observed streamflow variations (Figure 13). SHUD configurations were also developed for the Central Valley (Figure 14) and the entire state of California. To realize good model performance in a variety of settings, both open and closed boundary conditions were implemented (which is needed, for instance, because the state boundary is not a watershed boundary). Calibration on the Sacramento Watershed was slow, partly because computing resources were limited. In addition, due to the large area, the model, particularly groundwater,

required a very long spin-up period. Because of the number of elements and resolution required in such a vast region, the model's current OpenMP parallelization strategy may need to be hybridized with MPI. Nonetheless, a functional configuration of the model was found and is being explored to better understand California's hydrologic system.

5.2 Machine Learning Approach to Hydrologic Modeling

Using a novel data-driven approach, we have demonstrated the utility of temporal convolutional neural networks (TCNNs) for streamflow prediction and projection in California. Specifically, this machine learning (ML) based approach was used to capture the functional relationship between input fields and streamflow using training data derived from CAMELS, which provides daily precipitation, temperature, solar radiation and streamflow from 1980 to 2014. This prototype system used daily data from the 20 basins in California with consistently high-quality observations over the entirety of the study period. The TCNN model was trained separately for each basin with a single RTX 2080Ti GPU, requiring around 44 seconds per basin. The final trained model could then ingest a time series of input fields and produce a corresponding streamflow time series in seconds. This model was then compared with alternatives using linear regression, ANN, GRU and LSTM; an analysis of the resulting historical streamflow time series over a separate testing period showed that the TCNN model achieves comparable or superior performance to the GRU and LSTM systems, while requiring a shorter training time.

The trained model further showed evidence of capturing relevant physical processes, including groundwater and snow dynamics. It also exhibited stability under climate change, producing reasonable results even under extreme forcing. The trained model was then used for producing streamflow projections under climate change by taking the LOCA (Localized Constructed Analogs) dataset as input. In particular, precipitation, temperature and solar radiation at daily timescales was used from four global climate models (CanESM2, MIROC5, HadGEM2-ES and CNRM-CM5) that were selected for use in the Fourth California Climate Change Assessment (Pierce et al. 2018). Work is currently underway to compare the quality of these projections against analogous results produced using the SHUD model discussed in the previous section. If comparable performance is observed, then the ML approach would provide a rapid statistically-based method for downscaling meteorological inputs for streamflow prediction and projection.

The TCNN model is relatively general and can be easily applied to any other basin. Since the model is trained separately for each basin, a rapid ensemble projection over the entire state would also be possible with GPU acceleration.

5.3 Integrated Surface Water-Ground Water Modeling

A version of the ParFlow.CLM model (an integrated groundwater-land surface model (Kollet and Maxwell, 2008)) has been configured and validated for the Kaweah River watershed in the Southern Sierra Nevada and Central Valley California, covering an area of 12,276 km^2 (Figure 15). ParFlow.CLM simulates variably saturated subsurface flow using the 3D Richards' equation and it is coupled to the Common Land Model (CLM 3.0) to solve water and energy budgets at the land surface at an hourly time step. After validating the model against a range of in situ

(streamflow) and remotely sensed observations (evapotranspiration and snow water equivalent), we assess the impact of uncertainty in model forcing datasets, focusing on the precipitation and air temperature uncertainty. Gridded precipitation and air temperature datasets are used as historical records for downscaling climate projections, but they suffer from high levels of uncertainty themselves. We perform simulations of the Kaweah River watershed using precipitation and air temperature from four common gridded products, NLDAS-2, PRISM, Daymet, and Gridmet.

Figure 15: (a) Location of the study domain within the state of California. (b) Digital elevation model used to generate slope parameters for the ParFlow.CLM model.
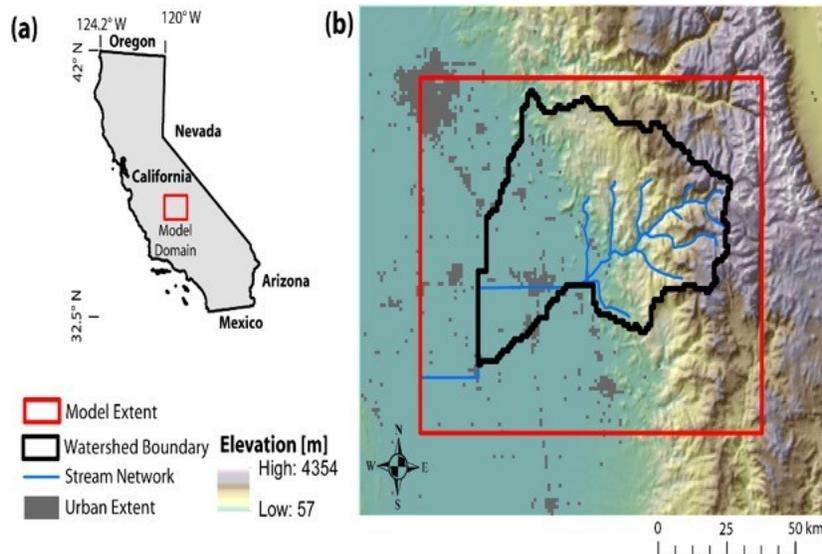
The watershed boundary and stream network derived from the 4-directional routing scheme used within ParFlow.CLM are shown as well as locations of urban development.

To quantify the impact of uncertainty in meteorological forcings on the simulated water budget, we calculate the relative uncertainty of model forcing data and simulated hydrologic fluxes using the three-cornered hat (3CH) method (Premoli & Tavella, 1993).

We recognize that this is not a total measure of uncertainty because it does not account for the uncertainty created by the model structure and parameters. However, this is a measure of model output uncertainty caused specifically by the uncertainty in precipitation inputs. We first calculate the uncertainty (Figure 16) in basin averaged time series of precipitation (P), streamflow (Q), ET, SWE, soil moisture (Θ), and potential groundwater recharge (R). We find that the datasets with the highest uncertainty in P do not result in the highest uncertainty for all output variables. It is particularly interesting to note that although the Gridmet forcing has the highest uncertainty in P, it has the lowest uncertainty in Q. Hydrologic models are typically calibrated and validated using only streamflow estimates. Our results suggest that this may not be an adequate approach if scientists are interested in the rest of the water budget where the Gridmet dataset has higher uncertainty (ET, Θ, or R). We also calculate the spatially distributed uncertainty in each of the simulated water balance components. We demonstrate that although the highest uncertainty in precipitation inputs is at the highest elevations of the Sierra Nevada, the highest uncertainty in simulated fluxes is found in the mid elevation regions along the transition zone between the Sierra Nevada range and the Central Valley. We attribute this transformation of uncertainty to the impact

of topography (Schreiner-McGraw and Ajami, *in review*). Quantifying uncertainty of simulated hydrologic fluxes caused by climate forcing is important as these gridded climate products are used to train or validate downscaled climate projections for climate change impact studies.

Figure 16: Uncertainty in domain average time series of multiple hydrologic variables.
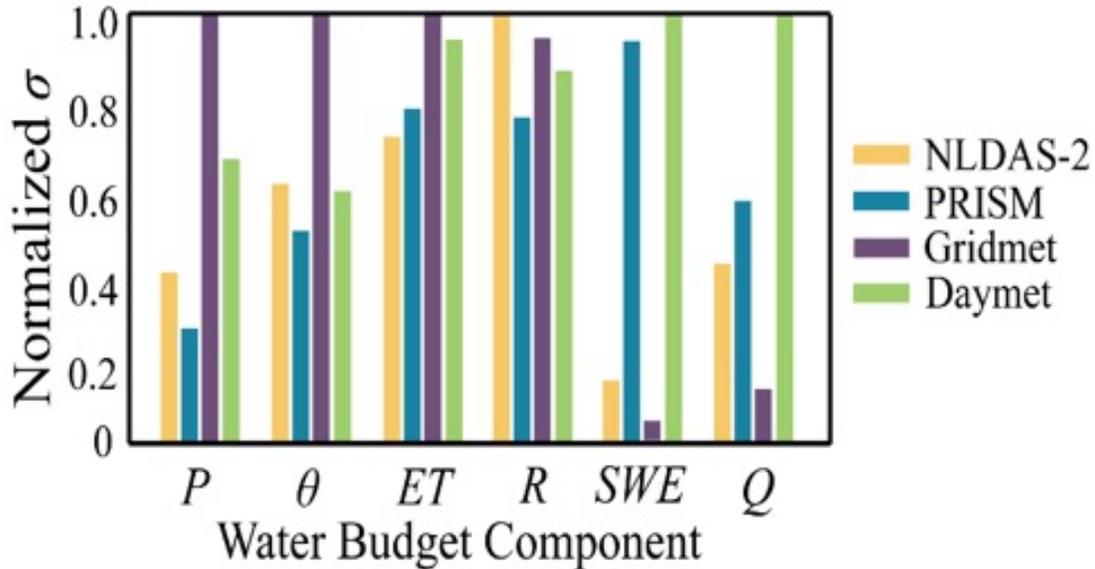
For each variable the uncertainty has been normalized by the dataset with the largest uncertainty to allow different variables to be displayed on the same chart.

Quantifying groundwater response time and recovery to climate change and droughts is one of the main objectives of this CEC project, and has important implications for groundwater management under sustainable groundwater management act (SGMA) in California and energy use for groundwater pumping. Our focus is on utilizing groundwater observations as well as idealized simulation domains to quantify these impacts while developing new metrics to quantify groundwater response time to droughts. To explore the groundwater response time to climate change, we utilize droughts as examples of meteorological changes imposed on the hydrologic system. A key challenge is associating observed changes in groundwater levels with corresponding observed changes in precipitation that caused the groundwater drought. Through this project we utilize precipitation observations from the PRISM dataset and groundwater level data from the Climate Response Network to test the efficacy of different drought classification methods on quantifying the groundwater response time to droughts. We find that previously used lagged-correlation between the standardized precipitation index (SPI) and the standardized groundwater index (SGI) is a reliable method (Bloomfield and Marchant, 2013) to quantify the impacts of aquifer properties on the response time, while other metrics are required to understand the impact of precipitation properties on groundwater drought (Schreiner-McGraw and Ajami, *In prep.*). In addition to the observational experiment, we develop synthetic numerical experiments using an integrated groundwater-land surface model, ParFlow.CLM, forced by synthetic climate realizations, to test drought classification metrics for groundwater. The model allows us to precisely quantify changes in groundwater storage to ensure that drought classification metrics are reliable (Schreiner-McGraw et al., *in prep.*).

So far in this project we have evaluated how uncertainty in the precipitation datasets propagates through a process-based hydrologic model resulting in uncertainty in the simulated hydrologic outputs. Next, we are investigating how combined uncertainty in precipitation and air temperature forcings impact simulated hydrologic fluxes in the Kaweah River watershed. Additionally, we plan to test the model for the larger Tulare Basin watershed. We focused on the Kaweah watershed (Figure 15) while learning how to implement this modeling approach and developing an approach with reasonable simulation times. We have determined the correct parameter sets that facilitate the rapid simulation of Kaweah watershed and we will expand this to the larger Tulare Basin region to test the feasibility of running a larger area using our limited high performance computational resources. Now that we have been able to improve our run-time limitations for the Kaweah River study basin using ParFlow,CLM, we will complete our simulations before the end of this CEC project. The model has a horizontal resolution of 1 km with the subsurface thickness of 622 m. This means that the model has the potential to simulate the impacts of groundwater pumping and irrigation on surface water-groundwater exchange.

After validating the model, we calculate the uncertainty of model outputs using the 3CH method. We recognize that this is not a total measure of uncertainty because it does not account for the uncertainty created by the model structure and parameters. However, this is a measure of model outputs uncertainty caused specifically by the uncertainty in precipitation inputs. We first calculate the uncertainty in basin averaged time series of precipitation (P), streamflow (Q), ET, SWE, soil moisture (Θ), and potential groundwater recharge (R), shown in Figure 16. We find that the datasets with the highest uncertainty in P do not result in the highest uncertainty for all output variables. It is particularly interesting to note that although the Gridmet forcing has the highest uncertainty in P, it has the lowest uncertainty in Q. Hydrologic models are typically calibrated and validated using only streamflow estimates. Our results suggest that this may not be an adequate approach if scientists are interested in the rest of the water budget where the Gridmet dataset has higher uncertainty (ET, Θ, or R).

Although the ParFlow.CLM modeling approach is computationally expensive relative to simpler hydrologic models commonly applied for climate change impact assessment, there are important advantages. Utilizing a process-based integrated modeling approach to simulate the climate change impacts on water resources has the advantage that the model performance is not tied to historical conditions. By representing the processes via which water moves through a landscape, changes in the meteorological conditions will be propagated through the system via physically-based equations describing shallow surface water and subsurface flow, rather than empirical methods derived from the statistical analysis of historic data. The second advantage of our approach is that we use a 3D representation of groundwater coupled with a land surface model. This allows us to study the impacts of hydrologic connectivity between the Sierra Nevada mountains, where most of the precipitation falls, and the Central Valley aquifer, where most of the water is used. Furthermore, it is possible to assess the impacts of pumping, irrigation and snowmelt processes in a fully integrated manner.

For the Kaweah River watershed, we use a domain that is 99x124x15 pixels and simulate this with 35 computational nodes. The slowest simulations take 12 hours of simulation time (~420 CPU hours) per year of simulation. Assuming that climate projections require 2,500 years of simulation time (25 GCMs with 100 years of data), this would require ~1 million CPU hours. This equates to

about 1 month of run time on 1,500 computational nodes. Therefore, this approach is feasible for the entirety of California if we have sufficient computational resources. A tradeoff could be achieved by using less than 25 scenarios of 100-year duration. If we ran 100 years of simulation for a historical period and 3 projected climates (representing low, median and high) precipitation projections, we would need less computational resources.

The ParFlow.CLM modeling approach provides multiple benefits. The ParFlow,CLM model simulates the entire terrestrial hydrologic cycle from the top of the mountains to the deepest part of the aquifer system using a 3D discretization. Furthermore, the ParFlow.CLM is coupled to the WRF model. Such an integrated atmospheric-hydrologic modeling system (ParFlow.WRF) can provide an ideal platform for regional climate downscaling while embedding detailed hydrological processes in the modeling framework. However, a fully integrated climate-groundwater framework is computationally expensive at this moment.

6. Observational Datasets

Weather, climate and hydrologic observational datasets have been employed in **this project** to model and investigate historical fluctuations and extreme events. The modeling and analysis in this project have used these data both as training and validation. Several of these dataset contain a small number of variables, e.g. Livneh or PRISM precipitation and maximum and minimum temperature, which may be commonly observed over much of California. However, as needs for downscaled data grow, datasets are required that include several climate variables, e.g. temperature, precipitation, winds, humidity, clouds or solar radiation. Because climate varies over a range of time and spatial scales, these datasets should ideally include many samples over a sufficient time span to represent climate variability, including short period extremes and long period fluctuations. The following Tables identify and comment on several of the datasets that were used in EPC-16-063 or in previous related studies.

Table 2: *In situ* **observational datasets employed in EPC-16-063 or commonly employed in climate studies.** (commonly used, for example, not comprehensive)

Data Set	Sampling	Duration	Coverage	Quality
COOP observer Temperature and Precipitation	Daily	Several decades	Statewide, but irregular	Uneven quality, use w/ care
First order (Airport) meteorological observations	Hourly	Several decades	Statewide, sparse, mostly airports	Generally good, but some are contaminated by station or instrument relocation and instrumental changes

RAWS wind, humidity, other variables	Hourly	Few-decades	Statewide in fire prone areas	Uneven quality, Use with care
SDGE and California other Utility station wind, humidity, barometric pressure	Hourly	Few to 10 years	Regional	Mostly good
Ocean weather buoy meteorology and upper ocean	Hourly	Few decades	Sparse locations near coast and offshore	Mostly good
Balloon-borne upper air observations	Twice-daily	Several decades	Few locations in CA	Good, but time of observation has changed
Snow course observations	Monthly	Several decades	Mountain locations that have seasonal snowpack	Good
Snow Sensor (California DWR) SNOTEL (Federal NRCS) and associated met observations	Hourly	Few decades	Mountain locations that have seasonal snowpack	Good but availability and quality varies depending time period and variable
Stream gage observations	Hourly-daily	Several decades	Statewide	Good
GOES (remote sensed) cloud and irradiance imagery	Half-hourly	Few decades	West-wide	Good

6.1 Gridded Datasets and Reanalysis

Most observed measures are not uniformly observed. Historical records are spatially patchy, temporally limited, and often contain errors. Thus, historical reanalyses are employed, using a dynamical model to develop a more complete, dynamically consistent set of observational records needed to investigate mechanisms controlling various forms of variability, including extreme events.

Table 3: Gridded Datasets employed in EPC-16-063 or commonly employed in climate studies

Data Set	Sampling	Duration	Coverage	Quality
Livneh Tmax, Tmin, Precipitation	Daily	Several decades	CONUS	Good

NLDAS-2 meteorology, fluxes	Daily and monthly	Several decades	CONUS	Good
PRISM Precipitation temperature, dew point temperature	Daily and monthly	Several decades	CONUS	Good but PRISM cautions not to rely on trends
Gridmet Precipitation, Tmax, Tmin, downward shortwave radiation, vapor pressure	Daily	Several Decades	CONUS	Mixed. Some variables are only sparsely measured and hence derived from regional model results.
Daymet Precipitation, Tmax, Tmin, downward surface shortwave radiation,	Daily	Several Decades	CONUS	Mixed. Some variables are only sparsely measured and hence derived from regional model results.
ERA5 Reanalysis	Hourly	1979-current to be extended to 1950-near current	Global	Good
NARR Reanalysis	3-hourly	1979-near present	North America and adjacent oceans	Good
NCEP Reanalysis	6 hourly	1948-present (NCEP R2, and improved version, is 1979-present)	Global	Ok
MERRA-2 Reanalysis	Hourly	1980-near present	Global	Good
Margulis Snow Reanalysis daily snow water equivalent, snow covered area	Daily	1985-2016	Sierra Nevada	Good

6.2 Climate and Hydrologic Variables Feasible to Downscale Using Dynamical, Statistical of Hydrologic Models

How feasible it is to downscale a climate variable depends on a number of factors. At one extreme, a variable that is well observed, spatially coherent, has a predictable or systematic diurnal variation, and is realistically simulated by both GCMs and WRF is easily downscaled. Temperature is the obvious variable satisfying nearly all these criteria. At the other end are

variables that are poorly observed, have strong spatial and temporal variability, and are poorly simulated by both GCMs and WRF. Vector wind is an example of a problematic variable. Such variables can be downscaled by the statistical, dynamical, and hybrid approaches described herein, but with poor observations and questionable model simulations the correspondence of the downscaled fields to what actually happens on the ground cannot be very well known. The following table parses out the research team's opinions, generated from this project as well as from prior experience, relating to the quality and feasibility for different variables.

Table 4: Variables of Interest for Downscaling

Variable	Quality and Feasibility Remarks
Temperature	Good. Observations are plentiful (although better for daily than hourly), spatial scales of temperature anomalies are large, the elevation dependence of temperature is strongly constrained, and GCMs agree in predicting a future warming trend. However, traditional statistical downscaling miss feedbacks in some areas, such as high-elevation regions that lose year-round snow cover as temperatures warm.
Precipitation	Ok. Observations are plentiful but mostly daily (not hourly) and not uniformly distributed across the landscape, and, precipitation has high space-time variability that may not be inadequately captured by statistical or dynamical models. Precipitation change in GCMs varies widely across GCMs and across ensemble members of same GCM, indicating that chaotic and unpredictable weather/climate fluctuations will play an important role in determining what actually happens in the state in the future
Wind	Problematic. Observations are sparse and quality is problematic because of changes in site surroundings, location, and instrumentation. Dynamical modeled winds may have closer agreement with observed wind gusts than with time averaged winds. Both global and regional dynamically modeled winds vary considerably from each other and from the available observations, suggesting that models generally simulate wind poorly and we have little basis for identifying good models.
Humidity	Ok. Near surface atmospheric humidity is only sparsely observed, but has broad spatial scale (like temperature). Dynamical model humidity appears to be ok in replicating sparse observations. Statistical downscaling performs reasonably well in replicating regional dynamical model structure, although there may be smaller scale features that statistical techniques miss.
Solar Radiation	Ok. Satellite observations are available over the region, although for a more limited period than temperature and precipitation observations. Cross validation studies show that statistical downscaling performs well for this field, but model-observation comparisons show that GCMs and WRF have significant problems simulating cloudiness, especially marine and coastal stratus layers.

	The quality of future downscaled projections will be hindered by GCM foibles in future projections of cloudiness, and dynamical (and therefore hybrid) downscaling approaches do not appear to alleviate this problem.
ET	Problematic. ET is rarely measured, especially over extensive periods of time. ET can be calculated but depends upon multiple variables, some of which are not themselves commonly observed or easily modeled. Furthermore, depends upon land surface cover.
Streamflow	Ok. Runoff is ok, but must be routed through stream channel network, which is not routinely provided by macroscale hydrological models. Additionally, these models don't routinely represent human manipulations—many (most) streams are affected by dams and diversions.
Snow water equivalent	Ok. Models such as ParFlow.CLM can provide good estimates using the energy balance approach. However, observations for validation is limited. Recent SWE and SCA products from Margulis across Sierra Nevada are used for validation.
Groundwater level elevation	Problematic. High resolution groundwater level observations across CA particularly in the mountain regions are not available. Models like ParFlow.CLM can estimate this information and incorporate the impact of pumping. Accurate estimates of groundwater levels are impacted by the lack of detailed subsurface characterization.
Soil moisture	Problematic. Soil Moisture is extremely heterogeneous and hard to measure and model. Hydrologic models can estimate soil moisture using physically based approaches. However, model validation is limited given the limited availability of in situ observations and coarse resolution of remotely sensed products for catchment scale evaluations.

7. Lessons Learned (from EPC-16-063 and Related Studies)

Increasingly, California's utilities and other decision makers are finding needs for greater specificity in the climate and weather events that drive extreme impacts as well as deliver water and other vital resources to the State. Here, we summarize some key findings that relate to effective climate model downscaling and hydrological modeling that have come into focus in our ongoing EPC-16-063 research, and also informed from prior studies such as Pierce, et al. 2018, which underpinned California's Fourth Climate Change Assessment. Realistic atmospheric downscaling and hydrological modeling is highly dependent upon data used to drive the models as well as the constructs of the models themselves. Input data may be sets of direct observations, analyzed observations, and model generated data such as global and regional reanalyses. Atmospheric downscaling models may be either dynamical or statistical (or combinations thereof), and therefore carry different sorts of advantages and limitations. Hydrologic models contain another set of techniques, advantages and drawbacks. Tradeoffs between model sophistication and model simplicity and efficiency arise because of limitations in computational resources. Without going into great depth, these are some of the issues that are discussed below.

- The quality of dynamical and statistical downscaling is dependent upon the availability of global or regional modeling output to provide large scale guidance on projected or historical patterns. and high resolution historical data for validation or for training statistical models,

- Downscaling different variables may call into play different fine scale observation or regional dynamical model training datasets. Observed training or validation data is represented with varying quality and uneven spatial and temporal coverage for different variables. Observational datasets invariably contain some degree of sampling and instrumental errors.

- Biases in statistical or dynamical model output occurs in all variables, but is usually addressed by bias correcting individual variables without consideration of other variables and their biases. Full multi-variate bias correction schemes may provide a more consistent, more accurate representation of regional climate but are more than an order of magnitude more computationally expensive than the existing conditional bias correction methods and may not be affordable if large numbers of models, ensemble members, and emissions scenarios are desired.

- Dynamical modeling of some phenomena (e.g. marine boundary layers and coastal clouds) is problematic with standard regional model codes. Further, the effects of influences such as aerosols on California's coastal marine clouds is not well understood and new modeling and diagnostics study is required.

- In conducting statistical downscaling, consideration of multiple variables may be useful in identifying the patterns employed in statistically downscaling individual variables. For

example, it was found that the addition of sea level pressure to regional wind patterns added skill when LOCA downscaling wind over the California region. This underscores the value of the full suite of dynamically consistent atmospheric variables provided by atmospheric reanalyses or climate GCMs as parts of the statistical downscaling toolkit.

- There are a number of less well used and well measured variables for which statistical downscaling procedures have not been developed. A noteworthy example is a set of variables of interest to the wildfire prevention and management community that may be available from regional dynamical models but not from statistical downscaling techniques.

- Regional dynamical model downscaling of high volumes of GCM simulations is not practical because of computational resource limitations. Even statistical downscaling of the broad suite of CMIP6 simulations that is emerging will require substantial computer resources and data storage. Procuring necessary computer resources for comprehensive hybrid dynamical/statistical downscaling is a critical requirement for such an effort.

- Statistical downscaling schemes are most commonly run using historical training and historical analog datasets. However, "stationarity", wherein it is assumed that historical statistics will still apply to future periods which have undergone significant global climate changes, may not always hold up. Therefore it is worth considering a development under which regional dynamical downscaling is conducted on mid- and end-21st Century GCM output in order to produce alternative training data for statistical model downscaling within these future periods.

- Global and regional reanalyses differ in their scale and quality of historical patterns, evolution. Limited evidence suggests that finer scale global reanalyses, and those which more effectively assimilate remote sensed observations, may provide better large scale guidance in finer scale dynamical or statistical downscaling of historical patterns.

- Even though the global atmospheric reanalysis variables may have amplitudes that are diluted in comparison to high resolution dynamical results, they may nonetheless provide patterns that are sufficient to produce quite accurate downscaled results. Downscaling wind over California using the ERA5 reanalysis is a noteworthy example.

- Some observational datasets are relatively short or confined to a limited domain (e.g. SDGE winds and humidity), and thus may be inadequate for regionwide statistical downscaling training datasets. Nonetheless, such data may provide extremely valuable evaluation or calibration of downscaled methods and results.

- Surface water and groundwater resources, which are often treated as separate systems in climate impact assessments, ideally should be treated as linked components of the same hydrologic system.

- Like atmospheric models, different hydrologic models contain different kinds of physical representations and consequently different sets of uncertainties and errors. Consequently,

it is desirable to include multiple hydrologic models as part of regional climate change projections to provide better understanding of projection uncertainties.

- Landcover and irrigation practices are not commonly represented in most (any?) hydrologic models suited for climate assessment and will require further development. ParFlow.CLM has the irrigation scheme of CLM but further refinements are needed to represent irrigation types and scheduling relevant to California agriculture. Further, hydrologic models that represent other human manipulations such as dams, diversions and managed flows affect the hydrology of many catchments in California but are not represented broadly across the California Landscape.

- Routed runoff to drainage channels (to produce streamflow) is sought after from hydrologic model assessments. However, stream channel flow and routing is not included in some hydrologic models.

- Modeling groundwater, aquifer recharge and withdrawal is hampered by having to include important physical processes that operate on a very broad range of short to long timescales, necessitating large amounts of computational power. Modeling groundwater variation and change statewide using integrated groundwater-land surface models is possible but require investments in computational resources to include in the upcoming California Climate Change Assessment. The benefits of such an approach is in capturing hydrologic processes from the atmosphere to the bottom of the aquifer, removing the need for implementing multiple hydrologic models to capture different aspects of the system.

- Different global climate model simulations provided to the CMIP6 GCM archive have differing degrees of completeness in providing different future projected socioeconomic global changes (SSPs), numbers of ensemble members, spatial resolution and temporal sampling, and variables that are saved (Table 5). Currently there are only about a dozen GCMs that offer a full suite of SSPs and variables that cover the historical past and the 21st Century, although additional model output is still being contributed.

Table 5: Current CMIP6 Data Status (data downloaded by David Pierce at SIO)

```
The  4  dots are these experiments, in order: (x: one ens member, is incomplete; X: >1 ens member, at least 1 incomplete
historical  ssp245  ssp370  ssp585
```

	hurs	hursmax	hursmin	huss	pr	ps	psl	rlds	rsds	tasmax	tasmin	uas	vas
ACCESS-CM2	1111	1111	1111	1111	1111	....	1111	....	1111	1111	1111	1111	1111
ACCESS-ESM1-5	1111	1111	1111	1111	1111	....	1111	....	1111	1111	1111	1111	1111
AWI-CM-1-1-MR	....	....	....	....	2...	....	....	....	....	112.	212.	111.	111.
BCC-CSM2-MR	....	....	....	1111	1111	....	3111	....	1111	1111	1111	1111	1111
BCC-ESM1	....	....	....	....	3.X.	....	3.X.	....	1.X.	3.X.	3.X.	1.X.	1.X.
CESM2	1111	....	....	2112	2122	1.2.	1111	1112	2121	.xXx	.xXx	....	....
CESM2-FV2	1...	....	....	1...	1...	1...	X...	....	....	....	....	....	....
CESM2-WACCM	11X3	....	....	22X4	2212	..X.	22X3	23x3	32X1	.1.1	.3.2	....	....
CNRM-CM6-1	8666	1616	1616	1616	1111	3.X.	9666	....	1616	1111	1111	1616	1616
CNRM-CM6-1-HR	1111												
CNRM-ESM2-1	6333	....	....	3333	41X1	1.X.	5333	....	53X3	51X1	51X1	1333	53X3
CanESM5	9919	1918	1919	1919	1919	....	9998	.9.7	1919	1915	1919	1919	1919
EC-Earth3	11x1	112.	11x.	1111	1111	....	1111	X...	1111	x111	1111	X111	1111
EC-Earth3-Veg	1...	1...	1...	1...	1...	....	1...	X...	....	....	....	....	....
FGOALS-f3-L	....	....	....	2...	2...	....	1...	....	1...	1...	1...	1...	1...
FGOALS-g3	211.	11x.	11..	11..	11..	....	11..	.1..	.1..	.1..	.1..	....	....
GFDL-CM4	11.1	11.1	11.1	11.1	11.1	....	....	....	.1.1	.1.1	.1.1	.1.1	.1.1
GFDL-ESM4	1111	....	....	111.	1111	1111	1111	....	1..1	1111	1111	1111	1111
GISS-E2-1-G	....	1...	1...	....	....	....	....	....	....	1...	2...	....	....
HadGEM3-GC31-LL	1X.2	....	....	2X.2	3X.2	3...	3X.3	.X.3	3X.3	2X.1	3X.3	3X.2	3X.3
HadGEM3-GC31-MM	1...	1...	1...	1...	1...	1...	1...	1...	1...	1...	1...	1...	1...
IITM-ESM	....	....	....	....	....	....	....	....	X...	....	....	....	....
INM-CM4-8	1111	1111	1111	1111	1111	....	1111	1111	1111	1111	1111	1111	1111
INM-CM5-0	1151	5131	5141	3111	2141	....	5131	2141	8141	2151	8111	3151	3121
IPSL-CM6A-LR	9993	9291	9291	....	9291	9231	9291	....	9291	9291	9291	9291	9291
KACE-1-0-G	.331	.33.	.33.	.331	.331	....	.331	.331	.331	.21.	.12.	.331	.331
MIROC-ES2L	2...	....	....	....	....	....	....	....	....	....	....	....	....
MIROC6	2231	....	....	....	3121	....	X211	....	4322	3311	5122	3231	2113
MPI-ESM1-2-HR	3141	.1..	.1..	3141	6151	2141	2111	....	2131	4121	5151	2131	1121
MPI-ESM1-2-LR	.34.	.43.	.53.	.54.	.54.	.63.	.51.	.33.	.43.	.75.	.44.	.34.	.44.
MRI-ESM2-0	1111	1111	1111	1111	1111	....	....	....	1111	1111	1111	1111	1111
NESM3	....	....	....	....	22.1	....	31.1	....	31.1	12.1	12.1	11.x	22..
NorCPM1	....	....	....	....	6...	....	8...	....	6...	9...	6...	9...	6...
NorESM2-LM	33X1	....	....	2111	22X1	2311	22X1	.3X1	11x1	21x1	31X1	....	....
NorESM2-MM	1111	....	....	1111	1111	1111	1111	1111	1111	1111	1111	....	....
SAMO-UNICON	1...	1...	1...	1...	1...	....	1...	1...	X...	1...	1...	....	....
TaiESM1	1...	....	....	1...	1...	....	1...	1...	1...	1...	1...	....	....
UKESM1-0-LL	3211	....	....	3111	3111	1.1.	1211	....	5211	3111	4111	3211	2211

The variables are across the top:
- hurs = daily average surface relative humidity
- hursmax, hursmin: daily min/max surface relative humidity
- huss = surface specific humidity
- pr = precipitation
- ps = surface pressure
- psl = sea level pressure
- rlds = downward surface longwave radiation
- rsds = downward surface shortwave radiation
- tasmax, tasmin = daily min, max temp
- uas, vas = surface (10 m) U, V

For each set of 4 digits, for example "...." or "1111" or "22X4", the digits from left to right are for the historical, ssp245, ssp370, and ssp585 runs. If there is a digit, it is the number of complete ensemble members I have downloaded. A period "." means there is no data downloaded. A lower case "x" means a single, incomplete ensemble member. An upper case "X" means multiple ensemble members, at leastone of which is incomplete.

References

Bloomfield, J. P. and Marchant, B. P., 2013. Analysis of groundwater drought building on the standardised precipitation index approach, *Hydrol. Earth Syst. Sci.,* 17, 4769–4787, https://doi.org/10.5194/hess-17-4769-2013.

Brown, T. J., Mills, G., Harris, S., Podnar, D., Reinbold, H. J., Fearon, M. G., 2016: A bias corrected WRF mesoscale fire weather dataset for Victoria, Australia 1972-2012, *Journal of Southern Hemisphere Earth Systems Science*, 66 (3), 281-313.

Clemesha, R.E.S., A.Gershunov, S.F. Iacobellis and D.R. Cayan, 2017: Daily Variability of California Coastal Low Cloudiness: A Balancing Act between Stability and Subsidence. *Geophysical Research Letters*, 44, 3330–3338, doi:10.1002/2017GL073075.

Farthing, M. W. and Ogden, F. L.: Numerical Solution of Richards' Equation: A Review of Advances and Challenges, *Soil Science Society of America Journal*, 81, 1257, https://doi.org/10.2136/sssaj2017.02.0058, 2017.

Guzman Morales, J., A. Gershunov, J. Theiss, H. Li and D.R. Cayan, 2016: Santa Ana Winds of southern California: Their climatology, extremes, and behavior spanning six and a half decades. *Geophysical Research Letters*, 43, doi:10.1002/2016GL067887.

Hersbach, H., and Coauthors, 2019: Global reanalysis: goodbye ERA-Interim, hello ERA5. ECMWF, doi:10.21957/vf291hehd7. https://www.ecmwf.int/node/19027.

Hinkelman, L. (2019), The Global Radiative Energy Budget in MERRA and MERRA-2: Evaluation with Respect to CERES EBAF Data, J. Clim, 32, 1973-1994, doi:10.1175/JCLI-D-18-0445.1

Kollet, S. J., and Maxwell, R.M., 2008. Capturing the influence of groundwater dynamics on land surface processes using an integrated, distributed watershed model, *Water Resour. Res.*, 44, W02402, doi:10.1029/2007WR006004.

Kumar, M., Bhatt, G., and Duffy, C. J., 2009: An efficient domain decomposition framework for accurate representation of geodata in distributed hydrologic models, *International Journal of Geographical Information Science,* 23, 1569–1596, https://doi.org/10.1080/13658810802344143

Martin, A., Clemesha, R., DeHaan, L., and D. Cayan, (*in-prep*) Hybrid Downscaling of California Coastal Cloudiness at Decadal and Longer Timescales. *Journal of Climate*.

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P.C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., Berbery, E.H. and Ek, M.B., 2006: North American regional reanalysis. *Bulletin of the American Met. Society*, 87, 3, 343-360. doi:10.1175/BAMS-87-3-343.

Molod, A., L. L. Takacs, M. J. Suarez, J. T. Bacmeister, I.-S. Song, and A. Eichmann, 2015: Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA2. *Geosci. Model Dev.,* 8, 1339–1356, https://doi.org/10.5194/gmd-8-1339-2015.

Pierce, D. W., T. P. Barnett, B. D. Santer, and P. J. Gleckler, 2009: Selecting global climate models for regional climate change studies. *Proceedings of the National Academy of Sciences*, doi:10.1073/pnas.0900094106.

Pierce, D. W., D. R. Cayan, and B. L. Thrasher, 2014: Statistical downscaling using Localized Constructed Analogs (LOCA). *J. of Hydromet*, volume 15, page 2558-2585.

Pierce, D. W. and D. R. Cayan, 2015: Downscaling humidity with Localized Constructed Analogs (LOCA) over the conterminous United States. *Climate Dynamics,* DOI 10.1007/s00382-015-2845-1.

Pierce, D. W., D. R. Cayan, E. P. Maurer, J. T. Abatzoglou, and K. C. Hegewisch, 2015: Improved bias correction techniques for hydrological simulations of climate change. *J. Hydromet,* v. 16, p. 2421-2442. DOI: http://dx.doi.org/10.1175/JHM-D-14-0236.1

Pierce, D. W., J. F. Kalansky, and D. R. Cayan, (Scripps Institution of Oceanography). 2018. Climate, Drought, and Sea Level Rise Scenarios for the Fourth California Climate Assessment. California's Fourth Climate Change Assessment, California Energy Commission. Publication Number: CNRA-CEC-2018-006.

Pierce, D. W., and Cayan, D. R., 2019: Future projections of hourly surface temperatures in California. A report to the California Energy Commission, 48 pp.

Premoli, A., and Tavella, P., 1993. A revisited Three-Cornered Hat method for estimating frequency standard instability. IEEE Transactions on instrumentation and Measurement, 42(1), 7–13. https://doi.org/10.1109/19.206671

Sapsis, D., Brown, T. J., Low, C., Moritz, M., Saah, D., Shaby, B., 2016: *Mapping Environmental Influences on Utility Fire Threat*, 55 pp., Final report to the California Public Utilities Commission Pursuant to R.08 - 11-005 and R.15-05-006.

Schreiner-McGraw, A.P., and H. Ajami. Impact of uncertainty in precipitation forcing datasets on the hydrologic budget of an integrated hydrologic model in mountainous terrain. *Water Resources Research, In Review*.

Schreiner-McGraw, A.P., and H. Ajami. Propagating combined precipitation and temperature uncertainties in integrated hydrologic models. *Water Resources Research*, *In Preparation*.

Schreiner-McGraw, A.P. and H. Ajami. Delayed response of groundwater to droughts. *Geophysical Research Letters*, *In Preparation*.

Schreiner-McGraw, A.P., H. Ajami, and J. Palmer. Quantification of aquifer and meteorologic controls on groundwater response to drought. *Water Resources Research*, *In Preparation*.

Shu, L., Ullrich, P and C. Duffy (2019). Simulator for Hydrologic Unstructured Domain (SHUD): Numerical modeling of watershed hydrology with the finite volume method. *Geoscientific Model Development*.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. *NCAR Tech*. Note NCAR/TN-475+STR, 113 pp.,

Vanderstraeten, D. and Keunings, R., 1995: Optimized partitioning of unstructured finite element meshes, *International Journal for Numerical Methods in Engineering*, 38, 433–450, https://doi.org/10.1002/nme.1620380306,

Wood, R. (2012) Stratocumulus Clouds. *Monthly Weather Review*, 140(8), pp.2373-2423

EPC-16-063 Research Team

LOCA Downscaling: David Pierce, Dan Cayan – Scripps Institution of Oceanography (SIO), University of California, San Diego (UCSD)

Wind and Humidity diagnostics: Janin Guzman-Morales, Alexander (Sasha) Gershunov – SIO, UCSD

Modeling and diagnostics of Marine Stratocumulus Clouds: Andrew Martin - Portland State University; Laurel DeHaan, Rachel Clemesha – SIO, UCSD

Hydrological modeling using SHUD and Machine Learning techniques: Paul Ullrich, Lele Shu – University of California, Davis

Ground water modeling using PARFLOW.clm: Hoori Ajami, Adam Schreiner-McGraw – University of California, Riverside