

A 32-year species-specific live fuel moisture content dataset for southern California chaparral

Received: 2 May 2025

Accepted: 3 February 2026

Cite this article as: Varga, K., Jones, C. A 32-year species-specific live fuel moisture content dataset for southern California chaparral. *Sci Data* (2026). <https://doi.org/10.1038/s41597-026-06794-3>

Kevin Varga & Charles Jones

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

A 32-year species-specific live fuel moisture content dataset for southern California chaparral

Kevin Varga^{1,2,3} (jwnxr9@jmu.edu) and Charles Jones^{2,3}

¹School of Integrated Sciences, James Madison University

²Department of Geography, University of California Santa Barbara

³Earth Research Institute, University of California Santa Barbara

Abstract

Live fuel moisture content (LFMC) strongly affects the behavior of wildland fire, resulting in its incorporation into wildfire spread models and danger ratings. In this study, over ten thousand LFMC observations are combined with predictor variables from Landsat imagery and the Weather Research and Forecasting model to train species-specific random forest models that predict the LFMC of four fuel types—chamise, old growth chamise, black sage, and bigpod ceanothus. These models are then utilized to create a historical, 32-year long, LFMC dataset in southern California chaparral. Additionally, the high spatial and temporal sampling frequency of chamise allowed for quantile mapping bias correction to be applied. The final chamise output, which is the most robust, has a mean absolute error of 9.68% and an R^2 value of 0.76. The LFMC dataset successfully captures the variability in the annual cycle, the spatial heterogeneity, and the interspecies differences, which makes it applicable for better understanding varying fire season characteristics and landscape level flammability.

Background & Summary

As a wildland fire burns across a landscape, the behavior is often highly heterogenous, with stretches of scorched canopies intermixed with unburned islands. One of the many factors influencing this heterogeneity is live fuel moisture content (LFMC), calculated as the ratio of water to dry matter within vegetation. LFMC is typically expressed as a percentage. It can dip down to 30%, below which is considered dead fuel, or reach over 300%, with variations across different ecosystems, species, and even within individual plants¹. LFMC is influenced by a myriad of factors, including atmospheric conditions²⁻⁵, plant physiology⁶⁻¹¹, and soil characteristics^{12,13}. Soil moisture fundamentally affects plant water potential, while plant hydraulic traits dictate how much water moves through the system. Additionally, water is crucial for photosynthesis, which in turn creates more plant matter¹⁴. Atmospheric conditions play a large role by affecting water availability, vapor pressure deficit, evapotranspiration, and solar radiation¹⁵. By understanding how these variables affect LFMC, it is then possible to model LFMC dynamics, which can further aid in understanding wildland fire behavior.

Fire behavior starts at the combustion level, which is directly affected by LFMC. If enough heat is applied to a solid fuel, the molecules break down into gases, a process known as pyrolysis. Continued heating then causes combustion of those gases^{1,16}. The reaction rate of pyrolysis and combustion escalates exponentially with temperature, and greater rates of combustion increase a fire's heat release rate, which drives wildfire spread. In order for combustion to occur, water must be driven off, with the high latent heat of vaporization acting as a thermal sink, slowing these processes. The release of water vapor also dilutes the pyrolyzed combustible gases, thereby reducing combustion efficiency¹⁶. A flammability study of California chaparral species quantified LFMC-driven combustion delays by calculating an empirical

relationship between the moment of full combustion and maximum flame height, where every increase of 10% LFMC caused an additional four second delay¹⁷.

Due to its importance, LFMC has long been considered in fire danger rating systems, including the United States National Fire Danger Rating System (NFDRS)¹⁸. Initially, LFMC was simply modeled as very wet dead fuels, disregarding the dynamic nature of plant physiology where moisture and carbon are constantly in flux⁶. Currently, LFMC within the NFDRS is modeled with the Growing Season Index, which aims to predict ecophysiological limits of photosynthesis by combining the limiting effects of temperature, water availability, and day length¹⁸. Other studies have used meteorological variables or drought indices to predict LFMC with some success^{4,19}, but these methods do not adequately represent vegetation conditions or soil moisture. Furthermore, remote sensing-based LFMC modeling methods are able to monitor vegetation across vast regions^{20–22}, but problems persist, including heterogeneity within pixels, differing relationships between spectral indices and vegetation types, and varying instrumentation spatiotemporal resolution and lifespan²³. More recently, mechanistic models of LFMC have been developed that use physiological metrics for plant water content, such as relative water content or leaf water potential, and for plant dry mass, such as leaf density or leaf mass area^{9,11,24,25}. The separation of the LFMC components helps to understand species-specific drivers of LFMC variation.

Many wildfire prone areas across the United States, and some across the world, sample LFMC, but the locations are sporadic and often only focus on one species²⁶. As mentioned before, wildfires, and LFMC, can be heterogenous across a landscape, making spatially varying LFMC values important in wildfire behavior models, both operationally and in hindcast modeling analyses^{27–31}. McCandless et al.²¹, sought to better inform real-time wildland fire management by using remote sensing, soil moisture, and surface weather variables to model daily LFMC and dead fuel moisture across the Conterminous United States beginning in 2015. While the methodology is sound, the LFMC predictions are species-agnostic, which does not represent the large interspecies variations and potentially caused larger errors. Additionally, the satellite that was used aged out, causing an end to data production²¹. Similarly, another promising species-agnostic gridded LFMC product, created by Rao et al.²⁰, only produced data between 2016 and 2021, due to failure of the Sentinel-1B satellite. There are still many unanswered questions about how LFMC affects fire spread in different fuels, as well as how different species' LFMC responds to varying atmospheric conditions. For this reason, we set off to create a species-specific, long-term LFMC gridded dataset that could also be applied to other species in other regions.

In this study, we combine numerical weather modeling and remote sensing data to create predictive random forest models that output 32-year (Dec 1987 – Jun 2019) historical LFMC datasets of four shrubland fuel types: new growth chamise (*Adenostoma fasciculatum*), old growth chamise, new growth black sage (*Salvia mellifera*), and new growth bigpod ceanothus (*Ceanothus megacarpus*). New growth refers to the foliage and terminal stems of the current growing season, while old growth is foliage and terminal stems from past growing seasons³². The spatial domain stretches from San Luis Obispo County to the edge of Los Angeles County, in California (Fig 1), aligning with a 1 km climatology created with the Weather Research and Forecasting (WRF) model³³, which is used for model predictors. Imagery from NASA's Landsat mission is also used for a model predictor³⁴.

This California coastal mountain area has a Mediterranean climate with a pronounced summer dry season. Shrublands, specifically the highly flammable California chaparral ecosystem, are the dominant vegetation type, covering 35% of the land area³⁵. Chamise shrubs often comprise 50-90% of chaparral stands, leading

to fire agencies using chamise LFMC as a fire danger indicator. Black sage commonly contributes 20-50% of shrub canopy cover, while bigpod ceanothus is often 10-30% of cover³⁶. LFMC measurements of these species have been recorded throughout this area and time period, mostly by fire agencies to assess fire danger. Over 10,000 LFMC observations were obtained from the US National Fuel Moisture Database³⁷ and Santa Barbara County Fire Department³⁸ for model training and testing. The fuel specific models output LFMC values on the first and the fifteenth of each month, following the typical sampling schedule of fire agencies. The long data records of the products, especially the bias-corrected chamise dataset, provide an important resource to understand the spatiotemporal variability of LFMC and fire behavior in southern California chaparral.

Fig 1 – map of domain with elevation and observation sites

Methods

The following sections will describe dataset development, from initial predictor testing to cross validation checks. To start, LFMC observations were acquired as the model predictands. Then, predictors were calculated, including long-term lag variables, such as 90-day precipitation, 90-day mean temperature, and 150-day mean insolation, as well as short-term lag or instantaneous predictors, such as day length, 7-day mean soil moisture, and near-infrared reflectance of vegetation (NIRv). Species-specific random forest models were then trained and tested with two techniques: 5-fold cross validation using all observations and observation site specific cross validation, where each site was individually withheld as the test data. After testing, models were trained on all predictors/predictands and used to create the species-specific LFMC datasets. Lastly, quantile mapping bias correction was applied to the modeled chamise LFMC dataset, which was possible due to the high spatial frequency of observation sites. A visual flowchart summary of the dataset creation is shown in Figure 2.

Fig 2 - Flowchart

LFMC observation predictands

All LFMC observation sites within the spatiotemporal domain were identified (Fig 1). All fuel types with more than 500 observations were used, which included chamise, old growth chamise, black sage, and bigpod ceanothus, with chamise being by far the most numerous (Table 1). Black sage (4) and bigpod ceanothus (2) were sampled at significantly fewer sites, but their observation records covered most of the temporal domain. On the other hand, old growth chamise has been sampled at more locations (15) but has a shorter observation record. Most of the observations were provided by the US National Fuel Moisture Database program via the pyNFMD repository on GitHub³⁷. We also retrieved additional LFMC observations from Santa Barbara County Fire Department³⁸, as they were not included in pyNFMD. Santa Barbara County Fire Department (FD), Ventura County FD, Los Angeles County FD, Kern County FD, CAL FIRE San Luis Obispo Unit, and the Los Padres National Forest were the agencies that collected samples. LFMC sampling protocols are based on the US Forest Service Fuel Moisture Collection Methods. According to the Collection Methods, new growth is typically sampled, unless otherwise specified, such as with chamise old growth. New growth consists of the current growing season's fresh foliage and stems, while old growth consists of any past foliage and stems³².

Table 1 – Fuels, number of observations and sites

Sampling sites had observations labeled chamise, chamise new growth, and chamise old growth. After comparing distributions of the three, the chamise labeled observations corresponded to typical chamise new growth values, so chamise new growth labels were converted to just chamise. When sites had two observations of the same fuel type on the same day, the mean was used. One sampling site, Frazier Park, was removed due to poor data quality, evident when comparing the values to spatially proximal sites. Another site was removed due to it having less than 25 observations.

WRF predictors

For every observation, predictor variables were calculated from a previously constructed WRF model, (Advanced Research WRF version 4.0.1) 32-year (1987-2019) high-resolution downscaling climatology³³, available via the CLIVAC lab at UC Santa Barbara³⁹. The WRF downscaling dataset utilized 4 two-way nested grids (27 km, 9 km, 3 km, 1 km) and numerical parameterizations for microphysics (Single moment 6-class)⁴⁰, long-wave and short-wave radiation transfer (RRTMG)⁴¹, Noah multi-parameterization options (Noah-MP)⁴², surface layer physics and planetary boundary layer (MYNN)^{43,44}. The choice of parameterizations, as well as 1 km resolution inner domain, have proven to produce accurate surface weather variables in other studies that investigate fire danger in association with downslope Sundowner winds in the area⁴⁵⁻⁴⁸. European Center for Medium-Range Weather Forecasts Interim Reanalysis (ERA Interim) was used for initial and boundary conditions, with updates on sea surface temperature every 6 hours. The 27 km outer WRF domain included grid nudging, which forces WRF to stay aligned with the reanalysis data. For predictor variable pre-processing, WRF hourly outputs from the 1 km domain were processed into daily accumulated precipitation and mean daily outputs of temperature, relative humidity, incoming shortwave radiation, soil moisture, actual evapotranspiration (AET), surface pressure, net radiation, saturation vapor pressure, actual vapor pressure, and wind speed.

Vegetation physiological dynamics are affected by preceding meteorological conditions. As such, we tested a number of short term and long-term lag variables from WRF, inspired by a previous study¹⁹ (Table 2). Testing involved comparing the influence each predictor had on model accuracy by calculating Gini importance and cross validation statistics. All WRF predictors are direct outputs of the model, except reference evapotranspiration (ETo), which was calculated with the Penman-Monteith equation using WRF outputs⁴⁹. WRF modeled AET is then subtracted from ETo to get climatic water deficit, which has been shown to affect LFMC⁴. We mostly utilized direct outputs from WRF, rather than composite drought indices, to enable direct attribution of LFMC response to individual meteorological drivers via Gini importance. Drought indices have previously been used to estimate vegetation moisture content, but they are an imperfect method⁵⁰.

For investigating predictor variable lag times, we conducted an autocorrelation analysis for each initial tested predictor (Table 2), where LFMC observations, as well as each potential predictor, were resampled to a semi-monthly temporal frequency. Precipitation, ETo, and climatic water deficit were summed over the semi-monthly periods, while all other variables were averaged. The highest correlation lag time was averaged across each LFMC observation site for each variable. For example, when averaging all sites, accumulated precipitation showed the greatest correlation after 3.9 periods, or about 58 days, so 58-day accumulated precipitation was tested as a predictor. Interestingly, after model testing, the initial lag times that were used, mainly 90- and 150-day, performed better than the computed lag times, resulting in the use of the initial lag times for the final predictors.

After final predictor selection, daily values for each predictor were calculated across the entire domain. Due to seasonal cycles, many of the predictors are correlated with each other (Fig 3). However, random forests are robust to multicollinearity because predictors are randomly selected at each split within each decision tree, preventing correlated predictor pairs from dominating the ensemble. Bootstrapped sampling further reduces any negative effects of collinearity by randomly selecting a subset of predictors to use in each decision tree⁵¹. The WRF downscaling dataset extends from June 1987 through June 2019, but due to the 150-day lag variables, our predictors, and thus final LFMC dataset, stretch from December 1987 through June 2019.

Table 2 – Initial and final predictors

Figure 3 – Predictors and predictand at one LFMC sampling site

Landsat predictor

To incorporate vegetation conditions, we calculated vegetation indices from Landsat 4-8 imagery³⁴, due to the historical temporal coverage and expected future operational availability. Initially, the normalized difference vegetation index (NDVI), normalized difference moisture index (NDMI), and near-infrared reflectance of vegetation (NIRv) were all tested as predictors. Each of these three vegetation indices influenced the model in similar ways, with NIRv increasing model accuracy the most. Therefore, NDVI and NDMI were dropped (Table 2). NIRv is calculated by multiplying near-infrared reflectance (NIR) of a scene by NDVI. NIRv was first introduced to more accurately calculate terrestrial photosynthesis, partially by alleviating the mixed-pixel problem—remote sensing pixels including non-vegetated surfaces⁵². The mixed-pixel problem is important in our domain, which includes urban areas, bare ground, and agriculture. Due to vegetation having high reflectance of NIR, multiplying NIR by NDVI effectively pulls out the vegetation signal. We calculated semi-monthly NIRv values across our spatiotemporal domain using a previously published Google Earth Engine (GEE) based algorithm⁵³.

The algorithm goes as follows. First, Landsat images within individual years, and within our spatial domain, are passed through the GEE provided cloud filter before calculating NIRv. Another cloud filter focuses on cloud shadows by removing values that substantially deviate below a linear regression of the neighboring NIRv values. Semi-monthly values are then calculated by looking at all values within a 120-day window, centered on the desired time step and overlapping preceding and proceeding years. If there are more than five Landsat images in that window, a quadratic regression is used to estimate the NIRv value at the desired time step. If there are between three and five images, a linear regression is used. If there are less than three values, the median value is used. These NIRv values are then upscaled from 30 m to 1 km resolution and exported from GEE. The NIRv geoTIFF files are converted to the WRF latitude/longitude grid, masked with a land mask, and cropped to the temporal domain of the WRF predictor variables. NIRv values below zero were removed, as these values represent water, and one outlier above 0.5 was removed. For the final step, temporal linear interpolation was used to fill any NIRv data gaps less than 90 days.

Random forest model development

With predictor variables finalized, values were linearly interpolated from their gridded output to the location and date of each LFMC observation. The predictors were then normalized by subtracting the mean and dividing by the standard deviation, which was done for every model iteration. Random forest

models were chosen as the predictive model, due to the efficient computing, interpretability, and LFMC prediction success in other studies^{19–21,54}. Random forest hyper parameter tuning was conducted by creating a parameter grid of bootstrap use setting, maximum depth, maximum features, minimum samples per leaf, minimum samples per split, and number of trees. For each of the four fuel types, 5-fold cross validation was performed on 100 different randomized iterations of the parameter grid. The parameter settings that resulted in the lowest R^2 value for the test sets were saved as the tuned parameter combination for each fuel type and used for every model iteration (Table 3).

Table 3 – Random forest hyperparameters

Using all available predictors—normalized as described above—and predictands for training, fuel specific random forest models were fit using the derived hyper parameter settings. The models were then used to calculate fuel specific LFMC values at 1 km spatial resolution on the first and the fifteenth of every month from December 1987 through June 2019. Many of the time steps in the modeled LFMC data included small spatial gaps, especially along the coast. These spatial gaps are attributable to constant cloud cover occurring during Landsat passes, which results in missing NIRv values. All of the LFMC spatial gaps were identified for every time step and determined to be small enough to be fillable via spatial linear interpolation, providing dataset consistency across space and time.

Chamise quantile mapping bias correction

When comparing the modeled chamise LFMC against the observations, the modeled results at some observation sites exhibited consistent biases, especially during the peak periods of the wet and dry seasons. We determined that quantile mapping bias correction could be applied to the chamise LFMC dataset, due to the high spatial coverage of observation sites (Fig 1). To do so, chamise LFMC observations were resampled to the semi-monthly temporal resolution of the modeled results, allowing for quadratic interpolation of two time steps. In order to optimize the bias corrected results, different numbers of quantiles were tested: 10, 20, 30...100, with 30 quantiles causing the greatest reduction in mean bias error across the sites. As such, 30 quantile values were calculated for the LFMC observations and their corresponding model predictions at each site. The differences between the quantile values were calculated in order to determine the typical bias correction that was needed for the range of modeled values. Modeled values over the full temporal domain were then fit into the aforementioned model quantile bins, with the corresponding correction factor then applied to each bin.

With temporally varying bias correction factors calculated at all 28 chamise LFMC observation sites, spatial interpolation of the corrections at every time step was performed using latitude, longitude, and elevation, across the 1 km resolution grid of modeled chamise LFMC values. A thin plate spline driven radial basis function was used, due to the efficient implementation across multiple dimensions. Some sites are close in proximity, which makes it difficult for the interpolation to exactly fit the input data. Therefore, a smoothing factor of 0.01 was used to allow for the fitting of the interpolated corrections to be based on a least-squares fit. Once gridded bias correction factors were calculated for every time step, these were added to the original modeled chamise LFMC values to produce a bias corrected dataset. Figure 4 visualizes the bias correction of the time step with the greatest correction value.

Fig 4 – Maps showing chamise bias correction at time step with greatest correction

Data Records

The modeled LFMC of four fuel types—new growth chamise (*Adenostoma fasciculatum*), old growth chamise, new growth black sage (*Salvia mellifera*), and new growth bigpod ceanothus (*Ceanothus megacarpus*)—and the bias corrected chamise LFMC are available as a single NetCDF file on Dryad, with 1 km, semi-monthly spatiotemporal resolution and latitude, longitude, and time (December 1987 – June 2019) as the dimensions⁵⁵. There are LFMC values for every land grid point in the domain, even though these fuel types do not occur in many locations. The chamise dataset, especially the bias corrected chamise dataset, is considered to be the most robust, due to fire agencies preferentially sampling chamise in this domain, as well as the species prevalence in chaparral ecosystems⁵⁶ (Fig 1, Table 1).

Technical Validation

For final validation, the modeled outputs were linearly interpolated, both spatially and temporally, to the time and location of every LFMC observation of the respective fuel types. Figure 5 shows scatter plots between the observations and predictions, with associated statistics. The models all tend to over predict at low observational values and under predict at high observational values. This tendency is common in random forest models, due to the averaging effect across many decision trees⁵⁷. Prediction accuracy at the lower end of LFMC is more important for fire danger prediction. Specifically for chamise, a past study determined that large fire growth is more common below a 79% LFMC threshold³. For our bias corrected chamise model predictions of observed LFMC below 79%, the MAE is 5.9%, the MBE is 1.47%, and the RMSE is 8.22%. The positive MBE in this lower range and the overall negative MBE of all models shows that the bias on the higher end of predicted values is more pronounced.

Fig 5 – Scatter plots between observed and predicted

Model prediction uncertainty is also quantified by using the variance among all the decision trees employed for each individual prediction. Specifically, the standard deviation is computed using the range of predictions from every tree. A 95% prediction interval is then calculated as the mean prediction ± 1.96 standard deviation. This approach leverages the natural ensemble spread of the forest to estimate local predictive uncertainty. The method follows the quantile regression forest framework, where tree variance reflects epistemic uncertainty due to model structure and training data sampling⁵⁸. Uncertainty fields are provided in the dataset as separate variables with the same spatiotemporal dimensions as the LFM predictions.

Usage Notes

Limitations

The dataset presented in this study comes with a few limitations. First, LFMC sampling locations are spatially sparse, especially for black sage and bigpod ceanothus (Fig 1). Therefore, the potential spatial heterogeneity of intraspecies LFMC behavior may not be captured during model training. Old growth chamise has greater spatial coverage, but the observational record was temporally shorter. Due to these limitations, the chamise dataset is considered the most robust, although 28 sampling sites is still not widespread spatial representation. Second, the datasets provide LFMC values for every land grid point, even though any particular fuel type does not occur in many areas. It is recommended that land cover data is used to mask the LFMC data in order to focus on regions where the results are more appropriate, such as shrublands. Specifically, modeled LFMC results in agricultural or developed areas may not be accurate, due to differences in NIRv values between native shrubs and irrigated crops or vegetation.

Additionally, the construction of this dataset was based on the spatiotemporal domain and data availability of a previously constructed, 32-year downscaled climatology. While this type of historical, downscaled data is not available everywhere, the predictor variables can also be calculated using other climatological datasets. We specifically sought to combine meteorological predictors, which represent drivers of vegetation green-up and senescence, and current condition predictors—NIRv, day length, soil moisture—that would allow for this LFMC predictive methodology to be applied to other fuel types in other regions. It is the authors' intent to expand this historical dataset to the entire state of California.

Application

The potential usage of this dataset spans different disciplines and sectors. For example, fire agencies consistently sample LFMC to determine current fire danger. In the Mediterranean climate of California, the wet-up and dry-down behavior of LFMC determine when fire season starts and ends in different areas. With a historical LFMC dataset, additional analyses can lead to a better understanding of how meteorological and climatological factors influence LFMC behavior on a year-to-year basis. Those factors can then be used to help predict and prepare for upcoming fire/dry seasons. Other widespread LFMC datasets do exist^{20,21,59}, but the combination of locally produced climatological data, the long Landsat lifespan, and the species-specific outputs, make this LFMC dataset valuable for local studies of fire dynamics.

The authors of this paper also contribute to a holistically-minded Regional Wildfire Mitigation Program (RWMP), which promotes living with fire instead of fighting it⁶⁰. The landscape level management vital to creating fire resilient communities requires an understanding of the varying flammability of the surrounding natural fuel types. Historically, an emphasis has been placed on understanding when chamise LFMC reaches critical flammability thresholds^{3,4}, but it is also important to understand how other species' LFMC varies in space and time. For example, Figure 8 indicates chamise has the shortest wet season, when comparing fuel types on the same LFMC scale. Could other, less-flammable species be promoted around communities to decrease the fire risk? And how will flammability change as chaparral species distribution evolves with climate change? These are questions that a historical LFMC dataset can help answer.

Data availability

The LFMC dataset described in this work is available on Dryad (<https://doi.org/10.5061/dryad.rjdfn2zkw>).

The LFMC observations used for training the models are also available on Dryad. The observations were downloaded from the US National Fuel Moisture Database (<https://github.com/wmjolly/pyNFMD>) and the Santa Barbara County Fire Department (<https://sbcfire.com/wildfire-predictive-services/>).

The data used for LFMC model predictors is publicly available via the UCSB CLIVAC Lab (<https://clivac.eri.ucsb.edu/clivac/SBCWRFD/index.html>) and the NASA Landsat program (<https://doi.org/10.1016/j.srs.2023.100103>).

Code availability

The codebase for the creation of this dataset is publicly available as Jupyter Notebooks on GitHub (https://github.com/kcvarga7/sba_lfm_1987-2019). All code was written in Python, with the exception of the GEE script used for downloading Landsat imagery. That GEE script is referenced in the GitHub readme, as well as the applicable Jupyter Notebook.

Acknowledgements

This research was supported by the NASA Future Investigators in NASA Earth and Space Science and Technology program (Award No. 80NSSC21K1630), the University of California Office of the President Laboratory Fees Program (Grant ID: LFR-20-652467), and The Nature Conservancy's Jack and Laura Dangermond Preserve. We would also like to acknowledge high-performance computing support provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation Prediction of and Resilience against Extreme Events program (Award No. 1664173). Lastly, we thank Matt Jolly, United States Forest Service Ecologist, for providing essential live fuel moisture observations.

Author Contributions

Kevin Varga: Conceptualization, methodology, data curation, formal analysis, visualization, writing-original draft preparation, writing-review and editing

Charles Jones: Conceptualization, methodology, writing-review and editing, supervision

Competing Interests

The authors declare no competing interests.

References

1. Keane, R. E. *Wildland Fuel Fundamentals and Applications*. <https://doi.org/10.1007/978-3-319-09015-3>. (Springer International, 2015).
2. Drucker, J. R., Farguell, A., Clements, C. B. & Kochanski, A. K. A live fuel moisture climatology in California. <https://doi.org/10.3389/ffgc.2023.1203536>. *Front. For. Glob. Change* **6**, 1203536 (2023).
3. Dennison, P. E. & Moritz, M. A. Critical live fuel moisture in chaparral ecosystems: a threshold for fire activity and its relationship to antecedent precipitation. <https://doi.org/10.1071/WF08055>. *Int. J. Wildland Fire* **18**, 1021 (2009).
4. Park, I., Fauss, K. & Moritz, M. A. Forecasting Live Fuel Moisture of *Adenostema fasciculatum* and Its Relationship to Regional Wildfire Dynamics across Southern California Shrublands. <https://doi.org/10.3390/fire5040110>. *Fire* **5**, 110 (2022).
5. Ma, W. *et al.* Assessing climate change impacts on live fuel moisture and wildfire risk using a hydrodynamic vegetation model. <https://doi.org/10.5194/bg-18-4005-2021>. *Biogeosciences* **18**, 4005–4020 (2021).

6. Jolly, W. & Johnson, D. Pyro-Ecophysiology: Shifting the Paradigm of Live Wildland Fuel Research. <https://doi.org/10.3390/fire1010008>. *Fire* **1**, 8 (2018).
7. Cakpo, C. B. *et al.* Exploring the role of plant hydraulics in canopy fuel moisture content: insights from an experimental drought study on *Pinus halepensis* Mill. and *Quercus ilex* L. <https://doi.org/10.1186/s13595-024-01244-9>. *Ann. For. Sci.* **81**, 26 (2024).
8. Grossiord, C. *et al.* Plant responses to rising vapor pressure deficit. <https://doi.org/10.1111/nph.16485>. *New Phytol.* **226**, 1550–1566 (2020).
9. Griebel, A. *et al.* Specific leaf area and vapour pressure deficit control live fuel moisture content. <https://doi.org/10.1111/1365-2435.14271>. *Funct. Ecol.* **37**, 719–731 (2023).
10. Pivovarov, A. L. *et al.* The Effect of Ecophysiological Traits on Live Fuel Moisture Content. <https://doi.org/10.3390/fire2020028>. *Fire* **2**, 28 (2019).
11. Jolly, W. M., Conrad, E. T., Brown, T. P. & Hillman, S. C. Combining ecophysiology and combustion traits to predict conifer live fuel moisture content: a pyro-ecophysiological approach. <https://doi.org/10.1186/s42408-025-00361-8>. *Fire Ecol.* **21**, 19 (2025).
12. Qi, Y., Dennison, P. E., Spencer, J. & Riaño, D. Monitoring Live Fuel Moisture Using Soil Moisture and Remote Sensing Proxies. <https://doi.org/10.4996/fireecology.0803071>. *Fire Ecol.* **8**, 71–87 (2012).
13. Brown, T. P. *et al.* Decoupling between soil moisture and biomass drives seasonal variations in live fuel moisture across co-occurring plant functional types. <https://doi.org/10.1186/s42408-022-00136-5>. *Fire Ecol.* **18**, 14 (2022).
14. Kozlowski, T. T., Kramer, P. J. & Pallardy, S. G. The Physiological Ecology of Woody Plants. <https://doi.org/10.1093/treephys/8.2.213>. *Tree Physiol.* **8**, 213 (1991).
15. Wallace, J. M. & Hobbs, P. V. *Atmospheric Science: An Introductory Survey*. <https://doi.org/10.1016/C2009-0-00034-8>. (Elsevier, 2006).

16. Finney, M. A., McAllister, S. S., Grumstrup, T. P. & Forthofer, J. M. *Wildland Fire Behaviour: Dynamics, Principles, and Processes*. <https://doi.org/10.1071/9781486309092>. (CSIRO Publishing, 2021).
17. Sun, L., Zhou, X., Mahalingam, S. & Weise, D. R. Comparison of burning characteristics of live and dead chaparral fuels. <https://doi.org/10.1016/j.combustflame.2005.08.008>. *Combust. Flame* **144**, 349–359 (2006).
18. Jolly, W. M., Freeborn, P. H., Bradshaw, L. S., Wallace, J. & Brittain, S. Modernizing the US National Fire Danger Rating System (version 4): Simplified fuel models and improved live and dead fuel moisture calculations. <https://doi.org/10.1016/j.envsoft.2024.106181>. *Environ. Model. Softw.* **181**, 106181 (2024).
19. Capps, S. B., Zhuang, W., Liu, R., Rolinski, T. & Qu, X. Modelling chamise fuel moisture content across California: a machine learning approach. <https://doi.org/10.1071/WF21061>. *Int. J. Wildland Fire* **31**, 136–148 (2021).
20. Rao, K., Williams, A. P., Flefil, J. F. & Konings, A. G. SAR-enhanced mapping of live fuel moisture content. <https://doi.org/10.1016/j.rse.2020.111797>. *Remote Sens. Environ.* **245**, 111797 (2020).
21. McCandless, T. C., Kosovic, B. & Petzke, W. Enhancing wildfire spread modelling by building a gridded fuel moisture content product with machine learning. <https://doi.org/10.1088/2632-2153/aba480>. *Mach. Learn. Sci. Technol.* **1**, 035010 (2020).
22. Peterson, S., Roberts, D. & Dennison, P. Mapping live fuel moisture with MODIS data: A multiple regression approach. <https://doi.org/10.1016/j.rse.2008.07.012>. *Remote Sens. Environ.* **112**, 4272–4284 (2008).
23. Yebra, M. *et al.* A global review of remote sensing of live fuel moisture content for fire danger assessment: Moving towards operational products. <https://doi.org/10.1016/j.rse.2013.05.029>. *Remote Sens. Environ.* **136**, 455–468 (2013).

24. Balaguer-Romano, R. *et al.* Modeling fuel moisture dynamics under climate change in Spain's forests. <https://doi.org/10.1186/s42408-023-00224-0>. *Fire Ecol.* **19**, 65 (2023).
25. Nolan, R. H. *et al.* Drought-related leaf functional traits control spatial and temporal dynamics of live fuel moisture content. <https://doi.org/10.1016/j.agrformet.2022.108941>. *Agric. For. Meteorol.* **319**, 108941 (2022).
26. Yebra, M. *et al.* Globe-LFMC, a global plant water status database for vegetation ecophysiology and wildfire applications. <https://doi.org/10.1038/s41597-019-0164-9>. *Sci. Data* **6**, 155 (2019).
27. Finney, M. A. *FARSITE: Fire Area Simulator-Model Development and Evaluation*. <https://doi.org/10.2737/RMRS-RP-4>. RMRS-RP-4 <https://www.fs.usda.gov/treearch/pubs/4617> (1998) doi:10.2737/RMRS-RP-4.
28. Finney, M. A. An Overview of FlamMap Fire Modeling Capabilities. *Andrews Patricia Butl. Bret W Comps 2006 Fuels Manag.- Meas. Success Conf. Proc. 28-30 March 2006 Portland Proc. RMRS-P-41 Fort Collins CO US Dep. Agric. For. Serv. Rocky Mt. Res. Stn. P 213-220* **041**, <https://research.fs.usda.gov/treearch/25948>. (2006).
29. Tymstra, C. *Development and Structure of Prometheus: The Canadian Wildland Fire Growth Simulation Model*. <https://publications.gc.ca/site/eng/380448/publication.html>. (Northern Forestry Centre, Edmonton, 2010).
30. Varga, K. *et al.* Megafires in a Warming World: What Wildfire Risk Factors Led to California's Largest Recorded Wildfire. <https://doi.org/10.3390/fire5010016>. *Fire* **5**, 16 (2022).
31. Seto, D. *et al.* Simulating Potential Impacts of Fuel Treatments on Fire Behavior and Evacuation Time of the 2018 Camp Fire in Northern California. <https://doi.org/10.3390/fire5020037>. *Fire* **5**, 37 (2022).
32. Zahn, S. & Henson, C. *Fuel Moisture Collection Methods - A Field Guide*. <https://www.fs.usda.gov/t-d/pubs/pdf/11511803.pdf> (2011).

33. Jones, C., Carvalho, L. M. V., Duine, G.-J. & Zigner, K. Climatology of Sundowner winds in coastal Santa Barbara, California, based on 30 yr high resolution WRF downscaling.
<https://doi.org/10.1016/j.atmosres.2020.105305>. *Atmospheric Res.* **249**, 105305 (2021).
34. Crawford, C. J. *et al.* The 50-year Landsat collection 2 archive.
<https://doi.org/10.1016/j.srs.2023.100103>. *Sci. Remote Sens.* **8**, 100103 (2023).
35. CALFIRE FRAP. Vegetation (fveg). <https://map.dfg.ca.gov/metadata/ds1327.html>. (2015).
36. Sawyer, J. O., Keeler-Wolf, T. & Evens, J. *A Manual of California Vegetation, Second Edition*.
<https://vegetation.cnps.org/>. (California Native Plant Society, 2009).
37. United State Government. National Fuel Moisture Database. <https://github.com/wmjolly/pyNFMD>.
38. Santa Barbara County Fire Department. Live Fuel Moisture Program. <https://sbcfire.com/wildfire-predictive-services/>.
39. Jones, C. Santa Barbara County WRF Downscaling Data.
<https://clivac.eri.ucsb.edu/clivac/SBCWRFD/index.html>.
40. Hong, S.-Y. & Lim, J.-O. J. The WRF Single-Moment 6-Class Microphysics Scheme (WSM6).
https://www2.mmm.ucar.edu/wrf/site_linked_files/phys_refs/micro_phys/WSM6.pdf. *Asia-Pac. J. Atmospheric Sci.* **42**, 129–151 (2006).
41. Iacono, M. J. *et al.* Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. <https://doi.org/10.1029/2008JD009944>. *J. Geophys. Res. Atmospheres* **113**, (2008).
42. Niu, G.-Y. *et al.* The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements.
<https://doi.org/10.1029/2010JD015139>. *J. Geophys. Res. Atmospheres* **116**, (2011).

43. Nakanishi, M. & Niino, H. Development of an Improved Turbulence Closure Model for the Atmospheric Boundary Layer. <https://doi.org/10.2151/jmsj.87.895>. *気象集誌 第2輯* **87**, 895–912 (2009).
44. Olson, J. B. *et al.* A Description of the MYNN-EDMF Scheme and the Coupling to Other Components in WRF–ARW. <https://repository.library.noaa.gov/view/noaa/19837> (2019).
45. Duine, G.-J., Jones, C., Carvalho, L. M. V. & Fovell, R. G. Simulating Sundowner Winds in Coastal Santa Barbara: Model Validation and Sensitivity. <https://doi.org/10.3390/atmos10030155>. *Atmosphere* **10**, 155 (2019).
46. Zigner, K. *et al.* Evaluating the Ability of FARSITE to Simulate Wildfires Influenced by Extreme, Downslope Winds in Santa Barbara, California. <https://doi.org/10.3390/fire3030029>. *Fire* **3**, 29 (2020).
47. Zigner, K. *et al.* Wildfire Risk in the Complex Terrain of the Santa Barbara Wildland–Urban Interface during Extreme Winds. <https://doi.org/10.3390/fire5050138>. *Fire* **5**, 138 (2022).
48. Duine, G.-J., Carvalho, L. M. V. & Jones, C. Mesoscale patterns associated with two distinct heatwave events in coastal Santa Barbara, California, and their impact on local fire risk conditions. <https://doi.org/10.1016/j.wace.2022.100482>. *Weather Clim. Extrem.* **37**, 100482 (2022).
49. Allen, R., Pereira, L., Raes, D. & Smith, M. *FAO Irrigation and Drainage Paper No. 56*. <https://www.fao.org/4/x0490e/x0490e00.htm> (1998).
50. Ruffault, J., Martin-StPaul, N., Pimont, F. & Dupuy, J. L. How well do meteorological drought indices predict live fuel moisture content (LFMC)? An assessment for wildfire research and operations in Mediterranean ecosystems. <https://doi.org/10.1016/j.agrformet.2018.07.031>. *Agric. For. Meteorol.* **262**, 391–401 (2018).
51. Breiman, L. Random Forests. <https://doi.org/10.1023/A:1010933404324>. *Mach. Learn.* **45**, 5–32 (2001).

52. Badgley, G., Field, C. B. & Berry, J. A. Canopy near-infrared reflectance and terrestrial photosynthesis. <https://doi.org/10.1126/sciadv.1602244>. *Sci. Adv.* **3**, e1602244 (2017).
53. McMahon, C. A. *et al.* A river runs through it: Robust automated mapping of riparian woodlands and land surface phenology across dryland regions. <https://doi.org/10.1016/j.rse.2024.114056>. *Remote Sens. Environ.* **305**, 114056 (2024).
54. Camprubí, À. C., González-Moreno, P. & Resco de Dios, V. Live Fuel Moisture Content Mapping in the Mediterranean Basin Using Random Forests and Combining MODIS Spectral and Thermal Data. <https://doi.org/10.3390/rs14133162>. *Remote Sens.* **14**, 3162 (2022).
55. Varga, K. & Jones, C. A 32-year species-specific live fuel moisture content dataset for southern California chaparral. <https://doi.org/10.5061/dryad.rjdfn2zkw>. *Dryad* (2025).
56. Hanes, T. L. Succession after Fire in the Chaparral of Southern California. <https://doi.org/10.2307/1942434>. *Ecol. Monogr.* **41**, 27–52 (1971).
57. Hastie, T., Tibshirani, R. & Friedman, J. Random Forests. in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (eds Hastie, T., Tibshirani, R. & Friedman, J.) 587–604, https://doi.org/10.1007/978-0-387-84858-7_15. (Springer, New York, NY, 2009).
58. Meinshausen, N. Quantile Regression Forests. <http://jmlr.org/papers/v7/meinshausen06a.html>. *J. Mach. Learn. Res.* **7**, 983–999 (2006).
59. Quan, X. *et al.* Global fuel moisture content mapping from MODIS. <https://doi.org/10.1016/j.jag.2021.102354>. *Int. J. Appl. Earth Obs. Geoinformation* **101**, 102354 (2021).
60. Moritz, M. A. *et al.* Beyond a Focus on Fuel Reduction in the WUI: The Need for Regional Wildfire Mitigation to Address Multiple Risks. <https://doi.org/10.3389/ffgc.2022.848254>. *Front. For. Glob. Change* **5**, 848254 (2022).

Figure Legends (Captions)

Fig 1. Elevation of spatial domain, which stretches from sea level to over 2500 meters above mean sea level (mamsl), with LFM observation sites shown, differentiated by the number of observations and the fuel types that were sampled.

Fig 2. Flowchart depicting the process of dataset creation.

Fig 3. Live fuel moisture observations at one sampling site, Sisar Canyon Upper Ojai Valley, plotted with the nine model predictors.

Fig 4. Initial modeled chamise live fuel moisture (a), bias correction applied (b), and bias corrected chamise LFM (c) at the time step with the greatest bias correction value.

Fig 5. The observations for each fuel type are plotted against modeled values that are linearly interpolated to the time and location of each observation. Mean absolute error (MAE), root mean square error (RMSE), mean bias error (MBE), and R^2 are shown for each model. The red line indicates the ordinary least squares regression.

Table 1. Live fuel moisture observations.

Fuel Type	# Obs	# Sites	First Obs	Last Obs
Chamise	6404	28	1/12/1990	6/29/2019
Old Growth Chamise	1426	15	5/1/2006	6/29/2019
Black Sage	1247	4	1/12/1990	6/28/2019
Bigpod Ceanothus	580	2	1/12/1990	11/13/2018

Table 2. Initial and final model predictors.

Initial Tested Predictors		Final Predictors
Day Length	90-day total precip	90-day total precip
3-day max temp mean	90-day mean temp	90-day mean temp
7-day max temp mean	150-day mean temp	150-day solar radiation
7-day max RH mean	150-day mean RH	NIRv
3-day total precip	150-day solar radiation	Day Length
7-day total precip	Elevation	7-day mean soil moisture
7-day mean soil moisture	Slope	150-day mean RH
30-day wind speed	Aspect	90-day climatic water deficit
30-day mean VPD	NDVI	30-day total precip

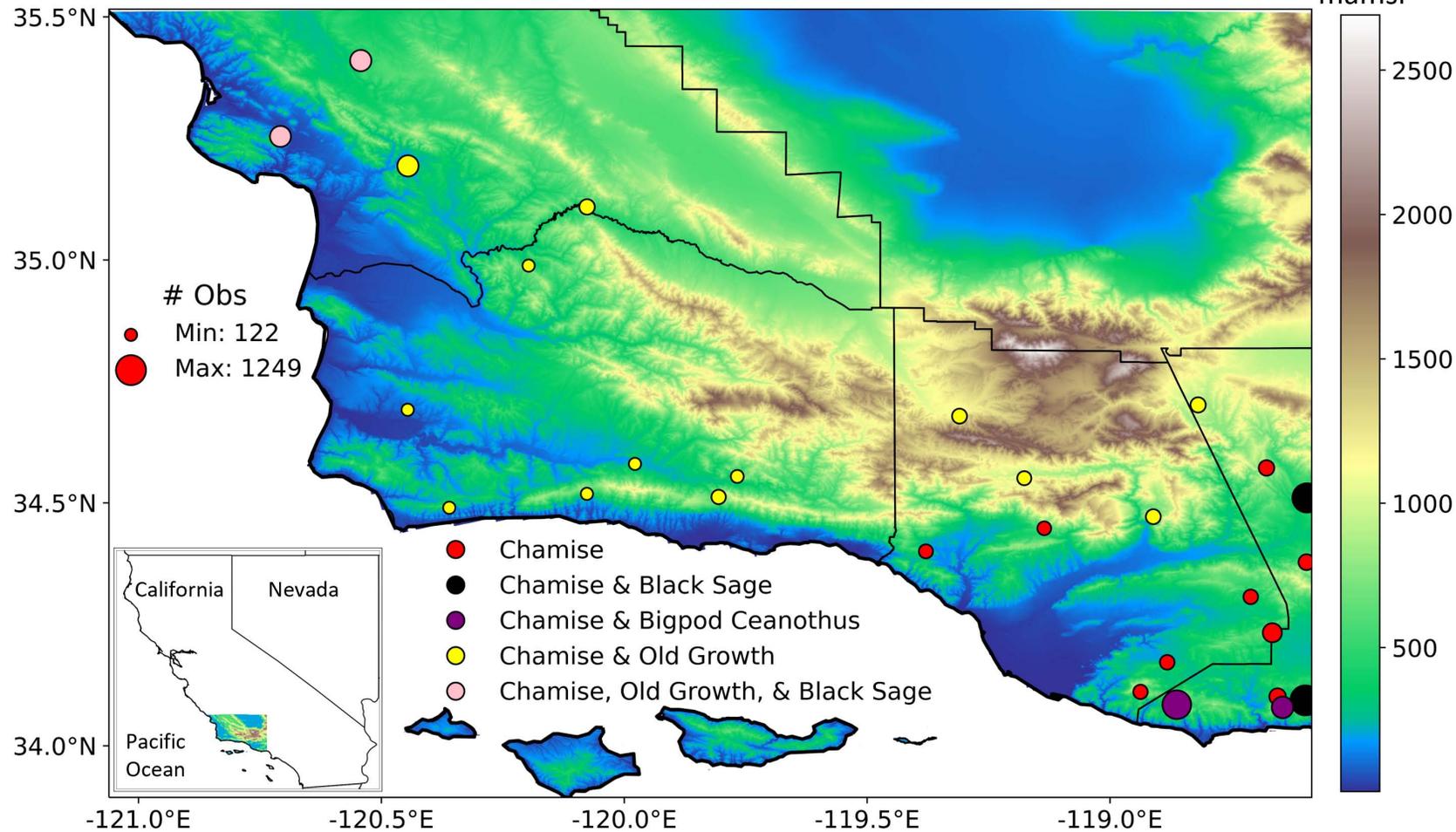
30-day total precip	NDMI	
90-day ref evapotranspiration	NIRv	
90-day climatic water deficit		

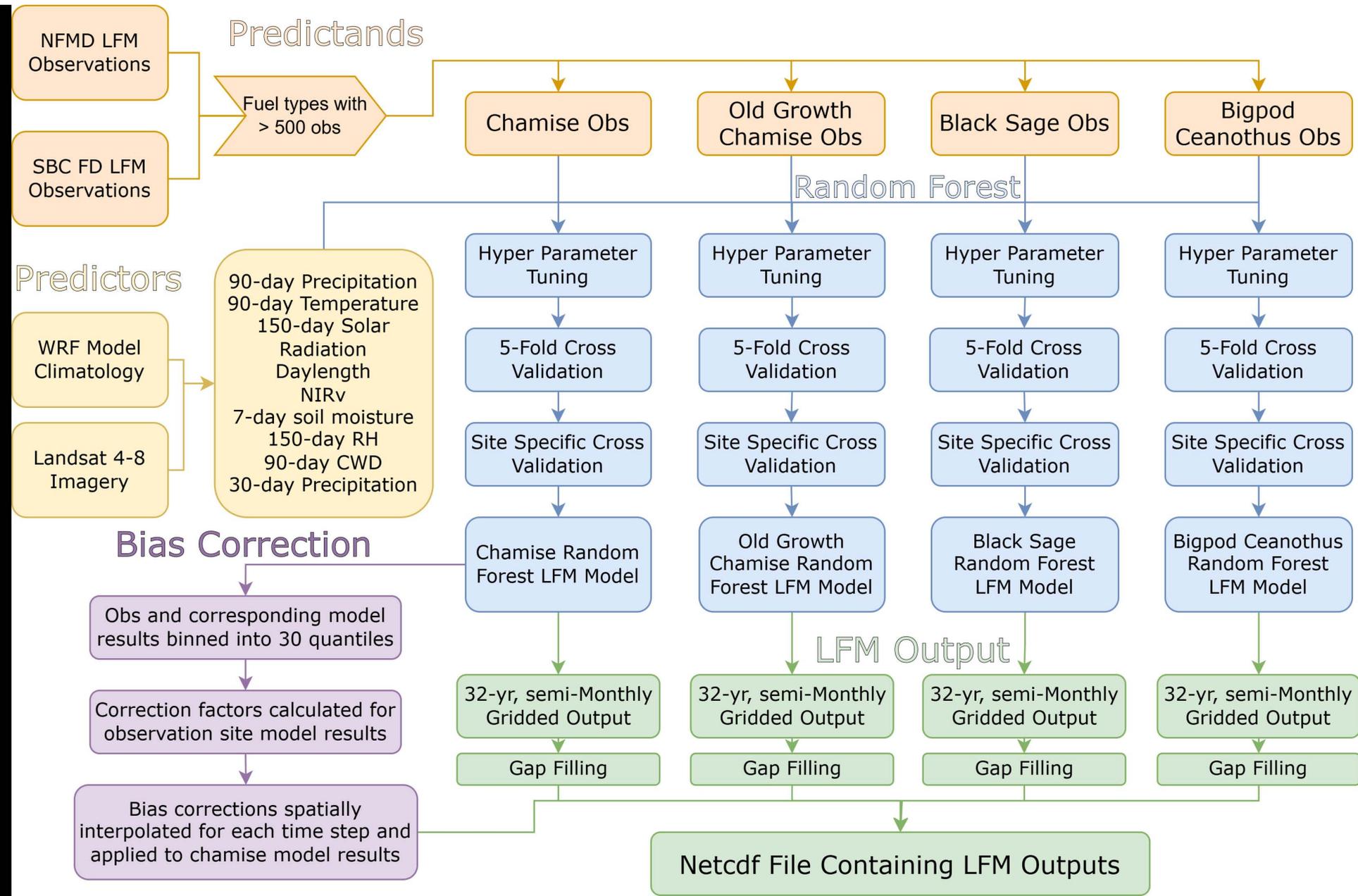
ARTICLE IN PRESS

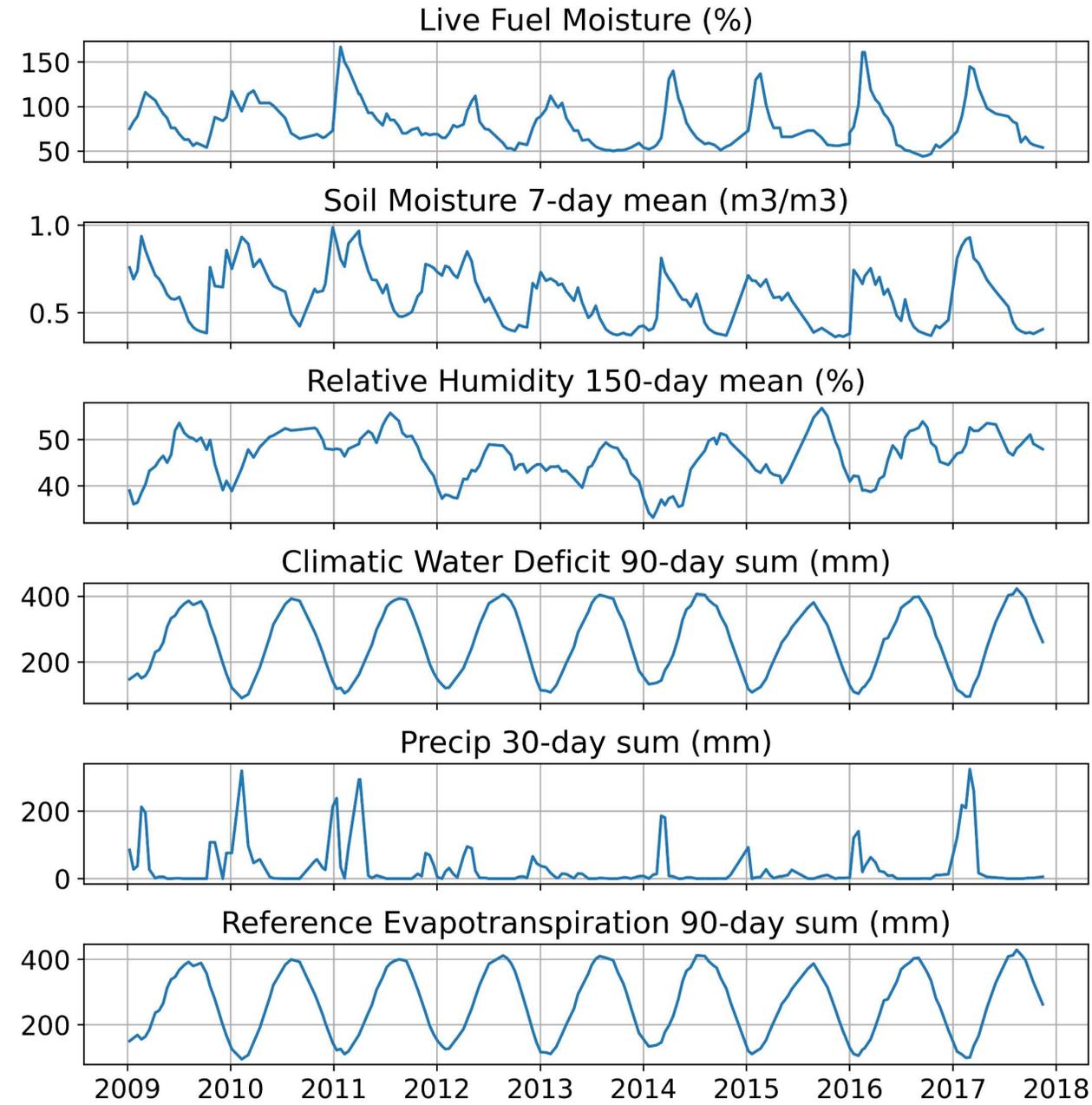
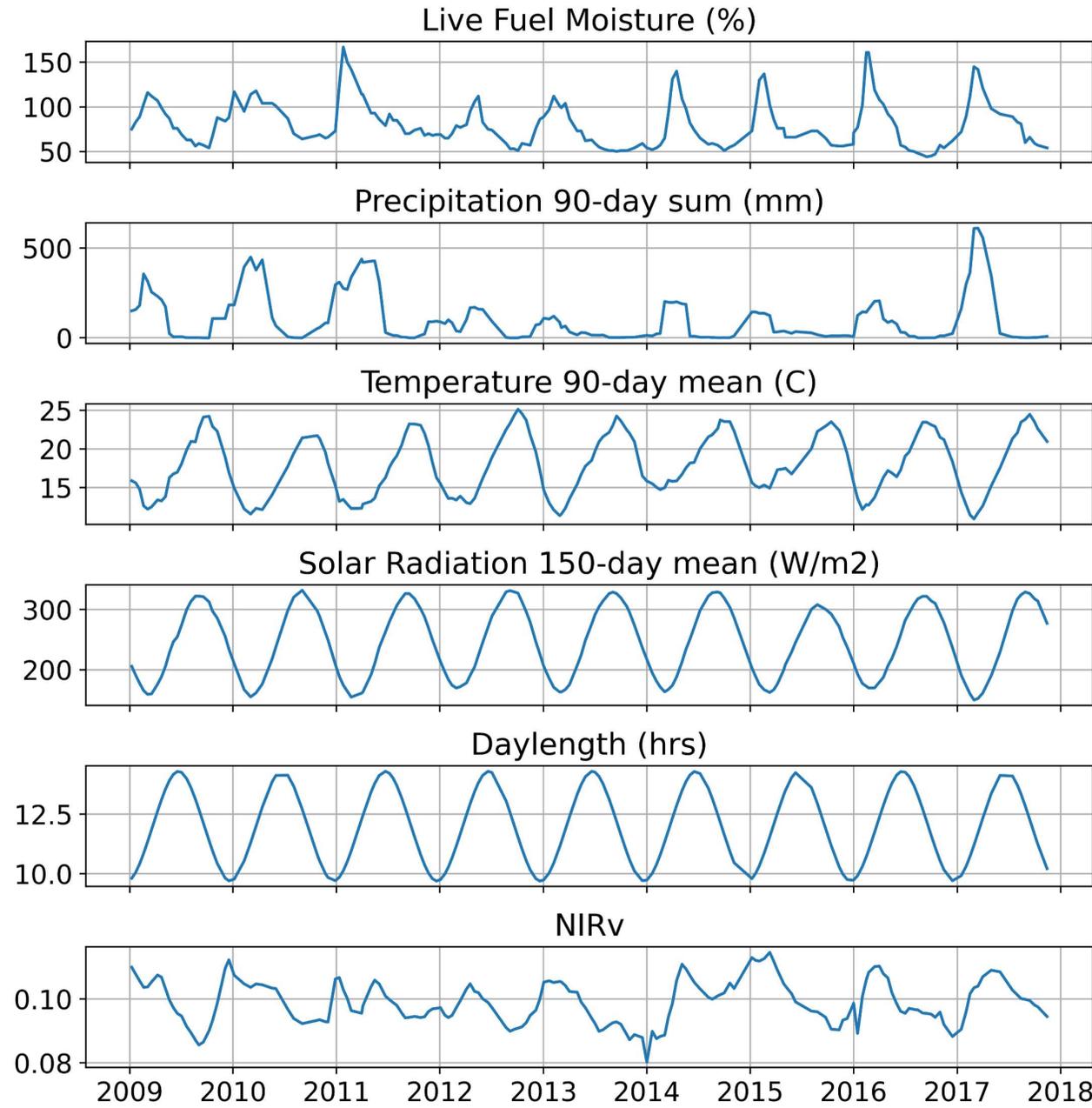
Table 3. Hyper parameters for random forest models determined via randomized grid search.

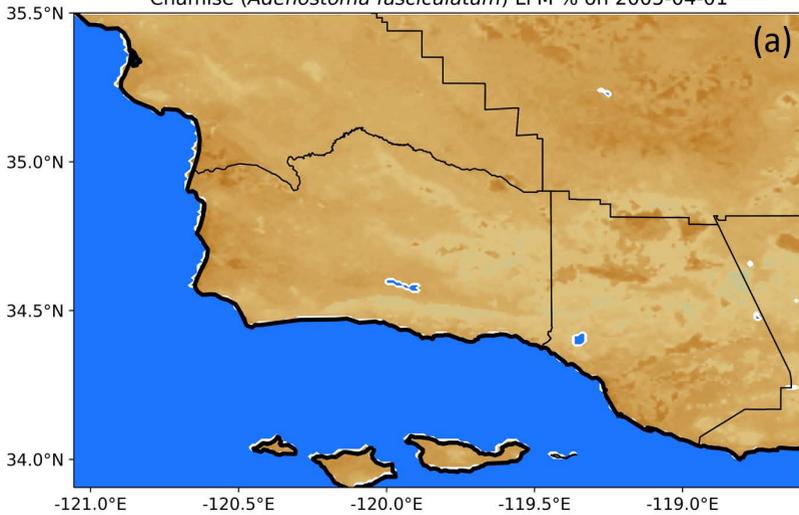
Fuel Type	# Trees	Min Samples Split	Min Samples Leaf	Max Depth	Max Features	Bootstrap
Chamise	800	2	2	50	3	FALSE
Old Growth Chamise	1000	2	1	20	3	TRUE
Black Sage	1000	2	1	110	3	TRUE
Bigpod Ceanothus	1000	2	1	20	3	TRUE

Live Fuel Moisture Dataset Domain

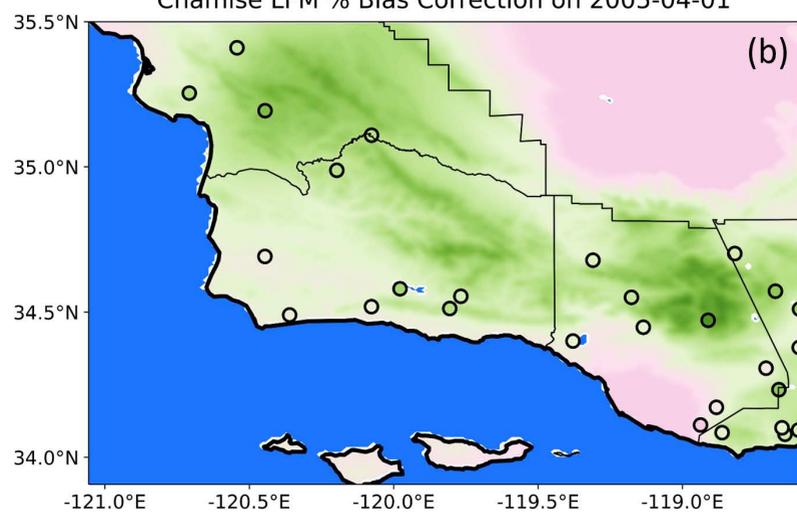
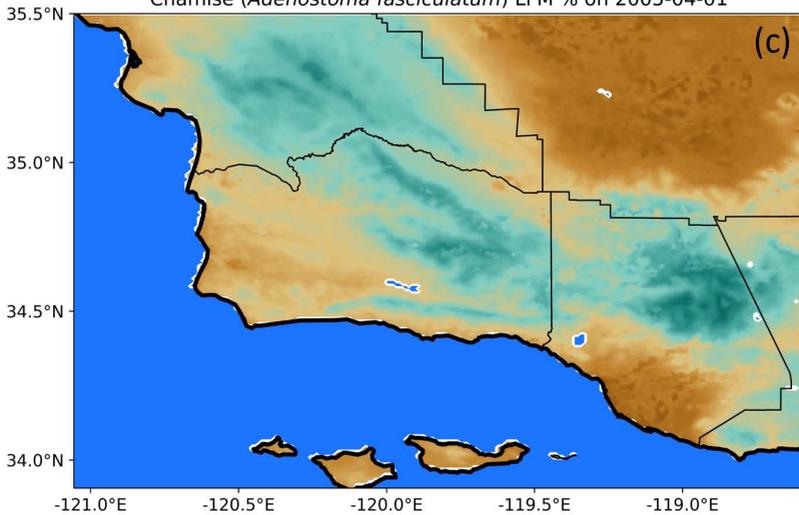




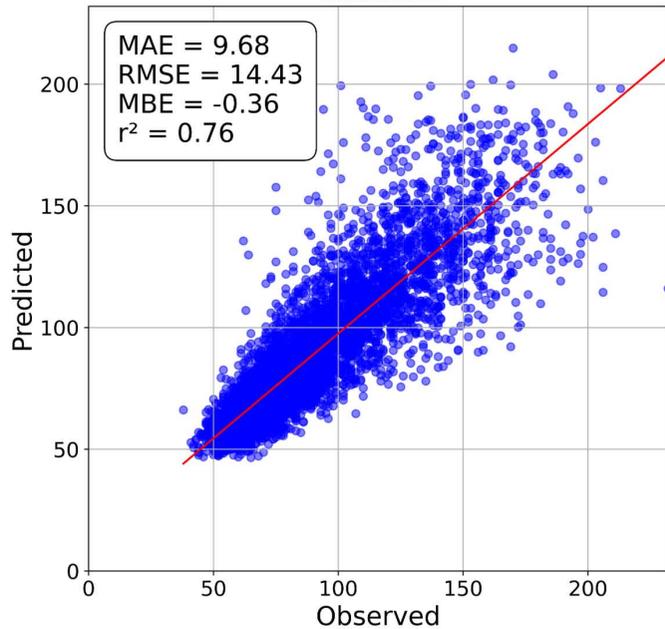


Chamise (*Adenostoma fasciculatum*) LFM % on 2005-04-01

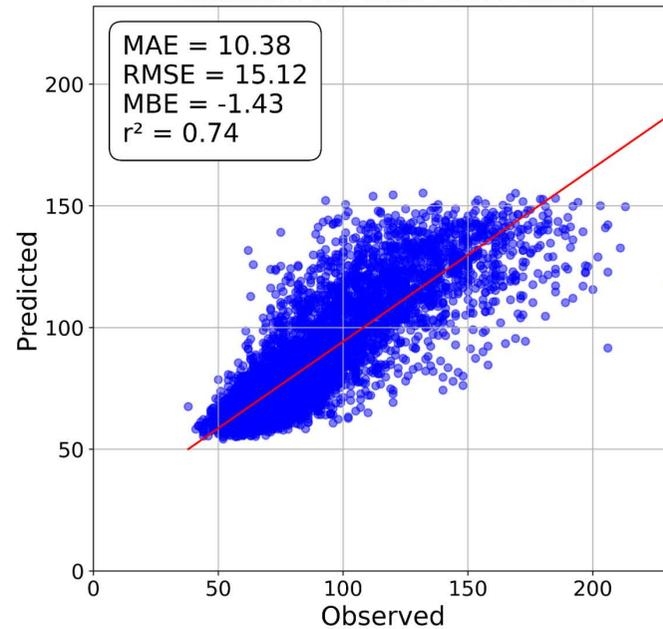
Chamise LFM % Bias Correction on 2005-04-01

Chamise (*Adenostoma fasciculatum*) LFM % on 2005-04-01

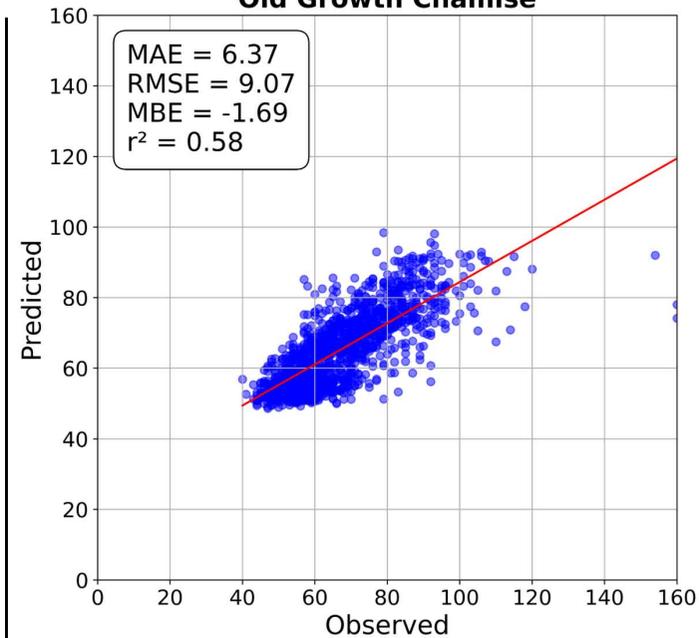
Chamise



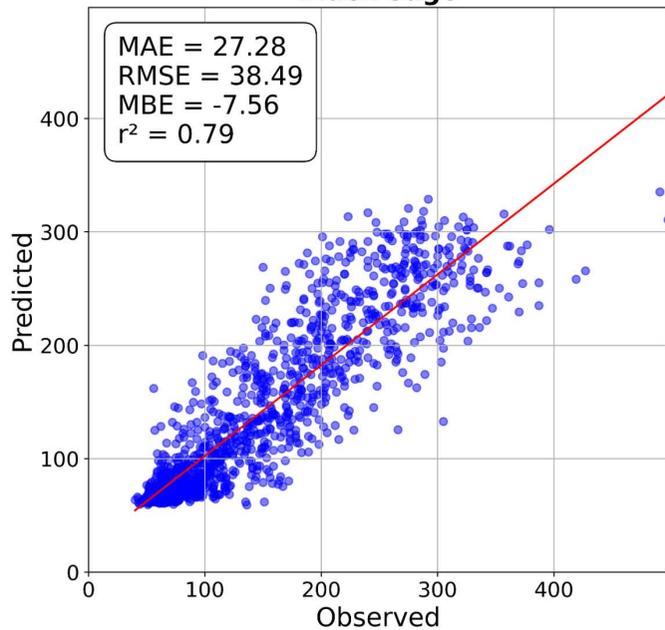
Chamise No Bias Correction



Old Growth Chamise



Black Sage



Bigpod Ceanothus

