Title
Complexity in Climatic Controls on Plant Species Distribution: Satellite Data Reveal Unique Climate for Giant Sequoia in the California Sierra Nevada

Author
Waller, Eric Kindseth

Publication Date
2014

Peer reviewed|Thesis/dissertation

Complexity in Climatic Controls on Plant Species Distribution: Satellite Data Reveal Unique Climate for Giant Sequoia in the California Sierra Nevada

By

Eric Kindseth Waller

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Environmental Science, Policy, and Management

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Dennis D. Baldocchi, chair
Professor Max A. Moritz
Professor David D. Ackerly

Fall 2014

ABSTRACT

Complexity in Climatic Controls on Plant Species Distribution: Satellite Data Reveal Unique Climate for Giant Sequoia in the California Sierra Nevada

by

Eric Kindseth Waller

Doctor of Philosophy in Environmental Science, Policy, and Management

University of California, Berkeley

Professor Dennis D. Baldocchi, Chair

A better understanding of the environmental controls on current plant species distribution is essential if the impacts of such diverse challenges as invasive species, changing fire regimes, and global climate change are to be predicted and important diversity conserved. Climate, soil, hydrology, various biotic factors (e.g., competition, disease, dispersal), fire, history, and chance can all play a role, but disentangling these factors is a daunting task. Increasingly sophisticated statistical models relying on existing distributions and mapped climatic variables, among others, have been developed to try to answer these questions. Any failure to explain pattern with existing mapped climatic variables is often taken as a referendum on climate as a whole, rather than on the limitations of the particular maps or models. *Every* location has a unique and constantly changing climate (challenging our definition of climate), so that *any* distribution *could* be explained by some aspect of climate. This is not a rationale for doing so – such models would certainly be overly fit to the given conditions and not generalizable. It is an argument for a more complete evaluation of possible climatic controls and the interaction of climate with other important variables (e.g., soil). A lack of adequate maps for various climate and climate-derived variables plays a large role in the failure of modelers to sufficiently address climatic complexity. In particular, site water balance, radiation, humidity, wind, and temporal variability in all of these factors may be poorly understood factors in controlling distributions. Weakness in the mapping of these variables is well recognized in the water balance modeling field, but is less emphasized in the species distribution modeling field, despite the fact that variables that affect the water balance are also likely to play a major role in species distribution. In this dissertation, I 1) improve the mapping of cloud cover from satellite imagery, in order to generate accurate monthly cloud frequency variables; 2) use these cloud frequency variables to demonstrate the possible importance of cloud cover, a previously overlooked climate variable, to the distribution of a particularly charismatic plant species, giant sequoia; and 3) further investigate the climate associated with the frequent cloud cover in the vicinity of giant sequoia, to identify other climatic factors that may also be important to the tree's distribution.

Chapter 1 of this dissertation reviews some of the major flaws in species distribution modeling (with existing climate data) and addresses concerns that climate may therefore not be predictive of, or even relevant to, species distributions. Despite problems with climate-based models, climate and climate-derived variables still have substantial merit for explaining species distribution patterns. Additional generation of relevant climate variables and improvements in other climate and climate-derived variables are still needed to demonstrate this more effectively. Satellite data have a long history of being used for vegetation mapping and even species distribution mapping. They have great potential for being used for additional climatic information, and for improved mapping of other climate and climate-derived variables.

Improving the characterization of cloud cover frequency with satellite data is one way in which the mapping of important climate and climate-derived variables (e.g., water balance) can be improved. An important input to water balance models, solar radiation maps could be vastly improved with a better mapping of spatial and temporal patterns in cloud cover. Chapter 2 of this dissertation describes the generation of custom daily cloud cover maps from Advanced Very High Resolution Radiometer (AVHRR) satellite data from 1981-1999 at ~5 km resolution and Moderate Resolution Imagine Spectroradiomter (MODIS) satellite reflectance data at ~500 meter resolution for much of the western U.S., from 2000 to 2012. Intensive comparisons of reflectance spectra from a variety of cloud and snow-covered scenes from the southwestern United States allowed the generation of new rules for the classification of clouds and snow in both the AVHRR and MODIS data. The resulting products avoid many of the problems that plague other cloud mapping efforts, such as the tendency for snow cover and bright desert soils to be mapped as cloud. This consistency in classification across cover types is critically important for any distribution modeling of a plant species that might be dependent on cloud cover.

In Chapter 3, monthly cloud frequencies derived from the daily classifications were used directly in species distribution models for giant sequoia (*Sequoiadendron giganteum* (Lindley) Buchholz) and were found to be the strongest predictors of giant sequoia distribution. A high frequency of cloud cover, especially in the spring, differentiated the climate of the west slope of the southern Sierra Nevada, where giant sequoia are prolific, from central and northern parts of the range, where the tree is rare and generally absent. Other mapped cloud products, contaminated by confusion with high elevation snow, would likely not have found this important result. The result illustrates the importance of accuracy in mapping as well as the importance of previously overlooked aspects of climate for species distribution modeling. But it also raises new questions about why the clouds form where they do and whether they might be associated with other aspects of climate important to giant sequoia distribution. What are the exact climatic mechanisms governing the distribution? Detailed aspects of the local climate warranted more investigation.

Chapter 4 investigates the climate associated with the frequent cloud formation over the western slopes of the southern Sierra Nevada: the “sequoia belt”. This region is climatically distinct in a number of ways, all of which could be factors in influencing the distribution of giant sequoia and other species. Satellite and micrometeorological flux tower data reveal characteristics of the sequoia belt that were not evident with surface climate measurements and maps derived from them. Results have implications for species distributions everywhere, but especially in rugged mountains, where climates are complex and poorly mapped.

Chapter 5 summarizes some of the main conclusions from the work and suggests directions for related future research.

Table of Contents

Acknowledgments

I am extremely thankful for the tremendous assistance that I have received throughout this dissertation. First and foremost I would like to thank the team of Max Moritz and Dennis Baldocchi for shepherding me through this rather extended, yet very fruitful, period. Each stepped up in a major way during a couple critical times of need. Max gave me space, both literally and figuratively, to pursue my ideas, and was incredibly patient. At the same time, he was always willing to poke and prod my ideas and provide constructive feedback. Max also helped create a great work environment, with a consistently fun and savvy group of labmates. Dennis helped get me through these last few years, both by taking over in an advisory capacity and financially with funding for a fog mapping project that ended up providing the foundation for key areas of this work. I'm glad that Dennis could get as excited about cloud cover patterns as I, and that he was patient with my perfectionism in getting the cloud classification just right. He was also instrumental in improving my communication, even though many of the figures in my dissertation still might not meet his standards.

I would also like to thank David Ackerly for his important role as my outside committee member. He provided incredibly constructive ideas whenever required, and gave me quick turnaround and incisive feedback that proved very helpful for my eleventh hour finish. He is a model faculty member, even if he's not in the ESPM department. Other faculty from my qualifying exam that have my tremendous appreciation include Richard Dodd and Perry DeValpine. Richard was always willing to chat about giant sequoia and other topics in the hall. Perry was an inspiration for his intellectual rigor – I got the most out of his classes of any in the program. Other faculty that provided useful feedback over the years include Sally Thompson, Todd Dawson, Allen Goldstein, and Greg Biging. The Carlson lab and the Huntsinger / Bartolome "Range" lab provided fun alternative field study and social homes when needed. The ESPM department and staff were very supportive in social, academic, and financial matters. The College of Natural Resources' Geospatial Innovation Facility (GIF) provided helpful research tools and services.

I had tremendous labmates over the years, all of whom I consider close friends. Marc Parisien, Meg Krawchuk, Tadashi Moody, Erica Newman, Mike Mann, and Enric Batllori have all been great to work and play with over the years. Graduate students in ESPM also helped provide sanity checks. From the cohort, I'd like to specifically acknowledge the friendship of Kevin Krasnow, Melina Kozanitas, Steve Bayes, Shasta Ferranto, Alison Forrestel, and Kyle Holland. And I can't not mention other ESPMites Erica Spotswood, Sheri Spiegal, Alan DiVittorio, Kate Wilkin, Stella Cousins, and Lauren Hallett, both for their friendship and for academic discussion related to my research. Brad Balukjian wasn't in my cohort and didn't discuss my research, but he did provide a very entertaining introduction to my closing talk and some fun trips to Santa Barbara.

Other researchers who have provided a lot of dissertation-related help include folks from NASA Ames (Rama Nemani, Jennifer Dungan, Petr Votava, and Sangram Ganguly) and those involved in the Southern Sierra Critical Zone Observatory project (Roger Bales, Michael Goulden, and Matt Meadows). I've also collaborated with several researchers focusing on coastal

fog – Alicia Torregrosa and Ian Faloona have always been very willing to chat about various research topics. The California Energy Commission (Guido Franco, in particular) was instrumental in providing funding for the fog mapping project which was very helpful for my pursuit of cloud cover mapping, in general. Microsoft also provided funds that supported this research.

Outside of the ivory tower, I am very thankful for my parents, Deb and Gene, for putting up with and supporting me all these years. I'm glad that they're excited about the PhD (Phinally!) and actually reading it. I'm also super happy that I've been able to share a lot of common interests and adventure with my brother Nat, the Guero Jardinero, who is in the Lofoten Islands of Norway as I write this. I may have to go there and study their cloud patterns.

Chapter 1. Introduction

Complexity of Climate Controls on Plant Species Distributions

This dissertation involves the use of a variety of sources, including satellite and meteorological tower data, to reveal overlooked climatic controls on the distribution of a notable plant species, highlighting the need for greater exploration of climatic dimensionality in order to understand why plants grow where they do. Giant sequoia (*Sequoiadendron giganteum* (Lindley) Buchholz) of the California Sierra Nevada, with a very well mapped yet previously puzzling distribution (Hartesveldt *et al*, 1975; Harvey *et al*, 1980), exemplifies the lack of knowledge about climatic control, even for the most charismatic of species. Giant sequoia is found to have a unique and unusually cloudy climate. Cloud frequency and other important aspects of this distinct climate would not be included in typical distribution models used to project plant response to climate change. There are certainly a variety of biotic and abiotic controls on the current distribution of plant species, not to mention historical factors and chance, but essential aspects of current climate are neglected both by relatively simplistic statistical analyses relating species distributions to conventional climate variables like temperature and precipitation and by those who justifiably critique such models but then downplay climate's role in plant species distribution.

This introductory chapter addresses some recent issues in the field of plant species distribution modeling (SDM) and sets the stage for improved treatment of climate variables. It is well established that plants respond to a complex integration of environmental factors (Walter, 1973; Chapin III *et al*, 1987; Woodward, 1987; Givnish, 2002; Wright *et al*, 2004). It is logical that species distributions, too, would have a complex relationship with climate and the overall environment. Extensive measurement of plant physiological response to environmental variation is still needed to improve our understanding of controls on distribution and to project response to future climate change. Assessment with respect to complex environmental variables that integrate a variety of factors, such as those related to the water balance (Thornthwaite and Mather, 1957), should be especially beneficial in the understanding of controls on plant species performance and distribution (Austin, 2002). Water balance variables such as evapotranspiration (ET) and deficit have been shown to be strongly predictive of broad patterns of vegetation (Stephenson, 1990) and species (Gavin and Hu, 2006) distributions. But there is still much to be improved in the mapping of relevant components of these important integrating variables in order to move from statistical correlations to accurate forecasting of species and ecosystem response. Accurate characterization of the contribution of soil (depth, texture, etc.) is likely the most relevant and the most challenging (Lathrop Jr *et al*, 1995; Schenk and Jackson, 2002; Schenk and Jackson, 2005; Ichii *et al*, 2009) – it may be problematic for some time. In the meantime, improved mapping of other important components of the water balance such as solar radiation, humidity, and wind is a more manageable goal.

The improved accuracy and resolution of cloud cover classification from satellite imagery described in Chapter 2 of this dissertation should help to improve spatial modeling of solar radiation. It could also help with the modeling of spatial patterns of humidity, due to the

obvious correlation between cloud cover and humidity (Walcek, 1994; Tompkins, 2003), or at least guide the siting for increased measurement of surface humidity. The same may even be argued for the improved modeling of wind, as there appears to be some negative correlation between wind and cloud cover, at least in some areas (Singer and Smith, 1953). Each of these variables is an important input to complex ET models (Monteith, 1965) and ET is a critical element of the water balance. Until improvements are made to the handling of soil, also necessary for accurate ET estimation, these variables – solar radiation, humidity, and wind - are worth exploring individually for their utility in modeling species distribution.

In Chapter 3, monthly cloud frequency variables derived from the custom cloud cover classification developed in Chapter 2 are shown to be the strongest predictors of giant sequoia distribution from a large set of climate variables. This would not have been demonstrated without the critical improvements in resolution and accuracy of the custom cloud frequency products. Some portion of cloud cover's apparent importance with respect to giant sequoia distribution may relate to its better mapping compared to other climate variables, but it is clear that the west slope of the southern Sierra Nevada, where giant sequoia is very common, is very cloudy, especially in spring, in comparison to the northern Sierra Nevada, where giant sequoia is quite rare, if not absent.

Further investigation of the climate of the main giant sequoia "belt" on the west slope of the southern Sierra Nevada using satellite data, mapped data sets, and meteorological towers in Chapter 4 revealed many other related and unusual climatic features of the region. The cloud product helped expose these climatic characteristics that may have otherwise escaped notice, revealing that cloud cover may just be one part of an overall climate environment amenable to giant sequoia dominance. Determining the precise controls would require further investigation, including improvements in the mapping of other variables. In any case, the results demonstrate that the relationship between climate and plants is complex, that accuracy and resolution are important in identifying climatic controls, and that there are essential aspects of climate not included in standard species distribution models.

Background

It has long been recognized that climate plays a large role in the distribution of plant species (von Humboldt and Bonpland, 1807). Temperature and precipitation are obvious aspects of the climate that likely influence species distributions. But determining the exact manner in which temperature and precipitation affect distribution is not so easy. Simply circumscribing the minimum and maximum levels of each with respect to a species' distribution would vastly overestimate actual distributions. This is due to the many other constraining factors, as well as the fact that the interaction of temperature and precipitation probably matters more than individual values of either. These are just some of the reasons that a simple treatment of temperature and precipitation will not be sufficient to describe the distribution of plant species. Nevertheless, improvements in computing power and digital data (e.g., species locations and climatic maps), as well as concerns about climate change, have led to a proliferation of simple

(mostly temperature- and precipitation-based) models relating species locations to environmental variables that are assumed to capture the "niche" (Hutchinson, 1959) of species. [1]

These "niche" requirements are often applied under modeled future climates to project the future distribution of a given species. Models used for projecting into the future are often based only on temperature and precipitation, because these are the only variables that are projected into the future (and usually at only coarse spatial and temporal resolution). The generally more restricted extents of predicted future distributions are usually quite alarming. Problems with this type of approach have been identified (Davis *et al*, 1998; Austin, 2002; Austin, 2007; Pearson and Dawson, 2003; Guisan and Thuiller, 2005; Guisan *et al*, 2006; Randin *et al*, 2006; Austin and Van Niel, 2011), and include admonitions to use meaningful climatic variables that are thought to strongly relate to the distribution of the species. Supposedly more biologically meaningful climatic variables of common resort (e.g., BIOCLIM) include those related to combinations of temperature and precipitation, such as average temperature of the wettest quarter (3 months) and driest quarter.

This type of attempt at capturing the interaction of precipitation and temperature is still far from capturing the complexity of the overall environment governing the distribution of plant species. Variable selection in SDM is a challenge (Araujo and Guisan, 2006; Austin and Van Niel, 2011; Barve *et al*, 2011) and likely related to a lack of well-mapped environmental variables with a more mechanistic relationship with plant distribution (Gavin and Hu, 2005). A more sophisticated and plant-relevant treatment of the interaction between precipitation and temperature might involve the use of water balance data (Stephenson, 1990; Gavin and Hu, 2006; Franklin *et al*, 2013). But there remains the question of whether more sophisticated variables related to the water balance are mapped well enough to be of robust utility. As mentioned, accurate water balance modeling requires good maps of soil characteristics and a good understanding of the way in which soil influences the interaction of environmental variables.

The fact that climate and climate-based derivatives (e.g., the water balance) are likely influencing many species distributions and climate-based distribution models appear to perform well does not mean that most climate-based models have robust predictive ability under changing conditions. Many SDM approaches with climate data are sufficient to make good *maps* of species distributions but a good map does not ensure good prediction. Several researchers (Bahn and McGill, 2007; Beale *et al*, 2008; Chapman, 2010) have shown that common approaches to SDM may simply be taking advantage of autocorrelation or spatial structure of the climate variables to yield apparently good results (see below). The apparent confirmation of good results on test data in most studies is due to the associated lack of independence of testing data (Hampe, 2004; Hijmans, 2012). The application of such models under projected future climates can be misleading. But this weakness should not be grist for those who would dismiss the importance of climate. The fact that many climate-based models have obvious problems does not mean that climate and climate-based derivatives do not influence many species distributions.

[1] The use of "niche" is admittedly a stretch, and not even very clear, given the difficulty for modelers in determining whether they are modeling the "fundamental" niche or the "realized" niche (Araujo and Guisan, 2006). That it is always closer to the latter, even when claims are made otherwise, seems obvious.

Addressing a Critique of Species Distribution Modeling

While the impact of autocorrelation on ecological studies has long been recognized (Legendre, 1993), only recently did researchers call attention to the fact that much SDM relied on the spatial structure of climate variables for models that made good maps but were not necessarily useful for extrapolating to different conditions, such as those associated with future climates. Bahn and McGill (2007) somewhat surprisingly found that simple spatial models (i.e., interpolation) performed better than environmental variables at modeling the distribution of various bird species. Ignoring the fact that spatial models would not be able to explain spatially-outlying populations that climate-based distribution models are often able to explain (Currie, 2007), their results raised some flags.

Other recent studies (Beale *et al*, 2008; Chapman, 2010) used null models to demonstrate that species distribution models (for birds and plants, respectively) using climatic variables often perform little better than chance. In the former case, artificial species distributions with the same prevalence and spatial structure as real distributions form null model distributions. In the latter case, artificial climate gradients with the same values and spatial structure of real climate variables were used for null distribution models. In each case, the fact that "real" models did not tend to perform significantly (or any) better than null models is used as evidence that many models purportedly relating species distributions to climate variables are misleading.

Beale et al.'s work, in particular, stirred a strong reaction and prompted a series of rebuttal attempts (Araujo *et al*, 2009; Aspinall *et al*, 2009; Peterson *et al*, 2009) because most ecologists have the strong sense that climate influences species distributions. These rebuttals were not necessarily wrong, but they did not refute Beale et al., as they were easily parried (Beale *et al*, 2009). This is because the approaches and arguments of the distribution modeling critics are fairly sound, if slightly misleading. Beale et al. (as well as the others) employed typical, and bad, practices of SDM to demonstrate its problems. But the fact that models that use certain coarse-scale climate variables are bad (and they can be, as shown by Beale et al.) does not exclude climate as a control on distributions. These researchers included the caveats that they are just using the coarse-scale climate variables that other studies have commonly used. But these are at spatial resolutions of 10 and 50 km! And they are very careful with their wording that "*evidence* for climatic limitation… may have been overstated" (Chapman, 2010) or "distributions of most birds in our study are not strongly associated with climate variables *currently available*" (Beale *et al*, 2008) – italics added. These authors are not definitively arguing that climate is unimportant. They are criticizing modeling approaches and indicating that climate has not been *proven* to matter at *macro-scales*, but these nuances could be overlooked by those looking to dismiss climate.

These critiques of climate-based models, while valid, have not adequately emphasized the issue that commonly used predictor variables simply don't cover the complexity of climate-driven control. Araujo et al. (Araujo *et al*, 2009) tried to get at this issue with their rebuttal, but their climate variables alone were likely not good enough to make the case for climate-driven control. Rather, those who would emphasize non-climatic factors such as dispersal, geographic variation in adaptation, and other biotic factors (Hampe, 2004) are bolstered. There is nothing inherently wrong with considering these other possibilities, but alternative modeling approaches

– for example, partitioning biotic, abiotic, and movement factors - the "BAM" framework (Soberon and Nakamura, 2009), endorsed by Beale et al. (2014), could underestimate the importance of climate, especially with a failure to focus on the right climate and climate-based variables. This is similar to notions of hierarchical controls on species distributions, with different factors allegedly operating at different scales, and climate ostensibly only operating at relatively large scales (Pearson and Dawson, 2003). Lennon (Lennon, 2000) even argues that the supposedly high autocorrelation of climate variables (particularly "temperature-related climatic factors" of actual and potential evapotranspiration) contributes to their misleading importance in distribution models. He argues that more "spatially discontinuous factors" (i.e., not climate) are given too little import.

In fact, climate and its water balance (e.g., evapotranspiration) derivatives are quite variable spatially. Climate, to the extent that such a (fixed) thing exists, is unique everywhere (especially considering wind, radiation, and humidity patterns) and these notions of scales of climate relevance are simple conceptualizations. Climate doesn't just consist of the occasional microclimate within a macroclimate. Since every location has a unique climate in some respect, climate *could* be used to explain every distribution. This does not mean that climate variables *should* be used to explain every distribution - such models would almost certainly be overfit (i.e., strong fit to the particulars of the given data, with poor performance on independent data). But it's an argument for a more thorough evaluation of climate before dismissing or downplaying climate's relevance.

Climate-based distribution models capture the realized niche of plants as defined by their climate space. Yet climate is generally operating in tandem with soil, topography, hydrology, disturbance, and biotic factors (at least) to shape this realized niche. Models attributing all of this to climatic control will likely be overfit and unable to track changes in climate that are not matched by concomitant changes in other variables. Hence the argument is made for using water balance variables; these (ideally) integrate soil, topography, and hydrology with climate. While disturbance and biotic factors (e.g., competition) are often invoked as reasons that climate-based distribution models are overfit and would fail under changing conditions, one could argue that these relationships may be largely governed by the water balance as well. Ultimately, the chosen climatic variables or climate-based derivatives may not need to be mechanistic for distribution models to be useful, but the more closely correlated they are with variables that truly matter, the better (i.e., useful for future projection). Soil, topography, and hydrology would likely be the main covariates of climate that could affect this correlation and future projections.

Satellite Data and Plant Species Distribution Modeling

Satellite data have the potential to improve many of the challenging aspects of distribution modeling. They have been used extensively for vegetation mapping and can provide maps at the species level in some circumstances. Climate and other ancillary data are often incorporated into the mapping to improve discrimination. Switching the roles, satellite data can help to improve species distribution models based on climatic and topographic data (Kerr and Ostrovsky, 2003; Bradley and Fleishman, 2008; Buermann *et al*, 2008). For example, satellite-derived indices of vegetation such as the Normalized Difference Vegetation Index (NDVI; (Tucker, 1979)) can help to identify areas of vegetation. Furthermore, vegetation indices can be

paired with satellite-derived land surface temperature (LST) data, often in combination with meteorological data, to characterize surface ET and the productivity of vegetation (Mu *et al*, 2007; Mu *et al*, 2011). Time series of these values can also be used to determine the phenology or seasonality of vegetation that can be helpful for species identification. But these approaches serve to improve plant species distribution *mapping* more than they necessarily improve *understanding* of the underlying controls on the distribution – in fact, they could serve as a crutch to make a good map and reduce a more "mechanistic" understanding. They do not provide a model that could conceivably be projected into the future.

However, the same information could be tapped for improved understanding. Spatial and temporal patterns of vegetation, NDVI, and ET could be used to facilitate the development of better models of their control – making sure to avoid the circularity that could be introduced from land-cover characteristics influencing models of land-cover. Vegetation productivity and phenology could be used as indirect indicators of climatic patterns and annual meteorological conditions. Alternatively, soil and hydrological characteristics such as rooting depth could be considered the effective residual in a model relating plant productivity, as measured from satellite, and climate (Ichii *et al*, 2009). These results could be used to improve simple water balance models which have already been shown to predict broad patterns in plant functional type (Stephenson, 1990) and species (Gavin and Hu, 2006). It is likely that the models could explain fine-scale patterns in species distribution with improved resolution in mapping and improved inputs provided by a sophisticated analysis of satellite data.

Atmospheric properties characterized by satellite are another potential source of improvement in mapping. In addition to LST, a better characterization of atmospheric properties (e.g., temperature and humidity) relevant to plants may now be possible with satellite data. Clouds are an obvious atmospheric property that can be mapped - satellite data have been used to generate standardized cloud climatologies for the world, but only at a coarse resolution (e.g., International Satellite Cloud Climatology Program (Schiffer and Rossow, 1983)). The coarse resolution of these products has likely discouraged their use in plant SDM. More recently, relatively fine-scale (~1 km) mapping of clouds and/or fog has been used to predict or explain plant species distributions (Fischer *et al*, 2009) and mortality (Baguskas *et al*, 2014) across limited areas. Extending the relatively high resolution classification of clouds from satellite data in space and time (Chapter 2) to help explain species distribution – in particular, giant sequoia (Chapter 3), is a primary focus of this dissertation.

Meteorological Data from Satellite and Tower

In Chapter 4 I use additional satellite, map, and meteorological tower data to investigate the climate of the giant sequoia. MODIS thermal data, in particular, reveal spatial patterns in land surface temperature that would not be captured with the existing sparse network of climate stations. The "Sequoia belt" has among the lowest range of average day-to-night (diel) LST in California, including coastal areas. This observation was so surprising that it was suspected that atmospheric or land surface characteristics other than temperature were affecting the measurements. Even the surface climate station data that do exist in the area did not appear to confirm the satellite result. Only with measurements from the tops of 55 meter tall meteorological towers located at the northern edge of the "Sequoia belt" were similar climatic

patterns observed. Temperatures measured at the top of a tower at 2015 meters (closest to the "Sequoia belt") recorded *lower* ranges in diel temperature than even the satellite data. This related mostly to unusually low daytime temperatures, especially in the spring months. When combined with unusual afternoon peaks in dew point temperature (essentially, absolute humidity), this site had extremely high afternoon relative humidity values at height, especially in the spring. This coincides with the spring peak in cloud cover in this vicinity, and is obviously related, but illustrates that there is more to the unusual climate than just cloud cover. The combination of high humidity, cloud cover, and extremely low average winds, also characteristic of the area, suggests relatively low evaporative demand.

Observations of Chapter 4 also suggest the importance of climate data at surprisingly fine temporal resolution. Climate-based distribution models are often criticized for their use of annually averaged and summed values of temperature and precipitation, respectively. Generally, the criticism is that finer temporal resolution (e.g., monthly or quarterly) should be considered, as well as extremes and variability (Easterling *et al*, 2000), since species don't necessarily respond to average conditions. The half-hourly measurements of the meteorological towers here reveal that climatic patterns at temporal resolutions finer than even daily may be important as well. For example, regular peaks in relative humidity in the afternoon (found at the 2015 meter tower), when plants are active, would have different effects than regular pre-dawn peaks. This climatic behavior would not be seen in a daily record of minimum and maximum relative humidity, let alone longer time scales.

Implications for Plant Species Distribution Modeling

This research demonstrates the complexity of the giant sequoia climate, showing that the dimensionality of plant-relevant climate is obviously high. Interactions with other factors make the relationship between climate and plants even more complex. It would be prohibitively difficult to try to identify the unique characteristics of the environment for every species. This is another argument for data reduction in the form of optimally integrating variables. In any case, not all plant species will have a particular climate as unique as that for giant sequoia. Giant sequoia may have a globally unique climate. Nevertheless, this work indicates that improved identification and mapping of a range of plant-relevant climate variables may help to explain the distribution of many species, even without a perfect understanding of optimally integrating variables. While climate's relevance to plant species distribution occurs in tandem with many other variables (making climate-based predictions alone potentially problematic), the difficulty in disentangling these factors should not be confused with the notion that climate does not matter to plant species distribution.

References

Araujo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677-1688.

Araujo, M.B., Thuiller, W. & Yoccoz, N.G. (2009) Reopening the climate envelope reveals macroscale associations with climate in European birds. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, E45-6; author reply E41-3.

Aspinall, R.J., Miller, J.A. & Franklin, J. (2009) Calculations on the back of a climate envelope: addressing the geography of species distributions. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, E44; author reply E41-3.

Austin, M. (2007) Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological Modelling*, **200**, 1-19.

Austin, M.P. & Van Niel, K.P. (2011) Improving species distribution models for climate change studies: variable selection and scale. *Journal of Biogeography*, **38**, 1-8.

Austin, M. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101-118.

Baguskas, S.A., Peterson, S.H., Bookhagen, B. & Still, C.J. (2014) Evaluating spatial patterns of drought-induced tree mortality in a coastal California pine forest. *Forest Ecology and Management*, **315**, 43-53.

Bahn, V. & McGill, B.J. (2007) Can niche-based distribution models outperform spatial interpolation? *Global Ecology and Biogeography*, **16**, 733-742.

Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., Soberón, J. & Villalobos, F. (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, **222**, 1810-1819.

Beale, C., Lennon, J. & Gimona, A. (2009) European bird distributions still show few climate associations. *Proceedings of the National Academy of Sciences*, **106**, E41-E43.

Beale, C.M., Lennon, J.J. & Gimona, A. (2008) Opening the climate envelope reveals no macroscale associations with climate in European birds. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 14908-14912.

Bradley, B.A. & Fleishman, E. (2008) Can remote sensing of land cover improve species distribution modelling? *Journal of Biogeography*, **35**, 1158-1159.

Buermann, W., Saatchi, S., Smith, T.B., Zutta, B.R., Chaves, J.A., Milá, B. & Graham, C.H. (2008) Predicting species distributions across the Amazonian and Andean regions using remote sensing data. *Journal of Biogeography*, **35**, 1160-1176.

Chapin III, F.S., Bloom, A.J., Field, C.B. & Waring, R.H. (1987) Plant responses to multiple environmental factors. *Bioscience*, **37**, 49-57.

Chapman, D.S. (2010) Weak climatic associations among British plant distributions. *Global Ecology and Biogeography*, **19**, 831-841.

Currie, D.J. (2007) Disentangling the roles of environment and space in ecology. *Journal of Biogeography*, **34**, 2009-2011.

Davis, A.J., Jenkinson, L.S., Lawton, J.H., Shorrocks, B. & Wood, S. (1998) Making mistakes when predicting shifts in species range in response to global warming. *Nature*, **391**, 783-786.

Easterling, D.R., Meehl, G.A., Parmesan, C., Changnon, S.A., Karl, T.R. & Mearns, L.O. (2000) Climate extremes: observations, modeling, and impacts. *Science (New York, N.Y.)*, **289**, 2068-2074.

Fischer, D.T., Still, C.J. & Williams, A.P. (2009) Significance of summer fog and overcast for drought stress and ecological functioning of coastal California endemic plant species. *Journal of Biogeography*, **36**, 783-799.

Franklin, J., Davis, F.W., Ikegami, M., Syphard, A.D., Flint, L.E., Flint, A.L. & Hannah, L. (2013) Modeling plant species distributions under future climates: how fine scale do climate projections need to be? *Global Change Biology*, **19**, 473-483.

Gavin, D.G. & Hu, F.S. (2005) Bioclimatic modelling using Gaussian mixture distributions and multiscale segmentation. *Global Ecology and Biogeography*, **14**, 491-501.

Gavin, D.G. & Hu, F.S. (2006) Spatial variation of climatic and non-climatic controls on species distribution: the range limit of Tsuga heterophylla. *Journal of Biogeography*, **33**, 1384-1396.

Givnish, T.J. (2002) Adaptive significance of evergreen vs. deciduous leaves: solving the triple paradox. *Silva Fennica*, **36**, 703-743.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993-1009.

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., OVERTON, J., Aspinall, R. & Hastie, T. (2006) Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology*, **43**, 386-392.

Hampe, A. (2004) Bioclimate envelope models: what they detect and what they hide. *Global Ecology and Biogeography*, **13**, 469-471.

Hartesveldt, R.J., Harvey, H.T., Shellhammer, H.S. & Stecker, R.E. (1975) The giant sequoia of the Sierra Nevada. *The giant sequoia of the Sierra Nevada*.

Harvey, H.T., Shellhammer, H.S. & Stecker, R.E. (1980) Giant sequoia ecology. Fire and reproduction. *Giant sequoia ecology.Fire and reproduction*.

Hijmans, R.J. (2012) Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology*, **93**, 679-688.

Hutchinson, G.E. (1959) Homage to Santa Rosalia or why are there so many kinds of animals? *American naturalist*, **93**, 145.

Ichii, K., Wang, W., Hashimoto, H., Yang, F., Votava, P., Michaelis, A.R. & Nemani, R.R. (2009) Refinement of rooting depths using satellite-based evapotranspiration seasonality for ecosystem modeling in California. *Agricultural and Forest Meteorology*, **149**, 1907-1918.

Kerr, J.T. & Ostrovsky, M. (2003) From space to species: ecological applications for remote sensing. *Trends in Ecology & Evolution*, **18**, 299-305.

Lathrop Jr, R.G., Aber, J.D. & Bognar, J.A. (1995) Spatial variability of digital soil maps and its impact on regional ecosystem modeling. *Ecological Modelling*, **82**, 1-10.

Legendre, P. (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology*, **74**, 1659-1673.

Lennon, J.J. (2000) Red-shifts and red herrings in geographical ecology. *Ecography*, **23**, 101-113.

Monteith, J. (1965) Evaporation and environment., **19**, 4.

Mu, Q., Heinsch, F.A., Zhao, M. & Running, S.W. (2007) Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sensing of Environment*, **111**, 519-536.

Mu, Q., Zhao, M. & Running, S.W. (2011) Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment*, **115**, 1781-1800.

Pearson, R.G. & Dawson, T.P. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, **12**, 361-371.

Peterson, A.T., Barvea, N., Binib, L.M., Diniz-Filhob, J.A., Jiménez-Valverdea, A., Lira-Noriegaa, A., Loboc, J., Mahera, S., de Marco Jr, P. & Martınez-Meyerd, E. (2009) The climate envelope may not be empty. *PNAS*, **106**, E47.

Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M. & Guisan, A. (2006) Are niche-based species distribution models transferable in space? *Journal of Biogeography*, **33**, 1689-1703.

Schenk, H.J. & Jackson, R.B. (2002) The global biogeography of roots. *Ecological Monographs*, **72**, 311-328.

Schenk, H.J. & Jackson, R.B. (2005) Mapping the global distribution of deep roots in relation to climate and soil characteristics. *Geoderma*, **126**, 129-140.

Schiffer, R. & Rossow, W.B. (1983) The International Satellite Cloud Climatology Project(ISCCP)- The first project of the World Climate Research Programme. *American Meteorological Society, Bulletin*, **64**, 779-784.

Singer, I.A. & Smith, M.E. (1953) Relation of gustiness to other meteorological parameters. *Journal of Meteorology*, **10**, 121-126.

Soberon, J. & Nakamura, M. (2009) Niches and distributional areas: concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences of the United States of America*, **106 Suppl 2**, 19644-19650.

Stephenson, N.L. (1990) Climatic control of vegetation distribution: the role of the water balance. *American Naturalist*, 649-670.

Thornthwaite, C.W. & Mather, J.R. (1957) Instructions and tables for computing potential evapotranspiration and the water balance.,.

Tompkins, A. (2003) Impact of temperature and humidity variability on cloud cover assessed using aircraft data. *Quarterly Journal of the Royal Meteorological Society*, **129**, 2151-2170.

Tucker, C.J. (1979) Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, **8**, 127-150.

von Humboldt, A. & Bonpland, A. (1807) *Essai sur la géographie des plantes*.

Walcek, C.J. (1994) Cloud cover and its relationship to relative humidity during a springtime midlatitude cyclone. *Monthly Weather Review*, **122**, 1021-1035.

Walter, H. (1973) *Vegetation of the earth in relation to climate and the eco-physiological conditions.*, English Universities Press.

Woodward, F.I. (1987) *Climate and plant distribution*, Cambridge University Press.

Wright, I.J., Reich, P.B., Westoby, M., Ackerly, D.D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J.H. & Diemer, M. (2004) The worldwide leaf economics spectrum. *Nature*, **428**, 821-827.

Chapter 2. Developing a Custom Cloud Climatology of the Western United States from Satellite Data, 1981-2012

Abstract

Most currently available cloud climatology data from satellite are at very coarse spatial resolution, on the order of tens of kilometers. This renders them of low utility for understanding solar radiation and other climatic patterns necessary for a variety of applications, including hydrological and ecological modeling. Higher resolution cloud mask products accompanying satellite data are found to be of insufficient accuracy to generate accurate higher resolution cloud climatologies. Here, over 30 years of satellite data were processed with novel cloud classification algorithms to generate a custom cloud climatology for the western United States. These data consisted of a time series of Advanced Very High Resolution Radiometer (AVHRR) data for 1981-1999 and Moderate Resolution Imaging Spectroradiometer (MODIS) data for 2000-2012 (MOD09GA and MYD09GA products, from MODIS Terra and Aqua satellites, respectively). Standardized AVHRR data at .05 degree resolution obtained from NASA's Land Long Term Data Record (LTDR) project (AVH09C1 data) enabled the development of a relatively accurate product from older sensor data that suffer from inconsistency in data quality, viewing geometry related distortion, limited spectral and thermal band information, and overpass time drift. The LTDR project's development of a novel mid-infrared reflectance product was found to be especially helpful for the discrimination of clouds from snow, as was a custom attempt at deriving mid-infrared reflectance. MODIS thermal data were not used, but MODIS reflectance data at 500 meter resolution were found to generally be sufficient to separate most clouds from snow, especially with the use of band 5 shortwave infrared data in addition to one of the other shortwave bands. The three true color bands available in MODIS (compared to only one in AVHRR) were found to be especially helpful for separating clouds and snow from bright desert soils. Unfortunately, recent problems of striping in both Terra band 5 and Terra band 3 (blue) led to the need for classification algorithm adjustments for that sensor that did not completely redress resulting cloud classification problems. Nevertheless, these newly developed products provide vast improvements in cloud climatology compared to existing datasets that often have problems consistently discriminating clouds from snow and desert soils, or are mapped at much coarser spatial resolution. These improvements are critical for analyses requiring accurate relatively high resolution mapping of cloud cover, such as the distribution modeling of plant species that might be dependent on clouds.

Introduction

Cloud cover plays a major role in the Earth's global climate (Ramanathan *et al*, 1989) and its role in regulating finer scale biological pattern is increasingly appreciated (Gentry, 1992; Goldsmith *et al*, 2013; Sklenár *et al*, 2008; Sugden and Robins, 1979). Future trends in cloud

cover are uncertain (Hansen *et al*, 1998), but will both affect and be affected by climate change (Ackerman and Stokes, 2003), with implications for the biodiversity of cloud forests (Foster, 2001), among other areas. Besides providing useful climatological information for a variety of current applications (e.g., solar energy generation, ecological and hydrological modeling), improved mapping of historic global patterns and trends in cloud cover (i.e., a cloud climatology) might help toward the forecasting of future cloud cover regimes associated with climate change. The spatial resolution of cloud cover maps can be improved due to recent advances in daily global satellite imagery, with Moderate Resolution Imaging Spectroradiomter (MODIS) data allowing a better understanding of patterns and trends since 2000, and advances in Advanced Very High Resolution Radiometer (AVHRR) processing permitting a better understanding (but at a coarser resolution than for MODIS, despite the names) back to 1981.

Historic patterns of cloud cover across large areas have not been consistently mapped with accuracy at a scale useful for most ecological purposes (i.e., less than tens of km). Historic cloud cover maps have generally been of very coarse resolution, especially for any products with global coverage. For example, the International Satellite Cloud Climatology Project (ISCCP) has used historic satellite imagery back to 1983 (mostly AVHRR) to generate a global record of cloud cover at the coarse spatial resolution of 32 to 280 km (2.5 degrees) and temporal resolution of days to months (Rossow and Schiffer, 1991; Schiffer and Rossow, 1983). Similarly, the Pathfinder Atmosphere (PATMOS) project has generated various cloud products at 110 km (~ 1degree) resolution using AVHRR data from 1981 to 1999 (Jacobowitz *et al*, 2003; Stowe *et al*, 2002). Slightly finer resolution products have been generated from NOAA's High Resolution Infrared Radiation Sounder (HIRS) for 1979-2001 at approximately 20 km resolution (Wylie *et al*, 2005; Wylie and Menzel, 1999). The spatial resolution of these products fails to match that of many other important climate variables that are now commonly available at resolutions on the order of a kilometer to several kilometers, despite the availability of historic satellite data at these resolutions. Furthermore, the ability of ISCCP data to detect long-term trends in solar radiation at the surface (Pinker *et al*, 2005) has been found to be suspect due to satellite orbital drift and resulting viewing geometry inconsistency that affects cloud cover classification (Dai *et al*, 2006; Evan *et al*, 2007; Norris, 2000). While ISCCP finds decades-long declines in cloud cover, processing of HIRS data from 1979-2001 suggests little change in global cloud cover over a similar period (Wylie *et al*, 2005).

More recent processing of both AVHRR (NOAA Pathfinder Atmosphere's Extended climate dataset: PATMOS-x; (Heidinger *et al*, 2012)) and MODIS (NOAA Cloud Climatology product; (Douglas *et al*, 2010)) provide improvements to cloud climatology in resolution and accuracy, but still have some limitations: the PATMOS-x product tends to classify bright desert soils as clouds (see Appendix 2.A.1) and the NOAA Cloud Climatology is a relatively simple brightness thresholding product that tends to classify anything very bright, such as snow, as clouds. Even the MODIS normalized difference snow index ($NDSI = (\rho_{band\ 4} - \rho_{band\ 6})/(\rho_{band\ 4} + \rho_{band\ 6})$; (Hall *et al*, 1995; Hall *et al*, 2002)) is not sufficient for consistent cloud and snow separation, primarily due to confusion between snow and ice clouds. The utility of MODIS band 5 for discrimination of difficult-to-separate snow and clouds - due to a relative peak in band 5 for clouds, even those that have low reflectance in bands 6 and 7 (Riggs and Hall, 2004) - has been more recently noted. But their "peak band 5" ratio $[(\rho_{band\ 4} / \rho_{band\ 5})/(\rho_{band\ 5} / \rho_{band\ 6})]$ has not been incorporated into standard processing of MODIS data. MODIS data are accompanied by more

elaborate cloud masks (Platnick *et al*, 2003), but these have been found to have some systematic error when processed over time (see below and Appendix 2.A.2). Significant advances have been made with MODIS data in the identification of clouds through spatial and temporal context (Lyapustin *et al*, 2008) and in the separation of clouds from snow using radiometric and topographic rules (Miller *et al*, 2005), but these improvements have not been employed to generate a long term cloud climatology, or incorporated into the standard MODIS cloud products.

Improvements to and standardization of historic AVHRR data as part of NASA's Land Long Term Data Record (LTDR) project (Pedelty *et al*, 2007) have helped to set the stage for a more accurate long term cloud climatology. This includes a novel band 3 (middle infrared) "reflectance" product (provided by the LTDR project as "experimental" (Pedelty *et al*, 2007) and derived by "removing" the thermal component of the band 3 radiance (Roger and Vermote, 1997; Roger and Vermote, 1998)) that helps to improve the discrimination of snow and clouds, in particular. The LTDR project strives for a degree of consistency among diverse sources of historic imagery (e.g., AHVRR, 1981-1999; MODIS, 2000-present) to generate a now thirty-plus year continuous record of daily land observations from space (http://ltdr.nascom.nasa.gov). Daily global LTDR Version 3 AVHRR data at 0.05 degree resolution (7200 x 3600 pixels) can be browsed and downloaded from their website (ltdr.nascom.nasa.gov/ltdr/productSearch.html; AVH09C1 for radiometric data) and daily tiled MODIS reflectance data at ~500 meter resolution can be downloaded from NASA's LAADS Web (http://ladsweb.nascom.nasa.gov/data/search.html) or from the USGS LPDAAC "Data Pool" (https://lpdaac.usgs.gov/get_data/data_pool; MOD09GA and MYD09GA, respectively, for MODIS Terra and Aqua reflectance data).

Each of these satellite data sets is accompanied by cloud mask products - flagged in the various bits of the Quality Assessment (QA) field for AVHRR (Pedelty *et al*, 2007) and the "state_1km" Scientific Data Set (SDS) field for MODIS (http://modis-sr.ltdri.org/products/MOD09_ UserGuide_v1_3.pdf). These products are not well established for cloud climatology purposes. The AVHRR LTDR cloud mask product employs the same "Clouds from AVHRR" (CLAVR-1; (Stowe *et al*, 1999)) algorithm of the original PATMOS project, but doesn't appear to have been used to generate a higher resolution cloud climatology. The value in the second bit (technically bit number 1, "cloudy") of AVH09C1's QA field appears to represent all cloud cover (the first bit is supposed to be "partly cloudy" but appears to be always empty). The MODIS MOD09GA and MYD09GA cloud state bits (first two, "0 and 1", of state_1km SDS) are derived from the MOD35 cloud mask, which does not always agree with the MOD06 cloud product (Leinenkugel *et al*, 2013; Liu and Liu, 2013). There is also an "internal cloud flag" in bit 10 (counting from zero, otherwise it's 11th) that does not always agree with the cloud state bits. The first bit (technically bit 0) appeared to best represent cloud cover for MODIS data (but see Appendix 2.A.2). These different MODIS products have variable uses, but are often designed for a focus on land cover and a desire to ensure cloud-free viewing. The resulting liberal approach to classification of clouds can result in haze, ocean glint, and even relatively bright surfaces such as snow (Hall and Riggs, 2007) or bare soils errantly classified as cloud. MODIS cloud masks have been found to be poor over winter snow cover in China (Wang *et al*, 2008) and even to have large errors of omission and commission of cloud cover in the Amazon (Hilker *et al*, 2012). Inconsistency in MODIS cloud classification related to underlying background land

cover has also been identified in an analysis across a whole year of daily data, regardless of which bits in the state_1km SDS are used to identify clouds (Wilson *et al*, 2014). Much of the confusion is likely due to the fact that different cloud products have been optimized for different purposes – which generally don't include "consistency for cloud climatology purposes". Identification of various problems has caused frequent changes in algorithms, with new "Collections" every few years (Ackerman *et al*, 1998; Frey *et al*, 2008). Yet problems persist, even in the most recent "Collection 6" (Wilson *et al*, 2014). Despite differences in sensors and cloud cover identification methodology, similar classification confusion issues occur for both AVHRR and MODIS, with frequent problems with snow and other bright surfaces.

Some error in cloud cover classification may be unavoidable, especially for some thin or partial clouds (where variable degrees of mixing with variable background land cover can create a confounding variety of spectral signatures) and what might be "ice clouds" – those with low shortwave infrared reflectance, which results in a spectral signature much like snow. (Clouds with these characteristics are often assumed to be cirrus, but they can include the tops of cumulonimbus and perhaps other types of clouds. The common characteristic is that they likely contain frozen water. They will hereafter be simply referred to as ice clouds.) The overlooking of thin or partial clouds may not be a gross error for cloud climatology, especially if there are not substantial land cover related biases, but the latter problem of ice clouds may be a bigger one. These clouds with low shortwave infrared reflectance can be quite common. They can be errantly omitted in a cloud classification due to their spectral similarity to snow (Miller *et al*, 2005), or contribute to misclassification of snow as cloud (Hall *et al*, 2002), depending on the classification algorithm. These issues can lead to substantial error, and even those errors that seem subtle can be magnified across a long-term climatology if systematic. This appears to particularly be the case for ice or snow and some bright desert playas. Confusion between snow and cloud is especially concerning if one is interested in identifying cloud cover frequency in mountainous areas of frequent snow cover, such as is the case for much of the western United States.

Clearly there is a need for accurate and consistent (across space and time) cloud cover characterization. Optimal cloud cover products likely vary by user. A user interested in subtle nuances of land cover (e.g., change in plant productivity over time) would want all hints of cloud identified and removed and might not care if clouds were easier to identify and remove over particular cover types. A cloud climatology product, on the other hand, should have a particular focus on overall accuracy and consistency, minimizing any systematic bias. Spatial (e.g., underlying cover types) or temporal bias is concerning for anyone looking for cloud trends in space or time.

Given that cloud classification accuracy likely declines for partial and the thinnest of clouds, and varies substantially by underlying cover types (e.g., high over water vs. lower over bright surfaces), research here did not put a major focus on always capturing the thinnest of clouds or partial clouds. Omitting a small portion of clouds in order to avoid commission error on oft-confused land-cover types results in an overall increase in accuracy according to an equal weighting of those errors – i.e., closer to the [0,1] point (perfect classification) on a receiver operating characteristic (ROC) curve (Coffin and Sukhatme, 1997). In addition to the issues of improved accuracy and consistency, this approach is defensible for several other reasons.

Frequent classification of partial cloudiness (i.e., subpixel clouds) as cloud would result in an overestimate of cloud frequency. Additionally, particularly thin clouds do not have an impact on surface radiation comparable to thick clouds. Furthermore, each of these hard-to-consistently-identify cloud cover types might occur in reasonable proportion to the easily mapped cloud cover for a given pixel, so the resulting omission might be proportional and the resulting biases minimal. Only areas subject to unusually frequent hazy or partial clouds would be exceptions.

The goal of this work is to improve upon existing satellite masks to generate consistent cloud classification and resulting climatology. This involves improved discrimination of cloud from snow and bright soils, in particular, for both MODIS and AVHRR data. This discrimination, and consistency in classification across cover types, may be easier with a slightly more conservative cloud classification. In a sense, improved discrimination could be thought of as creating a different and steeper ROC curve (higher true positive classification for a given false positive classification) compared to the standard product, with decreased commission error (i.e., false positives) enhanced by the more conservative classification (Figure 1). Additionally, the cloud classification is a balancing act of maximizing this overall accuracy, capturing overall frequency accurately, and doing so with consistency across underlying cover types. Figure 2 illustrates these issues, and shows how accuracy in cloud classification can vary by underlying cover type: for example, classification over water has a steep curve because of water's great contrast with cloud, while classification over deserts (and other bright surfaces) will have lower accuracy due to the reduced spectral contrast. Technically, these curves are not independent of each other unless land-cover specific classification is performed; otherwise classification decisions benefiting performance over one cover type will often be at the expense of performance over another cover type.

This different performance according to underlying cover type and lack of independence in performance among cover types are reasons that many cloud mask algorithms have land-cover specific classification according to different "processing paths" (MODIS;(Frey *et al*, 2008)) or "cloud detection surface types" (PATMOS-x; (Heidinger *et al*, 2012)) – stratification of the landscape. This should allow optimization for different cover types and overall improvement, but it is dependent on an accurate cover map. Stratification can result in spurious abrupt boundaries in cloud frequency that are obvious problems (see Appendix 2.A.2 for MODIS examples), but it can also make it more difficult to identify problematic areas. Egregious errors may still be present, but relatively concealed, adding an extra layer of complexity in understanding and quantifying the accuracy of the product. As a result of some of these concerns, the work described here involved a universal mapping of clouds without stratification.

Data and Methods

Assistance was provided by NASA Ames personnel in acquiring historic satellite data. These data consisted of NASA's LTDR daily AVHRR radiance and reflectance products (AVH09C1; 1981-1999), and MODIS Terra and Aqua daily reflectance data (MOD09 and MYD09, respectively, for 2000-2012 and 2002-2012, respectively). The NASA Earth Exchange Web Portal (NEX) was used for the storage and processing of the satellite data, and the storage of preliminarily processed products. AVHRR products are provided in a geographic projection

globally at .05 degree (~5 km) resolution. Terra and Aqua products were collected at a nominal resolution of 500 meters for 3 gridded tiles (h08v05, h08v04, h09v04) in sinusoidal projection (Figure 3). This tile selection was intended to cover all of California, but also covers much of the western United States, parts of Mexico and Canada, and a portion of the Eastern Pacific Ocean. While AVHRR data are collected at a nominal resolution of 1.1 km, and thus somewhat close to that of MODIS, the LTDR data have been processed and distributed at a coarser 0.05 degree (~5-km) resolution, as a means of data reduction and standardization of the product (e.g., off-nadir viewing can greatly enlarge the true ground-sampled area of a given pixel for this wide field of view sensor).

Table 1 and Table 2 provide band characteristics for the AVHRR and MODIS sensors, respectively. Table 3 gives exact dates for the different NOAA satellites associated with the AVHRR data. Table 4, Table 5, and Table 6 summarize the number of good days per month from each sensor (AVHRR, MODIS Terra, and MODIS Aqua, respectively), at least according to whether the images existed and were not empty or otherwise compromised in a major way. Zero values in Table 5 and Table 6 for February, 2000 and August, 2002 for MODIS Terra and Aqua, respectively, are the result of not processing the months at the inception of the data since they were incomplete. Terra began collecting data in late February of 2000, and Aqua in early August (August 8th) of 2002. Zero values in the months before these indicate the absence of data for these periods. Empty values after March 2012, on the other hand, correspond to the termination of processing for this work.

The tables of "good days per month" indicate the number of days for which images existed. However, it was not completely unusual for images – especially early AVHRR data - to have gaps/areas of completely missing data within otherwise useful images. These areas would be errantly classified as entirely cloud free. These partial images were initially used with no attempt at correction for data gaps within otherwise apparently good data. Later processing involved determining rules for these gaps and making maps of their frequency. These could be used to correct the preliminary cloud frequency calculations. These bad (generally missing) data tended to occur more frequently over higher latitudes, especially in the winter months, and therefore did not substantially affect preliminary outputs for the initial area of interest of California. It was not clear if there was a "bad or missing data" flag accompanying the LTDR data that could be processed in any automated manner.

All image data were stored on NASA's NEX computers for processing. The daily (7200x3600) AVHRR images are ~170 Megabytes, and each (2400x2400) MODIS tile is ~70 Megabytes, so this is approximately 60 Gigabytes per year for AVHRR data and 70 Gb/yr total for the 3 tiles of each MODIS sensor. (60 Gb/yr * 18 years) + (70 Gb/yr * 12 years) + (70 Gb/yr * 10 years) = ~2.6 Terabytes for the entire data record. Even processing of the data to simple daily cloud masks added hundreds of Gb of data. AVHRR data were initially only processed for a (2000x1000) window of North America south of 55 degrees latitude, including much of the eastern Pacific and western Atlantic, so as to keep the cloud classification optimized for areas of interest and to reduce dataset size (by over an order of magnitude compared to global). This main subset consisted of pixels 600:2599 in the horizontal and 700:1699 in the vertical, from top left. Given .05 degree resolution, this corresponds to ~5.05° N to 55.05° N, and 150.05° W to 50.05°

W. Far northern latitudes were also not considered because of limited sunlight during the winter months and increased snow and cloud confusion.

Data sources of note that were not used include MODIS thermal bands and other MODIS 1 km resolution data, such as the "Cirrus" band (channel 26, 1.38 um). MODIS radiance data for Channel 26 and thermal wavelengths are only available as "Level 1" swath data, meaning they require a "swath to grid" conversion and the use of the MODIS Reprojection Tool. The difficulty in applying these conversions to thousands of images, uncertainty about their utility in improving cloud classification, and concerns about their coarser 1 km resolution contributed to the decision not to explore their use.[1] MODIS thermal data are also available in a more processed, gridded format as "land surface temperature" (LST), but there are still concerns associated with the 1 km resolution, the lack of any data provided where a potentially inaccurate cloud mask has been applied, and its utility in improving cloud classification consistently. Thermal information is not a consistent physical characteristic of a surface, and can be unreliable for classification purposes, especially without accompanying meteorological information.

AVHRR Processing

The standard cloud mask was extracted from the AVHRR dataset, but as described in the AVHRR section of "Problems with Cloud Masks" (Appendix 2.A.1), was found to be very liberal in its definition of cloud, and included many areas of non-cloud, such as haze, sun glint on water bodies, and snow. A custom classification was developed to remove many of these problem areas. This custom classification focused on the three major problems associated with the standard cloud mask (haze/thin clouds, sun glint, and snow) and other problems associated with unusual band behavior (e.g., reflectance values in various bands at or below zero, or above one).

The custom classification of clouds for AVHRR relied upon the manual determination of a series of thresholds (rules) in various bands and band ratios. These were derived in MATLAB® (The MathWorks, 2011) through extensive querying of values from clouds, snow, and other surfaces evident in multiple dates of AVHRR imagery. These surfaces were identified on individual images from a combination of reflected, thermal, and contextual information. See Figure 4 for a basic overview of the processing and Appendix 2.B.1 for MATLAB code containing the exact rules.

First, hazy atmospheres (or possibly some very thin or partial clouds) were removed from consideration as cloud by requiring band 1 (red) reflectance to be greater than 0.12. Secondly, it was observed that what is commonly referred to as ocean glint (sun glint) tended to have unusually high "radiance" values for band 3 compared to band 1 reflectance. (Here referring to the raw band 3 product as radiance, because it has both the thermal and reflected contribution.) Clouds were therefore considered to have band 1 reflectance exceeding a threshold related to their band 3 radiance (band1 > (band3*9-2.45)), or, to account for some clouds with unusually

[1] Miller et al. (Miller *et al*, 2005) used the MODIS Cirrus band in their cloud/snow discrimination, in combination with a digital elevation model; water vapor in the lower atmosphere tends to absorb most of the radiance at 1.38 um, allowing high clouds (often cirrus) to stand out as brighter. But elevated areas can sometimes show surface returns in the Cirrus bands due to their thin - and often relatively dry – atmospheres.

low band 1 reflectance, a normalized difference of band 2 (NIR) and band 1 (red) [$(\rho_{NIR} - \rho_{red})/(\rho_{NIR} + \rho_{red})$] greater than 0.1. This ratio (Rouse Jr *et al*, 1974) is generally called the Normalized Difference Vegetation Index, or NDVI. It is not intended as a vegetation index here for these exceptional cases of radiometric problems. Clouds should generally have NDVI ≈ 0 ($\rho_{NIR} = \rho_{red}$), but if their red band is anomalously low, their NDVI will be anomalously high, assuming that the near infrared band is not similarly anomalous. Other rules prevent the selection of vegetated areas by this rule alone. Thirdly, clouds were differentiated from snow based primarily on their band 3 "reflectance" component as well as a normalized difference ratio of band 3 radiance and band 4 radiance [$(rad_{band3} - rad_{band4})/(rad_{band3} + rad_{band4})$] (which should also get at the reflectance component of band 3, since the thermal portion of band 3 should be somewhat proportional – if relatively gray body behavior in emission - to band 4, which is essentially all thermal). The LTDR-provided band 3 "reflectance" product and my normalized band 3 vs. band 4 ratio are very similar but not duplicates. Much more effort went into the creation of the band 3 "reflectance" product (Roger and Vermote, 1997; Roger and Vermote, 1998), but there's still some useful independent information in the normalized ratio. (See Table 7 for a list of various normalized ratios used in processing imagery.) Snow, with low reflectance in the spectral region covered by band 3, tends to have lower values than clouds in these products (although optically thick cirrus clouds with large ice crystals can be an exception (Miller *et al*, 2005)). Relatively high band 1 reflectance and the typically low NDVI of clouds were additionally used along with some variants on these band 3 related rules to assist with particularly problematic pixels.

Special exceptions to the above set of rules were made for pixels with unusual band behavior, often covering large areas of AVHRR data. In the cases where band 1 or band 2 anomalously exceeded 100% reflectance or had reflectance less than or equal to zero, band 3 rules (high values in band 3 "reflectance" and the normalized difference band 3 vs. band 4 ratio) were incorporated to capture clouds. In cases where band 3 had reflectance less than or equal to zero, relatively high values in band 1 reflectance and the normalized difference band 3 vs. band 4 ratio were indicative of clouds. Final cloud cover was determined according to those pixels that met the criteria of either of these two exceptions, or for those that met all three criteria in the previous paragraph.

The above-described custom cloud classification fixed most of the AVHRR cloud mask problems (or at least created a reasonably accurate conservative cloud cover classification) but was still imperfect, due to some of the limitations of coarse-resolution, band-limited, and noisy AVHRR data. Noticeable problems persisted in select desert areas – playas and those that happen to have reddish soils ("happen to" because their true color is not evident with AVHRR – red soils must have other characteristics that make them confounding in AVHRR bands). Subtracting the standard LTDR cloud frequency from my custom classification cloud frequency led to a mask for these areas [within a 570:1010, 160:670 geographic box (x,y, from the top left) within my already existing subset; this corresponds to a northwest corner of 47.05 N, 121.55 W and a southeast corner of 21.55 N, 99.55 W: encompassing arid lands of the western U.S. and northern Mexico]. These arid lands were the only areas where I mapped clouds with a greater frequency than the LTDR cloud mask. This was clearly due to an overestimate in my classification rather than an underestimate in the LTDR product. Consequently, I recoded my monthly cloud frequency values of pixels within this mask area to those of the LTDR product if my monthly value exceeded the value of the standard product. This post-processing led to a halo

around some desert areas (particularly obvious over playas) in resulting cloud frequency products that were similar to halos in the products derived from the standard cloud masks. (See Appendix 2.A.1: It is speculated that the standard cloud product took advantage of a land cover-based mask for these areas, and that their mask was not perfect – either not as extensive as it should have been, or slightly offset in its registration.)

The custom cloud classification algorithms were applied in MATLAB to generate daily .mat cloud-mask files. These were aggregated to monthly and annual products (both as cumulative sums and as cloud frequency as a percentage of the good data for a given month). Monthly products were exported to ASCII grids that could be incorporated into ArcGIS for easier manipulation (e.g., reprojection) and analysis. (Daily .mat cloud masks could be analyzed in MATLAB, but were too numerous to systematically convert to other more manageable formats.) Various derived products could then be generated in ArcGIS from these individual monthly products. For example, linear trends in monthly and annual cloud frequency were generated for both the standard AVHRR LTDR cloud product and the custom cloud product. Monthly regressions, ignoring any missing month, were calculated first, and the annual trend was determined as the average of the monthly regressions. This averaging approach can lead to slight biases related to the differing numbers of good data (and trend duration) in each month, but it also makes use of the most possible data, including different start years.

MODIS Processing

The lack of a truly standard cloud mask for MODIS data, and concerns about the accuracy of the various cloud masks that were assessed (see Introduction and Appendix 2.A.2), led to the development of custom cloud masks for MODIS data as well (see Appendix 2.B.2 for MATLAB code). The final process for standard MODIS imagery (i.e., those data without substantial problems - discussed below) consisted of two major stages with several steps. In the first stage, the very spectrally similar clouds and snow were first separated from everything else on imagery using visual examination of images and manual generation of classification rules (see Figure 5 for a basic overview). With clouds and snow separated from everything else, focus could be turned to a second major stage comprising three steps to separate clouds from snow (see Figure 6 for a basic overview): 1) a series of classification trees (Breiman *et al*, 1984) was generated from a progressively increasing set of manually collected training samples to separate clouds from snow; 2) the resulting preliminary classification of clouds and snow was used to automate placement of a large number of cloud and snow training samples, with visual/manual quality assurance and control; and 3) the final classification tree model was generated with the vastly increased set of points and an increased set of predictor variables. Before describing these steps in greater detail, it is worth discussing some of the challenges encountered in discriminating clouds (from snow, especially) and their attempted remedies.

Extensive analysis of color composites of multiple dates of MODIS Terra and Aqua scenes was initially conducted (in ENVI version 4.8 (Exelis Visual Information Solutions, Boulder, Colorado) and MATLAB) to ascertain whether spectral rules could be manually generated for consistent discrimination of cloudy and clear pixels. The use of color composites (generally false color composites) allows the simultaneous display of multiple dimensions (i.e., reflectance at difference wavelengths) necessary for identifying combinations of spectral information helpful in differentiation. Figure 7 shows a variety of MODIS Aqua color

composites over southern California from a mostly cloudy day in February of 2010. The Sierra Nevada (top-center of each image) has extensive snow cover, as do a few other scattered ranges. Discriminating cloud and snow is not possible in Figure 7.A (top), a true color composite (i.e., red, green, blue (RGB) channels = band 1, band 4, and band 3, respectively (1-4-3)). Even Figure 7.B (middle), a 7-2-1 band combination (RGB = SWIR, NIR, Red), would fail to consistently distinguish all clouds from snow. Many clouds in this image have the strongly reduced band 7 reflectance often associated with ice clouds, resulting in the same cyan color as the snow covered areas. (Note the bright clouds on the west edge that are magenta in this particular composite. This is a result of saturation in the NIR band, causing an unusual negative value that got treated as a zero in the green band in the color composite.) Only with information in the 7-5-1 color composite in Figure 7.C (bottom) can most types of clouds be consistently separated from snow covered areas. This is because ice clouds have a relative peak in band 5 compared to snow covered areas (Figure 8). Ice clouds continue to have a cyan appearance in a 7-5-1 display, while snow covered areas appear blue, as they are dark in both of the SWIR bands (bands 5 and 7).

The use of many dates of imagery is important for understanding the amount of variability that needs to be accounted for. Figure 9 reveals the spectral variability that can be found among different clouds – their spectral "signatures" in MODIS (technically, bands 3 and 4 (blue and green, respectively) should be moved to the left of band 1 for a more true signature; this should not matter much, as clouds should also be very "flat" in their reflectance from bands 1 through 4 – the fact that they are not here appears problematic). This variability demonstrates that spectral processing techniques dependent on unique signatures (e.g., spectral unmixing) are not appropriate for cloud characterization. The use of many dates is also essential to ensure that image variability, atmospheric variability, and seasonal patterns in land cover type and illumination are accounted for. It also helped to reveal the various sensor problems that have arisen over time – particularly with MODIS Terra bands 3 and 5 (see "MODIS Imagery Problems" section below for discussion of how these impacted the cloud classification). Consistent discrimination in space and time is critical – it seems that many cloud classification algorithms only solve the simplest situations. For example, as alluded to above, snow is easily distinguished from clouds in MODIS bands 6 or 7 if one never considers the difficulty posed by ice clouds. Ice clouds reveal that band 5 is just as important as the conventionally recognized bands 6 and 7 for consistent snow and cloud separation. And it's not just clouds that show variability. Spectral characteristics of snow can change over time due to factors such as dust and variable snow grain size (Painter *et al*, 2009). Snow and clouds can also be spectrally mixed with a variety of surfaces, making their discrimination even more difficult.

Snow and clouds together, however, are quite spectrally distinct from most other land cover types. Few other land cover types exhibit such "grayness", or spectral flatness, across the true color bands. Bright desert areas, and even bright desert playas, tend to be separable from clouds and snow based on their relatively high red reflectance compared to their reflectance in other visible bands. Even dark water looks very blue in a modified true color display when average true color reflectance [$\rho_{average} = (\rho_{red} + \rho_{green} + \rho_{blue})/3$] is used to calculate the normalized difference for each of the visible channels (e.g., $ND_{red} = (\rho_{red} - \rho_{average})/(\rho_{red} + \rho_{average})$ – see Table 7). Figure 10 and Figure 11 show the difference between a standard true color composite and a composite with brightness-normalized color deviation, respectively, for a MODIS Aqua image in January of 2010 over much of California and Southwestern deserts.

These images have both clouds and snow. Each is bright in the former and dark in the latter. But desert soils, which can often be confused with snow and cloud due to their brightness, are clearly very red in the brightness-normalized color deviations. Unusual green halos are seen on the far periphery of thin clouds over the ocean in the brightness-normalized product. Comparison with the standard true color image reveals that these areas are not evident as clouds – they are not bright. Their unusual green color suggests a problem with MODIS atmospheric correction – perhaps these areas were not (or were) (in)appropriately identified as clouds during atmospheric processing.

Separating clouds from snow is a greater challenge, especially due to the potential for confusion between snow and ice clouds. Since each tends to be spectrally flat in the visible and near-infrared bands, shortwave infrared bands (5-7) are very useful for their discrimination. Unfortunately, MODIS Aqua band 6 had frequent unusual radiometric values. This has been noted by others (Wang *et al*, 2006) and band 7 is now used as a substitute for band 6 in standard Aqua processing for snow cover, etc. This is often considered a problem for Aqua data, but I have not noted discernible difference in the utility of band 6 and band 7 – they are very highly correlated over all cover types and clouds. As a result, this project focused on the use of band 7 instead of band 6 for both Aqua and Terra.

Figure 12 shows that bands 1 and 7 (like the aforementioned NDSI based on the normalized difference of the respectively similar bands 4 and 6) are not sufficient to separate all snow from all clouds. Isolines of a normalized difference of bands 1 and 7 (i.e., nd17, analogous to NDSI) would plot as diagonal lines radiating from the origin in the scatterplot in Figure 12, with higher values toward the upper left of the scatter. Those pixels with extremely low band 7 reflectance (x-axis in scatterplot) but moderate to high band 1 reflectance (y-axis in scatterplot) have high nd17 values and are mostly snow (highlighted by the software tool as a big patch of red over Sierra Nevada of California in upper-middle portion of image) but include some clouds (highlighted by the software tool as small spots of red over ocean off the coast of Point Concepcion on left-middle portion of image). Furthermore, some mixes of snow, rock, and vegetation can have spectral similarities to clouds - probably because they are similar to hazy clouds over vegetation and/or soil. Adding some amount of shadow to either of these can make discrimination even more difficult.

The generation of other normalized band ratios, in addition to nd17, can assist with the discrimination. Given that band 5 was an important addition for separation of cloud and snow in color display, it is logical that indices involving band 5 would be useful. The normalized difference of bands 1 and 5 (nd15), as well as a version with an offset to band 5 (nd15sub) were developed. The offset (subtraction of 0.2 from band 5) provides a different origin to the isolines of the ratio in band 1 vs. band 5 spectral space (essentially providing different axes of separation), helping with the separation of clouds and snow under different conditions. The normalized difference true color bands also proved somewhat useful in snow and cloud separation, as snow mixing with vegetation and/or soil/rock was slightly more likely to have a green or red tint than spectrally similar clouds, which tend to have a slightly blue tint. (One could speculate that this relates to less of the blue light reflecting off the cloud having been scattered by the atmosphere on its return path; shorter wavelengths such as blue are scattered more and a cloud has less atmosphere above it; atmospheric correction of the imagery might not have

perfectly handled all of these factors.) Finally, using various summed ratios as variables in the classification trees was also found to be helpful – it essentially allows partitions to occur on multi-dimensional diagonal surfaces (hyperplanes) with respect to the original ratios, somewhat like a support vector machine; otherwise partitions can only be orthogonal to the original ratios. See Figure 13. This allows for various degrees of exceptions to thresholds according to magnitudes in other criteria.

Given the potential for greater spectral similarity between snow and clouds, the first step in cloud classification, as mentioned above, involved generating a product including both cloud and snow (i.e., separating clouds and snow from everything else). Relatively small deviations in each of the true color band reflectance values from average true color reflectance (as described above), provided the first set of rules used to separate clouds and snow from other land cover types (i.e., clouds and snow are more gray, with lower deviations). Slightly larger positive blue deviations and more negative red deviations were allowed for clouds and snow, due to the aforementioned tendency of desert soils to be more red than blue. Only small deviations were allowed in the green band, due to potential confusion with vegetated areas. A few additional rules are required to augment the true color rules to ensure accurate capturing of snow and cloud. For example, near-infrared reflectance is required to be greater than 0.12. This prevents very hazy areas and sun-glint on water from being called cloud. (This may remove some very thin clouds over the ocean from consideration as cloud, but the goal is a more conservative and consistent cloud cover product, and this level of cloudiness would not be evident over bright surfaces.) Additionally, pixels have to either be moderately bright in the visible bands ($\rho_{red} + \rho_{green} + \rho_{blue} > 0.9$) or have moderately high normalized difference ratios of blue reflectance to band 5 reflectance or band 7 reflectance (i.e., nd35 or nd37). [2] These rules capture the vast majority of moderately obvious clouds and snow without including other cover types. (One final set of rules was required to add unusually bright pixels that had aberrant negative or zero NIR reflectance values and non-zero values for red reflectance; these pixels that had likely saturated the NIR detector are almost always clouds, though there may be rare exceptions of very bright snow.)

With the accurate separation of clouds and snow from everything else achieved, the focus switched to the necessarily more automated discrimination (due to the difficulty) of clouds from snow. The separation of clouds and snow was performed in a three-stage process. In the first stage, values in an initial set of six indices for clouds and snow were manually collected in MATLAB through the querying of pixels on images from multiple seasons and years. A series of classification trees (Breiman *et al*, 1984) was generated in R (with mostly default settings of the rpart function (Therneau *et al*, 2010)) from a progressively increasing set of manually collected training samples. Ultimately, a limit in accuracy appeared to be reached with this initial set of six indices when the training set reached approximately two thousand training samples from a variety of clouds and snow. The tree developed from this training set was used to generate a preliminary classification of clouds and snow. The second stage of the process involved placing

[2] Band 3, blue, was used here instead of band 1, red, which was later used in similar ratios optimized for snow-cloud separation. This is because bright deserts tend to be reddish and relatively low in band 3, contributing to their desired lower values in the ratio. Band 1 is used for the snow-cloud separating versions, because the greater atmospheric contamination in band 3 can make snow look more cloud-like.

random points within each of these classes, in several image dates, in an automated fashion (using Hawth's Tools (Beyer, 2004) in ArcGIS 9.3). A higher density of points was placed in areas associated with greater confusion between cloud and snow (from the classification tree's terminal node classification statistics). All points were evaluated for correct assignment based on spectral characteristics and context using visual interpretation. Some were manually re-assigned to their correct class, if necessary, and some were deleted if their correct class was too difficult to determine (e.g., very thin or partial cloud over a snow covered area). The third stage involved using the approximately 22,000 resulting points in a second round of classification trees. This second set of classification trees incorporated additional variables such as the summed ratios described above and included slight variations in variable sets as well as the rpart function's complexity parameter. The optimal tree for the final rules was chosen from this final suite based on apparent performance on a variety of MODIS images.

In addition to a custom daily cloud mask, a custom daily snow mask was generated for MODIS (both Aqua and Terra) data. This was easily generated, as the initial cloud and snow mask was derived first and subsequent derivation of clouds left a snow mask as the remainder.

Just as was done for the AVHRR data, classification algorithms were applied in MATLAB to generate custom daily .mat cloud-mask files. These were aggregated to monthly and annual products (both as cumulative sums and as cloud frequency as a percentage of the good data for a given month). Monthly products were exported to ASCII grids that could be incorporated into ArcGIS for easier manipulation (e.g., reprojection) and analysis. This involved first merging each of the monthly sets of 3 MODIS tiles in their native sinusoidal projection in MATLAB before export. Various derived products could then be generated in ArcGIS from these individual monthly products, including reprojections to more commonly used formats, such as the geographic projection of the AVHRR data. Daily .mat cloud masks could be analyzed in MATLAB, but were seen as too numerous to convert to other more manageable data types.

MODIS Imagery Problems

Beginning in 2007, striping in band 5 of MODIS Terra data became apparent in viewing false color composites. This is concerning, given the importance of band 5 in separating ice clouds from snow. The contaminating effect could be difficult to detect if the resulting pixel values (i.e., appearance on imagery) were not particularly anomalous. Special spectral rules were developed to identify bad band 5 areas that were likely over cloud or snow (Figure 14 for a basic overview of the rules and the latter part of Appendix 2.B.2 for detailed MATLAB code), as band 5 is used in the first stage of processing to separate clouds and snow from everything else. Additional spectral rules were generated for the affected pixels that could be identified, to try to attain reasonable discrimination of cloud and snow without the use of band 5 (Figure 15 for a basic overview of the rules and the latter part of Appendix 2.B.2 for MATLAB code). The correction was moderately successful, and rendered the problem minimal, since the striping only affected a portion of the imagery, and ice clouds aren't as common as other clouds. However, beginning near the very end of 2007, a more serious problem became evident for Terra's band 3 (blue channel). This had an inordinate effect on cloud discrimination, as the model was heavily dependent on clouds (and snow) having spectral flatness in the true color bands. The problem became progressively worse over the 2008 year. A rough fix of band 3 (rather than any changing of the model) on data beginning January 1, 2008 was attempted based on scan angle. The

magnitude of the problem tended to be correlated with sensor zenith angles in the backscatter (high sensor azimuth) direction. A simple linear correction (subtraction) to the blue band according to sensor zenith angle in the backscatter direction was applied (Figure 16 for a basic overview of the rules and the end of Appendix 2.B.2 for MATLAB). [3] The correction reduced the faulty blue band issue, but fairly substantial banding patterns remained; this is problematic, as the cloud and snow masking is heavily reliant on, and sensitive to error in, spectral flatness in the true color bands. The effectiveness of the correction across all images was difficult to assess. Alternatives, such as removing the true color criterion or relying on only the green and red bands to establish true color grayness, were not investigated thoroughly, but were not thought to offer improvement. In the end, substantial blanket omission of clouds was evident on select images. Overall cloud frequencies did not appear to be affected in a substantial way at first, but lowered frequencies in comparison with Aqua results led to likely spurious long-term trends. (See the Results section.)

Accuracy Assessment

Solar radiation values from three micrometeorological towers (Soaproot Saddle = -119.256 , 37.031 , 1160 meters ; Providence Creek = -119.195 , 37.067 , 2015 meters ; Short Hair Creek = -118.987 , 37.067 , 2700 meters) collecting ½ hourly weather data at 55 meter height on an elevation transect on the west slope of the central Sierra Nevada were used to determine daily cloud conditions near satellite overpass times. These data were collected as part of the Southern Sierra Critical Zone Observatory project (http://criticalzone.org/sierra/; (Goulden *et al*, 2012)). The data available from various multi-year windows from 2008 to 2012 could be compared with some of the more recent data from MODIS for a ground-based attempt at accuracy assessment of the cloud classification.

Ideally it would be possible to get a ground-based estimate of cloud cover at the exact time of satellite overpass. (Even more ideally, to be fair to the classification, cloud cover would be assessed from the ground with respect to the satellite's position at the time of image collection!) This is difficult to acquire without advance knowledge of satellite overpass time, as solar radiation measurements, for example, may not be made with sufficiently fine temporal resolution. Even those from flux towers with historic half-hourly measurements will generally not perfectly coincide with satellite overpasses. Selection of the optimal half-hourly measurement for comparison is made more difficult by the fact that the MODIS overpass times are quite variable: MODIS Terra tends to pass over the Sierra Nevada of California, for example, at about 11 AM +/- 1 hour. Similarly, MODIS Aqua passes over the mountain range at about 1 PM +/- 1 hour. Selecting the solar radiation measurement that most closely coincides with each of the two daily overpasses is a complex task for automated processing. It is much easier to select a particular time for each overpass each day. Given the potential time window of overpass for each sensor, the median of three half-hourly measurements (11 AM +/- 30 minutes and 1 PM

[3] Unfortunately, MODIS viewing geometry information is only available at ~1 km resolution, so even with resampling to the finer resolution of reflectance data, the data are inherently coarser than the spectral information. There is also an apparent problem in that the viewing geometry data do not always perfectly align with brightness differences in the imagery that should be related to viewing geometry differences, considering common bidirectional reflectance behavior; this suggests a disconnect between the two (i.e., likely error in the viewing geometry information).

+/- 30 minutes) that most closely coincide with each mean overpass time was selected. This approach should both filter any anomalous data and provide an approximation of the most likely level of solar radiation over this hour long period.

The computed level of solar radiation can then be compared against the monthly maximum (i.e., maximum of the medians) for that time period to get a percentage index of radiation for a given overpass (100*daily value/monthly maximum). It was expected that clear days would have a value of >90 or >95 percent according to this index, with the differences related to variable illumination over the course of a month and some level of atmospheric variability. In fact, clear-day (determined from same-day MODIS imagery) values could be a bit lower than this expected value – and were allowed to be lower - often down to 65 to 80 percent, depending on the site. It is unlikely that this is entirely due to monthly illumination variability and atmospheric variability. It seems more likely that much of this is the result of "artificially" inflated monthly maxima due to reflectance off of nearby cloud formations: some of the days of monthly maxima appeared to be partly cloudy near the flux tower locations.

Other climate station data were used to evaluate the AVHRR and MODIS cloud classifications as well. For example, as part of an effort to use the cloud classifications in an analysis of trends in California Central Valley winter Tule fog (Baldocchi and Waller, 2014), the satellite-data-derived fog product was compared with climate station estimates of fog. This included estimates of fog days based on a custom interpretation of California Irrigation Management Information System (CIMIS; www.cimis.water.ca.gov/cimis/data.jsp) climate station data for Shafter, California (35.533° N, 119.281° W), in the southern San Joaquin Valley near Bakersfield, which has daily climate data spanning the duration of imagery processed in this study. This custom model was manually developed, based on values during obvious fog days as determined from satellite data. Fog days were modeled according to daily solar radiation values averaging less than 96.9 watts/m^2 (200 Langleys/day), precipitation less than 1 mm, and maximum minus minimum temperatures less than 10 °C. Relative humidity measurement was deemed too inconsistent and error-prone to be included in the model.

The snow cover "climatology" that was simultaneously generated for MODIS data could also be used as a form of cloud cover classification accuracy. All classification as snow for areas that should never have snow cover (e.g., water bodies, low elevations in much of California) can be considered commission error in the snow product. Other than some bright desert playas that might be classified as snow, these errant snow classifications could also be considered as omission errors in the cloud product. Obviously, this assessment only covers a portion of possible cloud classification error, but it is likely to be a large one. Another partial assessment of classification on cloud vs. snow (for MODIS only) includes classification tree performance on training data for clouds and snow. This assessment occurred for both particularly difficult (i.e., spectrally similar) training pixels of cloud and snow, as well as a more general set of cloud and snow pixels.

Classification accuracy can be measured in a variety of ways, even for a simple binary classification of "cloud" vs. "not cloud". In addition to simple overall accuracy (quantity correctly classified divided by total), user's and producer's accuracy (related to errors of commission and omission, respectively) can also be reported. These measures do not account for

marginal frequencies (i.e., class proportions) and the resulting possibilities of chance agreement that have nothing to do with the ability of the classifier. The modified Kappa statistic (Foody, 1992) accounts for the chance agreement that would arise based on the proportions of each class, both in truth and in the classification. The Tau statistic (Ma and Redmond, 1995) accounts for the same chance agreement due to actual class proportions, but with equal probabilities for each class in the classification, given that the classifier may have no knowledge of actual class proportions *a priori*. Each of these statistics is used in various parts of the accuracy assessment performed here.

Results

AVHRR

Long-term annual cloud frequency derived from the main subset of AVHRR data over southern North America plus adjacent oceans is shown in Figure 17. This is the product after correction of aberrant values over southwestern deserts through comparison with the standard LTDR cloud product. Subtracting the corrected cloud frequency from cloud frequency derived from the LTDR cloud product for the month of January (Figure 19) is revealing. Areas where the desert mask/comparison was applied are black because the frequencies differ little, if at all, here. Water bodies north of 50 degrees are also black (actually negative values) due to the LTDR cloud mask not applying to these areas in winter (and not north of 60 degrees over land). Bright areas in the subtraction product indicate locations where the LTDR standard product mapped clouds with much greater frequency. This includes snowy areas of the Sierra Nevada, Central Rockies and the Great Plains. These should not be mapped with such high frequency in the standard LTDR product. For example, quick perusal of multiple dates of daily MODIS winter imagery over the Great Lakes reveals that the lakes should be mapped with a greater frequency of cloud cover in January than the adjacent land areas of the upper Midwest and northern Great Plains (i.e., “Lake Effect” clouds). The standard LTDR product is doing the opposite of this, obviously because of confusion between snow and cloud. Many other land cover boundaries also turn up in this subtraction product, including coastlines; these are the result of unusual patterns in the LTDR product. The LTDR product also mapped clouds with much greater frequency over oceans. These were often thin or partial clouds that the custom product (somewhat intentionally) missed, but they also tended to include a lot of sun glint and other non-clouds.

Spatial patterns in the custom AVHRR-based annual cloud frequency product in Figure 17 and Figure 18 do not show the many obvious patterns related to abrupt land cover boundaries that plague the standard LTDR cloud product (see Figure 38 at the end of Appendix 2.A). There is obvious variability related to broad ecoregional boundaries (North America CEC Level I ecoregions (Commission for Environmental Cooperation (Montréal, Québec). and Secretariat, 1997)), but this is to be expected. The Great Lakes no longer stand out as an area of strikingly low cloud cover compared to their surroundings, primarily because snowy surrounding land is no longer classified as cloud with such high frequency in the winter. The custom AVHRR-derived product is at a much coarser resolution than the MODIS-derived product, but allows the analysis of cloud frequencies back to 1981. For example, it helped to extend the analysis of trends in California Central Valley winter Tule fog back to the 1981-82 fog season (Baldocchi and Waller, 2014). AVHRR data are also useful for the synoptic viewing of patterns across large areas that is

more difficult to achieve through MODIS processing. For example, particularly low values of cloud frequency can be seen for the Caribbean, the Gulf of Mexico north of the Yucatan Peninsula, off the west coast of Mexico, and off the east coast of the United States. The north Atlantic stands out for its high cloud frequency, in addition to other areas that are not as surprising (e.g., Pacific Northwest and tropical mountains). The west coast of Newfoundland, in particular, is especially high in the winter months, and this is likely fog as well. Other areas with high winter cloud cover that are likely fog include the Great Lakes, the Rogue River Valley of southern Oregon around Grants Pass and the Port Alberni area of Vancouver Island, British Columbia. Many of these patterns are not evident in other products that frequently confuse clouds and ice/snow.

MODIS

Figure 20 shows monthly cloud frequency from MODIS Aqua from approximately 2002-2012 (beginning and end years vary slightly by month) for California and areas of adjacent Oregon. Even though it does not have as long of a record as the Terra product, the Aqua product is likely the best for average cloud climatology purposes, because it did not have any of the problems that were associated with the Terra data. While the MODIS products do not have quite the temporal extent of AVHRR, limiting long-term trend analyses, they do have the improved spatial resolution and likely improved accuracy necessary for more detailed analysis of spatial patterns in cloud frequency. These products capture fine resolution in mountainous areas like the Sierra Nevada (outlined within California in Figure 20 with the Jepson floristic ecoregion (Hickman, 1993)) necessary to capture the mountain range's high frequency variability in clouds. The west slope of the southern Sierra Nevada, in particular, stands out as an area of high cloud frequency from February to May, and again during October and November. Higher elevations to the east (evident as high cloud frequency during July and August summer monsoon period) are not mapped as high cloud frequency during the winter, despite their frequent snow cover. California's Central Valley (just west of the Sierra Nevada) Tule fog is notable, especially in the southern part – the San Joaquin Valley, in the January cloud frequency image, and to some extent in the December image. The fog can occur in November and February as well, but is swamped by overall cloud frequency during these months. After applying a spatial filter to separate fog from general cloud cover (Baldocchi and Waller, 2014), the fine resolution and accuracy of the MODIS data allows a portion of the San Joaquin Valley between Hanford, Corcoran, and Tulare to be identified as the hotspot of Central Valley fog, at least from 2000-2012.

A direct comparison between April cloud frequencies derived from the Aqua custom classification and the Aqua standard cloud mask (Figure 21 and Figure 22, respectively) reveals differences that appear subtle, at first, but that have large implications. The standard cloud mask maps clouds with greater frequency over snow-covered areas of the Sierra Nevada, especially over areas with a lot of bare rock, such as northern Yosemite National Park. That much of this extra cloud cover is likely error due to the effects of bright snow and rock is revealed by dividing the standard cloud frequency product by the custom cloud frequency product (Figure 23). The tendency for bright deserts to be errantly classified as clouds in the standard data is also evident in this quotient product as bright areas across the bottom-right portion of the image. (Subtracting the two products instead of dividing yields a similar figure, but doesn't highlight as well the greater tendency for the standard product to classify desert areas as cloud, due to the lower

frequencies and absolute differences.) The distribution of giant sequoia (*Sequoiadendron giganteum* (Lindl.) Buchholz) is overlaid in red on these figures to demonstrate the problem posed by cloud classification error: the correspondence of the tree with the high frequency of cloud cover on the west slope of the southern Sierra Nevada seen in Figure 21 is not nearly so evident in Figure 22 due to general increases in cloud cover mapped all over the snowy (and rocky) regions of the Sierra Nevada. This would obviously hinder an assessment of the extent to which the tree's distribution might be related to cloud cover (see Chapter 3).

Comparison between the cloud frequency results for the MODIS Aqua and Terra custom classifications reveals some problems with the latter. Subtracting Terra-derived monthly and annual cloud frequency from Aqua-derived equivalents (see Chapter 3, figures 25 and 26 for March and April versions) reveals that, even before band problems with Terra began around 2008, the Terra classification was slightly more likely to classify portions of desert playas, as well as rugged snow and rock covered areas of the high Sierra Nevada, as cloud. (There is a slight possibility that the latter is due to an underestimate over these areas by the MODIS Aqua classification.) Another area with consistently low or negative values in these difference products occurs where there is real recession of coastal fog off the coast of California, but this is not substantial, due to the limited ~2 hour window between Terra and Aqua. In general, Aqua has slightly more cloud cover, and this is to be expected, given typical slight cloud increases in the afternoon globally (Stowe *et al*, 2002). Figure 24 is a graph of annual cloud frequencies from each satellite, including the annual difference between the frequencies from each satellite, for MODIS tile H08,V05. The difference is steady and small until 2008, when problems with the sensor on-board Terra increased. Interestingly, the differences are larger in 2008 and 2009 than in 2010 and 2011. It's possible that my Terra correction was optimized for the most recent Terra data, and "over-corrected" 2008 and 2009 Terra data, when the problems weren't quite as bad, creating a bigger problem. MODIS band 3 has two problems: one related to the "right side" of a swath at high zenith angles (actually a relatively continuous increase with zenith that can be ameliorated with a simple linear trend correction) and another that is more of a striping, which appears much harder to correct (and perhaps not as critical to correct). It's difficult to determine whether the true color information with corrections is providing enough benefit to overcome the problems.[4]

Overlaying annual cloud frequency from the custom classification of the 3 sinusoidal tiles of MODIS Aqua on top of the annual cloud frequency from the custom classification of AVHRR (Figure 25), one is challenged to see the boundary between the two. The three tiles appear as subtle parallelograms due to their original sinusoidal projection draped onto the geographic projection of the AVHRR data. This is a good sign for the accuracy of each product – especially the AVHRR product, which comes from a more problematic data set and brings lower expectations. Even the relatively low frequency in the AVHRR product off the coast of California (west of the diagonal MODIS boundary) from 1981-1999 data can be explained by the

[4] MODIS band 3 has two problems: one related to the "right side" of a swath at high zenith angles (actually a relatively continuous increase with zenith that can be ameliorated with a simple linear trend correction) and another that is more of a striping, which appears much harder to correct (and perhaps not as critical to correct). It's difficult to determine whether the true color information with corrections is providing enough benefit to overcome the problems.

relatively lower frequency of fog during a couple-decade long positive phase of the Pacific Decadal Oscillation that switched back to a negative phase in 1998 (Johnstone and Dawson, 2010). Regressions between MODIS cloud frequency (degraded to AVHRR resolution) and AVHRR cloud frequency for a grid of points for December and January support the general agreement (Figure 26 and Figure 27, respectively). The relationship is quite close to 1:1 for December (slope of regression 0.995, intercept of 0.01), with a fairly tight fit ($R^2 = 0.79$) except for a set of points with values of 0.2 to 0.4 in the MODIS result that have values of 0.3 to 0.55 in the AVHRR classification. These are mostly high desert areas in which the AVHRR product is likely overestimating cloud frequency, possibly due to confusion with the bright soils and/or snow cover. The fit is slightly tighter for January ($R^2 = 0.83$), but the relationship is no longer 1:1 (slope = 1.25) and the intercept is -0.13 (AVHRR classification overestimates at low end and underestimates at high end). Overall, coarse-level agreement is encouraging, but far from perfect, with obvious spatial and temporal variability.

Accuracy Assessment

The above comparison with MODIS data is a sort of validation of the AVHRR product, although not technically an accuracy assessment. No standard accuracy assessment of the AVHRR classification was performed. Its coarse resolution makes comparison with climate station data somewhat inappropriate. But visual comparisons with another MODIS cloud frequency product (i.e., NOAA Cloud Climatology Product) further support its accuracy. For example, in a comparison with the NOAA MODIS product over southern Mexico (Figure 28), the custom AVHRR classification (Figure 29) captures very similar patterns of higher values over mountain ranges, with similar but slightly blurrier detail, due to its much coarser resolution. This is an area where the NOAA product can be expected to be quite accurate, even with its simple brightness threshold-based methodology, due to the fact that few other bright cover types (e.g., deserts and snowy areas) are found in this area. It is encouraging that the result from the totally independent custom algorithm for AVHRR, with little reliance on brightness alone, shows very similar patterns.

Estimates of fog days from AVHRR compared somewhat favorably with those from the CIMIS climate station data. Overall agreement at the Shafter climate station was 76.9%. But much of this agreement was on non-foggy days, which are very easy to agree upon, simply by chance. If CIMIS data suggested fog, the AVHRR processing did not omit too many of these days, capturing these days 67.0% of the time as fog. AVHRR processing may have more substantial commission error though: only 37.3% of AVHRR fog days were considered fog by CIMIS processing. These numbers are all a little higher in the comparison for MODIS Terra fog days. Overall agreement is 84.7%, 72.4% of CIMIS fog days are classified as fog, and 48.3% of MODIS Terra fog days were considered fog by CIMIS processing. A corresponding Kappa value of 49.0% and a Tau value of 69.3% for (compared to 84.7% for overall agreement) reflect that a large portion of the agreement could be due to chance, especially the easy agreement on the many non-fog days. In general, the image-based approaches mapped much more fog than the CIMIS-based classification decision. It could be argued that the image processing is including a lot of cloudy days that aren't foggy at the surface. But visual comparison with some daily imagery suggested that the image-based model performs better than the CIMIS-based model. The CIMIS-based model may be hindered by the inability to use relative humidity data due to the inconsistency and inaccuracy of relative humidity measurement (Dai, 2006).

The classification trees that were used to assist with final rule generation for discriminating clouds from snow on MODIS data provide an indication of classification performance, at least on training data. Performance on training data tends to depend on how challenging the training pixels are. When approximately 20,000 snow and cloud pixels were selected from a variety of dates of imagery in a relatively automated manner (i.e., not targeted to challenging pixels, but dropped over preliminary snow and cloud classifications, with manual edits over any errantly classified pixels), over 95% of pixels were correctly classified as snow or cloud. On one hand, this could be considered an overestimate of general classification accuracy because the model is honing in on these training pixels (i.e., overfitting) and might perform worse on independent imagery. On the other hand, these were relatively challenging dates of imagery (with complex mixes of clouds and snow) and the model was not allowed to overfit; this model using R's default complexity parameter of .01 only had 6 splits and terminal nodes still had hundreds of samples. When classification trees were fed only 1,947 manually selected spectrally challenging pixels (e.g., ice clouds, shaded ice clouds, thin snow cover, and shaded/mixed snow cover), performance was a bit lower, at 84.3%. This is certainly a substantial underestimate of overall performance because it was restricted to challenging pixels, but does indicate that there is still some spectral inseparability of particularly problematic snow and clouds. It also demonstrates the sensitivity of these results to sampling.

Table 8 shows the average annual number of days with snow commission error in the custom snow classification over a variety of sites that should never be classified as snow. These values are generally only about 2 days per year, suggesting low overall confusion rates, given approximately 100 (over land) to 200 (over ocean) cloudy days per year, and the fact that this is likely the main source of cloud error. But these values can vary by site and by year. MODIS Aqua, in particular, appears to have a bit of a problem with the well-forested Santa Cruz Mountains of the central California coast, with average commission errors of over 4 per year. It is not surprising that certain cover types, such as bright playas, are more likely to be mis-classified. Confusion over darkly forested mountains is more surprising. It is likely due to weakness in separating snowy forested areas from some partially cloudy forested areas. [The standard MODIS snow product can suffer from the same problem of snow classification over California's coastal forests on cloud edges (personal observation: https://earthdata.nasa.gov/labs/worldview/).] My custom classification of MODIS data generally performs better for Terra than for Aqua over this particular forested area (and slightly better at most other sites) in terms of this type of cloud omission, except after 2009, when various radiometric problems with Terra became more pronounced and likely affected the results. It is unclear why the classification with early Terra performed better than that with early Aqua for this particular cloud omission, given the same classification algorithm. It is possible that problematic (i.e., ice) clouds are more common later in the day, during the Aqua overpass.

Only the MODIS products underwent a truly ground-based attempt at accuracy assessment. Table 9 shows the results for comparison between the three meteorological towers and the MODIS daily cloud classifications from both Terra and Aqua, both for the custom classification developed here and the standard product that accompanies the data. Accuracy here is simply defined as percentage overall agreement on a binary cloud classification. The accuracy figures are quite comparable among the sites, sensors, and classifications, generally hovering just

above 70%. These numbers are surprisingly low, and almost certainly not reflective of the performance of the cloud classification algorithms. Mountainous areas with frequent snow cover are difficult areas to map well consistently, but extensive visual assessment of individual daily cloud masks, as well as the final climatological products, suggests that these numbers are very misleading. In fact, the accuracy of the cloud classification from automated processing of the tower radiation data is probably much lower than that from the satellite data. Further disagreement also likely arises from the various slight spatial and temporal mismatches that can occur. It is probably more appropriate to use this data as a relative comparison among sensors and classification approaches than as an absolute assessment of classification performance.

Relative comparison reveals that the classification algorithms are performing comparably at the two lower elevation tower sites, but that the custom classification is performing 2-3% better (in absolute difference) than the standard classification at the highest site, both for Terra and Aqua. This is likely related to the known problems that the standard classification has with snow, which is quite common at this 2700 meter site. Both classifications perform notably worse (~3%) for Aqua than for Terra at this site. Otherwise, sensor and algorithm accuracy is quite comparable within a site. Overall cloud classification frequency does differ between the algorithms, across the sites. At the lower, 1160 meter site, the custom classification maps clouds with 4% and 7% less frequency than the standard classification, for Terra and Aqua respectively. Next, at the 2015 meter site, these figures jump to 19% and 14% less, respectively. Finally, at the highest site, these figures climb to 28% less for both sensors. Much of this may reflect the tendency for the standard algorithm to map snow cover as cloud. However, this could also be explained by the custom algorithm underestimating cloud cover at higher elevations – for example, if thin clouds over snow are classified as snow, or if particularly problematic types of ice clouds (classified as snow) are especially prevalent at the highest elevations.[5] But the actual cloud frequency totals indicate that the custom algorithm captures frequency accurately at the highest site – only at the lowest site does it substantially underestimate overall frequency (mostly for Aqua), at least in comparison to the tower processing.

Long Term Data Record (Consistency and Trends)

One of the goals of the LTDR project is to develop a relatively consistent daily global record over the 30-plus year period. Detecting land cover trends over time, such as changes in a vegetation index like NDVI, requires that radiometric fidelity does not change over time. Atmospheric effects need to be accounted for in this attempt, including the removal of clouds. Even if consistency in cloud classification were achieved (see below), inherent differences among the satellites would make direct comparison of cloud cover frequencies among satellites problematic. Most significantly, AVHRR's local overpass time does not perfectly correspond to a MODIS sensor's overpass time (it is closest to Aqua, but even MODIS overpass times vary a

[5] A visual assessment of the custom cloud frequency products of MODIS does suggest a possible underestimate of cloud frequency over many high elevation snowy areas - generally the highest ridges without substantial tree cover - of much of the western United States. These areas surprisingly have lower mapped cloud frequency than surrounding lower elevations. But comparison with some daily imagery suggests that these lower values over the highest elevations could be real – clouds often hover below the highest peaks. It is difficult to confirm whether this is a real result, or whether there is indeed a bias against classifying thin or partial clouds if there is apparent snow underneath.

little by day), and drifted later in the day over time until a reset with each new NOAA satellite launch (Figure 30 and Figure 31). Different NOAA satellites launched with local overpass times of ~1:30 to 2:30, but each tended to drift later ~3 minutes per month. This inconsistency might not matter for a variable like land cover NDVI, which shouldn't change much over the course of a day (but can because of artifacts introduced by varying sun-sensor viewing geometry). It can, however, substantially affect cloud cover frequencies, as demonstrated by the strong correlation observed by Stowe et al. (Stowe *et al*, 2002) in Figure 31. Some of this could be real increase due to meteorological factors, but this is likely a relatively small factor (Stowe *et al*, 2002). It is likely mostly due to both sun-sensor viewing geometry (higher solar zenith angles) and lower thermal radiance. It is difficult to standardize for all of these differences.

Both the LTDR project's standard cloud frequency products and the custom cloud frequency products developed here show yearly annual frequencies (averaged across the ~10,000 km by 5,000 km subset spanning southern North America plus adjacent oceans) that also closely correspond to yearly trends in AVHRR zenith angle and overpass time (Figure 32): increasing frequencies with increasing zenith angle and later in the day. The varying amounts each year are so closely correlated with the varying overpass times of the sensor that the variability itself is unlikely to contain any real signal. Accordingly, it is hard to have confidence that the computed linear increases over time (0.1% per year for LTDR and 0.2% per year for custom) indicate a real trend. The large differences in frequencies between the two approaches might be concerning as well: the LTDR product maps clouds with 57% greater frequency on average (25% in absolute terms), with the linear trend suggesting an average from 68% in 1981 to 70% in 1999, and an average of 42% in 1981 to 46% in 1999 for the custom product. Mapped versions of the trends suggest that inaccuracy in the LTDR project's standard classification of clouds (see Appendix 2.A.1) is hindering its ability to accurately portray changes.

Figure 33 shows that the 18 year (1982-1999) trend in February cloud frequency over much of North America according to the standard LTDR cloud mask has problems with snow-covered areas of southern Canada, the northern U.S., the Rocky Mountains and the Sierra Nevada. The consistent gray colored areas indicate an approximate 10% (absolute) increase over the 18 years that is certainly due to an increased tendency to classify February snow cover as cloud toward the end of the period. Any real trend would be unlikely to be so correlated with areas of frequent snow cover. Similarly, the 1981 to 1999 trend in annual cloud frequency for this product (Figure 34) shows some increases in the far northern plains (mostly purple color west of Lake Superior) that are likely biased by the snow problem. But more striking is the likely spurious increasing cloud frequency over southern Texas and northeastern Mexico (mostly white color). This "trend" occurs in most of the LTDR monthly versions, and is likely due to a developing problem with bright desert soils. (Green areas over far northern water bodies relate to the LTDR project not mapping cloud cover above 50 degrees latitude over water bodies.)

The LTDR product's mapped increase of cloud frequency over lower elevation desert areas of the southwest and northern Mexico makes the detection of a possible real decreasing trend in the region more difficult. The custom cloud frequency trend product maps this region as having fairly substantial decline in cloud cover across the board, from the Texas panhandle and southwestern Kansas through the southwest to the Pacific Ocean off southern and Baja California. These two trend products are otherwise somewhat similar across much of North

America and adjacent oceans. (It seems reasonable that two different cloud classification algorithms might agree on trends more than they agree on absolute frequencies, as the percentage disagreement may be fairly consistent across years.) It is encouraging that the custom product also shares some patterns with another long-term cloud trend product – that from a processing of ISCCP data (Figure 36): increases over the Pacific Northwest and offshore, as well as over the Great Plains of the U.S., with decreases over the Desert Southwest into northwestern Mexico and offshore, as well as for the Gulf of Mexico. (Some brown areas in Figure 35 match blue colors in Figure 36, indicating decreases, and some purple and white areas in Figure 35 match beige areas in Figure 36, indicating increases.) The ISCCP product might be viewed with some skepticism, however, due to noted problems with its consistency. The abrupt discontinuities between increases and decreases in the ISCCP trend product following geometric patterns are obviously problematic (e.g., Indian Ocean).

Other than through a comparison with AVHRR data for long-term trends in California Central Valley Tule fog frequency (Baldocchi and Waller, 2014), long-term trends were not seriously investigated with MODIS data. Part of this relates to the problems with Terra data after 2008, making close to a third of its data in this study suspect. The nine-plus year record for Aqua data here did not seem sufficient to warrant serious long term analysis; regular interannual variability would likely swamp the ability to detect any meaningful long-term trends.

Discussion

Most quantitative assessments performed in this study did not reveal the custom classifications to be substantial improvements over standard products. But that may relate to limitations in the assessments. Determining the accuracy and consistency of cloud products is as challenging as developing cloud products. The accuracy assessments need their own accuracy assessment. Qualitative assessment suggests that the custom classifications are, at a minimum, more consistent across cover types, an important consideration for a climatology product (e.g., one that is used for a relative comparison among areas, as for species distribution modeling). Consistency over time is not as clear. The Aqua-based custom product appears to have temporal consistency, but the Terra-based custom product was sensitive to radiometric problems in the data in recent years. Trends in the custom AVHRR classification appeared be less sensitive to spatial errors in classification than the standard AVHRR LTDR product, but have a different source of error related to insufficient treatment of variable and missing/bad AVHRR data.

It was determined that small overall increasing trends in cloud cover in the custom AVHRR product were somewhat suspect due to obvious interannual relationships with overpass times and solar zenith angles, but these are not the only factors that might be influencing apparent long term trends. While the number of days with missing images (Table 4 demonstrates that the LTDR dataset has more missing AVHRR data earlier in the record; there are many fewer missing days after 1990) was controlled for by computing trends based on monthly cloud percentages (monthly sums divided by days of available imagery) rather than monthly cloud sums, this correction does not account for other data issues that varied over time. Altogether missing days are not the only problem in the AVHRR record. As mentioned earlier (Data and Methods), there are many days on which imagery only covers some parts of the world, leaving

huge gaps with no data. These would appear to be "good" days in my processing because they were not empty or absent files and the data have no flag indicating their presence. These types of spatial gaps were more common earlier in the record, and bias any trend calculation that does not correct for these factors. There were also extensive periods when the thermal channels were problematic, which also affected the critical channel 3 "reflectance" product (most notably in the first 18 days of January, 1995). These problems were not among those anomalous data that I was able to account for (see Data and Methods) and have implications for the computed cloud frequency and the detection of subtle trends, in particular. The LTDR standard cloud mask attempted to map clouds even when the thermal problem occurred (the reasonable performance probably indicates some awareness of the problem; otherwise I would expect any algorithm to have major error, as for mine). But the LTDR cloud masks in these cases appear to have higher error than usual. In terms of generating accurate cloud climatologies, it may be more appropriate to avoid and keep track of these bad data days than to try to map clouds in spite of problems.

An accurate mapping of the frequency of "good" or "bad" data would be helpful. But it was difficult to determine the best set of rules for tracking these problem data. They were often characterized by particular values in different channels (e.g., zero, -9999, 32768). Generating frequency products for these different values in the different channels resulted in slightly different products. (For some reason, the difference between two of these "bad data" frequency products made for a pretty good desert mask.) It was difficult to determine which of these products was optimal. Corrections to the computed long-term trends based on these "bad data" frequency products was considered but never applied because of concerns about their accuracy. An alternative correction to the AVHRR trend products based on the assumption that there was a 9.5% increase in good data over the 18 year record (using the increase from 42% to 46% and assuming that there was no true change: (46-42)/42 = .095) did not substantially alter spatial patterns in the trend map, although it did decrease the trend values slightly across the board.

The custom classification of AVHRR cloud cover improved upon the standard LTDR product over snow-covered regions, but there is likely remaining confusion between cloud and snow. Custom MODIS products, benefiting from the various SWIR channels, further improve upon this discrimination. But even with MODIS, some confusion remains, especially for rugged canyons of mountainous regions, where shadowed mixes of rock and vegetation with snow are difficult to separate from likely similar mixes with clouds. Similar problems are evident in other MODIS-derived snow and cloud products, such as those for a Sierra Nevada hydrologic study group with their own custom classification (http://zero.eng.ucmerced.edu/snow/SN_view/SN_remote/sn.php). A quick query of this product's January, 2010 "Snow and Cloud" over the Sierra Nevada, for example, shows the rugged eastern edge of the snowy Sierra Nevada to often be errantly classified as cloud. These products focus on snow, but generate cloud maps in the process (Dozier *et al*, 2008). The error in classifying sparse snow cover on the edge of the Sierra as cloud is not of paramount concern for them. The MOD09 cloud state variable (from the MOD35 cloud product) appears to avoid this particular spectral ambiguity – probably through the use of thermal information – one reason that they're only available at the 1 km resolution of their thermal data. (The MOD35 product does not tend to have this particular error, but has no shortage of others.)

Cloud mask products generated here may be substantial improvements over existing products, but still have some subtle errors. Unfortunately, even subtle errors, if systematic, can add up over time. Sometimes these errors are only evident after the processing of multiple years of imagery. For example, there may be a slight bias toward greater classification of cloud cover over water bodies than over land. This is likely because thin clouds are much easier to see over water than over land. It makes sense to try to capture these clouds. But a failure to capture corresponding levels of clouds over land leads to a bias. This bias is especially noticeable over select inland water bodies – those where it is particularly unlikely to have higher frequencies of cloud cover than surrounding areas (e.g., Salton Sea and Gulf of California; unlike more continental water bodies where water temperatures often exceed air temperatures, leading to clouds forming specifically over the water, thick clouds do not appear to form specifically over these desert water bodies – although they do get some kind of haze that, in combination with either a low sun or perhaps high satellite zenith angle, looks a bit like a cloud. This same haze would not get called cloud if it occurred over land.). Similar subtle biases appear to occur over other land cover types, such as dark green vegetation (perhaps for similar reasons? – hazy clouds over dark green vegetation are easier to see). Some of these biases can be difficult to discern, as forested areas (and some water bodies) might be expected to have higher cloud cover than other areas anyway. On the other hand, some areas may be biased low in their cloud cover frequency. For example, with extensive focus on differentiating ice clouds and snow, there may not have been enough focus on capturing thin clouds over snow, with the result that frequently snow-covered areas have underestimates in their cloud frequency.

The background-caused bias could be reduced by restricting the classification to more obvious clouds (although possibly introducing a negative bias if thin, hazy clouds have a tendency to form over particular cover types). That is because there is little confusion with bright white clouds. This might lead to better looking products, in terms of avoiding land cover bias. But, of course, it would involve a substantial overall underestimate of cloud cover frequency. Optimizing the trade-off among accuracy in overall cloud cover frequency (balancing omission and commission errors), overall classification accuracy (avoiding omission and commission errors), and consistency in cloud cover classification among background cover types (i.e., avoiding or balancing any systematic errors of omission or commission) is a delicate balancing task.

Isssues with MODIS Compositing Along Scene Overlap Boundary

Problems with the various MODIS cloud products manifest themselves in some scene overlap areas, where clouds are not always removed by the compositing algorithm. Clouds are sometimes selected instead of snow, and worse, sometimes selected instead of water. Part of the problem may relate to the use of different cloud mask products by different MODIS processing groups. Pronounced brightness variability due to viewing geometry effects is also evident. These factors suggest that it might just be preferable to have scenes stitched together along an obvious line. This would also avoid an artificial bias low in cloud frequency that is caused by a cloud removing algorithm for scene overlap areas.

Cloud compositing problems are not the only issue in the scene overlap area. 1 km sensor zenith and azimuth products do not perfectly align with the viewing geometry related brightness effects apparent in the 500-m data, and this does not appear to be related to resampling and/or

resolution differences. It seems that different algorithms may have been applied at different stages. Appropriate application of the viewing-geometry related corrections for the problematic MODIS Terra blue band is difficult in scene overlap areas due to the problems with the mis-alignment of the 1 km sensor geometry products and the 500-m reflectance products.

Cloud Shadows

This work did not attempt to map cloud shadows. These are challenging to separate from water and shadowed terrain. Time series rules, reliable water masks, and the use of a digital elevation model could help with refinement, but would be a challenge, and mapping both clouds and cloud shadow would be a form of double counting. The exception would be cloud-shadowed clouds, which would not be a form of double counting in terms of vertical projection onto the land surface. An attempt was made to map these particular cloud shadows. Technically, the location of cloud shadows is more important to plants than the vertical projection of clouds, but this is likely a very minor difference spatially, and unlikely to lead to any major systematic error. Recognition of cloud shadow is more important for land surface observations than for cloud climatology.

Conclusions

New cloud climatologies have been generated from custom cloud classification of AVHRR and MODIS (both Terra and Aqua) satellite data. Three sinusoidal grid tiles of MODIS data for 2000 to 2012 were processed, covering much of the western United States at 500 meter resolution. Additional tiles could be similarly processed, space permitting, providing more extensive spatial coverage. The AVHRR product, while at a coarser ~5-km resolution, extends the record back to 1981, allowing the long-term confirmation of patterns observed in the MODIS products or the broad assessment of trends, and providing convenient global coverage if desired (although with limited confidence over certain desert and polar areas, and the focus here was on a subset covering most of North America and adjacent ocean areas). Each custom product appears to offer improvements to existing cloud classifications derived from the respective sensors, at least in terms of consistency among land surface types, a very important consideration in a cloud climatology. Problems in all sensors have contributed to the need for substantial quality assurance and quality control, but most issues have been somewhat fixable. MODIS Aqua data, in particular, appear to have the fewest problems, and the most reliable cloud classification. Substantial differences across AVHRR and MODIS (spectral sampling differences, data quality, overpass times, spatial resolution, etc.) make the detailed assessment of trends across the satellites somewhat suspect, but not inconceivable (e.g., these products have been used for a long term assessment of trends in California Central Valley Tule fog). Even the reliability of trends within the AVHRR LTDR (1981-1999) may be questionable. Trends for MODIS data would be more reliable, but the record (especially for the even more reliable Aqua data) may not yet be quite long enough to detect meaningful climatic signals compared to regular interannual variability. Nevertheless, these data capture spatial patterns of cloud cover not previously observed due to previous efforts' greater confusion with snow and other bright surfaces, such as desert soils. These problems would affect the performance of other cloud climatology products in ecological analyses, such as the prediction of plant species distributions. Chapter 3 of this dissertation explores the use of the custom monthly cloud frequency products

developed here from MODIS Aqua data for the prediction of giant sequoia distribution. It is likely that only with an accurate classification of clouds, without confusion between clouds and snow or rock in the snowy Sierra Nevada, could a cloud climatology demonstrate the importance of cloud cover to giant sequoia distribution.

References

Ackerman, S.A., Strabala, K.I., Menzel, W.P., Frey, R.A., Moeller, C.C. & Gumley, L.E. (1998) Discriminating clear sky from clouds with MODIS. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **103**, 32141-32157.

Ackerman, T.P. & Stokes, G.M. (2003) The Atmospheric Radiation Measurement Program-To predict reliably what increased greenhouse gases will do to global climate, we have to understand the crucial role of clouds. *Physics Today*, **56**, 38-46.

Baldocchi, D. & Waller, E. (2014) Winter fog is decreasing in the fruit growing region of the Central Valley of California. *Geophysical Research Letters*.

Beyer, H.L. (2004) Hawth's Analysis Tools for ArcGIS., **2012**.

Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. (1984) Classification and regression trees. Wadsworth & Brooks. *Monterey, CA*.

Coffin, M. & Sukhatme, S. (1997) Receiver operating characteristic studies and measurement errors. *Biometrics*, 823-837.

Commission for Environmental Cooperation (Montréal, Québec). & Secretariat (1997) *Ecological regions of North America: toward a common perspective*, The Commission.

Dai, A. (2006) Recent climatology, variability, and trends in global surface humidity. *Journal of Climate*, **19**.

Dai, A., Karl, T.R., Sun, B. & Trenberth, K.E. (2006) Recent trends in cloudiness over the United States: A tale of monitoring inadequacies. *Bulletin of the American Meteorological Society*, **87**.

Douglas, M., Beida, R. & Dominguez, A. (2010) Developing high spatial resolution daytime cloud climatologies for Africa., **2**.

Dozier, J., Painter, T.H., Rittger, K. & Frew, J.E. (2008) Time–space continuity of daily maps of fractional snow cover and albedo from MODIS. *Advances in Water Resources*, **31**, 1515-1526.

Evan, A.T., Heidinger, A.K. & Vimont, D.J. (2007) Arguments against a physical long-term trend in global ISCCP cloud amounts. *Geophysical Research Letters*, **34**.

Foody, G.M. (1992) On the compensation for chance agreement in image classification accuracy assessment. *Photogrammetric Engineering and Remote Sensing*, **58**, 1459-1460.

Foster, P. (2001) The potential negative impacts of global climate change on tropical montane cloud forests. *Earth-Science Reviews*, **55**, 73-106.

Frey, R.A., Ackerman, S.A., Liu, Y., Strabala, K.I., Zhang, H., Key, J.R. & Wang, X. (2008) Cloud detection with MODIS. Part I: Improvements in the MODIS cloud mask for Collection 5. *Journal of Atmospheric & Oceanic Technology*, **25**.

Gentry, A.H. (1992) Tropical forest biodiversity: distributional patterns and their conservational significance. *Oikos*, 19-28.

Goldsmith, G.R., Matzke, N.J. & Dawson, T.E. (2013) The incidence and implications of clouds for cloud forest plant water relations. *Ecology Letters*, **16**, 307-314.

Goulden, M., Anderson, R., Bales, R., Kelly, A., Meadows, M. & Winston, G. (2012) Evapotranspiration along an elevation gradient in California's Sierra Nevada. *Journal of Geophysical Research: Biogeosciences (2005–2012)*, **117**.

Hall, D.K., Riggs, G.A. & Salomonson, V.V. (1995) Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data. *Remote Sensing of Environment*, **54**, 127-140.

Hall, D.K., Riggs, G.A., Salomonson, V.V., DiGirolamo, N.E. & Bayr, K.J. (2002) MODIS snow-cover products. *Remote Sensing of Environment*, **83**, 181-194.

Hall, D.K. & Riggs, G.A. (2007) Accuracy assessment of the MODIS snow products. *Hydrological Processes*, **21**, 1534-1547.

Hansen, J.E., Sato, M., Lacis, A., Ruedy, R., Tegen, I. & Matthews, E. (1998) Climate forcings in the industrial era. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 12753-12758.

Heidinger, A.K., Evan, A.T., Foster, M.J. & Walther, A. (2012) A Naive Bayesian Cloud-Detection Scheme Derived from CALIPSO and Applied within PATMOS-x. *Journal of Applied Meteorology & Climatology*, **51**.

Heidinger, A.K., Foster, M.J., Walther, A. & Zhao, X. (2013) The pathfinder atmospheres extended (PATMOS-x) AVHRR climate data set. *Bulletin of the American Meteorological Society*,.

Hickman, J.C. (1993) *The Jepson manual: higher plants of California.*, University of California Press.

Hilker, T., Lyapustin, A.I., Tucker, C.J., Sellers, P.J., Hall, F.G. & Wang, Y. (2012) Remote sensing of tropical ecosystems: Atmospheric correction and cloud masking matter. *Remote Sensing of Environment*, **127**, 370-384.

Jacobowitz, H., Stowe, L.L., Ohring, G., Heidinger, A., Knapp, K. & Nalli, N.R. (2003) The Advanced Very High Resolution Radiometer Pathfinder Atmosphere (PATMOS) climate dataset: A resource for climate research. *Bulletin of the American Meteorological Society*, **84**.

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 of the United States of America*, **107**, 4533-4538.

Leinenkugel, P., Kuenzer, C. & Dech, S. (2013) Comparison and enhancement of MODIS cloud mask products for Southeast Asia. *International Journal of Remote Sensing*, **34**, 2730-2748.

Liu, R. & Liu, Y. (2013) Generation of new cloud masks from MODIS land surface reflectance products. *Remote Sensing of Environment*, **133**, 21-37.

Lyapustin, A., Wang, Y. & Frey, R. (2008) An automatic cloud mask algorithm based on time series of MODIS measurements. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **113**.

Ma, Z. & Redmond, R.L. (1995) Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing*, **61**, 435-439.

Miller, S.D., Lee, T.F. & Fennimore, R.L. (2005) Satellite-Based Imagery Techniques for Daytime Cloud/Snow Delineation from MODIS. *Journal of Applied Meteorology*, **44**.

Norris, J.R. (2000) What can cloud observations tell us about climate variability? *Space Science Reviews*, **94**, 375-380.

Painter, T.H., Rittger, K., McKenzie, C., Slaughter, P., Davis, R.E. & Dozier, J. (2009) Retrieval of subpixel snow covered area, grain size, and albedo from MODIS. *Remote Sensing of Environment*, **113**, 868-879.

Pedelty, J., Devadiga, S., Masuoka, E., Brown, M., Pinzon, J., Tucker, C., Roy, D., Ju, J., Vermote, E. & Prince, S. (2007) Generating a long-term land data record from the AVHRR and MODIS instruments., 1021-1025.

Pinker, R.T., Zhang, B. & Dutton, E.G. (2005) Do satellites detect trends in surface solar radiation? *Science (New York, N.Y.)*, **308**, 850-854.

Platnick, S., King, M.D., Ackerman, S.A., Menzel, W.P., Baum, B.A., Riédi, J.C. & Frey, R.A. (2003) The MODIS cloud products: Algorithms and examples from Terra. *Geoscience and Remote Sensing, IEEE Transactions on*, **41**, 459-473.

Ramanathan, V., Cess, R.D., Harrison, E.F., Minnis, P., Barkstrom, B.R., Ahmad, E. & Hartmann, D. (1989) Cloud-radiative forcing and climate: results from the Earth radiation budget experiment. *Science (New York, N.Y.)*, **243**, 57-63.

Riggs, G.A. & Hall, D.K. (2004) Snow and cloud discrimination factors in the MODIS snow algorithm., **6**, 3714-3716.

Roger, J. & Vermote, E. (1997) Computation and use of the reflectivity at 3.75 μm from AVHRR thermal channels. *Remote Sensing Reviews*, **15**, 75-98.

Roger, J. & Vermote, E. (1998) A Method to Retrieve the Reflectivity Signature at 3.75< i> μ m from AVHRR Data. *Remote Sensing of Environment*, **64**, 103-114.

Rossow, W.B. & Schiffer, R.A. (1991) ISCCP cloud data products. *Bulletin of the American Meteorological Society*, **72**, 2-20.

Rouse Jr, J., Haas, R., Schell, J. & Deering, D. (1974) Monitoring vegetation systems in the Great Plains with ERTS. *NASA special publication*, **351**, 309.

Schiffer, R. & Rossow, W.B. (1983) The International Satellite Cloud Climatology Project(ISCCP)- The first project of the World Climate Research Programme. *American Meteorological Society, Bulletin*, **64**, 779-784.

Sklenár, P., Bendix, J. & Balslev, H. (2008) Cloud frequency correlates to plant species composition in the high Andes of Ecuador. *Basic and Applied Ecology*, **9**, 504-513.

Stowe, L.L., Davis, P.A. & McClain, E.P. (1999) Scientific basis and initial evaluation of the CLAVR-1 global clear/cloud classification algorithm for the Advanced Very High Resolution Radiometer. *Journal of Atmospheric & Oceanic Technology*, **16**.

Stowe, L.L., Jacobowitz, H., Ohring, G., Knapp, K.R. & Nalli, N.R. (2002) The Advanced Very High Resolution Radiometer (AVHRR) Pathfinder Atmosphere (PATMOS) climate dataset: Initial analyses and evaluations. *Journal of Climate*, **15**.

Sugden, A.M. & Robins, R.J. (1979) Aspects of the ecology of vascular epiphytes in Colombian cloud forests. I. The distribution of the epiphytic flora. *Biotropica*.

The MathWorks, Inc. (2011) MATLAB 2011a.

Therneau, T.M., Atkinson, B. & Ripley, B. (2010) rpart: Recursive partitioning. *R package version*, **3**.

Tucker, C.J. (1979) Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, **8**, 127-150.

Tucker, C.J., Pinzon, J.E., Brown, M.E., Slayback, D.A., Pak, E.W., Mahoney, R., Vermote, E.F. & El Saleous, N. (2005) An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *International Journal of Remote Sensing*, **26**, 4485-4498.

Wang, C., Qu, J.J., Xiong, X., Hao, X., Xie, Y. & Che, N. (2006) A new method for retrieving band 6 of Aqua MODIS. *Geoscience and Remote Sensing Letters, IEEE*, **3**, 267-270.

Wang, X., Xie, H. & Liang, T. (2008) Evaluation of MODIS snow cover and cloud mask and its application in Northern Xinjiang, China. *Remote Sensing of Environment*, **112**, 1497-1513.

Wilson, A.M., Parmentier, B. & Jetz, W. (2014) Systematic land cover bias in Collection 5 MODIS cloud mask and derived products—A global overview. *Remote Sensing of Environment*, **141**, 149-154.

Wilson, A.M., Parmentier, B. & Jetz, W. (2014) Systematic land cover bias in Collection 5 MODIS cloud mask and derived products—A global overview. *Remote Sensing of Environment*, **141**, 149-154.

Wylie, D., Jackson, D.L., Menzel, W.P. & Bates, J.J. (2005) Trends in global cloud cover in two decades of HIRS observations. *Journal of Climate*, **18**.

Wylie, D.P. & Menzel, W.P. (1999) Eight years of high cloud statistics using HIRS. *Journal of Climate*, **12**.

Tables

Table 1. AVHRR (LTDR AVH09C1) band characteristics.
Bands 1 and 2 are surface reflectance, hence LTDR term "SREFL". Bands 4 and 5 measure thermal emissions or brightness temperatures, hence LTDR term "BT". Band 3 measures a region of both reflected and emitted radiance, and accordingly, two products are provided.

Band	Wavelength Range (µm)	Common Term	LTDR Term
1	0.58-0.68	Visible or Red (VIS)	SREFL_CH1
2	0.725-1.1	Near Infrared (NIR)	SREFL_CH2
3	3.55-3.93	Middle Infrared (MIR)	BT_CH3 & SREFL_CH3
4	10.5-11.3	Longwave Infrared (LWIR)	BT_CH4
5	11.5-12.5	Longwave Infrared (LWIR)	BT_CH5

Table 2. MODIS surface reflectance (MOD09GA and MYD09GA) band characteristics.
All bands are surface reflectance products, hence MOD09GA (or MYD09GA) term "Sur_refl".

Band	Wavelength Range (µm)	Common Term	MOD09GA HDF Term
1	0.62-0.67	Red	Sur_refl_b01
2	0.841-0.876	Near Infrared (NIR)	Sur_refl_b02
3	0.459-0.479	Blue	Sur_refl_b03
4	0.545-0.565	Green	Sur_refl_b04
5	1.23-1.25	Shortwave Infrared (SWIR)	Sur_refl_b05
6	1.628-1.652	Shortwave Infrared (SWIR)	Sur_refl_b06
7	2.105-2.155	Shortwave Infrared (SWIR)	Sur_refl_b07

Table 3. AVHRR LTDR: NOAA satellite dates.
NOAA satellites have gradually reduced utility for cloud mapping over time due to continuous drift later in day. Data in later periods have overpass times when solar zenith angles high – extreme case ending with a data gap at the end of 1994.

Satellite	Start Date	Finish Date
NOAA-7	6/14/1981	12/31/1984
NOAA-9	1/4/1985	11/7/1988
NOAA-11	11/8/1988	9/13/1994
NOAA-14	1/1/1995	12/31/1999

Table 4. AHVRR LTDR: "Good" days by month.
Data regularly available after late July of 1981. Within windows of data availability, data gaps occur. Other than the big gap between satellites at the end of 1994, these gaps become less frequent by March of 1991, although there are still occasional missing days in a few months.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1981	0	0	0	0	0	5	11	31	30	31	30	31
1982	30	28	30	23	30	30	29	30	28	31	30	31
1983	31	28	31	30	31	30	30	30	30	31	30	31
1984	29	27	29	28	31	30	30	31	30	31	30	28
1985	26	28	30	30	31	30	31	31	30	31	29	31
1986	31	27	29	30	31	30	31	31	30	31	30	31
1987	31	28	31	30	31	30	31	31	30	31	30	31
1988	30	29	27	30	29	29	25	30	29	30	26	31
1989	31	28	29	29	31	30	31	31	30	31	30	31
1990	29	27	31	30	31	30	23	30	30	30	28	31
1991	22	25	31	30	31	30	31	31	30	31	30	31
1992	31	28	31	30	31	30	31	31	30	31	30	27
1993	31	28	30	30	31	30	31	31	30	31	30	31
1994	30	28	31	30	31	30	31	31	13	0	0	0
1995	31	28	31	30	31	30	31	31	30	31	30	31
1996	31	29	31	30	31	30	31	31	30	31	30	30
1997	31	28	31	30	31	30	31	31	30	31	30	31
1998	31	28	31	30	31	30	31	31	30	31	30	31
1999	31	28	31	30	31	30	31	31	30	30	30	31

Orange: 1-4 missing days

Red: >= 5 missing days

Table 5. MODIS MOD09 Terra data: "Good" days by month.
Regularly available after March, 2000, but with a few months early in the record (through end of 2003) with large data gaps. Occasional missing days after 2003.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2000	0	0	30	26	31	29	31	18	30	30	29	31
2001	31	28	31	30	30	14	29	31	30	31	30	31
2002	31	28	21	28	31	30	31	31	30	31	30	31
2003	31	27	31	30	31	30	31	31	30	31	30	22
2004	31	28	31	30	31	29	31	31	30	31	30	30
2005	31	28	31	29	31	30	31	31	29	31	30	31
2006	31	28	31	30	31	30	31	29	30	31	30	31
2007	31	28	31	30	31	30	31	31	30	31	30	30
2008	31	29	31	29	31	30	31	31	29	31	30	28
2009	31	28	31	30	31	30	31	30	29	31	30	31
2010	31	28	31	30	31	30	31	31	30	31	30	31
2011	31	28	31	30	31	30	31	31	30	31	30	31
2012	31	29	31									

Table 6. MODIS MYD09 Aqua data: "Good" days by month.
Regularly available after early September, 2002, with very few missing days.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2000												
2001												
2002	0	0	0	0	0	0	0	0	28	31	30	31
2003	31	28	31	30	30	30	31	31	30	31	30	31
2004	31	29	31	30	31	30	31	31	30	31	30	31
2005	31	28	31	30	31	30	31	31	30	31	30	31
2006	31	28	31	30	31	30	31	31	30	31	30	31
2007	31	28	31	30	31	30	31	31	30	31	30	30
2008	31	29	31	30	31	30	31	31	30	31	30	31
2009	31	28	31	30	31	30	31	31	30	31	30	31
2010	31	28	31	30	31	30	31	31	30	31	30	31
2011	31	28	31	30	31	30	31	31	30	31	30	31
2012	31	29	31									

Table 7. Common and custom normalized difference reflectance (or radiance) ratios.

Index Name	Equation	Sensor	Source
Vegetation Index (NDVI)	$\frac{\rho NIR - \rho Red}{\rho NIR + \rho Red}$	AVHRR,MODIS	(Rouse Jr *et al*, 1974; Tucker, 1979)
Snow Index (NDSI)	$\frac{\rho Band4 - \rho Band6}{\rho Band4 + \rho Band6}$	MODIS	(Hall *et al*, 1995)
nt	$\frac{radBand3 - radBand4}{radBand3 + radBand4 - 0.444}$	AVHRR	Custom
nd35	$\frac{\rho Band3 - \rho Band5}{\rho Band3 + \rho Band5}$	MODIS	Custom
nd37	$\frac{\rho Band3 - \rho Band7}{\rho Band3 + \rho Band7}$	MODIS	Custom
nd15	$\frac{\rho Band1 - \rho Band5}{\rho Band1 + \rho Band5}$	MODIS	Custom
nd15sub	$\frac{\rho Band1 - \rho Band5 - 0.2}{\rho Band1 + \rho Band5 - 0.2}$	MODIS	Custom
nd17	$\frac{\rho Band1 - \rho Band7}{\rho Band1 + \rho Band7}$	MODIS	Custom
b1ndev, b4ndev, b3ndev	$\frac{\rho Bandx - \rho avg}{\rho Bandx + \rho avg}$	MODIS	Custom

ρ = reflectance
NIR = near infrared (this is band 2 in both AHVRR and MODIS; red is band 1 in both)
rad = "radiance" (digital number proportional to radiance)
x = MODIS band1, band4, or band3 (red, green, and blue)
avg = (ρband1 + ρband4 + ρband3)/3

Table 8. MODIS custom classification average snow commission errors (days misclassified) by year, locations without snow, for both Terra and Aqua satellites.
These snow commission errors are likely omission errors in the cloud product. Given that snow is likely the largest source of error for the cloud mapping, these strongly relate to accuracy of cloud product.

Terra, Year	Ocean	Santa Cruz Mountains	San Joaquin Valley	Mono Lake	Lake Tahoe	Average
2001	0.71	1.34	0.78	0.43	0.25	**0.70**
2002	0.48	1.91	1.53	0.29	0.35	**0.91**
2003	0.45	1.9	2.15	0.12	0.4	**1.00**
2004	0.65	1.95	0.37	0.17	0	**0.63**
2005	1.09	2.26	1.88	0.73	0.92	**1.38**
2006	2.69	5.26	0.92	1.41	1.53	**2.36**
2007	1.69	2.38	1.66	0.56	0.31	**1.32**
2008	0.83	1.71	1.71	0.73	1.03	**1.20**
2009	1.93	4.06	2.08	0.9	2.57	**2.31**
2010	3.47	4.82	3.42	0.58	2.01	**2.86**
2011	3.55	3.8	2.37	2.2	2.8	**2.94**
Average	**1.59**	**2.85**	**1.72**	**0.74**	**1.11**	**1.60**

Aqua, Year	Ocean	Santa Cruz Mountains	San Joaquin Valley	Mono Lake	Lake Tahoe	Average
2003	1.54	4.38	1.23	1.91	0.99	**2.01**
2004	1.27	2.97	2.47	1.71	0.69	**1.82**
2005	1.2	2.56	2.72	2.33	1.16	**1.99**
2006	1.77	5.43	2.36	2.25	0.96	**2.55**
2007	1.26	2.19	0.44	1.68	1.4	**1.39**
2008	1.5	5.18	2.44	2.11	1.61	**2.57**
2009	1.62	3.66	1.03	2.16	2.31	**2.16**
2010	2.28	6.45	2.5	2.2	1.76	**3.04**
2011	1.55	6.49	3.15	1.49	1.69	**2.87**
Average	**1.55**	**4.37**	**2.04**	**1.98**	**1.40**	**2.27**

Table 9. Meteorological tower-based accuracy assessment of MODIS cloud classifications.
Generally similar performance except for increased accuracy (agreement) in custom classification may be apparent at 2700 meter site, likely due to better performance handling snow cover.

Tower Site	Sensor	Standard Classification	Custom Classification
1160 m	Terra	71.6	72.2
	Aqua	71.6	71.4
2015 m	Terra	73.5	73.0
	Aqua	75.0	74.4
2700 m	Terra	70.5	73.2
	Aqua	67.9	69.8

Figures

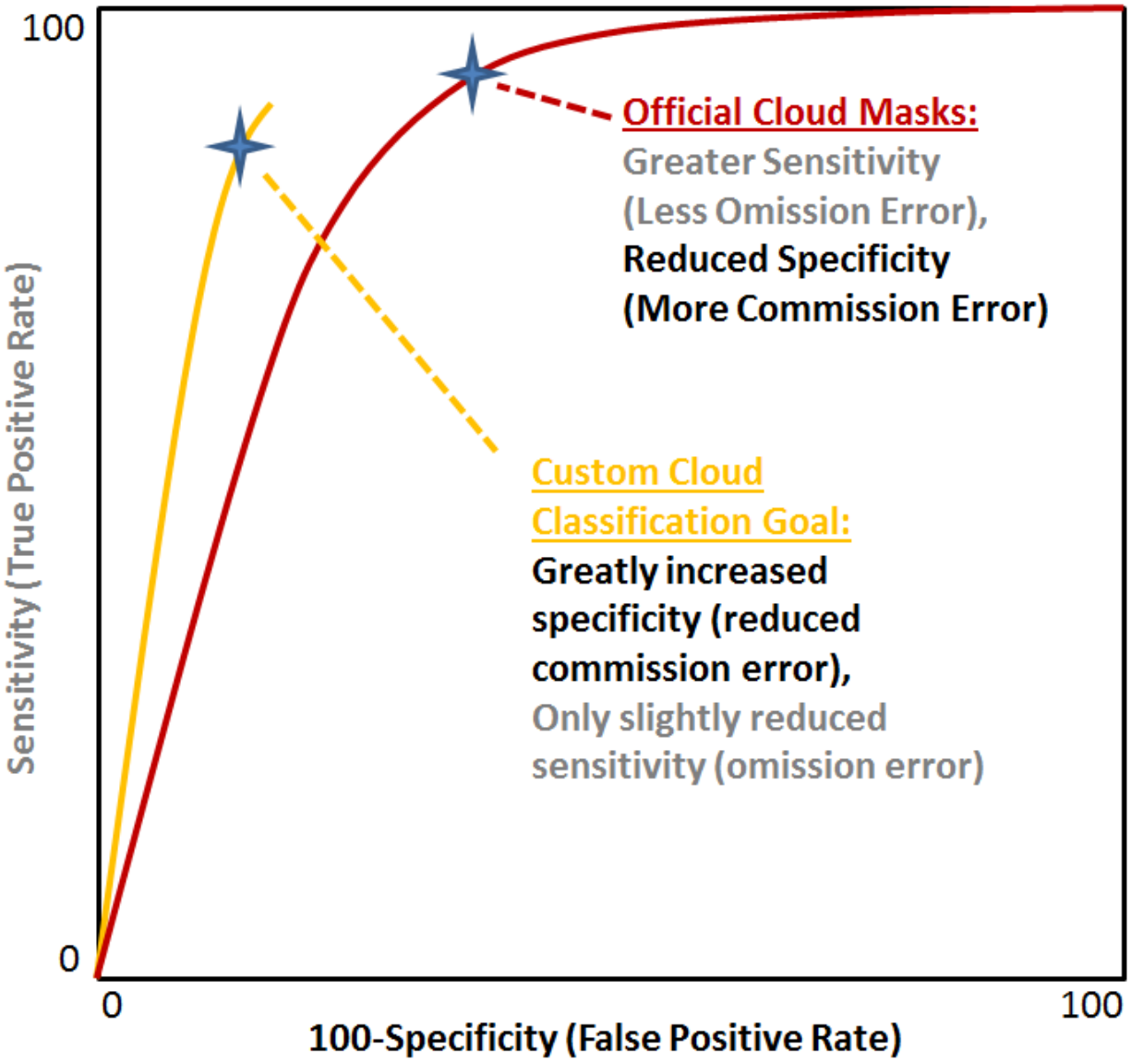


Figure 1. Conceptual receiver operating characteristic curves in relation to different cloud classification algorithms and thresholds.
Different algorithms are the red vs. orange curves, and different possible thresholds chosen are shown as stars. The goals of the custom cloud classification were to both improve discrimination, regardless of chosen thresholds (i.e., plot a steeper curve), and reduce commission error with confident identification of clouds. Omission error might increase slightly (on partial or thin clouds) but the net result would be closer to a perfect classification (top left corner).

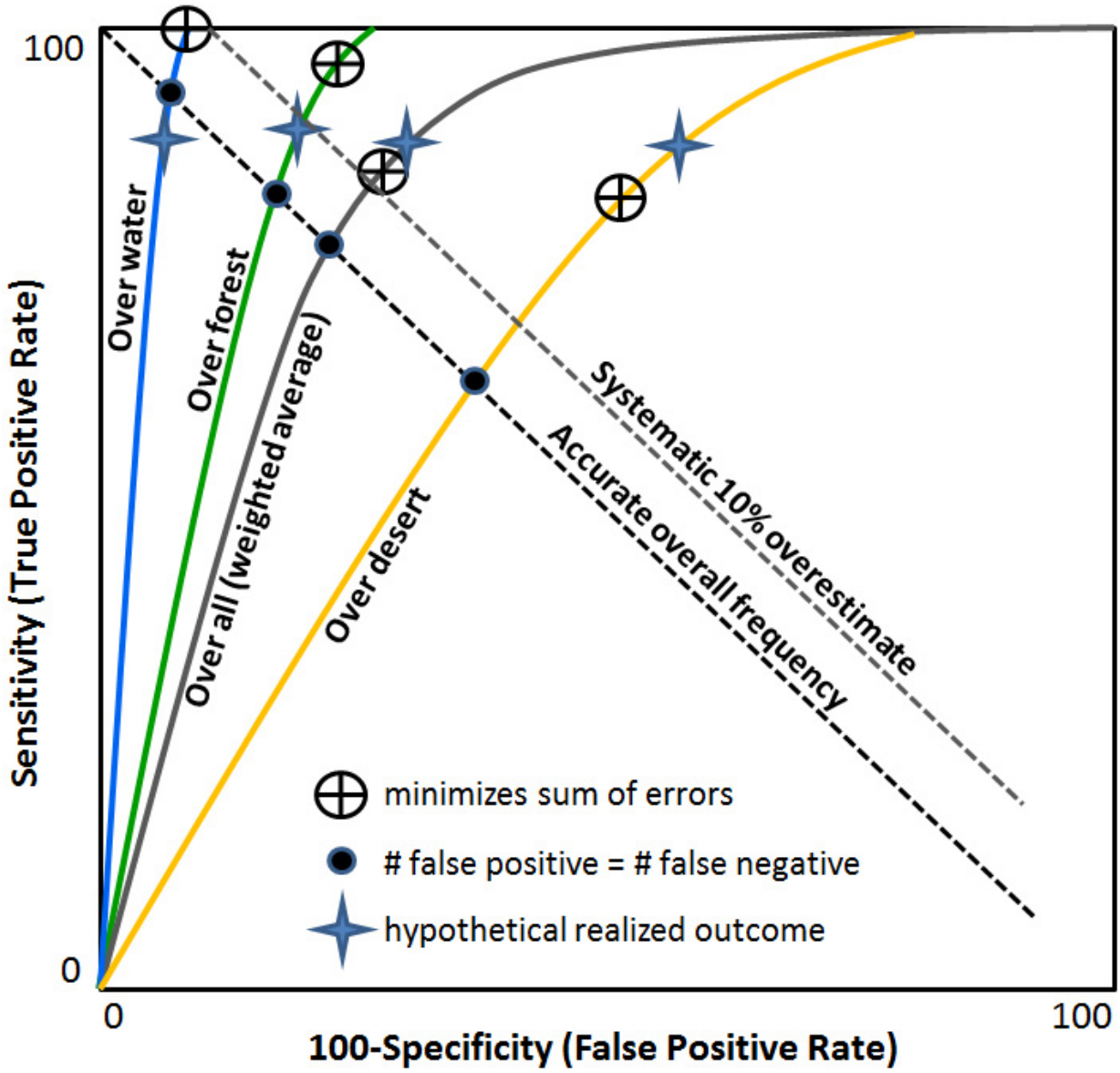


Figure 2. Balancing act in cloud classification goals and realizations: conceptual collection of receiver operator characteristic (ROC) curves for clouds over different surface types.
Steeper curves indicate better classification performance, with reduced omission error for a given commission error. Curves, and classification positions along the curve, are not independent of each other in practice, as different classification algorithms can be optimized for different underlying surfaces. Maximizing classification accuracy (minimizing the sum of errors, that is, false positives and false negatives) over different cover types does not necessarily balance false positives and false negatives (necessary to yield accurate overall frequencies - black dashed line). Actual realized outcomes may have relatively similar levels of omission error over different surfaces, but vastly different levels of commission error (as pictured with stars). This results in biased frequencies among surface types. If outcomes all fall along a given dashed diagonal line there will be no particular bias related to underlying cover types (whether balanced errors – black dashed line, general underestimates, or general overestimates – gray dashed line).

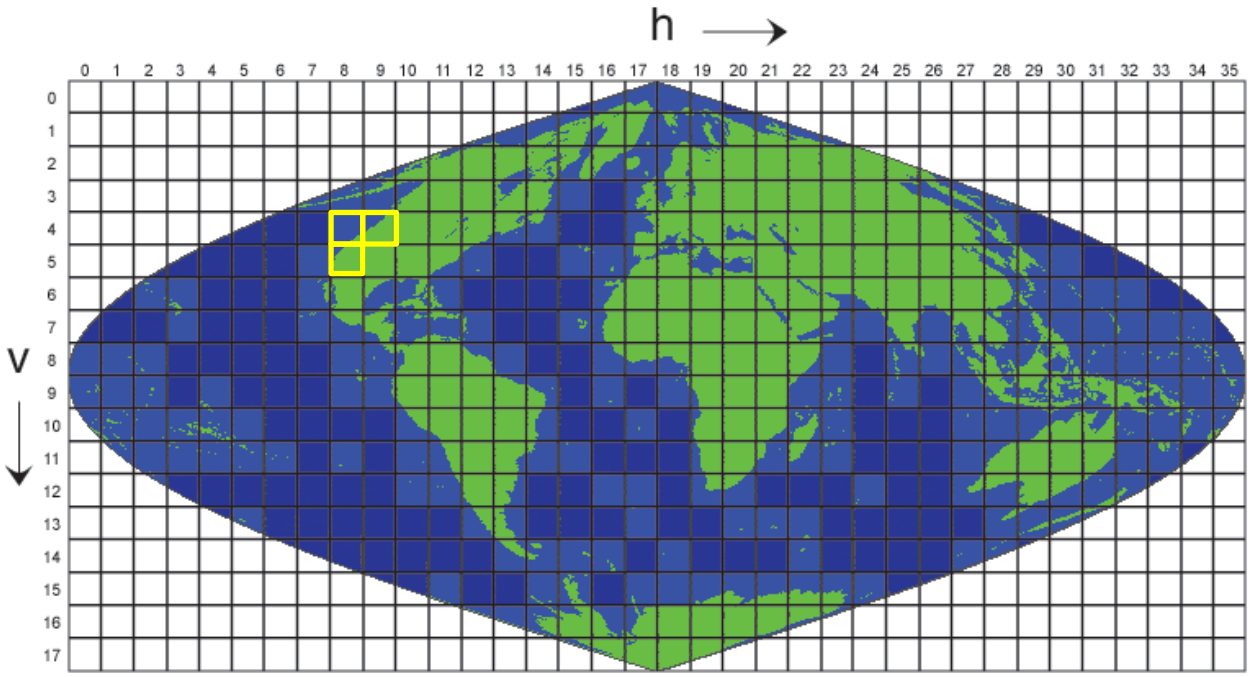


Figure 3. MODIS sinusoidal grid, with study tiles in yellow.
http://nsidc.org/data/modis/data_summaries/landgrid.html

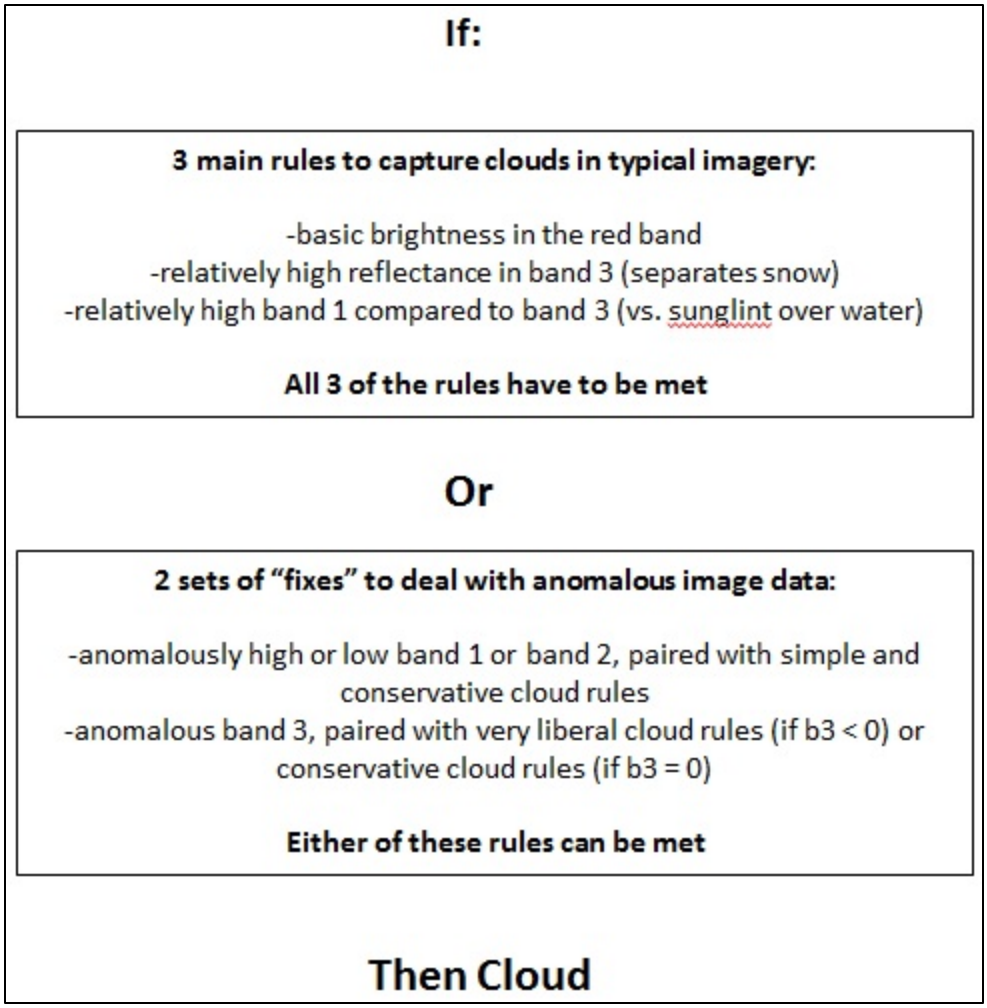


Figure 4. AVHRR processing: manually developed cloud rules from multi-image analysis. Basic overview of decision making process for discriminating clouds in AVHRR data. The first box comprises the standard processing rules. The second box comprises exceptions for anomalous data.

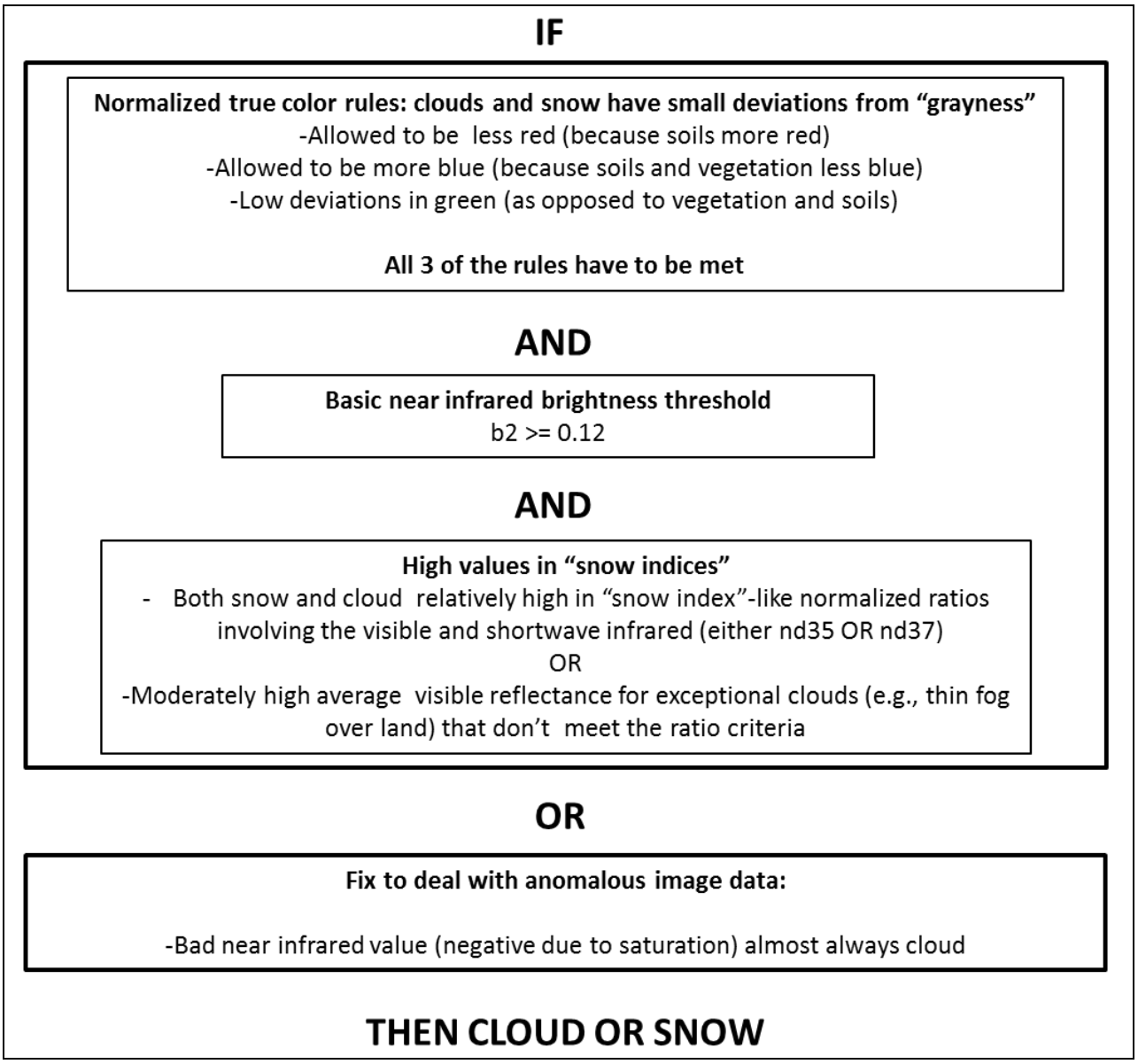


Figure 5. MODIS processing, 1st stage: manually developed rules for cloud OR snow from multi-image analysis.
First three boxes comprise rules for standard imagery. The second box accounts for anomalous data.

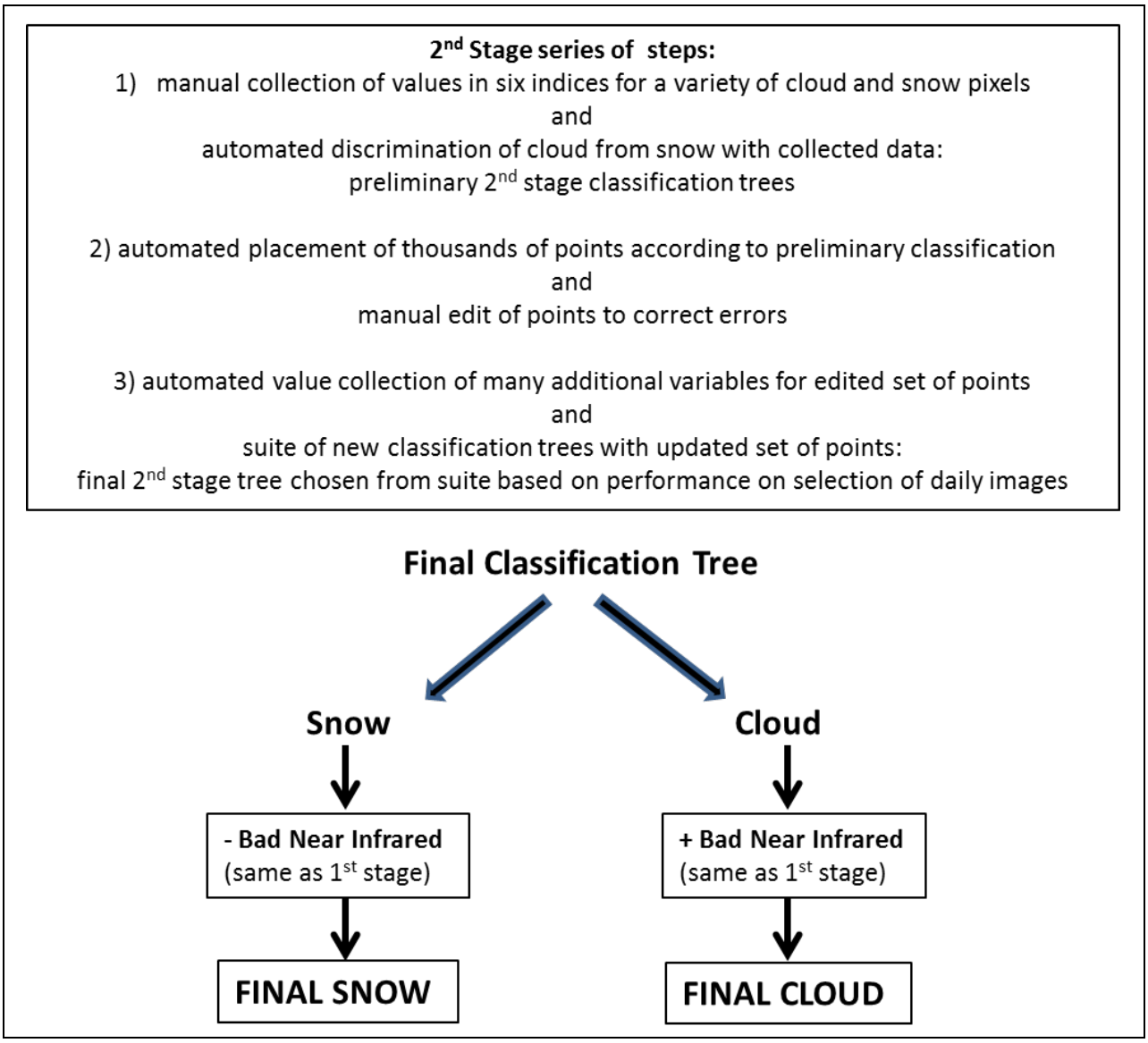


Figure 6. MODIS Processing, 2nd stage: a more automated rule-generation process with classification trees.

Big box comprises relatively automated steps with classification trees. Latter steps related to ensuring that bad (saturated) near infrared data classified as cloud rather than snow.

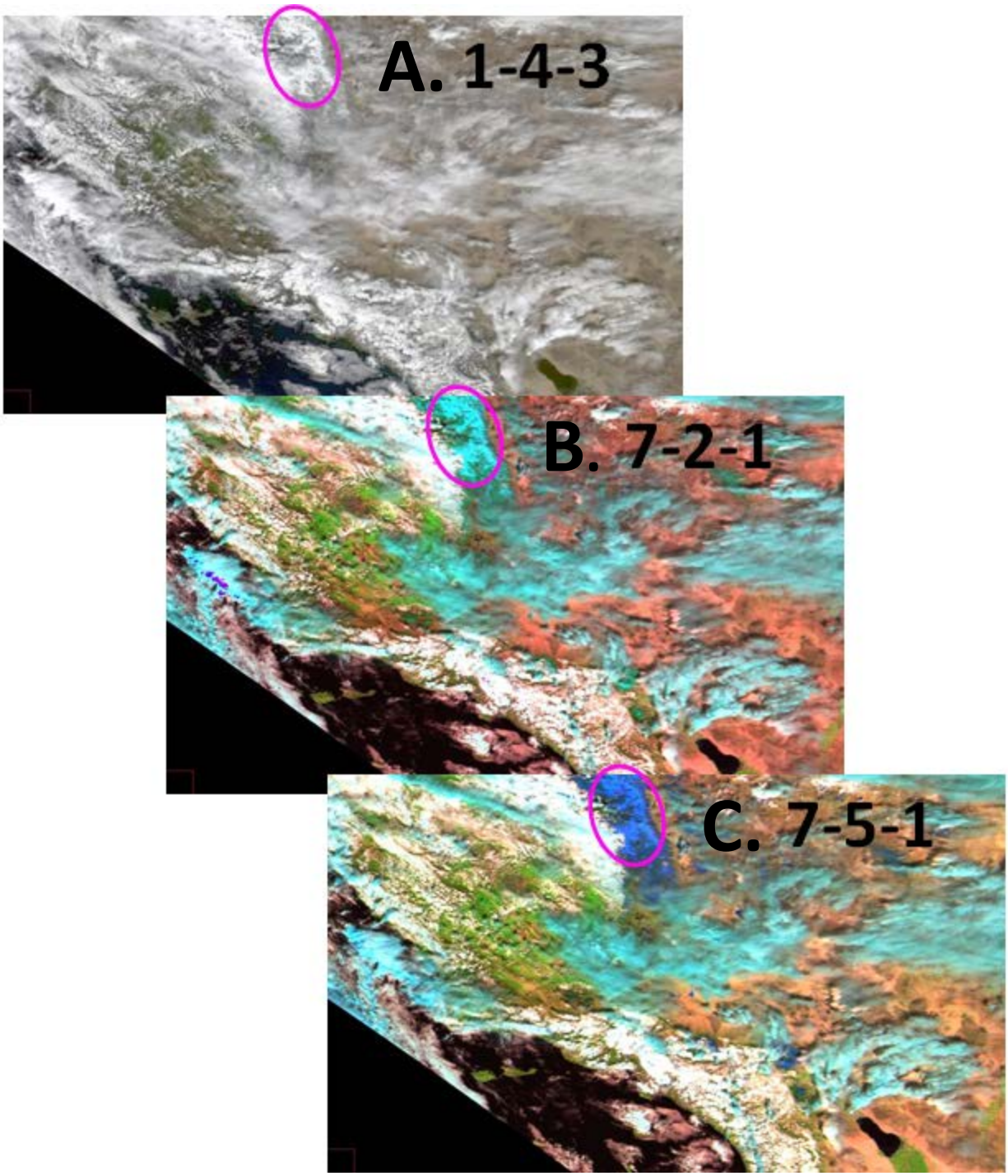


Figure 7. MODIS Aqua, southern California, February 19, 2010; color composite comparison. Comparison of true color display (A. 1-4-3) and false color (B. 7-2-1 and C. 7-5-1). Only in the 7-5-1 combination are snow-covered areas (within magenta ellipse, over the Sierra Nevada mountains) separable from virtually all clouds.

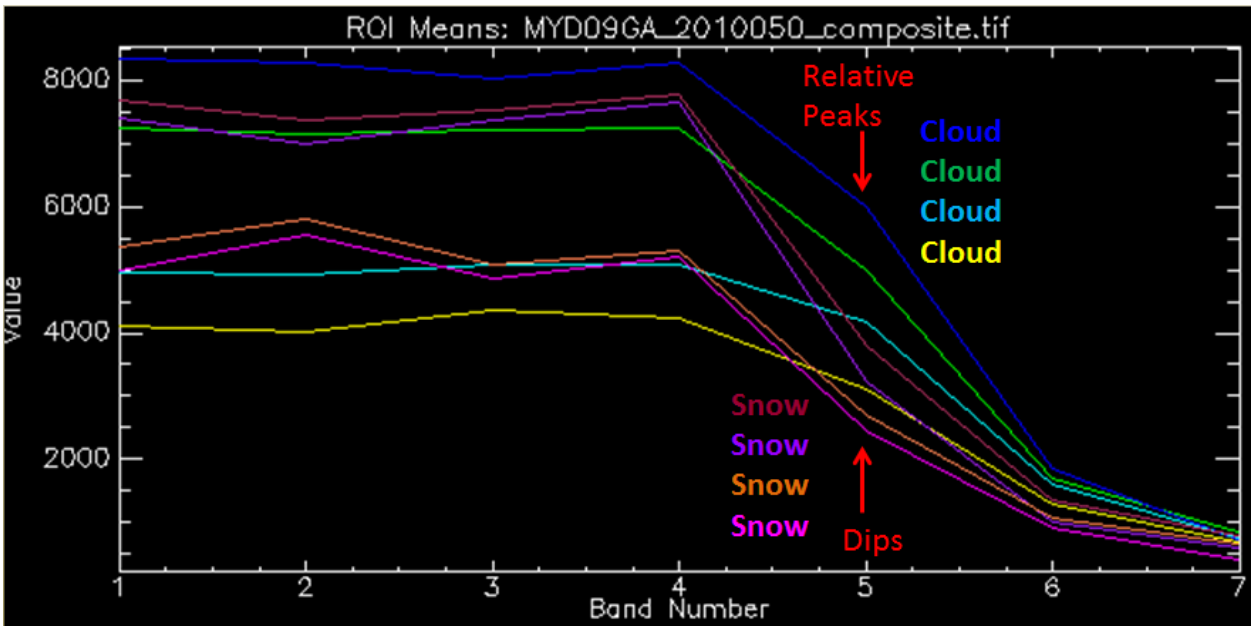


Figure 8. MODIS Aqua reflectance spectra, "ice" clouds and snow.
Percentage reflectance has been multiplied by 10000 to obtain the y-axis reflectance "Value". Note that the four ice cloud spectra have relative peaks in band 5, whereas the four snow spectra have relative dips.

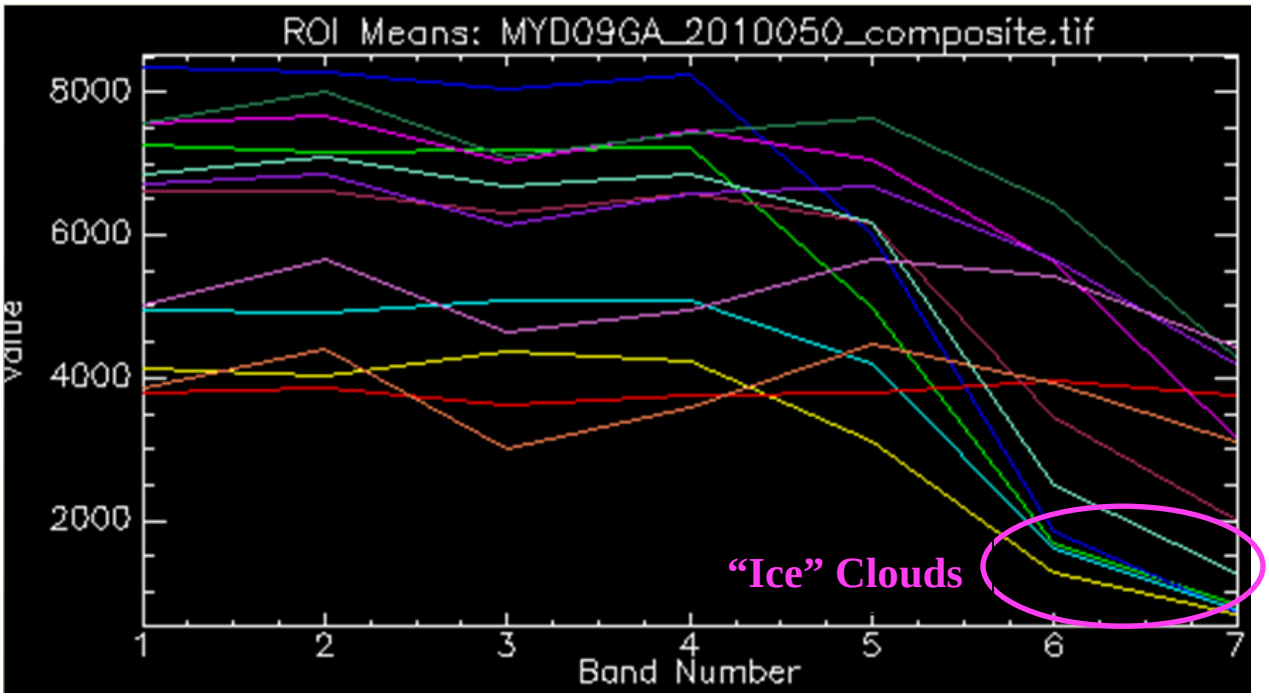


Figure 9. MODIS Aqua reflectance spectra, variety of clouds, including "ice" clouds.
As for above figure, percentage reflectance has been multiplied by 10000 to obtain the y-axis "Value". Ice clouds are unusually dark in bands 6 and 7, just like snow. There is no unique spectral signature of cloud.

Figure 10. MODIS Aqua, January 23, 2010. Tile H08,V05, true color.
Brightness alone might appear to be sufficient to separate clouds and snow from unobscured surface. In fact, many desert soils and urban areas are brighter than many snowy forests and darker clouds.

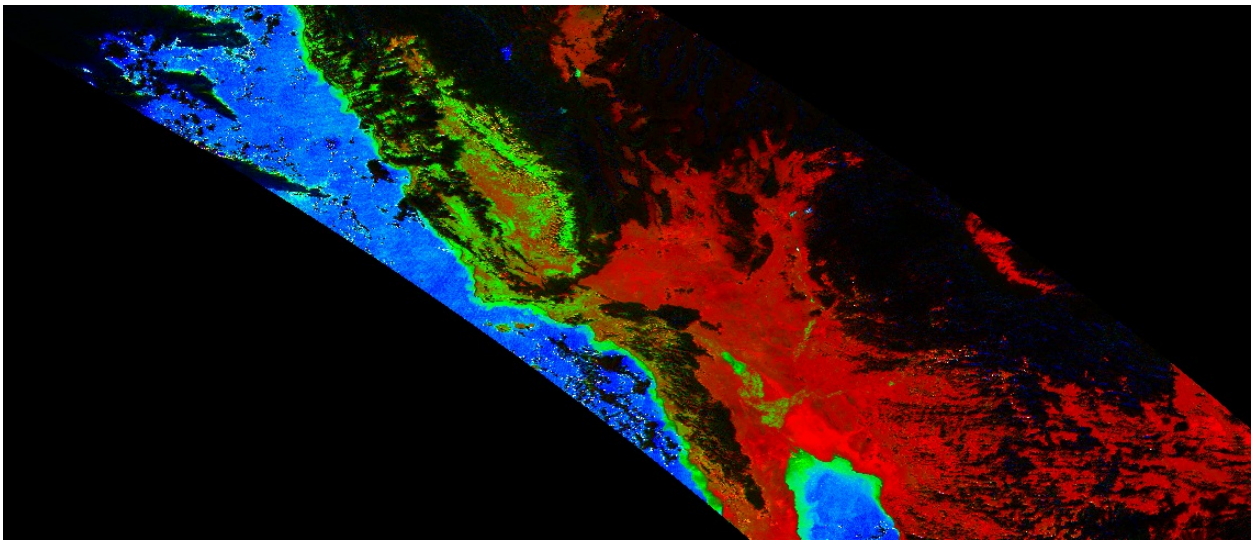


Figure 11. MODIS Aqua, January 23, 2010. Tile H08,V05, true color normalized deviations.
Few surfaces are as spectrally flat (i.e., grayscale) in the true color bands as clouds and snow. Stretching true color information by normalizing differences from average true color brightness highlights this. Dark areas indicate clouds and snow – this would already be a good mask without other spectral information.

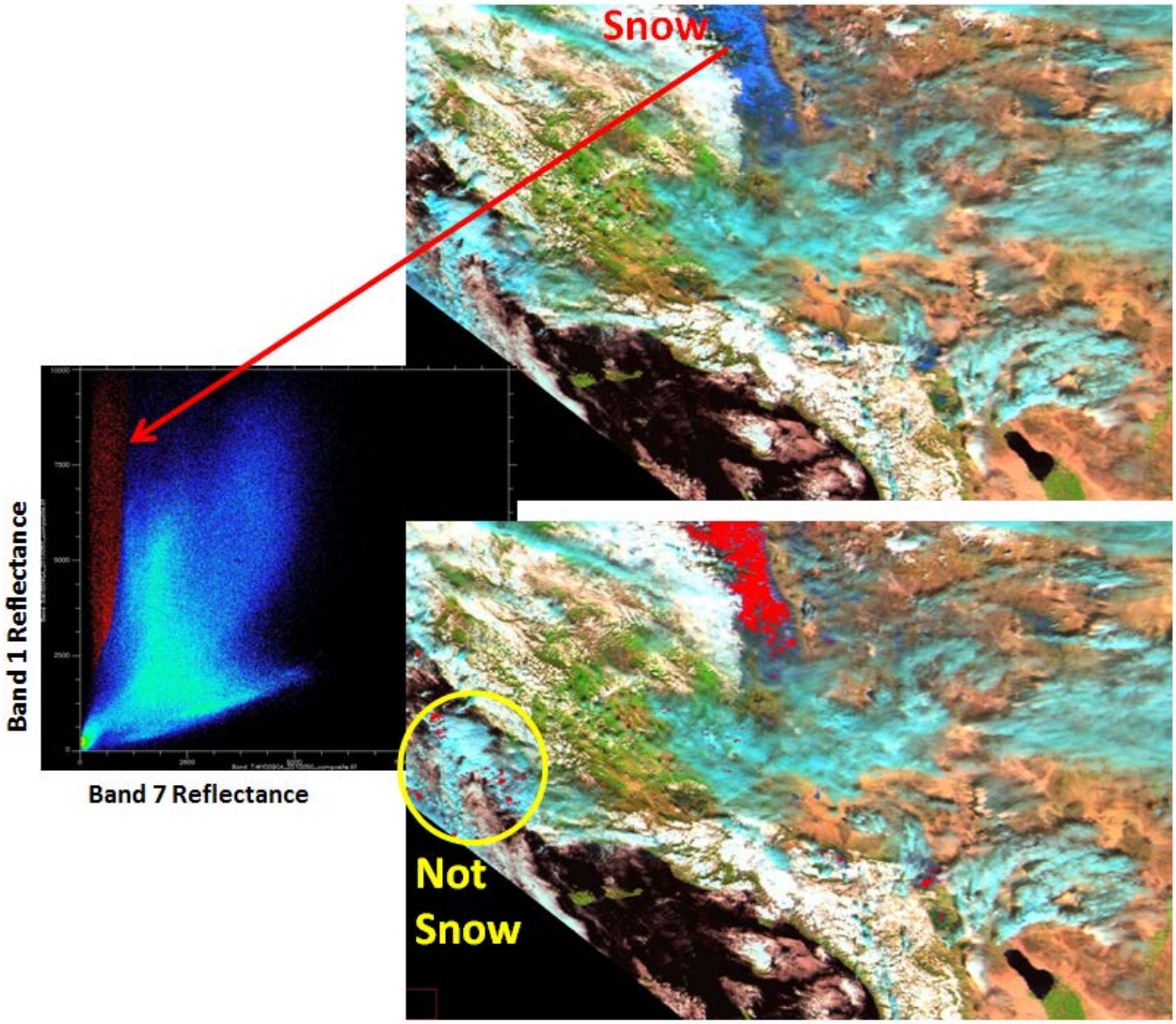


Figure 12. MODIS Aqua, southern California, February 19, 2010; snow vs. cloud in 7-5-1 display. Top right figure shows Sierra Nevada snow and a variety of clouds in a 7-5-1 RGB color display. Scatterplot on left shows image pixels in Band 1 and Band 7 spectral space, with ostensible "snow" values (medium to high band 1 reflectance, low band 7 reflectance) highlighted in red. Bottom right figure shows same 7-5-1 display with associated pixels now masked in red. This spectral space is mostly snow but includes some darker "ice clouds" (shown in yellow circle). Band 5 information can resolve some of this confusion.

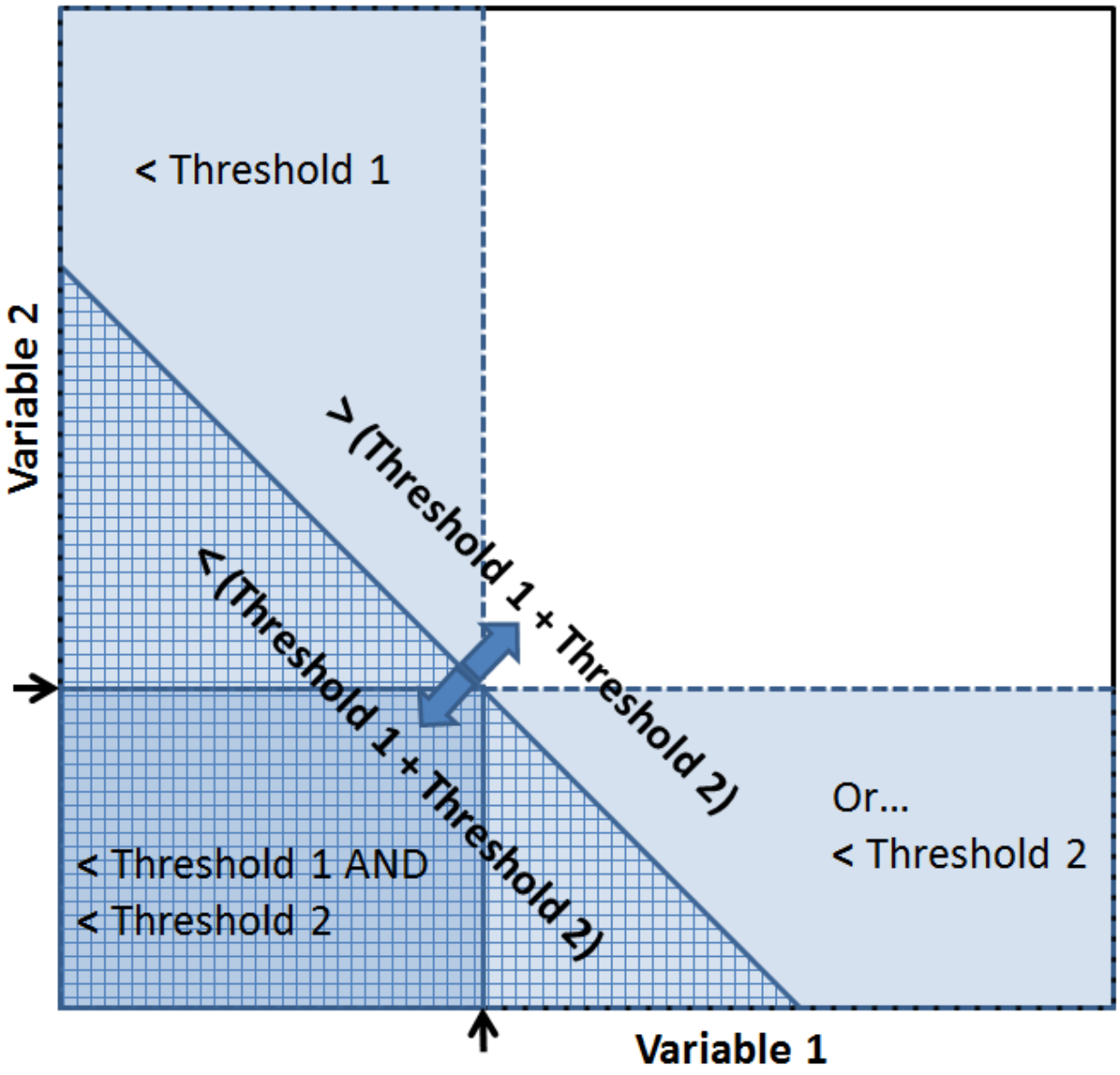


Figure 13. Conceptual diagram: variable thresholding.
Black arrows indicate some arbitrary threshold value in a given variable. Different Boolean treatment ("and" vs. "or") of various thresholds (common in cloud mask algorithms, and the typical output of classification trees) can result in dramatic classification differences, as indicated by different shading (light gray vs. dark gray). Merging, or adding, thresholds, allows models to consider non-orthogonal (diagonal) treatment of predictor variables (with resulting classification indicated by hatched pattern). In multiple dimensions, this would then be a multi-dimensional separating plane along a diagonal. This can improve classification accuracy, for example by allowing for values associated with lower probabilities in certain variables if complemented by values associated with higher probabilities in other variables.

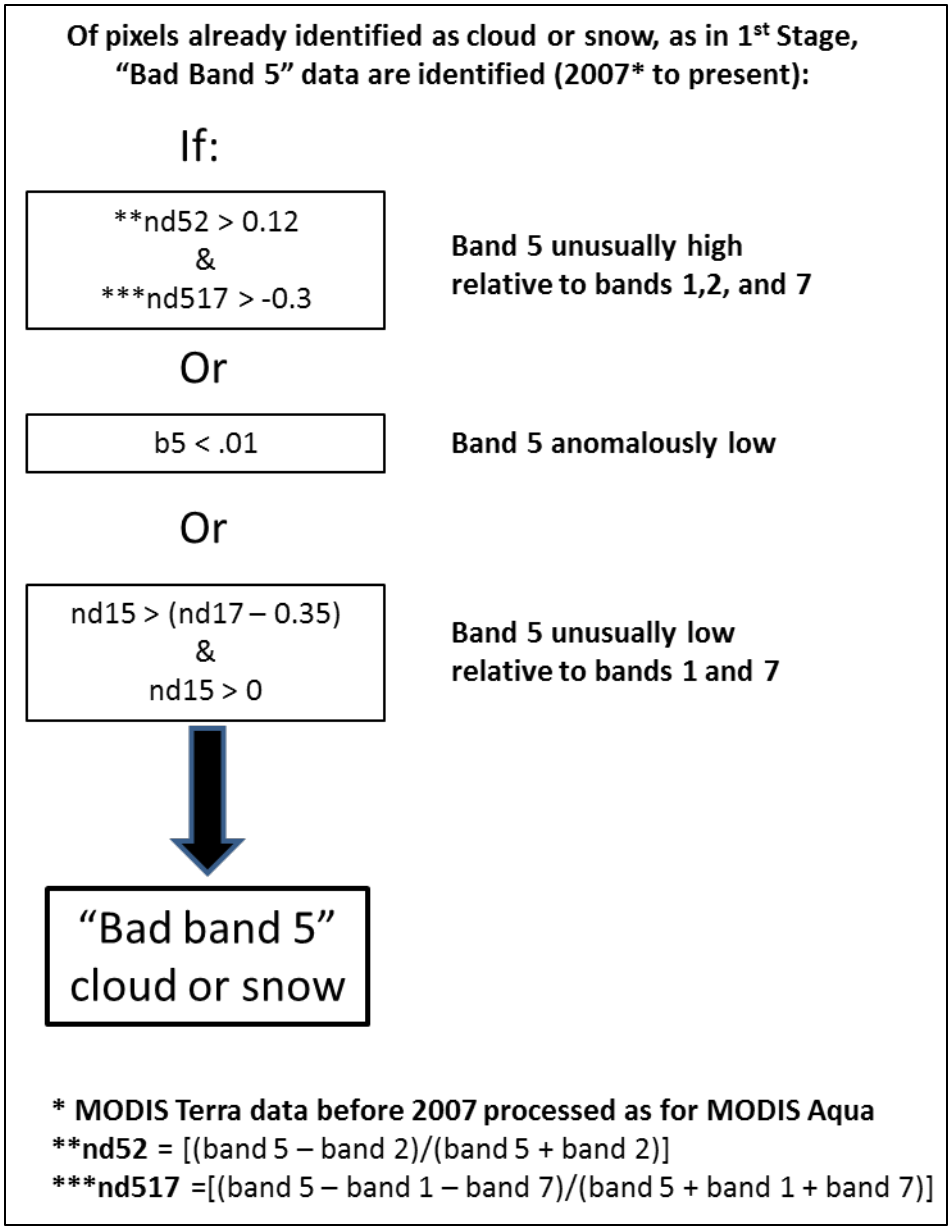


Figure 14. MODIS Terra identification of bad band 5 data.

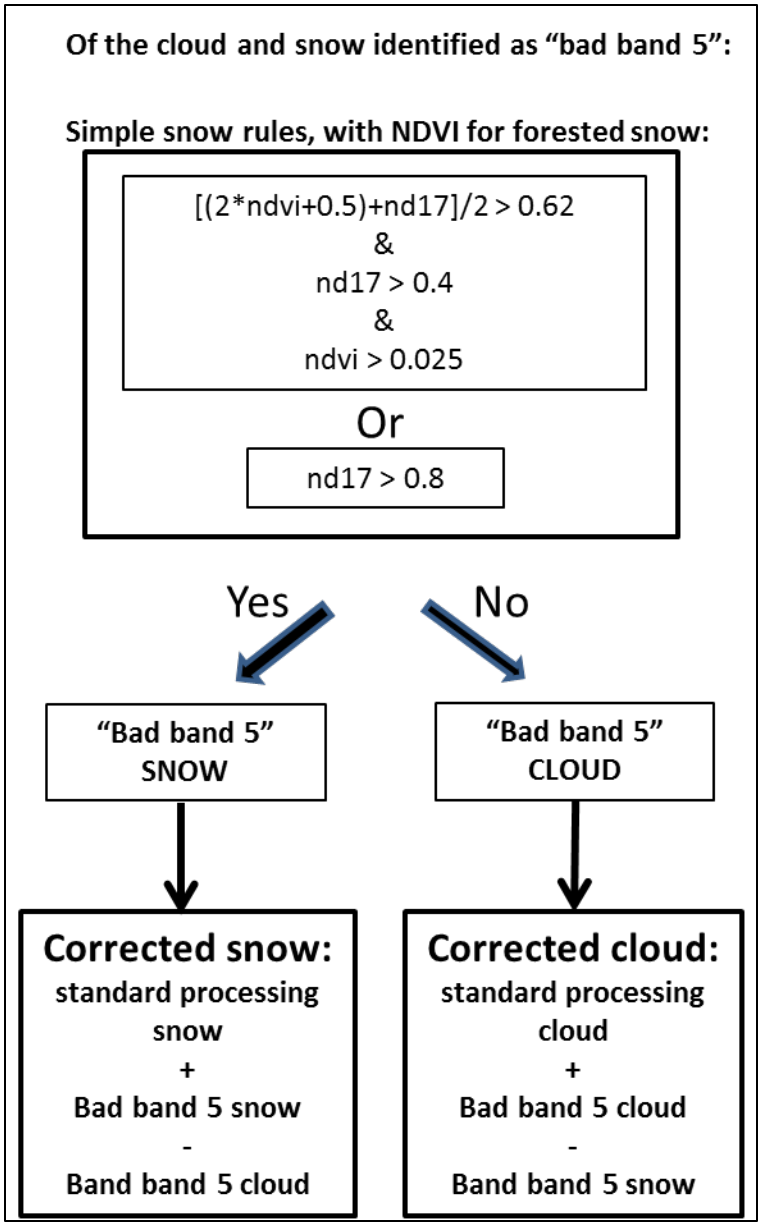


Figure 15. MODIS Terra processing of cloud and snow from bad band 5 data.

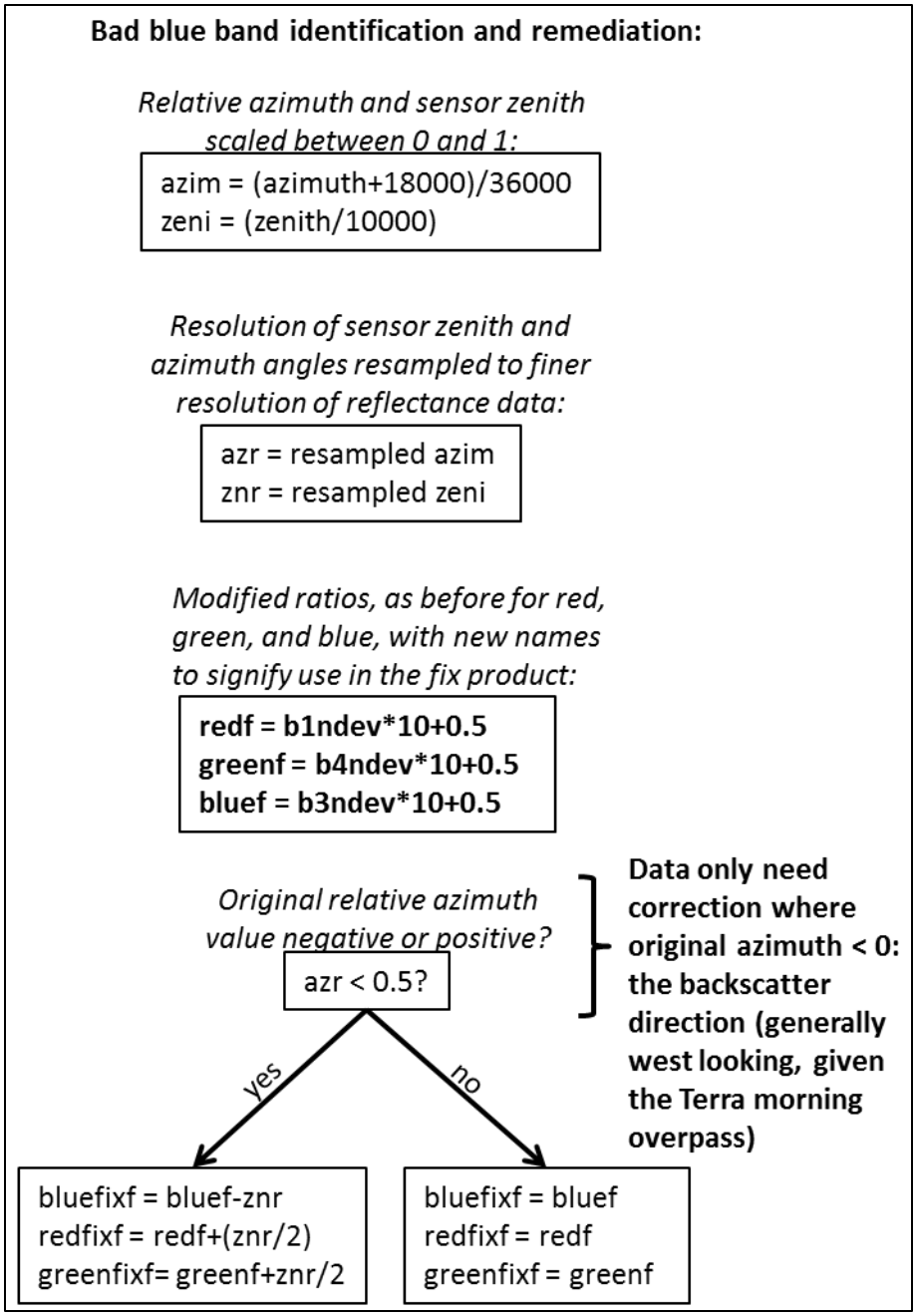


Figure 16. MODIS Terra identification of bad blue band (band 3) and basic correction.

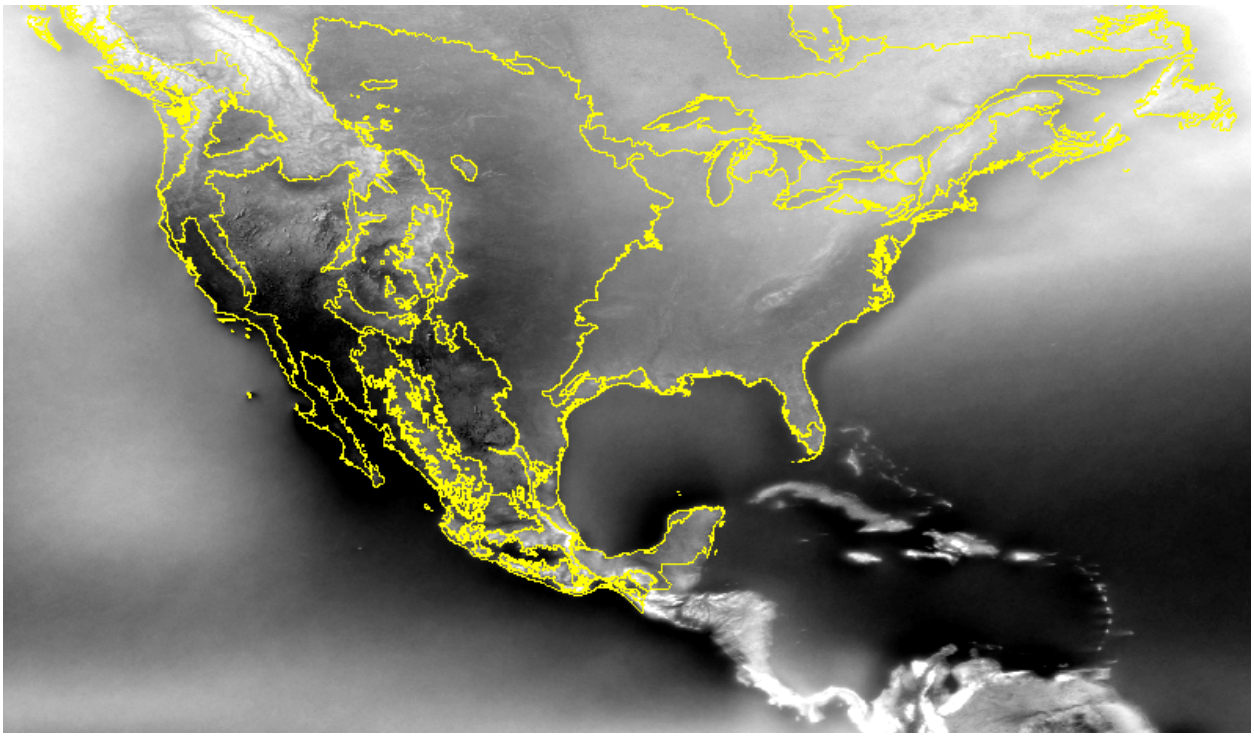


Figure 17. AVHRR custom cloud classification: annual cloud frequency, 1981-1999.
Compare with Figure 38 (Appendix 2.A).Correspondence with North American CEC ecoregions (yellow) is evident, but less high frequency variability related to contamination from underlying land cover.

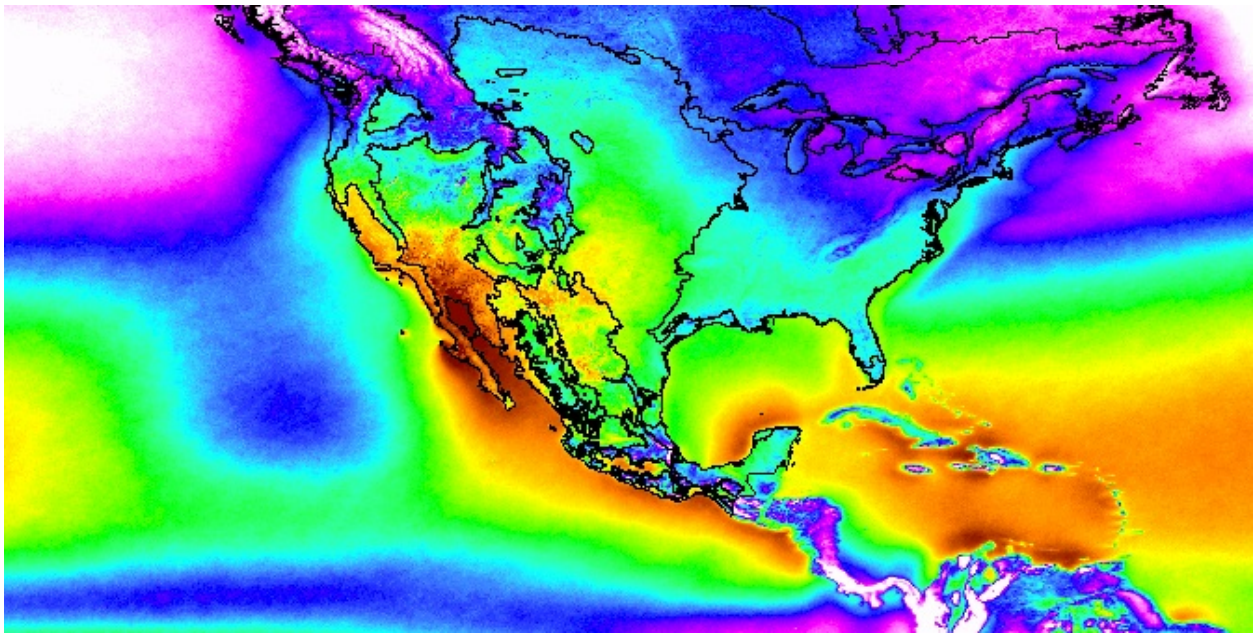


Figure 18. AVHRR custom cloud classification: annual cloud frequency, 1981-1999, color.
Color version helps to highlight some spatial patterns and spatial window expanded slightly over Pacific Ocean. North American CEC ecoregions in black. Otherwise, same as for Figure 12.

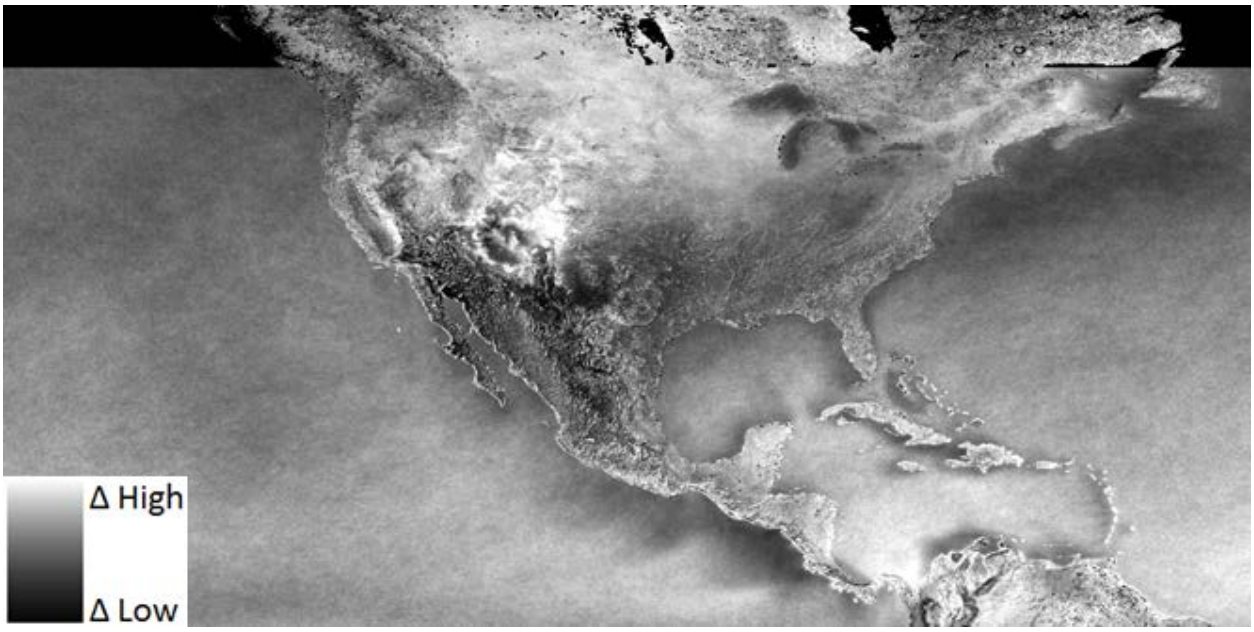


Figure 19. AVHRR LTDR cloud frequency minus custom cloud frequency, January 1982-1999.
Areas with much greater values in the LTDR product relate to a lower threshold for clouds (especially tropics) and LTDR confusion between snow and cloud (Sierra Nevada, southern Rockies, northern plains)

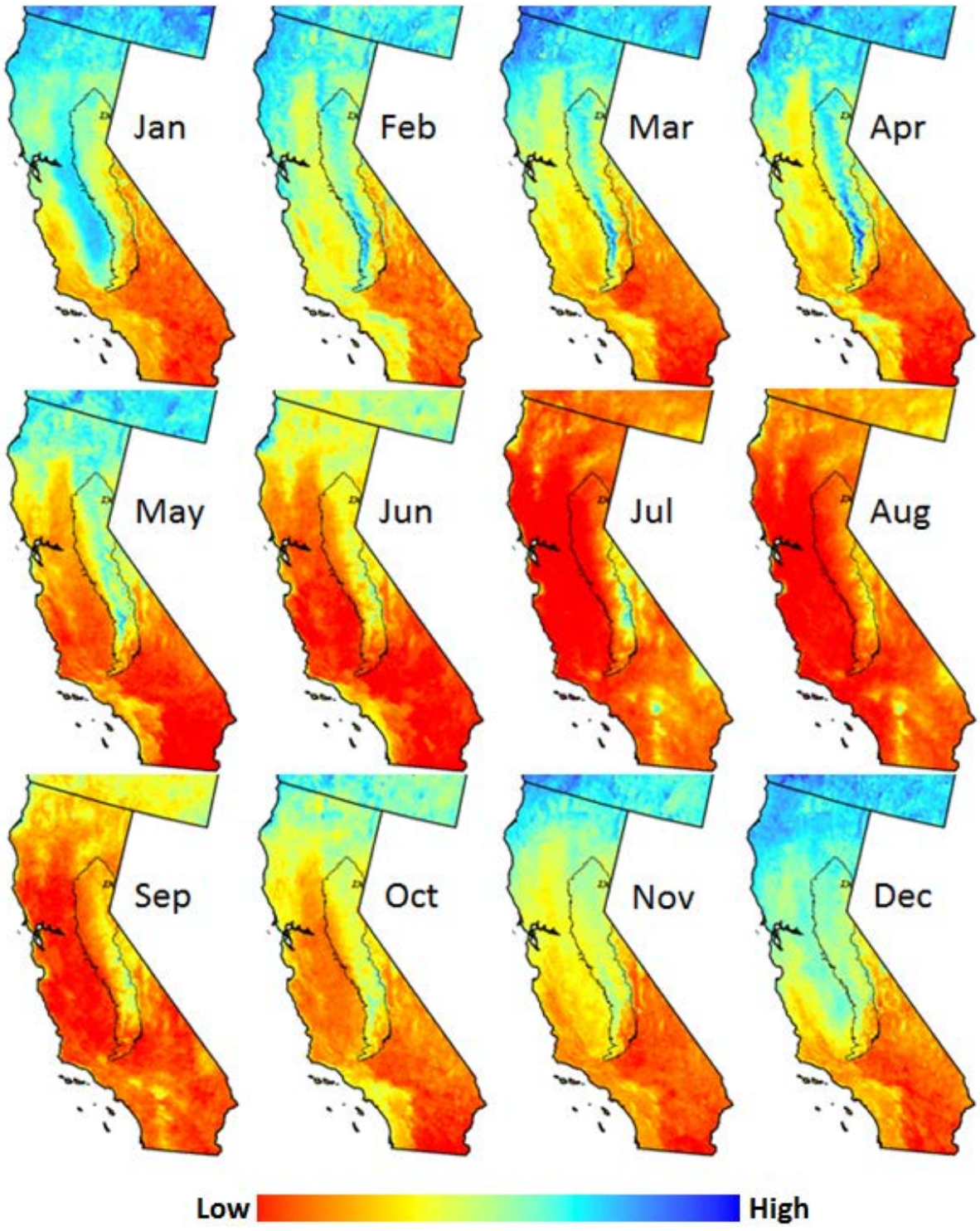


Figure 20. MODIS Aqua custom classification monthly cloud frequency, California.
The Sierra Nevada Jepson ecoregion (Hickman, 1993) has been included for reference. Each month has been individually color stretched to improve contrast.

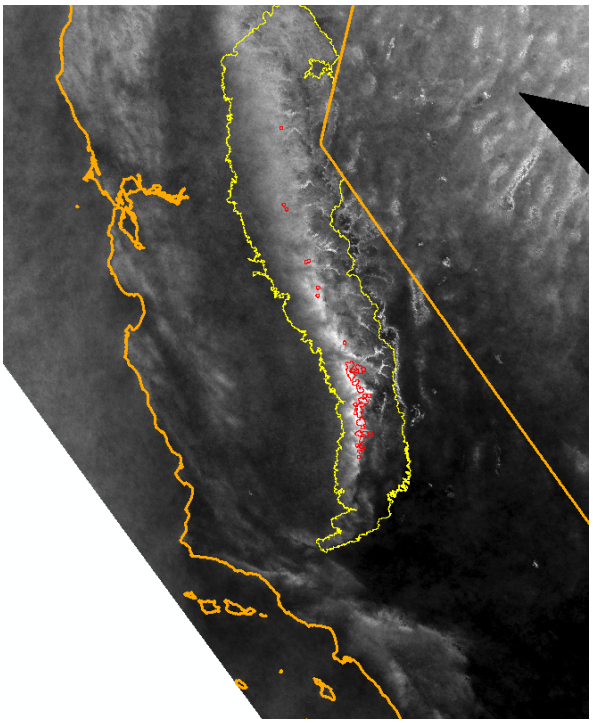


Figure 21. MODIS Aqua custom classification cloud frequency, April 2003-2011.
Giant sequoia distribution highlighted in red. Good correspondence with high April cloud frequency.

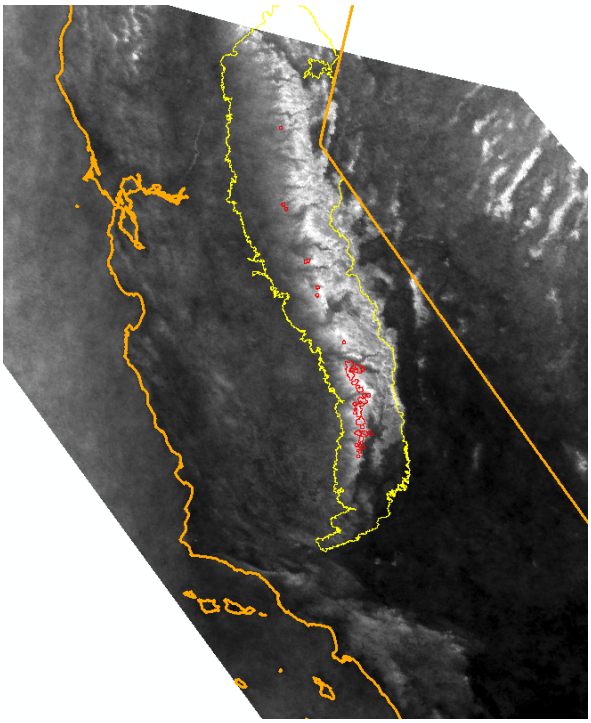


Figure 22. MODIS standard mask (cloud state variable) cloud frequency, April 2003-2011.
Giant sequoia region in red no longer stands out with particularly high April cloud frequency.

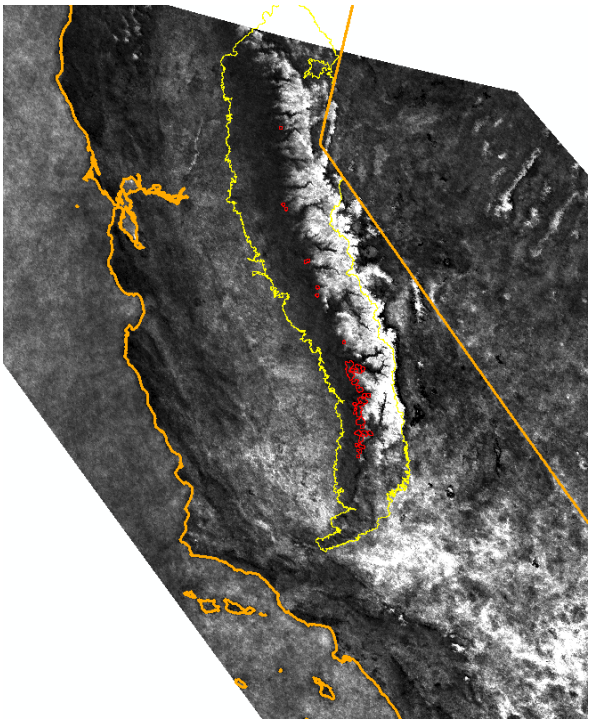


Figure 23. MODIS standard mask product divided by custom classification, April 2003-2011.
High values over the higher Sierra Nevada mountains relate to misclassification of snow as cloud.

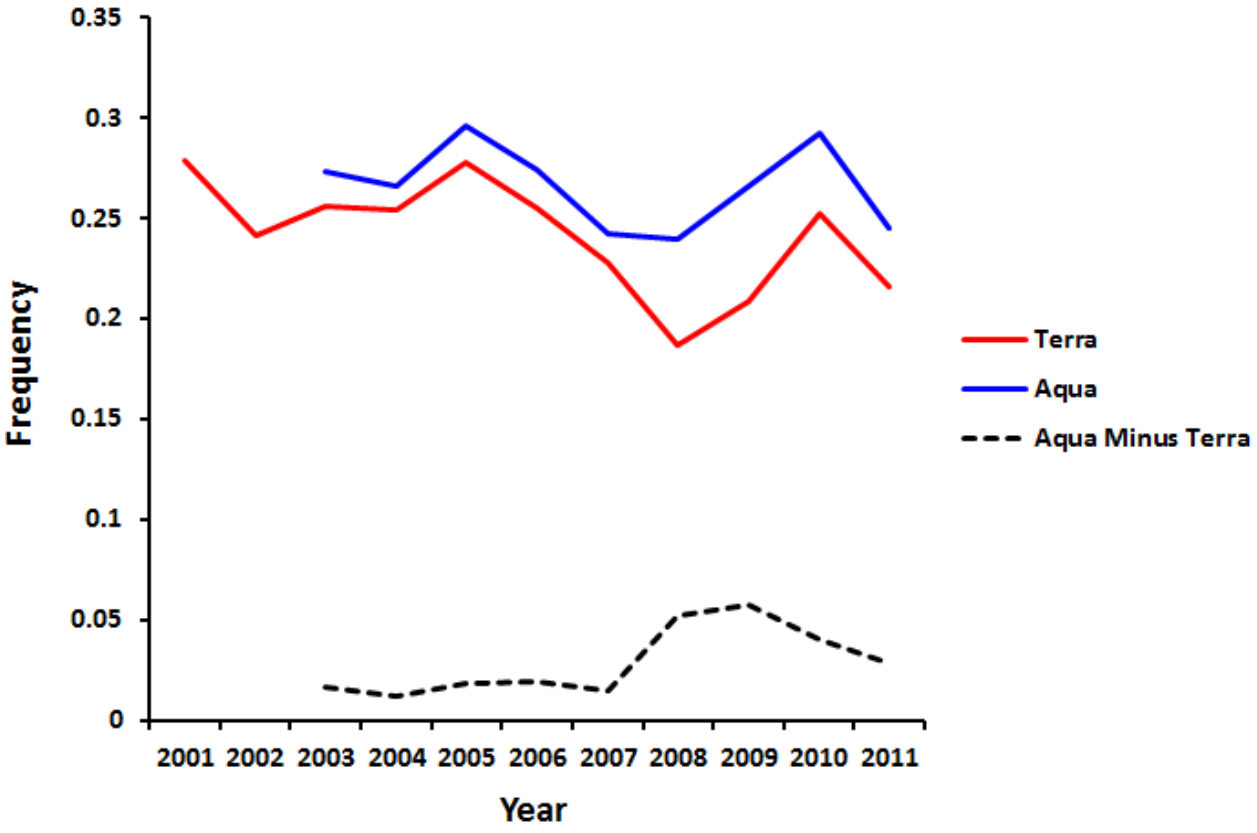


Figure 24. Custom MODIS Aqua and Terra annual cloud frequency and difference, tile H08V05

MODIS Aqua (afternoon) will naturally have slightly higher values than Terra (morning). The shift in 2008 to larger magnitude differences is associated with problems with the Terra data.

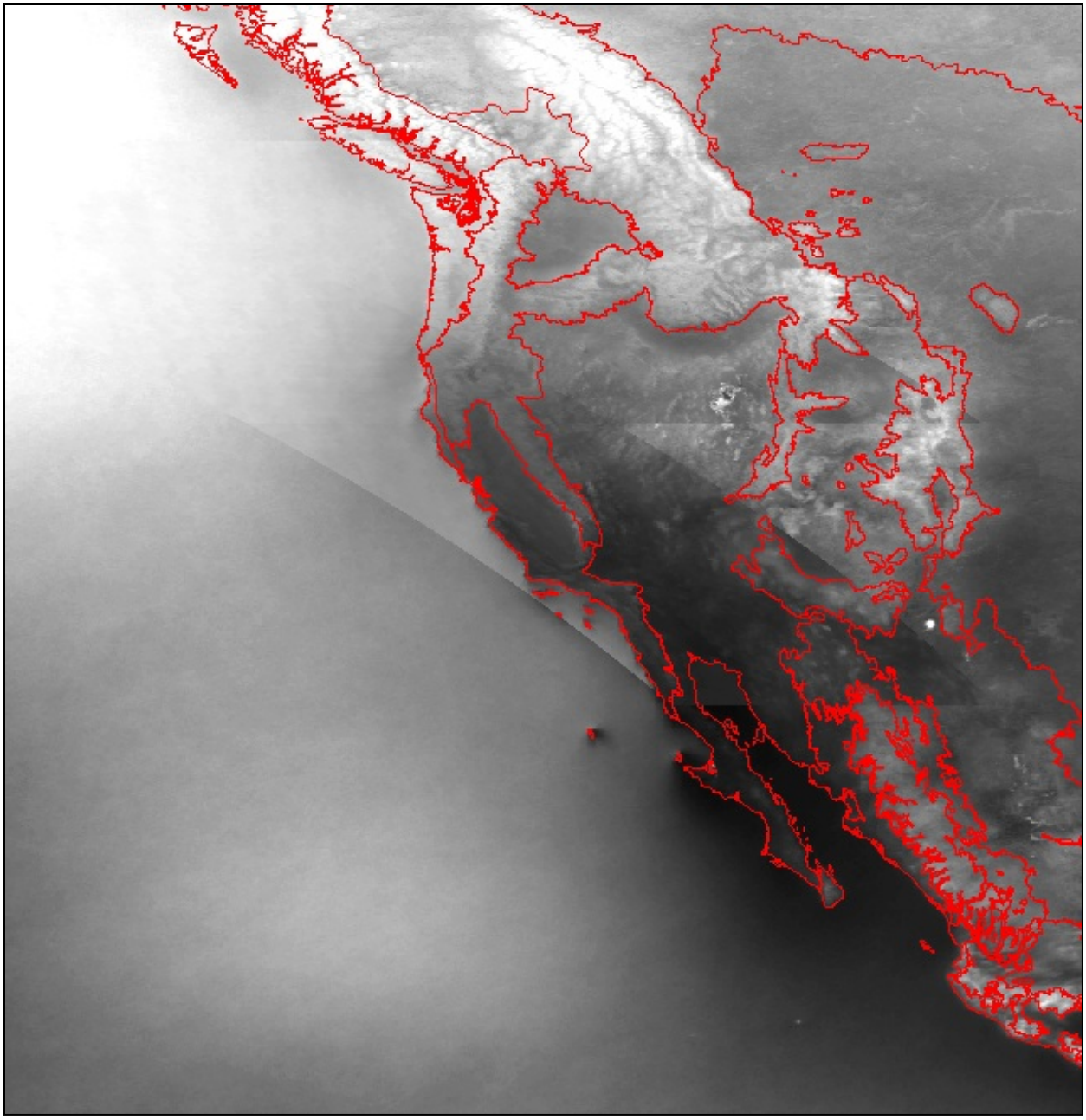


Figure 25. Overlay of MODIS Aqua cloud frequency on AVHRR cloud frequency (annual).
North America CEC Level I ecoregions in red. Mapped frequencies for each of the two products are comparable enough that the boundaries of the three sinusoidal tiles of MODIS (parallelograms in the geographic projection of AVHRR) are subtle.

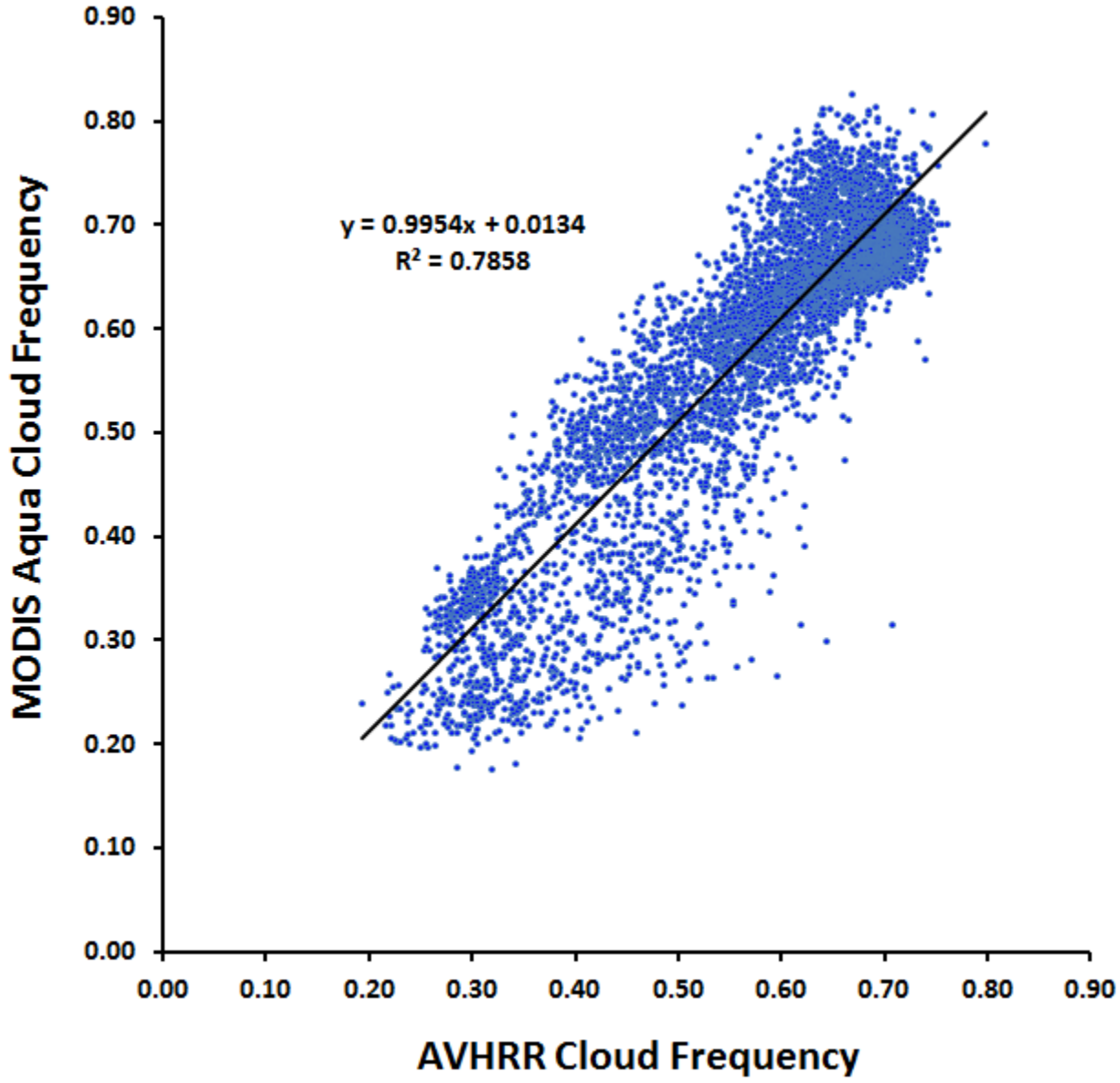


Figure 26. MODIS Aqua vs. AVHRR December cloud frequency comparison.
The linear relationship is quite close to 1:1, with an offset very close to zero, obviating the need for an accompanying 1:1 line for comparison. Deviations from the close 1:1 relationship mostly relate to AVHRR estimates over desert areas (either due to bright soils or snow).

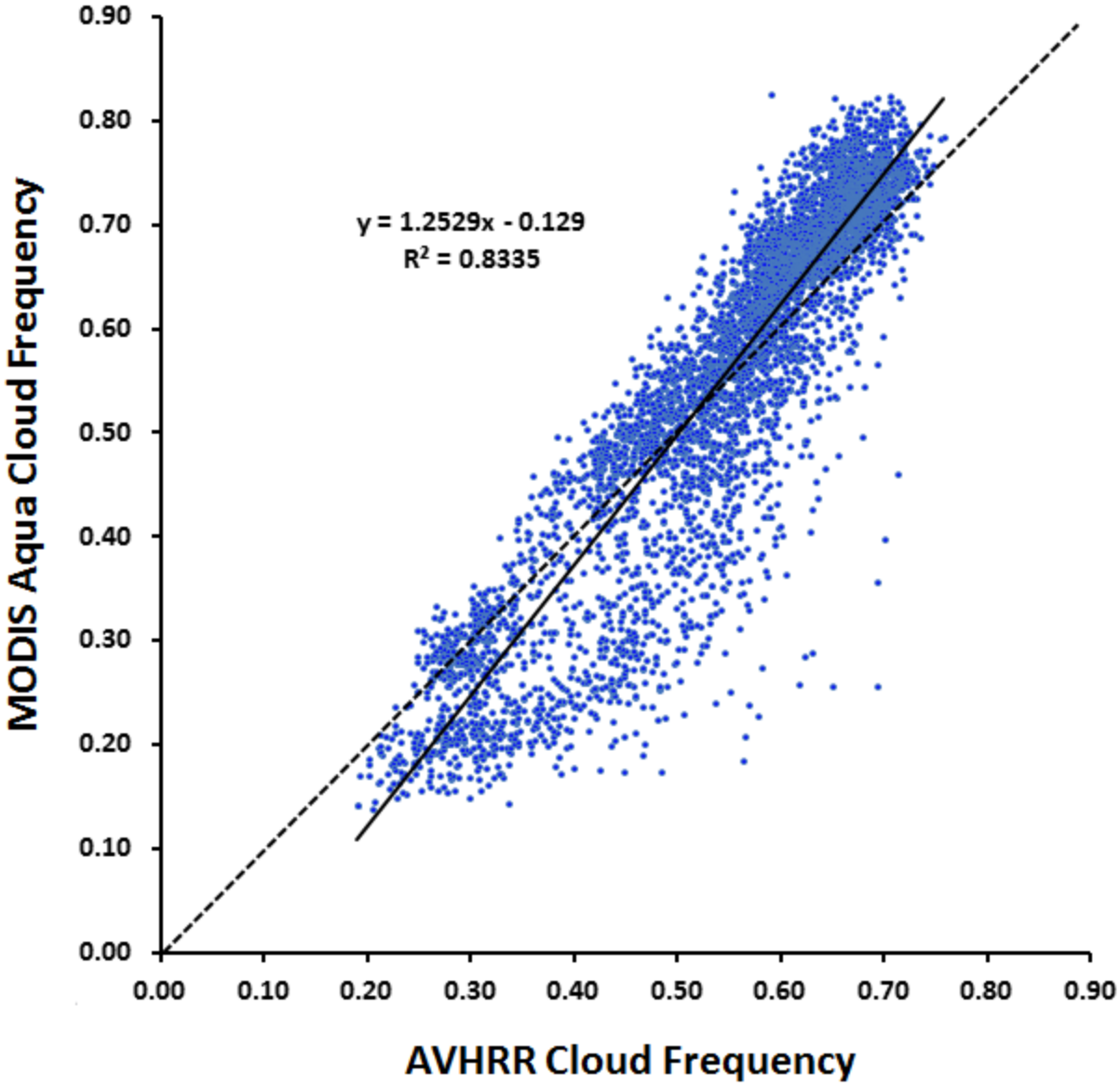


Figure 27. MODIS Aqua vs. AVHRR January cloud frequency comparison.
The relationship is slightly tighter than for December (indicated by slightly higher R-squared value of 0.83), but the slope falls off the 1:1 line (dashed), with an intercept far from zero.

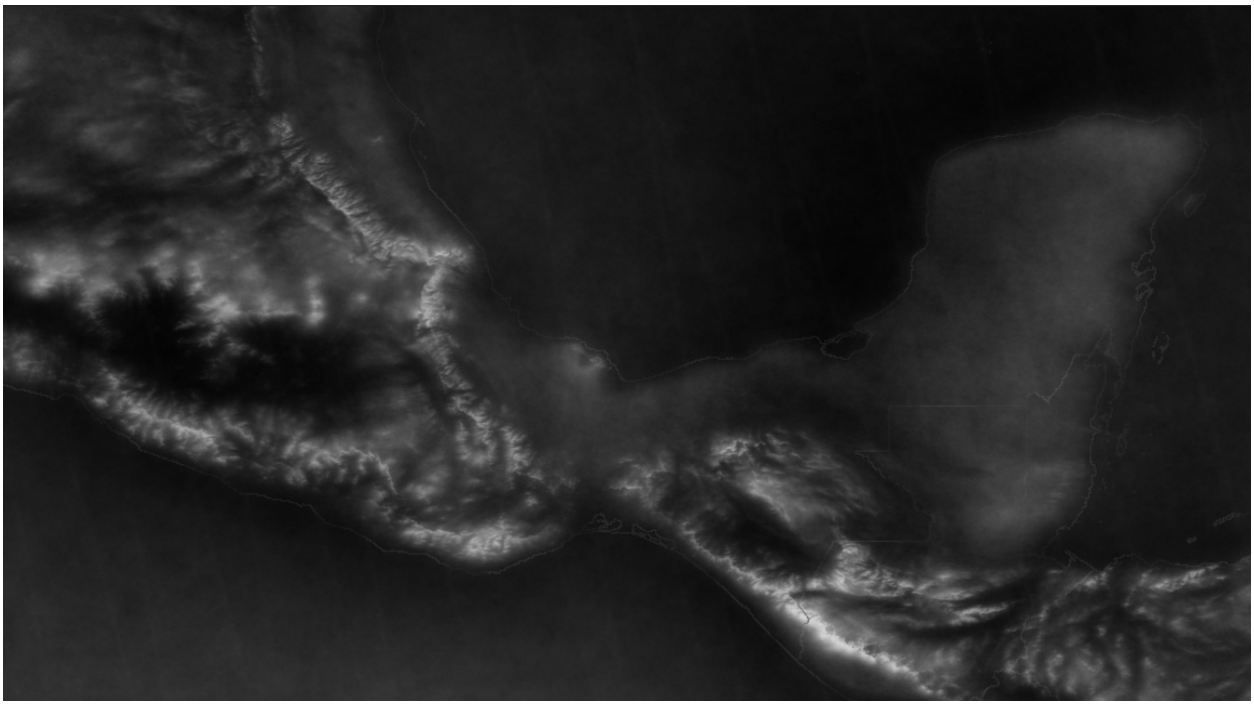


Figure 28. Southern Mexico cloud frequency, MODIS Aqua, May to October (2006-2012) – NOAA Cloud Climatology Project.
http://www.nssl.noaa.gov/projects/pacs/web/MODIS/

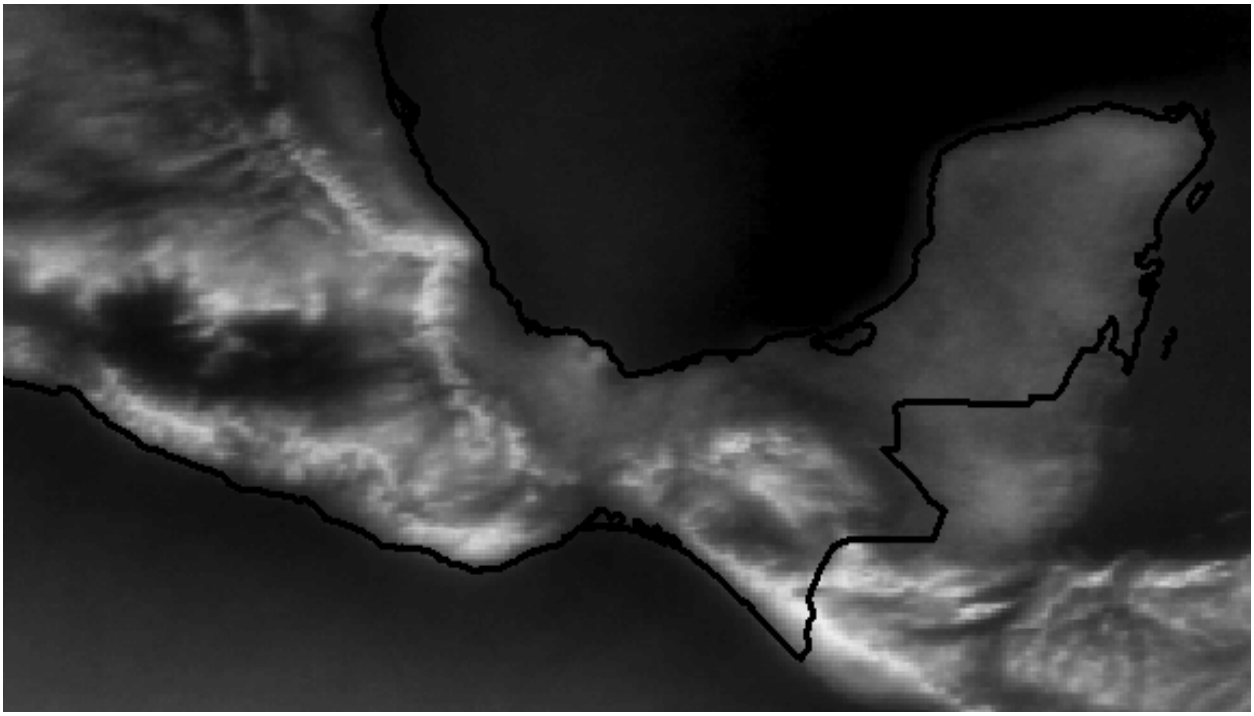


Figure 29. Southern Mexico cloud frequency, May to October (1981-1999), custom classification of AVHRR LTDR data.
Gross patterns very similar to those in Figure 28 despite older imagery at coarser resolution.

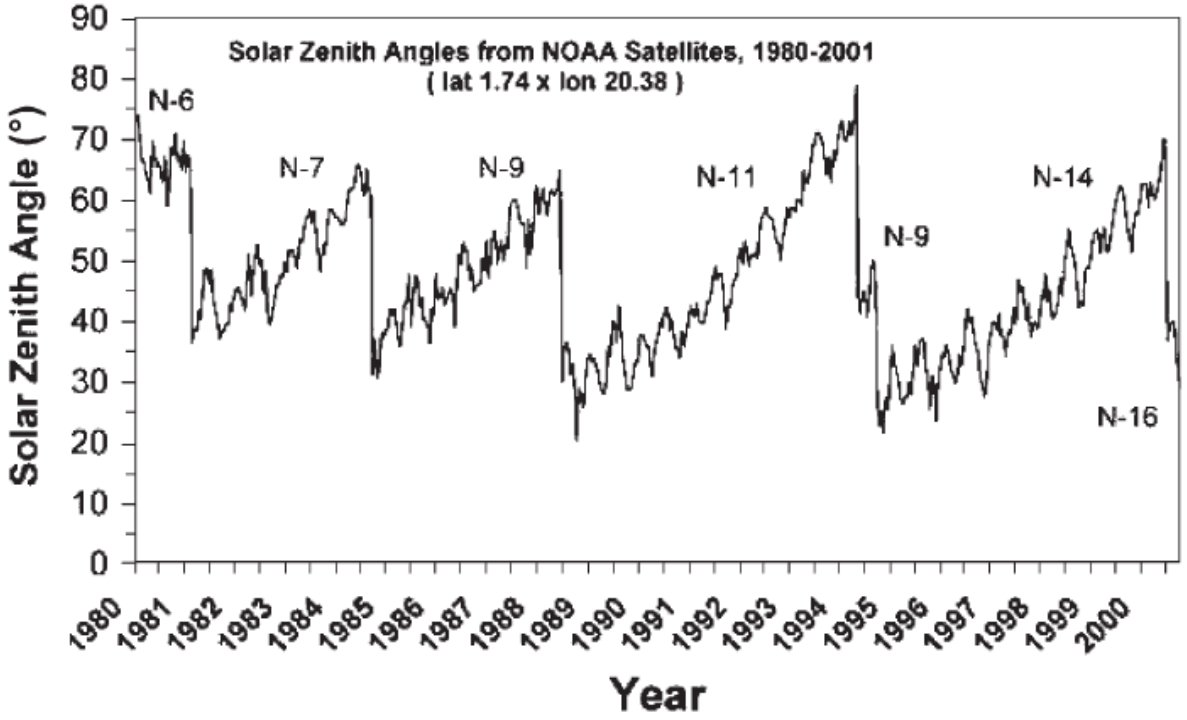


Figure 30. AVHRR solar zenith angles over LTDR period (Tucker *et al*, 2005).
Increasing trend in solar zenith angle over time after each satellite launch related to later overpass times.

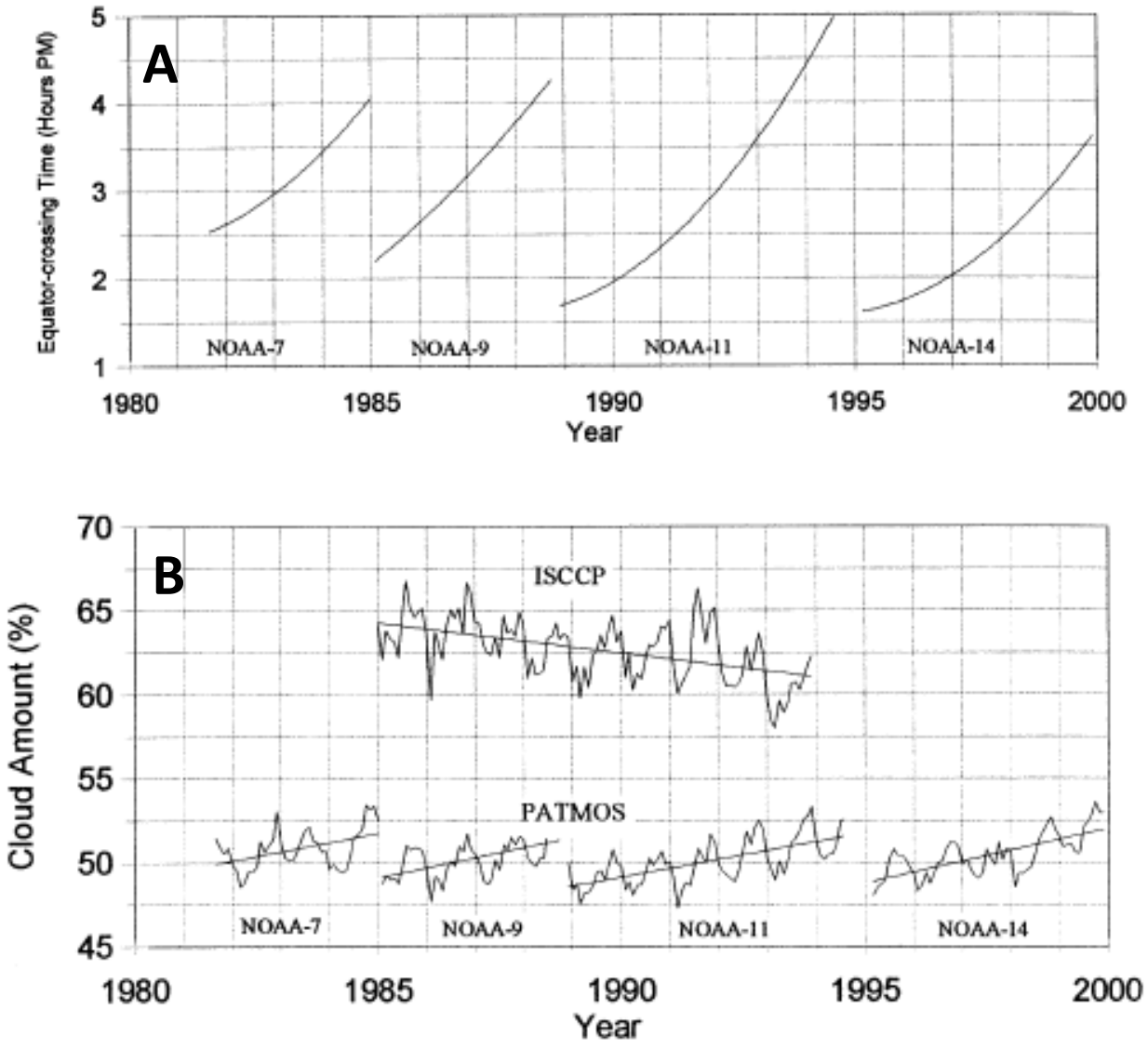


Figure 31. (A) AVHRR equator crossing time for various NOAA satellites; (B) Annual cloud classification percent from ISCCP and original PATMOS (taken from Stowe et al., 2002). Increasingly later overpass times and higher zenith angles (Figure 30) appear to bias cloud frequency, at least for PATMOS product. Decline seen in ISCCP data is questionable for other reasons (Figure 36).

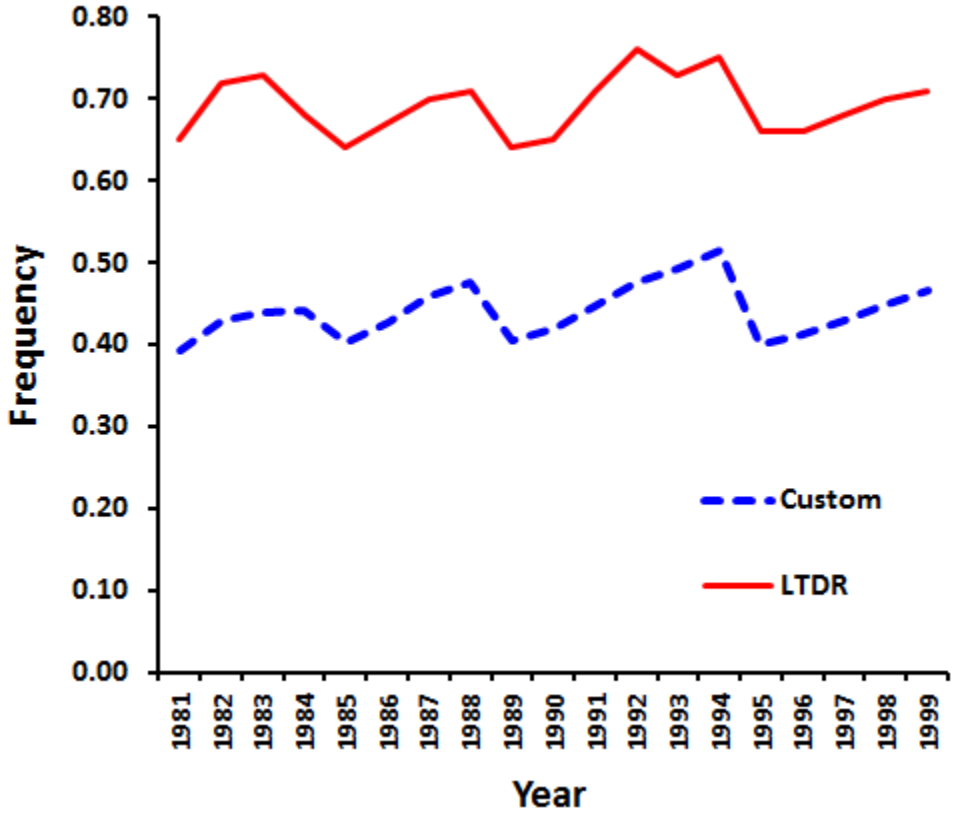


Figure 32. AVHRR annual custom cloud fraction vs. LTDR standard cloud fraction, averaged across ~10000x5000 km window.

Multi-year trends in cloud frequency for both classifications are disturbingly similar to trends in satellite overpass times (Figure 31) and associated solar zenith angle values (Figure 30).

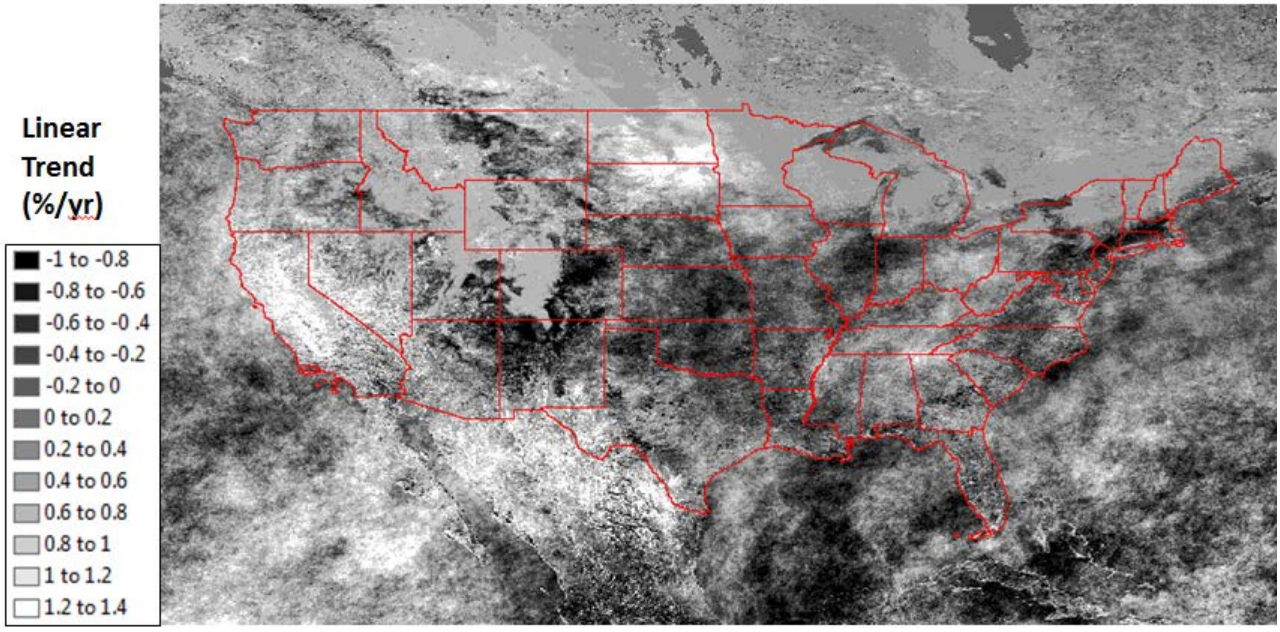


Figure 33. AVHRR LTDR standard February cloud frequency trend, percent per year, 1982-1999. Continuous gray values (trend ~0.6%/yr) throughout the Rocky Mountains and northern tier of U.S. states into Canada indicative of problem with increase in classifying winter snow as cloud cover. Increasing trend in northern Mexico may be related to a problem with desert soils.

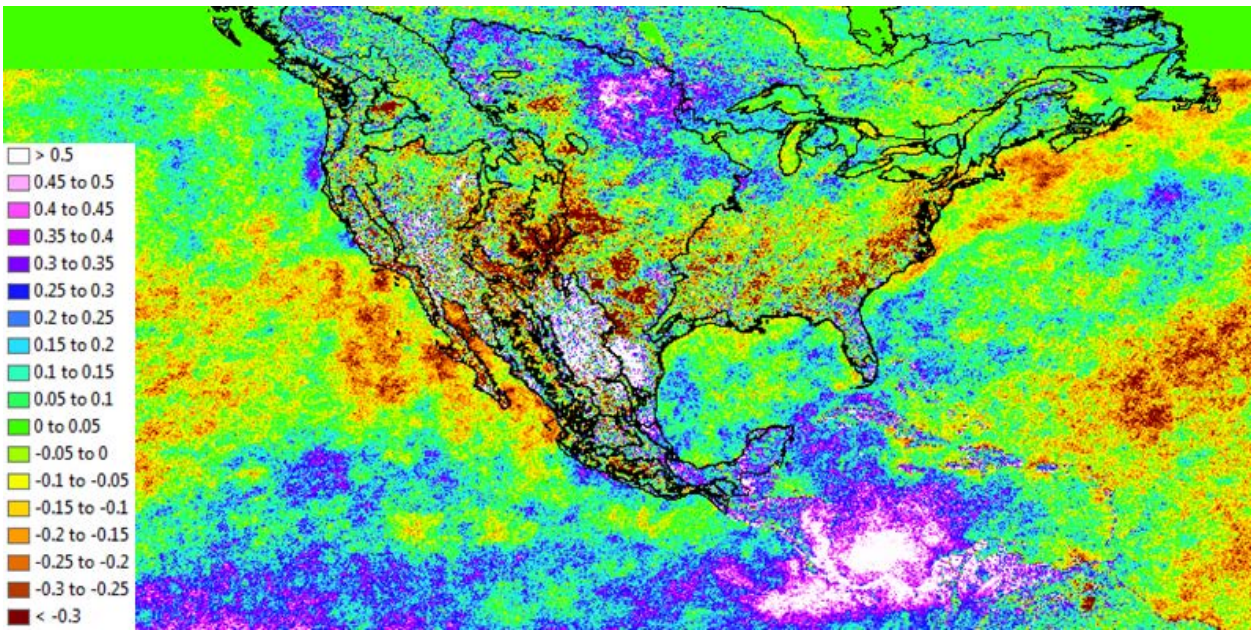


Figure 34. AVHRR LTDR standard annual cloud frequency trend, percent per year, 1981-1999. Snow problem less obvious in the annual product than in the February trend (Figure 33), but problem with deserts of northeastern Mexico and adjacent Texas apparent.

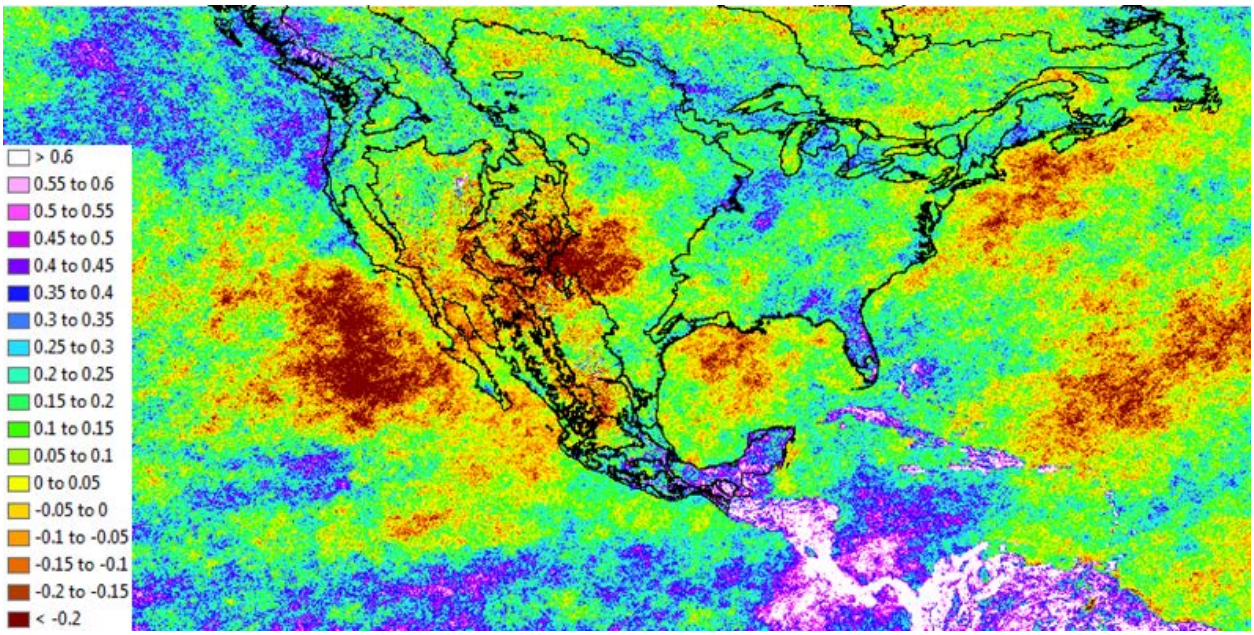


Figure 35. AVHRR custom annual cloud frequency trend, percent per year, 1981-1999.
Avoiding problem with deserts helps to highlight large area of decrease from eastern Pacific off southern and Baja California through the southwestern U.S. to the southern High Plains. Otherwise, there is good agreement with Figure 34, including increases in Central America, northern South America and adjacent oceans. Reasonable agreement with ISCCP product in Figure 36 as well (see text).

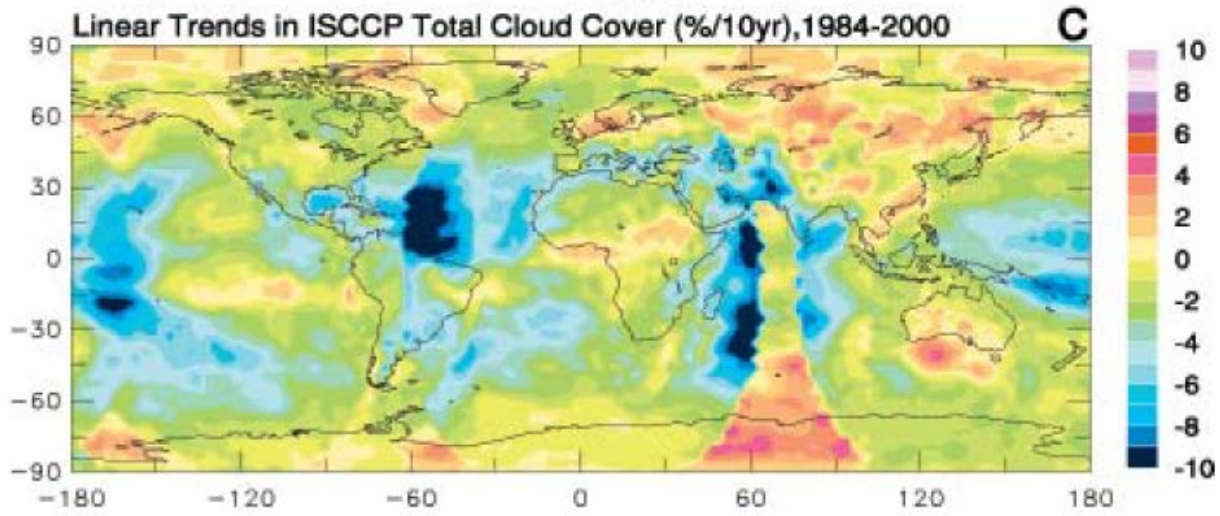


Figure 36. 1984-2000 trend in cloud cover (taken from Dai et al., 2006).
This product has noted problems related to satellite viewing geometry (Norris, 2000; Evan et al., 2007). This is most obvious over the Indian Ocean. Sun glint off ocean surfaces may also be a problem.

Appendices

Appendix 2.A. Problems with Cloud Masks

2.A.1. AVHRR Cloud Mask Problems

The new PATMOS-X classification (Heidinger *et al*, 2013) may be the most sophisticated publicly available classification of clouds from AVHRR imagery. These data are available on NOAA's Comprehensive Large Array-Data Stewardship System (CLASS) website (www.class.noaa.gov). While the product performs well in many areas, bright deserts of the southwestern U.S. and northern Mexico are errantly classified as cloud in its April, 1993 cloud frequency product, as seen in Figure 37. (This figure was generated using both their daily "probably cloud" and "cloud" classes. The results are not very different if restricted to only "cloud".) The main reason for this is likely that AVHRR does not have the three true color bands required for color information to separate most desert soils from clouds. Furthermore, some frozen lakes and bayshores of northern Minnesota, Manitoba, and Ontario appear to be classified as clouds in this product. It is likely that some snow covered areas (e.g., high Sierra Nevada mountains) also tend to get classified as cloud in this product.

The standard LTDR cloud mask (second bit, i.e., bit number 1) that accompanies the AVHRR LTDR data generally avoids calling many of these same desert areas cloud. This is somewhat surprising, given the lack of sufficient true color information with AVHRR. It appears to be the result of customized processing specific to these desert areas. In fact, a halo effect is often seen in derived cloud frequency products on some edges of these desert areas – for example on the edge of the playa of the Black Rock Desert in northwestern Nevada (Figure 38). The most likely explanation is that the land cover product that was used to stratify processing of the AVHRR data was slightly offset from the raw data that were being classified. It is also possible that the LTDR cloud mask is able to handle these spectrally challenging desert areas with a spatial uniformity rule across several pixels (as employed by CLAVR-1). In particular, the use of a spatial filter might explain the halo effects. It is more difficult to imagine how this might explain the many other land cover patterns that appear in the LTDR cloud frequency product. These could either be the result of stratification and land-cover specific processing in other areas, or simply a very strong bias in cloud frequency due to confusion with the underlying land cover. The pronounced high cloud frequency along coastlines is especially notable. Snowy areas are also overly likely to be classified as cloudy.

2.A.2. MODIS (Terra and Aqua) Cloud Mask Problems

One problem with MODIS cloud masks is apparent before even looking at the data: it is not readily apparent which product is the most appropriate cloud mask. This may be partially explained by the fact that different users will have different definitions of clouds (e.g., those interested in land cover vs. atmospheric scientists). Even different MODIS processing teams use different cloud products. For example, the "cloud state" parameter found in the MOD09 product's "state flag" at 1 km resolution (bits 0 and 1) is derived from the MOD35 cloud product. The parameter differs substantially in spatial pattern from the "internal cloud algorithm

flag" in the same product's bit 10 (eleventh bit). The MOD35 bit 0-1 combination in the first two bits – that is, a 1 in the first ("number 0"), rightmost, bit and a zero in the second ("number 1") bit (i.e., bits should be read right to left) is called "cloudy" and looks like the best cloud mask, when comparing against raw reflectance data for a variety of land cover types. The supposedly "mixed" (i.e., mixed pixels containing some cloud) type (bit combination 1-0) also derived from MOD35 does not appear quite as good; in the southwestern U.S., it simply captures many bright desert areas. Similarly, the cloud shadow parameter (bit "number 2", but actually the third bit) of the MOD09 state flag product does not appear accurate – it looks like a large buffer around the periphery of cloudy and even snow covered areas.

Only the "cloudy" bit combination appears reasonably accurate in single scenes. However, even in this product, systematic errors are revealed when processed across long time periods. In Figure 39, showing annual cloud frequency derived from the standard cloud masks for MODIS Aqua, many desert areas have unusually high cloud cover – likely due to their bright reflectance. It is difficult to tell from this figure alone, but many frequently snow-covered areas also have spuriously high cloud frequency values. And other odd patterns turn up as well – an abrupt discontinuity in frequency is evident between the deserts and highlands of Arizona and northern Mexico. This is obviously the result of different cloud classification algorithms for water, coast, desert, and "land", according to the MOD35 "processing path" (Wilson *et al*, 2014). The fact that lower-elevation vigorous agriculture in Arizona also shows up as anomalously high cloud frequency compared to surrounding desert suggests that it was either a vegetation-index related mask that was used for the different processing or the algorithm over "desert" has a problem with vigorous vegetation. Figure 40 zooms in on the annual cloud frequency product from MODIS Terra over southwestern California and northern Baja California. Urban areas of Los Angeles and San Diego have unusually high cloud cover, particularly in areas of dense build-up. Some bright desert areas can still be seen on the east. Finally, an obvious buffer (likely the "coast" processing path of MOD35) along the coast and around the offshore Channel Islands is evident as areas of slightly lower cloud cover. The processing path for "coast" probably adjusts the cloud classification due to confusion from bright beach sand and surf. But it also results in an underestimate of cloud frequency in these areas.

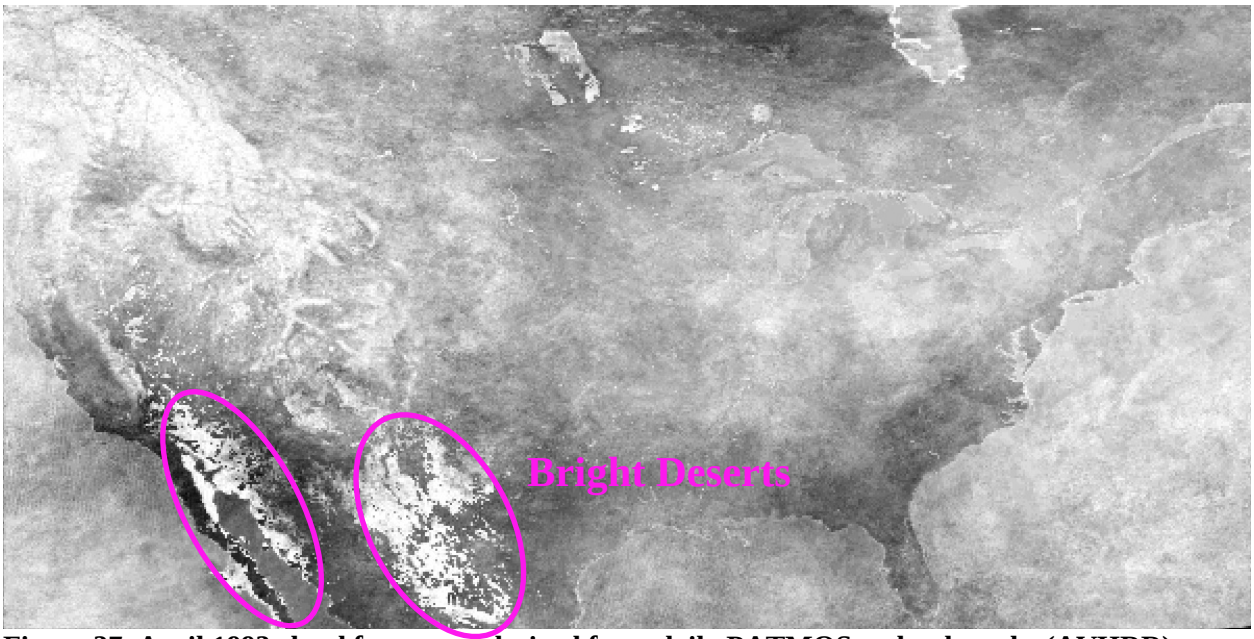


Figure 37. April 1993 cloud frequency derived from daily PATMOS-x cloud masks (AVHRR). Cloud mask algorithm has a problem with classifying bright deserts of the southwestern U.S. and northern Mexico (magenta ellipses) as cloud. This results in anomalously high cloud frequency for these areas.

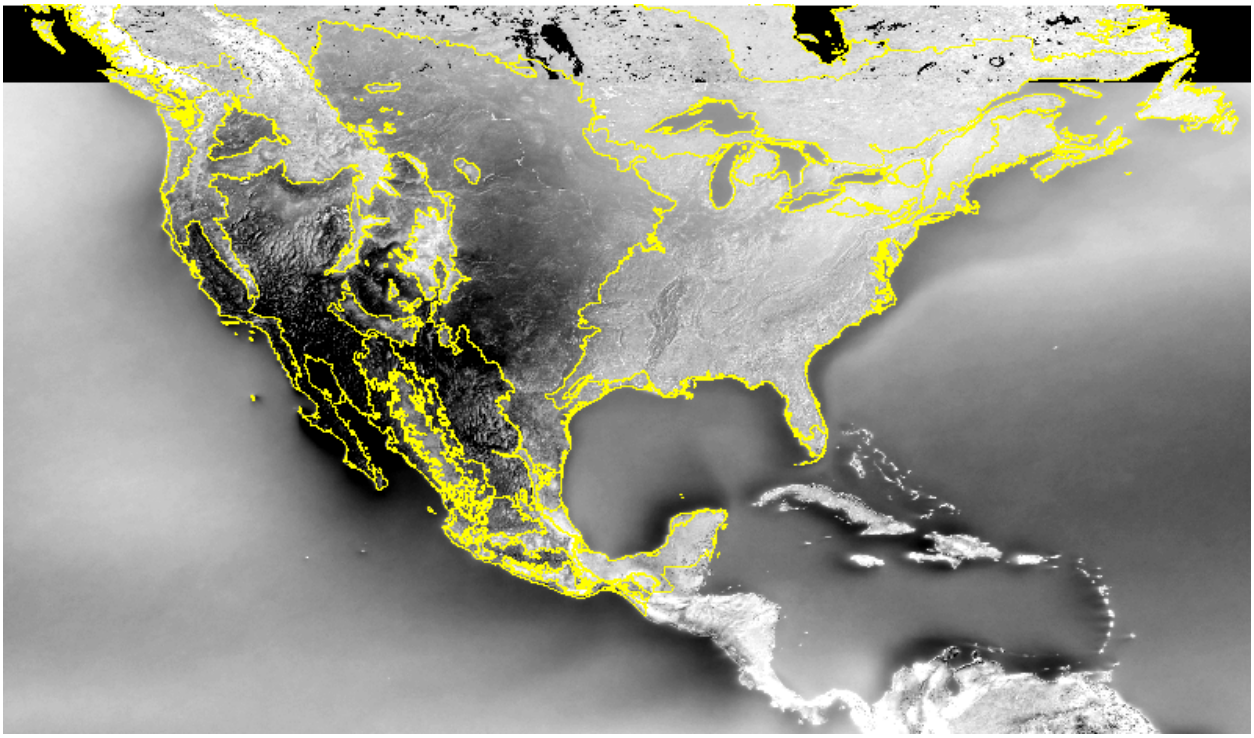


Figure 38. Annual cloud frequency derived from daily AVHRR LTDR standard cloud masks. Overlay of North American Commission for Environmental Cooperation Level I ecoregions in yellow shows the correspondence between cloud frequency and broad ecological patterns. Some high frequency variability related to land cover is evident, particularly along the lower Mississippi River Valley.

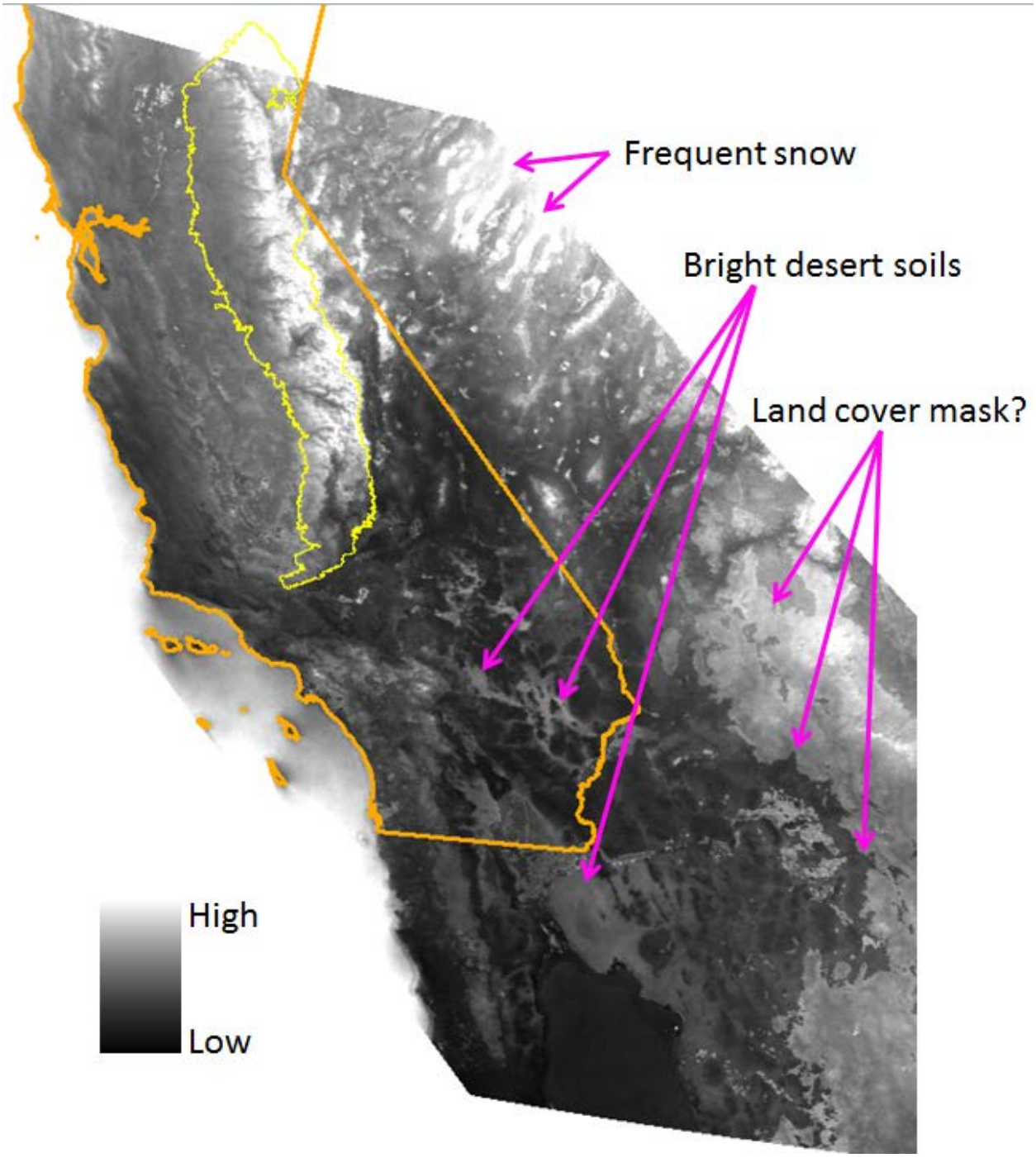


Figure 39. MODIS MYD09 standard cloud mask, annual frequency (2002-2012), tile H08, V05.
Bright desert soils are mapped with relatively high cloud frequency due to error in the individual daily cloud masks. Abrupt boundaries associated with land cover-specific processing are evident in eastern areas (Arizona and northern Mexico). Areas of bare rock and frequent snow are also mapped with excess cloud frequency, although this may be less obvious to the untrained eye.

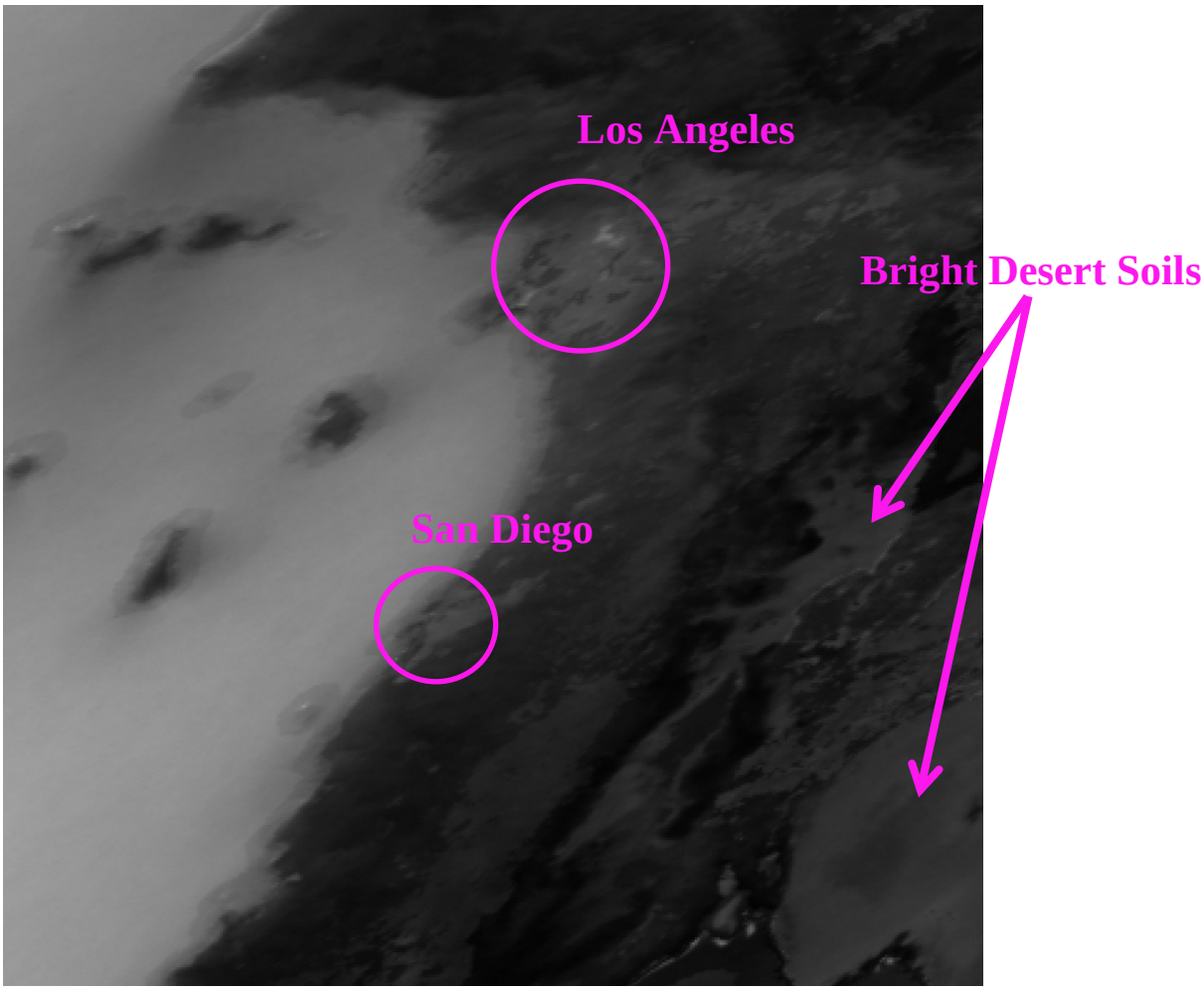


Figure 40. MODIS MOD09 standard cloud mask, annual cloud frequency (2000-2012), southern California (sinusoidal projection).

Same as previous figure, zooming in on southern California coastal and desert areas. Urban patterns are evident with excess cloud frequency, and buffers along the coast and around the Channel Islands (dark patches on left in the otherwise high cloud frequency Pacific Ocean) have anomalously low cloud cover compared to adjacent ocean.

Appendix 2.B. MATLAB code

2.B.1. AVHRR code

Code is provided below for deriving daily standard and custom cloud masks from AVHRR AVH09C1 daily reflectance HDF data and looping through multiple years in MATLAB. First are some examples for reading in some images from a single day into MATLAB. January 4th, 1984 data are used in the example. The year and calendar day are indicated following the "AVH09C1.A" in the import (hdfread) commands below. The NOAA-7 satellite is indicated by the "N07". The last set of numbers ("2010…") relates to a date and time of processing. For additional information, see:
http://ltdr.nascom.nasa.gov/ltdr/docs/workshop/PEDELTY_LTDR_data_formats.pdf

Each MATLAB line of code ends in a semi-colon (;).
Comments in the code below are either in bold text or, using the MATLAB convention for non-code comments, in green and following the pound (%) symbol.

[MATLAB's HDF Import Tool helped facilitate determining the commands for reading raster grids of data (or "Scientific Data Sets", SDS) from HDF files. The tool's importing is a relatively automated "point and click" process, with these next lines of code provided as output by the tool.]

Example code for reading in images within the "EOS" subset of the file data:

```
% Surface reflectance for channel 1 (red)
SREFL_CH1 = hdfread('C:\AVHRR_HDF\AVH09C1.A1984004.N07.003.2010062183040.hdf', 'Grid', 'Fields',
'SREFL_CH1');
% Surface reflectance for channel 2 (near infrared)
SREFL_CH2 = hdfread('C:\AVHRR_HDF\AVH09C1.A1984004.N07.003.2010062183040.hdf', 'Grid', 'Fields',
'SREFL_CH2');
% Quality Assessment field
QA = hdfread('C:\AVHRR_HDF\AVH09C1.A1984004.N07.003.2010062183040.hdf', 'Grid', 'Fields', 'QA');
```

Examples within the "HDF" subset of the file data (note the slightly different import commands):

```
% Brightness temperature, channel 3 [includes substantial reflectance component]
BT_CH3 = hdfread('C:\AVHRR_HDF\AVH09C1.A1984004.N07.003.2010062183040.hdf', '/BT_CH3',
'Index', {[1  1],[1  1],[3600  7200]});
% Brightness temperature, channel 4 [negligible reflectance component]
BT_CH4 = hdfread('C:\AVHRR_HDF\AVH09C1.A1984004.N07.003.2010062183040.hdf', '/BT_CH4',
'Index', {[1  1],[1  1],[3600  7200]});
% Brightness temperature, channel 5 [negligible reflectance component]
```

```
BT_CH5 = hdfread('C:\AVHRR_HDF\AVH09C1.A1984004.N07.003.2010062183040.hdf', '/BT_CH5',
'Index', {[1  1],[1  1],[3600  7200]});
% LTDR's "reflectance" component of channel 3 (mid infrared)
SREFL_CH3 = hdfread('C:\AVHRR_HDF\AVH09C1.A1984004.N07.003.2010062183040.hdf', '/SREFL_CH3',
'Index', {[1  1],[1  1],[3600  7200]});
```

Actual code used in processing full set of AVHRR imagery begins here:

```
% Program to loop through AVHRR data and generate official and custom cloud
% masks
path = '/AVHRR/';   % wherever the data are
prefix = 'AVH09C1.A';

% loop through years
for yr=1981:1999  % range of years processed

   if yr==1988 | yr==1992 | yr==1996 %left out 1984 due to missing data
   days=366;
   else
   days=365;
   end
%loop through days
      for day=1:days

         dastr=num2str(day);
         if day<10
         dstr=strcat('00',dastr);
         elseif day>=10 & day<100
         dstr=strcat('0',dastr);
         else
         dstr=dastr;
         end


   if yr==1981 | yr==1982 | yr==1983 | yr==1984
         satel='N07';
         elseif yr==1985 | yr==1986 | yr==1987 | (yr==1988 & day<313)
         satel='N09';
         elseif (yr==1988 & day>=313) | (yr > 1988 & yr < 1995)
         satel='N11';
         else
         satel='N14';
   end
yrstr=num2str(yr);
```

```
counter=day
fyle = dir(strcat(path,yrstr,'/',prefix,yrstr,dstr,'.',satel,'*hdf'));

if isempty(fyle)
   continue  % if the file does not exist (is empty) for the date, this returns to beginning of loop
end
if (fyle.bytes < 10000000)   % similarly, some files existed, but lacked much if any data: bad images
   continue
end

% HDF files
BT_CH3 = hdfread(strcat(path,yrstr,'/',fyle.name), '/BT_CH3', 'Index', {[1  1],[1  1],[3600  7200]});
BT_CH4 = hdfread(strcat(path,yrstr,'/',fyle.name), '/BT_CH4', 'Index', {[1  1],[1  1],[3600  7200]});
BT_CH5 = hdfread(strcat(path,yrstr,'/',fyle.name), '/BT_CH5', 'Index', {[1  1],[1  1],[3600  7200]});
SREFL_CH3 = hdfread(strcat(path,yrstr,'/',fyle.name), '/SREFL_CH3', 'Index', {[1  1],[1  1],[3600  7200]});
% EOS files (note slightly different read command structure)
SREFL_CH1 = hdfread(strcat(path,yrstr,'/',fyle.name), 'Grid', 'Fields', 'SREFL_CH1');
SREFL_CH2 = hdfread(strcat(path,yrstr,'/',fyle.name), 'Grid', 'Fields', 'SREFL_CH2');
QA = hdfread(strcat(path,yrstr,'/',fyle.name), 'Grid', 'Fields', 'QA');

% BASIC CONVERSION OF DATA

% Scaling values between 0 and 1 for 1000x2000 (row,column) pixel subset
% “b” used for reflectance channels, and “t” used for brightness temperature
b1 = double(SREFL_CH1(700:1699,600:2599))/10000;
b2 = double(SREFL_CH2(700:1699,600:2599))/10000;
t3 = double(BT_CH3(700:1699,600:2599))/10000;
t4 = double(BT_CH4(700:1699,600:2599))/10000;
t5 = double(BT_CH5(700:1699,600:2599))/10000;
b3 = double(SREFL_CH3(700:1699,600:2599))/10000;

% Getting bit 1 (second bit): LTDR “cloudy” field (standard cloud mask)
qau = uint16(QA(700:1699,600:2599));
xx = bitget(qau,2);
xxd = double(xx);

% better stretch of thermal bands for display purposes
t3s = t3*9-2;
t4s = t4*9-2;
t5s = t5*9-2;

% normalized difference ratios
nt = (t3s-t4s)./(t3s+t4s);
% the above is my custom attempt at band 3 “reflectance” by normalizing the
% difference between band 3 thermal and band 4 thermal
rat = (b3-nt)./(b3+nt);
% the above for some marginal clouds that seem to be higher in b3
```

```
ndvi = (b2-b1)./(b2+b1);

% SERIES OF RULES FOR CLOUDS

% Basic minimum low level of brightness in red band
r1 = b1 >= 0.12;

% Complex of rules involving various levels of derived b3 “reflectance” (clouds > snow) and custom
normalized thermal ratio (clouds > snow), with additional variations including b1 reflectance and NDVI.
r2 = (b3 >= 0.2 & nt>=.1) |b3>=.3 | nt>=.15 |(nt>=.1&(b3+nt)>=.25)
|((b3+nt>=.2)&(b1>=.35|(b1>=0.2&ndvi<.03)))|((b3>=.07&nt>=.07)&(b1>=.35|(b1>=0.2&ndvi<.03)))|((b
3+nt)>=.09&b3>=(nt-.01)&(b1>=.35|(b1>=.2&ndvi<.03)))|(b3>=.04&rat>.16&ndvi<.06);

% Rule to remove sunglint; NDVI for anomalous sunglint with very low b1 values
r3 = ndvi>=0.1|b1>(t3s-0.45);

% A COUPLE “FIXES” TO DEAL WITH ADDITIONAL ANOMALOUS DATA

% Anomalous data in bands 1 or 2, along with some basic rules to decide cloud
rfix = (b1>=1 | b2>=1 | b1<=0 | b2 <= 0) & (nt>=.15 | b3>=.3);

% Anomalous data in band 3 (<0) with relatively liberal pair of cloud rules or
% anomalous data in band 3 (=0) with more conservative pair of cloud rules
rfix2 = (b3<0 & b1>=0.2 & nt>0)|(b3==0 & (b1 >=0.35 & nt>=.15));

% GROUPING ALL THE RULES

% Have to meet all of the first 3 rules, OR one of the two fixes
rcld = double((r1 & r2 & r3)|rfix|rfix2);

% WRITING TO FILE
% Convert custom and standard cloud masks to int16
rcldint16 = int16(rcld);
xxdint16 = int16(xxd);

% write paths, names (yrstr and dstr relate to years and days derived from loops)
wrpathrc = '/AVHRR/rcld_masks/';
wrpathxx = '/AVHRR/xxd_masks/';
filerc = strcat(yrstr,'.',dstr,'.rcld.mat');
filexx = strcat(yrstr,'.',dstr,'.xxd.mat');
save(strcat(wrpathrc,filerc),'rcldint16');
save(strcat(wrpathxx,filexx),'xxdint16');

        end  % end of days loop
end   % end of years loop
```

2.B.2. MODIS Code

MODIS Aqua Data Processing:
Example code for deriving daily standard and custom cloud masks from MYD09GA daily reflectance HDF data and looping through multiple years in MATLAB.

Code is provided below for looping through multiple years of MODIS Aqua (and Terra until 2007) imagery for three sinusoidal grid tiles. First are some examples for reading in some images from a single day into MATLAB. June 17th, 2012 data are used in the example. The year and calendar day are indicated following "MYD09GA.A" in the import (hdfread) commands below. Horizontal and vertical tile numbers follow the calendar date. The last set of numbers ("2012…") relates to a date and time of processing.

Each MATLAB line of code ends in a semi-colon (;).
Comments below either in bold text or, using the MATLAB convention for non-code comments, in green and following the pound (%) symbol.

MATLAB's HDF Import Tool streamlines the reading of the individual raster grids of data (or "Scientific Data Sets", SDS) within the HDF file, making the initial import commands below (i.e., hdfread) a relatively automated "point and click" process once the file has been navigated to.

Within the "EOS" subset of the file data (rather than HDF, although little discernible difference):

```
% Surface reflectance for channel 1 (red)
sur_refl_b01_1 =
hdfread('C:\MODIS_MYD09_2012daily\MYD09GA.A2012169.h08v05.005.2012171055901.hdf',
'MODIS_Grid_500m_2D', 'Fields', 'sur_refl_b01_1');
% Surface reflectance for channel 2 (near infrared)
sur_refl_b02_1 =
hdfread('C:\MODIS_MYD09_2012daily\MYD09GA.A2012169.h08v05.005.2012171055901.hdf',
'MODIS_Grid_500m_2D', 'Fields', 'sur_refl_b02_1');
% Surface reflectance for channel 3 (blue)
sur_refl_b03_1 =
hdfread('C:\MODIS_MYD09_2012daily\MYD09GA.A2012169.h08v05.005.2012171055901.hdf',
'MODIS_Grid_500m_2D', 'Fields', 'sur_refl_b03_1');
% Surface reflectance for channel 4 (green)
sur_refl_b04_1 =
hdfread('C:\MODIS_MYD09_2012daily\MYD09GA.A2012169.h08v05.005.2012171055901.hdf',
'MODIS_Grid_500m_2D', 'Fields', 'sur_refl_b04_1');
% Surface reflectance for channel 5 (shortwave infrared)
sur_refl_b05_1 =
hdfread('C:\MODIS_MYD09_2012daily\MYD09GA.A2012169.h08v05.005.2012171055901.hdf',
'MODIS_Grid_500m_2D', 'Fields', 'sur_refl_b05_1');
% Surface reflectance for channel 6 (shortwave infrared)
```

```
sur_refl_b06_1 =
hdfread('C:\MODIS_MYD09_2012daily\MYD09GA.A2012169.h08v05.005.2012171055901.hdf',
'MODIS_Grid_500m_2D', 'Fields', 'sur_refl_b06_1');
% Surface reflectance for channel 7 (shortwave infrared)
sur_refl_b07_1 =
hdfread('C:\MODIS_MYD09_2012daily\MYD09GA.A2012169.h08v05.005.2012171055901.hdf',
'MODIS_Grid_500m_2D', 'Fields', 'sur_refl_b07_1');
```

Actual code used in processing begins here:

```
% Program to loop through MODIS Aqua data and generate custom cloud masks

path = '/modis/MYD09GA.005/';
header='/MYD09GA.A';

for t=1:3                    % Loop through 3 sinusoidal grid tiles
                if t==1
                   tile = 'h08v05';
                end
                if t==2
                   tile = 'h08v04';
                end
                if t==3
                   tile = 'h09v04';
                end
                %if t==4
                   %tile = 'h09v05';
                %end
for a=2002:2012          % Year loop for Aqua; Terra actually started in 2000
   yrstr=num2str(a);
   for b=1:12            %  Month of year loop to create two-digit month number
      mostr=num2str(b);
      if b<10
         monstr=strcat('0',mostr);
      end  %these just end if statement
      if b>=10
         monstr=mostr;
      end   %this just ends this if statement

         if (a==2004 | a==2008 | a==2012) & b==2
         days=29;
         elseif b==2
         days=28;
         elseif b==1 | b==3 | b==5 | b==7 | b==8 | b==10 | b==12
         days=31;
```

```
    else
    days=30;
    end

      for c=1:days     % Day of month loop to get 2-digit day number for each day
        dastr=num2str(c);
        if c<10
        dayestr=strcat('0',dastr);
        end
        if c>=10
        dayestr=dastr;
        end


  % Day of year through using Julian Date  – probably not the most sophisticated approach
  mjd = mjuliandate(a,b,c);
  if a == 2002
    jdate = mjd-52274;
  elseif a == 2003
    jdate = mjd-52639;
  elseif a == 2004
    jdate = mjd-53004;
  elseif a == 2005
    jdate = mjd-53370;
  elseif a == 2006
    jdate = mjd-53735;
  elseif a == 2007
    jdate = mjd-54100;
  elseif a == 2008
    jdate = mjd-54465;
  elseif a == 2009
    jdate = mjd-54831;
  elseif a == 2010
    jdate = mjd-55196;
  elseif a == 2011
    jdate = mjd-55561;
  elseif a == 2012
    jdate = mjd-55926;
  end

jstr=num2str(jdate);
if jdate<10
  daystr=strcat('00',jstr);
elseif jdate>=10 & jdate<100
  daystr=strcat('0',jstr);
elseif jdate>=100
  daystr=jstr;
```

```
   end


fyle = dir(strcat(path,yrstr,'.',monstr,'.',dayestr,header,yrstr,daystr,'.',tile,'*hdf'));

if isempty(fyle)
   continue % if the file does not exist (is empty) for the date, this returns to beginning of loop
end
if (fyle.bytes == 0) % some files existed in name, but lacked data: bad images
   continue
end

bd1raw =
hdfread(strcat(path,yrstr,'.',monstr,'.',dayestr,'/',fyle.name),'MODIS_Grid_500m_2D','Fields','sur_refl_b0
1_1');

bd2raw =
hdfread(strcat(path,yrstr,'.',monstr,'.',dayestr,'/',fyle.name),'MODIS_Grid_500m_2D','Fields','sur_refl_b0
2_1');

bd3raw =
hdfread(strcat(path,yrstr,'.',monstr,'.',dayestr,'/',fyle.name),'MODIS_Grid_500m_2D','Fields','sur_refl_b0
3_1');

bd4raw =
hdfread(strcat(path,yrstr,'.',monstr,'.',dayestr,'/',fyle.name),'MODIS_Grid_500m_2D','Fields','sur_refl_b0
4_1');

bd5raw =
hdfread(strcat(path,yrstr,'.',monstr,'.',dayestr,'/',fyle.name),'MODIS_Grid_500m_2D','Fields','sur_refl_b0
5_1');

bd6raw =
hdfread(strcat(path,yrstr,'.',monstr,'.',dayestr,'/',fyle.name),'MODIS_Grid_500m_2D','Fields','sur_refl_b0
6_1');

bd7raw =
hdfread(strcat(path,yrstr,'.',monstr,'.',dayestr,'/',fyle.name),'MODIS_Grid_500m_2D','Fields','sur_refl_b0
7_1');

% Removing scale factor to get reflectance back between 0 and 1
bd1 = double(bd1raw)/10000;
bd2 = double(bd2raw)/10000;
bd3 = double(bd3raw)/10000;
bd4 = double(bd4raw)/10000;
bd5 = double(bd5raw)/10000;
bd6 = double(bd6raw)/10000;
bd7 = double(bd7raw)/10000;
```

```
clear bd1raw bd2raw bd3raw bd4raw bd5raw bd6raw bd7raw

% various normalized difference ratios
ndvi = (bd2-bd1)./(bd2+bd1);
nd15 = (bd1-bd5)./(bd1+bd5);
nd17 = (bd1-bd7)./(bd1+bd7);
nd35 = (bd3-bd5)./(bd3+bd5);
nd37 = (bd3-bd7)./(bd3+bd7);
nd13 = (bd1-bd3)./(bd1+bd3);
nd31 = (bd3-bd1)./(bd3+bd1);
%cldresid = bd5-((bd1+bd7)/2);

maskn05 = (bd2 >= .05);  % just to mask out dark stuff – not cloud even if high ratio
nd15m = nd15 .* maskn05;
nd17m = nd17 .* maskn05;
nd35m = nd35 .* maskn05;
nd37m = nd37 .* maskn05;
ndvim = ndvi .* maskn05;

nd15m5 = nd15m + 0.5;  % I think this later called nd15tp
nd37mp = nd37m + 1;
nd37mpd = nd37mp/2;

% generating true color normalized deviations from average true color brightness
avg3 = (bd1 + bd4 + bd3)/3;
b1ndev = (bd1-avg3)./(bd1+avg3);
b4ndev = (bd4-avg3)./(bd4+avg3);
b3ndev = (bd3-avg3)./(bd3+avg3);

% adding NIR criterion to the NDSI-like vis vs. shortwave ratios
thr35 = nd35m > -0.15 & bd2 >= 0.12;
thr35d = double(thr35);
thr37 = nd37m >0 & bd2 >= 0.12;
thr37d = double(thr37);

thrstmult = thr35 & thr37;
thrstmultd = double(thrstmult);
thrstor = thr35 | thr37; % Should thresh change with or?
% probably should've added a criterion based on the sum of thr35&thr37
thrstord = double(thrstor);

% Rules related to deviations from grayscale in true color bands
tc1 = b1ndev > -.15 & b1ndev < .08;
tc2 = b4ndev > -.05 & b4ndev < .05;
tc3 = b3ndev > -.06 & b3ndev < .15;
tc4 = b3ndev > -.08 & b3ndev <= -.06 & (ndvim >= .03 | (nd17m > 0.4 & nd15m > 0.2 & bd2 >= 0.12));
```

```
tc1d = double(tc1);
tc2d = double(tc2);
tc3d = double(tc3);
tc4d = double(tc4);

tc4out = tc1d & tc2d & (tc3d | tc4d); % group of true color rules
tc4outd = double(tc4out);

tc4outnir = tc4outd & bd2 >= 0.12;  % adding NIR criterion to group of true color rules
tc4outnird = double(tc4outnir);

brite = (bd1 + bd4 + bd3) > 0.9;  % brightness rule for exceptional pixels that don't meet SWIR criteria

badnirfix = bd2 < .01 & bd1 > 0.2;

cloudsnowb = badnirfix | (tc4outnir & (thrstor | brite)); % Determines cloud&snow from rest
cloudsnowbd = double(cloudsnowb);

avg751 = (bd7 + bd5 + bd1)/3;

ndvim_p05 = ndvim + 0.5;

% NEW RATIOS FOR CLASSIFICATION TREE
ndvit = ndvi .* cloudsnowbd; %  multiplying serves as a mask to apply to cloud&snow
ndvita = (ndvit + 0.25)*2; % stretches low ndvi values, centers near 0.5
nd17t = nd17 .* cloudsnowbd;
nd15t = nd15 .* cloudsnowbd;
bd5sub = bd5 - 0.2;
nd15sub = (bd1 - bd5sub)./(bd1 + bd5sub); % gives slightly different axis of separation to nd15
nd15subt = nd15sub .* cloudsnowbd;
nd15tp = nd15t + 0.5;

red = ((b1ndev * 10)+0.5) .* cloudsnowbd;
green = ((b4ndev * 10)+0.5) .* cloudsnowbd;
blue = ((b3ndev * 10)+0.5) .* cloudsnowbd;

% variables consisting of added ratios
six = (nd17t + nd15tp + nd15subt + ndvita + red + green)/6;
%five = (nd17t + nd15tp + nd15subt + ndvita + red)/5;
four = (nd17t + nd15tp + ndvita + red)/4;
%fourb = (nd17t + nd15tp + nd15subt + ndvita)/4;
three = (nd17t + nd15tp + ndvita)/3;
%threeb = (nd17t + nd15subt + ndvita)/3;
%two = (nd17t + nd15tp)/2;
%twob = (nd17t + nd15subt)/2;
%twoc = (ndvita + nd15subt)/2;
%twod = (ndvita + nd17t)/2;
```

```
%twoe = (red + green)/2;

% TREE 23p: final rules
% These separate snow from clouds: to generate the snow from the cloud&snow
spp1 = six < .5707 & three >= .6238 & green >= .5332;
spp2 = six >= .5707 & nd17t >= .6396;
spp3 = six >= .5707 & nd17t < .6396 & four >= .5537 & green >= .4608 & three > .467;

allspp = (spp1|spp2|spp3)&(1-badnirfix);  % SNOW
allsppt = allspp .* cloudsnowbd;
allclppt = cloudsnowbd - allsppt; % CLOUD

% WRITING TO FILES
snowint16 = int16(allsppt);
cloudint16 = int16(allclppt);

wrpathcl = '/modis/MYD09GA/';
wrpathsn = '/modis/MYD09GA/';
filecl = strcat(yrstr,'.',monstr,'.',dayestr,'.cloud.mat');
filesn = strcat(yrstr,'.',monstr,'.',dayestr,'.snow.mat');
save(strcat(wrpathcl,tile,'/cloud_masks/',filecl),'cloudint16');
save(strcat(wrpathsn,tile,'/snow_masks/',filesn),'snowint16');


                    end % end of c=days loop

                end % end of b=months loop

            end % end of a=years loop

    end % end of t=tiles loop
```

Code for MODIS Terra problems: bad band 5 beginning in 2007 and bad band 3 beginning in 2008.

Rather than provide all of the redundant code, relevant changes are provided here. Incorporating the changes into the standard code would require some comparison with section of the code above, but should be manageable.

```
% FIX for bad band5

twod = (ndvita + nd17t)/2;  %  NOTE: NEED THIS ONE FOR TERRA (was commented out for Aqua)


% Some new ratios to help identify bad band 5
nd52 = (bd5 - bd2) ./ (bd5 + bd2);
nd52t = nd52 .* cloudsnowbd;
nd517 = (bd5 - bd1 - bd7) ./ (bd5 + bd1 + bd7);
nd517t = nd517 .* cloudsnowbd;

bad5m = (nd52t > 0.12 & nd517t > -0.3 & cloudsnowbd) | (bd5<.01 & cloudsnowbd) | (nd15tp > (nd17t
+ 0.15) & nd15tp > .5); % This line identifies the bad band 5 data from what is already classified as cloud
or snow

snowx = (twod>0.62 & nd17t>0.4 & ndvita>0.55) | nd17t>0.8;
snowxb = snowx.*bad5m;
cloudxb = (1-snowx).*bad5m;
allclpptx = allclppt + cloudxb - snowxb; % Adding newly classified cloud and subtracting newly classified
snow to/from original cloud
allspptx = allsppt + snowxb - cloudxb; % Adding newly classified snow and subtracting newly classified
cloud to/from original snow

snowint16 = int16(allspptx);
cloudint16 = int16(allclpptx);
```

```
% FIX for bad band 3

% need to import the sensor azimuth and sensor zenith angle
azim =
hdfread(strcat(path,yrstr,'.',monstr,'.',dayestr,'/',fyle.name),'MODIS_Grid_1km_2D','Fields','SensorAzimu
th_1');

zeni =
hdfread(strcat(path,yrstr,'.',monstr,'.',dayestr,'/',fyle.name),'MODIS_Grid_1km_2D','Fields','SensorZenit
h_1');
```

```
-----------------------------------------------


azimd = double(azim+18000)/36000;  % Scaling azimuth (-18000 to + 18000 to 0 to 1)
zenid = double(zeni/10000); % Scaling zenith 0 to 1
azr = zeros(2400,2400);  % just to establish dimensions for resampled reduced resolution
znr = zeros(2400,2400); % "" ""
for zx=1:1200  % reduce resolution loop for azim and zeni
   for zy=1:1200
      zx2=zx*2;
      zy2=zy*2;
      zx2m1 = zx2-1;
      zy2m1 = zy2-1;
      azr(zy2,zx2) = azimd(zy,zx);
      azr(zy2m1,zx2) = azimd(zy,zx);
      azr(zy2,zx2m1) = azimd(zy,zx);
      azr(zy2m1,zx2m1) = azimd(zy,zx);
      znr(zy2,zx2) = zenid(zy,zx);
      znr(zy2m1,zx2) = zenid(zy,zx);
      znr(zy2,zx2m1) = zenid(zy,zx);
      znr(zy2m1,zx2m1) = zenid(zy,zx);
   end
end

---------------------------------------------------

% stuff to "fix" the true color bad b3 problem
redf = ((b1ndev * 10)+0.5);  % no changes here from before: just adding f to indicate use
greenf = ((b4ndev * 10)+0.5);
bluef = ((b3ndev * 10)+0.5);
xx = azr < 0.5;  % these are negative (backscatter) azimuths originally
yy = azr >= 0.5; % these are the azimuths >= 0 (forward scatter) originally
bluefixxf = (bluef-znr).*xx;  % blue correction based on zenith, only applied where xx satisfied
bluefixyf = bluef .* yy;   % no need for zenith-based blue correction here where yy satisfied (not xx)
bluefixf = bluefixxf + bluefixyf; % merging the two above
redfixxf = (redf+(znr/2)).*xx; % required change to red as a result of changing the blue
redfixyf = redf .* yy;
redfixf = redfixxf + redfixyf;
greenfixxf = (greenf +(znr/2)).*xx; % required change to green as a result of changing the blue
greenfixyf = greenf .* yy;
greenfixf = greenfixxf + greenfixyf;

% values of same rules from before changed for adjusted variables
tc1 = redfixf > -1 & redfixf < 1.3;
tc2 = greenfixf > 0 & greenfixf < 1;
tc3 = bluefixf > -0.1 & bluefixf < 2;
```

```
tc4 = bluefixf > -0.3 & bluefixf < -0.1 & (ndvim >= .03 | (nd17m > 0.4 & nd15m > 0.2 & bd2 >= 0.12));

tc1d = double(tc1);
tc2d = double(tc2);
tc3d = double(tc3);
tc4d = double(tc4);

tc4out = tc1d & tc2d & (tc3d | tc4d);
tc4outd = double(tc4out);
clear tc1d tc2d tc3d tc4d

tc4outnir = tc4outd & bd2 >= 0.12;
tc4outnird = double(tc4outnir);

brite = (bd1 + bd4 + bd3) > 0.9;

badnirfix = bd2 < .01 & bd1 > 0.2;

cloudsnowb = badnirfix | (tc4outnir & (thrstor | brite));
cloudsnowbd = double(cloudsnowb);

clear tc4outnir tc4outnird tcout tcoutd
clear thr35 thr35d thr37 thr37d thrstmult thrstmultd thrstor thrstord
clear tc4out tc4outd tc1 tc2 tc3 tc4 tc1d tc2d tc3d tc4d

avg751 = (bd7 + bd5 + bd1)/3;

ndvim_p05 = ndvim + 0.5;

% NEW RATIOS FOR CLASSIFICATION TREE

ndvit = ndvi .* cloudsnowbd;
ndvita = (ndvit + 0.25)*2;
nd17t = nd17 .* cloudsnowbd;
nd15t = nd15 .* cloudsnowbd;
bd5sub = bd5 - 0.2;
nd15sub = (bd1 - bd5sub)./(bd1 + bd5sub);
nd15subt = nd15sub .* cloudsnowbd;
nd15tp = nd15t + 0.5;

% note that “fixed” variables replace equivalent non-fixed variables where necessary
%six = (nd17t + nd15tp + nd15subt + ndvita + red + green)/6;
sixfix = (nd17t + nd15tp + nd15subt + ndvita + redfixf + greenfixf)/6;
%five = (nd17t + nd15tp + nd15subt + ndvita + red)/5;
%four = (nd17t + nd15tp + ndvita + red)/4;
fourfix = (nd17t + nd15tp + ndvita + redfixf)/4;
%fourb = (nd17t + nd15tp + nd15subt + ndvita)/4;
```

```
three = (nd17t + nd15tp + ndvita)/3;
%threeb = (nd17t + nd15subt + ndvita)/3;
%two = (nd17t + nd15tp)/2;
%twob = (nd17t + nd15subt)/2;
%twoc = (ndvita + nd15subt)/2;
twod = (ndvita + nd17t)/2;  % NOTE: NEED THIS ONE FOR MOD09 (Terra)
%twoe = (red + green)/2;

% TREE 23p
%spp1 = six < .5707 & three >= .6238 & green >= .5332;
%spp2 = six >= .5707 & nd17t >= .6396;
%spp3 = six >= .5707 & nd17t < .6396 & four >= .5537 & green >= .4608 & three > .467;
spp1 = sixfix < .5707 & three >= .6238 & greenfixf >= .5332;
spp2 = sixfix >= .5707 & nd17t >= .6396;
spp3 = sixfix >= .5707 & nd17t < .6396 & fourfix >= .5537 & greenfixf >= .4608 & three > .467;

allspp = (spp1|spp2|spp3)&(1-badnirfix);
allsppt = allspp .* cloudsnowbd;
allclppt = cloudsnowbd - allsppt;

clear spp1 spp2 spp3 allspp

% FIX for bad band5

nd52 = (bd5 - bd2) ./ (bd5 + bd2);
nd52t = nd52 .* cloudsnowbd;
nd517 = (bd5 - bd1 - bd7) ./ (bd5 + bd1 + bd7);
nd517t = nd517 .* cloudsnowbd;
bad5m = (nd52t > 0.12 & nd517t > -0.3 & cloudsnowbd) | (bd5<.01 & cloudsnowbd) | (nd15tp > (nd17t
+ 0.15) & nd15tp > .5);
clear nd52 nd52t nd517 nd517t
snowx = (twod>0.62 & nd17t>0.4 & ndvita>0.55) | nd17t>0.8;
snowxb = snowx.*bad5m;
cloudxb = (1-snowx).*bad5m;
allclpptx = allclppt + cloudxb - snowxb;
allspptx = allsppt + snowxb - cloudxb;

snowint16 = int16(allspptx);
cloudint16 = int16(allclpptx);
```

Chapter 3. The Sierra Nevada's Cloud Forest: Cloud Frequency the Strongest Predictor in Model of Climatic Controls on the Distribution of Giant Sequoia

Abstract

The restriction of giant sequoia (*Sequoiadendron giganteum* (Lindl.) Buchholz) to primarily the southern Sierra Nevada in California has long puzzled researchers (Hartesveldt *et al*, 1975). There were no obvious differences in current climate to explain the distribution. New, relatively high resolution cloud climatologies derived from AVHRR and MODIS satellite imagery demonstrate that the main part of the sequoia belt is the cloudiest area of California in March and April on average, and among the cloudiest areas throughout the year. Species distribution models reveal that April cloud cover frequency, in particular, is the strongest predictor of sequoia distribution from a large set of climate variables, but many other monthly cloud frequencies are strongly correlated with April. Frequent cloud formation appears to be associated with regular pooling of relatively humid air along the abrupt western slope of the Sierra Nevada in this region throughout the year. Cloud cover could be important for its increased ratio of diffuse to direct solar radiation, reduced daytime temperature and vapor pressure deficit, and retention of winter snowpack, all of which would contribute to a reduction of soil water deficit late in the relatively dry growing season. An imperfect match with all groves suggests that additional factors may be involved, but cloud cover and its associated conditions likely play a role in favoring sequoia in most cases. Determining the exact mechanistic controls would require substantial further inquiry, but it is now clear that the climate on the western slope of the southern Sierra Nevada, the sequoia belt, is substantially different from other portions of the mountain range.

Introduction

Giant sequoia (*Sequoiadendron giganteum* (Lindl.) Buchholz), the largest tree in the world by volume, is a narrow endemic on the western slope of the Sierra Nevada Mountains of California, restricted to a mid-elevation belt of approximately 75 groves (Hartesveldt *et al*, 1975). Most of these groves lie in the southern Sierra, south of the Kings River, with only sporadic outlying groves to the north (Figure 1). In particular, the highest and lowest elevation trees (2700 m and 850 m), tallest of the species (95 m), biggest by volume (General Sherman, ~1500 m^3), and biggest grove of trees (Redwood Mountain Grove), all occur within the Kaweah River drainage near the center of the southern concentration of groves. The limited distribution pre-dates logging, as the present boundaries have been relatively fixed for over 500 years (Rundel, 1971). The restricted distribution contrasts with that of its cousin, the coast redwood (*Sequoia sempervirens* (D. Don) Endl.) – the *tallest* tree in the world, which is relatively common in its Coast Range distribution from Central California to the southwestern corner of Oregon (Figure 1), despite its much more extensive logging history. Vigorous reproduction and growth of giant sequoia following logging and fire (Harvey *et al*, 1980) has revealed that the

current climate continues to favor the tree, and that it is not just a relict of historic cooler climates, despite its capability of living for several thousand years.

The dearth of giant sequoia (hereafter, "sequoia", not to be confused with the genus of its cousin) in the northern Sierra has long puzzled researchers, as appropriate elevations and more than sufficient precipitation (Figure 2 and Figure 3, respectively) appear to exist in the north. As reviewed by Hartesveldt et al. (Hartesveldt *et al*, 1975), early explanations that varied from historic factors (e.g., elimination by glaciation during Pleistocene (Muir, 1876)) to climatic factors such as winter cold (Wilson, 1928) do not hold up to scrutiny. For example, not all areas of the northern Sierra were glaciated (Figure 4) and the winter cold in the high southern groves is more extreme than that of lower elevation groves in the north. Sequoia has had a long association with fire (Harvey *et al*, 1980), but the relationship is complex (Stephenson, 1999), and there are no clear latitudinal differences in fire regime across the Sierra (e.g., LANDFIRE data; (Rollins, 2009)). The most commonly accepted hypothesis for the distribution has been that the species is restricted to areas of appropriate temperature, with adequate groundwater (Rundel, 1972), as part of a "hydrophytic community" (Axelrod, 1959). However, even Muir observed that sequoia outcompetes its rivals on both wet and dry soils (Muir, 1876). To explain the frequent association with moisture, Muir argued that sequoia were responsible for water's frequent presence, rather than vice-versa. Rundel (Rundel, 1972), alternatively, suggested that this groundwater came from summer thunderstorms over the "High Sierra" (higher elevations of the Sierra Nevada) via percolation. A close examination of grove boundaries with respect to topography and the often distant location of the High Sierra and summer precipitation calls this assertion into question. Why are the groves so concentrated in the southern Sierra, on the western slopes, rather far removed and hydrologically isolated from the High Sierra? Particularly, why is sequoia common from the Tule River watershed south to the Greenhorn Mountains - areas which have no possibility of hydrologic connection to the High Sierra? Why is the sequoia belt (~1500-2400 meters in the south & ~1400-2000 meters in the north) so mesic in general, accompanied by many wet meadows and relatively rich herb and shrub diversity (Barbour and Minnich, 1988), yet it does not have higher precipitation than comparable (or even much lower) elevations in the northern Sierra? Given that groundwater percolation from High Sierra thunderstorms is unlikely, other factors would need to explain this mesic setting and the associated sequoia distribution.

The frequent concentration of clouds over mid-elevations of the west slope of the southern Sierra (what I will refer to as the "sequoia belt") seen in individual satellite images (Waller, 2006) and custom cloud frequency products (Chapter 2) indicated that the climate of the southern Sierra is quite different from that of the northern Sierra, and were thought to be a possible factor in the puzzling scarcity of sequoia in the latter. Clouds or fog (clouds at the surface) are known to be an important control on species distributions in many ecosystems, including tropical montane cloud forests (Sugden and Robins, 1979; Gentry, 1992; Sklenár *et al*, 2008), and have been related to the distribution of sequoia's cousin, the coast redwood (Cannon, 1901; Dawson, 1998). For sequoia, cloud cover could be important for its moderation of temperature, including reduced daytime temperatures, which would also increase relative humidity and reduce vapor pressure deficit. Sequoia could also benefit from the increased ratio of diffuse to direct light associated with cloud cover. Winter and early spring cloud cover would contribute to the retention of snowpack by reducing sublimation and delaying snowmelt. All of

these factors would contribute to a reduction of soil water deficit late in the relatively dry growing season.

In this study, species distribution modeling is used to investigate climatic controls on the distribution of sequoia. In particular, custom cloud frequency products derived from satellite data are incorporated into the analysis to assess their importance in comparison to more conventionally used climate variables such as temperature and precipitation. Individual monthly frequencies are included to determine whether evident seasonality in cloud frequencies is relevant to sequoia distribution. Additional processing of satellite data allows the improved assessment of seasonal and longer-term trends in a particular pattern of cloud formation that is relatively restricted to the sequoia belt.

Data and Methods

An earlier analysis (Waller, 2006) using periodic observations from Landsat Thematic Mapper and other satellite data found that cloud banks frequently formed in the region of sequoia groves (Figure 7), especially in the spring. To explore whether this related to any systematic differences in (cloud) climate, custom-generated cloud climatologies from two different historic satellite sensors with daily coverage were assessed: the Advanced Very High Resolution Radiometer (AVHRR, 1981-1999) and the Moderate Resolution Imaging Spectroradiometer (MODIS, 2000-2012). The length of the satellite record was not absolutely essential to estimate cloud frequencies, but was already processed for a study looking at long-term trends in California's Central Valley winter Tule fog (Baldocchi and Waller, 2014). As the core of the sequoia distribution conveniently, and perhaps not coincidentally, lies due east of the core of Central Valley Tule fog formation, this dataset was perfectly suited for sequoia analysis as well. Chapter 2 of this dissertation describes the custom classification of daily cloud cover through the processing of daily raw radiance and reflectance data from the AVHRR and MODIS sensors, respectively, in more detail. This creation of custom products was deemed necessary, as existing cloud cover masks for these satellite data were found to have problems of confusion between clouds and bright desert areas (Figure 5) and between clouds and snow (Figure 6). Even recent more sophisticated cloud classification algorithms (Heidinger *et al*, 2012) continue to have problems with discriminating clouds from snow and other bright areas [see Chapter 2] – a critical concern in temperate mountainous regions. Other cloud climatology data sets, such as the International Satellite Cloud Climatology Project (ISCCP) (Schiffer and Rossow, 1983) are at such a coarse resolution (32 km at finest) as to prevent determining whether they confuse cloud and snow, at least over the Sierra Nevada. In any case, such datasets would be too coarse to be of much use in relatively fine scale understanding of species distributions, and species distribution modeling at the ~1 km resolution of many large area climate data sets.

While AVHRR data are collected at a nominal resolution of 1.1 km, and thus reasonably close to that of MODIS (nominally 500 meters for the reflectance data used here), the Long-Term Data Record (LTDR) (Pedelty *et al*, 2007) versions used here are processed and stored at a coarser 0.05 degree (~5 km) resolution, as a means of data reduction and standardization of the product (e.g., off-nadir viewing can greatly enlarge the true ground-resolution of the data for this

wide field of view satellite). Products derived from MODIS data (Aqua, in particular), however, were the main source of cloud climatological data used in sequoia species distribution models described below. These products are more accurate, at finer resolution (finer even than the other climate variables used in species distribution modeling here (1 km)), and of sufficient duration for a representative monthly cloud climatology.

"Sequoia Cloud Formation" (SCF) frequency maps were also created to highlight days on which cloud cover was focused over the sequoia belt specifically. Daily imagery can sometimes highlight the relationship with sequoia groves better than climatological averages (which can get somewhat swamped by more general cloud cover), especially for some of the outlying groves (Figure 7). Applying a spatial filter to daily cloud masks allows the assessment of the precise spatial patterns in cloud frequency associated with the particular "sequoia belt" pattern, and allows a better understanding of the pattern's climatology and trends over time, relatively independently of the climatology and trends in overall cloud cover. Twenty-five manually selected points inside the region of frequent cloud cover were compared daily with 75 manually selected and well distributed points surrounding the sequoia region (Figure 8) in MATLAB. If a greater number of the 25 points within the region were cloudy than there were of the 75 points outside the region, this was considered a SCF day. This spatial filter is analogous to an approach that was used to identify days with winter Tule fog in California's Central Valley (Baldocchi and Waller, 2014) – another cloud pattern with a distinctive spatial signature. Monthly and annual SCF climatologies were generated from these daily products. This approach has the obvious potential for spatial bias (with sensitivity to point numbers and placement), but was only used as a relative index of approximate SCF frequency for comparing among months and years – NOT for any of the predictive modeling in this study.

Various other processing of the monthly cloud frequency products was performed to further investigate relevant aspects of the cloud climatology of the sequoia belt. This included principal component analyses (across the 12 monthly averages for both MODIS Terra and Aqua data, within California and Oregon), the computation of monthly standard deviations (across years) and the associated monthly coefficients of variation (across years). The MODIS-based cloud frequencies from Terra were subtracted from those for Aqua to assess the development of clouds over the course of a day.

Assessment of controls on sequoia distribution involved the use of several statistical approaches. Maxent (Phillips *et al*, 2004; Phillips *et al*, 2006; Phillips and Dudík, 2008), a species distribution modeling program predicated upon maximum entropy principles, has been found to perform very well in comparisons of many models (Elith* *et al*, 2006). An advantage of Maxent is that it only requires species presence locations as the dependent variable, along with various predictor variables (mostly climate here, including monthly cloud cover frequency; Table 1). Multiple Maxent models of sequoia were performed at the scale of California and Oregon and at the scale of the Sierra Nevada (Jepson ecoregion (Hickman, 1993)) with different sets of predictor variables, both to explore how modeling extent affected the results, and to better ascertain reasons for sequoia absence in the Pacific Northwest.

Classification trees (Breiman *et al*, 1984; Ripley, 1996) and logistic regression were also performed in the R software package as alternative statistical approaches. These statistical

approaches require species absence data in addition to presence data; absence data can be harder to acquire than presence data, but in the case of a well-documented species like sequoia, there is certainty that areas that have not been mapped for species presence can be considered reliable absences. Similar to the Maxent model runs, analyses were performed at the extent of both California and Oregon, as well as at the extent of the Sierra Nevada ecoregion.

Sequoia grove locations were derived from combining two different data sets. One high resolution map of grove boundaries was obtained for the main (southern Sierra) portion of the sequoia distribution from the National Park Service. A digital coverage of points from the Jepson Online Interchange – California Floristics (http://ucjeps.berkeley.edu/interchange/) was used for the remaining groves (which tend to be much smaller). Single sample points were randomly placed within each sequoia grove polygon so as to maximize independence among samples and these were combined with the points from the coverage (edited to ensure accuracy and native locations only). For statistical modeling requiring absence data, absence locations were generated by first buffering the grove locations (5 km buffer to account for a little bit of sample location error and climatic autocorrelation), and then ensuring that absence points fell outside of these buffers. Absence points were selected from a variety of locations for different models (extent of absences can vastly influence model results; source), including all of California and Oregon, the entire Sierra Nevada, and the Sierra Nevada with an extra concentration of absence points from just outside of the buffered polygons.

In addition to the monthly cloud cover frequencies, other climate variables were derived from PRISM (Daly *et al*, 1994) and Daymet (Thornton *et al*, 1997). PRISM data have been found to be more reliable than Daymet (Daly, 2006; Daly *et al*, 2008), so its data are used for temperature and precipitation. For other variables not provided by PRISM – at least not at a relatively high 30 arc-second (~800 m) resolution free of charge (e.g., radiation and humidity), Daymet data at 1-km resolution were used. Similarly, in a custom-generated potential evapotranspiration (PET) product using a modified Priestley-Taylor model (Priestley and Taylor, 1972), PRISM data were used for the monthly temperature and precipitation, whereas Daymet data were used for monthly radiation and humidity (vapor pressure). This PET dataset, and water balance products derived from it, were also used in the modeling. These included minimum and maximum monthly water balance (precipitation – PET), cumulative annual water balance (i.e., all monthly water surplus carried through), cumulative annual deficit (defined as no carry-over of monthly water surplus), and months of deficit (i.e., PET> precipitation). Other humidity-related variables (e.g., relative humidity and vapor pressure deficit) were derived from the monthly Daymet humidity (vapor pressure) and mean temperature (average of minimum and maximum temperature) data. Elevation was derived from a U.S. Geological Survey Digital Elevation Model (DEM). All PRISM, elevation, and cloud frequency data were resampled and reprojected to the 1-km resolution and Lambert Conformal Conic projection of Daymet data. In general, minimum and maximum monthly values, in addition to annual sums, averages, and ranges were used instead of all monthly values, in order to reduce the total number of variables. Table 1 contains a list of the 45 variables used in statistical modeling, including 12 monthly cloud frequency variables.

Results

Cloud Cover Mapping

Custom mapping of daily cloud cover resulted in climatology products quite different from those generated from available cloud mask products. As described in Chapter 2, the differences between the custom mapping of AVHRR clouds (Figure 9) and the Long Term Data Record (LTDR) cloud masking (Figure 10) for March of 1982-1999 mostly relate to the classification of snow as cloud in the LTDR product. Resolving the confusion between snow and cloud in the custom product is critical, as only with consistent discrimination is it evident that cloud frequency is particularly high at mid-elevations over the southern (main) part of the sequoia distribution, especially in the early spring months. In fact, the sequoia belt is the cloudiest area of California in the months of March and April (50-70% frequency), whether the data come from a custom classification of AVHRR for 1982-1999 (March; Figure 9) or MODIS for 2003-2011 (April; Figure 11). Conversely, cloud cover is not particularly high over the sequoia belt (or adjacent highlands) in the summer months (e.g., ~10% in August, Figure 12), when summer thunderstorms would ostensibly provide water. [Precipitation data from climate stations and PRISM maps indicate that even the highest peaks of the southern Sierra receive less than 3 cm of precipitation in both July and August – for a total far less than 10% of their annual precipitation (e.g., http://www.wrcc.dri.edu/pcpn/west.frac.julaug.gif) and less than higher areas of the central and northern Sierra. Nor is there evidence of consistent snowmelt availability in the mid-summer months of July and August in river hydrographs for the Kaweah and Tule Rivers – two major watersheds entirely within the sequoia region (www.epa.gov/region9/water/wetlands/tulare-hydrology/tulare-fullreport.pdf).] The seasonal trajectories of monthly cloud frequency within the sequoia belt (buffered sequoia coverage) for both Terra and Aqua, in comparison with corresponding values for coast redwoods, are shown in Figure 13. It is noteworthy that cloud frequency in the sequoia region tracks that in the Redwood region pretty closely, and is higher on April and May afternoons. A relative downtick in MODIS data for March may be anomalous, as this was the cloudiest period for sequoia in AVHRR data (1982-1999).

Figure 14 focuses on the core of the distribution and demonstrates the close correspondence between afternoon cloud frequency in April and the distribution of the groves - from the relatively isolated McKinley Grove in the north through the Kings Canyon and Sequoia National Park groves south to the southernmost grove of Deer Creek. The MODIS classification is able to capture detail at the scale of individual large NW-SE ridges trending away from the higher elevation Sierra crest. Clouds tend to pool on the west/southwest sides of these large ridges, and this may help explain the surprisingly high frequency of sequoia groves on west and southwest slopes (Hartesveldt *et al*, 1975). There are several groves that don't correspond to the highest levels of cloud cover – notably those on the north-facing slopes of Kings Canyon (south wall) and the Freeman Creek and Cunningham groves east of the Great Western Divide to the south. (Other groves that cross to the east of the Western Divide near the southern limit do have high cloud frequency – here, this spur of the main Sierra Nevada, also called the Greenhorn Mountains, are low enough to allow frequent cloud passage over the ridge. The clouds often approach the Freeman Creek grove as well [webcam observations, http://ssgic.cr.usgs.gov/dashboards/WebCam.htm, see Bald Mtn WebCam #1] but not with sufficient frequency to register strongly on the cloud frequency maps.) It should be noted that the rugged canyons of the

Kings (Middle and South Fork) and Kern Rivers, as well as the steep eastern escarpment of the Sierra Nevada (paralleling, but just to the west of, Highway 395), are errantly mapped as having high cloud frequency in the custom product due to confusion with mixed snow/rock/vegetation. Portions of the Owens Lake Bed are also errantly classified as cloudy.

The effect of the spatial filtering for "Sequoia Cloud Formation" (SCF) days on the April MODIS Aqua cloud mask is seen in the difference between Figure 14 (no spatial filter applied) and Figure 15 (filter applied). Figure 16 plots the frequency of SCF by month within the sequoia belt (buffered coverage of the sequoia groves) as detected by this technique, in comparison to overall cloud frequency by month within the sequoia belt. April has the highest frequency, at about 16%, while August has the lowest frequency, at about 4%. These are likely an underestimate and overestimate, respectively, due to omission and commission error. This determination of an overestimate for summer months is based, in part, on evaluation of the resulting monthly patterns [Appendix 3.A.1; see Appendix 3.A.2. for annual patterns]. August, and other summer months, capture a few days when general high cloud cover happened to fall disproportionately over the sequoia belt (tricking the spatial filtering algorithm). SCF in non-summer months, on the other hand, is more likely to miss days due to the more frequent general cloud cover during this period, especially across northern California. Figure 17 shows the computed SCF frequency by year within the sequoia belt, again in comparison to overall cloud frequency within the belt. The consistency at around 7% is likely a slight to moderate underestimate for aforementioned reasons, but the likelihood of an interannual bias is slight, and indicates that the formation appears steady over time. Interannual variability computations support this consistency: standard deviations of monthly cloud cover across years is low compared to adjacent areas of the Sierra Nevada and southern California, and the coefficients of variation are among the lowest in California.

Principal component analyses (Table 2 for loadings from first 6 principal components on MODIS Aqua monthly cloud frequency for California and Oregon) reveal that the sequoia belt is not just an area of high frequency of cloud cover (1st P.C. with similar weights on fall through spring frequencies; image looks like the overall frequency product), but is also one of the dominant modes of intra-annual variability in California. While the 2nd P.C. captures the winter (Dec & Jan positive loadings) vs. summer (June-Sept negative loadings) contrast, the strong contrast between a spring peak and a winter peak in the 3rd P.C. (large negative weights in order of April, May, March, and February and large positive weights on January and December) captures the sequoia region with low values (due to negative loadings on spring months). Alternatively, the sequoia region has relatively high values in the 4th P.C due to positive weights on February and March (and negative weights on May and November - apparently getting at some contrast with far northeastern California and much of Oregon). These results are evident in a color composite of the 2nd through 4th P.C.'s (Figure 18). The sequoia belt turns up as orange – what looks like a sort of extreme of the California Floristic Province (Stebbins and Major, 1965) foothills - due to its very low blue values (4-2-3 RGB puts the 3rd P.C. in the blue channel) and high red values (4th P.C in red). The orange color indicates that the sequoia belt also has moderate values in the 2nd P.C. (high mid-winter and low in summer). The 4th P.C. includes positive weights on summer (July and August) as well, which results in a strange mix of the spring-peaking areas with monsoonal summer-peaking areas of southern California. These higher P.C.'s begin to make less clear sense, and the percentage of variance explained drops

quickly as well. The first seven P.C.'s appear to have some spatial coherence, with much less in the last five.

Statistical Analyses and Distribution Modeling

Statistical analyses reveal the strong association between cloud cover frequency and sequoia distribution. They are also potentially useful for disentangling other climatic factors that might be involved in sequoia distribution. However, species distribution modeling is sensitive to many choices, including the study area extent. An initial extent of California and Oregon was explored to help ascertain any factors that might be responsible for large scale patterns of the distribution (e.g., spring cloud cover is high in the Pacific Northwest, but there are no sequoia there), but this large extent caused variables with broad geographic trends (i.e., large autocorrelation range) to appear unusually important. For example, in the associated classification tree (see Figure 8), elevation and minimum (winter) monthly vapor pressure deficit (vpdm_mi in figure) simply allow all low elevation areas (elev<1518 meters) and most of Oregon, northern California, and high elevation areas to essentially be removed from consideration. It seems unlikely that those particular variables were chosen because of a very direct impact on sequoia distribution. Once these broad areas are essentially removed by the model, April cloud cover frequency is found to be important, and has the greatest reduction in deviance (indicated by length of descending "branches") of any of the climate variables. Interestingly, a relatively low level of August cloud cover frequency is the last split to separate remaining non-sequoia areas from sequoia areas. Yet high summer cloud cover is not evident within the sequoia groves – it is more common at the highest elevations. Whether this last rule is ecologically meaningful, or just a way of removing some of the lower elevation areas with high April cloud cover, is not clear.

Removing all but the Sierra Nevada ecoregion from the spatial extent of modeling consideration allows the highlighting of variables that are more locally correlated with sequoia distribution. Figure 9 shows that April cloud frequency is now the first split, still with a very high reduction in deviance, followed by many other monthly cloud cover variables (although not always with sequoia corresponding to high values – in fact, the winter and summer cloud cover variables may be serving as proxies for other elevation-related factors). It is noteworthy that another summer month cloud frequency (this time July) shows up as an important final split, with low values separating out areas without sequoia.

Maxent models facilitate the evaluation of the relationship between different predictor variables and sequoia distribution. In modeling the distribution of sequoia at the extent of California and Oregon, Maxent shows similar trends to the classification tree, with April cloud frequency an important, but slightly secondary, variable. However, in confining modeling to the Sierra Nevada ecoregion, April cloud frequency is clearly the most important predictor (Table 3). Among 45 climate variables, it contributes to almost half of the model's explanatory power. [This did not change in several model runs with subtle variations in Maxent's modeling choices; the following comments generally refer to the model associated with Table 3.] It performs twice as well as most variables when used alone (i.e., highest solitary training gain). Only several other variables that are heavily correlated with April cloud frequency, such as February, May, and October cloud frequency, along with elevation or variables correlated with elevation (minimum temperatures and vapor pressure), consistently perform more than half as well as April cloud

frequency in this sort of simple bivariate analysis. While collinearity among the cloud frequency variables may impact the ability to relate Maxent's results to the relative importance of each month's cloud frequency, Maxent does not completely break down due to the collinearity - it is able to tease out the subtle differences that suggest that April cloud frequency is the most strongly related to sequoia distribution.

Collinearity amongst the many climate variables explains why there are only negligible drops in training gain when any one variable is alternately removed in a jackknife approach – even April cloud frequency. It also likely explains why most of the cloud frequency variables do not contribute largely to the full model, despite many having high solitary training gain values – most of the cloud frequencies are heavily correlated with April cloud frequency within the Sierra Nevada (monthly correlations with April: Jan:0.7, Feb: 0.93, Mar:0.94, May:0.94, Jun:0.81, Jul:0.37, Aug:0.56, Sep:0.74, Oct:0.85, Nov:0.83, Dec:0.78). Somewhat surprisingly, removing only November cloud frequency (also quite correlated with April) resulted in the largest drop in model gain of any predictor variable, albeit a small one. On its own it is not considered a strong predictor, nor is it a large contributor to the full model (0.4% contribution). It must be contributing some unique information, however small. Maxent's predictor variable response curves that are conditional upon other variables being held at their mean suggest (but do not ensure, as these do not account for interactions) that November cloud frequency has an inverse relationship with sequoia distribution, after accounting for April cloud frequency's positive and strong relationship. November cloud frequency also has reasonably high permutation importance (8%) in the full model, indicating that the model is sensitive to randomly changing (permuting) its values in presence and background sites. Minimum monthly precipitation frequency has similar behavior: low solitary training gain and a somewhat low contribution to the full model (3.5%), with a negative relationship with Sequoia, but very large permutation importance. However, these permutation importance values fluctuate among slight variations in model runs much more than the percent contribution results, suggesting that the latter are more robust and informative.

Maxent also provides response curves for each predictor variable that are independent of other variables in the model. These curves (Figure 21) are not sensitive to multicollinearity. The response curves for February through May show the importance of high cloud cover frequency at this time: the responses are monotonically increasing to the highest levels of cloud frequency and peak (logistic?) responses are also higher. On the other hand, sequoia groves do not correspond to highest cloud cover levels in other months of the year; for example, in the summer months (July-Sept) the highest cloud frequencies are at elevations well above the sequoia belt and so the response curves peak at the much lower values common at middle elevations. The response values (peak height) in the summer months are also not as high as they are for the spring months. Similarly, the late fall and early winter (Nov-Jan) have peak responses at moderate cloud frequencies, with lower peak response values. In these months, the highest cloud frequencies are at elevations below the sequoia belt or in the far northern Sierra and sequoia groves only have moderate cloud frequencies.

Another statistical approach, logistic regression, has slightly different performance when all predictor variables are simultaneously evaluated, due to its greater sensitivity to multicollinearity. With most of the monthly cloud cover frequencies (especially spring) highly

correlated, the expected spring months are not found to be especially important when all 45 climate variables are included. In fact, many of the other climate variables are heavily correlated as well. However, confining a logistic regression analysis to just a few simple climate variables (April cloud frequency, annual precipitation, maximum monthly temperature, and minimum monthly temperature) reveals that April cloud frequency is by far the most important ($p<<.001$ -> actually, $3.77*10^{-11}$: i.e., very significant) of the four (Table 4). Maximum temperature is also found to be significant in this particular model, and its negative slope indicates that colder, possibly higher elevation areas of the high April cloud frequency are more likely to have sequoia (perhaps where the clouds are near the surface terrain?). This may be analogous to the selection of summer cloud frequency values by the classification tree. Interestingly, given the Maxent results for April and November cloud frequency, forward stepwise logistic regression modeling finds a two-variable model of just April cloud frequency and November cloud frequency to be a much more parsimonious (lower Akaike's Information Criterion, AIC (Akaike, 1974)) model than this four variable model. November cloud frequency may be somewhat interchangeable with monthly maximum temperature in these particular models, as they both have negative slopes in the model resulting in comparable impacts, in terms of weighing higher elevations (due to lower cloud cover) more heavily (although, notably, both perform better with April cloud cover, in terms of AIC, than does elevation). [1]

Discussion

Cloud Cover Frequency

The sequoia belt is among the cloudiest areas of California over the course of a year, and is the cloudiest area of California in March and April, particularly in the afternoon. Clouds form specifically over the sequoia belt at any time of year, but with less frequency in the heart of the summer (i.e., July through September), when clouds tend to form at higher elevations, and to some extent the middle of winter (i.e., December and January), when clouds tend to form at slightly lower elevations – likely related to the Central Valley Tule Fog formation, as discussed below. Sequoia-specific cloud formation (SCF) shows great interannual consistency and the belt has particularly low monthly coefficients of variation in cloud cover frequency across years. April is the peak month for a sequoia-specific cloud formation (SCF) and April cloud frequency is consistently selected by various statistical approaches (i.e., Maxent, classification tree, and logistic regression) as the variable most strongly related to giant sequoia distribution, particularly when the analysis is confined to the Sierra Nevada. Yet April cloud frequency is also heavily correlated with other months so it is problematic to absolutely differentiate April from other months. It is not clear whether April cloud cover is particularly critical to sequoia distribution, or whether this is somewhat coincidental, as it happens to be the time of year when the cloud formation is slightly more restricted to areas directly over the elevations at which sequoia tend to grow.

[1] The model with the lowest AIC resulting from continuing the forward stepwise approach has nine more variables for a total of 11, 4 of which are monthly cloud frequency variables (April, positive slope; November, negative; February, negative; October, positive). This model explains 67% of the null deviance. It's likely overfit but may also demonstrate the multivariate complexity of the controls on the distribution. Backward stepwise logistic regression results in a more complex model with even lower AIC, but it is even more likely to be overfit. Some type of cross-validation would be required to demonstrate the overfitting, but sequoia is too restricted in its distribution to provide a relatively independent validation sample.

Cloud cover modifies local climate in a variety of ways (Sklenár *et al*, 2008), all of which could be factors in its having a relatively direct control on the distribution of giant sequoia. Clouds can moderate temperatures: cooling during the day, but also reducing cold at night. The reduction in radiation, daytime temperature, and vapor pressure deficits reduces evapotranspiration and sublimation and delays snowmelt, ameliorating drought stress. All of these could be especially helpful at the seedling stage, when the plant is particularly sensitive to drought conditions, especially with post-fire clearing and exposure. They could also be especially helpful during unusually dry years. (Many of these factors were included in distribution models but their importance might be missed if they are not mapped well, or if maps fail to capture important temporal aspects.) Direct input of fog water is also possible (Dawson, 1998; Burgess and Dawson, 2004) but cloud base height was not a focus of this study due to difficulties in measuring remotely. Web cam observations reveal that these clouds are generally fairly stratiform (Figure 22); the eastern edge of formation, at a minimum, can be assumed to be intersecting mountain slopes and forming a fog. Cloud cover arriving in the afternoon might be particularly beneficial, as plants can take advantage of morning sun when temperatures, and thus vapor pressure deficits, are typically lower, and gain benefits of clouds when water demand is typically higher. Cloud cover also increases the proportion of diffuse to direct light, which can increase photosynthetic light use efficiency (Rocha *et al*, 2004; Alton *et al*, 2007) and water use efficiency, particularly during periods of low soil moisture (Rocha *et al*, 2004). Finally, clouds may simply alter the competitive balance among species, allowing those that perform better under these conditions to outcompete species that perform more poorly under these conditions (see Species Distributions section below).

A more indirect role for cloud cover would relate to its correlation with other climatic variables that might more directly control the distribution. These would include high humidity and low vapor pressure deficit near the surface (of course, these aren't just statistically correlated with cloud cover as there is a direct link at 100% humidity and zero VPD at cloud level), and perhaps historically mesic (higher precipitation?) climates, of which the clouds might be a sort of relict. Humidity and vapor pressure did not tend to have high importance as predictors in our statistical analysis, but this may be due to the fact that, as interpolations of sparse climate station data, they were not mapped as accurately, with fine spatial detail, as cloud cover. (Daymet data were used for humidity due to their availability at finer resolution – PRISM 800 meter humidity data were not freely available.)

Humidity

The particular image of cloud formation in Figure 7 is unique in the sense that the clouds are forming in mid-summer – but on a relatively cool day. This kind of pattern is not unusual, on the other hand, in the spring months. The fact that it can also occur on rare cold days in the summer suggests that humid air is often pooling up against the abrupt escarpment of the western slope of the Sierra in this region – perhaps throughout the year – and that only when this coincides with periodic cooler air (often associated with storm front passage) is the humidity obvious in the form of a cloud. Frequently higher humidity on the west slope of the southern Sierra is hinted at by a variety of sources (Chapter 4), even in the summer, but a lack of climate stations in the region likely contributes to a lack of accuracy and detail in mapping the spatial

pattern of this humidity (Feld *et al*, 2013), and possibly the lack of apparent importance in giant sequoia distribution modeling.

The fact that a few sequoia groves occur well outside of the sequoia cloud belt (e.g., Placer County Grove, and both groves of Calaveras Big Trees) suggests that high cloud cover is not an absolute requirement in all locations. Other groves (e.g., those on the south wall of the South Fork of the Kings River canyon and those just east of the Greenhorn Mountain divide) fall just outside of the main belt of highest cloud frequency, suggesting that there may be an important variable that is very highly, but not perfectly, correlated with spring cloud frequency. Surface relative humidity (and low VPD) is a good candidate, and many of these outlier groves may be subject to frequently high humidity but just outside of the zone of frequent cloud formation.

Of course, the perceived climate is an integration of many factors (Woodward, 1987; O'Brien *et al*, 2000), and sequoia distribution could just be a question of having an appropriate balance of all the other main factors (e.g., temperature, precipitation, wind, hydrology, and soil), where cloud cover or humidity tilts the balance in many cases. Alternatively, historic legacies could be invoked: perhaps these outlier areas are relicts with a cloudier, more humid history, and are holding on in particularly amenable sites or are on their way out. (This latter argument might be the case for the northernmost, Placer County grove, which has very few trees and no reproduction, but not the case for the next-most northernmost Calaveras groves, which are doing well (Cook, 1942); www.nps.gov/history/history/online_books/cook/groves.htm). Whatever the case might be, the existence of outlier groves does not refute the potential importance of cloud cover or humidity to sequoia survival in the majority of its groves.

Endogenous vs. Exogenous Moisture

What about extending Muir's argument that the sequoias are creating the mesic environment – perhaps their transpiration is leading to the higher humidity and thus cloud formation? Several lines of evidence argue against this "endogenous" explanation of a local source of moisture. Evapotranspiration at these elevations is still quite low in early spring when clouds are most frequent – generally about 1-2 mm/day at most, while it is higher in the adjacent lower elevation foothills (Goulden *et al*, 2012). [See Chapter 2 for related discussion involving satellite and meteorological tower data.] Furthermore, related patterns of cloud development are not always entirely restricted to the sequoia belt – the clouds have a tendency to vary in their formation over elevation according to temperatures, generally occurring over progressively higher elevations from winter through late spring, from elevations well below the sequoia belt, to elevations just above. They also tend to move up latitudinally and in elevation over the course of a day, with higher frequencies over the more southern Tehachapi Mountains early in the day (e.g., Figure 23) to higher frequencies over the sequoia belt later in the day (e.g., Figure 24). Subtracting average morning cloud cover (i.e., MODIS Terra) from average afternoon cloud cover (i.e., MODIS Aqua) for the months of March (Figure 25) and April (Figure 26) confirms that the cloud cover over the sequoia belt regularly builds up in the afternoon. Bright areas of these figures indicate locations in which there is greater cloud cover in the afternoon than in the morning. Dark areas indicate locations in which there is greater cloud cover in the morning than in the afternoon – such as the daytime receding fog of the southern California coast in these spring months. (Alternatively, and unfortunately, dark areas over the eastern High Sierra and

White Mountains indicate areas where the custom processing of MODIS Terra, in particular, had a problem with snow and desert/rock mixes classified as cloud.) The Tehachapi Mountains likely do not show up as a dark region of decreased cloud cover despite the aforementioned cloud movement because they can also occasionally be an area of cloud build-up over the course of the afternoon, just as for the sequoia belt. The cloud movement and the cloud build-up have a net canceling effect.

The movement up in elevation over the course of a day may also be evidence against an endogenous source of moisture. This movement is common throughout the year (and might be related to the convective boundary layer height – see Chapter 4), but especially noticeable in the winter, when the low elevation manifestation may be the obvious Tule fog blanketing the Central Valley floor. On some days it appears that this Tule fog layer may transition into clouds over the sequoia belt region (see Figure 27 and Figure 28; though there is no certainty that the clouds in Figure 26 over the Central Valley sit at the surface). These two different manifestations of "fog" (stratus that may or may not intercept the ground) may be related in the sense that they tend to follow the passage of storm fronts (the moisture source), are associated with low-level convergence and low winds as a result of topographic blockage by the Sierra Nevada, and capped by an inversion layer. They are rarely, if ever, seen at the same time, due to the vastly different boundary layer heights associated with the two events. See Chapter 4 for further discussion.

Amelioration of Summer Drought?

It may seem that if cloud cover were important to sequoia distribution, it would be in terms of relieving summer drought-stress, so that the cloud cover should show up in the drier summer months to be of benefit. But cloud cover does not have to be a summer-drought reliever to be important to sequoia distribution - an explanation that operates across summer months is unnecessary. Any difference in climate at any time of year among locations is sufficient to alter the competitive balance among species. Plant distributions are not just a matter of survival - plants do not simply grow where they do not die – something akin to their fundamental niche (Hutchinson, 1959). They have to outcompete other plants, among other things, resulting in their realized niche (Connell, 1961). Different climate regimes favor different plants, and cloud cover (or humidity) at any time of year is part of the climate regime.

Soils, Hydrology, and Water Balance

Sequoia is generally found in unglaciated regions with soils of granitic origin but this is not a sufficient explanation for their particular pattern of distribution - there is some variation in parent material and soil type among groves (Harvey *et al*, 1980; Elliott-Fisk *et al*, 1996). It is true that only southward from Yosemite is the west slope of the Sierra almost entirely underlain by granite, but this still does not explain why sequoia becomes much less common north of the Kings River. Nor does it explain why the highest elevation, lowest elevation, tallest, biggest by volume, and biggest groves of sequoia are all in the Kaweah River drainage. Sequoia is clearly faring best near the center of the core of its distribution. This sort of Gaussian (normally distributed around its center) behavior does not suggest a strong response to an edaphic factor. While hydrology and soils are likely to play an important role in sequoia distribution, their importance is mostly manifested at finer scales, likely explaining why forests of giant sequoia are not more continuously distributed within the core area. Unfortunately, accurate data on these

variables at fine scales tends to be lacking. The species distribution analysis performed here was a relatively coarse-scaled climate modeling effort that did not consider hydrology or soil and its potential effects on water storage and water deficit, among other factors. One study (Lutz *et al*, 2010) found that giant sequoia had the highest (modeled) average actual evapotranspiration among the tree species of Yosemite National Park, and an associated relatively low water deficit (given that deficit is the difference between potential evapotranspiration and actual evapotranspiration). This resulted, in part, from their use of a soil map that gave soils of sequoia the highest average water holding capacity. While this is consistent with the finding (Goulden *et al*, 2012) that the mid-elevation forest belt of the west slope of the Sierra Nevada is likely able to sustain high productivity due to rooting into deep soils, there is a potential problem of circularity in using a soil map to predict sequoia distribution. Given that soil formation is a product of climate, organisms, relief, parent material, and time (i.e., CLORPT, sensu Jenny (Jenny, 1961)), the soils of sequoia groves may be deep because of the productive climate and the associated sequoia growth, and not the other way around. Restricting the analysis to climate variables (or, at least, standardizing soil depth) avoids this sort of circularity. Water balance is likely an important factor in sequoia distribution, but it would be useful to determine whether this is the case irrespective of soil depth (i.e., without trying to include soil depth in a water balance model).

Wind

Relatively complex water balance models (e.g., Penman-Monteith (Monteith, 1965)) also utilize information on wind, as it can play a large role in the evapotranspiration process. Two natural resources energy companies, 3Tier (http://www.3tier.com/en/support/resource-maps/) and AWS Truewind (now AWS Truepower, https://www.awstruepower.com/products/maps-and-resource-data/), have produced maps of average winds at 80 meter height (Figure 29 and Figure 30, respectively), that indicate that the eastern San Joaquin Valley and the foot of the sequoia belt are among the least windy portions of the country. A relative lack of wind at 80 meters would reduce the water stress felt by tall trees, in particular. It also may be a factor in allowing moist air and clouds to pool up near the sequoia belt, rather than being dispersed over the Sierra Nevada. The transition of the Sierra Nevada west slope to a north/south orientation here, rather than a north-northeast/south-southwest orientation, may also be a factor, especially due to the prevailing northwesterly winds in the San Joaquin Valley from spring through fall.

Unfortunately, water vapor may not be the only atmospheric constituent pooling in the region due to relatively low winds. It may also be the reason air pollution is extremely bad in the San Joaquin Valley, and why other tree species in the vicinity of the sequoia belt are suffering from ozone chlorosis (Barbour and Minnich, 1988). Maps of chlorotic damage in the Sierra correspond to the sequoia belt (Panek *et al*, 2013), although generally peaking at a lower elevation. In any case, the lack of wind may be a large factor in the distribution of sequoia, but (just as for humidity) wind may not be mapped at the level of detail necessary to demonstrate its importance beyond the extent that it can be seen to be somewhat correlated with cloud cover frequency.

Species Distributions

The distribution of other plant species may also support the notion that the sequoia belt, or at least the western slope of the southern Sierra, has its own unique climate regime. While the Sierra Nevada is known to have high plant species diversity, it is not particularly high in

endemism. This has been attributed to the lack of topographic barriers with the similar environments of the Coast Ranges (Shevock, 1996). On one hand, this relative lack of endemism is an argument against a unique climate – on the other hand, it's noteworthy that the Sierra Nevada has such a similar climate to the Coast Ranges. Sierra bleeding-heart (*Dicentra nevadensis* (Eastw.)) and purple fairly lantern (*Calochortus amoenus* (E. Greene)) have distributions nearly identical to that of the core distribution of giant sequoia, although at generally higher and lower elevations, respectively. Sierra bladdernut (*Staphylea bolanderi* (A. Gray)) also has a similar distribution, although also at a slightly lower elevation (with outliers in the Shasta Lake region). Alternatively, gray pine (*Pinus sabiniana* (Douglas)) has a long-noted (Griffin and Critchfield, 1972) gap in its distribution that corresponds nearly exactly (although at a lower elevation) to the main belt of sequoia (Figure 31). Researchers (Schwilk and Keeley, 2006) have surmised that this gap is related to rugged topography and lack of adequate fire refugia due to the absence of large river valleys in the area. The rugged topography cannot be denied, and there is probably a connection between the steep escarpment in this portion of the Sierra and the sequoia distribution. However, it seems more likely that the relevance of the steep topography is related to the effect on the compactness of the cloud formation and a variety of other climatic factors, such as humidity and precipitation. That this different topography AND climate regime creates a unique fire regime is likely - it's just difficult to disentangle the relative importance of climate and fire for species distribution. In any case, the pine grows in many fire prone areas without large river valley refugia (e.g., throughout the southern Central Coast Ranges). Yet it does not grow in other humid and cloudy areas, such as those near the coast. The pine may simply be outcompeted by other species in the different climate (and fire) regime in the foothills adjacent to the sequoia belt.

Other species distributions in the vicinity of the sequoia belt are notable, although not as distinctive as the above-mentioned patterns. For example, the Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) Sierra distribution reaches its southern limit near the northern edge of the main sequoia belt, yet it extends farther south in the Coast Ranges; other species share this trend: e.g., Pacific madrone (*Arbutus menziesii* (Pursh)) and tanoak (*Notholithocarpus densiflorus* (Hook. & Arn.) Manos, Cannon, & Oh). Conversely, many other species have the opposite pattern in that they extend their range south along the west slope of the Sierra much farther than they do in the Coast Ranges, for example, Pacific dogwood (*Cornus nuttallii* (Audubon)), western redbud (*Cercis occidentalis* (Torrey)), mock orange (*Philadelphus lewisii* (Pursh)) and western azalea (*Rhododendron occidentale* (Torrey & A. Gray) A. Gray) - all of which have extreme southern California disjuncts (possibly due to summer monsoonal moisture, although general high frequency of cloudiness and rainfall could be factors as well), as well as beaked hazelnut (*Corylus cornuta ssp. californica* (Marsh.)), Sierra plum (*Prunus subcordata* (Benth.)), California buckeye (*Aesculus californica* (Spach) Nutt.), Oregon oak (*Quercus garryana* (Hook.)), and California nutmeg (*Torreya californica* (Torrey)). Something about the environment in the vicinity of the sequoia belt clearly allows some species to flourish and other species to be at a disadvantage. Including cloud cover frequencies in models of these species' distributions could demonstrate how important cloud cover, or relative lack thereof, might be in controlling their distribution.

Conclusions

A major hurdle for early hypotheses on the distribution of giant sequoia was in identifying differences in climate between that of the southern Sierra and that of the northern Sierra, besides the South's expected slightly greater warmth and aridity at a given elevation (factors which would not be expected to favor sequoia). In fact, there appears to be much greater variability to the Sierra climate (for given elevations) than shown by simple climate variables interpolated from a sparse network of climate stations in rugged mountain terrain. Both the climate variables that need to be considered and the spatial pattern of those variables are difficult to understand with limited station data. Satellite data, in particular daily imagery with reasonably high resolution (i.e., MODIS), is revealing on both of these fronts. Unusually high cloud cover frequency currently stands out as the most salient environmental feature of the sequoia belt. But a host of factors suggests that the environment on the west slope of the southern Sierra Nevada in the vicinity of the giant sequoia groves is unique in a variety of ways.

The fact that cloud frequency correlates with the giant sequoia distribution far better than any other climatic variables indicates that it is either a strong driver in the distribution or is heavily correlated with other driving environmental variables that are either not mapped, or not mapped very well. Further research is needed to clarify this. It should involve collection of high resolution information on humidity, radiation, wind, and snowpack (possibly along with precipitation and temperature), with a focus on several sequoia groves just inside and outside of the main zone of high cloud cover frequency. Chapter 4 touches upon some of these topics with the use of an elevation transect of micrometeorological tower measurements from a 55 meter height on half-hour time steps. These provide important insights but do not address the problem of meteorological data scarcity. Despite the needs for more data and further investigation to determine the exact mechanisms involved in the sequoia distribution, the correlation between cloud cover and sequoia is strong enough to suggest that cloud cover is indeed a factor. It may be fair to call the giant sequoia belt the cloud forest of the Sierra Nevada.

References

Akaike, H. (1974) A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, **19**, 716-723.

Alton, P., North, P. & Los, S. (2007) The impact of diffuse sunlight on canopy light-use efficiency, gross photosynthetic product and net ecosystem exchange in three forest biomes. *Global Change Biology*, **13**, 776-787.

Axelrod, D.I. (1959) Late Cenozoic Evolution of the Sierran Bigtree Forest. *Evolution*, **13**, 9-23.

Baldocchi, D. & Waller, E. (2014) Winter fog is decreasing in the fruit growing region of the Central Valley of California. *Geophysical Research Letters*.

Barbour, M.G. & Minnich, R.A. (1988) Califomian Upland Forests and Woodlands. *North American terrestrial vegetation*, 131-164.

Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. (1984) Classification and regression trees. Wadsworth & Brooks. *Monterey, CA*.

Burgess, S. & Dawson, T. (2004) The contribution of fog to the water relations of Sequoia sempervirens (D. Don): foliar uptake and prevention of dehydration. *Plant, Cell & Environment*, **27**, 1023-1034.

Cannon, W. (1901) On the relation of redwoods and fog to the general precipitation in the redwood belt of California. *Torreya*, 137-139.

Connell, J.H. (1961) The influence of interspecific competition and other factors on the distribution of the barnacle Chthamalus stellatus. *Ecology*, **42**, 710-723.

Cook, L.F. (1942) giant sequoias of California.

Daly, C., Neilson, R.P. & Phillips, D.L. (1994) A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology*, **33**, 140-158.

Daly, C. (2006) Guidelines for assessing the suitability of spatial climate data sets. *International Journal of Climatology*, **26**, 707-721.

Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J. & Pasteris, P.P. (2008) Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, **28**, 2031-2064.

Dawson, T.E. (1998) Fog in the California redwood forest: ecosystem inputs and use by plants. *Oecologia*, **117**, 476-485.

Dawson, T.E. (1998) Fog in the California redwood forest: ecosystem inputs and use by plants. *Oecologia*, **117**, 476-485.

Elith*, J., H. Graham*, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G.,

Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M. & E. Zimmermann, N. (2006) Novel methods improve prediction of species? distributions from occurrence data. *Ecography*, **29**, 129-151.

Elliott-Fisk, D.L., Stephens, S.L., Aubert, J.E., Murphy, D. & Schaber, J. (1996) Mediated settlement agreement for Sequoia National Forest, Section B, Giant Sequoia Groves: An evaluation.

Feld, S.I., Cristea, N.C. & Lundquist, J.D. (2013) Representing atmospheric moisture content along mountain slopes: Examination using distributed sensors in the Sierra Nevada, California. *Water Resources Research*.

Gentry, A.H. (1992) Tropical forest biodiversity: distributional patterns and their conservational significance. *Oikos*, 19-28.

Goulden, M., Anderson, R., Bales, R., Kelly, A., Meadows, M. & Winston, G. (2012) Evapotranspiration along an elevation gradient in California's Sierra Nevada. *Journal of Geophysical Research: Biogeosciences (2005–2012)*, **117**.

Griffin, J.R. & Critchfield, W.B. (1972) The distribution of forest trees in California. *USDA Forest Service Research Paper, Pacific Southwest Forest and Range Experiment Station*.

Hartesveldt, R.J., Harvey, H.T., Shellhammer, H.S. & Stecker, R.E. (1975) The giant sequoia of the Sierra Nevada. *The giant sequoia of the Sierra Nevada*.

Harvey, H.T., Shellhammer, H.S. & Stecker, R.E. (1980) Giant sequoia ecology. Fire and reproduction. *Giant sequoia ecology.Fire and reproduction*.

Heidinger, A.K., Evan, A.T., Foster, M.J. & Walther, A. (2012) A naive Bayesian cloud-detection scheme derived from CALIPSO and applied within PATMOS-x. *Journal of Applied Meteorology and Climatology*, **51**, 1129-1144.

Hickman, J.C. (1993) *The Jepson manual: higher plants of California.*, University of California Press.

Hutchinson, G.E. (1959) Homage to Santa Rosalia or why are there so many kinds of animals? *American naturalist*, **93**, 145.

Jenny, H. (1961) Derivation of state factor equations of soils and ecosystems. *Soil Science Society of America Journal*, **25**, 385-388.

Lutz, J.A., van Wagtendonk, J.W. & Franklin, J.F. (2010) Climatic water deficit, tree species ranges, and climate change in Yosemite National Park. *Journal of Biogeography*, **37**, 936-950.

Monteith, J. (1965) Evaporation and environment., **19**, 4.

Muir, J. (1876) *On the post-glacial history of Sequoia gigantea*.

O'Brien, E.M., Field, R. & Whittaker, R.J. (2000) Climatic gradients in woody plant (tree and shrub) diversity: water-energy dynamics, residual variation, and topography. *Oikos*, **89**, 588-600.

Panek, J., Saah, D., Esperanza, A., Bytnerowicz, A., Fraczek, W. & Cisneros, R. (2013) Ozone distribution in remote ecologically vulnerable terrain of the southern Sierra Nevada, CA. *Environmental Pollution*, **182**, 343-356.

Pedelty, J., Devadiga, S., Masuoka, E., Brown, M., Pinzon, J., Tucker, C., Roy, D., Ju, J., Vermote, E. & Prince, S. (2007) Generating a long-term land data record from the AVHRR and MODIS instruments., 1021-1025.

Phillips, S.J., Dudík, M. & Schapire, R.E. (2004) A maximum entropy approach to species distribution modeling., 83.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231-259.

Phillips, S.J. & Dudík, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161-175.

Priestley, C. & Taylor, R. (1972) On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review*, **100**, 81-92.

Ripley, B. (1996) Neural networks and pattern recognition. *Cambridge University*.

Rocha, A.V., Hong-Bing, S., Vogel, C.S., Peter, S.H. & Curtis, P.S. (2004) Photosynthetic and water use efficiency responses to diffuse radiation by an aspen-dominated northern hardwood forest. *Forest Science*, **50**, 793-801.

Rollins, M.G. (2009) LANDFIRE: a nationally consistent vegetation, wildland fire, and fuel assessment. *International Journal of Wildland Fire*, **18**, 235-249.

Rundel, P.W. (1971) Community Structure and Stability in the Giant Sequoia Groves of the Sierra Nevada, California. *American Midland Naturalist*, **85**, 478-492.

Rundel, P.W. (1972) Habitat Restriction in Giant Sequoia: The Environmental Control of Grove Boundaries. *American Midland Naturalist*, **87**, 81-99.

Schiffer, R. & Rossow, W.B. (1983) The International Satellite Cloud Climatology Project(ISCCP)- The first project of the World Climate Research Programme. *American Meteorological Society, Bulletin*, **64**, 779-784.

Schwilk, D.W. & Keeley, J.E. (2006) The role of fire refugia in the distribution of Pinus sabiniana (Pinaceae) in the southern Sierra Nevada. *Madroño*, **53**, 364-372.

Shevock, J.R. (1996) *Status of rare and endemic plants*, Centers for Water and Wildland Resources, University of California.

Sklenár, P., Bendix, J. & Balslev, H. (2008) Cloud frequency correlates to plant species composition in the high Andes of Ecuador. *Basic and Applied Ecology*, **9**, 504-513.

Stebbins, G.L. & Major, J. (1965) Endemism and speciation in the California flora. *Ecological Monographs*, 2-35.

Stephenson, N.L. (1999) Reference conditions for giant sequoia forest restoration: structure, process, and precision. *Ecological Applications*, **9**, 1253-1265.

Sugden, A.M. & Robins, R.J. (1979) Aspects of the ecology of vascular epiphytes in Colombian cloud forests. I. The distribution of the epiphytic flora. *Biotropica*.

Thornton, P.E., Running, S.W. & White, M.A. (1997) Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology*, **190**, 214-251.

Waller, E. (2006) Topoclimatic Controls on the Distribution of Giant Sequoia: A Satellite Perspective. Poster session presented at the 2nd CIRMOUNT-sponsored MTNCLIM conference, MTNCLIM 2006 Conference, September 19-22, 2006, Timberline Lodge, Mt. Hood, Oregon, USA..

Wilson, H.E. (1928) The Lore and the Lure of Sequoia. *Wolfer Printing Co*.

Woodward, F.I. (1987) *Climate and plant distribution*, Cambridge University Press.

Tables

Table 1. Predictor variables for species distribution models

Annual	Source	**Monthly Maxima**	Source	**Monthly Minima**	Source
Precipitation	P	Precipitation	P	Precipitation	P
Precipitation Frequency	D	Precipitation Frequency	D	Precipitation Frequency	D
Mean Maximum Temp	P	Temperature	P	Temperature	P
Mean Minimum Temp	P	Relative Humidity	D	Relative Humidity	D
Temperature Range (Max Mo.-Min Mo.)	P	Vapor Pressure	D	Vapor Pressure	D
Average Relative Humidity	D	Vapor Pressure Deficit	D	Vapor Pressure Deficit	D
Average Vapor Pressure	D	Radiation	D	Radiation	D
Average Radiation	D	PET	D/P	PET	D/P
Potential ET (PET)	D/P	Water Balance	D/P	Water Balance	D/P
PET Range	D/P				
Water Balance (cumulative annual PPT-PET, monthly carryover)	D/P				
Water Balance Range	D/P				
Deficit (cumulative annual PPT-PET,no monthly carryover)	D/P				
Months of Deficit	D/P				
Other					
Monthly Cloud Frequency	C (12)				
Elevation	DEM				

D = Daymet
P = PRISM
D/P = Jointly Derived
C = Custom Derived
(12 separate months)

Table 2. Principal component loadings: MODIS Aqua monthly cloud frequency (California & Oregon).
Values for months with strong loadings in bold (negative = black; positive = dark gray) in the 2nd through 4th components (those P.C.'s included in Figure 18; P.C. 1 strongly correlated with overall frequency).

	Principal Component					
	1	**2**	**3**	**4**	**5**	**6**
January	0.34	**0.41**	**0.53**	0.07	0.27	-0.33
February	0.25	0.16	**-0.20**	**0.62**	-0.10	-0.15
March	0.38	0.12	**-0.24**	**0.29**	-0.14	0.20
April	0.36	0.05	**-0.50**	0.08	0.24	0.07
May	0.36	-0.19	**-0.31**	**-0.50**	0.43	-0.02
June	0.28	**-0.35**	0.01	-0.12	-0.17	-0.70
July	0.06	**-0.48**	**0.24**	**0.31**	0.30	0.16
August	0.12	**-0.39**	**0.25**	**0.23**	0.25	0.23
September	0.17	**-0.32**	**0.17**	0.07	-0.09	0.07
October	0.26	-0.23	0.06	-0.07	-0.57	-0.06
November	0.36	0.06	**0.20**	**-0.27**	-0.36	0.49
December	0.32	**0.28**	**0.29**	-0.16	0.14	0.11
% Variance Explained	**89.7**	**5.8**	**1.8**	**0.7**	**0.5**	**0.4**

Table 3. Maxent variable comparison, sequoia distribution model, Sierra Nevada ecoregion
Solitary gain is an evaluation of an individual variables predictive ability, while percent contribution and permutation importance relate to the final model. The latter reflects the sensitivity of the final model to randomly changing (permuting) the value of the predictor in presence and background sites. Variables are sorted according to their solitary gain, which highlights the strength of the monthly cloud frequencies.

Variable	Solitary Gain	Percent Contribution	Permutation Importance
Cloud frequency, April	1.69	45.2	0.38
Cloud frequency, May	1.18	1.2	0.98
Cloud frequency, February	0.97	0.0	0.09
Mean minimum temperature	0.97	0.3	6.57
Minimum monthly temperature	0.96	0.0	0.00
Cloud frequency, October	0.93	0.2	2.26
Maximum monthly absolute humidity	0.86	0.2	2.29
Cloud frequency, March	0.86	2.9	10.34
Elevation	0.86	13.4	5.30
Annual average absolute humidity	0.82	0.0	0.00
Maximum monthly VPD	0.81	1.6	2.37
Maximum monthly water balance	0.76	0.0	0.00
Maximum monthly precipitation	0.76	0.0	0.00

Annual precipitation	0.76	0.0	0.00
Minimum monthly VPD	0.74	3.8	0.00
Monthly maximum temperature	0.70	0.2	0.56
Cloud frequency, August	0.69	0.5	1.76
Annual water balance range	0.69	1.9	0.36
Mean maximum temperature	0.67	0.0	0.00
Annual precipitation frequency	0.65	10.2	2.45
Annual water balance	0.63	0.0	0.00
Minimum monthly precipitation	0.62	3.1	2.61
Minimum monthly absolute humidity	0.61	0.4	1.81
Cloud frequency, July	0.60	0.3	0.00
Cloud frequency, September	0.59	0.3	2.17
Maximum monthly precipitation frequency	0.51	0.0	0.00
Annual deficit	0.51	0.0	0.00
Annual temperature range	0.51	1.1	2.32
Months of deficit	0.46	5.4	7.19
Annual relative humidity	0.43	0.1	0.10
Minimum monthly precipitation frequency	0.37	3.5	31.67
Minimum monthly water balance	0.32	0.1	0.64
Minimum monthly deficit	0.32	0.0	0.14
Cloud frequency, June	0.31	0.8	3.26
Maximum monthly PET	0.29	0.0	0.00
Cloud frequency, January	0.29	0.0	0.00
Minimum monthly relative humidity	0.18	0.0	0.00
Minimum monthly radiation	0.17	2.1	0.00
Maximum monthly radiation	0.16	0.4	2.54
Minimum monthly PET	0.14	0.1	0.64
Annual PET	0.10	0.1	0.70
Cloud frequency, November	0.10	0.4	8.27
Cloud frequency, December	0.08	0.0	0.00
Annual radiation	0.06	0.1	0.21

Table 4. Four variable logistic regression results in R software, Sierra Nevada extent
April cloud frequency (apr0311) by far the strongest predictor in this simple 4 variable model, with climate variables commonly used in simple models of species distributions (annual precipitation, maximum temperature of the warmest month, minimum temperature of the coldest month).

```
> summary(mod1)

Call:
glm(formula = Outcome ~ apr0311 + ppt_ann + tmaxm_1 + tminm_1,
    family = binomial("logit"), data = sn.ign)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.14652  -0.51465  -0.26215  -0.07783   2.59123

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.336e+00  4.054e+00  -0.329   0.7418
apr0311      1.732e+01  2.620e+00   6.613 3.77e-11 ***
ppt_ann     -8.538e-06  9.488e-06  -0.900   0.3682
tmaxm_1     -2.695e-03  1.132e-03  -2.380   0.0173 *
tminm_1      1.505e-03  1.360e-03   1.107   0.2684
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 268.96  on 255  degrees of freedom
Residual deviance: 170.60  on 251  degrees of freedom
AIC: 180.6

Number of Fisher Scoring iterations: 6
```

Figures

Figure 1. Giant sequoia distribution in California (from Shirley, 1947).
(Sequoia more restricted than its cousin, the coast redwood, which is common in the central and northern Coast Ranges.)

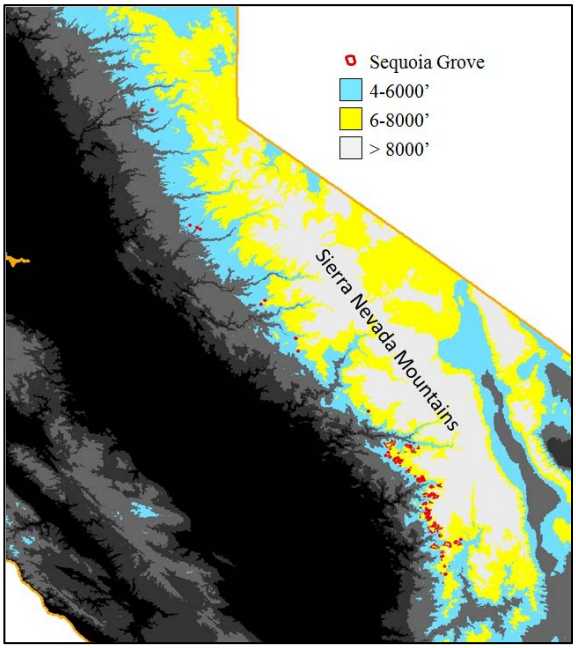


Figure 2. Giant sequoia distribution in relation to elevation belts.
Generally found at higher elevation belt in the south, lower elevations for outlier groves in north. No lack of appropriate elevations in the north.

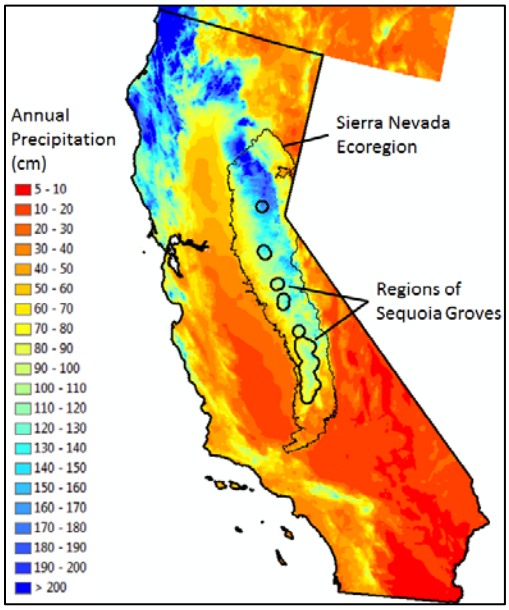


Figure 3. Giant sequoia distribution in relation to annual precipitation.
No lack of adequate precipitation for sequoia in the northern Sierra Nevada.

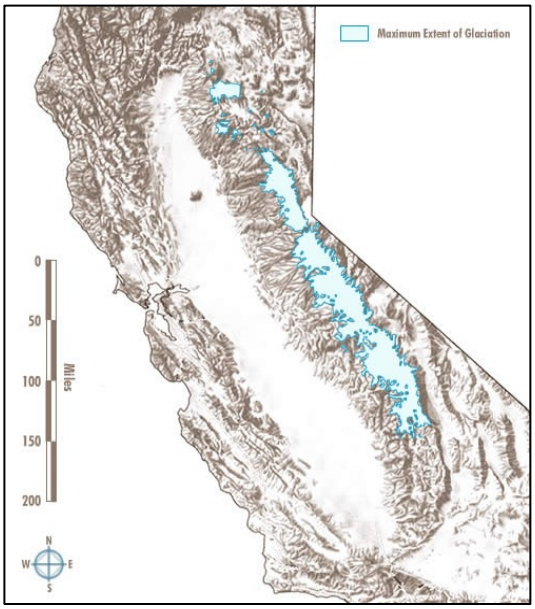


Figure 4. Maximum extent of Pleistocene glaciation in California (from Donley, 1979).
Glaciers not extensive enough to explain absence of sequoia in the northern Sierra Nevada.

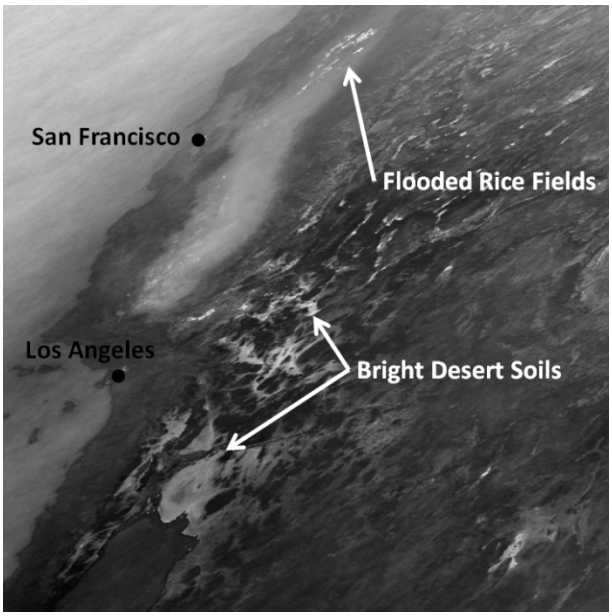


Figure 5. MODIS Terra official cloud mask, January average (2001-2012)
Tile H08,V05 (Sinusoidal Projection). Strong evidence of error related to land cover.

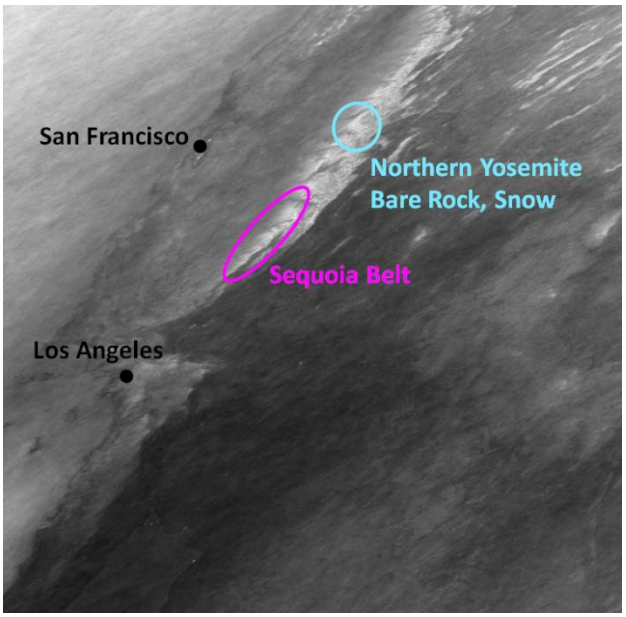


Figure 6. MODIS Terra official cloud mask, April average(2000-2011)
Tile H08,V05 (Sinusoidal Projection). More subtle confusion with bright bare rock and snow mix.

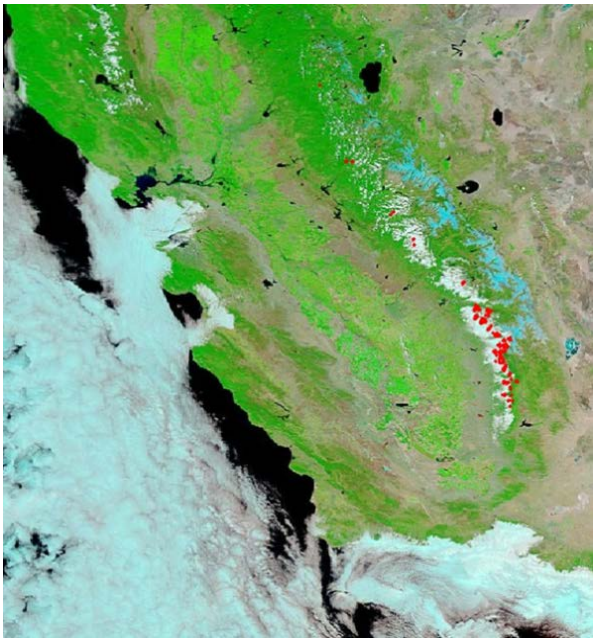


Figure 7. MODIS Aqua (~1:30 PM), 7/14/11, false color composite (RGB=7-2-1).
Giant sequoia distribution highlighted in red. This color composite shows contrast between clouds (white) and snow (cyan).

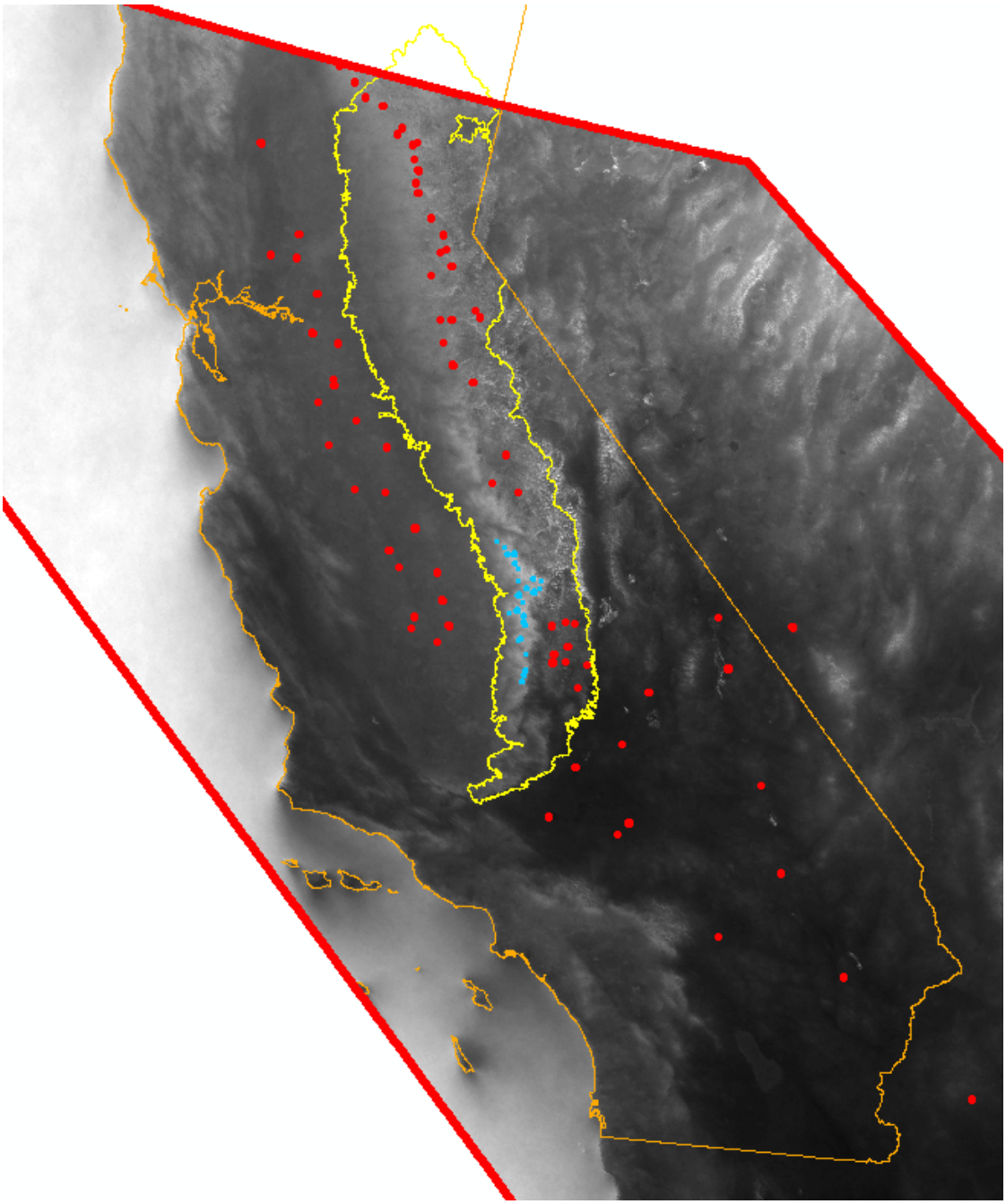


Figure 8. Spatial filtering technique for identifying days of "Sequoia Cloud Formation" (SCF). Backdrop image is annual cloud frequency from MODIS Aqua, 2002-2012, for sinusoidal grid tile H08, V05 and only serves to help with spatial referencing. Spatio-temporal filtering is performed on a daily basis with individual MODIS cloud cover masks. If more of the 25 cyan points were identified as cloudy on a given day than of the 75 surrounding red points, the day was determined to have the SCF. All cloud cover on this day is then retained and contribute to the SCF frequency products in Figure 15 and Appendices 3.A.1 and 3.A.2. This approach has the obvious potential for spatial bias (with sensitivity to point numbers and placement), but was only used as a relative index of approximate SCF frequency for comparing among months and years – NOT for any of the predictive modeling in this study.

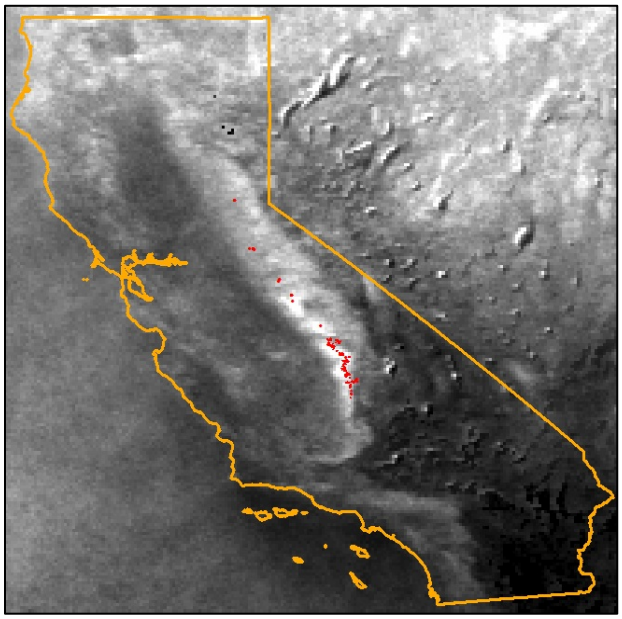


Figure 9. AVHRR March cloud frequency (Custom) with giant sequoia distribution
Sequoia (red) show strong relationship with high cloud frequency in March (1982-1999 average).

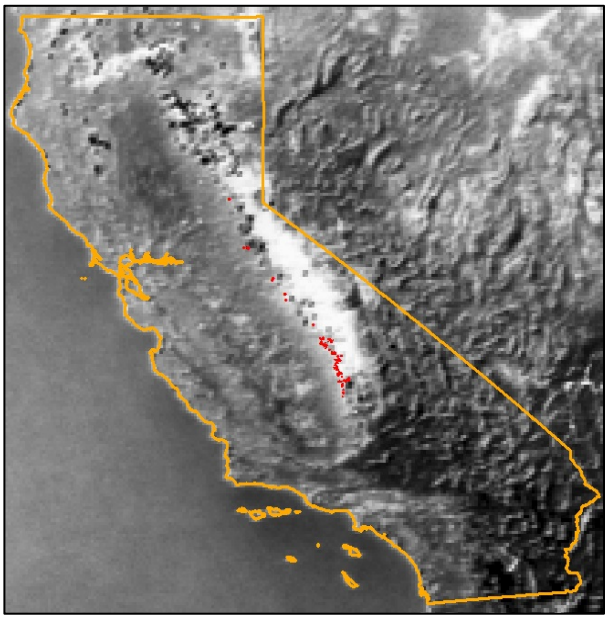


Figure 10. AVHRR March cloud frequency (LTDR) with giant sequoia distribution
Misclassification of snow as cloud obscures relationship between Sequoia (red) and cloud.

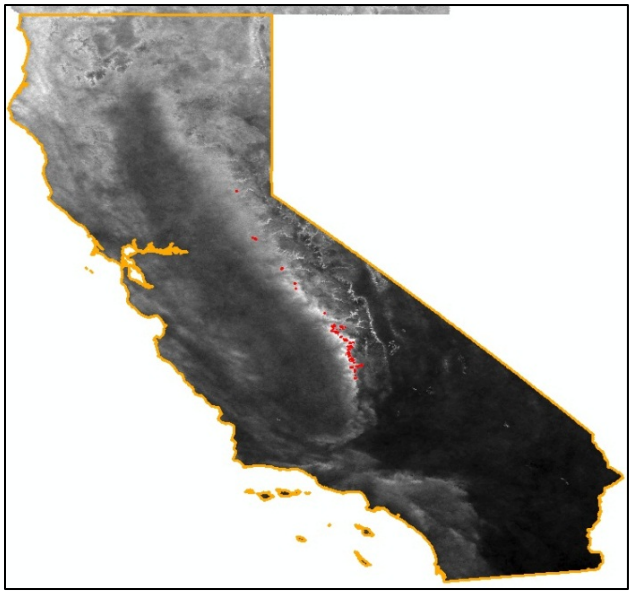


Figure 11. MODIS Aqua April cloud frequency with giant sequoia distribution
Sequoia (red) show strong relationship with high cloud frequency in April (2003-2011 average).

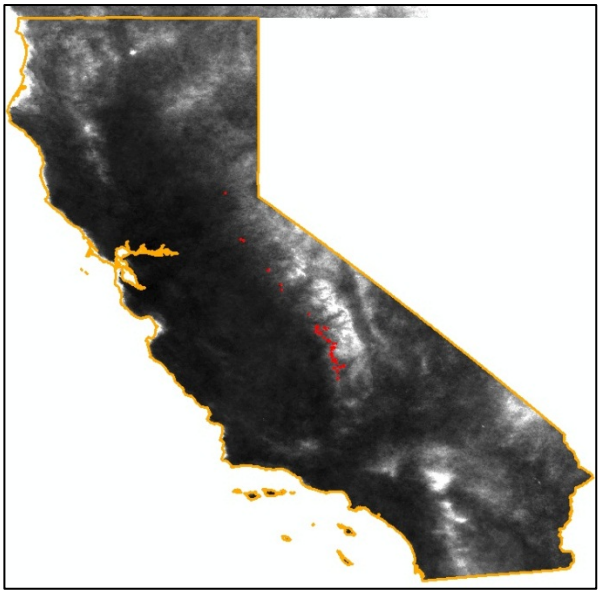


Figure 12. MODIS Aqua August cloud frequency with giant sequoia distribution
Sequoia (red) show weak relationship with high cloud frequency in August (2003-2011).

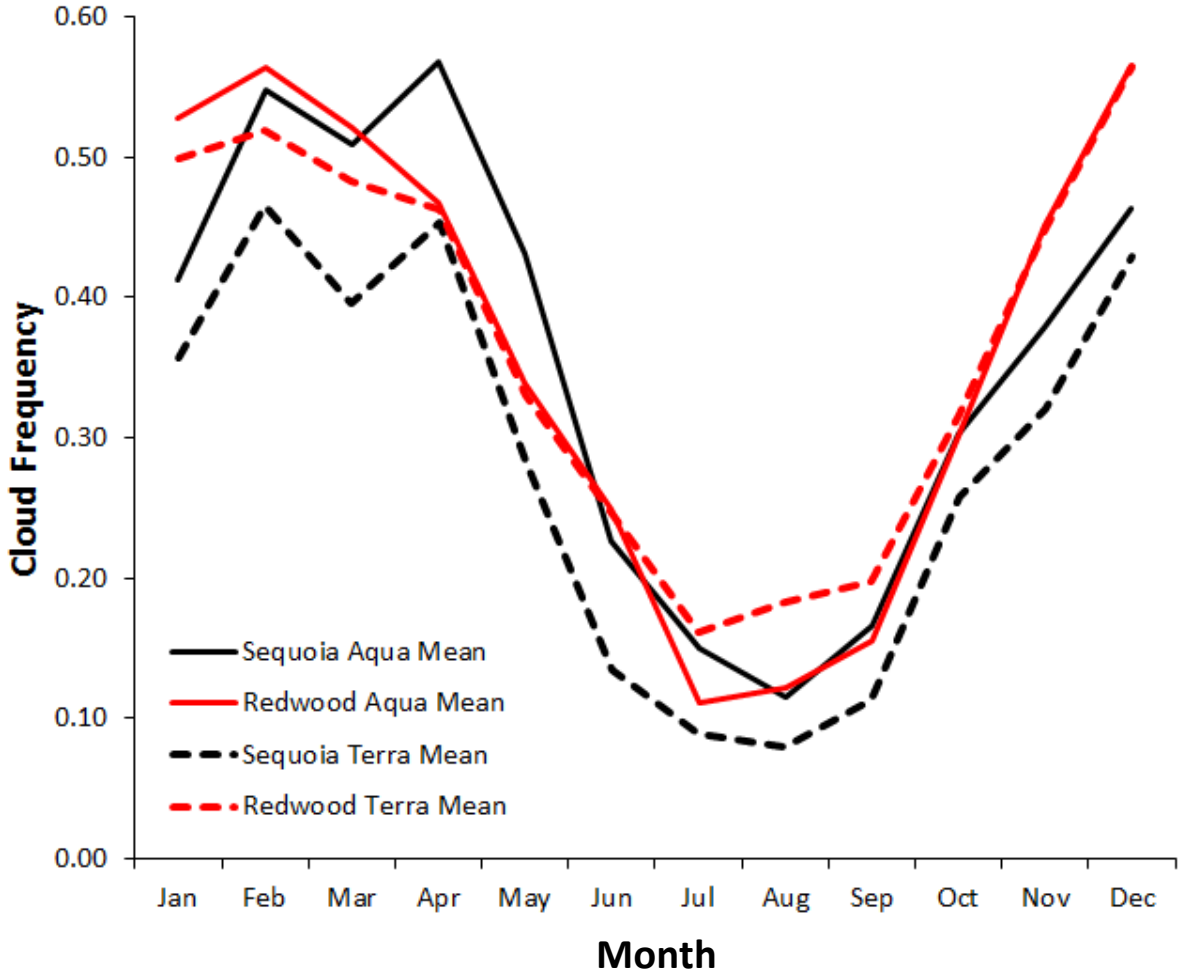


Figure 13. Giant sequoia and coast redwood mean monthly cloud frequency comparison (MODIS Terra and Aqua).
Redwood (red lines) has greater frequency on summer mornings (Terra, ~ 11 AM, dashed) and early- to mid-winter morning and afternoon, but sequoia (black lines) not far behind, and exceeds redwood on spring afternoons (Aqua, ~ 1 PM, solid).

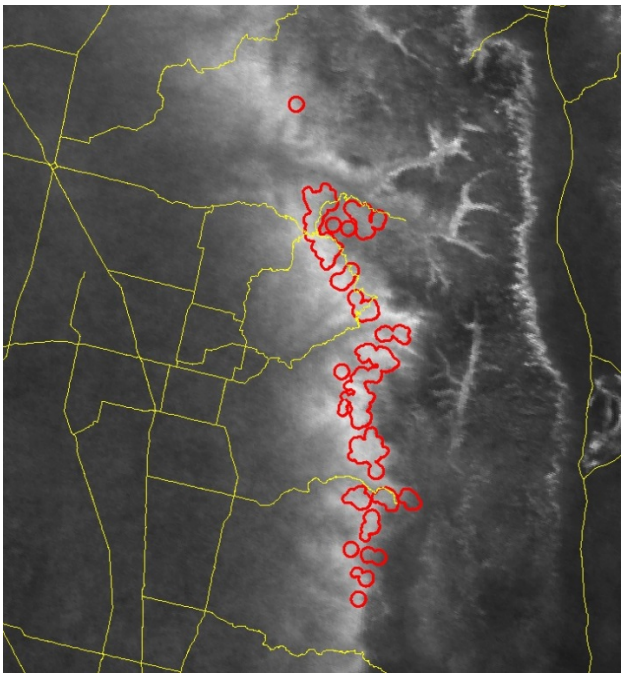


Figure 14. MODIS Aqua Cloud Frequency, April 2003-11 (main sequoia belt).
Giant sequoia (red) show good correspondence with April cloud frequency.

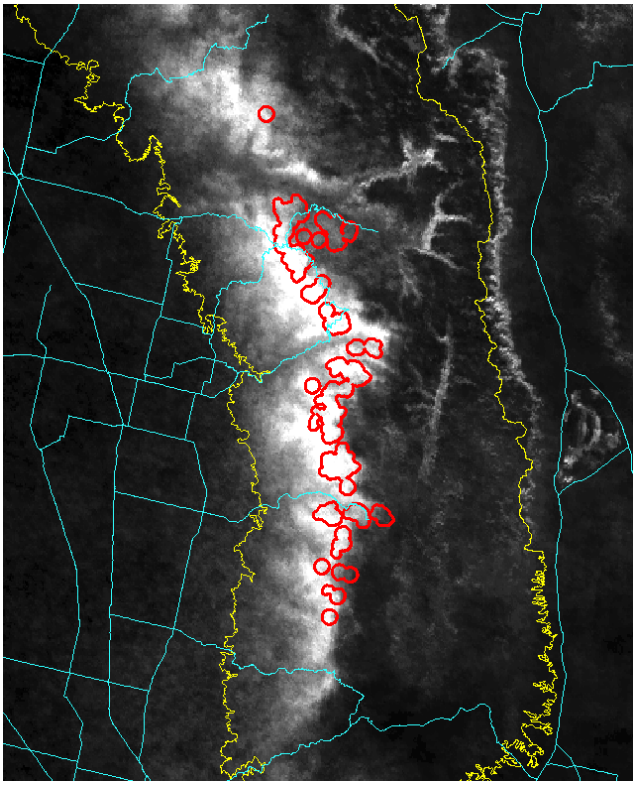


Figure 15. MODIS Aqua "Sequoia Cloud Formation" frequency, April 2003-11
Spatial filtering approach enhances relationship between sequoia (red) and clouds from Figure 14.

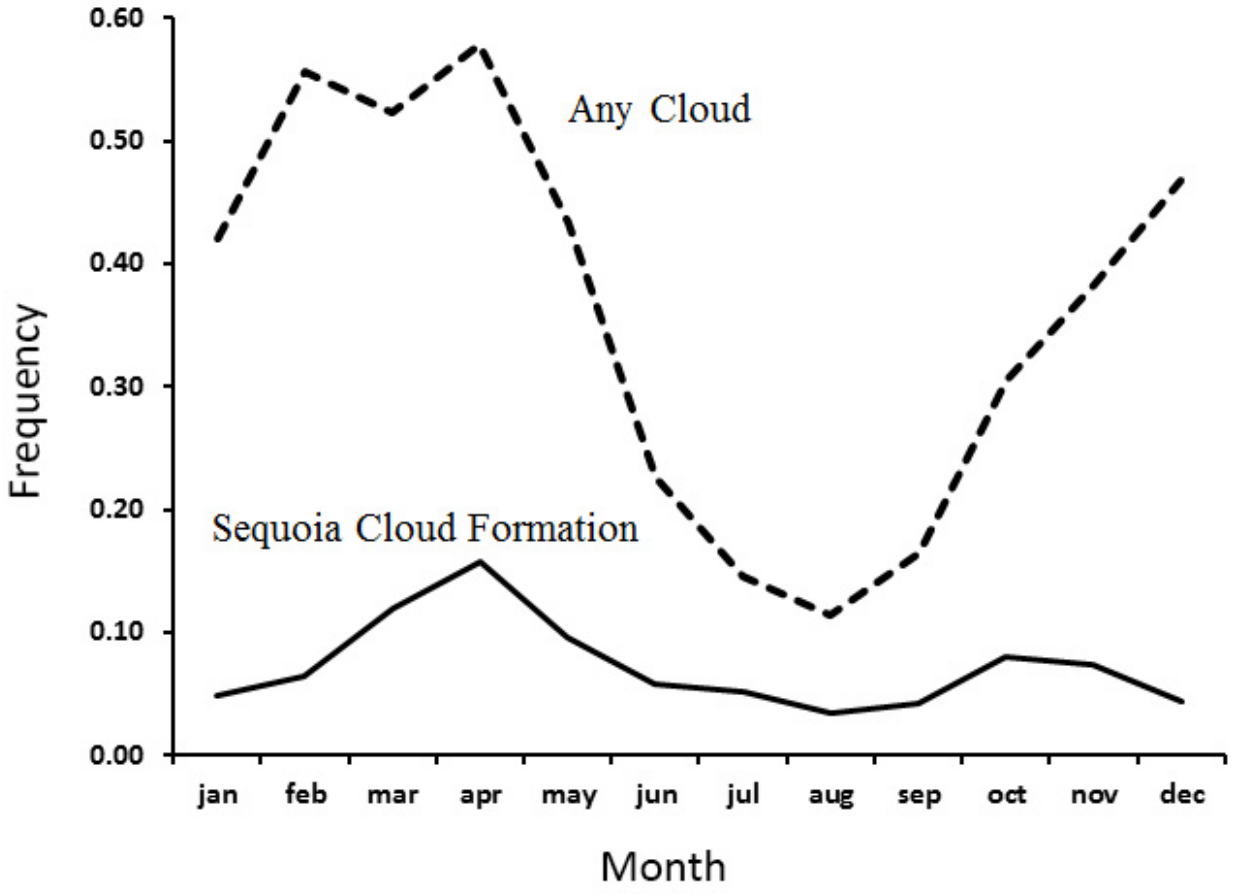


Figure 16. MODIS Aqua monthly cloud frequency (2003-2011) and frequency of "Sequoia Cloud Formation" within sequoia belt.
Both show an obvious late winter and spring peak, but the specific formation over the giant sequoia area occurs with a much higher relative frequency in April.

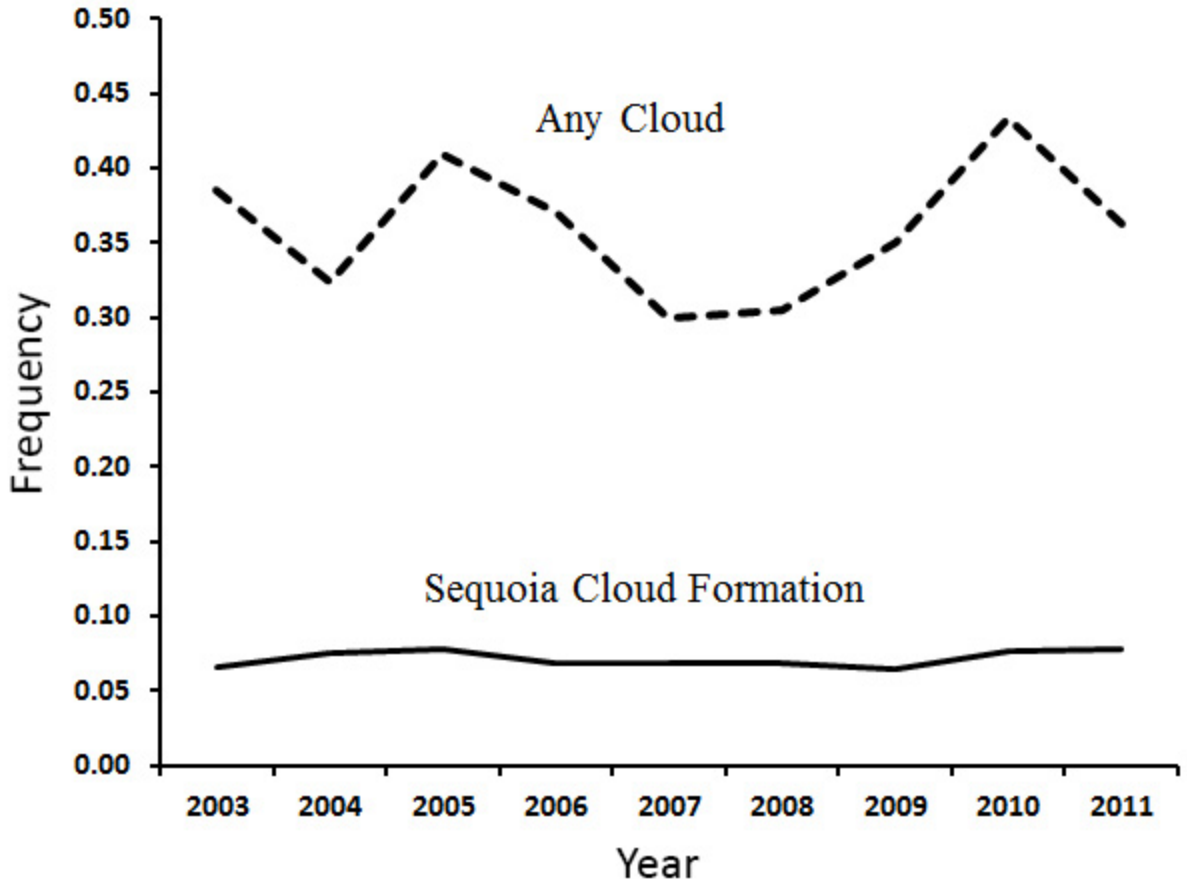


Figure 17. MODIS Aqua annual cloud frequency, 2003-2011, sequoia belt.
The frequency of specific "Sequoia Cloud Formation" appears steadier than the overall cloud frequency. This nine-year record is not long enough to establish long-term trends, but there is no obvious long-term trend apparent in either.

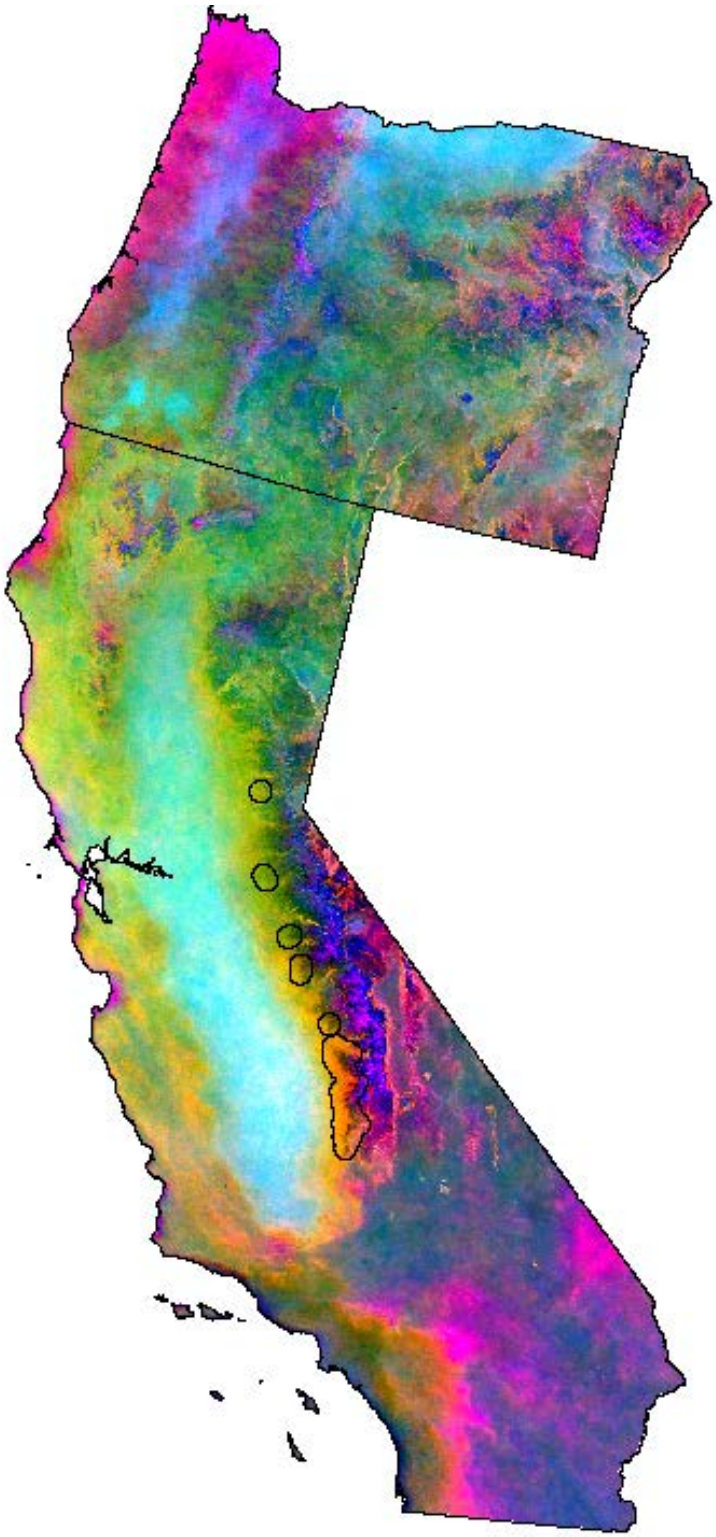


Figure 18. Color display of 2nd-4th principal components of MODIS Aqua monthly cloud frequency (2003-2011) for California and Oregon. RGB= PC4, PC2, PC3.
Buffered giant sequoia groves outlined in black. Sequoia region notable for relatively high values in PC's 2 and 4, and distinctly low values in PC 3, hence orange color in display. Principal component loadings can be found in Table 2.

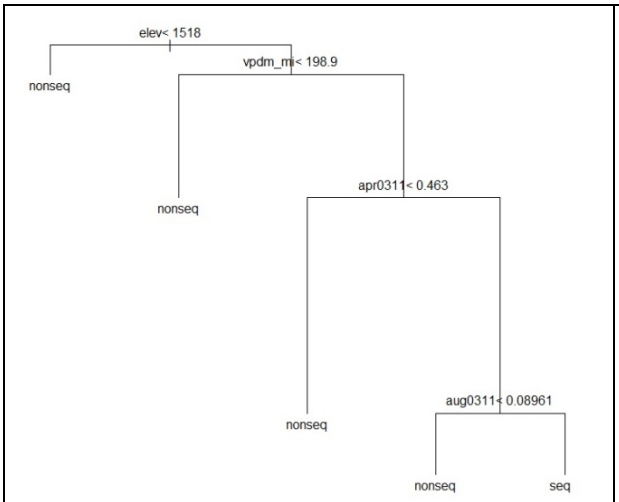


Figure 19. Giant sequoia classification tree in R, California and Oregon model extent
April cloud frequency of secondary import.

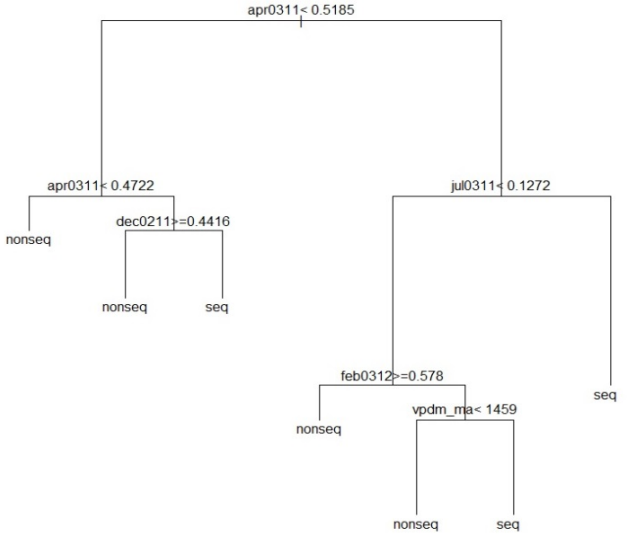


Figure 20. Giant sequoia classification tree in R, Sierra Nevada model extent
April cloud frequency now of primary importance.

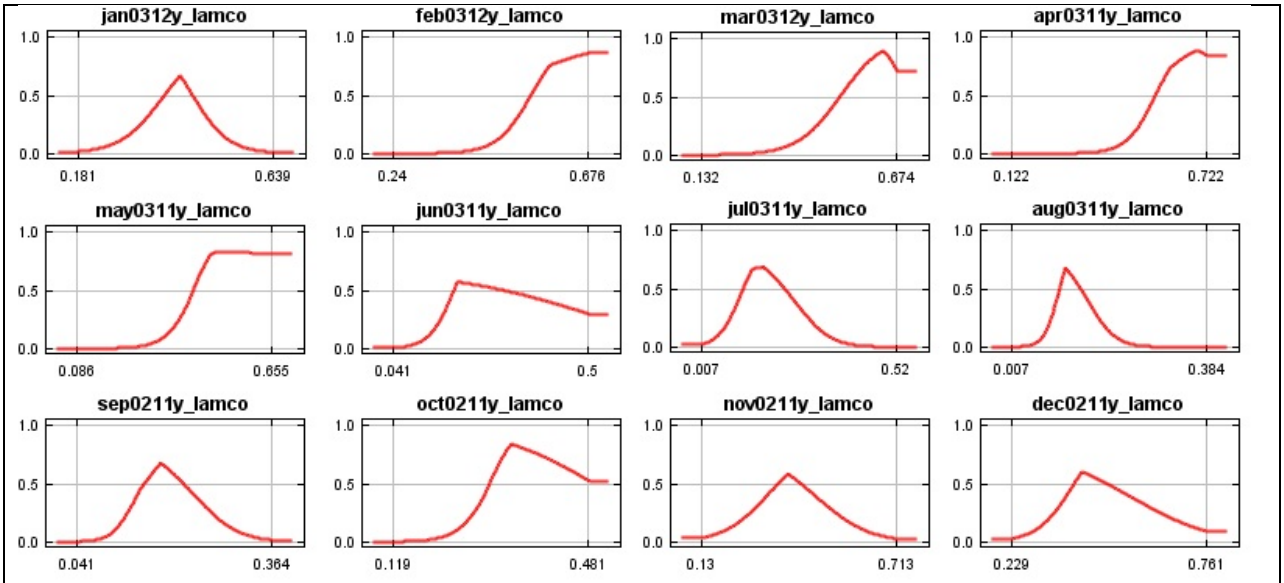
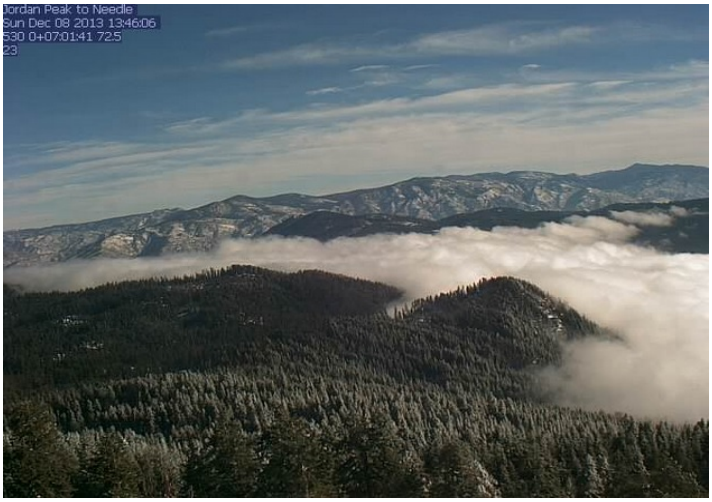


Figure 21. Maxent response curves for the monthly cloud frequency predictor variables in sequoia distribution model.
X-axis cloud frequency for given month, and y-axis sequoia response (~ probability). Sequoia distribution correlates with high levels of February, March, April, May, and October cloud cover.



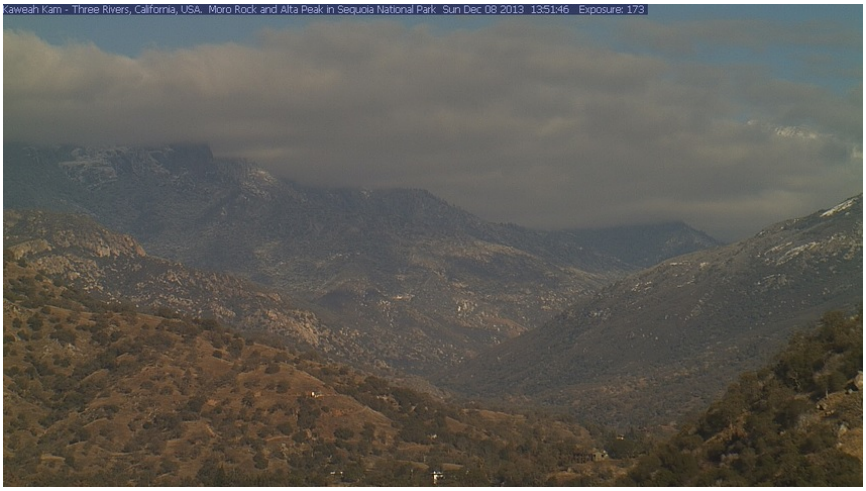


Figure 22. Webcam images for giant sequoia region, Dec. 8, 2013

A. Milk Ranch to Castle Rocks (Sequoia National Park)

B. Jordan Peak to Needles (Sequoia National Forest)

C. Three Rivers toward Moro Rock (Sequoia National Park)

(http://ssgic.cr.usgs.gov/dashboards/WebCam.htm)

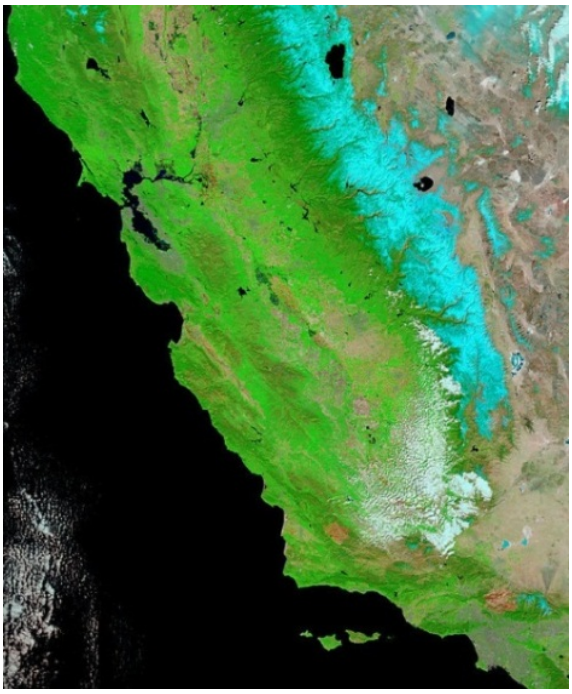


Figure 23. MODIS Terra (~11 AM) clouds
Clouds frequently form over the extreme southern Sierra and Tehachapi Mountains in the morning.

Figure 24. MODIS Aqua (~1 PM) clouds
Later in the same day as Figure 23, the clouds have migrated to center over sequoia belt.

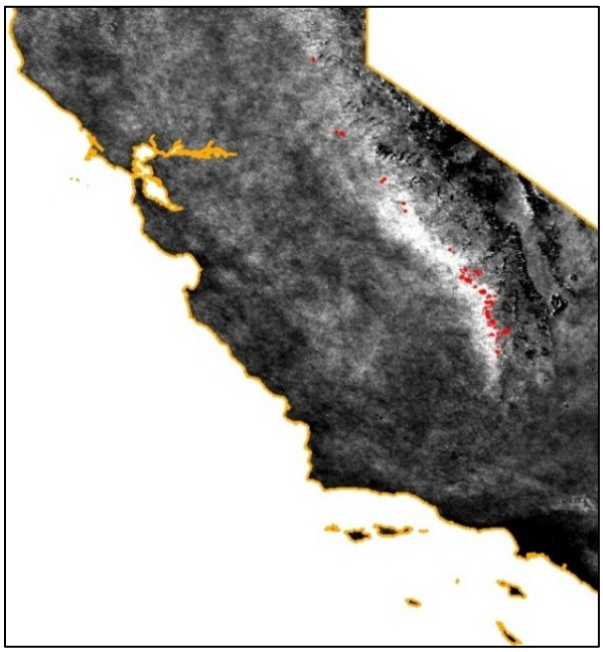


Figure 25. MODIS March afternoon minus morning (Aqua-Terra) cloud frequency.
Clouds clearly increasing over the day near the sequoia belt.

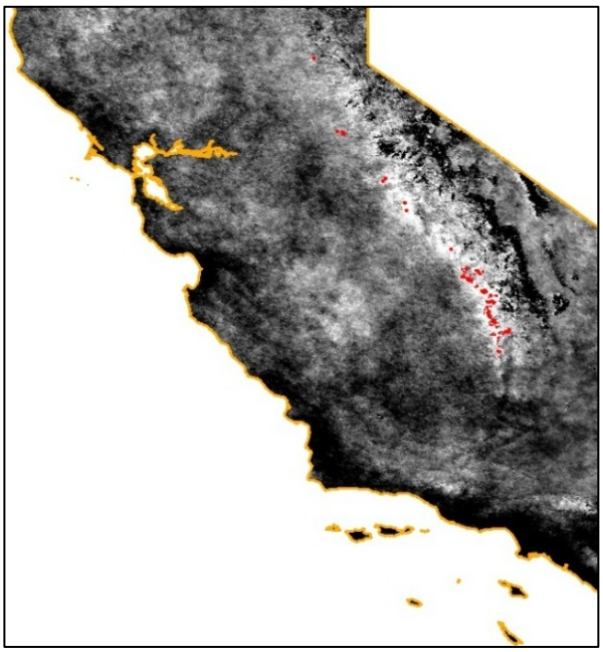


Figure 26. MODIS April afternoon minus morning (Aqua-Terra) cloud frequency.
Same as for March in Figure 25, but fog retreat near coast more evident.

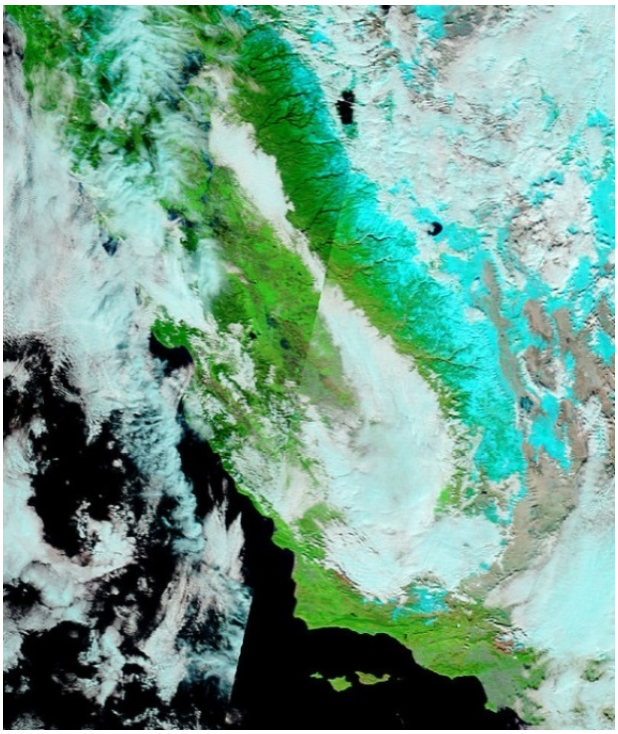


Figure 27. MODIS,Terra (~11:00 AM),1/27/2010
Possible winter Tule fog in the morning. No certainty of cloud base height at ground.

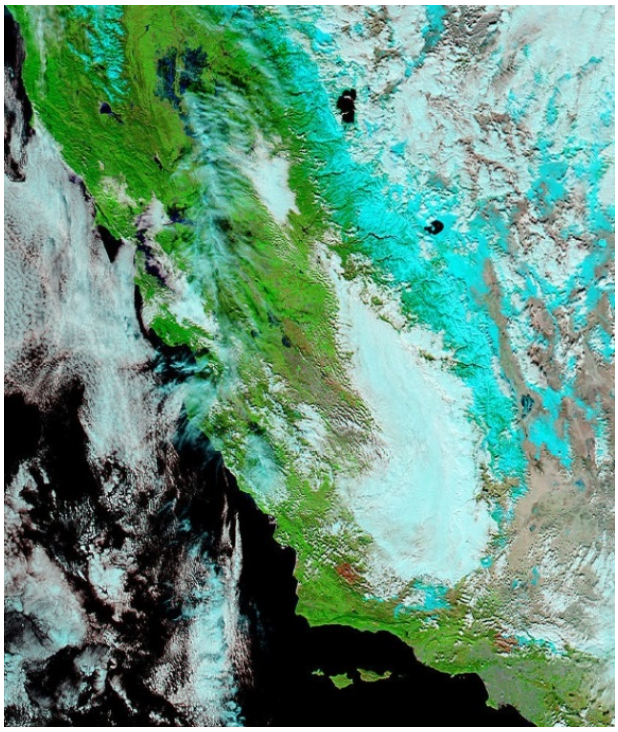


Figure 28. MODIS, Aqua (~1:00 PM), 1/27/2010
Later in same day as Figure 27, appears to hint at Sequoia Cloud Formation in the afternoon.

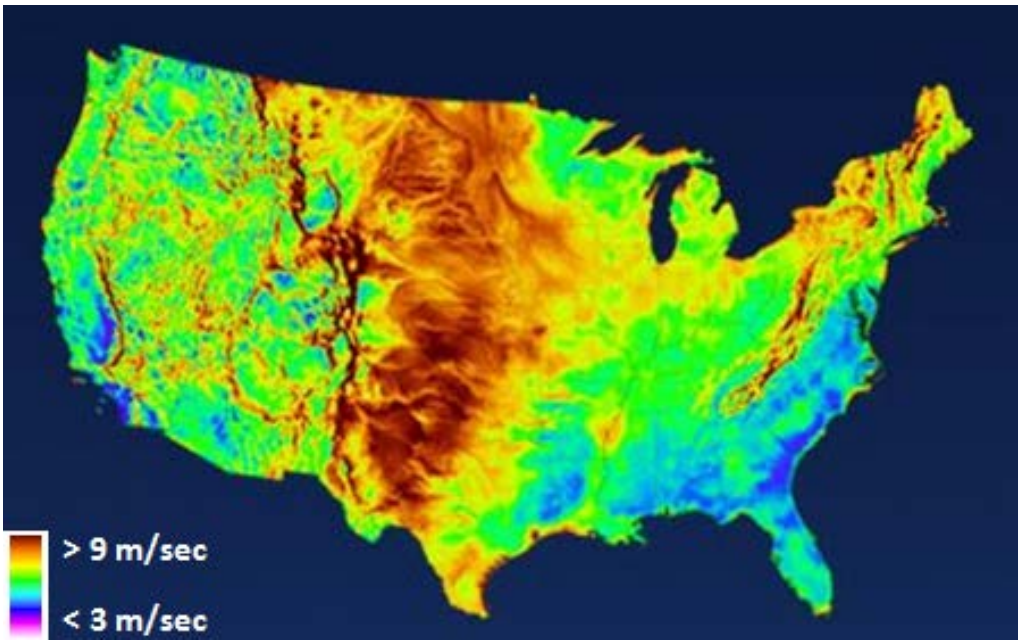


Figure 29. 3Tier U.S. average wind speed at 80 meter height.
The sequoia region and adjacent areas of the San Joaquin Valley have some of the lowest average wind speeds in the country.

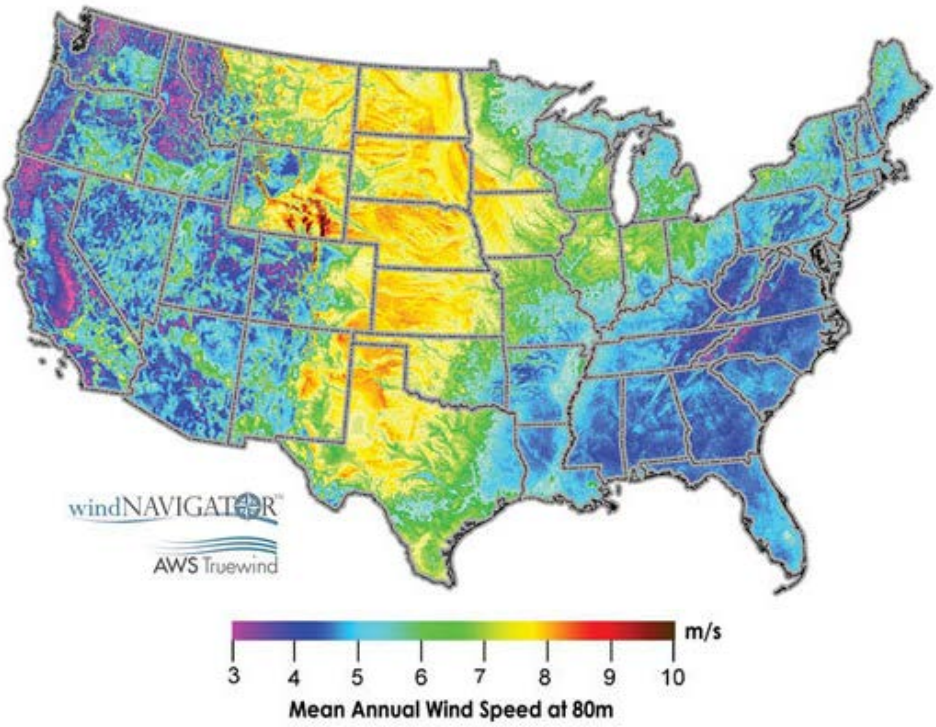


Figure 30. AWS Truewind (now Truepower) U.S average wind speed at 80 meter height.
This product does not agree with the 3Tier product everywhere (e.g., extreme differences over Appalachian Mountains) but it does agree that winds are low in the sequoia region.

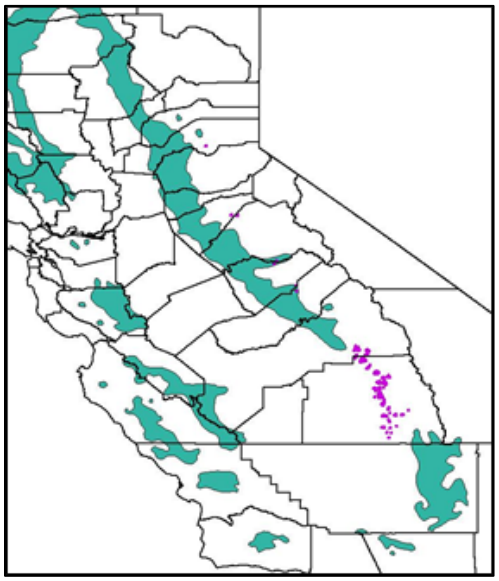
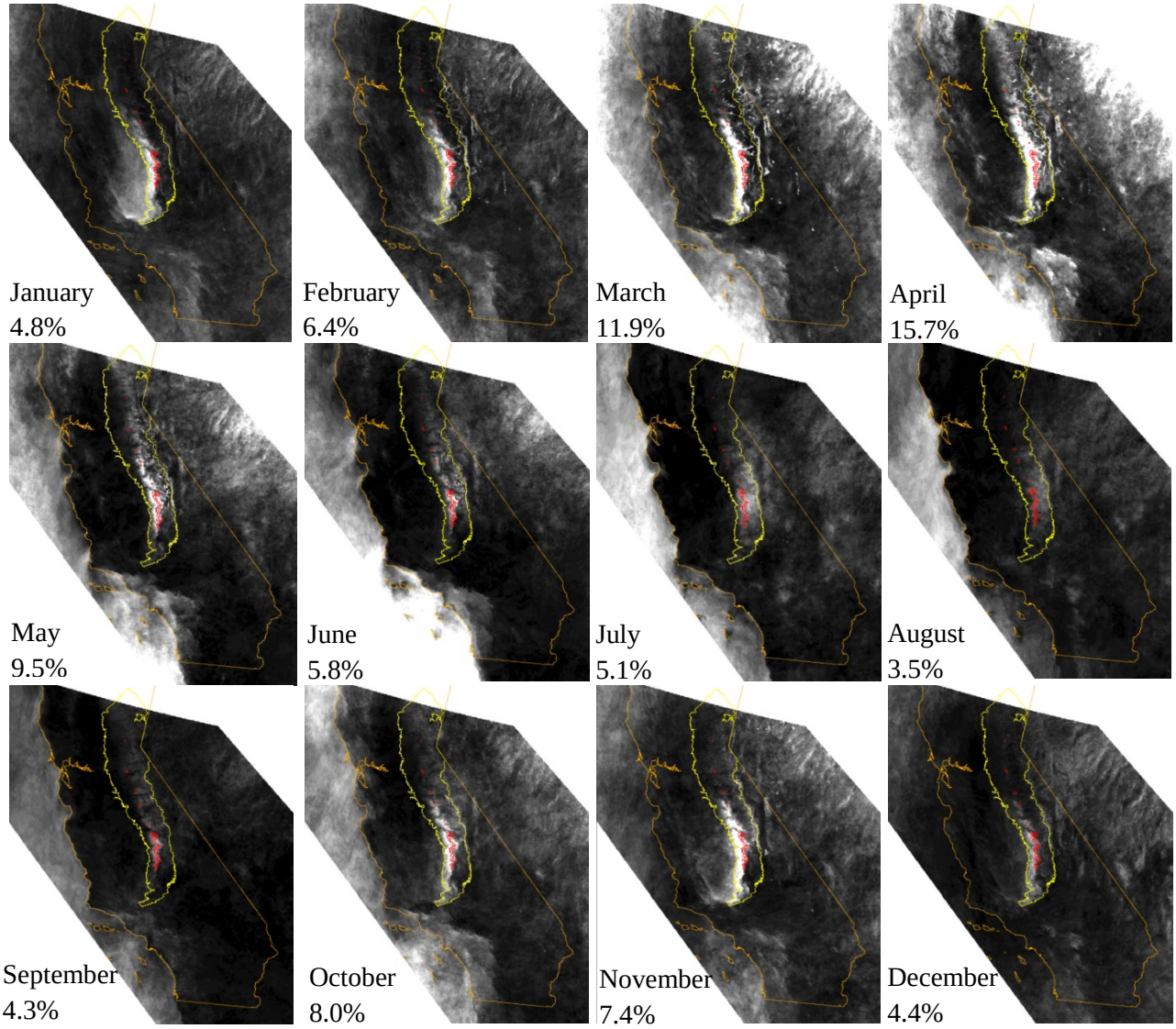


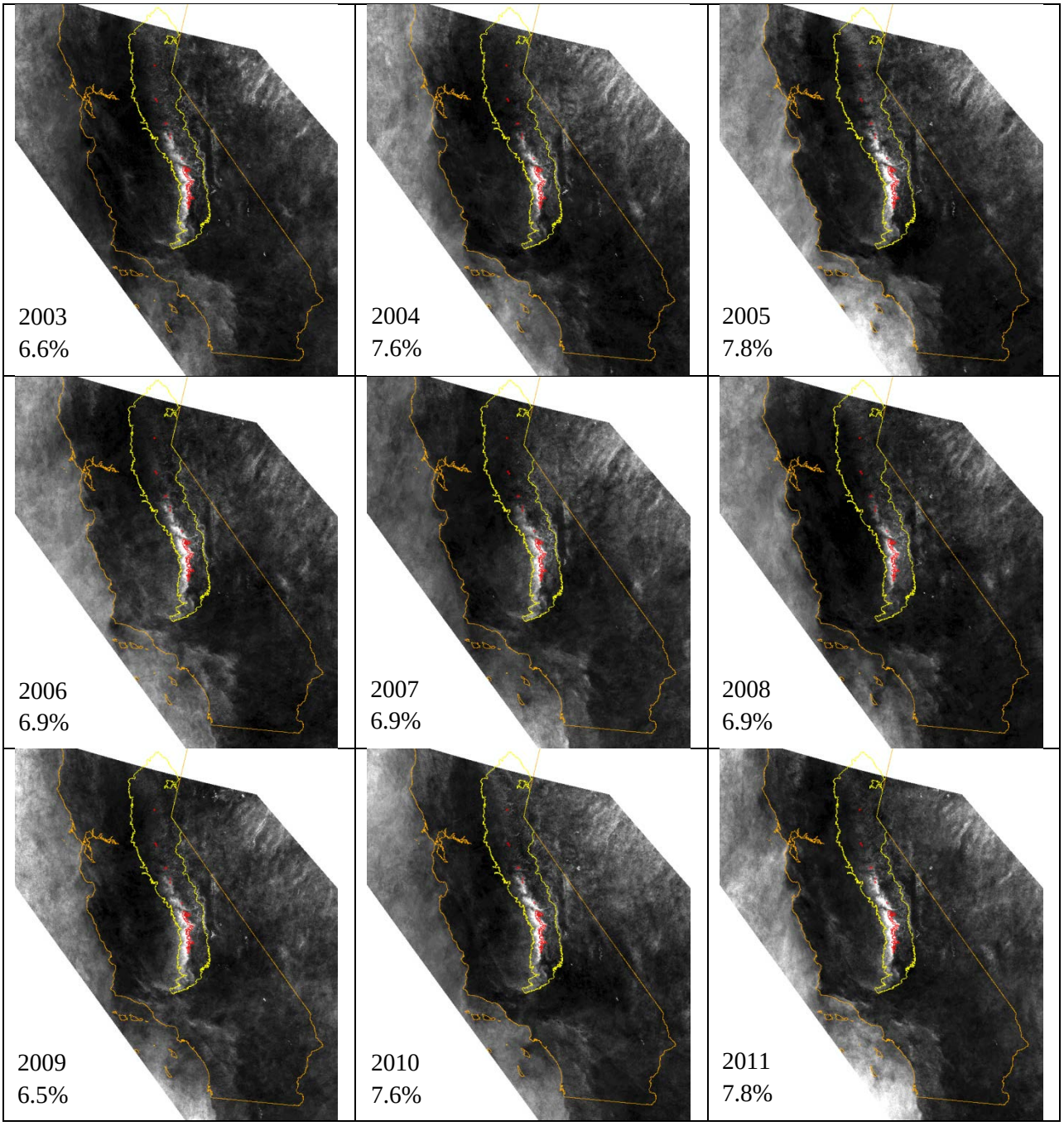
Figure 31. Giant sequoia distribution in relation to gray pine (*Pinus sabiniana*).
Sequoia (magenta) occupies the long-noted gap in the distribution of the pine (teal), although at a higher elevation.

Appendices

3.A.1. MODIS Aqua monthly frequency on days determined to have the "Sequoia Cloud Formation" for sinusoidal grid tile H08, V05 according to the spatial filter. The spatial filter of daily cloud cover masks helps to highlight the frequency of the specific formation over the region of giant sequoia. The product also indicates other areas that tend to be cloudy when the formation is occurring (e.g., southern California mountains and Great Basin ranges of Nevada). Monthly frequency percentages within the giant sequoia distribution (red) indicated. California is outlined in orange and the Sierra Nevada ecoregion is in yellow.



3.A.2. As for Appendix 3.A.1, but annual frequency, 2003-2011.



Chapter 4. Satellite, Mapped, and Meteorological Tower Data Reveal Unique Climate for Giant Sequoia

Abstract

Giant sequoia (*Sequoiadendron giganteum* (Lindley) Buchholz) occurs in scattered mid-elevation (~1400-2400 m) groves across the west slope of the Sierra Nevada of California, with a much higher concentration in a relatively confined southern belt. The west slope of the southern Sierra Nevada in this vicinity, the "sequoia belt", has a unique climate evidenced by a high frequency of cloud cover. Cloud cover can occur here at almost any time of year, but a sequoia belt formation is especially distinct on afternoons in spring months following the passage of cold fronts, while most of the region is otherwise cloud-free. A custom cloud mapping and climatology (Chapter 2) allowed the identification of this cloud cover as the strongest predictor of sequoia distribution from a large set of climatic variables (Chapter 3). This chapter describes the use of ancillary satellite data, climate station data (including data collected at 55 meters above the surface), and other data sources to investigate various aspects of the associated climate and ascertain the mechanisms behind the frequent cloud cover. In the process, unusual behavior in several other climatic covariates is identified to contribute to a particularly outstanding climate in the region that is not simply unusual for being cloudy. Many of these unusual characteristics are particularly evident at the 55 meter height of a micrometeorological tower. These include a consistently and extremely low range in diel temperature (particularly in spring), anomalously low spring temperatures (particularly during the day), dew point temperatures and relative humidity values that peak during the afternoon, and consistently and unusually low wind speeds. Any number and combination of these climatic covariates could also play a large role in the distribution of giant sequoia.

Introduction

Frequent cloud cover on the west slope of the southern Sierra Nevada (hereafter, Sierra) is a potential contributor to the distribution of giant sequoia (*Sequoiadendron giganteum* (Lindley) Buchholz), as described in Chapter 3 of this dissertation, but the clouds may be more of a symptom than a cause of a unique climate. While clouds increase surface relative humidity due to their impact on temperature, their actual formation is related to high vapor pressure relative to temperature (i.e., at saturation; conversely, low temperature relative to vapor pressure) at whatever elevation they are forming. Values of these variables at the surface, where they would have a direct impact on giant sequoia (hereafter, sequoia) growth, may be correlated with their values at elevation, irrespective of cloud formation. Identifying climatic drivers of the frequent cloud formation could help to clarify controls on sequoia distribution and, in the process, clarify the unique climatic conditions associated with the west slope of the southern Sierra.

Little attention has been paid to latitudinal patterns in climate in the Sierra. Differences in climate with elevation have begun to get more assessment, especially due to concerns about climate change and the impact on the rain vs. snow elevation threshold that can influence winter snowpack and summer runoff amounts (Jeton *et al*, 1996; Dettinger *et al*, 2004; Lundquist *et al*, 2008). Some recent studies have touched upon characteristics of the Sierra environment that hint at the unique atmosphere of the west slope of the southern Sierra – the sequoia belt. One (Feld *et al*, 2013) found that the spatial patterns of dew point temperature lapse rates of Yosemite National Park in the central Sierra often behave differently from the steadier lapse rates of the American River basin to the north. Dew points of Yosemite drop less steeply to 2000 meters, especially in late spring and early summer (when they can drop at only 3 °C/km), and more rapidly above 2000 meters. A 5.3 °C/km lapse rate is more typical for the American River Basin. They attribute most of this difference between basins to the different topographic trend on the west slope of Yosemite (a steep rise to 2000 meters, followed by a descent into the Tuolumne River basin) and its effects on air flow patterns. They speculate that "flow separation" caused by the Clark Range (a spur off the main Sierra crest) keeps moister air to the west. The topographic configuration in Yosemite may be more similar to the abrupt topography of the west slope of the more southern Sierra, where dew point lapse rates might show a similar, if not more extreme, pattern.

Feld et al. found that PRISM climate data (Daly *et al*, 1994) did a reasonably good job of capturing average dew point lapse rates in the Sierra, but performed better in the American River basin than in Yosemite where the complex topography and lack of sufficiently dense climate stations likely posed problems. PRISM and Daymet (Thornton *et al*, 1997) dew point maps sometimes hint at areas of relatively high dew points on the west slope of the southern Sierra, but patterns lack coherence – in the case of PRISM, abrupt discontinuous patterns in minimum temperature (resulting from their attempt to capture complex microclimates) have cascading effects on relative humidity. Inconsistency in relative humidity measurement at and among climate stations ((Dai, 2006); personal observation), whether due to sensor inconsistency or microclimatic effects, respectively, may further contribute to problems in mapping complex micro- and topo-climates in highly variable terrain.

Other climate variables besides dew point have been found to show notable variation around the 2000 meter threshold on the west slope of the southern Sierra. A study of evapotranspiration across an elevation gradient (Goulden *et al*, 2012) found that winter productivity is maintained up to this elevation, which they attribute to plant functioning at relatively moderate winter temperatures up to this elevation. At some point above 2000 meters, but below 2700 meters, trees appear to go dormant in winter, without recovery during any periods of warm-up. This elevation also coincides with a marked turnover in species composition (Goulden *et al*, 2012). Atmospheric pollutants rising up from the Central Valley of California have also been found to reach a limit at approximately this 2000 meter threshold. For example, Barbour (Barbour and Minnich, 1988) mentions that west-slope forests of the southern Sierra between 1300 and 1900 meters are especially impacted by ozone exposure (and the resulting ozone chlorosis / chlorotic mottle disease). Other authors (Van Ooy and Carroll, 1995; Panek *et al*, 2013) have related this exposure to topographic situation and regional wind patterns that impact the southern Sierra more than similar areas to the north.

This chapter describes the use of a variety of satellite, surface, micrometorological (micromet) tower, and mapped (including modeled/assimilated reanalysis) climate data to investigate the unusual climate associated with frequent cloud formation over the sequoia belt on the west slope of the southern Sierra. Satellite and mapped data provide the synoptic view required to ascertain how the climate of the southern Sierra specifically differs from that of the northern Sierra, where sequoia is generally absent. Climate maps are used to determine whether there are any unusual long-term patterns in spring temperature, humidity, or precipitation associated with the frequent cloud formation during this period. Modeled reanalysis data allow similar analyses, but also provide a larger view (both horizontally and vertically) on whether the sequoia region is subject to unique (cold) airmasses in the spring or associated with particular vertical movements of air in the spring. Satellite land surface temperature data help to address whether unusual temperature patterns associated with the sequoia belt, not captured by mapped climate data, might contribute to its frequent cloud formation. Atmospheric data from satellite allow the assessment of whether atmospheric humidity and water vapor are frequently high in the sequoia belt even when clouds are not forming. Measurements at the top of 55 meter micromet towers are especially helpful for giving highly accurate measurements of the extent to which climate above the immediate surface differs from that near the surface – an important consideration for both tall trees and an unusual atmosphere. An elevation transect of four 55 meter tall micromet towers spanning ~2300 meters in the south-central Sierra includes one tower roughly within the northern end of frequent "Sequoia Cloud Formation" (hereafter, SCF). This tower and transect allow detailed investigation of meteorological factors inside and just outside the SCF. Surface and tower top measurements of ½ hourly temperature, dew point, relative humidity, and wind speeds and direction, in particular, are assessed on days with and without afternoon cloud formation, and help to identify the climate and meteorological conditions associated with the frequent cloud formation. All of these disparate data sources help to clarify, in turn, possible controls on sequoia distribution.

Data and Methods

Chapter 2 of this dissertation described the generation of cloud climatologies from satellite data, including MODIS reflectance data. These were used to identify the particularly cloudy climate associated with sequoia distribution on the west slope of the southern Sierra in Chapter 3. Analysis here uses a variety of sources to ascertain climatic features associated with this particularly cloudy climate: 1) additional MODIS satellite data, including the land surface temperature product and various atmosphere products; 2) climate data from mapped and modeled reanalysis data sets, as well as surface climate stations; and 3) meteorological data from an elevation transect intersecting the northern end of the particularly cloudy belt, including surface measurements and 55 meter micrometeorological tower top measurements.

MODIS Land Surface Temperature Data

MODIS land and atmospheric data are easily accessible (http://ladsweb.nascom.nasa.gov/), generally with twice-daily coverage (Terra, with prefix MOD; and Aqua, with prefix MYD), and include many georegistered derived products. For example, MOD11 and MYD11 data provide a total of 4 daily (2 day and 2 night) observations of land surface temperature (LST, an emitted radiance-based estimate of "skin" temperature, not to be confused with air

temperature). These sensors have equatorial crossing times of approximately 10:30 AM and 1:30 PM daily, but due to the northern latitudes of the area of interest, local overpass times are closer to 10:00 AM +/- 50 minutes and 1:00 PM +/- 50 minutes. Night overpass times are ~12 hours later for each. Daily night time Aqua (~ 1 AM) and day time Aqua (~ 1 PM) data were downloaded from 2012 and "differenced" to determine an approximate range of daily diel land surface temperature (dLST) for cloud free days. These data were averaged over months and the entire year. In addition to the daily data, the 46 eight-day MYD11 LST composites for 2012 were processed and averaged for the year.

MODIS Atmosphere and Vegetation Products

Two other sources of MODIS satellite data were evaluated for remotely-sensed information on the climate of the sequoia belt: MOD05/MYD05 (Precipitable Water Vapor: PWV) and MOD07/MYD07 (Atmospheric Vertical Profile). MOD05/MYD05 have day (thermal infrared- and near infrared-derived) and night (only thermal infrared) products as well, and can be easily downloaded and viewed, but are only Level 2, so they do not come in a georegistered format, nor are easily reprojected. Similarly, MOD07/MYD07 provide 5 km resolution non-georegistered vertical profiles of the atmosphere, which include temperature and dew point temperature at 20 atmospheric pressure levels, extending well up into the atmosphere, from which corresponding vapor pressures and relative humidity were determined for lower atmosphere levels. These pressure levels are related to height, which can vary somewhat in relation to regional air pressure patterns and more substantially with topography (they are generally at greater heights over elevated areas, although generally lower with respect to the surface terrain height – until the highest pressure values don't even exist over higher terrain). Difficulties associated with the mass reprojection of these datasets discouraged their processing, but they are still useful for viewing purposes.

Other satellite products of interest include MOD16 (actual evapotranspiration (AET) and potential evapotranspiration (PET)) and MOD13 vegetation indices (e.g., the Normalized Difference Vegetation Index (NDVI) and the Enhanced Vegetation Index (EVI)), although the EVI associated with MOD13 has notable viewing geometry problems; MCD43B4 Nadir BRDF-Adjusted Reflectance (NBAR) provides important modeled standardization of measured reflectance to a common nadir-viewing geometry (Schaaf *et al*, 2011) with which improved vegetation indices were calculated. Solar geometry is not standardized in the NBAR product. Table 1 summarizes the various MODIS data used and referenced here.

Mapped, Modeled Reanalyis, and Surface Station Climate Data

Climate data included PRISM (Daly *et al*, 1994) and Daymet (Thornton *et al*, 1997) maps. Analysis focused on patterns of relative humidity, temperature and precipitation. Relevant seasonal patterns of temperature and precipitation, in particular, were identified through the simple differencing of values between months. This included comparisons of spring (April) and fall (October) temperatures, late winter/early spring (March) and winter (February) temperatures, and late winter/early spring (March) and late fall/early winter (December) precipitation. Surface climate information was obtained from Remote Automated Weather Station (RAWS; www.raws.dri.edu) data. This included data on average monthly snowfall and total precipitation for a variety of mid-elevation sites across the Sierra. A ratio of monthly snowfall (not snow water equivalent) to total monthly precipitation (likely very close to snow water equivalent in

winter months) provided a relative index of temperatures aloft during precipitating storm events. Higher ratios suggest a higher percentage of precipitation falling as snow and/or a colder and drier snow.

Other climate data were derived from National Center for Environmental Prediction/North American Regional Reanalysis (NCEP/NARR) (Kalnay *et al*, 1996) data (available at http://www.esrl.noaa.gov/psd/cgi-bin/data/narr/plotmonth.pl). Atmospheric vertical velocity (Omega) values, at a variety of atmospheric pressure heights, show average spatial patterns of rising and descending air in various months. NCEP/NARR data were also used to calculate monthly integrated precipitable water vapor (PWV) across the atmosphere. Finally, additional NCEP/NARR mapped datasets (e.g., geopotential heights, vector winds and meridional winds) provide supporting information on larger-scale climate patterns.

Meteorological Data

An elevation transect of flux towers (Goulden et al, 2012), part of an ecohydrological study associated with the Southern Sierra Critical Zone Observatory (http://criticalzone.org/national/), has been a source of micrometeorological measurements for several years on the west slope of the south-central Sierra at an elevation interval of ~800 m (405, 1160, 2015, and 2700 m) with ½ hourly climatic measurements (including temperature, relative humidity, incoming solar radiation, sensible heat flux, latent heat flux, wind speed, and wind direction) at 55 meters above the ground surface (Table 2 and Figure 1). The data span various 2-4 year windows, from 2009-2012, depending on the site, with some inconsistency in data acquisition. The transect crosses elevations of the sequoia belt near the northern extent of the main belt of high cloud frequency. (Sequoia start to become much less common this far north, but the McKinley Grove of trees is quite close to the transect, occurring at ~1900 meters, less than 10 air miles southeast of the 2015 m tower and about 12 miles west/southwest of the 2700 m tower. The SCF is still commonly present here, and the 2015 m tower, in particular, falls within the zone of high cloud frequency.) These tower data are accompanied by surface measurements, along with an additional network of climate data loggers scattered throughout the Providence Creek watershed (mostly elevated within tree canopies) just south of the 2015 m tower.

Dew point, vapor pressure deficit, and dew point depression (temperature minus dew point temperature; proportional to lifting condensation level) were additionally calculated from the tower data. These, along with the above-mentioned climate variables of interest were then averaged (when data were available – technically, using NanMean in Matlab to ignore empty values; NanMedian was used for wind direction – these should have binned the into directional categories and then processed with "Mode", but trends were still evident and these data were not a focus) for the half hour intervals across the days of the month to generate average ½ hourly trajectories by month. Individual months were then averaged across the multiple years to generate longer-term monthly averages. Relative humidity poses a unique situation for averaging values, given that the saturation vapor pressure (and thus relative humidity) is an exponential function of temperature. See Appendix 4.A.1 for a brief discussion of the issue.

Climate averages on days without afternoon cloud cover were also generated. This was an attempt to evaluate whether climate patterns on clear days were different from those days on

which the SCF was occurring. These clear afternoons were determined by selecting the days on which afternoon radiation (8 measurements from 12 to 3:30) was within 85% of an estimated daily afternoon maximum for a month, determined by taking the maximum value from each of those 8 ½ hour increments over the course of a month. This daily afternoon maximum value may overestimate a given day's potential afternoon radiation, both because of solar variability over a month (e.g., late January will be higher than early January, near winter solstice), and because reflection off of nearby clouds can enhance local radiation values. The 85% figure accounts for this possible overestimate and was determined as a compromise between accidentally including cloudy days and obtaining too little data. [This method does not discriminate between the SCF days and general afternoon cloud cover. If this were the goal, more complex daily analysis would be required, including comparison with satellite data. The SCF filter was intended as a relative index for comparison among months and years – it is less appropriate for absolute analyses. See Chapter 3 for more details.]

Results

MODIS Satellite Land Surface Temperature Data

The sequoia belt has notably low diel land surface temperature (dLST) range throughout the year on MODIS LST data. These areas have low values in the annual average for 2012 and even stand out as low in the difference between summer dLST (~8 °C) and winter dLST (~ 4 °C), meaning that their summer dLST values, while higher than winter, are especially low relative to other areas. That is, they have a low daily range, and a low seasonal range in their daily range. Figure 2, a classification of the values in color, highlights the low average range of temperature in the sequoia belt compared to other areas of California. Figure 3, with the same information but in continuous grayscale, highlights a similarity in appearance with a true color band, indicating a relationship with vegetation density. The dLST has a very strong (negative) correlation with cloud cover frequency in the Sierra (i.e., lower temperature range with more cloud cover), even though all the LST values come from clear days (Figure 4). A belt of low dLST extends a bit beyond the main range of sequoia, both south and north, from the southern end of the Greenhorn Mountains east of Bakersfield to west of the Mariposa Grove of southwestern Yosemite National Park. An outlier of low dLST matches the Tuolumne and Merced Groves of western Yosemite. Only the far northern groves of sequoia do not have strikingly low dLST values, just as they do not have strikingly high cloud cover frequency.

MODIS Satellite Atmospheric Data

MODIS PWV (MOD05) integrated water vapor products and vertical profile data (MOD07) were found to have some consistency and accuracy issues. The thermal and near-infrared versions of the MOD05 water vapor products do not tend to agree, with the thermal product apparently more sensitive to upper atmosphere moisture, and the near-infrared (NIR) product more sensitive to lower atmosphere moisture. Most obviously, the thermal product maps areas adjacent to cloud formation as high water vapor, whereas the NIR product maps these areas as low water vapor. The thermal product is also at a much coarser resolution (5 km vs. the 1 km resolution of the NIR product). Similarly, the vertical profile data are very coarse (also at 5 km), often noisy-looking, and have coarse patches of missing data ostensibly related to atmospheric pressure levels that are absent at various elevations.

Despite disagreement between the two water vapor products, they tend to agree that the edge of California's Central Valley in the vicinity of the sequoia belt is a region of frequently high integrated water vapor. The NIR-based PWV data (Figure 5), at a finer resolution and possibly more sensitive to lower atmosphere moisture, show this pooling of moisture at the foot of the sequoia belt especially well. Because it integrates across the atmosphere, there is a strong tendency for lower elevation areas to have higher values, and it is difficult to disentangle the inverse relationship between elevation and water vapor without intensive processing. The thermal PWV product appears slightly less reliable for capturing the humidity of the sequoia region – this may be due to its sensitivity to upper atmospheric moisture, as well as its coarseness and noisiness. It has inconsistent behavior when the actual cloud formation is occurring: sometimes showing high values adjacent to clouds when the formation occurs, sometimes overly masking clouds in the region (probably due to high PWV values), and sometimes showing low PWV values and missing the clouds altogether.

The MODIS vertical profile data show consistently high relative humidity in the vicinity of the sequoia belt, at a variety of pressure levels (Figure 6). These high humidity values are a result of unusually low temperatures compared to surroundings and high dew point temperatures. These patterns were evident at a variety of the higher pressure levels (mostly observed at 700, 780, and 850 mb, very roughly corresponding to ~3000, 2200 and 1500 meters, respectively) typically available for the area. The highest pressures (e.g., 1000 mb, approximately sea-level) are missing data due to the elevation. In general, areas of higher surface elevations counterintuitively tend to have relatively higher temperatures and dew point temperatures at the various elevated pressure levels – this is likely related to the fact these areas have respective pressure heights that are relatively low compared to their surface elevations (i.e., not as high above the surface as they would be elsewhere).

In contrast to the adjacent high elevations, MYD07 vertical profile data consistently show low elevations for the 850 hPa pressure heights over the foothills of the sequoia belt (evaluated in January, April, and July). These are expected to be ~1500 meters, and yet values of ~200 meters are observed here. Similarly, 700 hPa heights of 1700-1800 meters are found here, rather than the expected ~3000 meters. This is evidence of a consistent low pressure trough due to rising air. Low atmospheric pressure is, in turn, associated with cold air aloft (Ahrens, 2011).

Climate Maps and Station Data

PRISM climate data show patchy areas of west slope Sierra ridge-tops with ranges in monthly maximum minus minimum (i.e., diel) temperature (~10-11 °C) that are almost as low as those indicated by the satellite data. But sequoia groves often sit just below these ridge-tops and tend to have slightly larger ranges in temperature (~12-14 °C) in the mapped data. Many other ridges and higher elevations throughout the Sierra are mapped with lower temperature ranges. The lowest temperature ranges (~7-9 °C) in California are mapped along the immediate coast and highest elevations of some coastal ranges and the Sierra. (Contrastingly, some of these coastal ranges have extremely large temperature variations in Daymet climate data. This is likely due to problems with extrapolating to high elevations based on elevation lapse rates derived from comparing coastally influenced low elevations with more continental higher elevation stations.)

Long-term NCDC climate data from Grant Grove and Giant Forest comport with these PRISM values, with monthly diel ranges of temperature from 10 °C (winter) to 13-15 °C (summer).

Patches of high relative humidity are usually associated with these areas of low maximum temperature. Overall, they hint at some high values on the western slope of the southern Sierra, but they do not stand out compared to the high humidity values near the highest elevations of the Sierra. Nor do the patches of higher relative humidity perfectly coincide with the sequoia groves, likely for the same reason that the areas of low maximum temperature do not perfectly coincide.

Subtracting April maximum temperature data (PRISM) from October and November maximum temperature data indicate that California stands out from the rest of the United States in terms of relatively cool spring temperatures. Most of the U.S. is warmer in April than in November. The immediate coast of California, the tops of some coast range mountains, the highest parts of the Sierra, and parts of the sequoia belt are exceptions to the nationwide pattern. Similarly, only California's coastal mountain ranges and the western front of the Sierra are substantially colder in April than in October (Figure 7). Zooming in a bit on California in Figure 8 (March maximum temperature minus February maximum temperature), March has little to no warm-up from February on the west slope of the southern Sierra – one of the few areas other than a coastally influenced zone (of primarily southern and central California) with this behavior.

Cold spring storms may explain these relatively cold spring temperatures in the mountains of California, and the west slope of the southern Sierra, in particular. PRISM data indicate that the latter has a relatively high frequency of spring storms – for example, it gets much more precipitation in March than in December (Figure 9) – like the mountains of southern California, and unlike areas of the Pacific Northwest, which have a peak or near-peak of precipitation in December. The fact that spring storms in California are particularly cold, especially at elevation, is supported by climate data showing monthly snowfall in relation to monthly total precipitation. Dividing the former by the latter yields a ratio that peaks in March and April for a variety of sites at mid-elevations (5000-7000') in the Sierra, including the Giant Forest and Grant Grove areas of the sequoia region (Figure 10). These high ratios in spring indicate either of two possibilities: 1) of the precipitation that falls in spring, it is more likely to fall as snow, and 2) the snow is drier and less dense in spring (i.e., more snow for a given water equivalent). Each of these possibilities indicates colder conditions during storms.

Figure 8 shows the Giant Forest and the Grant Grove as behaving like other mid-elevation areas of the Sierra in terms of snowfall in relation to total precipitation. But this is not shown in combination with the information from Figure 9 – the map that shows that a relatively high portion of the sequoia region's precipitation comes in March (and April too, although not shown). The result is that these sequoia sites get about as much snow in April (25.6" and 25.0" for Giant Forest and Grant Grove, respectively) as they do in December (28.9" and 26.9" for the two sites, respectively), despite getting ~60% of the total precipitation. They also get about twice as much snow in April as in November (12.8" and 15.7", respectively), despite very similar amounts of total precipitation in these months. These relative amounts are not matched by other areas of the Sierra, where April snowfall is comparable to November snowfall.

Like the MODIS atmospheric pressure height data, NCEP/NARR data show consistent rising air in the southern San Joaquin Valley, with negative vertical velocity (Omega) values at 850 hPa (~1500 m) on average throughout the year except for December and January. Appendix 4.A.2 shows these products for January and April, as well as 500 hPA (~5500 m) products for the same months showing consistent descending air aloft over the southern Sierra associated with subsidence. The coarsely-interpolated resolution of 0.3 degrees may explain an imperfect spatial match with the sequoia foothills (seen in MODIS data) in the pattern of rising air in the 850 hPa product.

NCEP/NARR data were also used to calculate integrated precipitable water vapor (PWV). (See the Appendix 4.A.3 for exemplary months.) These data give confirmation that the Central Valley of California has unusually high PWV in spring months, relative to air above adjacent areas of the Pacific Ocean. PWV values tend to vary seasonally with temperatures, as warmer air can hold more moisture, hence different value scaling in the legends. The adjacent Pacific provides a sensible standard for comparison with the Central Valley, as consistent westerly and northwesterly winds result in its strong influence on Central Valley airmasses.

Tower Air Temperature Data

Tower measurements of temperature at 55 meters show even lower diel ranges in temperature than the satellite and surface measurements. At the 2015 meter tower, within the cloud belt, average diel air temperature range varies from only 1 °C in the fall through spring to about 2 °C in the summer months (Figure 11). Outside of the main cloud belt, at the 2700 meter tower, these ranges go up slightly, but still remain quite low, from about 2-4 °C, with the lowest ranges in winter (Figure 12). This difference among tower sites is less than the difference between 55 meters and the surface; surface measurements near the 2015 meter tower have seasonal variation in diel temperature ranges from 3-8 °C, with the lowest ranges (and lowest temperatures) December through April (Figure 13). It is notable that this latter range is still quite a bit lower than values for the NCDC climate stations associated with the sequoia groves. It is in line with ranges in dLST from MODIS for this area. It is also in line with seasonal variation in diel temperature ranges for the lower two towers, where the lowest ranges are restricted to the winter months.

Air temperatures at the 2015 meter tower are relatively warm in January and quite cold in the late winter and early spring, with February to April colder than January. Substantial warm-up occurs in late spring through summer, but late spring months are still colder than would be expected (e.g., May<< October and June<<September) for non-coastal locations (where thermal inertia associated with a large adjacent body of water should be less of a factor). The upper (2700 meter) tower showed similar seasonal behavior, with unusually cold spring values. The lower two towers show somewhat similar seasonal patterns, though January is progressively relatively colder with decreasing elevation, and late spring (i.e., April and May) values are progressively relatively warmer (Figure 14 and Figure 15). The fact that January is quite cold at the lower (especially lowest) towers demonstrates that the results at the upper towers were very likely not related to any especially anomalously warm January months and anomalously cold spring months; the upper elevations simply have unusual seasonal temperature patterns.

Slightly different monthly temperature patterns were evident on days without substantial afternoon cloud cover, as determined by the afternoon radiation algorithm (Figure 16). December, February, and March air temperatures at the 2015 meter tower were a bit warmer in these cases, in line with the previously (i.e., on averages from all days) warmer January temperatures. April is also a bit warmer – slightly warmer than January in these cases, although still much colder in comparison than would be expected of most areas of the northern hemisphere (e.g., much colder than October and November). These daily periods (midnight to midnight) finish warmer than they started because they are a select subset of the data.

Tower Humidity Data

Relative humidity measurements across the transect of four towers (Figure 17) were especially revealing, in terms of the timing of cloud formation. The 2015 meter tower, within the SCF belt, had annual peak relative humidity during the afternoon, as opposed to other sites which experienced maximum humidity in the early morning (405 meter) or early evening (other two). This afternoon peak in relative humidity is highly unusual and coincides with the timing of the SCF. The monthly patterns at the 2015 meter tower (Figure 18) reveal that this afternoon peak is strongest in the spring months (February-June, in particular April), followed by the fall months of October and November, just as for frequency of the SCF.

Relative humidity is determined from the ratio of vapor pressure (determined from dew point temperature) and saturation vapor pressure (determined from temperature). Dew point temperatures peak in the early to mid-afternoon at the 2015 meter tower (Figure 19). This is much earlier in the day than for other towers. And while seasonal patterns in dew point temperature look to be about what might be expected seasonally, with winter lowest and summer highest, actual temperatures at the 2015 meter tower (Figure 11) do not follow the same seasonal trends, and this has major implications for relative humidity. In combination with incredibly low swings in diel temperature, spring temperatures (February-April) are extremely low – lower throughout the day and night than in January. May and June are also colder than would be expected. These afternoon peaks in dew point and low afternoon temperatures combine to cause the unusually high February-June afternoon relative humidity values.

Many of these same climate patterns were evident on days without afternoon-specific cloud development, as determined by the afternoon radiation algorithm. Relative humidity is often high at the 2015 meter tower, even on days without afternoon cloud cover (Figure 20). Interestingly, however, February and March are notably lower (behaving more like winter months on these clear days in terms of both their overall lowering, and in only having a narrow later afternoon peak), as are October and November to some extent, while April through June continue to have particularly high relative humidity values, especially notable in the early afternoon.

Tower Wind Data

Tower winds might be slightly low (compared to U.S. maps at 80 meter height) due to their measurement at 55 m, but values are extremely low, especially at the Soaproot Saddle (1160 m) site, where average wind speed is less than 1 m/s (Figure 21). Towers record slightly increasing winds moving away from this site, with adjacent towers averaging about 2 m/s and the highest (2700 m) tower averaging just a bit over 2 m/s. The three highest towers all record

consistently low values at night (generally lowest in summer), with higher values during the day (peaking in the spring and summer at the upper two, with no clear seasonal trend at the 3rd highest). While daytime winds at the upper two sites are between 2-3.5 m/s and evening winds are 1-2 m/s, the 1160 m site has daytime winds of 1-1.4 m/s and night winds half of that. The lowest tower shows much more daily and seasonal variability, with generally lower winds in the winter (day and night) and generally higher winds in the summer (day and night). Most months at this lowest tower show a substantial early morning and early evening trough (abrupt drops in wind speed that coincide with a switch in wind direction, morning and night), although the early evening trough becomes very muted in the spring and summer (with late spring and summer having another near-midnight drop in wind speed).

Wind directions are pretty consistent throughout the year at each tower, but vary among towers and switch between day and night at all of the towers. Daytime winds are consistently from the southeast at 2015 meters. The towers just above and below (2700 m, and 1160 m) have daytime winds consistently from south/southwest. The lowest site, at 405 m, has the least consistency over the day in wind direction, with winds generally migrating from south early in the day in all seasons, to southwest (winter), west (spring), and northwest (summer) over the course of the afternoon to early evening. The low site has much greater consistency in direction at night, when winds are always from the northeast. The next tower up (1160 m) has winds consistently from the east/northeast at night. The upper towers tend to have winds from the north at night.

Discussion

MODIS Satellite Land Surface Temperature Data

MODIS LST has been found to have very little bias for night-time air temperature measurement. Daytime LST values, however, are often far from air temperature values, due the heating of the surface "skin". This is most extreme in hot deserts, where daytime LST values exceeding 50 °C often occur, even when air temperatures are far below this. This difference between radiant LST and air temperature is likely the main reason that many areas with low diel air temperature (i.e., coastal areas) do not also have extremely low diel LST. Only well-vegetated canopies in regions of low diel air temperature also have particularly low dLST. It has been suggested (Goulden, personal communication) that the particularly low dLST of the sequoia belt may essentially be mapping tree height, because the CZO towers up at heights of tree canopies (55 m) capture similarly low temperature range. More specifically, it was suggested that the low dLST might be an integration of temperature from surface to tree canopy height, and may include the effects of evaporative cooling and canopy shadowing. It is difficult to know the extent to which MODIS LST might be integrating radiant temperature across height in its LST measurement, because it is notoriously difficult to validate the daytime LST (Wang *et al*, 2008). Coarse pixel resolution and the diversity of land cover types, as well as a diversity of surface heights from which LST may be sensed, all pose problems for validation. It does seem likely that something cool above the surface (either canopy or air) is influencing the LST measurements in the sequoia belt, however, as this is one of the few areas where daytime LST values are on par with surface air temperature, and yet one would expect at least some additional contribution from surface heating, despite the dense vegetation. In any case, the spatial pattern of low dLST doesn't

perfectly coincide with consistently tall forest, but it does tend to follow a region of fairly continuous mostly forested vegetation cover, with a good deal of canopy shadowing. It also coincides so closely with the SCF (Figure 4), despite the fact that the LST measurements come from clear days, that there is likely a direct relationship between the two: these areas must be consistently cold above the surface, and more frequently reaching saturation vapor pressure as a result.

There is also the possibility that high humidity associated with the cloud belt is contaminating the LST measurement. Cloud cover can radically alter LST values – areas of clouds are usually masked (sometimes too much so), but thin clouds can be missed (personal observation). While long term processing can help to remove the effects of these anomalous data, the fact that thin clouds, cloud edges, and other areas near clouds often show up as low LST raises the concern that areas of high humidity, and not just thin clouds, might alter LST measurement. It has been shown that the product does not have large errors related to *surface* humidity (Wang *et al*, 2008), but processing to avoid humidity contamination involves view-angle specific correction (Wan and Dozier, 1996; Wan, 2008), and some residual view-angle related error remains (Wang *et al*, 2008). It seems likely that variable atmospheric humidity, or even just missed cloud edges and thin clouds, could be a factor.

Several factors suggest that humidity or cloud contamination should not be the dominant mechanism influencing low dLST values in the sequoia belt. The spatial pattern of low dLST is very consistent on a daily basis (during clear skies when measurements are reliable) – even in winter, when water vapor is not expected to be high at these elevations (above common boundary layer inversion heights). Yet other data sets (tower data, as well as MODIS precipitable water vapor and vertical profile of dew point) indicate that high humidity and water vapor are not present in the region on every day. Mostly, however, there appears to be confirmation that these low dLST values are real, contrary to what might have been suggested by established climate stations and maps of the area. The surface measurements in the forest (Upper Providence Creek) give comparable ranges to satellite-derived dLST, and the tower data at 55 meters have even lower temperature ranges. The substantial magnitude of differences with established climate stations may give more reason to be concerned about the representativeness of the established climate stations than for the satellite data. For example, the Shadequarter RAWS station just west of Sequoia National Park at 1240 meters includes a picture of the site in its metadata: the met station is positioned over a patch of dirt (http://www.raws.dri.edu/cgi-bin/rawMAIN.pl?caCSHQ) on a bare grassy ridge. This relative lack of vegetation will cause it to heat up much more than areas with more typical levels of vegetation, which will benefit from canopy shading and evaporative cooling. And this site has relatively low diel temperature ranges (compared to other established climate stations) of 7-11 °C.

Considering the above factors and extensive viewing of thermal data, the factors likely contributing to the especially low dLST of the “sequoia belt” include, in possible order of their dominance:

1) A truly low range in diel temperature at (~tree) height due to exogenous climatic factors, but also often perceived at height here due to pixels consisting of well-stocked stands of (often tall) trees

2) Low surface heating and surface radiative temperature due to relatively continuous canopy cover, as well as shadowing from canopy and terrain
3) Same as for factor 2, but leading to cooling of actual air
4) The evapotranspirative cooling effects of these well-vegetated stands on air and surfaces
5) Atmospheric (mostly water vapor) contamination of radiance data

These factors are not mutually exclusive. The perception of surface temperature at tree height described in factor 1 could apply to factors 2 through 4 as well. Factors 1, 3, and 4 are real air temperature effects. Factor 2 is likely a very large factor – there is certainly some uncertainty associated with this order.

MODIS Aqua-based dLST is calculated from night-time temperatures (at ~1 AM) that could be far from the minimum in some areas where substantial cooling occurs throughout the evening until early morning (pre-dawn). This potential bias would probably be far larger than any bias associated with not necessarily truly capturing the high temperature by measuring at ~1 PM. Surface measurements near the 2015 meter tower suggest that night-time temperatures there more or less bottom out not long after the sun sets, not tending to cool substantially through the evening. This behavior is not necessarily typical, however, and regional variability in the difference between 1 AM temperature and pre-dawn minima could play a role in the dLST patterns. Accounting for this bias would probably raise dLST values across the board, but would also likely increase the contrast between humid and dry areas, where greater night-time cooling might be expected in the latter.

MODIS Satellite Atmospheric Data

The MOD05 precipitable water vapor (PWV) and MOD07 atmospheric vertical profile products were useful for confirming that humidity and water vapor are frequently high in the vicinity of the sequoia belt but precision with which conclusions could be made is low. Just as the two different versions of the PWV product often fail to agree on amounts, the vertical profile data can look a bit different from the PWV products. While the vertical profile data are coarse and noisy, the regularity with which the sequoia belt is mapped as having high relative humidity at a variety of pressure levels (generally low temperature AND high dew point temperature) is concerning; tower measurements indicate that high humidity should not be ever-present. There is a slight concern that the unusual air temperatures or underlying vegetation characteristics could be artificially influencing the vertical profile humidity value.

The fact that the two different PWV products (thermal infrared and near-infrared) do not show the same patterns of water vapor is also a concern, but there still appears to be useful information, particularly in the NIR-based product. The daytime only MOD05 NIR-based PWV product is the finer resolution of the two PWV products and more sensitive to lower atmosphere moisture. As such, it is a better indicator of humidity in the region of the sequoia belt. This product is useful for giving an idea of the rough geographic location and frequency of moisture pooling, but the exact atmospheric height of the moisture is not provided. It appears to show lower elevations near the sequoia belt as having frequently higher levels of water vapor, but radical drop-offs with elevation. While this may reflect reality to some extent, due to the thinner atmosphere at higher elevations, it's possible that this product may underestimate vapor at higher altitudes, just as it appears to be less sensitive to upper atmosphere moisture. An underestimate

of vapor at higher altitudes seems to be supported by the high relative humidity values indicated by the MOD08 vertical profile product over mountains (red ellipses in Figure 6) that are mapped as having low integrated water vapor (compare same areas in Figure 5).

Climate Maps and Station Data

PRISM climate maps hint at patterns of interest in the sequoia belt, but fail to show strong local correlation between climate variables and sequoia distribution consistently. This can be seen through a visual comparison and also in the fact that PRISM climate were not strong drivers in modeling the distribution of sequoia (Chapter 3). The lack of strong relationship may be the result of any of several factors: inconsistency in surface station data (e.g., microclimatic effects) leads to inconsistency in maps; they capture surface patterns rather than above-surface patterns that sequoia is more responsive to; they do capture some relevant climatic patterns, but these are just slightly offset from species locations. These maps are the best available, more reliable than Daymet in the complex terrain of the Sierra, but it is likely that they would benefit still from a denser network of climate stations (Feld *et al*, 2013). It is important that these climate stations are judiciously placed, consider microclimatic differences, and perhaps even surface height differences.

PRISM climate data may not capture the fine-scale climate that sequoias perceive, for the above-mentioned reasons, but they may capture broad patterns that help to explain the story behind the unique climate of the west slope of the southern Sierra. For example, the unusually cold early spring (March and April) temperatures in the Coast Ranges of California and the west slope of the Sierra were highlighted by comparison with their October and November temperatures. These areas stand out nationally for their relatively (i.e., relative to fall) cold early spring temperatures. (The sequoia belt also stands out to some extent as having low values in the PRISM-derived diel temperature range for the same months of March and April, due to these low daytime values.) The evident differences between spring and fall temperatures are not as extreme as those observed in the tower data, because the maps are derived from surface measurements that are more sensitive to local surface warming with the relatively higher sun and increased radiation in the spring months. (At all four towers, solar radiation is comparable between March and October, but April is greater and November is less, even with greater spring cloud cover.) The relatively cold spring values are more extreme in the mountains than near the coast, suggesting that they are not related to local ocean temperature differences between spring and fall.

Especially cold spring temperatures above the surface are strongly indicated by the snowfall to total precipitation ratios for mid-elevation RAWS data for the Sierra. Surprisingly cold spring temperatures have also been reported for some of the highest elevations (RAWS at 3800 m) of the White Mountains in eastern California (Hall, 1991). These were attributed to a high frequency of closed low pressure systems in late winter and spring. This greater frequency of closed lows in the spring relates to more frequent meridional flow in spring, as the Pacific High migrates north. There is some evidence for this in northeastern Pacific Ocean patterns of geopotential heights, vector winds, and meridional winds in NCEP/NARR data. All indicate average flow arriving in California from the northwest in spring, as opposed to closer to due west in the winter.

Tower Air Temperature Data

Tower collected air temperature data suggest that the satellite-based LST data may, in fact, strongly reflect air temperatures just above the surface. Diel variability in air temperature at the towers is actually lower than the unusually low satellite-based dLST values. These ranges are consistently approximately 1-2 °C throughout the year for the 2015 meter tower. This site, like the other towers, is less subject to surface diel heating and cooling shown by adjacent surface measurements. But it is also outstanding compared to other towers: only the upper tower comes close to having similarly low diel temperature ranges. This 2015 meter tower appears to sit within the unique zone associated with cloud formation and consistently cold air (especially in spring) at height. It may be well-positioned with respect to local topography and winds, as well, which may help it capture conditions reflective of those above the surface layer. The lower 1160 meter tower is also close to the cloud and low temperature zone, but does not share the same low diel pattern as the 2015 meter tower. It is likely that while this site often sits under clouds, it tends to be farther under the clouds than the 2015 meter site. There may also be other aspects of its local situation that prevent it from capturing any elevated extremely low temperature ranges.

Tower Humidity Data

Just as the towers capture temperature conditions that differ from the immediate surface, they capture relative humidity values that differ. Colder daytime temperatures in comparison to the surface, in combination with dew point temperatures that are not proportionally lower, result in higher relative humidity values, at least for the 2015 meter tower. This tower is distinct from others in having a daytime (early afternoon) peak in relative humidity. This is quite unusual, and is the result of dew point temperatures peaking early in the afternoon, and the unusually low temperature rise during the day. This is also consistent with this tower better tapping into the atmosphere above the surface layer: cool moist air associated with cloud formation likely has a tendency to sit a bit above the surface, just as the clouds do not always form a ground fog (at least not at every elevation; it is more likely on their eastern edge where they may intersect the terrain).

Even days without the afternoon cloud development have unusually high relative humidity values throughout the afternoon, at least for April and May (and to some extent June). February, March, and November have notably different daily relative humidity trajectories on days without the afternoon-specific cloud development. The trajectories, with steep peaks only in the late afternoon with low solar elevation, look more like those for the winter months of December and January, although they occur on relatively warm days. It is difficult to speculate upon the reasons for the different behavior between late spring and early spring in this case.

The next tower above on the transect (2700 meters) has steadily rising relative humidity throughout the day from mid-morning on, that always peaks near the end of the day or early evening. This is the result of dew point peaks later in the day, combined with more substantial late afternoon drops in temperature. The next tower below (1160 meters) has similar patterns to the upper tower in the late fall through early spring. In the late spring and summer, however, this site has large drops in relative humidity throughout the day, associated with larger temperature increases, and late afternoon dew point decreases. This is very similar to the lowest (405 meter) tower, which has steady daily declines to late afternoon troughs in relative humidity consistently throughout the year, due to the fact that its dew points are highest early in the day (reaching

afternoon plateaus in fall through early spring, but dropping substantially in the afternoon in May through September).

Either the air at 55 meters above the surface at 2015 meters on the west slope of the southern Sierra, the sequoia belt, is quite a bit different from the air at 55 meters, just ~800 meters above and below this elevation, or the 2015 meter tower happens to be more suitably located for obtaining characteristics of the atmosphere above the immediate surface layer. Two RAWS climate stations on the west slope of the southern Sierra show daily dew point and relative humidity trends that are similar to that of the 2015 meter tower, with frequent afternoon peaks, but at 2300 meters (Park Ridge) and 2570 meters (Buck Rock). These stations are both on exposed ridge sites, suggesting that the siting with respect to local topography may be the key factor. But the appropriate elevation belt as the unique factor cannot be dismissed either, as these sites are not substantially above the sequoia belt, and the 2570 meter station appears to experience less frequent incursions of afternoon water vapor than the 2300 meter site.

Tower Wind Data

Two different wind maps for the United States (AWS Truepower and 3Tier) portray the east side of the San Joaquin valley and the adjacent southern Sierra foothills as having the lowest average wind speeds (at 80 meter height) in the country, at less than 4 meters/second. Finer patterns of the lowest winds within this low wind area are not detailed (wind rose data for larger cities in the valley, such as Fresno and Bakersfield, indicate averages in the low 3's, with slightly higher values in spring and slightly lower values in winter) but tower data here suggest that the lowest winds may actually occur somewhere between 400 and 2000 meters (likely near the 1160 meter tower) on the west slope of the Sierra. Other sources closer to the main sequoia belt (e.g., RAWS Park Ridge at 2300 meters) suggest that the elevation of minimum winds may be at a higher elevation within the heart of the sequoia belt – they are frequently very low (e.g., 1-2 mph = < 1 meter/second) at the 2300 meter station, at least at the surface.

Low winds may be the reason that water vapor pools in the region, along with other atmospheric constituents, such as aerosols and pollutants (ozone and nitrogen) that are known to show high values in the region (Van Ooy and Carroll, 1995; Panek *et al*, 2013). Vertical profiles of the atmosphere in the southern San Joaquin Valley show strong relations between specific humidity and ozone and possibly NO2 (Ian Faloona, PhD, personal communication), with peaks in each near the top of the summer boundary layer (800-900 meters). LIDAR profiles of the Central Valley atmosphere reveal spatial patterns of aerosols (De Young *et al*, 2005; Lewis *et al*, 2010) pooling at the top of the mixed layer (1200-2000 meters) against the slopes of the Tehachapi Mountains and extreme southern Sierra.

Many of the wind directions might relate to local topography with respect to mountain/valley breezes, but the steady low wind during the day from the southeast at the 2015 meter site is consistent with evidence that moisture (and resulting clouds) is additionally transported parallel to the Sierra, possibly as part of the Fresno Eddy – topographically induced counterclockwise circulation around the periphery of the San Joaquin Valley (Bao *et al*, 2008). This mountain parallel flow is sometimes observed on satellite time series to transport clouds in a north and northwesterly direction along the western front of the Sierra. This same circulation has been identified as a possible cause of ozone transport into the west slope of Sequoia and

Kings Canyon National Parks (Panek *et al*, 2013). That this southeasterly wind is only picked up at the 2015 meter tower may relate to the possibility that this tower is better positioned above surface processes and/or to the possibility that this air movement is fairly confined to this elevation belt. Wind profiler data in the Sierra foothills near Auburn have also been found to show extremely low winds from the southeast at altitude, although in this case it is between 900 and 1800 meters, or 500-1400 meters above the surface (Fast *et al*, 2012), and this is also well north of the influence of the Fresno eddy.

Boundary Layers and Subsidence Inversions

Topographic configuration of the Central Valley and surrounding mountains may play a role in the convergence of air at the surface, low winds, rising air, and high convective boundary layer heights in the Central Valley, which are especially high on afternoons in the spring months of March to June (Bianco *et al*, 2011; Jeong *et al*, 2013). The surprising timing of this increased boundary layer (and its secondary sub-peak in October and November), as opposed to summer when greater convection is expected, coincides perfectly with the timing of increased SCF (Figure 22). Boundary layers have been modeled as particularly high on the periphery of the Central Valley at this time (Jeong *et al*, 2013). Relative humidity is often higher near the top of the convective boundary layer in the afternoon (Fast *et al*, 2012) and this elevated humid layer may pool (and often form clouds) against the topographic blocking of the mid-elevations of the abrupt west slope of the southern Sierra, especially after the passage of cold fronts with residual moisture from the maritime polar airmass. The situation may be analogous to the cold winter *nortes* that bring cold (of continental origin, in this case) polar air to eastern Mexico leading to the formation of a stratus deck at intermediate (1000 to 2000 m) elevations on the east side of the Sierra Madre Oriental, under the capping influence of a subsidence inversion (Fitzjarrald, 1986).[1] This formation has also been described as upslope fog reinforced by stratus advection (Garcia-Garcia and Zarraluqui, 2008). California's boundary layers are often under the influence of strong subsidence inversions – these are known to influence the formation of Central Valley Tule fog in winter (Holets and Swanson, 1981) and California coastal fog in summer (Iacobellis *et al*, 2009).

The development of the convective boundary layer has been identified as a means to loft pollutants upward in the Central Valley and into the Sierra (Bao *et al*, 2007; Collett Jr *et al*, 1990; Fast *et al*, 2012) and it likely does the same for water vapor. The boundary layer over the adjacent Sierra is typically much shallower than that observed for the Central Valley in spring (Choi *et al*, 2011; Jeong *et al*, 2012; Jeong *et al*, 2013), and this may cause the stratus deck to be much closer to the surface in the Sierra (and even squeeze the water vapor into a shallower stratum?). A capping subsidence inversion at the top of the boundary layer (occasionally penetrated by convective activity and cumulus formation) keeps the moist air and clouds below the highest elevations of the Sierra, which generally sit high and dry in these conditions.

Cloud formation over the sequoia belt may not always be the result of the exact same mechanism. Larger cloud events that extend into the Central Valley, with the appearance of a flat-topped stratus deck, might be more indicative of a large-scale rising convective boundary layer (capped by a subsidence inversion) driving the formation. However, when cloud formation

[1] Modification of the cold, polar air by the warm water of the Gulf of Mexico (Cavazos, 1994) provides the moisture and heat at the surface to enhance convection, possibly making it more like the maritime polar air in California.]

is only directly adjacent to the slopes of the Sierra and less stratified, it might be more indicative of (shallow) upslope flow (anabatic flow due to local convection) and adiabatic cooling. Nor does every cloud formation coincide with an afternoon peak in relative humidity at the surface: after a major front passes through, there may be enough residual moisture in the atmosphere that an afternoon rise in the convective boundary layer is not required to provide the moisture for cloud formation.

Exogenous vs. Endogenous Sources of Moisture and Cooling

The sequoia belt, particularly the southern end, has among the lowest LST swings (at least from ~1:00 AM to ~1:00 PM) in California – rivaled only by a few coastal mountain regions dominated by coastal redwoods, such as the Santa Cruz Mountains. Various lines of evidence presented here suggest that this is the result of unique climate patterns in the region, with unusual levels of cool daytime temperature and high relative moisture that may often sit above the surface layer. An argument might be made that the relatively cool daytime temperatures and high relative moisture could be the result of a highly productive ecosystem with high levels of evapotranspiration and the associated evaporative cooling. This is similar to the argument addressed in Chapter 3 about whether the productive forest is the main source of moisture for the SCF development. It was argued in Chapter 3 that daily movement and seasonality of the clouds do not support this notion. (The clouds do not develop and hover over sequoia groves – they often move latitudinally over the course of a day in a predictable manner, often coalescing over the sequoia region in the afternoon; additionally, their most frequent season does not coincide with the most productive period.)

Of course, the trees are transpiring water, and the latent heat of vaporization is a cooling process, so they have some impact. The question would be whether the forests are transpiring more water in the southern Sierra than in the northern Sierra, ostensibly related to greater productivity in the south. The answer is that there is no evidence that the sequoia belt is more productive and water-using than other regions of the Sierra, at least according to ET mapping. For example, data from the Breathing Earth System Simulator (BESS (Ryu *et al*, 2011)) suggest an average of only 464 mm of ET for the main sequoia belt for 2009. Mixed conifer forests of the northern Sierra are found to be just as high, if not higher (generally 450-500 mm according to this product). The MOD16 MODIS ET product (which has been found to have problems in comparison with flux tower data (Goulden *et al*, 2012)) gives average ET values across the main sequoia belt of 570 mm in 2011 and 525 mm in 2012. These tend to be among the higher values of the southern Sierra, but the same MODIS products show many higher values in productive areas of the northern Sierra forests – of around 800 mm in 2011 (wet year) and 750 mm in 2012 (dry year). MODIS products show even higher values in coastal redwood (*Sequoia sempervirens* (D. Don) Endl.) forests with values of 800-1100 mm (compared to only 500-800 mm in BESS data).

These ET products have different methodologies and can be very sensitive to various remotely-derived inputs, which are often problematic. For example, vegetation indices such as the Normalized Difference Vegetation Index (NDVI; (Tucker, 1979)) and the Enhanced Vegetation Index (EVI; (Huete *et al*, 2002)) are known to be sensitive to solar illumination angles (Galvão *et al*, 2011), which can cause seasonally varying levels of the indices that have nothing to do with vegetation productivity. The Sierra, in particular, is a noted area in which ET

products do not agree very well (Velpuri *et al*, 2013), so the values determined for sequoia areas should be interpreted with caution. While the remotely-derived products appear to be giving similar ET values for the sequoia forests, with no sign of increased productivity over the west slope of the southern Sierra, much higher values have been estimated for Sierra mid-elevation forests from flux tower data (Goulden *et al*, 2012). Despite the lack of any mapped evidence, it is possible that sequoia forests are more productive than those in the northern Sierra, but if this were the case, there would still be the question of why the sequoia were able to be so productive in the south.

In any case, the tower-collected micromet data do not support the notion of an endogenous source for moisture and cooling. The 2015 meter tower shows latent heat flux (an indicator of evapotranspiration, ET) to be consistently low from November through March, with only a small uptick in April (Figure 23), the month with the highest relative humidity values and cloud formation frequencies. May through September have much higher levels of ET, with a peak in July, and yet these are not the months with the highest levels of relative humidity. The consistently low dLST throughout the year, with low seasonality, additionally suggests that local ET is not the strongest driver of remotely-measured surface temperature, given the high seasonality of ET.

Sublimation: Tower Latent Heat Flux and RAWS Climate Data

Given that plants are inactive at the 2700 meter tower in winter (Goulden et al, 2012), measured latent heat flux (Figure 24) at this time may be dominated by sublimation. With average latent heat values of 10-20 Watts/m2, this equates to 0.3 to 0.6 mm of sublimation per day and possibly 90 mm over the duration of snow cover (~November/December to May at this elevation). Cloud cover plays a role in keeping this number from being even larger. It is notable that latent heat values are skewed toward before noon (1/2 hour interval “25”) at the 2015 meter tower (Figure 23) and the 2700 meter tower (Figure 24) in the early spring months (March-May), but skewed toward after noon at the 405 meter tower (Figure 25), as would typically be expected based on typically warmer temperatures and increased vaporpressure deficits. The lower site is certainly experiencing afternoon transpiration during this period, but the uppermost site should not be under any kind of stomatal regulation at this dormant time of year – the afternoon decline there suggests that afternoon cloud cover indeed plays a large role in reducing evaporation (i.e., mostly sublimation at this time of year). The reductions in spring afternoon transpiration at the 2015 meter site, on the other hand, where plants are likely active (Goulden *et al*, 2012), is likely related to the local peaks in afternoon humidity and reduced vapor pressure deficits.

The reduced sublimation and snow loss due to cloud cover should be evident in the persistence and amount of snow on the ground in spring. Good maps of spring snow depth are lacking, but RAWS climate station data provide average snow depth by month. Giant Forest and Grant Grove (1945 meters and 2010 meters elevation, respectively), with ~480 cm of snowfall each year on average, have an average of 12.7 cm and 15.2 cm, respectively, of snow on the ground in May. Farther north in the Sierra, Donner Memorial (1815 meters) and Huntington Lake (2140 meters) have comparable levels of snowfall each year, but only 5.1 cm and 7.6 cm, respectively, of snow on the ground in May. As mentioned, however, the southern areas also receive a higher fraction of their precipitation during this later period in the wet season. An extensive analysis was beyond the scope of this work, but cursory research suggests that cloud

cover's impact on snow's persistence and contribution to soil moisture could be a non-negligible factor in summer water relations.

Other Observations: Doppler RADAR

San Joaquin Valley NEXRAD S-Band Doppler RADAR (HNX, Hanford, CA) near Fresno consistently picks up a reasonably strong return from the vicinity of the sequoia belt, even when clouds (and certainly not rain) are not evident. This "anomalous propagation" (AP) signal is captured at elevations just below the sequoia belt – a narrow more-or-less elevation-related band during the day that looks to be particularly associated with steep escarpments, often expanding to a much wider, but more diffuse-looking band in the evening and overnight. Many authors (Laurence *et al*, 2002; Weckwerth *et al*, 2005; Fabry, 2006) have described the potential for S-band RADAR to map relative humidity due to its sensitivity to water vapor. Apparently the best correlations are found between refractivity and the mixing ratio (specific humidity), which measures the mass of water vapor relative to air. Refractivity is a known AP problem in humid coastal areas and storm fronts due to high gradients of both temperature and moisture (Bech *et al*, 2000; Chang and Lin, 2011). It may be more plausible that the AP is picking up on the frequent temperature anomalies in the sequoia region, rather than humidity, as the temperature anomalies are likely more frequent.

Relationship Between Central Valley Tule Fog and Sequoia Cloud Formation

There may be a relationship between a common winter ground (radiative) fog in the Central Valley (Tule fog) and the SCF. Mornings with Tule fog sometimes segue into afternoons with clouds adjacent to or over the sequoia belt. Researchers have hypothesized that one of the sources of fog intercepting higher elevations in the Sierra might be a lifting of Tule fogs "previously trapped near the floor of the San Joaquin Valley" (L COLLETT *et al*, 1991). The two formations don't occur at the same time – Tule fog is common overnight and into the morning, tending to burn off during the day, while clouds (or fog) move over the sequoia vicinity during the day with clearing over the Central Valley. And seasonally, Tule fog is more common mid-winter. But some of the mechanisms may be similar. Tule fog benefits from a subsidence inversion, low winds, and some moisture provided by frontal passage. All of these factors appear to be involved in the sequoia belt cloud formation, as this is an area of unusually low winds and the inversion may be acting as a ceiling to trap the humid air. [There is a noteworthy correlation between inversion strength over the southern Sierra foothills and that (generally associated with coastal fog) over San Diego from March-May (Iacobellis *et al*, 2009).] The difference in timing (both daily and seasonally) between the two events may simply be a product of the different elevations and the timing of local boundary layer inversions. It is notable that the SCF is due east of the core of the Tule fog formation – the area with the highest frequency of the winter fog.

The fact that inversions have been found to be much more common during periods of atmospheric ridging (Iacobellis *et al*, 2009), as occurs during dry periods, leads to the possibility that inversions and resulting pooling of moist air at the base of the Sierra may be especially important for sequoia as a buffer against drought. The relationship between the SCF and sequoia distribution could be analogous to the importance of the Tule fog in promoting herbaceous cover in the Central Valley over what otherwise might be desert shrublands, based on the low precipitation levels there (Twisselmann, 1967).

Conclusions

It is now clear that the sequoia belt of the Sierra has a highly unusual climate, beyond simply having a high frequency of cloud cover. This may be confirmation that cloud cover is not the only unusual factor responsible for the general restriction of sequoia to the southern Sierra. High cloud frequency was an obvious symptom of this climate and encouraged investigation. Thermal data from satellites indicated that areas of high cloud frequency (and the sequoia distribution) appeared to be associated with unusually cold daytime temperatures and unusually low diel temperature ranges. Concerns about the accuracy of the satellite data led to an assessment of climate and weather data mapped and measured in the region. Data included recent measurements made at the tops of 55 meter tall micrometeorological towers across a southern Sierra elevation transect. The tower data revealed that the climate at 55 meters can be quite different from the climate at the surface. This may especially be the case in the sequoia belt, where tower data reveal astounding climatic characteristics. Surprising features include diel ranges of temperature throughout the year of only 1-2 °C, spring temperatures colder than mid-winter temperatures throughout the day, and large afternoon peaks in dew point, which, when combined with low diel temperature swings, contribute to unusual afternoon peaks in relative humidity throughout the year, even on clear days. These afternoon peaks in relative humidity (only observed at tower height) are very likely related to the frequent afternoon cloud formation in the belt.

A possible mechanism for afternoon peaks in dew point and relative humidity is large scale rising air near the surface transporting water vapor into cooler air aloft. This rising air could be due to surface heating and a rising convective boundary layer, or convergence of winds at a surface low and divergence aloft. The resulting clouds, in turn, are capped by a subsidence inversion aloft. Cold air aloft, especially in spring, may relate to the still-cold northern Pacific Ocean at this time, together with increased meridional flow and airmasses from the northwest, as the Pacific High moves north. Local topographic factors likely play a role as well, trapping the cold air and moisture against the steep west slope of the southern Sierra. All of these possible factors require more investigation.

The unusual climatic observations in the sequoia belt, particularly from tower data, suggest the need for a variety of additions and improvements to our climate monitoring network. Climate station and map data do not appear to consistently capture this unique climate. Much of this may relate to the use of climate station data sensitive to local siting and surface conditions that differ tremendously from those just above the surface. The collection of more information at a variety of heights above the surface could help to better characterize the climate that trees (especially large trees) perceive. This research suggests that satellite data may be helpful in identifying particular unique locations where measurements at height would be especially beneficial. In addition to more locations in two dimensions and three dimensions, more precision in the temporal aspects of humidity measurement would be beneficial. The timing of minimum and maximum relative humidity has strong implications for vegetation and should be reported (or collection of something like early AM and PM values should be standardized). The sequoia belt is unusual in having peak relative humidity during the day; the common assumption that peak relative humidity generally occurs early in the morning is probably violated anywhere near an afternoon cloud formation.

This study revealed unique climatic conditions associated with the sequoia belt. Strong explanatory hypotheses are generated, grounded in an array of data sources, but exact mechanisms continue to be unanswered. These include relationships with sequoia autecology and the synecology of the sequoia belt, as well as the mechanisms behind the climatic patterns.

References

Ahrens, C.D. (2011) *Essentials of meteorology: an invitation to the atmosphere*, Cengage Learning.

Bao, J., Michelson, S., Grell, E., Grell, G. & Djalalova, I. (2007) Investigation of Orographic Venting of Atmospheric Boundary Layer Air Using Observations and the WRF-Chem Model., **1**, 05.

Bao, J., Michelson, S., Persson, P., Djalalova, I. & Wilczak, J. (2008) Observed and WRF-simulated low-level winds in a high-ozone episode during the Central California Ozone Study. *Journal of Applied Meteorology & Climatology*, **47**.

Barbour, M.G. & Minnich, R.A. (1988) Califomian Upland Forests and Woodlands. *North American terrestrial vegetation*, 131-164.

Bech, J., Sairouni, A., Codina, B., Lorente, J. & Bebbington, D. (2000) Weather radar anaprop conditions at a Mediterranean coastal site. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, **25**, 829-832.

Bianco, L., Djalalova, I., King, C. & Wilczak, J. (2011) Diurnal evolution and annual variability of boundary-layer height and its correlation to other meteorological variables in California's Central Valley. *Boundary-Layer Meteorology*, **140**, 491-511.

Cavazos, T. (1994) The influence of the southern oscillation on the winter climate of Nuevo Leon state, Mexico. *Geofísica Internacional*, **33**.

Chang, P. & Lin, P. (2011) Radar Anomalous Propagation Associated with Foehn Winds Induced by Typhoon Krosa (2007). *Journal of Applied Meteorology & Climatology*, **50**.

Choi, W., Faloona, I., McKay, M., Goldstein, A. & Baker, B. (2011) Estimating the atmospheric boundary layer height over sloped, forested terrain from surface spectral analysis during BEARPEX. *Atmospheric Chemistry and Physics*, **11**, 6837-6853.

Collett Jr, J.L., Daube Jr, B.C., Gunz, D. & Hoffmann, M.R. (1990) Intensive studies of Sierra Nevada cloudwater chemistry and its relationship to precursor aerosol and gas concentrations. *Atmospheric Environment.Part A.General Topics*, **24**, 1741-1757.

Dai, A. (2006) Recent climatology, variability, and trends in global surface humidity. *Journal of Climate*, **19**, 3589-3606.

Daly, C., Neilson, R.P. & Phillips, D.L. (1994) A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology*, **33**, 140-158.

De Young, R.J., Grant, W.B. & Severance, K. (2005) Aerosol transport in the California Central Valley observed by airborne lidar. *Environmental science & technology*, **39**, 8351-8357.

Dettinger, M.D., Cayan, D.R., Meyer, M.K. & Jeton, A.E. (2004) Simulated hydrologic responses to climate variations and change in the Merced, Carson, and American River basins, Sierra Nevada, California, 1900–2099. *Climatic Change*, **62**, 283-317.

Fabry, F. (2006) The spatial variability of moisture in the boundary layer and its effect on convection initiation: Project-long characterization. *Monthly Weather Review*, **134**.

Fast, J.D., Gustafson Jr, W., Berg, L.K., Shaw, W.J., Pekour, M., Shrivastava, M., Barnard, J.C., Ferrare, R., Hostetler, C.A. & Hair, J. (2012) Transport and mixing patterns over Central California during the carbonaceous aerosol and radiative effects study (CARES). *Atmospheric Chemistry and Physics*, **12**, 1759-1783.

Feld, S.I., Cristea, N.C. & Lundquist, J.D. (2013) Representing atmospheric moisture content along mountain slopes: Examination using distributed sensors in the Sierra Nevada, California. *Water Resources Research*.

Fitzjarrald, D.R. (1986) Slope winds in Veracruz. *Journal of climate and applied meteorology*, **25**, 133-144.

Galvão, L.S., dos Santos, J.R., Roberts, D.A., Breunig, F.M., Toomey, M. & de Moura, Y.M. (2011) On intra-annual EVI variability in the dry season of tropical forest: A case study with MODIS and hyperspectral data. *Remote Sensing of Environment*, **115**, 2350-2359.

Garcia-Garcia, F. & Zarraluqui, V. (2008) A fog climatology for Mexico. *Die Erde*, **139**, 45-59.

Goulden, M., Anderson, R., Bales, R., Kelly, A., Meadows, M. & Winston, G. (2012) Evapotranspiration along an elevation gradient in California's Sierra Nevada. *Journal of Geophysical Research: Biogeosciences (2005–2012)*, **117**.

Hall, C.A. (1991) *Natural History of the White-Inyo Range, Eastern California*, Univ of California Press.

Holets, S. & Swanson, R.N. (1981) High-inversion fog episodes in central California. *Journal of Applied Meteorology*, **20**, 890-899.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X. & Ferreira, L.G. (2002) Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, **83**, 195-213.

Iacobellis, S.F., Norris, J.R., Kanamitsu, M., Tyree, M. & Cayan, D.R. (2009) Climate variability and California low-level temperature inversions. *California Climate Change Center*.

Iacobellis, S.F., Norris, J.R., Kanamitsu, M., Tyree, M. & Cayan, D.R. (2009) Climate variability and California low-level temperature inversions. *California Climate Change Center*.

Jeong, S., Zhao, C., Andrews, A.E., Bianco, L., Wilczak, J.M. & Fischer, M.L. (2012) Seasonal variation of CH4 emissions from central California. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **117**.

Jeong, S., Hsu, Y., Andrews, A.E., Bianco, L., Vaca, P., Wilczak, J.M. & Fischer, M.L. (2013) A multitower measurement network estimate of California's methane emissions. *Journal of Geophysical Research: Atmospheres*, **118**, 11,339-11,351.

Jeton, A.E., Dettinger, M.D. & Smith, J.L. (1996) *Potential effects of climate change on streamflow, eastern and western slopes of the Sierra Nevada, California and Nevada*, US Department of the Interior, US Geological Survey.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G. & Woollen, J. (1996) The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77**, 437-471.

L COLLETT, J., C DAUBE, B. & Hoffmann, M.R. (1991) Spatial and temporal variations in precipitation and cloud interception in the Sierra Nevada of central California. *Tellus B*, **43**, 390-400.

Laurence, H., Fabry, F., Dutilleul, P., Bourgeois, G. & Zawadzki, I. (2002) Estimation of the spatial pattern of surface relative humidity using ground based radar measurements and its application to disease risk assessment. *Agricultural and Forest Meteorology*, **111**, 223-231.

Lewis, J., De Young, R., Ferrare, R. & Allen Chu, D. (2010) Comparison of summer and winter California central valley aerosol distributions from lidar and MODIS measurements. *Atmospheric Environment*, **44**, 4510-4520.

Lundquist, J.D., Neiman, P.J., Martner, B., White, A.B., Gottas, D.J. & Ralph, F.M. (2008) Rain versus snow in the Sierra Nevada, California: Comparing Doppler profiling radar and surface observations of melting level. *Journal of Hydrometeorology*, **9**.

Panek, J., Saah, D., Esperanza, A., Bytnerowicz, A., Fraczek, W. & Cisneros, R. (2013) Ozone distribution in remote ecologically vulnerable terrain of the southern Sierra Nevada, CA. *Environmental Pollution*, **182**, 343-356.

Ryu, Y., Baldocchi, D.D., Kobayashi, H., Ingen, C., Li, J., Black, T.A., Beringer, J., Gorsel, E., Knohl, A. & Law, B.E. (2011) Integration of MODIS land and atmosphere products with a coupled-process model to estimate gross primary productivity and evapotranspiration from 1 km to global scales. *Global Biogeochemical Cycles*, **25**.

Schaaf, C.B., Liu, J., Gao, F. & Strahler, A.H. (2011) *Aqua and Terra MODIS albedo and reflectance anisotropy products. Land Remote Sensing and Global Environmental Change*Anonymous pp. 549-561. Springer.

Thornton, P.E., Running, S.W. & White, M.A. (1997) Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology*, **190**, 214-251.

Tucker, C.J. (1979) Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, **8**, 127-150.

Twisselmann, E.C. (1967) A flora of Kern county, California. *Wasmann Journal of Biology*, **25**.

Van Ooy, D.J. & Carroll, J.J. (1995) The spatial variation of ozone climatology on the western slope of the Sierra Nevada. *Atmospheric Environment*, **29**, 1319-1330.

Velpuri, N., Senay, G., Singh, R., Bohms, S. & Verdin, J. (2013) A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET. *Remote Sensing of Environment*, **139**, 35-49.

Wan, Z. & Dozier, J. (1996) A generalized split-window algorithm for retrieving land-surface temperature from space. *Geoscience and Remote Sensing, IEEE Transactions on*, **34**, 892-905.

Wan, Z. (2008) New refinements and validation of the MODIS land-surface temperature/emissivity products. *Remote Sensing of Environment*, **112**, 59-74.

Wang, W., Liang, S. & Meyers, T. (2008) Validating MODIS land surface temperature products using long-term nighttime ground measurements. *Remote Sensing of Environment*, **112**, 623-635.

Weckwerth, T.M., Pettet, C.R., Fabry, F., Park, S., LeMone, M.A. & Wilson, J.W. (2005) Radar refractivity retrieval: Validation and application to short-term forecasting. *Journal of Applied Meteorology*, **44**.

Tables

Table 1. MODIS satellite data products evaluated.

Short Name	MODIS Data Product	Spatial Resolution (m)	Temporal Granularity
MOD05/MYD05	Total Precipitable Water	1000 (Near-Infrared, Day) and 5000 (Thermal, Day and Night)	Daily
MOD07/MYD07	Atmospheric Profile	5000	Daily
MOD09GA/MYD09GA	Surface Reflectance	500 (bands 1-7)	Daily
MOD11A2/MYD11A2	Land Surface Temperature and Emissivity	1000	8-day
MOD13/MYD13	Vegetation Indices	500	Daily
MOD16	Evapotranspiration	1000	Year
MCD43B4	Nadir BRDF-Adjusted Reflectance	1000	16-day

Table 2. Sierra Nevada Critical Zone Observatory micrometeorological flux tower sites.

Tower	Elevation (m)	Latitude	Longitude
San Joaquin Experimental Range	405	37.109	-119.731
Soaproot Saddle	1160	37.031	-119.256
Providence Creek	2015	37.067	-119.195
Shorthair Creek	2700	37.067	-118.987

Figures

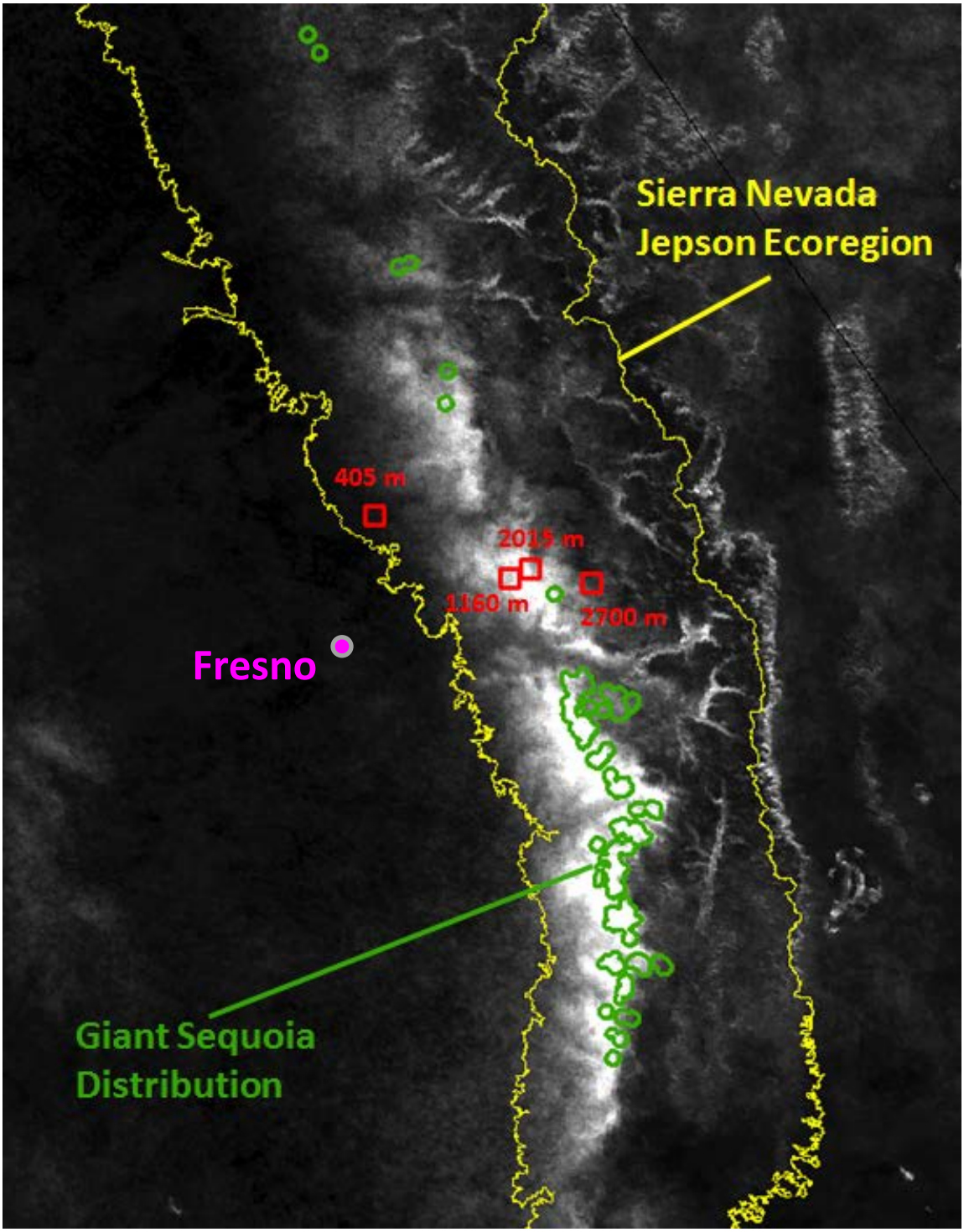


Figure 1. Southern Sierra Nevada Critical Zone Observatory biometeorological towers elevational transect with respect to giant sequoia distribution and April cloud formation (SCF) frequency.
The 2015 meter tower is at the upper and northern end of frequent cloud formation, but well within typical elevations of giant sequoia.

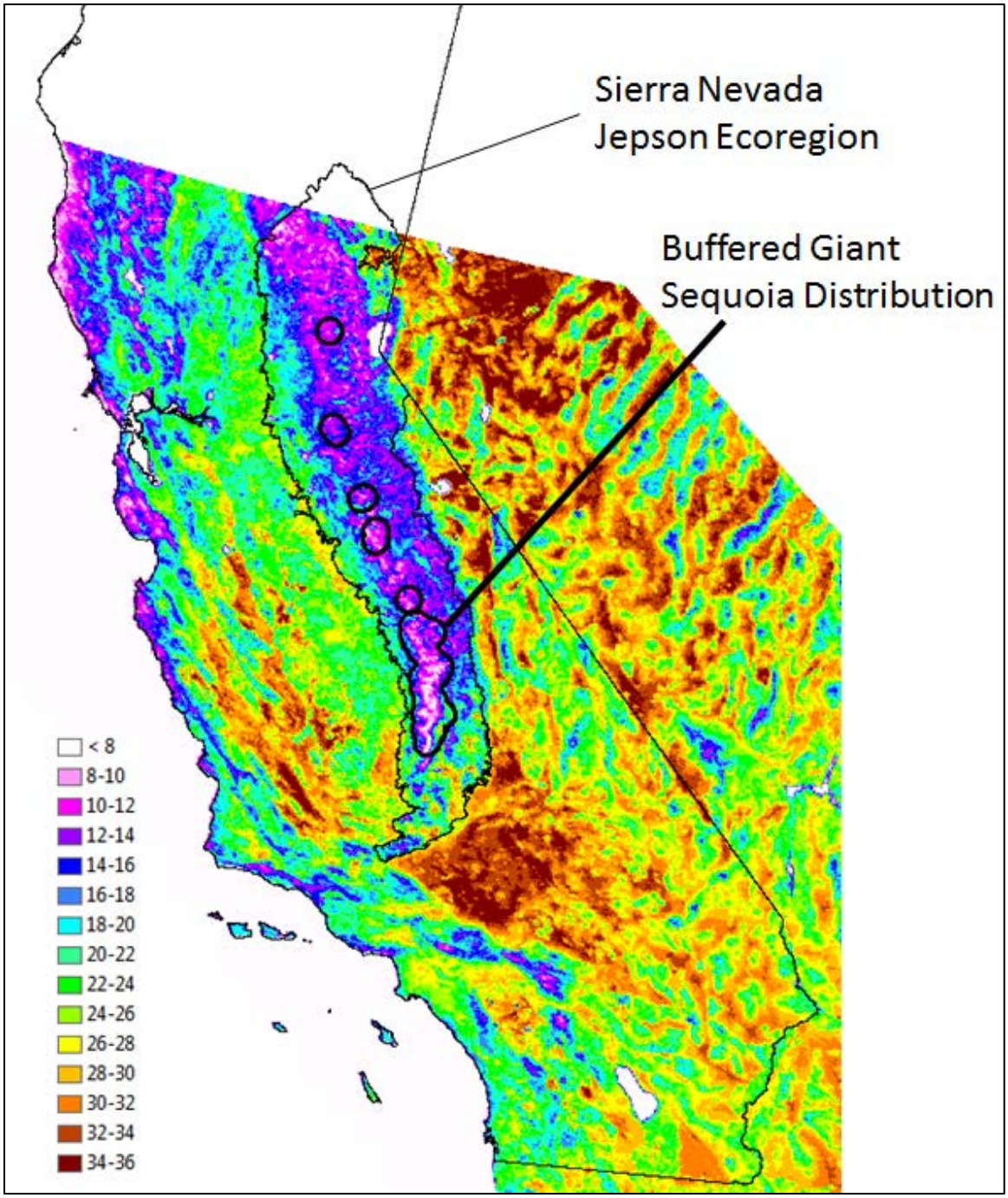


Figure 2. Average MODIS Aqua diel range in land surface temperature (°C), 2012 – classified. Portion of MODIS tile H08,V05 within California and adjacent areas. Buffered giant sequoia distribution in thick black line, along with Sierra Nevada ecoregion in thin black line. Southern portions of the sequoia belt have among the lowest average range of temperature in the state. Data are for clear days only, so results may be biased high, due to the moderating effects of clouds.

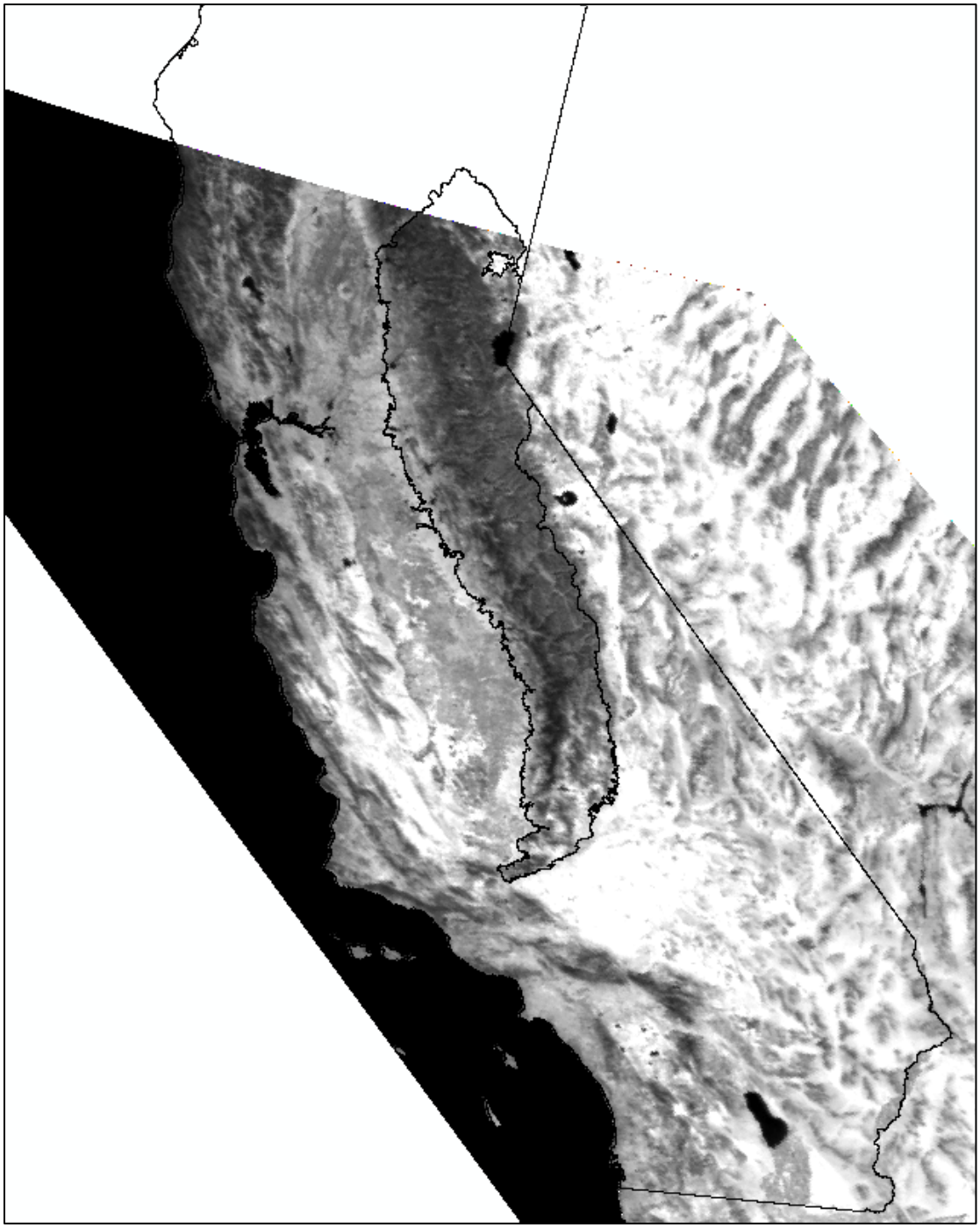


Figure 3. Average MODIS Aqua diel range in land surface temperature (°C), 2012 - grayscale.
Same as Figure 2 but continuous grayscale. Grayscale better conveys appearance of strong relationship (negative correlation) with vegetation density, as this looks similar to a visible band image.

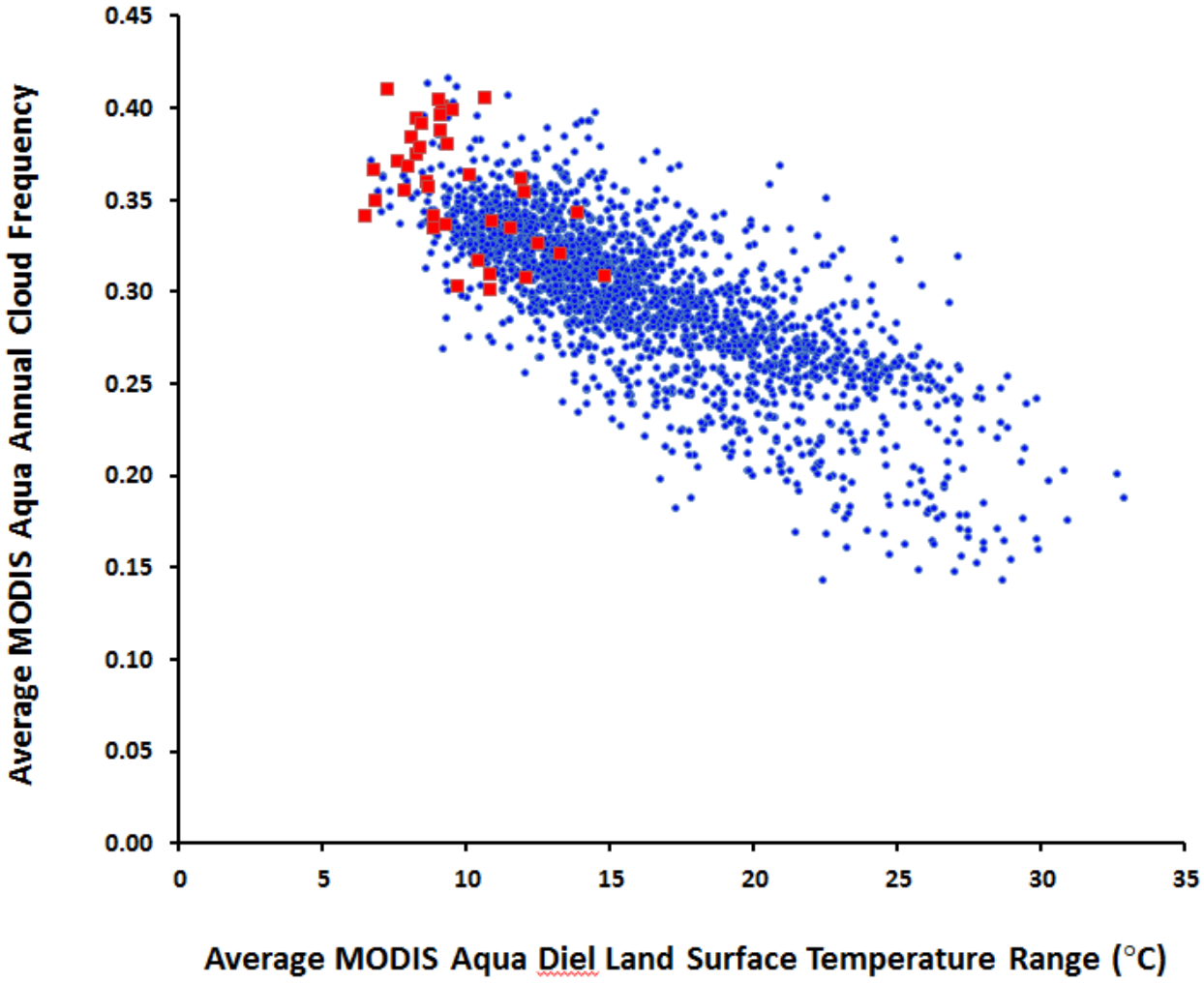


Figure 4. MODIS Aqua average afternoon annual cloud frequency (2003-11) vs. average diel range in MODIS Aqua land surface temperature (°C), 2012.
Random points selected within giant sequoia groves (red boxes) and throughout the Sierra Nevada (blue points).

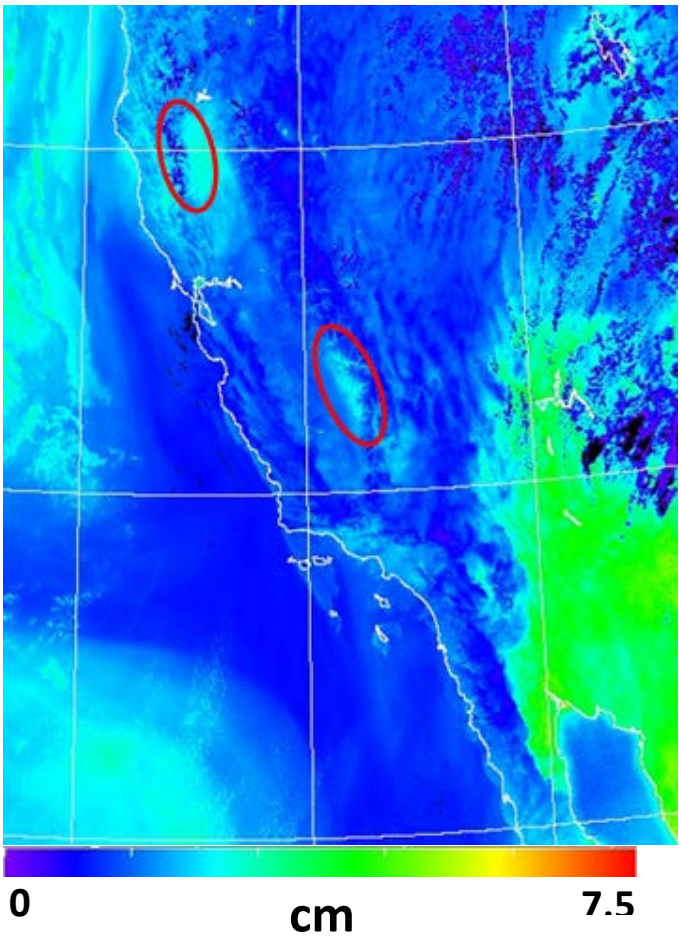


Figure 5. MODIS Aqua near-infrared precipitable water vapor (MYD05), California and adjacent areas: August 19, 2011

Red ellipses highlight steep water vapor gradient areas of northwestern Sacramento Valley (north) and eastern San Joaquin Valley (south), each with apparent "pooling" of water vapor in foothills. Water vapor is integrated across the atmosphere, so values low over mountains and areas of high clouds.

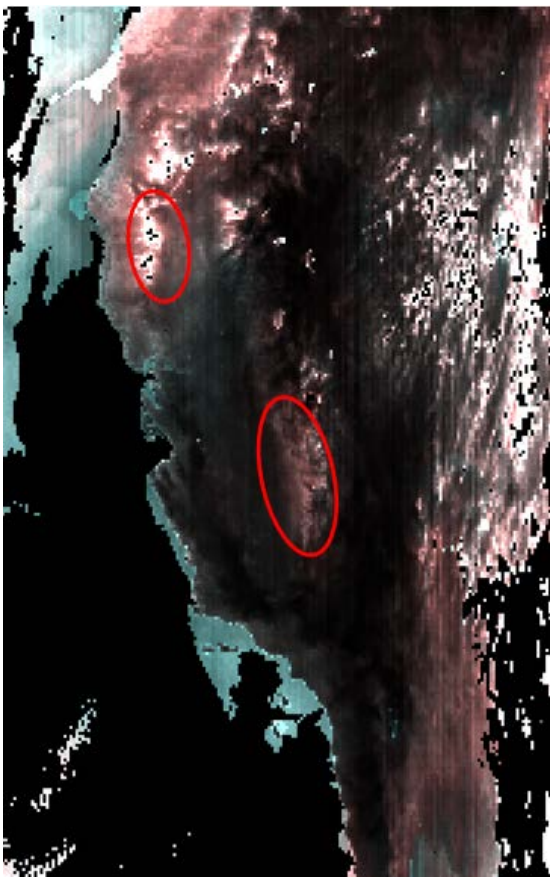


Figure 6. MODIS Aqua atmospheric vertical profile (MYD07) relative humidity: August 19, 2011

Same areas as Figure 5 on the same date appear much different in a humidity product. Higher elevations generally have higher values (note frequent reversal from water vapor product), but foothills of eastern San Joaquin Valley (sequoia belt) still high.

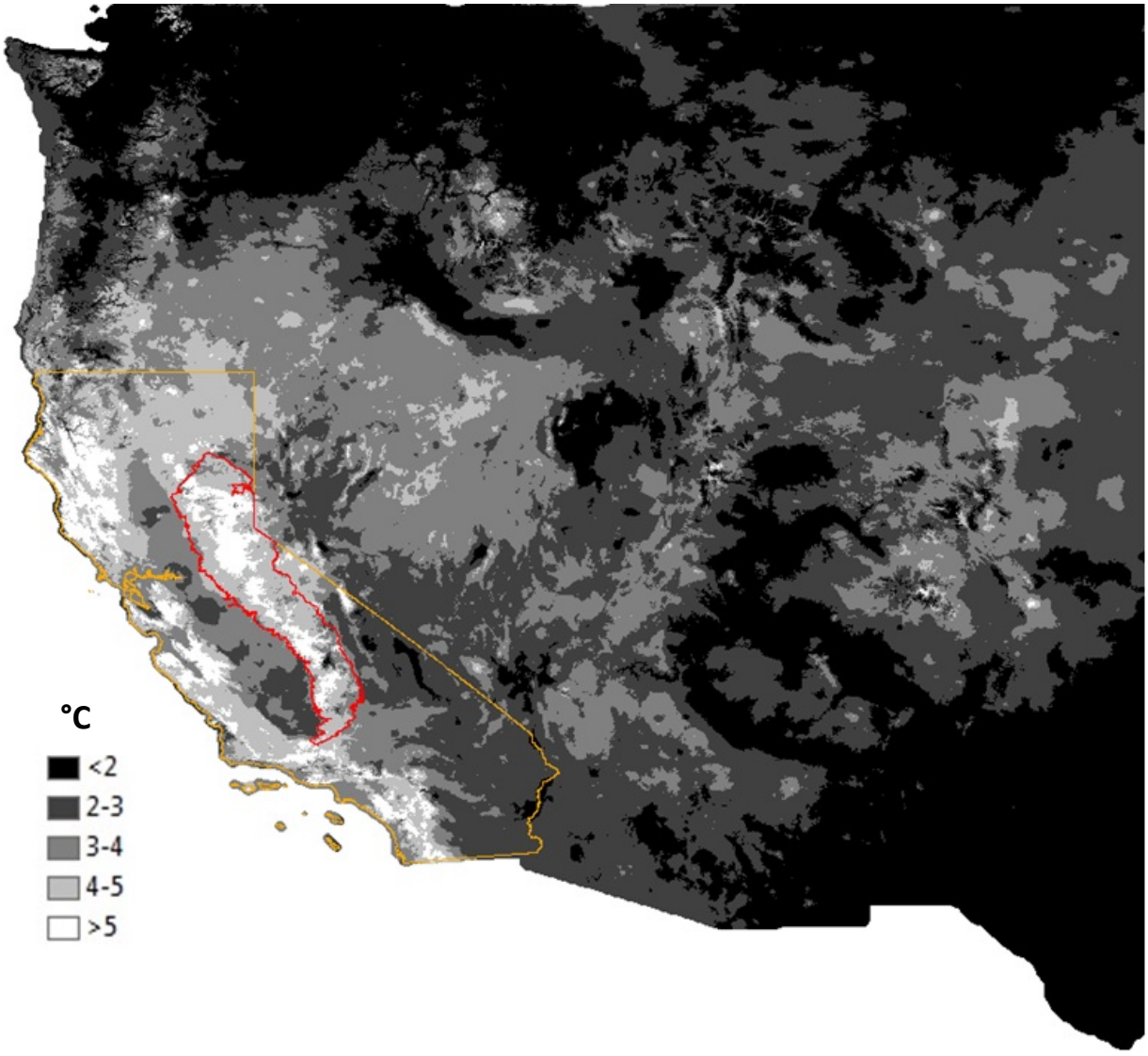


Figure 7. PRISM October maximum temperature minus April maximum temperature (°C). California outlined in orange and the Sierra Nevada Jepson Ecoregion outlined in red. California unusually cold in April compared with much of the United States, and particularly high values for California coast ranges and west slope of the Sierra Nevada indicate April much colder than October.

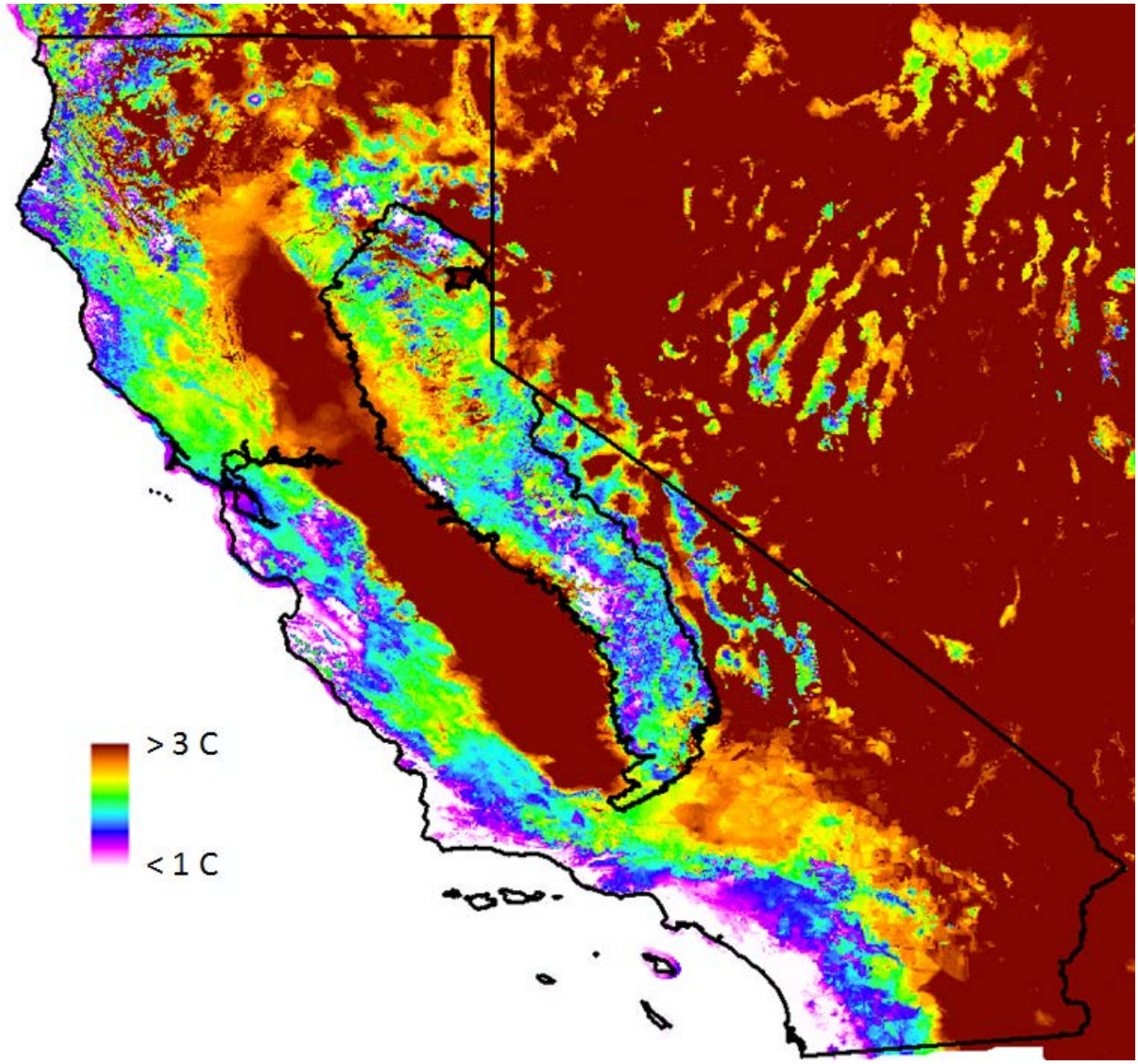


Figure 8. PRISM March maximum temperature minus February maximum temperature (°C).
Most areas tend to warm substantially between February and March (brown areas greater than 3 °C). Coastal California is a general exception (white areas less than 1 °C), warming very little due to the moderating influence of the cool ocean. The west slope of the southern Sierra Nevada joins the coastal areas in having very little warming in March, likely due to the impact of spring cloud cover. The patterns throughout the Sierra Nevada are generally quite noisy, however, likely due to the paucity of climate stations in the Sierra Nevada, and the complicating effects of the rugged topography.

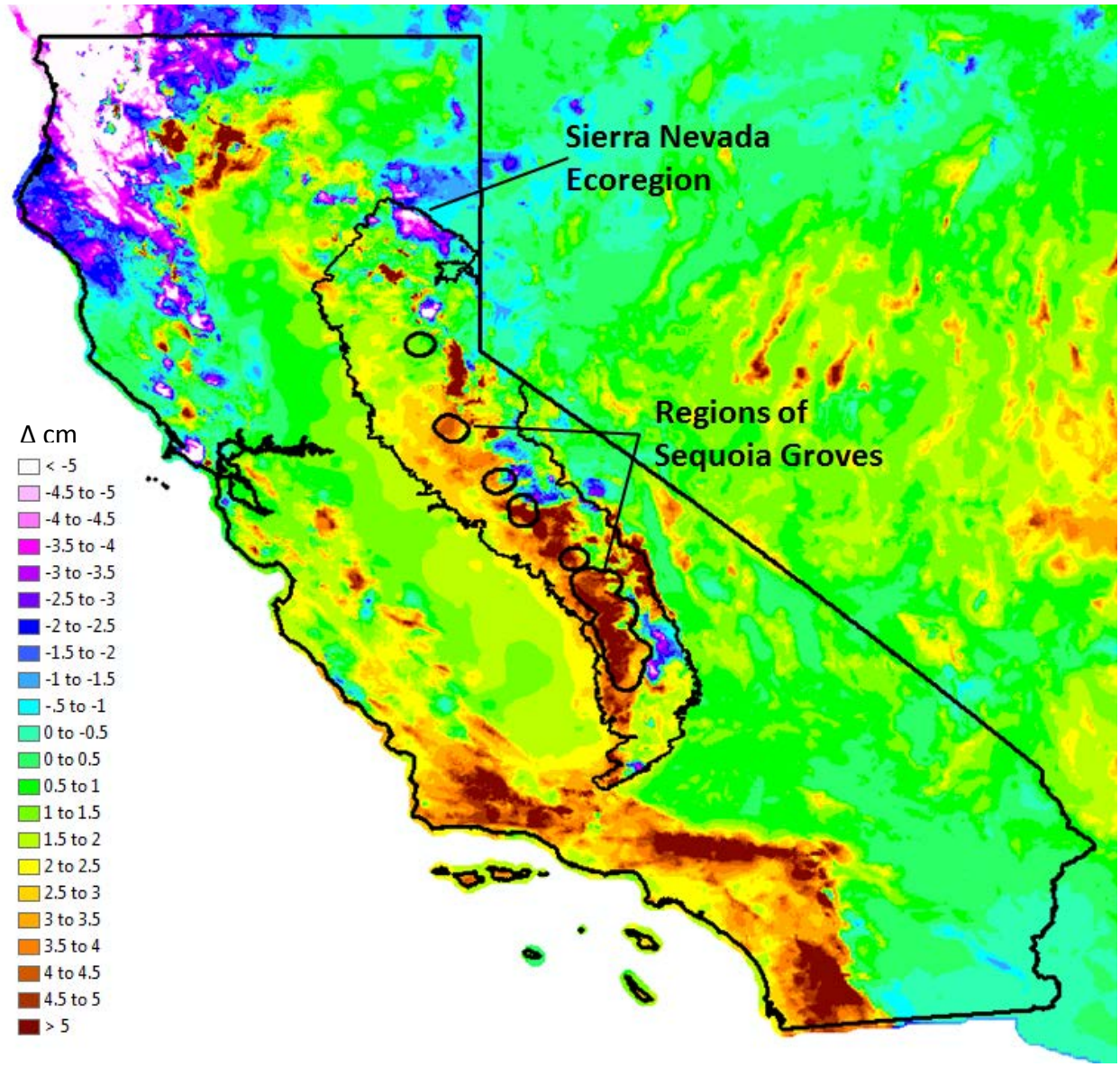


Figure 9. PRISM average March precipitation minus average December precipitation (cm)

Much of the western United States has a winter peak in precipitation but the timing can differ. A large portion has comparable precipitation in December and March (small difference areas are shades of green). The Pacific Northwest tends to have its precipitation skewed earlier in the "rain season", with much higher values in December than in March (white areas). Southern California mountains and the west slope of the southern Sierra Nevada are notably skewed toward precipitation later in the "rain season", with much higher values in March (brown areas).

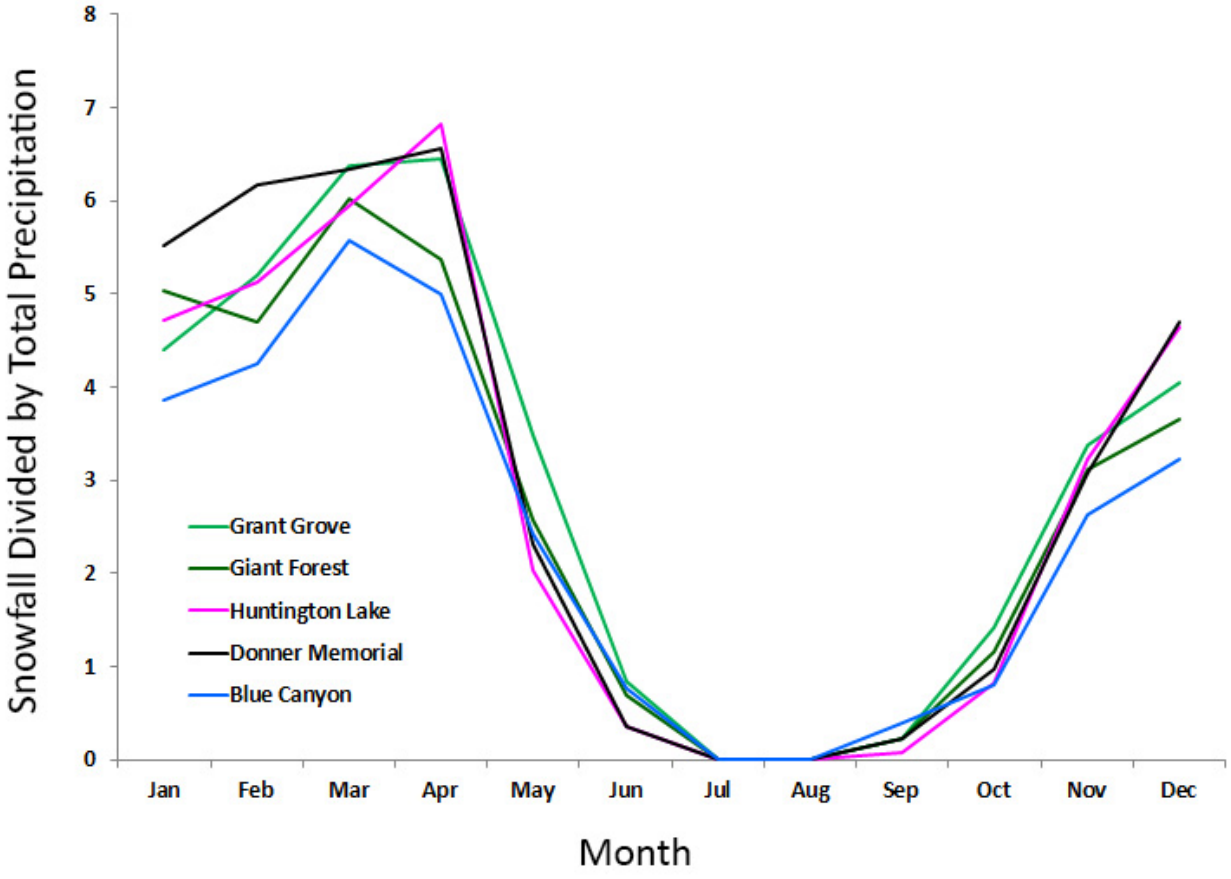


Figure 10. Monthly snowfall divided by precipitation, Sierra Nevada mid-elevation RAWS sites

March and April have a relatively high amount of snowfall relative to total precipitation at a variety of Sierra Nevada sites, including two sequoia groves. This could indicate a higher proportion of snow events to rain events, but it is more likely related to a colder and drier snow.

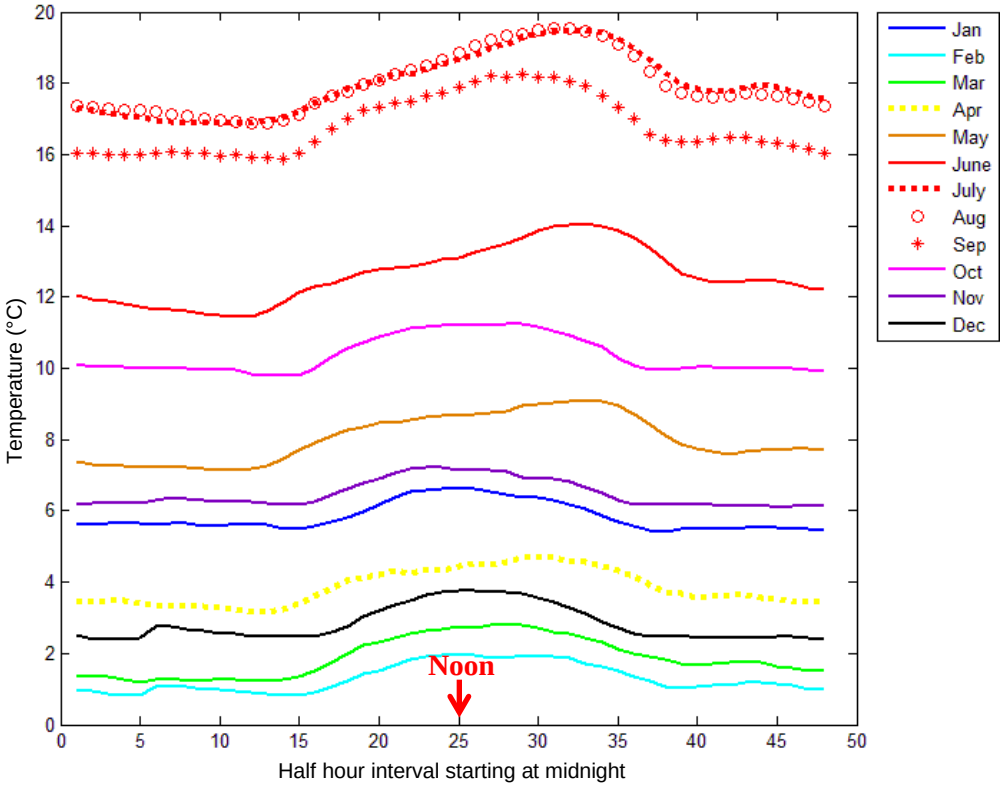


Figure 11. Average 1/2 hourly temperature by month: 2015 meter tower.
This site has a very low average range in diel temperature, of 1-2 °C.

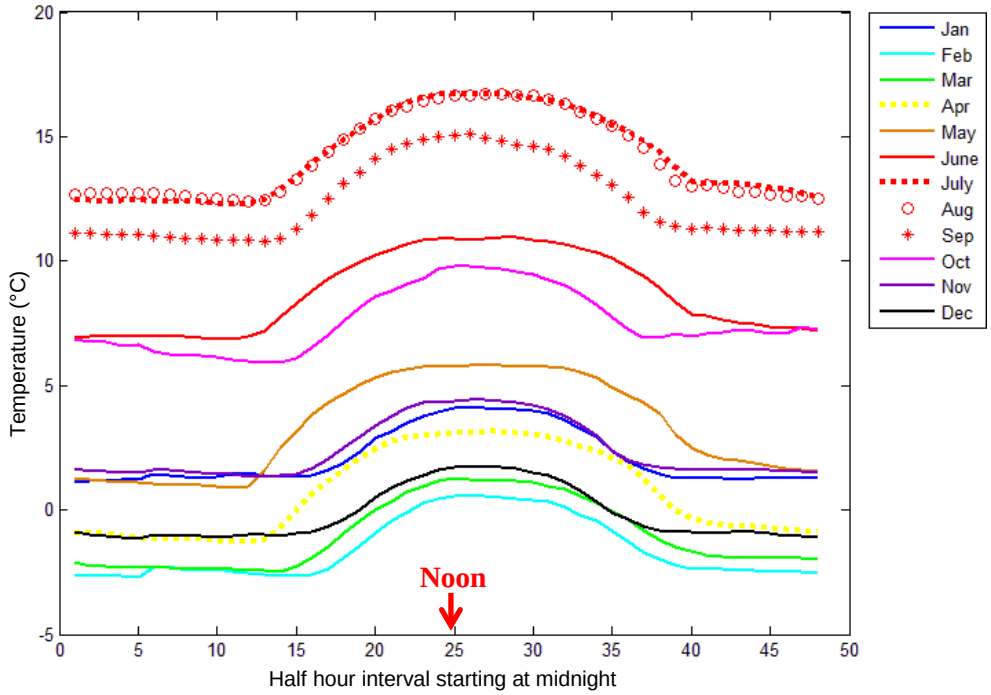


Figure 12. Average 1/2 hourly temperature by month: 2700 meter tower.
Site has a slightly higher range in diel temperature at 2700 m, but still only 2-4 °C.

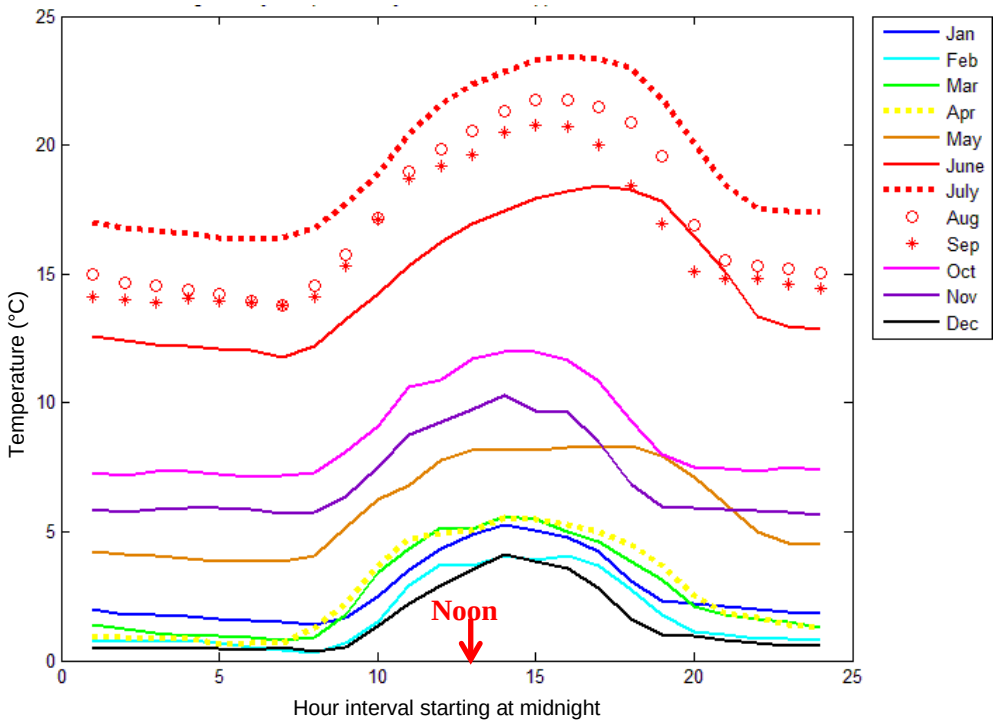


Figure 13. Average hourly temperature by month: surface measurement, 2015 meters.
Surface has higher ranges in temperature (3-8 °C) than tower, especially in the summer months.

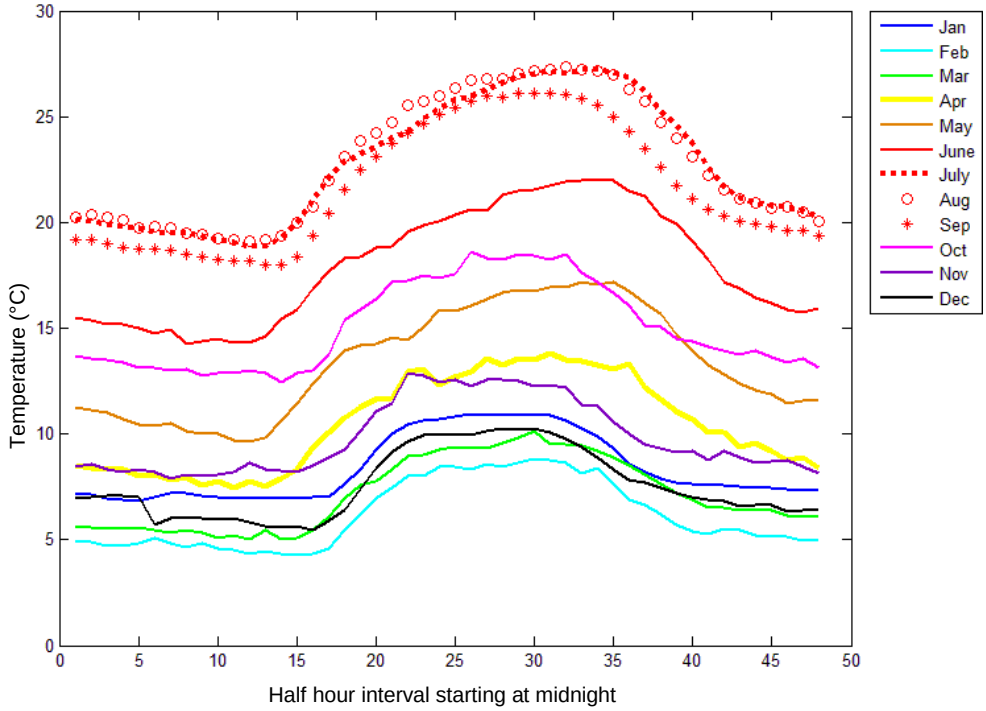


Figure 14. Average 1/2 hourly temperature by month: 1160 meter tower.
Much larger range in temperature (3-8 °C) at this tower, more like surface measurement above.

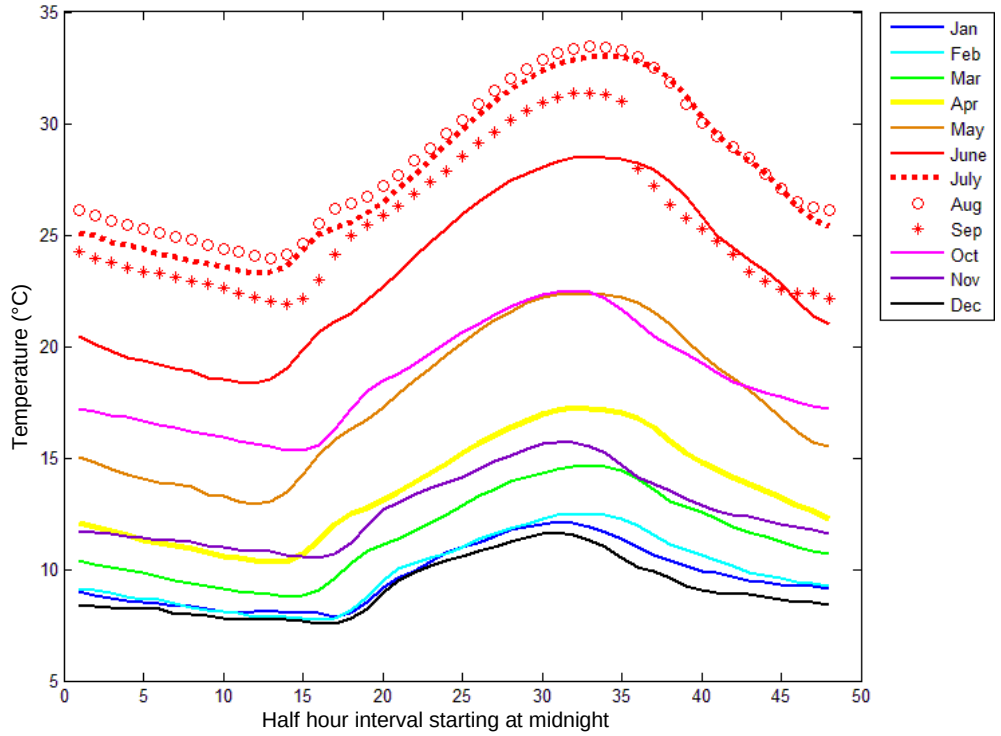


Figure 15. Average 1/2 hourly temperature by month: 405 meter tower.
Higher ranges in temperatures here too. Winter months colder than spring months at this lower tower.

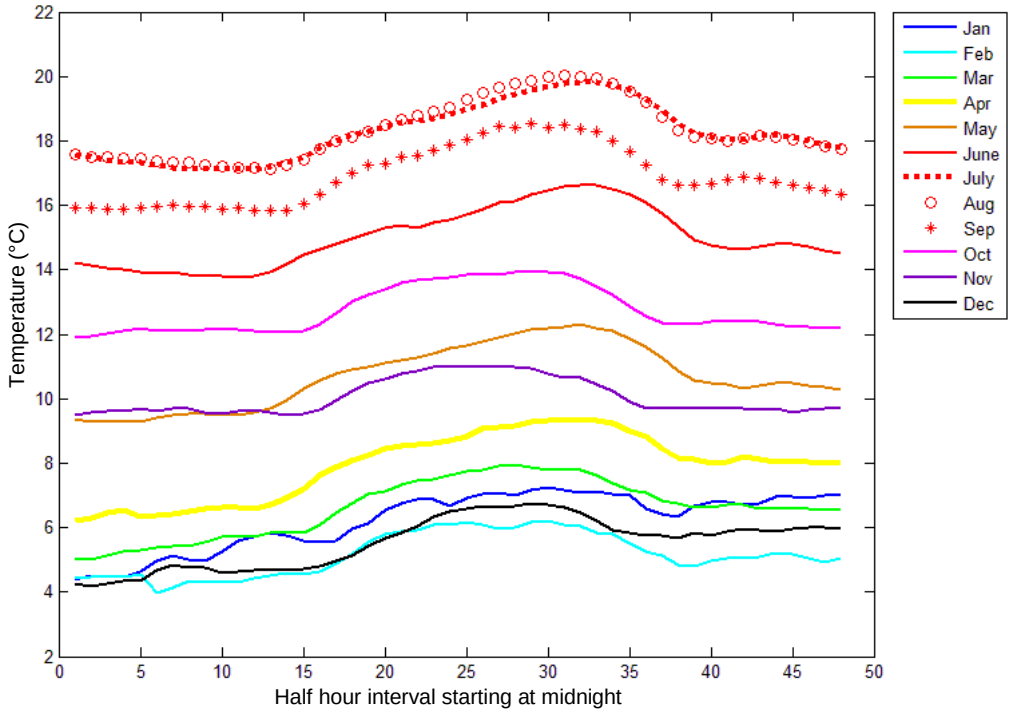


Figure 16. Average temperature by month: 2015 meter tower, days with clear afternoons.
Even days with clear afternoons have extremely low range in temperature, though spring now warmer.

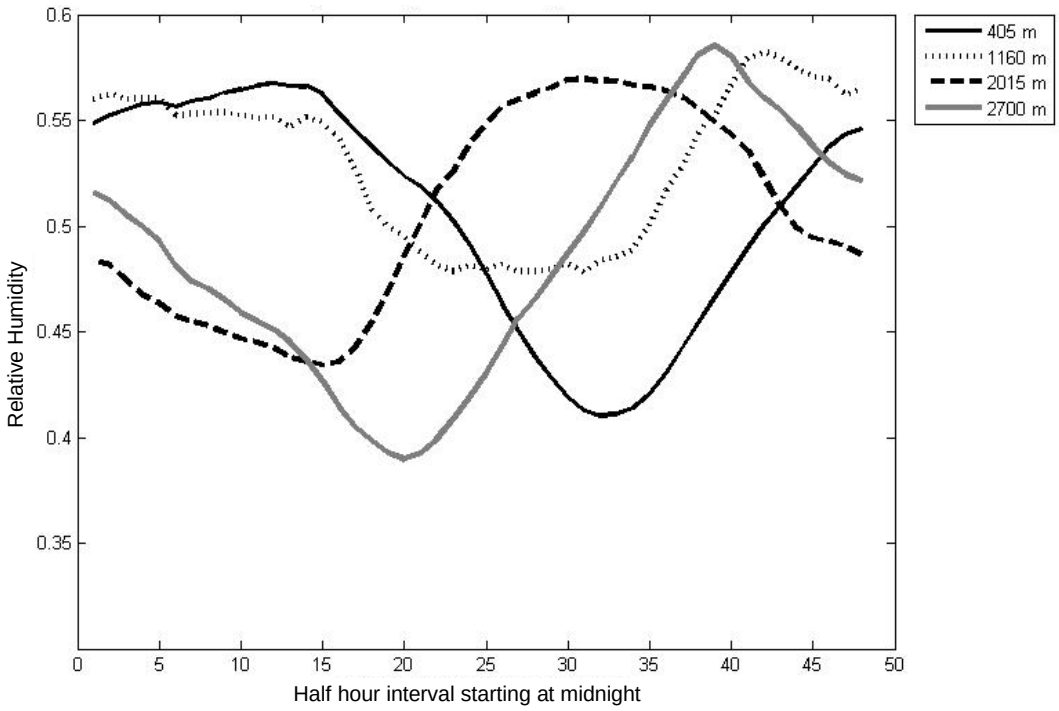


Figure 17. Annually averaged 1/2 hourly relative humidity across 4 tower elevational transect. Afternoon peak in afternoon at 2015 m highly unusual. Early evening peaks at others also notable.

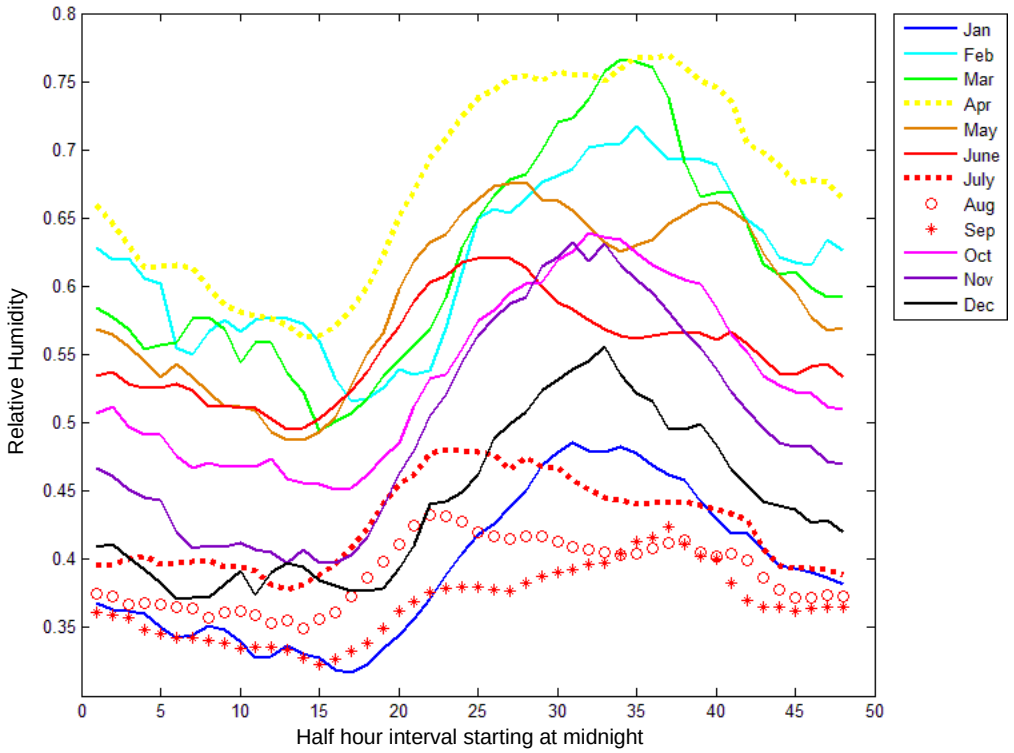


Figure 18. Average 1/2 hourly relative humidity by month: 2015 meter tower. Afternoon peaks throughout the year, but most pronounced in early spring months (Feb-April).

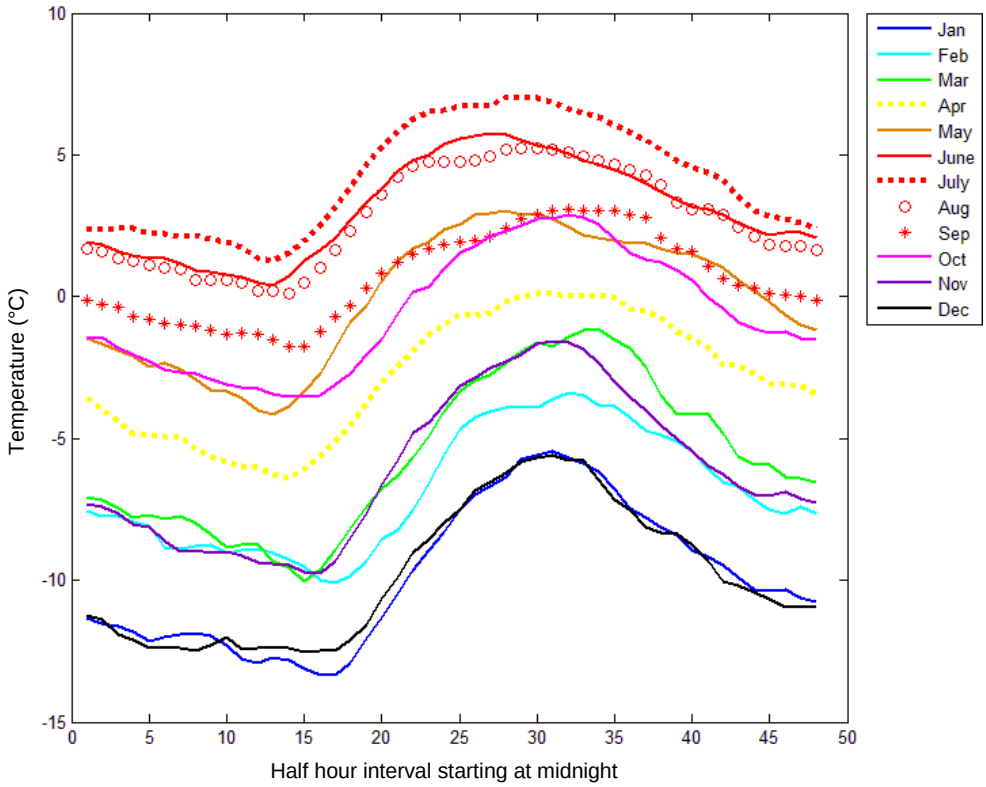


Figure 19. Average 1/2 hourly dew point temperature by month: 2015 meter tower.
Pronounced afternoon peaks in dew point, in contrast to small changes in air temperature.

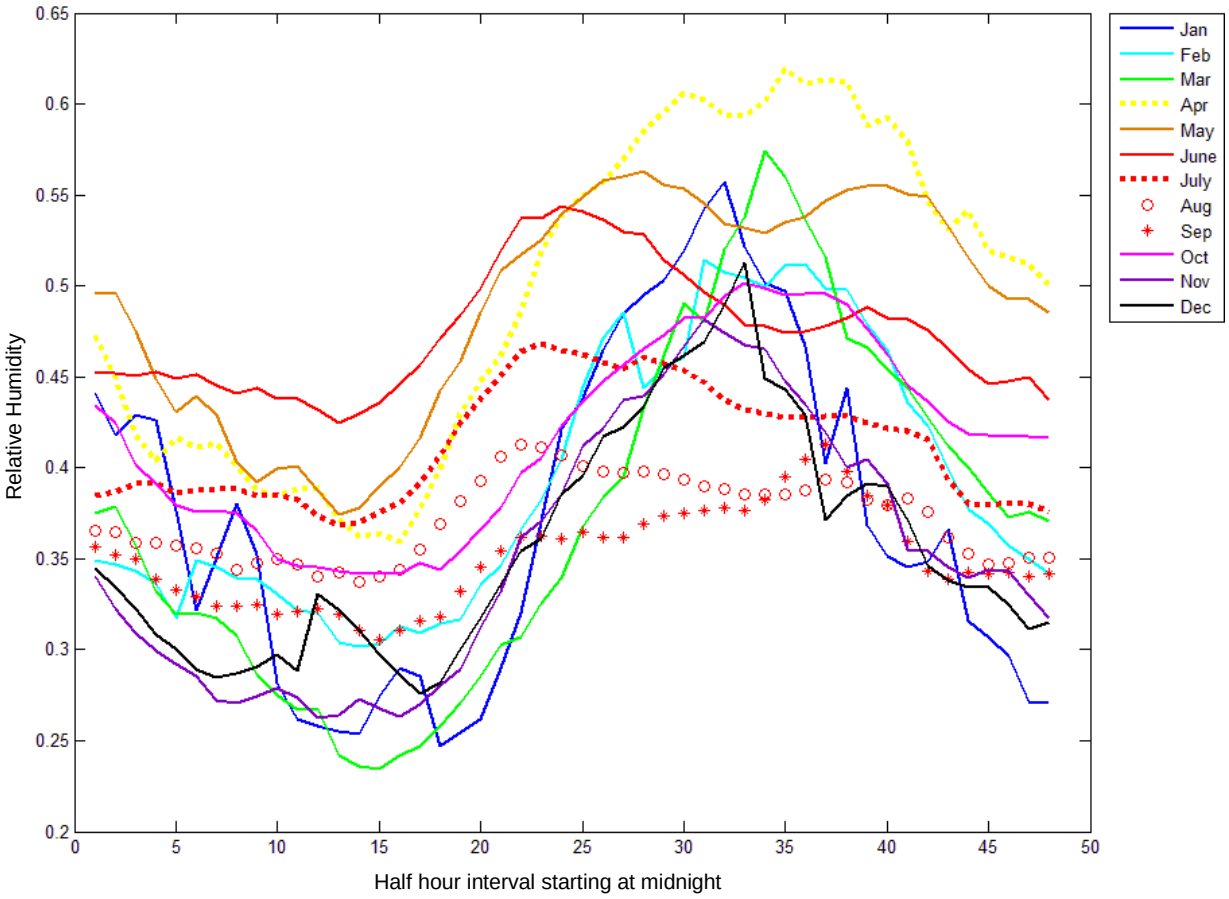


Figure 20. Relative humidity by month: 2015 meter tower, days with clear afternoons.
Even days with clear afternoons have afternoon peaks in relative humidity. More pronounced in late spring months (April-June), in comparison to early spring peak for all days
Figure 18.

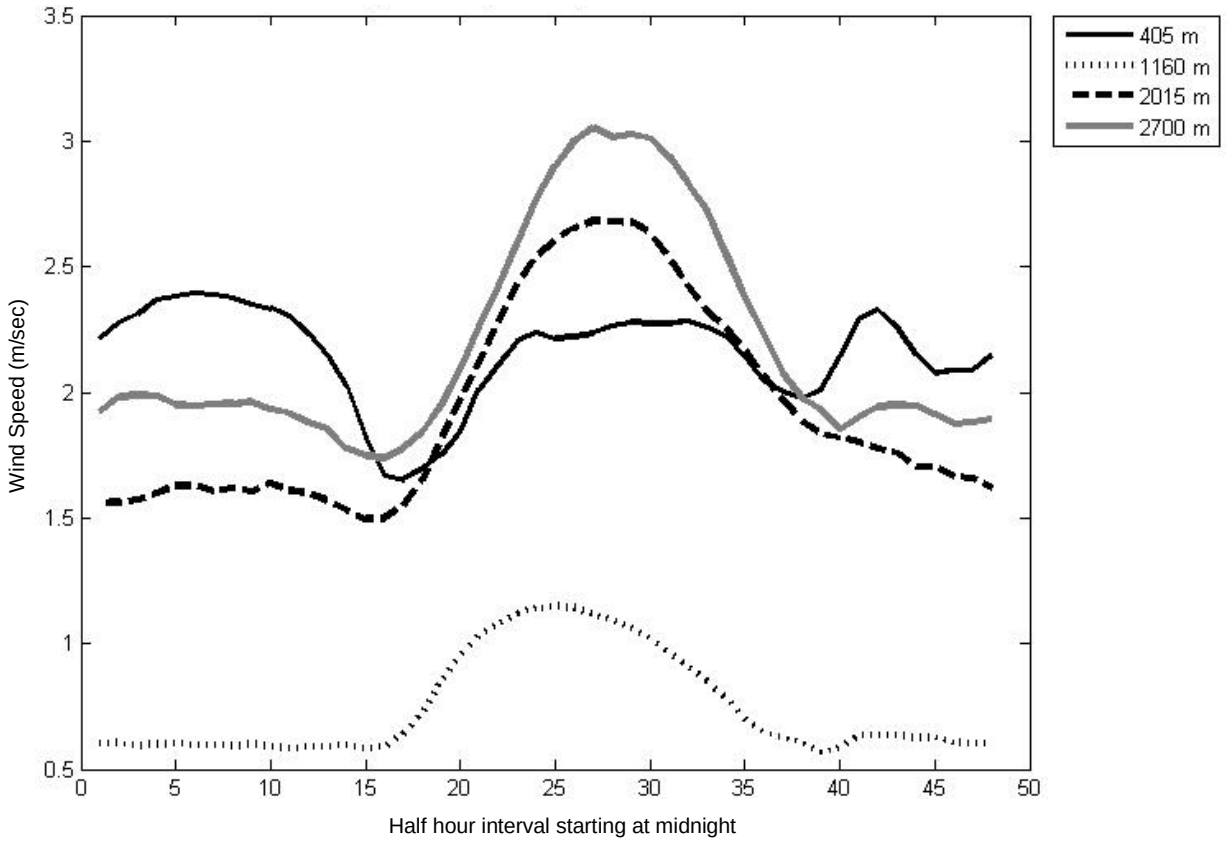


Figure 21. Annually averaged 1/2 hourly wind speed across 4 tower elevational transect of Sierra Nevada.

All sites have very low average wind speeds, with especially low winds at the 1160 meter site. Upper three sites have obvious early afternoon peaks, while the lower site has no pronounced peak.

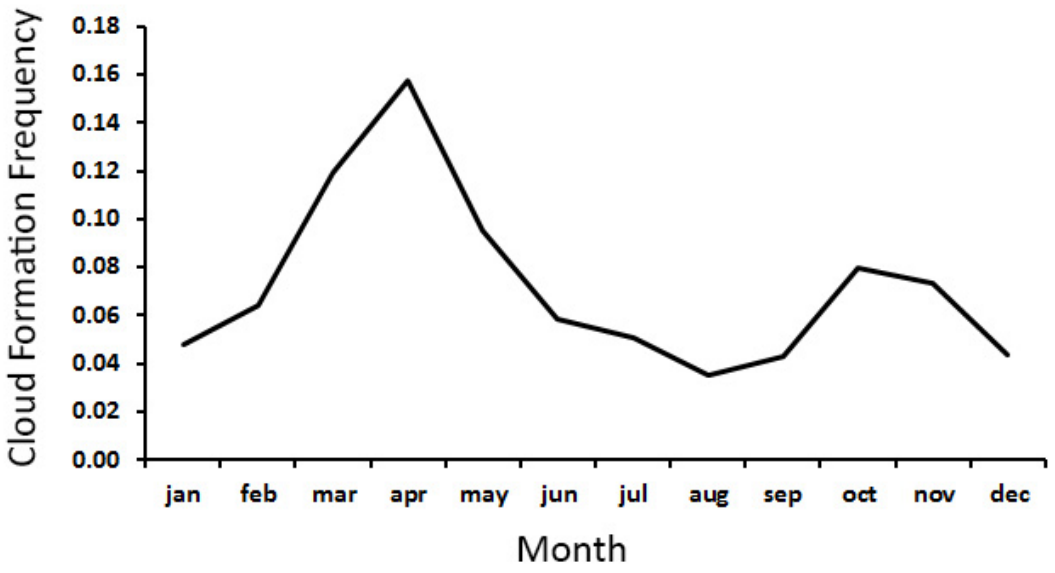


Figure 22. Average monthly frequency of "Sequoia Cloud Formation" within sequoia belt.
Spring peak in the particular cloud formation coincides with spring peaks in convective boundary layer height of adjacent areas of the Central Valley.

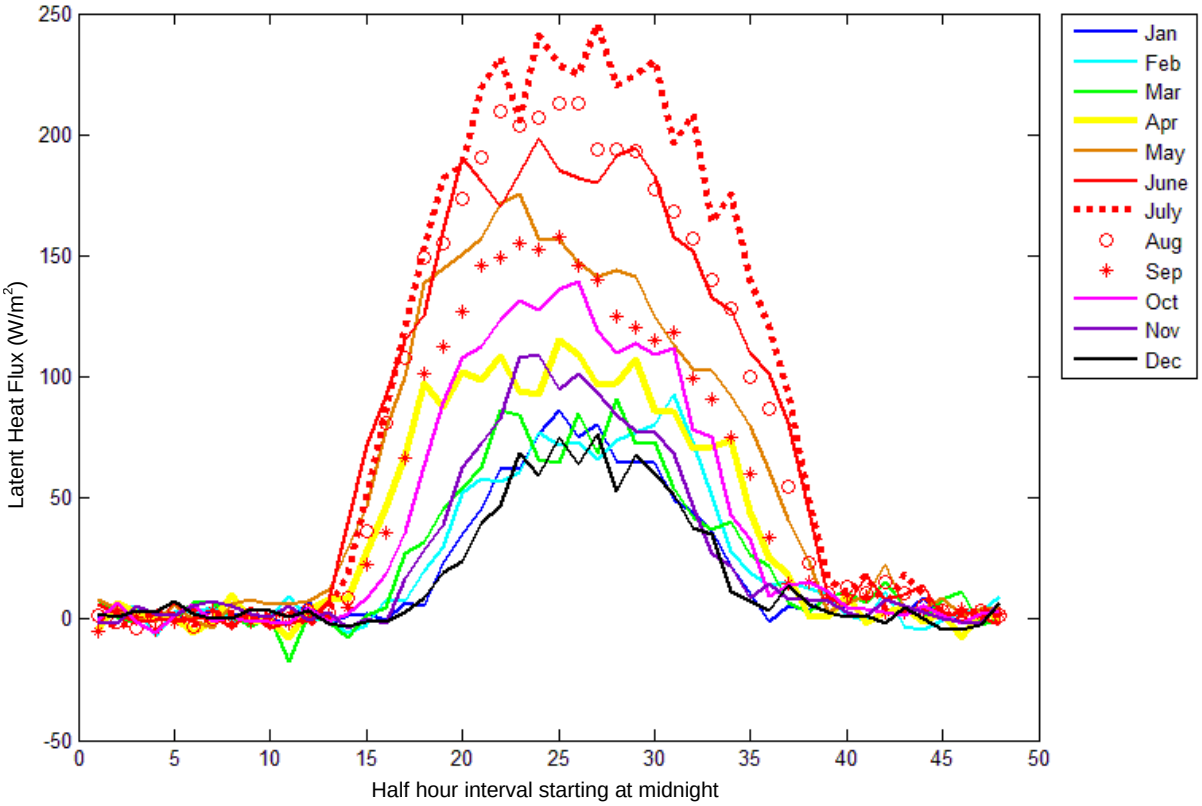


Figure 23. Average ½ hourly latent heat flux by month, 2015 meter tower.
Low spring values suggest local transpiration not source of spring clouds or low temperature range. Skew toward morning in spring months suggests afternoon clouds and humidity reducing evaporation.

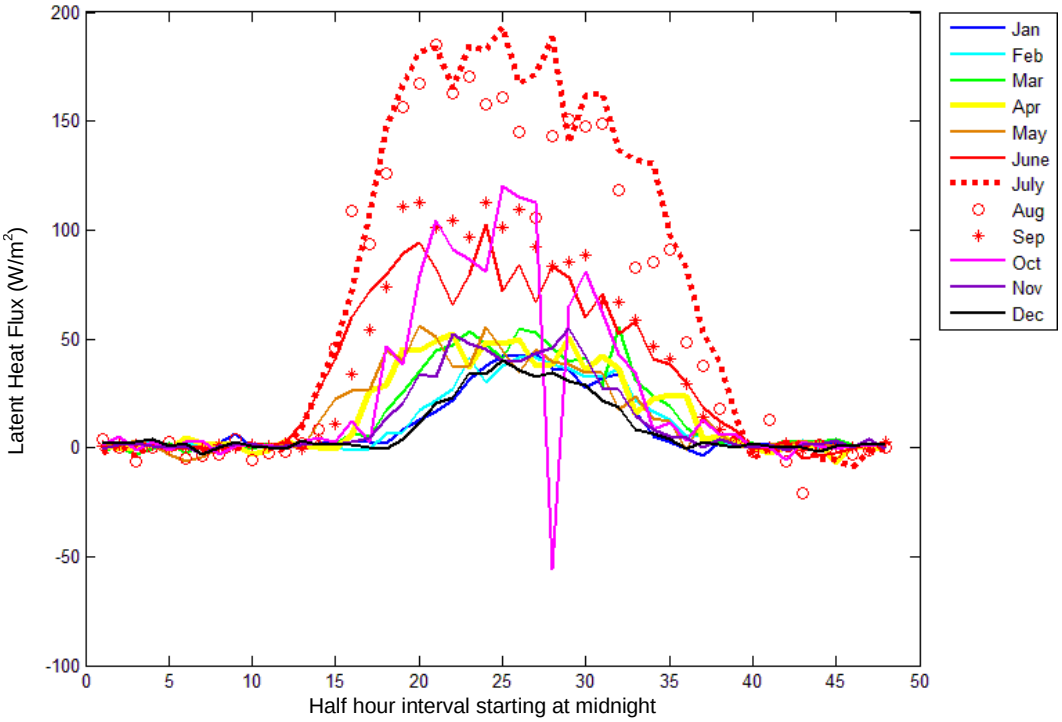


Figure 24. Average ½ hourly latent heat flux by month, 2700 meter tower.
Likely mostly sublimation in winter and spring. Skew toward morning in spring suggests cloud impact.

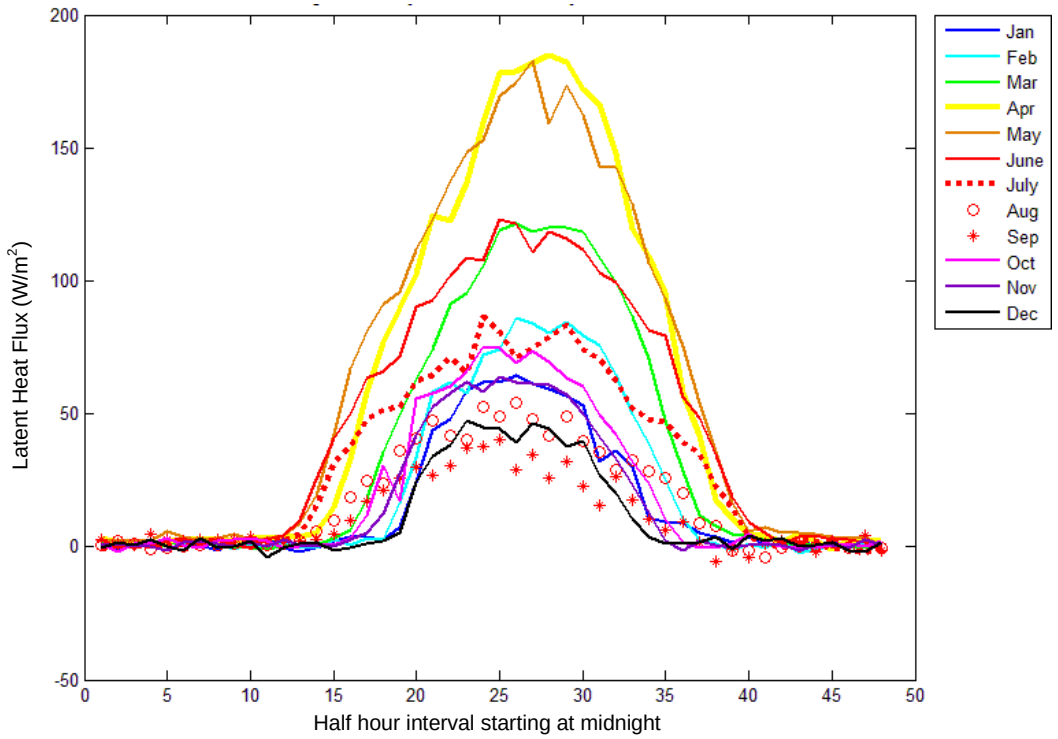


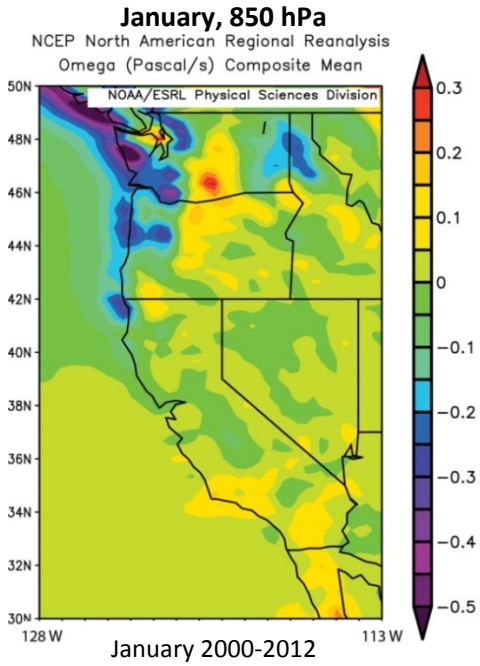
Figure 25. Average ½ hourly latent heat flux by month, 405 meter tower.
Skew toward afternoon suggests no reduction due to clouds or humidity at this lowest site.

Appendices

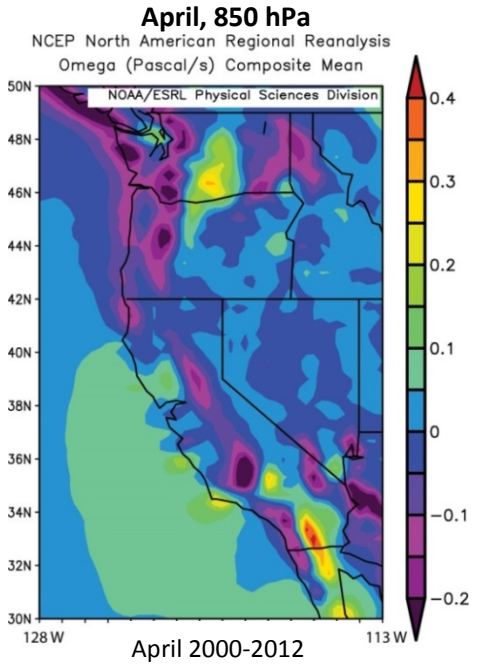
4.A.1. Averaging relative humidity values

There is an issue associated with averaging relative humidity values, because of the non-linear relationship between saturation vapor pressure and temperature. Technically, if dealing with the mixing of separate air parcels (and, for example, trying to determine if a mixing cloud will form) one should first convert relative humidity and temperature values to vapor pressure values, which can then be proportionally averaged along with temperature (essentially saturation vapor pressure). The (proportionally weighted) average of two relative humidity values (at different temperatures) is not the same as the average of the two vapor pressures divided by the average of the two different temperatures: it can be an underestimate (e.g., in the case of two non-saturated air parcels mixing to form a saturated air parcel, that is, a mixing cloud) or an overestimate, depending on the slope of the vapor pressure (y-axis) and temperature (x-axis) relationship between the two air parcels. If the slope is high it will be an underestimate, and if the slope is low it will be an overestimate. In the latter case, there will be greater overestimates for larger differences in relative humidity and for lower relative humidity values (because these are higher temperatures with greater differences in saturation vapor pressure for a given difference in relative humidity) and/or lower temperatures (where a given temperature difference has a greater effect on relative humidity for a given vapor pressure). But this does not mean that we should want the result that would be appropriate for the mixing of two air parcels; for example, we would not want to report values of 100% or more in the extreme case of warm humid days averaged with very cold days. That is, assessing the character of air at a given location over time (i.e., how close is it to condensation) is not the same thing as the mixing of air parcels. A simple averaging of relative humidity values is appropriate for assessing the character of air over time.

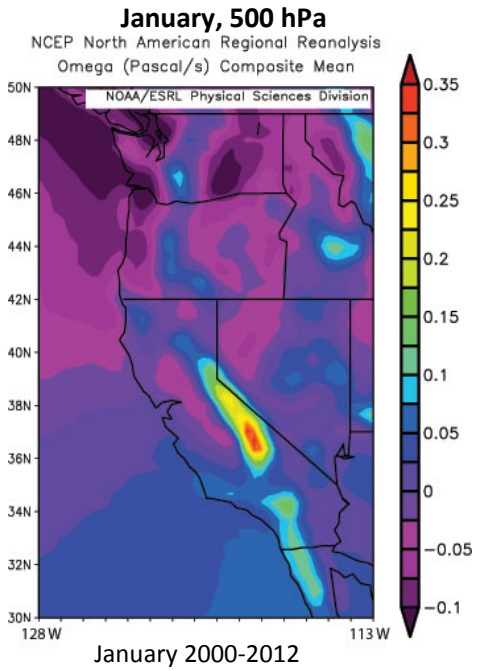
4.A.2. NCEP NARR Average Vertical Velocity (Omega) data. January and April (2000-2012) at 850 hPa (upper panels) and 500 hPa (lower panels). The southern San Joaquin Valley tends to have rising air (negative Omega values) at 850 hPa (~1500 m) in most months (represented by April here), except for mid-winter. Both months are characterized by descending air over the southern Sierra Nevada at 500 hPa (~5500 m). The associated subsidence likely contributes to the presence of capping inversions aloft.



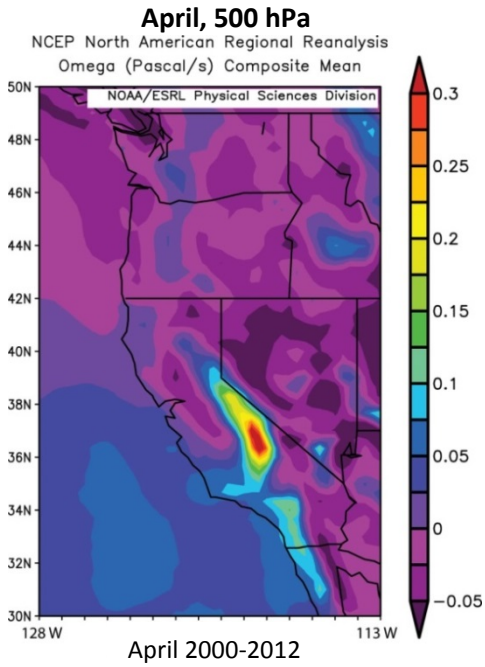
January 2000-2012



April 2000-2012

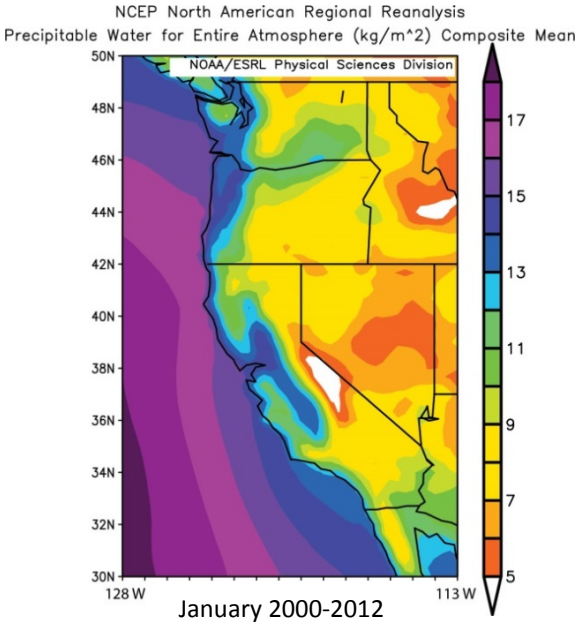


January 2000-2012

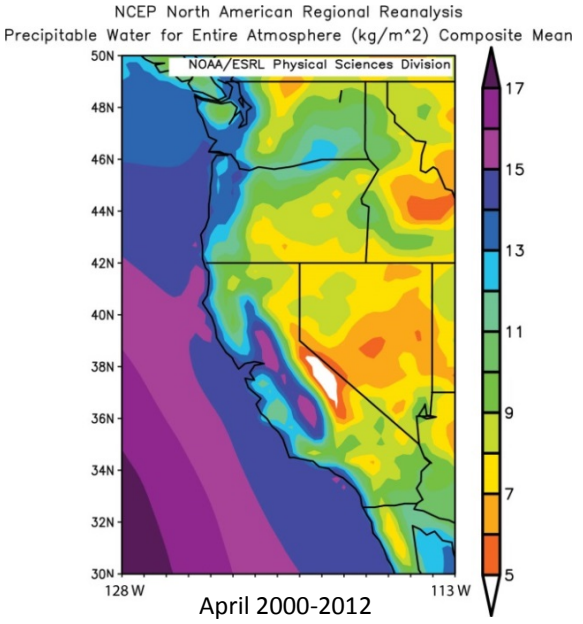


April 2000-2012

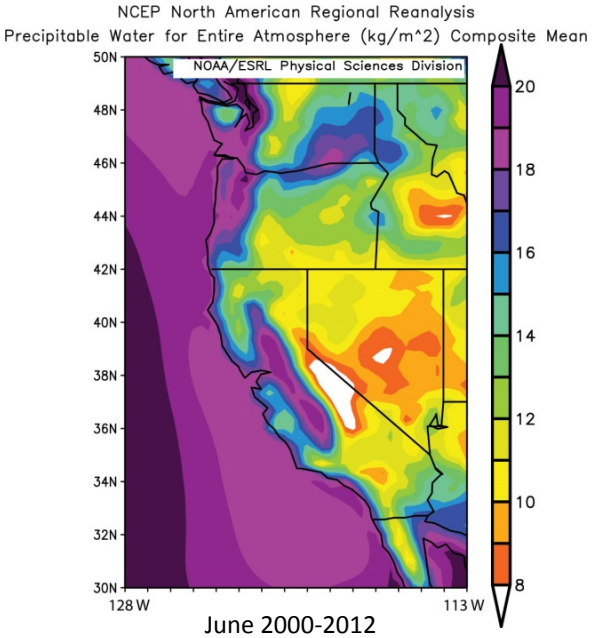
4.A.3. NCEP NARR integrated precipitable water vapor (PWV) data of select months. January is representative of winter conditions, with relatively low water vapor over land, due to cold temperatures, other than areas under some coastal influence, including the Central Valley of California. The legends change among months, so relative amounts are easier to note. In January the Central Valley PWV only reaches values for the southern and central California coast, lower than coastal values for the Pacific Northwest. By April the Central Valley exceeds all but the north coast of California. This may be related to the generally thicker boundary layer that forms during this period. This pattern more or less extends to June. The pattern changes radically in July (although so does the legend), when southern monsoonal influence dominates. The southern Central Valley is still quite high relative to much of California and the Pacific.



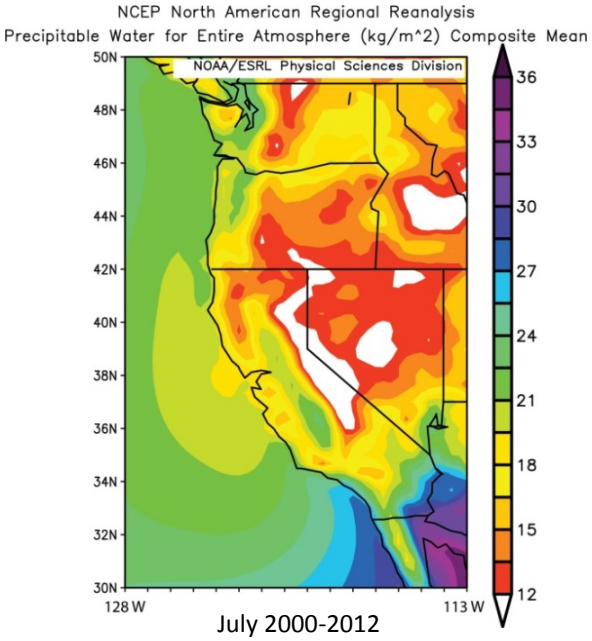
January 2000-2012



April 2000-2012



June 2000-2012



July 2000-2012

Chapter 5. Conclusions

Summary of Key Findings

Key aspects of this dissertation involved: 1) improved mapping of cloud cover from satellite imagery, in order to generate accurate monthly cloud frequency variables; 2) use of these cloud frequency variables to demonstrate the importance of cloud cover, a previously overlooked climate variable, to the distribution of a particularly charismatic plant species, giant sequoia; and 3) investigation of the climate in the vicinity of giant sequoia, to explain mechanisms behind the frequent cloud cover, and, in the process, identify other climatic factors that may also be important to the tree's distribution.

In the first major study (Chapter 2), novel algorithms were developed for improved classification of clouds in daily satellite imagery from 1981 to 2012. The main improvements were in terms of consistency across cover types, including deserts and snow covered areas. This is an important concern for species distribution modeling, in which relative differences among areas are more relevant than absolute amounts. The cloud classification included both Advanced Very High Resolution Radiometer (AVHRR) data for 1981 to 1999 at ~5 km resolution (.05°) and Moderate Resolution Imaging Spectroradiometer (MODIS) data for 2000 to 2012 at ~500 meter resolution. These daily cloud classifications were converted into monthly, annual, and longer-term cloud frequency products. The extended record allows the long-term assessment of patterns and trends in cloud cover, although limitations associated with the coarser resolution and older data of AVHRR make precision in long-term trends suspect. MODIS Aqua and Terra products, however, were more accurate, and their ~10 and 12 year records, respectively, covered here are long enough to generate accurate monthly cloud frequency climatology products.

In Chapter 3, the monthly cloud frequencies generated from MODIS Aqua were used in species distribution models for giant sequoia and found to be the strongest predictors of its distribution from a large suite of climate variables. April afternoon cloud frequency, in particular, was found to be the strongest predictor, but many of the monthly frequencies are highly correlated with April. A high frequency of cloud cover, especially in spring, differentiates the climate of the west slope of the southern Sierra Nevada, where giant sequoia are prolific, from northern parts of the range, where the tree is practically absent. Other mapped cloud products, contaminated by confusion with high elevation snow, would likely not have found this important result. The result illustrates the importance of accuracy in mapping as well as the importance of previously neglected aspects of climate for species distribution modeling.

In Chapter 4, a substantial number of unusual climate patterns was found to be associated with the frequent cloud formation on the west slope of the southern Sierra Nevada. MODIS land surface temperature (LST) data revealed that the same area of frequent cloud formation has unusually low diel LST range throughout the year. This is in spite of the fact that these satellite LST data can only come from days with clear skies, when temperature ranges would be expected to be higher. Surface and mapped climate data sets do not capture this behavior, but meteorological data from a 55 meter tall tower positioned within the northern limits of the frequent cloud belt captured even more extreme behavior. Here, diel temperatures vary only 1-2 degrees C on average each day. Furthermore, spring temperatures are unusually cold at this

tower top, with March and April temperatures colder than January. When combined with dew point temperatures (essentially absolute humidity) at the tower top that regularly, strongly, and unusually, peak in the afternoon, afternoon relative humidities, especially in the spring, can get very high. High afternoon relative humidity values are often present even on clear days. This provides some evidence that it might not just be the cloud cover responsible for sequoia distribution and that relative humidity values above the surface may need better mapping. Other aspects of the climate in the region were quite distinctive as well, including incredibly low wind speeds at tower top height all across a 2300 meter elevation transect of four meteorological towers. Surface climate station data also revealed unusually high ratios of snowfall to total precipitation in the spring months of March and April (higher than winter) across mid-elevations of the entire Sierra Nevada. When combined with the southern Sierra Nevada having a relative peak in late winter and early spring (March/April) moisture, this indicates that mid-elevations of the southern Sierra Nevada have an unusually high amount of spring snowfall relative to other areas of comparable annual precipitation in the mountain range. These snowfall results suggest that cold air aloft in spring is characteristic of much of the Sierra Nevada, but relatively high water vapor in the spring in the southern Sierra Nevada may be the particular key to frequent cloud formation there.

Directions for Future Research

A key aspect of this dissertation involved improvements in cloud frequency mapping with satellite data. Further improvements to the cloud frequency products could continue to be made. These would especially help with detecting subtle trends in cloud frequency over time. For example, while the southwestern deserts of the United States appeared to be an area of declining cloud frequency during the AVHRR record of 1981-1999, this was also an area in which AVHRR cloud classification has some challenges. The better detection of trends, with AVHRR in particular, would require the better treatment of its problematic data, including difficult to track data gaps and aberrations. With respect to clouds and giant sequoia, improvements could be made to the admittedly ad hoc spatial filtering method for detecting temporal trends in the particular cloud formation over the sequoia belt. The cloud formation is restricted enough that the coarse resolution of the AVHRR data may not be very helpful for assessing trends, but MODIS data are beginning to be, as the MODIS data record is now approaching 15 years. Surface station data may be helpful in this regard, as well as for validation of the satellite data, but climate stations are sparse in the area.

There may be limitations to what can be done with accurate cloud classification from the AVHRR LTDR data, but processing following land-cover related stratification would likely help optimize the classification over different areas. This approach has merit but can be sensitive to the accuracy of the land cover map. Similar strategies could be pursued with MODIS data as well, although the MODIS-derived cloud product developed here probably has less room for improvement. A more detailed accuracy assessment of the custom MODIS classification would help to identify areas that might need particular attention. Some snow-covered areas would be likely candidates. There is a possible underestimate of cloud frequency in snowy mountain locations in the custom MODIS cloud classification generated here, and a dearth of climate stations at these elevations makes accuracy assessment in these areas especially hard. The evaluation of daily cloud masks in relation to individual satellite images may be an effective way

of assessing the accuracy of the cloud product over these areas. The use of visual cues – essentially contextual classification – is extremely helpful for analyzing satellite data. For the same reasons, classification algorithms that can make better use of context (both spatial and temporal) may have substantial merit for further investigations. For example, areas that appear to be persistently cloudy may be more likely to be something else, such as snow or desert playa. Of course, care must be taken not to eliminate areas that just happen to be persistently cloudy.

The original daily cloud maps generated here are only a single snapshot (AVHRR) or two (MODIS) over the course of a day. Finer temporal resolution could be achieved with weather satellite data with higher frequency of image capture (e.g., every 15 minutes for the Geostationary Satellite System, GOES). The higher resolution of MODIS (and even Landsat data at ~30 meters, despite its 8-16 day return interval, if enough are available over its multi-decadal record) could potentially be used to spatially downscale the finer temporal-resolution, but coarser spatial resolution, GOES data. Alternatively, new sensors at finer spatial and temporal resolution may eventually make this notion of downscaling academic. Finer temporal resolution in cloud cover mapping would contribute to improvements in mapping patterns of solar radiation.

The distribution modeling of sequoia in Chapter 3 would benefit from the incorporation and/or improvement of maps of climate variables identified in Chapter 4 as possibly relevant to cloud formation. Spatial patterns of relative humidity at some height above the surface might be ideal, but even improved mapping of relative humidity at the surface would be helpful. The incredibly low winds speeds in the area may also be a factor in cloud formation and persistence, and highlight the need for better wind maps. Given that unique spatial and temporal cloud cover patterns may be obvious symptoms of unique climates, they, along with other patterns in satellite imagery, such as land surface temperature patterns and surface reflectance (i.e., vegetation) patterns, could be used to guide the siting for increased sampling of various climate variables. They could also be used to improve the interpolation among the sites for better mapping of the climate variables. The need for improved climate mapping includes temporal aspects as well as spatial aspects. Micrometeorological tower data investigated in Chapter 4 indicated that climate variables such as relative humidity can vary substantially over the course of a day, with the time of peak values varying substantially among sites. Even daily minimum and maximum values would not capture the important result that the tower site with peak relative humidity in the afternoon also had the most frequent afternoon cloud cover.

Given that cloud cover has a strong influence on solar radiation, temperature, humidity, and snowpack retention, and that all of these variables can factor into the overall site water balance, a better understanding of cloud cover has major implications for improved mapping of the water balance. Water balance variables are hypothesized to be among the best integrators of the plant-relevant environment and likely important in many species distributions, but their accurate mapping is heavily dependent on accurate characterization of the soil contribution. This accurate characterization is currently lacking, but relevant aspects of the soil contribution might potentially be solved if they are the only unknown in a comprehensive vegetation and water balance model.

Better maps of various climate and climate-derived variables, such as the water balance, would also be useful to hone in on the physiological ecology of giant sequoia, and the exact

manner in which sequoia is benefiting from its unique climate. Quantifying the water savings from its increased cloud cover in comparison to its increased humidity (and reduced vapor pressure deficits) might help to clarify the critical mechanisms, and also better understand sequoia sensitivity to future climate changes. A better understanding of sequoia's response to the increased ratio of diffuse radiation to direct sunlight provided by cloud cover is also needed, as increases in light use efficiency could contribute to increases in water use efficiency. Feedbacks among climate, sequoia, and fire certainly deserve better attention as well. A unique climate likely promotes a unique fire regime.

Some hypotheses were put forth in Chapter 4 related to the different meteorological factors that might be involved in the cloud formation associated with giant sequoia. These hypotheses were supported by important ground measurements. Continued micrometeorological data collection from the southern Sierra Nevada Critical Zone Observatory elevational transect would help to ascertain whether the very unusual meteorological patterns recorded (e.g., very cold spring temperatures at 55 meters aloft at the 2015 meter elevation tower) were not just a several-year aberration. It is also worth determining whether there is anything unusual about the particular siting of the tower at 2015 meters and whether it perceives characteristics of the free troposphere. It is highly unlikely that the data were an aberration, as they are consistent with aspects of the atmosphere suggested by many other sources. But continued measurement and comparison with other data sources would help to better understand the mechanisms behind the unusual climate. Clearly understanding the meteorological mechanisms and how they might change in the future, together with improved understanding of environmental controls on giant sequoia distribution and performance, is necessary in order to better forecast giant sequoia response to future climate changes. The cloud frequency maps and other climate patterns revealed in this dissertation should help to guide the further understanding of this globally outstanding species and its unique climate.