California's coastal fog:

modeled trends in space & time

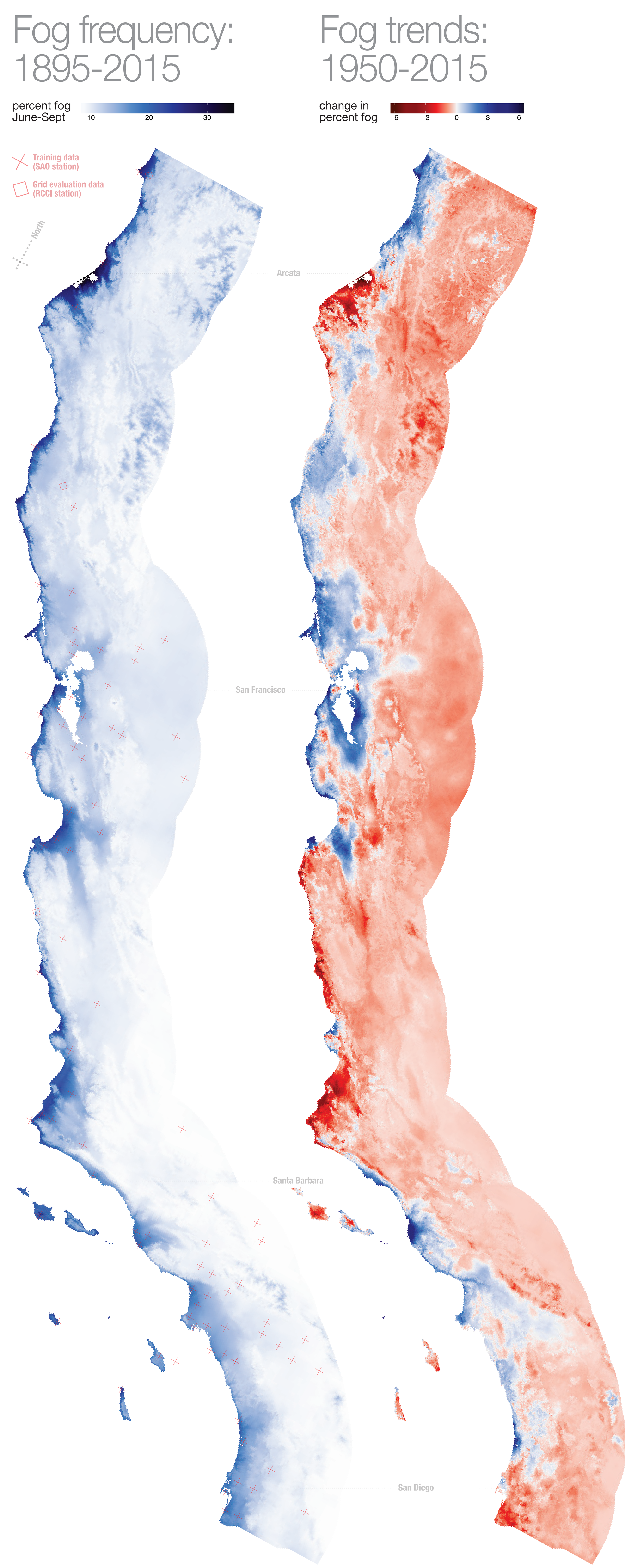
Summertime fog strongly influences human and ecological systems in coastal California, creating a cool, moist microclimate buffered from the hot, dry conditions that dominate the greater region[1-4]. Yet data on spatial and temporal fog patterns is scarce in comparison to data on other climate variables, severely limiting consideration of fog in applications such as ecological modeling or climate trend analyses.

In this study I use airport observations of fog and other climate variables to construct a predictive fog model, demonstrating that widely measured climate variables can be used to predict coastal fog. I then produce a gridded fog dataset with various potential applications, covering all of coastal California and comprising a time series that spans more than a century. Model evaluation indicates that this approach captures the dominant spatiotemporal patterns of monthly fog frequency in this region.

Recent multi-decadal trends in these fog estimates suggest that fog frequency has increased near the coast but declined inland, consistent with an intensification of coastal climate buffering. This helps to address open questions about the impacts of climate change on coastal fog[1]. Still, significant uncertainty remains, particularly in understanding fog patterns in remote areas and potential bias in long-term fog trends.

Matthew Kling
PhD student, UC Berkeley
mattkling@berkeley.edu

Acknowledgements:
David Ackerly, Todd Dawson,
Alicia Torregrosa, Healy Hamilton,
Emily Burns



Model specification and results

[1] Fog quantification

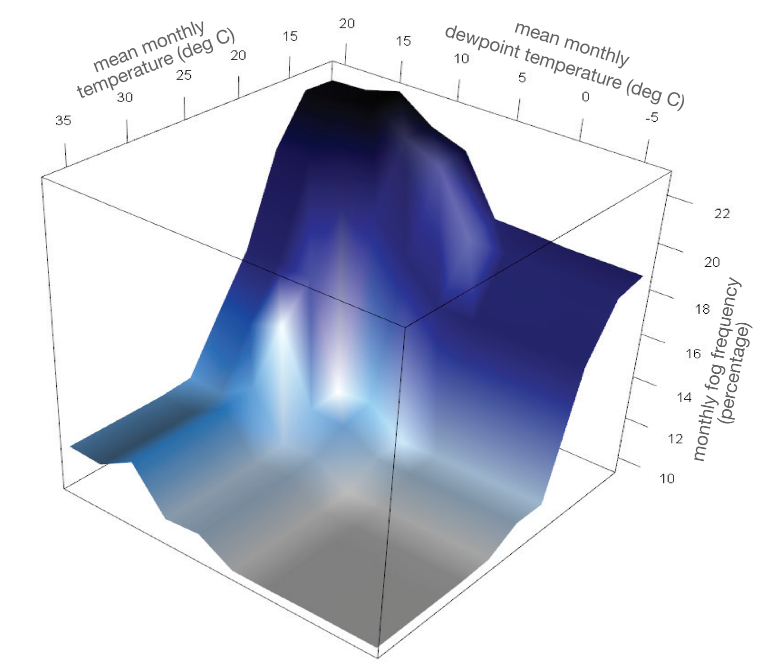
Informally, fog consists of stratus clouds near ground level. Operationally, I define fog as the presence of a cloud ceiling less than 400m above ground, a definition drawn from the literature[2]. I use this criterion to quantify fog as a binary variable for hourly cloud ceiling observations from the Surface Airways Observations (SAO) dataset[5] for 87 airfields in California within 50km of the coast (red crosses in the map at far left). This dataset spans 1935-2016, but observations at most individual stations are far less extensive, limiting its utility in directly assessing long-term fog trends.

For summer months (June-September) I converted hourly fog presence to monthly fog frequency (percent of hours with fog) time series for each airport. Corresponding hourly air temperature and humidity records were summarized in tandem. After discarding months missing more than 5% of hourly observations the final model training dataset included more than 8300 station-months, each with 6 climate variables: monthly fog frequency, monthly averages of daily minimum, mean, and maximum temperature, monthly mean dewpoint temperature, and monthly mean dewpoint depression. Dewpoint depression is the difference between air temperature and dewpoint temerature, and has strong physical ties to vapor condensation.

[3] Predictive model fitting

After cross-validation, a final random forest regression model was fit using the entire SAO training dataset, predicting monthly fog frequency based on the other four climate variables. Random forests are a widely used class of nonparametric machine-learning algorithm, and performed better in testing than generalized linear regression.

The chart below shows modeled fog frequency as a function of monthly mean temperature and dewpoint, a 2D projection of the complete 5-dimensional response surface. Both variables and their interaction are captured nonparametrically in the model. All 5 predictor variables improved accuracy and were retained during model selection.



[5] Gridded fog frequency prediction

The above statistical model trained on airport observations was used to predict monthly fog frequency for every 800m PRISM grid cell in California within 50km of the coast, for each of the four summer months from 1895 to 2015. The map at far left shows the mean of these predictions across the entire 484-month time series.

[8] Multi-decadal trend measurement

For each 800m grid cell I used ordinary least squares regression to estimate the trend in fog frequency across the 264 summer monthly values from 1950 to 2015. This timeframe was chosen to avoid higher-uncertainy PRISM data in the early 1900s and to emphasize trends associated with anthropogenic climate change. The map at left shows change in fog frequency between the beginning and end of those trend lines (expressed in percent of total summer hours, not percent of initial fog frequency).

These results suggest that fog declines in Monterey and Aracata, which have been rasied as evidence of regional coastal fog declines[2], may not be broadly representative. Significant heterogeneity is apparent in the direction and magnitude of estimated trends, with fog increasing along much of the coastline and decreasing farther inland.

This pattern would signify an intensification of the coastal fog gradient, hinting that increasing inland temperatures (demonstrated separately from this study) may be intensifying coastal upwelling-driven climate buffering[1,8], increasing onshore advection of sea fog but also dissipating it more rapidly as it progresses inland.

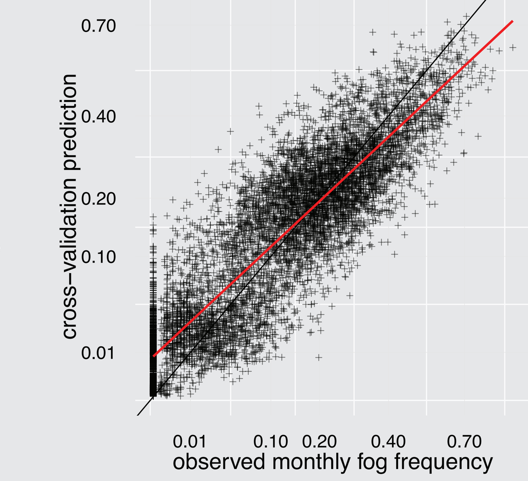
Model evaluation

[2] Monthly fog predictability

Physical first principles govern how temperature, pressure, and humidity relate to fog at an instantaneous timescale[3]. But statistics, not physics, governs whether those correlations hold at the monthly timescale necessary to predict fog from widely available monthly climate data sources.

Cross-validation within the airport dataset was used to evaluate the predictability of monthly summertime fog frequency from temperature and dewpoint. Observed fogginess for each airport-year was tested against the predictions of a random forest model trained on observations that had neither airport nor year in common with the testing data, minimizing false optimism from spatial and temporal autocorrelation.

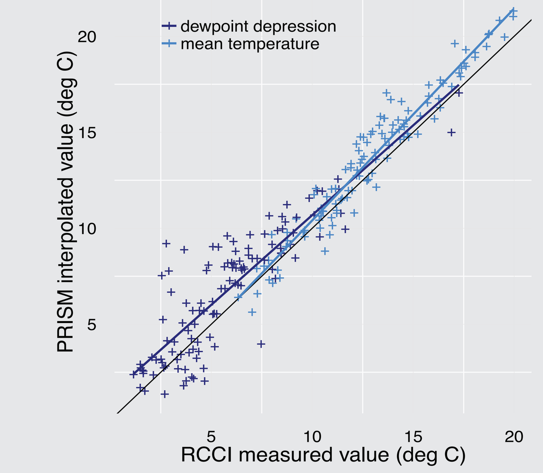
This validation yielded an r^2 of 0.78, illustrated at right, suggesting summer fog frequency is largely predictable from other climate variables in coastal California. While this simple model captures a large majority of the spatiotemporal variablity in fog frequency, it tends to overpredict the least foggy datapoints and underpredict the most foggy, indicating room for further model refinement.



[4] Gridded predictor data quality

Using this model to generate wall-to-wall fog predictions requires gridded predictor variables. Among the most widely used is PRISM, which comprises monthly time series of 800m grids interpolated from weather station data and is tailored to coastal California[6]. Uncertainty in interpolated PRISM data contributes to uncertainty in the maps at left. But virtually all publicly available weather station data is used in the creasion of PRISM[6], making evaluation with independent data a challenge.

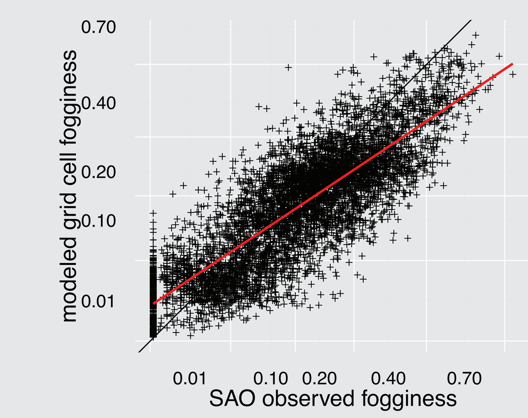
I used records from meterological stations installed in the tops of redwood trees by the Redwoods and Climate Change Initiative (RCCI)[7] at three remote loactions (red diamonds in the map at far left) to assess the reliability of PRISM data far from the weather stations used to create it. This comparison is shown at right, with a point for each redwood weather station for each month from 2011-2015. Decomoposing this overall correlation into spatial, seasonal, and interannual components still shows strong correlations in each dimension. This indicates PRISM interpolation performs well, but it does not address the reliability of multi-decadal trends in PRISM.



[6] Model predictions at airport locations

I compared gridded fog frequency predictions to observed SAO fog frequencies at the locations of airport weather stations. An r^2 of 0.69 for this correlation across all station-months, depicted at right, suggests that the majority of spatiotemporal variablity in fog frequency is captured by the model. Predictive accuracy far from airports is not evaluated by this test.

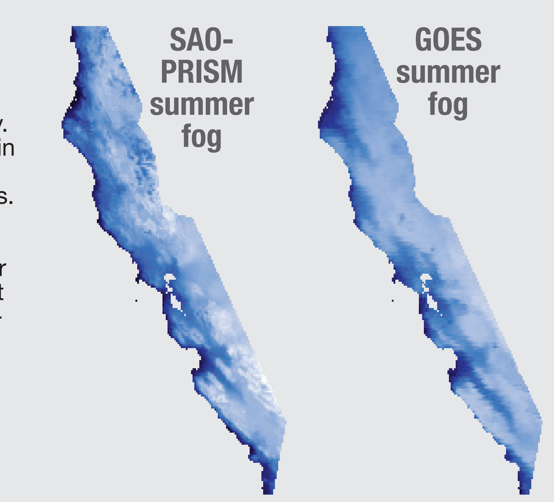
The predictive error in this relationship reflects climatic variability within PRISM grid cells and inaccuracy in PRISM interpolation, as well as uncertainty in modeled fog-temperature-humidity relationships. As in the cross-validation analysis above, the model tended to predict less extreme fogginess than observed, suggesting accuracy could be further improved.



[7] Comparison to satellite imagery

Beyond station-based fog observations like the SAO data used to fit this model, the best source of California fog data comes from satellite imagery. The maps at right compare the average summer fog frequency predicted in this analysis to a GOES-based low cloud frequency dataset[4] for northern and central California, each normalized to highlight relative spatial patterns.

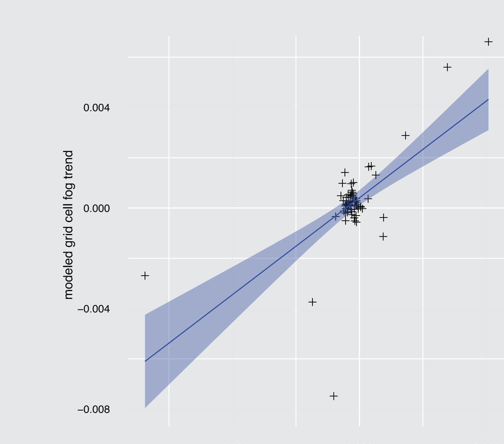
The dominant coastal gradient and the regions of highest and lowest fog frequency are clearly captured by both datasets. GOES data shows higher low cloud frequency overall and some differences in spatial pattern—but it captures a much broader vertical spectrum of stratus clouds than the narrow sub-400m fog category addressed in this study[4], making direct detailed comparison inappropriate.



[9] Evaluation of fog trends

Multi-decadal trends based on PRISM data may incude non-climatic artifacts from changes in weather station operations. Comparing the fog frequency trends presented here to trends in direct SAO fog measurements may help to clarify this uncertainty, although SAO trends are subject to similar biases since cloud ceiling measurement transitioned from eyeballed estimates to laser measurement during the study timeframe.

The figure at right compares the statistically significant subset of trends mesured directly on SAO observations to corresponding trends in grid cell predictions. Data were removed from the modeled time series to match gaps in airport observations. A clear positive correlation exists in trends across these station-months, suggesting that at locations with airports, multi-decadal fog trends roughly agree across the two datasets. However, error propagating from upstream stages of the modeling process means that bias and uncertainty in these fog trends remain difficult to quantify.



[1] Torregrosa, A., O'Brien, T. A., & Faloona, I. C. (2014). Coastal fog, climate change, and the environment. *Eos, Transactions, American Geophysical Union*, 95(50), 473-474.
[2] Johnstone, J. A., & Dawson, T. E. (2010). Climatic context and ecological implications of summer fog decline in the coast redwood region. *Proceedings of the National Academy of Sciences*, 107(10), 4533-4538.
[3] Koračin, D., Dorman, C. E., Lewis, J. M., Hudson, J. G., Wilcox, E. M., & Torregrosa, A. (2014). Marine fog: A review. *Atmospheric Research*, 143, 142-175.
[4] Torregrosa, A., Combs, C., & Peters, J. (2016). GOES-derived fog and low cloud indices for coastal north and central California ecological analyses. *Earth and Space Science*.
[5] U.S. Global Surface Airways Hourly Observations. *National Climatic Data Center, NESDIS, NOAA, U.S. Department of Commerce*.
[6] Daly, C., Gibson, W. P., Taylor, G. H., Johnson, G. L., & Pasteris, P. (2002). A knowledge-based approach to the statistical mapping of climate. Climate research, 22(2), 99-113.
[7] RCCI forest climate monitoring data. (2016). *Save the Redwoods League*, personal communication.
[8] Bakun, A. (1990). Global climate change and intensification of coastal ocean upwelling. *Science*, 247(4939), 198-201.