



**Center for Sustainability
and the Global Environment**
NELSON INSTITUTE FOR ENVIRONMENTAL STUDIES
UNIVERSITY OF WISCONSIN-MADISON

**Description of the Approach, Data and Analytical Methods used for the
Farms Under Threat Projections of Climate-related Crop and Land-use
Suitability and Sea Level Rise**

Technical Report

Oct 27, 2022

By

Seth A. Spawn-Lee and Tyler J. Lark

University of Wisconsin-Madison

Nelson Institute Center for Sustainability and the Global Environment

in partnership with

American Farmland Trust

and

Conservation Science Partners, Inc.

Recommended citation:

Spawn-Lee and Lark. (2022). Description of the Approach, Data and Analytical Methods used for the Farms Under Threat Projections of Climate-related Crop and Land-use Suitability and Sea Level Rise. Technical Report. May 31, 2022. Center for Sustainability and the Global Environment, University of Wisconsin-Madison.

Table of Contents

1. Introduction	3
2. Methods	4
2.1. Crop-specific Suitability Modeling	4
2.1.1. Training Data Selection	4
2.1.2. Gridded Covariates	8
2.1.3. Model specification and refinement.	10
2.1.4. Spatial Prediction	11
2.1.5. Validation.....	11
2.1.6. Variable Importance	13
2.1.7. Derivative Predictions	16
2.2. Broad Land Use Class Suitability Modeling	16
2.2.1. Training data selection	16
2.2.2. Gridded Covariates.....	17
2.2.3. Model Specification and Refinement.....	17
2.2.4. Spatial prediction	18
2.2.5. Validation.....	18
2.2.6 Variable Importance	19
2.2.7 Derivative layers	20
2.3. Sea-Level Rise Modeling	20

1. Introduction

This document describes the methods used to generate spatial projections of (1) the future suitable geographies of three indicator crops, (2) the future suitable geographies of broadly defined agricultural land use classes, and (3) areas inundated due to future sea level rise. Projections of each were generated for two climate change scenarios—IPCC Representative Concentration Pathways (RCPs) 2.6 and 8.5—in the years 2020, 2040, and 2060. These projections support American Farmland Trust’s Farms Under Threat (FUT) data series and our description of them presumes basic familiarity with past FUT research products and reports. We based the decision to generate spatial projections out to 2060 on the assumption that we might not see significant changes until 2050. However, this was not the case and the model showed significant changes by 2040. Since it is much more difficult to interpret the 2060 projections, the FUT products and reports focus on the changes by 2040. The data set generated using the methods described in this document include:

- Projected apple, corn, and winter wheat suitability in 2020, 2040, and 2060 under RCPs 2.6 and 8.5. Suitability here is defined as the “probability of cultivation” such that values range from 0 to 1 and is not an indication of expected yield or productivity. Corn and winter wheat projections reflect expected rain-fed geographies whereas projected apple geographies are agnostic to water source.
- Projected cropland, rangeland, pastureland, and other land use suitability in 2020, 2040, and 2060 under RCPs 2.6 and 8.5. Again, suitability is defined here as the probability of a given land use class being present such that values range from 0 to 1. Projected geographies for these land use classes reflect un-irrigated conditions.
- Changes in suitability for the above crops and land use classes from 2020 to 2040 and 2020 to 2060.
- Estimates of “the most” suitable geographies of apple, corn, and winter wheat in 2020, 2040, and 2060 under RCPs 2.6 and 8.5. These are simply binary (yes/no: 1, 0) projections, representing areas with the highest suitability. The suitability cut-off used was crop-specific and based on where a given crop currently occurs.
- Projected land use classes (cropland, pasture, rangeland, other) in 2020, 2040 and 2060 under RCPs 2.6 and 8.5 These projections are based on the aforementioned land use suitability projections and take the form of classified maps wherein each pixel is assigned to the most probable land use class.
- Geographies of sea-level rise inundation in 2020, 2040, and 2060 under RCPs 2.6 and 8.5.

In what follows, we describe the methods used to generate each of these projections and, where possible, provide supporting validation statistics.

2. Methods

2.1. Crop-specific Suitability Modeling

Our goal was to assess and illustrate how climate change will affect the US geographies of three regionally and nationally important indicator crops in 2040 and 2060, relative to 2020. Climate-driven shifts in crop suitability are akin to climate-driven species range shifts and can be assessed using similar methods. Various types of species distribution models (SDMs) have been used to assess species range shifts and vary in terms of the outcomes they aim to predict, the biological processes they address, and the ways in which they address those processes (Elith and Leathwick 2009, Briscoe *et al* 2019). Questions of ‘suitability’, *per se*, generally don’t require the treatment of these latter biological mechanisms and can thus be addressed using correlative approaches that relate presence/absence (P/A) data to spatially resolved covariates. Once specified, these models—sometimes referred to as “niche models”—can be applied to projections of those covariates to predict suitability under alternative climate regimes (e.g. Rehfeldt *et al* 2012).

We used random forest models (Breiman 2001) as correlative SDMs to assess how the probabilities of cultivating various indicator crops change in space and over time in response to global climate changes. Our models relate observed recent geographies of a crop or land cover’s presence, as indicated by various spatial data sources, to its co-occurring biophysical conditions. They ignore more dynamic, non-biophysical attributes of current plantings like socioeconomic factors and management practice for which future changes caused by climate changes cannot be known. We use these P/A and biophysical relationships to predict future crop geographies under various climate projections at select points in future time. In practice, this approach entails four sequential steps:

1. Generating a large, comprehensive set of high-quality training data representing the known P/A of a given crop;
2. Relating these locations to a suite of gridded covariates;
3. Using these data (and the relationships therein) to specify a random forest model that accounts for the relationships between P/A geographies and their covariates, used here as predictors; and
4. Spatially applying this model to gridded projections of the considered covariates to predict suitability at a future time.

2.1.1. Training Data Selection

We used the USDA Cropland Data Layer (CDL) to infer the current suitable geographies of each of our focal crops. The CDL is the most exhaustive and accurate remotely sensed crop type map of its kind due to its basis in a classifier trained using a well-matched, non-public spatial dataset of annual, farmer reported crop type locations (the USDA Common Land Unit). It is produced annually at a 30-m resolution for the conterminous US and has remarkably high user’s accuracy for many crops including those we focus upon in this analysis: apples, corn, and winter wheat (Lark *et al* 2017, 2021). In addition, the CDL is accompanied by a confidence layer that we used to filter training points such that we only consider those that are most confidently indicative of a focal crop’s presence. The CDL allows us to infer

conditions in which a crop is biophysically suited for production based on the geographic location of its observed presence. However, as discussed below, it is an imperfect approximation—one akin to inferring a species fundamental niche from its realized niche. As such the regions we predict to be suitable using this approach are almost certainly conservative as they additionally and implicitly reflect to unknown degrees non-biophysical factors such as infrastructure and economic constraints.

The predictions of correlative models are particularly sensitive to the quality of the presence/absence (P/A) data upon which they are trained (Hengl *et al* 2018). Ideally, training data for suitability modeling should encompass the full range of conditions that a given crop can tolerate. This ideal is challenged by the fact that crops are not always planted in all the regions in which they are suited, nor are they exclusively planted in areas in which suitability is high. We can reasonably infer that the conditions associated with locations where a crop is observed are tolerable to that crop. The corollary, though, is less trivial. Crops may be absent from a location either because it is biophysically intolerable or because socio-economic factors have precluded it from being planted there. For this reason, we cannot be certain that a crop's observed domain truly represents its *full* range of biophysical tolerances, nor can we assume the reason for a crop's apparent absence. P/A issues have long plagued all SDMs and ultimately require specialized treatment (Elith and Leathwick 2009, Briscoe *et al* 2019). For our application, these treatments include selecting focal crops for which we can reasonably assume that locations of observed presence encompass a broad swatch of the crop's full range of suitability and proactively seeking to resolve P/A ambiguities where possible. Nevertheless, these measures remain imperfect. While they may, for example, redress most P/A concerns for economically dominant crops like corn whose range of dryland production maps well to its actual suitability; but, for other crops like winter wheat, its absence may simply reflect the presence of an economically more dominant crop like corn, rather than inhospitable biophysical conditions. Thus, while we endeavor to address such issues where possible, the underlying complexity of crop-choice inhibits our ability to do so and this caveat should be considered when interpreting our projections.

Crop rotations are one ubiquitous factor in annual cropping systems that can contribute to the apparent absence of a crop in an otherwise suitable location simply because the crop's position in a crop rotation did not coincide with the year in which the P/A observation was made. To address this issue, we considered 5 years of crop P/A when selecting training data from the CDL. To do so, we stacked five consecutive years of CDL to reconstruct crop rotations and retain all field locations in which a focal crop was planted at least once, thereby considering these areas to have been suitable for that crop. Further, as with all remotely sensed products there is noise in the CDL classifications such that there may be stray single pixels that erroneously suggest the presence of a given crop when it was in fact never there. To further address this particular issue, we filtered the presence patches attained from the rotational analysis to only retain patches of at least one hectare in area (10 contiguous pixels). Finally, we considered the CDL's corresponding confidence associated with a focal crop's detected presence and removed from consideration those with a confidence score below 90%. Collectively, these treatments should more inclusively assess presence while preventing false positives that may result from noise and misclassification.

Once patches of recent presence were identified, for corn and winter wheat, we further filtered out those known to be irrigated using the annual irrigation maps of Xie *et al* (2021) and the USDA's Census of Agriculture (CoA). Irrigation represents an infrastructural investment meant to decouple suitability from variation or change in climatic suitability. The inclusion of irrigated locations in our training dataset would likely confound our predictions by, for example, suggesting that arid regions may be suitable without explicitly acknowledging that such predictions are predicated on the use of irrigation infrastructure that may or may not actually exist. By omitting irrigated locations from our training set, our suitability predictions thus represent the suitability of rain-fed crops. After experiment with the data of Xie *et al.* (2021) we found that, in some arid regions, it failed to completely mask out all irrigated croplands and that the presence of these instances—albeit rare—was suggesting high suitability in locations for which no rainfed instances of the crop were observed. To address this, we further filtered the training data set by removing any remaining instances of crop presence from counties in which the CoA reported that 50% or more of the given crop was irrigated in the given county. Based on qualitative assessment, this additional filter appeared to effectively reduce instances of spurious prediction. Apples were excluded from this filter as it was determined that the majority of commercial apple production is irrigated. By retaining all apple areas in our training set, it should be assumed that projected geographies presuppose the use of irrigation where necessary.

While the CDL is available through 2020, the irrigation maps of Xie *et al.* are only available through 2017 and thus constrained which years could be used to train the models. As such, we used 2015 as the training year and additionally considered the two years immediately preceding and following 2015 (2013-2017) to account for the potential confounding effects of rotations. The Census of Agriculture is conducted and reported every five years and the most proximate years to our training period were 2012 and 2017. When determining how much of a county's crop was irrigated, we assigned each county the highest value among the two years.

The steps described to this point generated binary maps of each crop's P/A circa 2015. To train the model we needed to generate a sample of these locations and, at those locations, extract covariate information. Because these locations are used to infer the range of each crop's biophysical tolerances, it is imperative that a sampling strategy be used that captures less common, but nonetheless real, conditions to carefully define the extremes of these ranges. To best ensure that our training set was inclusive to this effect, we used a stratified sampling scheme wherein we sampled an equal number of points from each strata. Our strata were derived from the gridded covariates described in the section below using k-means unsupervised classification such that each strata represents a region of covariate similarity. We generated 50 strata and then further stratified each using the P/A layers described above for a total of 100 strata (50 representing areas in which the focal crop was observed and 50 representing its absence). These strata can be viewed in Figure 1 wherein each is randomly assigned a unique color for visualization.

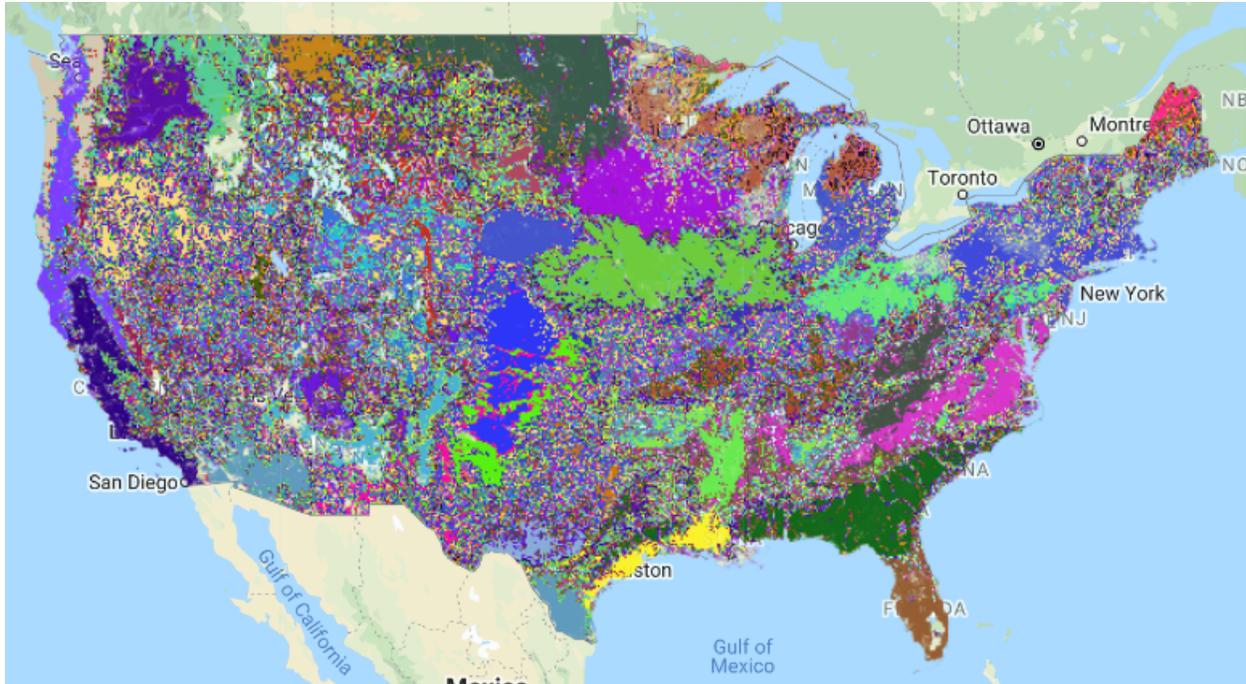


Figure 1. Strata ($n = 50$) used to stratify our random sample when training and validating our models. Each strata represents a region of covariate similarity as inferred from k-means unsupervised classification.

We aimed to sample 300 point locations within each stratum for corn and winter wheat and 500 for apples (apples required more because there were fewer acres of presence and thus, given the lower success rate, greater sampling effort was required to ensure an adequate sample size). The final P/A sample size for all crops is reported in Table 1.

Table 1. Compositional summary of the training data used for each of the crop-specific models. Due to computational constraints we aimed to attain ~30,000 points total for each crop. Because a given crop was sometimes not present in some strata, there were less ‘presence’ points than ‘absence’ points for all three crops.

Crop	Presence Points	Absence Points	Total Points
Apples	7,617	25,000	32,617
Corn	11,246	15,000	26,246
Winter Wheat	12,710	15,000	27,710

2.1.2. Gridded Covariates

As introduced above, our models predict a crop’s P/A probability based on the corresponding values of gridded biophysical covariates representing climate, soil, terrain, and orbital properties.

We used the climate projections of NASA’s Earth Exchange (NEX) Downscaled Climate Projections (NEX-DCP30) for each of the representative concentration pathways (RCPs) 2.6 and 8.5. These projections represent ensemble statistics of the 33 general circulation models used in the Coupled Model Intercomparison Project Phase 5 (CMIP5) and have been downscaled to a 30 arc second resolution for the conterminous US. The NEX-DCP30 dataset includes projections of three distinct climate variables: (i) monthly mean of the daily precipitation rate, (ii) monthly mean of the daily minimum near surface air temperature, and (iii) monthly mean of the daily maximum near surface air temperature. These projections are forecasted from 2006 to 2099 at annual increments for all RCPs. From these NEX-DCP30 variables, we further derived functionally important climate metrics as summarized in Table 2.

Table 2. Gridded covariates are used to train and spatially execute all our suitability models.

Category	Source	Temporal Resolution	Variable
Soil	gNATSGO	Static	<ul style="list-style-type: none"> • Bulk density • Cation exchange capacity • Clay (%) • Depth to restrictive layer • Electrical conductivity • Organic matter content • pH • Sand (%) • Water holding capacity
Terrain	National Elevation Database (NED) [derived]	Static	<ul style="list-style-type: none"> • Aspect • Elevation • Slope

Orbital	Sunrise equation [derived]	Static	<ul style="list-style-type: none"> • Maximum day length
Climate	NEX	Bimonthly average (Jan-Feb; Mar-Apr; May-Jun; Jul-Aug, Sep-Oct; Nov-Dec)	<ul style="list-style-type: none"> • Mean daily precipitation • Mean minimum surface temperature • Mean maximum surface temperature
	NEX [derived]	Bimonthly average (Jan-Feb; Mar-Apr; May-Jun; Jul-Aug, Sep-Oct; Nov-Dec)	<ul style="list-style-type: none"> • Potential evapotranspiration (Thornthwaite approximation) • Aridity index
		Annual	<ul style="list-style-type: none"> • Growing degree days (base temp: 10) • Extreme degree days (base temp: 30) • Total precipitation • Standard deviation of monthly mean daily precipitation • Mean of monthly mean surface temperature range • Standard deviation of monthly maximum surface temperature • Standard deviation of monthly minimum surface temperature

Since NEX-DCP30 projections are forecasted from 2006, we trained our models using the mean of a location’s 2013-2017 projections under RCP 8.5—which most closely tracks observed climate change to date (Schwalm *et al* 2020). This prevents inconsistencies that might have arisen had we instead trained our models using one set of climate observations and then applied it to a separate set of modeled projections.

Whereas climate variables were based on NEX projections, soil and terrain variables were assumed to remain static over the relatively brief interval of time captured by our analyses. All soil variables were taken from the gNATSGO and, with the exception of “depth to restrictive layer,” represent

conditions in surface soils (0-30 cm depth). gNATSGO contains data voids for some variables which we filled using a nearest-neighbor approximation.

Terrain variables were derived from the National Elevation Database, a 10m resolution digital elevation model for the conterminous US.

In practice, at each of the training points derived above, we extracted the corresponding values of all 50 of these covariates and joined them to the points as attributes. These data were then used to train the predictive models as described below.

2.1.3. Model specification and refinement.

The data generated above were used to specify and develop random forest models for each crop. Random forest is a machine learning method that identifies relationships between a dependent variable and any number of supplied covariates by generating and then averaging an ensemble of decision or, in this case, regression trees. This ensemble can then be used for prediction like any decision/regression model by supplying values for each of the covariates—the returned value is the average of the values produced for the supplied set of covariates by each of the trees within the ensemble. There exist a growing number of studies that employ this method for agricultural suitability and yield estimation (e.g. Hoffman *et al* 2020, 2018, Lark *et al* 2020).

When developing a random forest model, users must specify the number of trees and the number of variables that are considered at each split in those trees. To some extent, the model's accuracy is sensitive to these parameters and they can be refined experimentally to reduce prediction error. For each crop, we exhaustively assessed the error of all specifications ranging from 0 to 500 trees and with 1 to all variables being considered at each split. We ultimately selected the configuration that produced the lowest error. Error, in this case, was determined by randomly splitting the training set of a given crop such that 80% of points were used for training and the remaining 20% were withheld to be used for independent validation. Because our models are predicting a probability ranging from 0 to 1 whereas our training data are binary 0 = absences, 1 = presence, we converted our probability estimates to binary for model evaluation [only] by assuming values greater than or equal to 0.5 represented presence (1) and values less than 0.5 indicated absence (0). Doing so, we could then quantify accuracy using confusion matrix methods commonly used in classification that can be succinctly summarized as % accuracy. As such, we sought the random forest configuration for each crop that generated the highest % accuracy (and thus minimized error).

The results of our exhaustive refinement exercise determined the final model configuration that we subsequently used to project suitability. These specifications are summarized in Table 3.

Table 3. Model specifications determined from our exhaustive refinement exercise.

Model	Number of trees	Number of variables	Overall Accuracy (%)
Apples	150	15	99.2
Corn	200	6	90.5
Winter Wheat	200	4	91.8

2.1.4. Spatial Prediction

We next applied the refined models to gridded covariates of the corresponding year and RCP to spatially predict the suitable geographies of each given crop. The resulting maps represent the probability (values range from 0 to 1) of each crop’s P/A on a given date under the specified scenario.

2.1.5. Validation

The model evaluation reported in Table 3 supports the refinement decisions in section 2.1.3 and compares the model’s classification to observed patterns *within the training years (2013-2017)*. For purposes of projection, an assessment of the models’ performance outside of the training window adds more information. To that end, we compared each model’s 2020 RCP 8.5 projection to the geographic patterns observed in the 2020 CDL as a way of validating the model’s ability to project future planted geographies based upon observed circa 2015 P/A. At approximately 50,000 random non-irrigated points distributed evenly across the 50 strata used for model training we compared the observed P/A of the given model’s focal crop to the projected probability of its P/A. To facilitate quantitative comparison of a binary observation (P/A) to a continuous projected probability (0-1), we again classified the projected probabilities of each model as either present (prob. ≥ 0.5) or absent (prob. < 0.5) and then factorially compared these classifications to contemporaneous 2020 P/A observations in contingency matrices (Table 4).

Table 4. Confusion matrices assessing the accuracy of our crop-specific projections for 2020.

APPLES

		Observed (Truth)	
		Absent	Present
Prediction	Absent	23,441	7,010
	Present	159	4,499

CORN

		Observed (Truth)	
		Absent	Present
Prediction	Absent	19,677	6462
	Present	4122	8639

WINTER WHEAT

		Observed (Truth)	
		Absent	Present
Prediction	Absent	20,342	6,933
	Present	3,266	13,144

Using these matrices, we then derived and assessed for each model the following statistics: overall accuracy (%), which represents the percent agreement between predictions and observations; Cohen's kappa, which represents how predictions compare to what might be expected merely by chance (values range from -1 to 1 wherein 1 represents complete agreement, 0 represents no agreement, and negative values indicated agreement that is worse than chance); producer's accuracies of both presence and absence (%), which represents the percent at which observed instances are correctly classified; and user's accuracies of both presence and absence (%), which represents the consistency with which a predicted outcome matches that which was observed at that point. These statistics are summarized for all three models in Table 5.

Table 5. Summary statistics assessing the accuracy of our 2020 crop-specific projections.

	Accuracy (%)	Kappa	Producer's Accuracy (%)		User's Accuracy (%)	
			Absence	Presence	Absence	Presence
Apples	79.58	0.4533***	99.33	39.09	76.98	96.59
Corn	72.79	0.4105***	82.68	57.21	75.28	67.70
Wheat	76.65	0.5235***	86.17	65.47	74.58	80.10

***all Kappa statistics were significantly greater than zero ($p \ll 0.001$)

The results of our classification accuracy assessment indicate that the models' projections, overall, are moderately accurate for use in near term forecasting and produced agreement far greater than would be expected by chance alone. While the overall accuracy of all three models was fairly comparable, the apple model was the most accurate having an overall accuracy of nearly 80%—which is often considered excellent in the context of remotely sensed land cover classification—and a kappa suggesting that patterns aren't likely due to mere chance. Corn had the lowest classification accuracy (~73%) and though it's kappa was significantly greater than 0; $p < 0.001$). The user's and producer's accuracies of all three models suggest that the models consistently map presence where crops are observed to be present, but they do not capture all presences. A user of these projections can more-or-less (depending on the model) reliably expect the focal crop to be present where the model predicts it to occur. By contrast, from the producer's accuracy perspective, the models are less reliable at predicting and capturing all observed locations of each crop. This is most pronounced for apples and likely reflects the small sample size of apple presence in both the training and test data. Nonetheless, both producer's, users, and overall accuracy are remarkably high (generally $> 70\%$) – they are comparable to the accuracy of remotely sensed land cover maps that classify based on direct observation of contemporary earth surface reflectance, whereas our projections are devoid of any such direct observations and are instead based purely upon statistical associations between P/A and climatic, biophysical, and physiographic properties. Interestingly, the apple and corn models tend to predict presence where absences are observed more than wheat suggesting that the planted area of these two crops may not exhaust their suitable ranges.

2.1.6. Variable Importance

To cursorily assess the factors most prominently driving our projections, we quantified variable importance using two different metrics: (i) mean decrease in impurity and (ii) mean decrease in accuracy. Both metrics generally describe the degree to which the accuracy of the model's prediction decreases when the variable in question is removed from consideration—a large decrease in prediction accuracy or purity suggests that the withheld variable is more important than a variable whose removal less effected prediction accuracy or purity. While both metrics seek to quantify the same effect, they do so in slightly different ways and can thus yield different results. It is thus most informative to consider the results of both approaches when assessing variable importance. Important results are also sensitive to the data used

to train the model. To assess this sensitivity, we used 10-fold cross validation to quantify the mean and standard deviation of each variable's importance score for each of the three crop-specific models.

For all three models there was generally good agreement in the ranking of variables by both importance metrics. A pattern emerged in which a single terrain variable (elevation for apples and slope for corn and wheat) was either the most or one of the most important variables according to importance metrics (Figure 2.). We surmise that this reflects a prominent decision within each of the models that partitions high mountains (having both high slope and elevation) in which cropland is largely absent from the flatter landscapes in which most agriculture inhabits. Variables related to summer moisture (monthly precipitation, monthly aridity, and soil available water content) also proved to be particularly important by both metrics and for all three crops suggesting that crop geographies are largely constrained by moisture dynamics. As such our projections primarily reflect an expected response to changing moisture regimes.

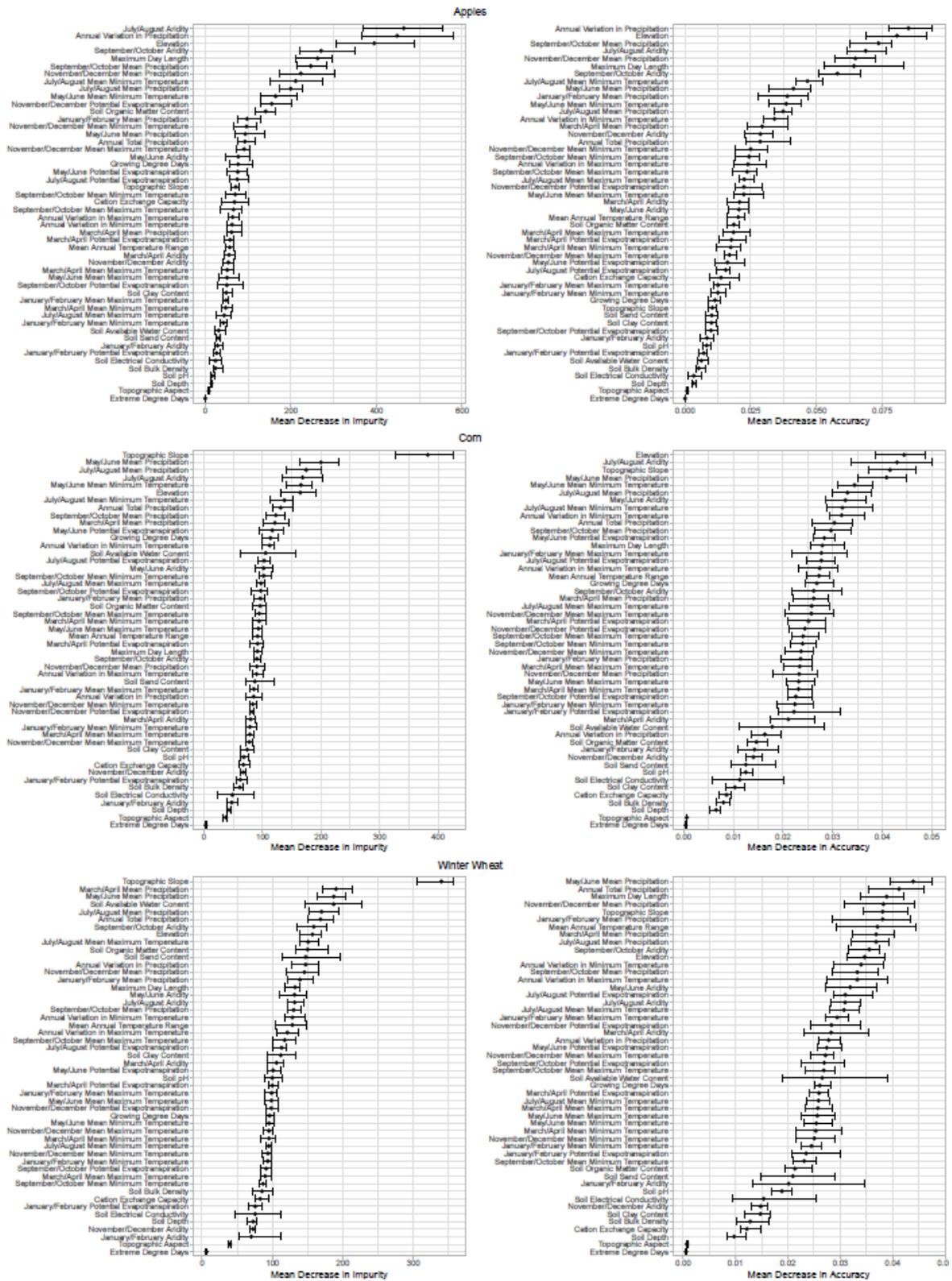


Figure 2. Variable importance plots for the apple, corn, and winter wheat random forest models. Both “mean decrease in impurity” and “mean decrease in accuracy” are reported for each model as two distinct

methods of quantifying importance. For both metrics and all three models, variable importance was assessed using 10-fold cross validation such that the mean importance score is shown as the center point and the standard deviation of the 10 score estimates is shown as the whiskers.

2.1.7. Derivative Predictions

To more succinctly illustrate and summarize the shifting geographies of our focal crops, we defined the most suitable areas as those having a suitability value greater than a derived, crop-specific threshold value. These thresholds were derived by first generating suitability projection for 2015 using the specified model and then masking this projection to areas in which the given crop was observed to be present circa 2015 (as described in section 2.1.1) to estimate the statistical distribution of suitability within planted pixels. The threshold was determined as the median value of these pixels and “most suitable” areas were mapped as those having a suitability value that exceeded this threshold. As such, these areas don’t identify areas that ‘significantly’ contribute to a crops total production, but rather are a heuristic representing the geographies of relatively high suitability. The “most suitable” layers delivered to AFT were further masked to areas identified by AFT as being either cropland, pastureland, or rangeland in their 2016 FUT map.

In addition, we also assessed the projected change in suitability for all three crops between 2020 and 2040 and 2020 and 2060 by differencing the respective maps.

2.2. Broad Land Use Class Suitability Modeling

The methods used to project the geographies of four broadly defined land use classes (i.e., cropland, pastureland, rangeland, other) were similar to those described above for projecting crop-specific geographies. In particular, the means of data selection, covariate treatment, model specification, model evaluation, and model execution were essentially the same aside from the distinctions noted in the subsequent text.

2.2.1. Training data selection

Rather than focusing exclusively on one of three focal crops, our methodology for broad land use classes instead aimed to predict the probabilities of all four land use classes simultaneously using a single model trained using one comprehensive dataset such that the probabilities of all four classes are directly comparable and sum to 1. This approach enables us to determine the most probable land use in each pixel at a given point in time as that having the highest predicted probability.

Since this application seeks to project the probable geographies of broad land use classes (not just their biophysical suitability) with an understanding that these classes compete for the same finite space, many of the P/A issues described above—particularly those related to presence as proxy of suitability and the interpretation of absence—are less applicable in this context. Instead, we accept that a number of unresolved factors (genetics, economics, tradition, etc.) likely dictate their arrangement on the landscape as we see it today and the model results assume that similar socioeconomic factors would operate in the future. As such, we caution that sharp changes in markets or biotic stresses (like diseases) would cause

deviations that are not accounted for in this random forest model. By taking a broad sample of each class's presence throughout the conterminous US we are implicitly "baking in" many unresolved mechanisms in our training data set such that our predictions will represent *the probable geographies due to climate change, all else being equal*.

To that end, we collected training data representing a class's observed presence for each of the broad land use classes as determined by the FUT 2016 landcover map with minor modifications. Upon inspection of the 2016 FUT map, we noticed that roads throughout much of the US were misclassified as either pasture or rangeland. To prevent this issue from confounding our projections, we used the 2016 NLCD impervious surface layer to identify roads, buffered them by two pixels to ensure full removal and then masked these areas out from further consideration. We also excluded irrigated cropland and pasture as described above using the irrigation maps of Xie et al. 2021.

The same stratified sampling scheme used for the crop-specific models—including the same 50 covariate-based strata—was used to ensure a comprehensive sample. However, rather than further partitioning these 50 strata into two subsets representing areas of a given crop's observed P/A, strata were instead partitioned into four substrata based on the observed presence of each of the four focal land use classes for a total of 200 strata. Thereafter, 100 samples were then taken from each stratum. The final training set included 19,885 points including 5,000 from croplands, pasturelands, and other classes and 4885 from rangeland areas. Note, too, that only 'presence' locations were collected since absence of one land use can be inferred from the concurrent presence of another (this is discussed further in section 2.2.3).

2.2.2. *Gridded Covariates*

The same covariates were used in the same manner as described for the crop-specific analyses (Table 2).

2.2.3. *Model Specification and Refinement*

Rather than predicting the probability of a given land use class's occurrence using crop-specific models that disregard the influence of other land uses, as was done for the crop-specific modeling, here we instead predicted the probabilities of all four land use classes using a single model such that their probabilities would be in direct competition and sum to one. To do so, we used the same training data to specify four separate (but intimately related due to their dependence on the same training dataset) sub-models. When running each of the four sub-models, the sub-model's focal class was assigned a value of one (i.e., "present") in the training data set and the remaining three classes were set to zero (i.e., absent).

For the purposes of model refinement, we assessed the overall model's *classification* accuracy—that is, its ability to assign the highest probability to the observed class—in search of the configuration that maximized prediction accuracy (and thus minimized error). Again, the dynamic variables in this exercise were the number of trees generated by the model and the minimum number of variables considered at each branching node within those trees. As with the crop specific models, we split the training data 80/20 in this exercise such that 80% of the training points were used to train the model and

the withheld 20% were used to assess the accuracy of its predictions. Our assessment found that the optimal configuration involved 300 trees in which 7 variables were considered at each split.

2.2.4. Spatial prediction

Rather than one output layer, our broad land use class methodology produced four layers (each representing the probability of one of the four classes) for each year and RCP on a scale of zero to one. The final layers delivered to AFT were further masked to areas identified by AFT as being either cropland, pastureland, or rangeland in their 2016 FUT map.

2.2.5. Validation

Because our land use model was trained using the FUT 2016 land cover map which was singularly produced using an expansive combination of land use/cover datasets, there exists no comparable product representing a later date that could be used to independently validate our model’s forecasting performance. As such, we instead compared the classification accuracy of our model back against that of the 2016 layer at roughly 200,000 random non-irrigated points distributed evenly amongst the 50 strata described above in section 2.1.1. We constructed a confusion matrix (Table 6) using these points from which we derived the same accuracy statistics as described above in section 2.1.5 (Table 7).

Table 6. Confusion matrix comparing the predictions of our broad land use model to the classes observed according to the 2016 land cover map of the previous FUT report.

		Observed			
		Cropland	Pasture	Rangeland	Other
Predicted	Cropland	18,752	7,978	2,072	4,699
	Pasture	10,393	25,693	6,163	8,131
	Rangeland	5,875	7,180	34,461	15,382
	Other	3,860	7,514	5,871	21,546

Table 7. Summary statistics derived from Table 6 describing the accuracy of our broad land use classification.

Overall Accuracy (%)	Kappa	Producer’s Accuracy (%)				User’s Accuracy (%)			
		Crop	Pasture	Range	Other	Crop	Pasture	Range	Other
54.13	0.3857***	48.23	53.12	70.96	43.30	55.97	51.00	54.79	55.54

***The Kappa statistics is significantly greater than zero (p << 0.001)

The model's overall classification accuracy is moderate at 54% and the kappa statistic confidently suggests that its predictions are distinct from those one might expect to result from chance alone. Producer's and user's accuracies were highest for the grassland classes (pasture and rangeland) and lowest for cropland and 'other'. Interestingly there was relatively low confusion between the two grassland classes. Instead, cropland was often confused with pasture more than any other class and may reflect similarity and thus underlying confusion in the training data sources between crops like alfalfa, hay, and even small grains and pasture grasses. Likewise, the catch-all 'other' class was confused with rangeland more than any other class and may reflect the breadth of both classes—our training data for rangeland (the 2016 FUT land cover map) classifies much of the American west as rangeland therein lumping numerous herbaceous and barren ecosystems together while 'other' is the broadest thematic class of the four and likely includes lands with similar characteristics to many of those encompassed by rangeland.

2.2.6 Variable Importance

The same variable importance assessment described above for the crop specific models (see section 2.1.6) were applied to the broad land use model as well (Figure 3).

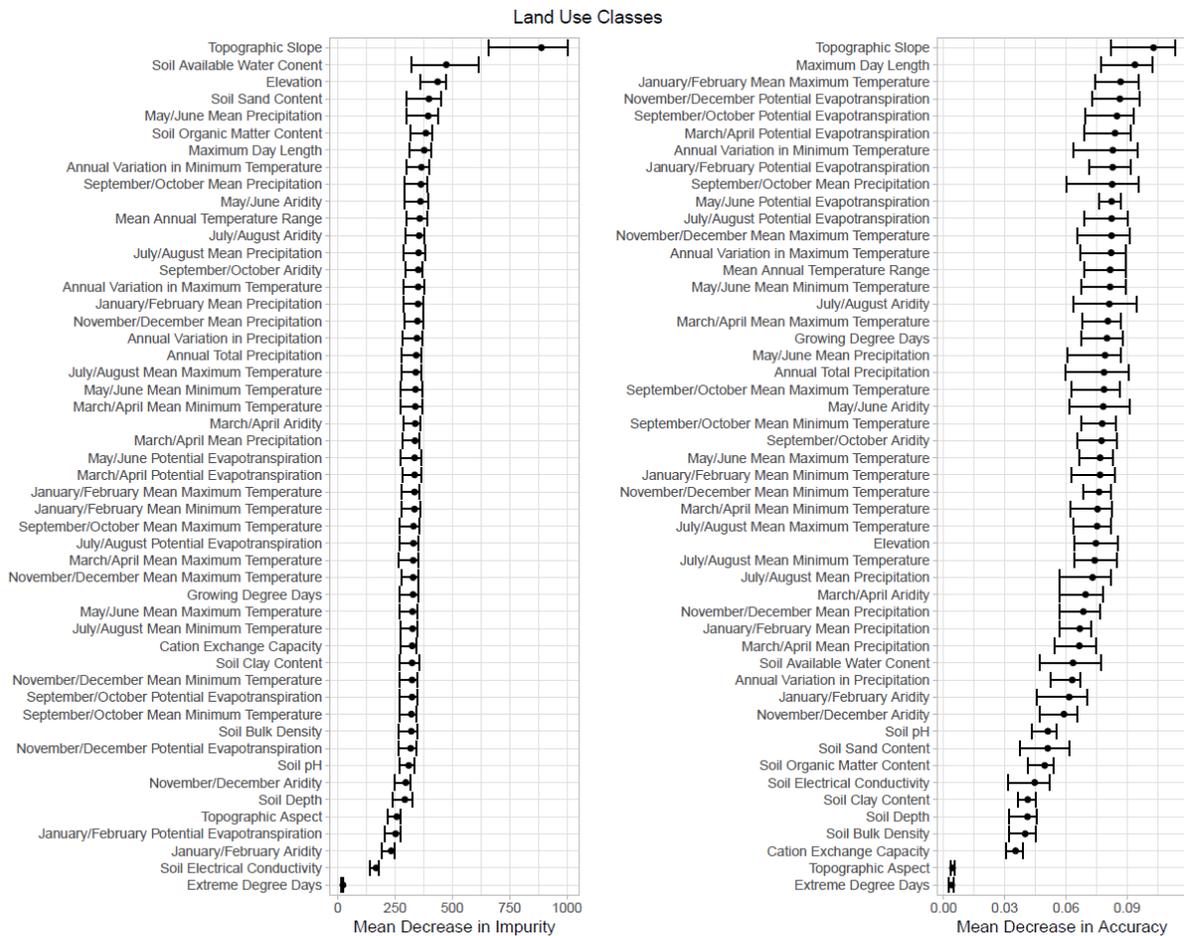


Figure 3. Variable importance plots for the broad land use random forest model. Both “mean decrease in impurity” and “mean decrease in accuracy” are reported for each model as two distinct methods of quantifying importance. For both metrics and all three models, variable importance was assessed using 10-fold cross validation such that the mean importance score is shown as the center point and the standard deviation of the 10 score estimates is shown as the whiskers.

Unlike the crop-specific models, however, we found less agreement between the two importance metrics. While slope was the dominant variable identified by both metrics, there was substantial disagreement in the metrics’ rankings of the subsequent variables. Mean decrease in impurity identified soil properties as the next most dominant variables whereas mean decrease in accuracy primarily highlighted evapotranspiration and, to a lesser degree, temperature. Notably, the scores for most variables were indistinguishable from one another when considering their variance. As such it’s less clear which factors are driving our broad land use predictions which is to some degree unsurprising considering the complexity of modeling four heavily aggregated land use types simultaneously and that are effectively competing against one another.

2.2.7 Derivative layers

Many additional layers were derived from the broad land use projections, including the most probable land use class for each year/scenario combination. These were derived pixel-wise as simply the class which had the highest predicted suitability.

In addition, we also assessed the projected change in suitability for all four landcover types between 2020 and 2040 and 2020 and 2060 by differencing the respective maps.

All of these derivative layers were also masked to areas identified by AFT as being either cropland, pastureland, or rangeland in their 2016 FUT map.

2.3. Sea-Level Rise Modeling

We mapped projected sea level rise in 2020, 2040, and 2060 under RCP 2.6 and 8.5 following NOAA methods (NOAA, 2017). In brief, we first used VDatum to map the mean higher high-water level (MHHW) which represents the average height of the highest tide recorded each day during the recording period. This was done for all CONUS coastal regions. We then identified the projected global mean sea level rise (m) expected by 2020, 2040, and 2060 under each scenario as reported in the IPCC’s AR5 (Table 8) and added these values as constants to the MHHW raster giving us projected estimates of MHHW in each year under each RCP. Finally, we mapped these projections to a DEM wherein an elevation of zero represented mean sea level at present—all regions of the DEM having elevations less than our projections were assumed to have been inundated (at least once, however briefly). This method identified two non-coastal areas as inundated despite no inundated fluvial connection to the ocean: near Death Valley, CA, and surrounding the Salton Sea in southern California. These are areas that lie below

sea-level but are not currently inundated—we thus assumed predicted flooding of these areas to be an artefact of the approach and removed these regions by masking out the counties containing them.

Table 8. Global mean sea level rise (m) projected by the IPCC’s AR5 in 2020, 2040, and 2060 under RCPs 2.6 and 8.5.

Year	RCP 2.6	RCP 8.5
2020	0.08	0.08
2040	0.17	0.19
2060	0.26	0.33

References

- Breiman L 2001 Random Forests *Machine Learning* **45** 5–32
- Briscoe N J et al 2019 Forecasting species range dynamics with process-explicit models: matching methods to applications *Ecology Letters* **22** 1940–56
- Conservation Science Partners (CSP). 2020. Description of the approach, data, and analytical methods used for the Farms Under Threat: State of the States project, version 2.0. Final Technical Report. Truckee, CA.
- Elith J and Leathwick J R 2009 Species Distribution Models: Ecological Explanation and Prediction Across Space and Time *Annual Review of Ecology, Evolution, and Systematics* **40** 677–97
- Hengl T, Nussbaum M, Wright M N, Heuvelink G B M and Gräler B 2018 Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables *PeerJ* **6** e5518
- Hoffman A L, Kemanian A R and Forest C E 2018 Analysis of climate signals in the crop yield record of sub-Saharan Africa *Global Change Biology* **24** 143–57
- Hoffman A L, Kemanian A R and Forest C E 2020 The response of maize, sorghum, and soybean yield to growing-phase climate revealed with machine learning *Environ. Res. Lett.* **15** 094013
- Lark T J, Mueller R M, Johnson D M and Gibbs H K 2017 Measuring land-use and land-cover change using the U.S. department of agriculture’s cropland data layer: Cautions and recommendations *International Journal of Applied Earth Observation and Geoinformation* **62** 224–35
- Lark T J, Schelly I H and Gibbs H K 2021 Accuracy, Bias, and Improvements in Mapping Crops and Cropland across the United States Using the USDA Cropland Data Layer *Remote Sensing* **13** 968
- Lark T J, Spawn S A, Bougie M and Gibbs H K 2020 Cropland expansion in the United States produces marginal yields at high costs to wildlife *Nature Communications* **11** 4295
- Rehfeldt G E, Crookston N L, Sáenz-Romero C and Campbell E M 2012 North American vegetation model for land-use planning in a changing climate: a solution to large classification problems *Ecological Applications* **22** 119–41
- Schwalm C R, Glendon S and Duffy P B 2020 RCP8.5 tracks cumulative CO2 emissions *PNAS* **117** 19656–7
- Xie Y, Gibbs H K and Lark T J 2021 Landsat-based Irrigation Dataset (LANID): 30-m resolution maps of irrigation distribution, frequency, and change for the U.S., 1997–2017 *Earth System Science Data Discussions* 1–32