Mapping agricultural tile drainage in the US Midwest using explainable random forest machine learning and satellite imagery

Luwen Wan [a, b, c, *], Anthony D. Kendall [a], Jeremy Rapp [a], David W. Hyndman [a, d]
[a] *Department of Earth and Environmental Sciences, Michigan State University, East Lansing, MI, 48824, USA*
[b] *Department of Earth System Science, Stanford University, Stanford, CA, 94305, USA*
[c] *Institute for Human-Centered Artificial Intelligence, Stanford University, Stanford, CA, 94305, USA*
[d] *Department of Geosciences, The University of Texas at Dallas, Richardson, TX, 75080, USA*

* *Corresponding author:*
Luwen Wan, 367 Panama St, Green Earth Sciences, Stanford, CA 94305
E-mail address: luwenwan@stanford.edu (L. Wan), kendal30@msu.edu (A. D. Kendall), rappjer1@msu.edu (J. Rapp), hyndman@utdallas.edu (D. W. Hyndman)

Abstract

There has been an increase in tile drained area across the US Midwest and other regions worldwide due to agricultural expansion, intensification, and climate variability. Despite this growth, spatially explicit tile drainage maps remain scarce, which limits the accuracy of hydrologic modeling and implementation of nutrient reduction strategies. Here, we developed a machine-learning model to provide a Spatially Explicit Estimate of Tile Drainage (SEETileDrain) across the US Midwest in 2017 at a 30-m resolution. This model used 31 satellite-derived and environmental features after removing less important and highly correlated features. It was trained with 60,938 tile and non-tile ground truth points within the Google Earth Engine cloud-computing platform. We also used multiple feature importance metrics and Accumulated Local Effects to interpret the machine learning model. The results show that our model achieved good accuracy, with 96 % of points classified correctly and an F1 score of 0.90. When tile drainage area is aggregated to the county scale, it agreed well ($r^2 = 0.69$) with the reported area from the Ag Census. We found that Land Surface Temperature (LST) along with climate- and soil-related features were the most important factors for classification. The top-ranked feature is the median summer nighttime LST, followed by median summer soil moisture percent. This study demonstrates the potential of applying satellite remote sensing to map spatially explicit agricultural tile drainage across large regions. The results should be useful for land use change monitoring and hydrologic and nutrient models, including those designed to achieve cost-effective agricultural water and nutrient management strategies. The algorithms developed here should also be applicable for other remote sensing mapping applications.

Keywords: agricultural tile drainage, random forest classification, feature importance, Google Earth Engine (GEE), Landsat, US Midwest

1 Introduction

Tile drainage is generally installed to remove excess water and enhance crop productivity in poorly drained and humid regions, particularly in areas with high precipitation and shallow groundwater tables (Hirt and Volk, 2011; ICID, 2018; Kokulan, 2019; Møller et al., 2018). Tile drainage installation has several perceived benefits including increased soil aeration, healthier and deeper root systems, optimal moisture conditions for crop growth, and more productive harvests (King et al., 2015; Schilling and Helmers, 2008; Skaggs et al., 1994). The hydrological effects of tile drainage are complex and depend on factors such as rainfall amount and intensity, soil types, and existing soil moisture conditions (Valayamkunnath et al., 2022). Miller and Lyon (2021) found that areas with a high percentage of tile drainage (>40 % of the watershed area) had flashy streamflow in 59 Ohio watersheds. In contrast, Adelsperger et al. (2023) analyzed 139 watersheds in agricultural regions of the US Midwest and found that tile drainage reduced flashiness. Studies have consistently shown that tile drainage can exacerbate nutrient losses from agricultural lands, which can enhance eutrophication in receiving water bodies such as the Gulf of Mexico and the Great Lakes (King et al., 2015; Ma et al., 2023; Rabalais and Turner, 2019; Ren et al., 2022; Smith et al., 2015). Accurately modeling streamflow and subsequent nutrient dynamics in tile-drained agricultural fields can be challenging due to the scarcity of detailed tile drainage data (White et al., 2022). Spatially-explicit tile drainage information is needed to quantify their environmental effects and inform more cost-effective management efforts.

Installing tile drains can help agricultural producers adapt to climate change, yet our ability to predict their effects is limited by the lack of accurate maps of tile drainage locations and practices. Information on tile drainage is often limited in spatial and temporal resolution, even in data-rich agricultural regions like the Midwestern United States. The USDA Census of Agriculture (Ag Census) estimates the area drained by tile through surveys conducted with farmers in counties across the continental United States (CONUS), which are aggregated at the county level every five years (NASS, 2017; USDA-NASS, 2012). The accuracy of these data may be affected by participation rates and respondent accuracy, potentially leading to inconsistencies (Jame et al., 2022). Existing tile drainage datasets often rely on Geographic Information Systems (GIS) analyses that assume agricultural areas with flat and poorly drained soils will likely have tile drainage installed (Sugg, 2007; Nakagaki et al., 2016; Nakagaki and Wieczorek, 2016; Valayamkunnath et al., 2020; Jame et al., 2022). Table 1 summarizes existing tile drainage products and methods. These estimates rely solely on geospatial analyses to identify *likely* tile-drained areas. Developing a cost-effective approach to map spatially explicit *actual* tile drainage installations would be a substantial improvement.

Table 1. Existing tile drainage products and methods.

Product Name	**Resolution and study area**	**Method Description**	**Publication**
Estimated subsurface drainage	County, 18 leading drainage states	GIS analysis based on row crops with poorly drained soil	Sugg, 2007
SubsurfaceDrainExtentUS_1990s	30-m, national scale	Used county area from Sugg, 2007, cropland and poorly drained soil	Nakagaki et al., 2016
USDA_NASS_2012	County, CONUS	Survey-based Ag census	USDA-NASS, 2012
SubsurfaceDrainExtentMW_2012	30-m, 12 Midwest States	Based on the 2012 NASS, cropland, poorly to moderately poorly drained soils	Nakagaki and Wieczorek, 2016

USDA_NASS_2017 *	County, CONUS	Survey-based Ag census	USDA-NASS, 2017
AgTile-US *	30-m, CONUS	Geospatial analysis based on cropland, slope, and soil drainage class	Valayamkunnath et al., 2020
TD-MostPD (TD-AllPD) *	30-m, 12 Corn Belt states	Areas with very poorly and poorly (and somewhat poorly) drained soils	Jame et al., 2022
SEETileDrain	30-m, 14 Midwestern and Great Lakes states	Derived from satellite and environmental datasets with machine learning	This study

Note: * datasets of tile drainage as of 2017 are compared in section 4.2 with the Spatially Explicit Estimate of Tile Drainage (SEETileDrain) product generated here.

Researchers have used thermal and aerial images to estimate tile drainage extent (Naz and Bowling, 2008; Prinds et al., 2019; Tilahun and Seyoum, 2020; Woo et al., 2019) and have mapped individual tile drains and estimated drainage spacing using high-resolution aerial imagery (Naz et al., 2009; Naz and Bowling, 2008). The extent and type of tile drains at a few edge-of-field sites in Michigan and Wisconsin were delineated using high-resolution imagery acquired with unmanned aircraft systems (Webber and Williamson, 2021). Another approach involved using an image differencing technique to delineate tile drainage area for a site in Indiana, comparing shortwave infrared reflectance (SWIR) before and after a ~2.5cm rainfall event (Gökkaya et al., 2017). SWIR is strongly related to soil moisture, and soil with tile drainage tends to dry faster resulting in higher SWIR values. However, high-resolution aerial imagery approaches can be costly, and the image differencing method is susceptible to variable weather conditions such as rainfall intensity and cloud cover. An application of convolutional neural networks has recently been developed to delineate tile drainage at the field scale, although this may limit its broad applicability (Redoloza et al., 2023; Song et al., 2021; Woo et al., 2023). Although several studies have estimated tile drainage at field- to watershed-scales, few have developed drainage maps over broad regions. A random forest model was developed to map tile drainage and reported reasonable overall accuracy rates in the Red River basin (87%) and the Bois de Sioux Watershed (77%) in Minnesota over multiple years (Cho et al., 2019). This model used vegetation indices from Landsat imagery, combined with thermal-moisture and climate-land variables, and assumed tile drainage permit records are ground 'truth' measurements.

There is insufficient spatially explicit and well-validated information regarding tile drainage extent in the US Midwest, which includes 93% of the tile drained area in the United States (USDA-NASS, 2017). Here, we mapped agricultural tile drainage by integrating satellite-derived, climate- and soil-related variables with comprehensive ground truth points in 2017 using the Google Earth Engine (GEE) cloud computing platform. This study aims to (1) provide spatially explicit 30-m estimates of tile drainage for the US Midwest in 2017, (2) identify important features for tile drainage classification across this region, and (3) provide an explainable framework to apply machine learning in agrohydrology. The spatially explicit tile drainage dataset, SEETileDrain, generated here has numerous potential applications in hydrological modeling, water quality assessment, and crop management. It offers valuable insights for environmental managers seeking to optimize agricultural water and nutrient management practices. The machine learning algorithms employed here can also map historical tile drainage with appropriate inputs, identify changes in drained area over time, and establish a baseline to predict future tile drainage installations in response to changes in climate and land use.

2 Study area

The study region includes 14 states in the central US (12 US Midwest states: Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin; and two Great Lakes states: Pennsylvania and New York). This region has a generally sub-humid to humid continental climate with warm to hot summers (Peel et al., 2007). The annual rainfall decreases from east to west, with average yearly precipitation and evapotranspiration of 860 and 634 mm, respectively (Abatzoglou et al., 2018).

The Midwest region is known for its deep, fertile soil with high concentrations of organic matter. It is suitable for cultivation of corn, soybeans, sorghum, alfalfa hay, cotton, wheat, and more. The Midwest has earned its reputation as the “agricultural heartland” of the US and is one of the most extensively cultivated agricultural areas globally (FAO, 2017). The United States produces over 30% of the world's soybeans and corn (USDA, 2023); the Midwest comprises 34 % of the country’s agricultural area (Fig. 1a). Agriculture is thus integral to the local economies of the US Midwest.

As shown in Fig. 1b, approximately 70% of the soil in the Midwest is classified as excessively- or well-drained (yellow-green), based on the natural drainage classification provided by Soil Survey Geographic Database, gSSURGO (USDA 2013, USDA 2017). The remaining 30% falls into soil drainage classes (blue), including very poorly drained, poorly drained, and somewhat poorly drained. The eastern Midwest is mainly lowlands, with elevations gradually increasing towards the west. The median slope across the region is 3.5° (Fig. 1c).

The US Midwest is among the most productive agricultural regions in the world, due in part to its extensive tile drainage systems (Fausey et al., 1995). The 14-state region identified in Figure 1 includes 92.9 % (208,358 km^2 of 224,190 km^2) of the tile drained land across CONUS; 21.6 % of the agricultural land in this region is tile-drained (USDA-NASS, 2017). However, tile drainage data for 183 counties within these states (of 1177 total counties) are not reported due to the absence of tile drains or withheld details to protect individual farm privacy (see NA-counties in Fig.1d). To maintain a continuous boundary while excluding most NA counties, our study region is confined to 737 counties delineated by the heavy blue line in Fig. 1d. The subset, hereafter referred to as the ‘US Midwest’, comprises 204,842 km^2, or 91.4%, of the tile-drained CONUS in 2017 (Fig. 1d).

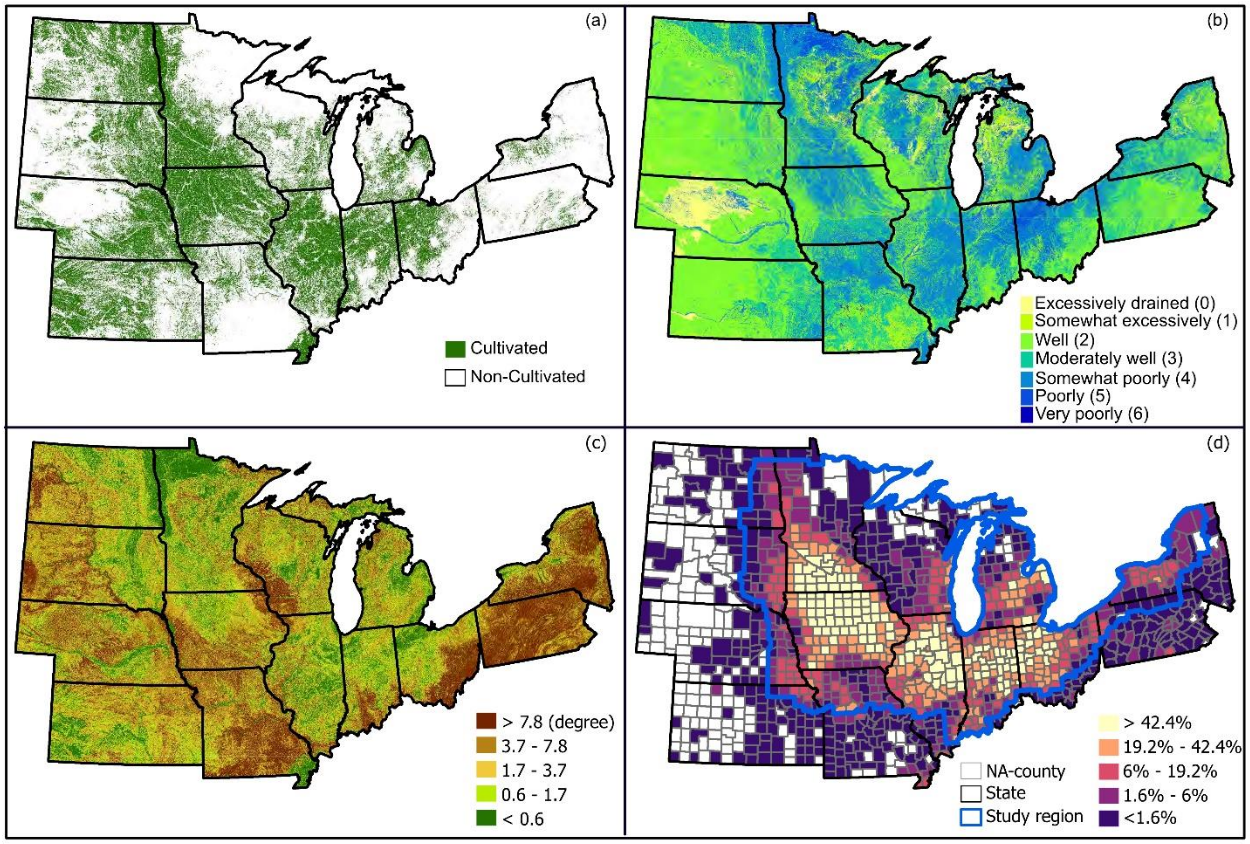

Fig. 1. Study region (737 counties delineated in (1d)) showing **(a)** cultivated and non-cultivated areas based on the 2016 National Land Cover Database and the 2017 Cropland Dataset Layer; **(b)** soil drainage class, classification is from natural drainage classes in the USDA SSURGO dataset, numbers in parentheses are used to visualize distribution difference and classification; **(c)** mean slope in degrees; and **(d)** percent of land area by county in tile drainage based on the 2017 Ag Census. In (d), areas with no tile drains or withheld data are shown as white NA counties due to privacy concerns, as these counties contain few farms.

3 Data and methods

The workflow we developed for creating SEETileDrain at 30-m resolution across the US Midwest is shown in Fig. 2. Geographic, remote sensing, and meteorological products at a variety of resolutions were first selected (light orange boxes). Then, a set of derived variables were computed from these primary sources (darker orange boxes). Next, we generated groups of tile and non-tile ground truth points from various sources across the US Midwest region (light green boxes). These ground truth points and input features were then utilized to train an initial random forest machine learning model that was used to select a final variable subset (light blue box). Additionally, feature importance and correlation were evaluated during this step. The final model was trained using only the more important feature from each highly correlated pair (dark blue boxes), thereby reducing multicollinearity. Through this process, tile drainage area across the Midwest was mapped, accuracy assessments were performed at pixel- and county levels (dark blue boxes), and feature importance was evaluated (dark green box).

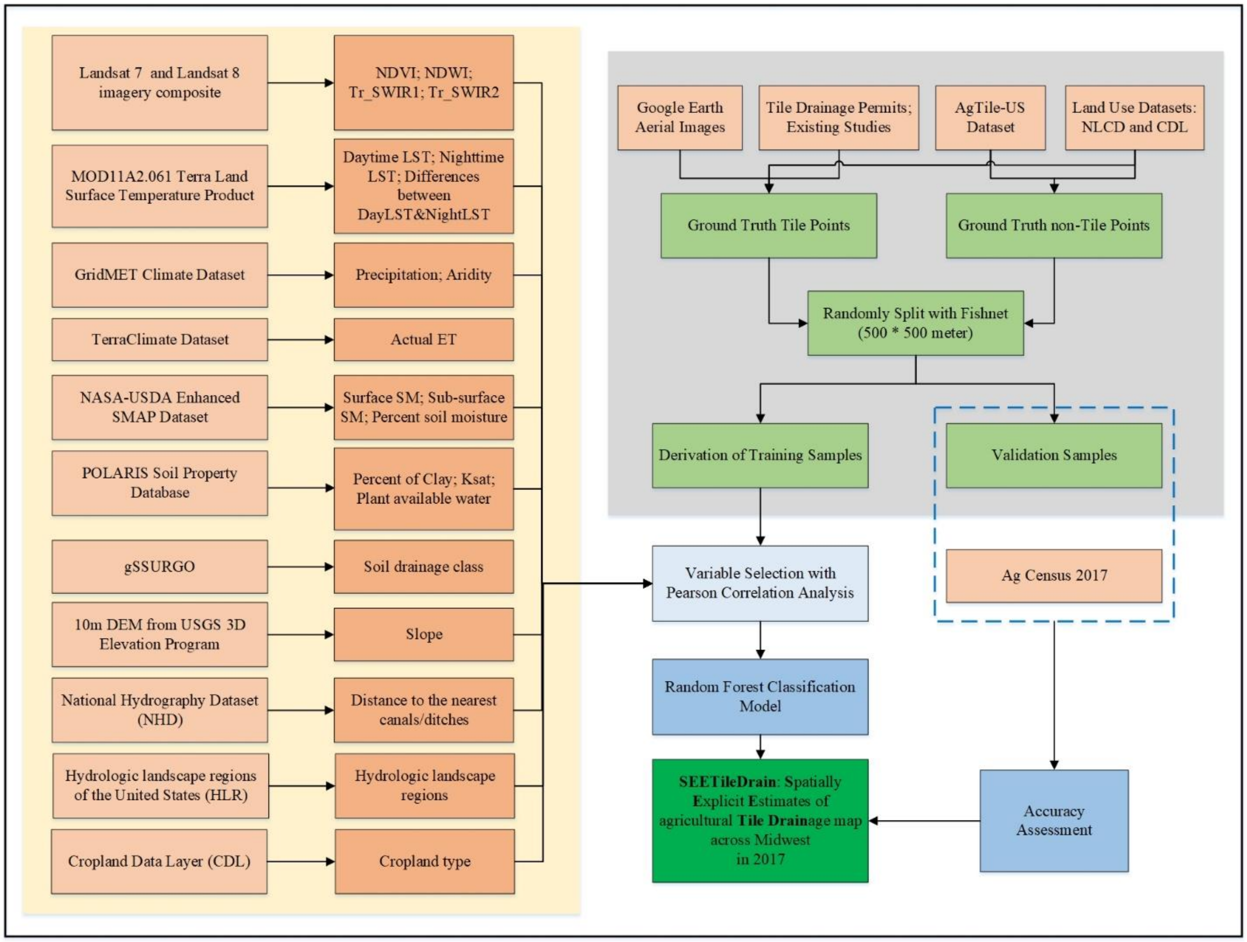

Fig. 2. Workflow diagram of the random forest classification method used to create agricultural tile drainage maps across the US Midwest. This method uses 11 remote sensing and environmental datasets (highlighted in lighter orange boxes) and 21 derived variables (shown in orange boxes). Data sources used to create ground truth points for tile and non-tile, as well as the 2017 Census of Agriculture, are indicated by lighter orange boxes. Ground truth points are shown in light green boxes, classification and assessment methods are shown in blue boxes, and the final product of this study (SEETileDrain) is shown in a green box.

3.1 Variables derived from remote sensing imagery and environmental datasets

Initially, 21 variables were identified from eleven satellite and environmental datasets, as shown in orange and lighter orange boxes. Then, 36 additional features capturing the distinguishing characteristics between tile and non-tile points were derived from these datasets and variables, guided by prior research (Cho et al., 2019; Jame et al., 2022; Valayamkunnath et al., 2020). For instance, fields with tile drainage are expected to have higher Normalized Difference Vegetation Index (NDVI) values throughout the growing season, as tile drains enhance crop growth, and NDVI correlates strongly with green biomass (Prinds et al., 2019). Details about each of these datasets, their sources, how they were computed, and the shortnames assigned to each variable are provided in the Supplemental Text S1, as well as Tables A1, A2 and A3.

These 57 initial variables can be grouped into five categories: 1) climate, 2) static landscape characteristics, 3) surface reflectance indices, 4) land surface temperature (LST), and 5) Soil Moisture Active/Passive (SMAP). With the exception of the static landscape characteristics, each of these variables are dynamic remotely-sensed or meteorological variables. The appropriate periods for these variables were identified by analyzing the differences between tile and non-tile points at approximately 14-day intervals based on two Landsat-derived indices: NDVI and Normalized Difference Water Index (NDWI). Both NDVI and NDWI have been demonstrated to help identify tile drained lands (Cho et al., 2019; Zhang et al., 2014). We then identified three periods - spring: 4/1–5/31, summer: 7/1–8/31, and growing season: 5/1–9/31, based on the mean of maximum (max among available Landsat images) NDVI and NDWI for all the ground truth points across the study region (Fig. A1). For instance, tile points exhibited lower NDVI but higher NDWI in April and May than non-tile points. Thus, we define the spring season as April 1st to May 31st. Periods for other input features match the specified time ranges for NDVI and NDWI as needed.

Except for a few static landscape variables (soil drainage class, HLR, distance to canals or ditches, plant available water (PAW)) that were computed locally and uploaded to GEE, all the data sources mentioned above can be accessed through the Earth Engine's public data archive (developers.google.com/earth-engine/datasets), as well as the awesome-gee-community-catalog (Roy et al., 2023). The native resolution of these data sources varies from 10 to 10,000 m (Tab. A2&A3). The input features were derived by aggregating or disaggregating them to 30-m resolution. The final classification was performed using assets that were reduced and interpolated with the built-in reduce and interpolation functions in GEE. See more details in Text S1.

3.2 Tile and non-tile ground truth point for classification

Tile drainage ground truth points (Fig. 3) were obtained through visual interpretation, literature, and tile drainage permit records to train the classification model. A significant proportion of the tile drainage points were identified through visual interpretation (purple squares in Fig. 3) using the aerial imagery base map from Google Earth Pro. Potentially drained points were randomly selected from likely tile drained areas (Valayamkunnath et al. 2020) within cultivated lands. Drainage status was then visually interpreted based on patterns of tile drains, spaces, and canal ditches around the fields using Google Earth Pro. From this process, 1753 tile drained points were positively identified. Those not positively identified as tile drained were omitted, as there are many conditions that can lead tiled fields to be visually indistinguishable from non-tiled ones. Another set of 15,118 visually interpreted tile drainage points was obtained from a previous study that used a geospatial model to determine the likely tile-drained areas of the CONUS (Valayamkunnath et al., 2020). More details about each of these datasets can be found in Supplemental Text S3. The workflow for generating ground truth points is shown in Fig. A2.

Additional tile drainage information from the literature (orange) and tile drainage permits (green) were available as polylines and polygons. The polylines include sources such as tile lines and drainage structures from three sites (Blissfield, Clayton, and Palmyra) in Michigan, US, as well as tile lines from the Story County Farm in Iowa, tile line permits from the Bois de Sioux Watershed District, and a photo of a tile-drained field in the southeast research farm in Iowa. The polygons used in this study were obtained from permits for agricultural subsurface drainage tile locations provided by USGS in North Dakota (Finocchiaro, 2016) and South Dakota (Finocchiaro, 2014). These polygons are assumed to represent ground truth tile drainage measurements. To ensure their accuracy, we randomly sampled points from these polygons and polylines areas and excluded any points too close to the edges of polygons and polylines.

To avoid biased classification due to clustered tile points from available data and the spatial correlation among adjacent points, a proximity limit between points selected for training and testing was set using a 500*500-m fishnet grid. Yan and Roy (2016) reported that US farms' average and median field sizes are 400 and 527 m, respectively. Here, a threshold of 500 m was selected to provide training and testing point separation, thus handling co-variation while maintaining an adequate number of points for model training and validation. Following this process, a total of 28,723 tile drainage points remained.

The potential tile-drained layer AgTile-US (Valayamkunnath et al., 2020) was also used to identify non-tiled points. We created a mask using a 120-m buffer around all likely tile drained areas. Those areas within agricultural land but outside the mask were defined as likely non-tiled. We then randomly selected 32,215 points, each at least 500 m apart from this layer and used as likely non-tile points in the classification. Although these points may not accurately reflect actual ground conditions, it is practical to utilize them for classification due to the limited availability of non-tile data and the associated labor costs for manual identification.

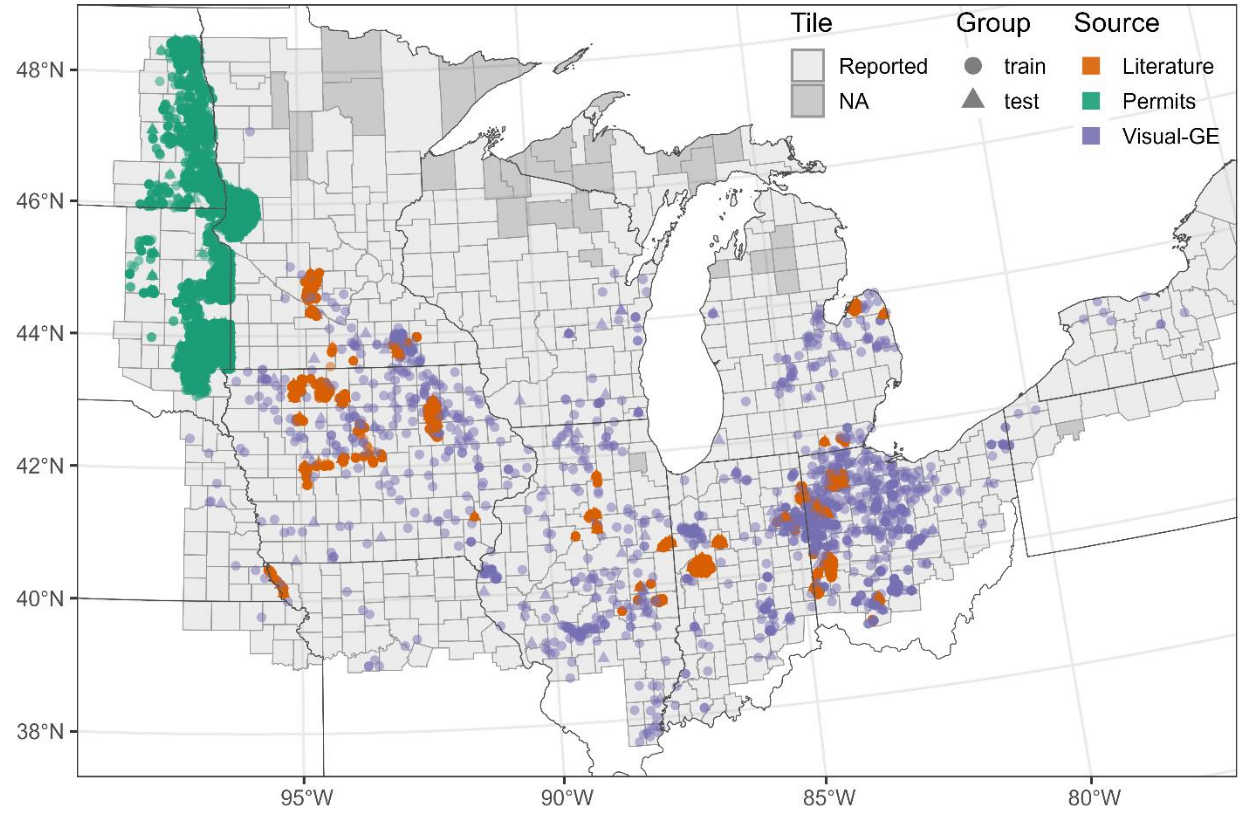

Fig. 3. Map of ground truth tile drainage points sourced from literature (orange), tile drainage permits (green), and visual interpretation from Google Earth Pro aerial imagery (purple). The training and testing tile drainage points are ≥ 500 m apart, as indicated with different point shapes. The study area's counties are highlighted in gray, with the ones reported by USDA-NASS to have tile drainage displayed in light gray. Areas marked as "NA" indicate no tile drains or withheld data due to privacy.

3.3 Random forest classification and accuracy assessment

Random Forest (RF) was chosen for classification because it effectively handles non-monotonic relationships between features and accommodates non-linear relationships. It also reduces the likelihood of overfitting, a common issue in machine learning, by generating random subsets of features and constructing multiple decision trees based on these subsets (Breiman, 2001). Therefore, RF classification has been widely used for various classification tasks, including irrigation mapping, flood risk assessment, and water quality predictions (Belgiu and Drăguţ, 2016; Cho et al., 2019; Deines et al., 2019, 2017; Wang et al., 2015; Xie and Lark, 2021; Gupta et al., 2023).

The RF classifiers are tuned by varying the number of decision trees (*ntree*) and the number of features randomly selected and tested for the best split when growing trees (*mtry*). A review of random forest in remote sensing found that most research sets *ntree* to 500 (Belgiu and Drăguţ, 2016). This is because the error rate stabilizes before reaching this number of trees (Lawrence et al., 2006), and the *randomForest* R package defaults to 500 for *ntree* (Liaw and Wiener, 2002). The classification accuracy is highly dependent on the *mtry* parameter, which is typically set to the square root of the number of input features. However, it can be adjusted between one and the maximum number of input features. Using the Google Earth Engine cloud computing platform, the classifier applies the knowledge gained from the training data to make predictions for areas in out-of-sample regions. We used the default settings of 500 trees (*ntree*) and the square root of the number of features as features per split (*mtry*) for classification.

Here, random forest classification was performed twice: an initial classification with all 57 features, and a final classification and mapping conducted with the 31 selected (Table A2) after eliminating highly covariate features (Table A3). Cross-validation techniques were employed to train and validate the machine learning model using different proportions of data. The model was trained on balanced data and validated on imbalanced data proportionate to tiled and non-tiled areas. To split our ground truth dataset, we first divided non-tile points, randomly selecting 30% for validation (22,540/9675 training/validation). Next, the number of validation tile drainage points was determined based on the overall area ratio (~1:4) of tile and non-tile in the US Midwest region of the Ag Census (USDA-NASS, 2017). Thus 2281 tiled points were selected for validation, leaving 26,442 for training. Overall, we used 49,982 points in training, and 11,956 in validation (81 % training/19 % testing).

An initial classification model was trained on all features, which was then used as the basis for reducing our total number of features, as many of our input features were highly correlated. This is not strictly necessary for accurate classification because RF is generally robust to multicollinearity among predictor features. The RF model naturally reduces the variance by building many decision trees based on a random subset of input features and ground truth points. However, the reliability and interpretability of feature importance can be affected by multicollinearity. For instance, multicollinearity can result in biased importance scores where the actual contribution of each feature to the model's predictive power isn't accurately reflected because the model may arbitrarily favor one correlated feature over the others in different trees. By reducing covariation, we can evaluate the unique contribution of each feature and improve the interpretability of the model. With this initial classification, we calculated the Pearson correlation coefficient (r) between each pair of features and used the *MeanDecreaseAccuracy* algorithm to determine feature importance in the initial classification, with *ntree* = 500 and *mtry* set by the square root of the number of features used in the model. If the Pearson correlation between pairs of features exceeded a threshold of 0.8, we retained only the feature with higher importance. For instance, in the first model run with *ntree* = 500 and *mtry* = 7 (~ square root of 57, the number of initial features), 50 pairs of features have a correlation coefficient higher than 0.8. For one example pair, SSM_mediean_spr and SMP_median_spr, the correlation coefficient (r = 0.99) exceeded our threshold; SSM_mediean_spr was dropped at this step based on its lower Mean Decrease in Accuracy. It is important to note that there is no standard threshold for correlations, as they are domain-specific and vary across studies (Schober et al., 2018). Here, a correlation of 0.8 or greater is considered to indicate a strong correlation.

To evaluate the model's accuracy on the testing dataset, we used three metrics: overall accuracy, recall (producer's accuracy), and precision (user's accuracy), see equations (1) – (3). These metrics are commonly used in classification models and directly show the number of correctly identified pixels (Congalton and Green, 2019). Accuracy is the percentage of pixels correctly identified (i.e., as tile or non-tile) in agricultural fields, recall is the percentage of known tile drainage points that were correctly identified, and precision is the number of true positive results (points identified as tile drainage) divided by the number of samples predicted to be positive, including those not identified correctly. Each metric ranges from 0 to 1, with higher values indicating the classification model is more reliable. These metrics were calculated as follows:

$$Overall\ Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$Recall = TP / (TP + FN) \quad (2)$$

$$Precision = TP / (TP + FP) \quad (3)$$

Where TP represents the number of true positives (tile drainage points), TN represents the number of true negatives (non-tile), FP represents the number of false positives (points predicted to be tile drainage that are not tile-drained), and FN represents the number of false negatives (points expected to be non-tile that are tile-drained).

We also calculated the F1 score and balanced accuracy (equations (4) & (5)) to evaluate the predictive performance. F1 score was calculated as the harmonic mean of precision and recall, which gives the same weight to precision and recall (Sasaki, 2007). For this case, the true proportion of tile and non-tile classes differs substantially, so balanced accuracy can provide a more robust metric.

$$F1\ score = 2 * (Precision * Recall) / (Precision + Recall) \quad (4)$$

$$Balanced\ Accuracy = (TP / (TP + FN) + TN / (TN + FP)) / 2 \quad (5)$$

The county-scale accuracy assessment used 2017 Ag Census statistical data. We calculate classified tile drainage areas for each county and then fit a linear regression model to compare them with reported ones.

To ensure the reliability of our tile drainage product for the US Midwest in 2017, we compared the total estimated tile drainage area to three other likely tile drainage maps (AgTile-US, TD-MostPD and TD-AllPD) and reported areas from the Ag Census for the study region in 2017. We also performed a random forest classification using the same training and validation points but only used slope and soil drainage classes, as they are commonly used in other products with GIS analysis. For this simple model, the number of trees (*ntree*) was set to be the same as our full classification (default: 500), and the number of *mtry* was specified as two since there are only two features. We compared the out-of-bag (OOB) error, accuracy and F1 score with our final classification.

3.4 Interpretability and explainability of the machine learning model

We applied four methods to assess and communicate the results of the tile drainage classification model: Mean Decrease in Gini (Gini) (Liaw and Wiener, 2002), Mean Decrease in Accuracy (MDA) (Archer and Kimes, 2008), Shapley Additive Explanations (SHAP) (Lundbery and Lee, 2017), and Accumulated Local Effects (ALE) (Apley and Zhu, 2020). These measures each provide different insights into how the model makes predictions, which input features are important, and how predictions change as input features vary. Applying multiple methods results in a more comprehensive understanding of the random forest machine learning model.

Feature importance helps quantify what features drive model performance and, more specifically, how much each feature improves the model's accuracy. We ran the *importance* function from the *randomForest* package in R (Liaw and Wiener, 2002), as GEE does not provide feature importance measures. To do this, we extracted feature values for training and validation points and developed a proxy random forest classification with the same training and testing data and identical parameter settings (*ntree* and *mtry*) using the *randomForest* packages in R; thus, identically specified, the feature importance values computed in R are likely to mirror their importance in GEE. The *importance* function includes *MeanDecreaseGini* and *MeanDecreaseAccuracy* measures. *MeanDecreaseGini* is an impurity-based importance measure, representing the total decrease in node impurities from splitting on the features, averaged over all trees. *MeanDecreaseAccuracy* is a permutation-based importance, measuring accuracy reduction on out-of-bag samples when the feature values are randomly permuted; *MeanDecreaseAccuracy* is considered a more reliable measure of importance than the *MeanDecreaseGini* (Strobl et al., 2008).

We also computed Shapley values, which assess the contribution of each feature to each individual prediction. Unlike the *MeanDecreaseGini,* the Shapley value measures the average contribution of a feature to the prediction across all possible combinations of features, not the difference in prediction when we remove the feature from the model. To explain the model as a whole (not just using one prediction), and for computational reasons, we randomly decomposed 489 (1 % of the training points) predictions, which allows us to examine the global feature importance using SHAP (the average absolute of 489 Shapley values).

Finally, we computed an overall importance by assigning each input feature a score equal to its rank within each of the three classification algorithms (Gini, MDA, and SHAP) and summing the three scores. Note the highest ranked feature in each metric was assigned a score of N = number of features used in the model, and the lowest a score of 1. The highest possible score (3*N) corresponds to a feature that ranked highest in all three measures of importance.

To visualize the relationship between the value of the predictor features and probability of the targeted class (i.e., tile), we computed and plotted ALE with the *iml* R package (Molnar, 2018). ALE plots produce output similar to the commonly used Partial Dependence Plots, but with distinct advantages. They are generally more robust and efficient, especially when dealing with correlated features and complex data landscapes. This robustness comes from the way ALE calculates and focuses on localized changes around actual data points, rather than over a large or entire range of a feature. By focusing on real data distributions, ALE reduces the risk of extrapolating beyond what the model has learned from the training dataset, improving both the interpretability and accuracy of the results.

4 Results and discussion

4.1 Feature selection, elimination and distribution differences between tile and non-tile

We initially collected 57 input features by selecting among the three time periods (spring, summer, and growing season) and statistics (mean, maximum, range) from 21 variables (orange boxes in Fig 2). Following our initial classification with 57 features, 31 were selected for final classification with paired correlation coefficients < 0.8 (Table A2 and Fig. A3); 26 features (Table A3) were removed by successively eliminating less important features from the highly correlated features pairs as detailed in Text S2. These features were not incorporated into our final classification model as they did not significantly improve prediction accuracy.

We compared bulk distribution differences between the two groups: tile and non-tile ground truth (Figures A4 & A5). Based on the Wilcoxon test in R, our analysis showed a significant difference ($p < 0.0001$) between the two groups for the 30 features, and only one feature (Tr_swir1_grow_max) with no significant difference (ns) between tile and non-tile groups used in the final classification. For median and range differences between tile and non-tile points, see Table A4. Feature distribution differences between tile and non-tile are shown in Text S4. Several alternate input variables were investigated for inclusion in this research, as described in Text S5.

4.2 Classified map from random forest classification and accuracy assessment

Random forest binary classification provides pixel-wise probabilities for a given class, and the probabilities for tile drainage are shown in Fig. 4a. Classified tile-drained areas (Fig. 4b) are identified with probabilities > 0.5 and are concentrated in the Corn Belt region, including the eastern Dakotas, southern Minnesota, north-central Iowa, northeastern Illinois and Indiana, northwestern Ohio, and the Michigan's thumb area. As examples, inset maps in Fig.4b illustrate that the machine learning model generally captures the reported tile fields in the western Lake Erie basin and tile permits in South Dakota.

The point-based assessment for the testing dataset indicates that the classification model achieved good overall accuracy, with a score of 0.96, thus 96 % of tile and non-tile points, are classified correctly. Recall and precision are 0.96 and 0.85, respectively. The balanced accuracy and F1 scores are 0.96 and 0.90, respectively. These metrics demonstrate the good quality of the classification model.

We summed the predicted tile drainage areas for each county from the classified map and compared these with the reported areas from the Ag Census in 2017, as illustrated in Fig. A6. The random forest classification model reasonably agreed with the reported area, with an R^2 value of 0.69. However, the model tended to overestimate tile drainage area, with a best-fit linear slope of 1.1, especially in counties with larger tile drainage reported by NASS. It often underestimates area in counties with smaller reported areas, such as points with reported areas < 500 km^2. Eight states (Illinois, Indiana, Iowa, Michigan, Minnesota, North Dakota, Ohio, and Pennsylvania) had R^2 values ≥ 0.65, while tile drainage areas in Michigan and Pennsylvania were underestimated (Fig. A7). The estimates for the remaining states (Kansas, Missouri, Nebraska, New York, South Dakota, and Wisconsin) are less accurate, likely because they are less heavily tile-drained or have fewer ground truth points.

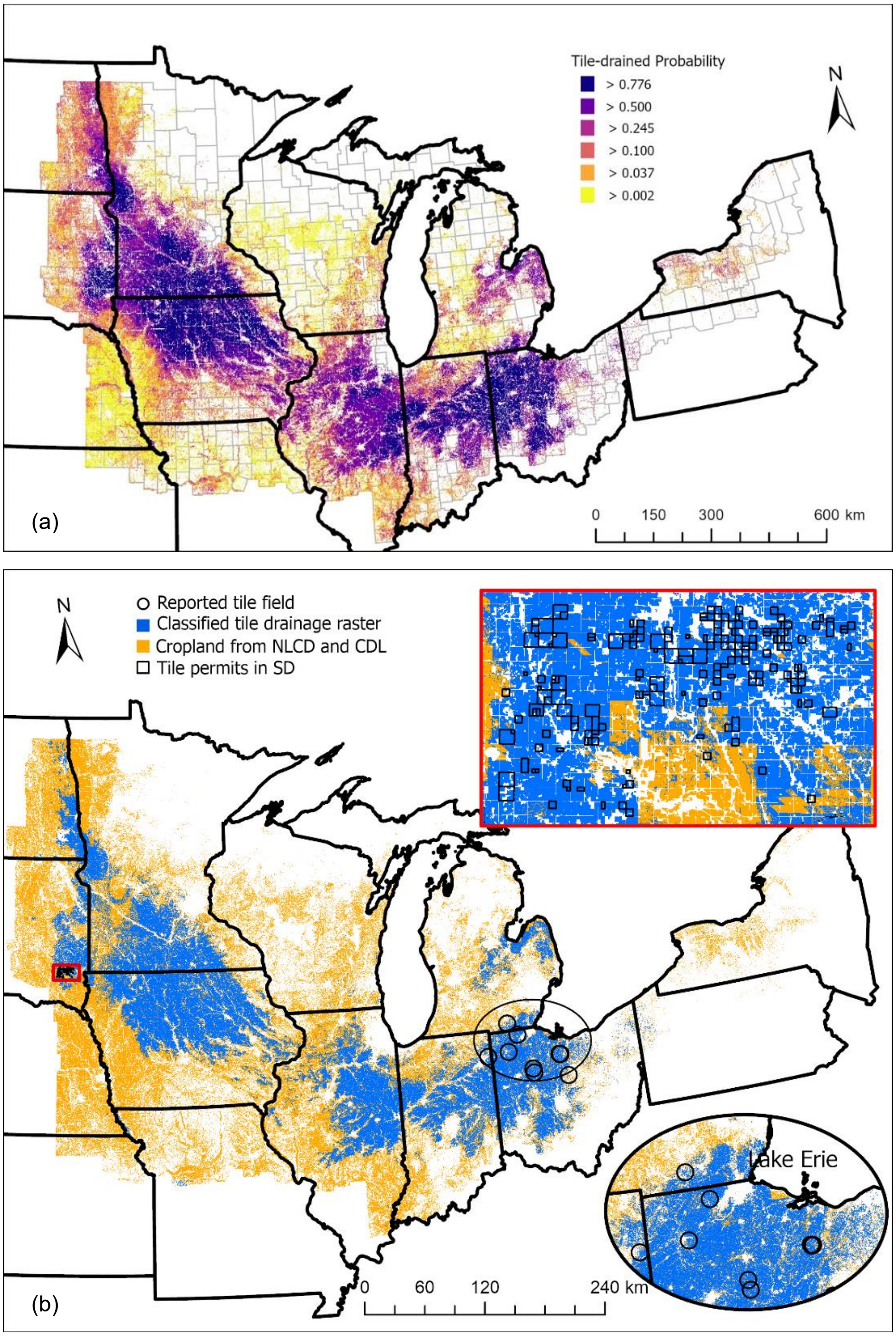

Fig. 4. Random forest classification from GEE showing **(a)** probability of tile drainage at 30-m resolution and **(b)** classified tile drainage map (SEETileDrain, blue) with agricultural land (yellow) as a background, corresponding tile drainage probability > 0.5 in (a). Two zoomed-in windows show the reported tile fields (black circles) in the western Lake Erie basin and tile drainage permits (black rectangles) in South Dakota (red rectangles).

We created a county-level residuals map, calculated as the tiled area percentage difference (predicted-reported) divided by the reported area (Fig. A6). The map revealed overestimation in counties with heavily tile-drained regions, primarily in Eastern Dakota, Southern Minnesota, the Des Moines Lobe in Iowa, Northeastern Illinois, Mid-northern Indiana, and Northwestern Ohio. This is likely because our tile drainage training points are concentrated in these areas, and we assumed tile drainage permits in South Dakota, North Dakota, and the Bois de Sioux Watershed District all represent actual tile installations for ground truth.

We also found that 27 % of counties were not predicted to have any tile drainage installation despite having reported tile drains by NASS (NA-Classified in Fig. A6). These counties have a relatively low reported area, with a median reported area of about 27 km^2 (5 % of agricultural lands). This indicates that our classification model would benefit from more ground truth information from regions with a low percent of tile drains. It is important to note that the area reported by farmers through surveys may be somewhat inaccurate. However, it is the only source available for this comparison across the region.

Table 1 lists various tile drainage products covering multiple Midwestern states, and they estimated tile drainage area in US Midwest (737 counties) as 204,842 km^2 (USDA-NASS), 201,206 km^2 (AgTile-US), 576,493 km^2 (TD-MostPD), 1,025,288 km^2 (TD-AllPD). Our product, SEETileDrain, estimated that 185,549 km^2 were tile drained in 2017, which is ~9.4 % lower than the estimates from survey-based statistics from USDA-NASS. AgTile-US utilizes information on soil drainage and topographic slope threshold within cropland areas to estimate tile drainage and constrained the geospatial model with statistical tile drainage area at the county level, trained on data from the Census of Agriculture in 2017 (Valayamkunnath et al., 2020). The estimate is ~1.8 % lower than the reported areas from USDA-NASS. Another study by (Jame et al., 2022) developed two Transforming Drainage (TD) extent products for the US Midwest based on soil drainage class. These classes were selected because they are related to crop production and are considered more suitable for estimating likely drained land than soil properties alone. The first product, TD-MostPD, includes areas with very poorly and poorly drained soils that are likely tile-drained. The second product, TD-AllPD, includes somewhat poorly drained soils in addition to the two categories used in TD-MostPD. Thus, it’s unsurprising that TD-MostPD and TD-AllPD estimated 2.8 and 5 times more tile drainage extent than the statistics from USDA-NASS for this region, respectively.

For better comparison to these maps (AgTile-US, TD-MostPD, TD-AllPD), which rely primarily on soil drainage class and slope information, we trained a random forest classification model based solely on these inputs of all ground truth points. This simplified model had an out-of-bag (OOB) error rate of 36 %, compared to 4 % for our full model. The model also exhibited lower precision (0.29) and recall (0.66), demonstrating a less accurate classification model based on only two features, compared to our final classification model which used 31 features. This simplified model's F1 score, and balanced accuracy were 0.4 and 0.64, compared to those of 0.9 and 0.96 in our final model with 31 features. This indicates that added features beyond soil drainage class and slope significantly improved the binary classification.

Overall, the random forest classification model used here provides substantial advantages over traditional estimation methods that rely on Ag Census data collected at the county level (USDA-NASS, 2017), which can introduce uncertainty due to the potential reporting errors and scale mismatches. Unlike these methods, which have been used to produce *likely* tile drainage products that were used for area comparison, our model estimates *actual* tile drainage locations and is completely independent of the Ag Census and its inherent errors. This independence allows the model to provide consistent and reliable estimates of tile drainage between survey years, bypassing the limitations typically associated with temporal gaps in data collection. In addition, the spatially explicit nature of SEETileDrain addresses the critical issue of scale by providing high-resolution information at the 30-m pixel scale, providing a more accurate tool to assess and manage agricultural drainage systems.

4.3 Feature importance and model explainability

The importance of 31 input features for tile drainage classification across the US Midwest was assessed using Gini, MDA, and SHAP. AET in the growing season (AET_grow) ranked the highest in the MDA, and the maximum nighttime LST in the summer (NightLST_max_summ) ranked top in the Gini index (Fig. A8); thus, they were in the top right corner of the multi-way importance plot (Fig. A9). Four additional features (aridity in spring, median soil moisture percent in summer, the range of daytime LST in growing season and soil drainage class) ranked high in both Gini and MDA, meaning that removing these features will substantially reduce purity and model accuracy.

SHAP, here average absolute Shapley values from 489 randomly selected predictions from training points, are shown in Fig. 5a. Shapley values determine if features positively or negatively affected the classification accuracy of our model for a given prediction (tile, in this case). It is important to note that Shapley values are the average contribution of an input feature to the target prediction. Thus, it might be biased due to slightly more non-tile points being used in model training. Our results reveal that the most important features are the maximum nighttime LST in the summer (NightLST_max_summ) and soil drainage class.

We computed the overall importance (Fig. 5b) from the three measures (Gini, MDA, SHAP). It revealed that features derived from MODIS products and soil- and climate-related features are the most important. The top-ranking feature is the maximum of nighttime LST during summer. This is somewhat consistent with (Cho et al., 2019), where the mean LST in the spring strongly contributed to the random forest classification for tile drainage in the Bois de Sioux Watershed District in Minnesota and the Red River basin (overlies portions of Dakotas and Minnesota). Accumulated Local Effects to evaluate whether high or low values of the features correspond to higher probabilities of tile drains are shown in Text S6.

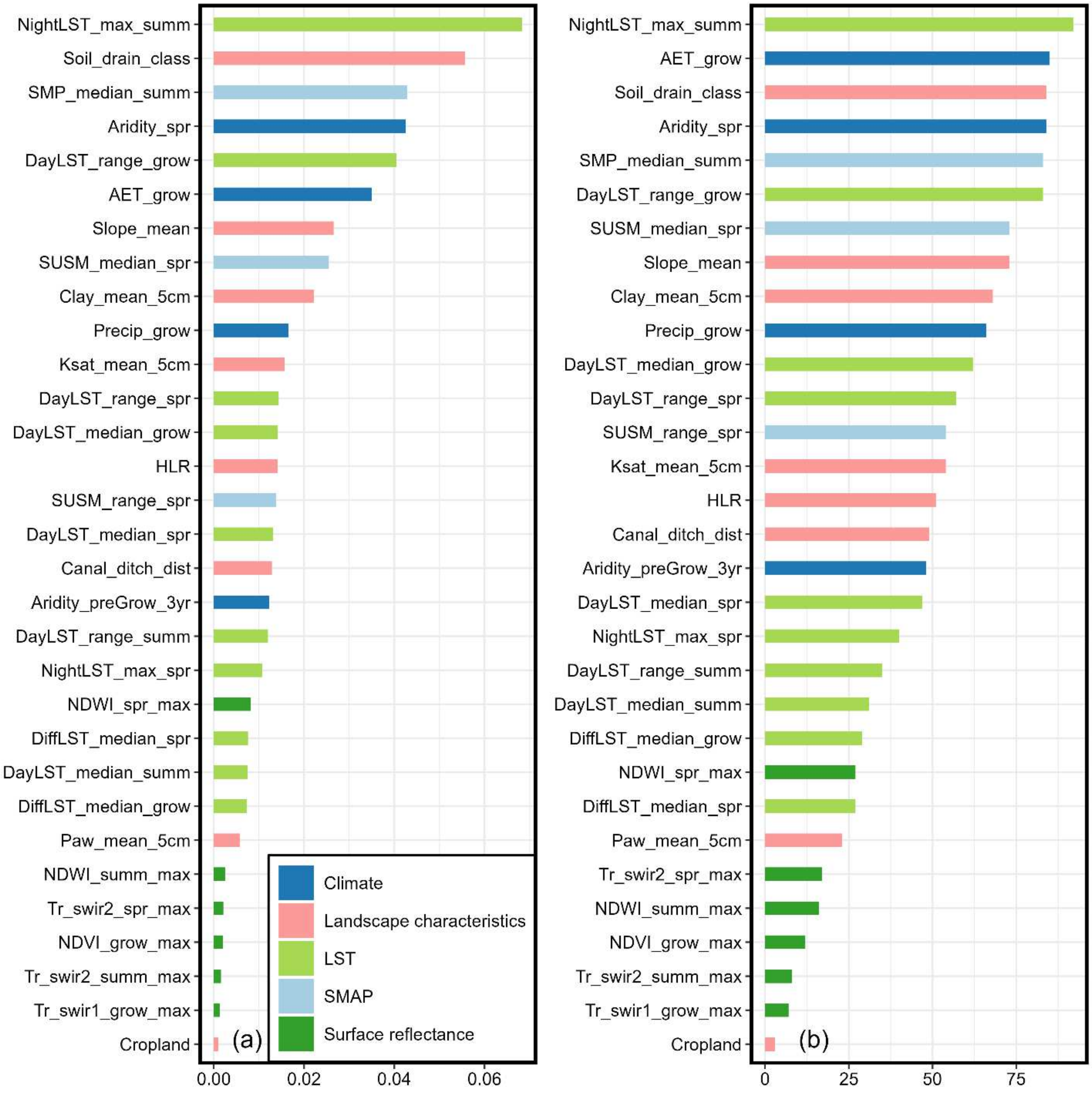

Fig. 5. **(a)** SHAP values, average absolute Shapley values from 489 (1%) predictions and **(b)** overall importance (summed score from MDA, Gini, and SHAP). Different colors in (a) & (b) denote original data category. Features are identified by their short names, which can be related to full names and other information using Table A2.

4.4 Limitations and future implications

The algorithm's accuracy depends on the coverage of ground truth information, which was not evenly spatially distributed and is imbalanced for tile and non-tile points. Other sources of bias include assumptions that all tile drainage permits represent actually drained areas, and potential issues during the visual interpretation. In addition, correlation between variables may still remain and influence the feature importance, though this study successively removed relatively highly correlated (paired correlation coefficients >0.8) but less important features. Although feature co-variation is not often explicitly evaluated in random forest classification, additional preprocessing could be used to reduce the effects of correlated features on the performance of the classification model and improve the robustness of feature importance.

Future work could improve accuracy and efficiency by including advanced deep learning, and visual transformer-based encoder-decoder architectures to identify tile drainage (Breitkopf et al., 2022; Redoloza et al., 2023). The use of high-resolution images, such as the Harmonization of the Landsat and Sentinel-2 data (HLS) (Claverie et al., 2018), images provided by the Planet Lab or Google's aerial imagery, would enable the identification of finer resolution tile drainage, which would be particularly useful at small scales (e.g., field to watershed). Incorporating such techniques and data sources could enhance our understanding of the effects of tile drainage on agricultural landscapes and facilitate the development of more effective management strategies.

The machine learning model developed here could be readily applied to other regions for past and future years, which would be helpful for hydrological, water quality, and crop modeling research. This information may also help watershed managers and stakeholders achieve cost-effective agricultural water and nutrient management strategies while maintaining optimal crop production. As one of the critical water and nutrient transport pathways, a more accurate tile drainage map would improve the estimates of the contributions of this pathway (Michaud et al., 2019; King et al., 2015; Ikenberry et al., 2014). In addition, hydrologic and water quality models across scales would benefit from the method and results. For example, the National Agroecosystems Model (NAM) framework is continually updated with improved techniques and new data (White et al., 2022). By integrating a machine learning algorithm and incorporating the SEETileDrain product, the NAM would better support USDA modeling efforts from the field to the national scale.

5 Conclusions

Spatially explicit agricultural tile drainage across the US Midwest was mapped at 30-m resolution using a random forest machine learning classification model. This model was implemented on the Google Earth Engine cloud computing platform, ultimately selecting 31 features from eleven data sources, using a novel collection of ground truth points. This map product, SEETileDrain, is the first estimate of actual tile drainage across the US Midwest independent of county-level surveys. The resulting classified map demonstrated good accuracy in point-based assessment and reasonable agreement with the reported area from USDA-NASS. Land surface temperature, soil moisture percent, actual evapotranspiration, and soil drainage class were strong predictors for tile drainage identification in the US Midwest.

Declaration of Competing Interest

None.

Acknowledgments

This research was jointly funded by NASA Grants NNX11AC72G and 80NSSC21K1652, USDA-NIFA/NSF award 2018-67003-2740, NOAA Grant NA12OAR4320071, and Michigan State University. Support for this research was also provided by the NSF Long-term Ecological Research Program (DEB 2224712) at the Kellogg Biological Station and by Michigan State University AgBioResearch. Luwen Wan was partially funded by the China Scholarship Council (CSC). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NASA, USDA-NIFA/NSF, NOAA, or CSC.

We are grateful to Eva Arvizu and Jordyn Porter, who visually interpreted tile drainage points via the Google Earth aerial images. We thank Dan Jaynes (Iowa State University), Ghane Ehsan and Jason Piwarshi (Michigan State University), and Nathan Trosen (Moore Engineering, Inc.) for providing ground truth points for this research. We also thank Eunsang Cho (NASA, Goddard Space Flight Center), who shared code and guidance on the machine learning algorithm and David Roy (Michigan State University), Yanhua Xie (The University of Oklahoma) and Jill Deines (Pacific Northwest National Laboratory) for their valuable suggestions to improve this work. The BdSW tile drainage permit records were obtained from the BdSW district in Minnesota (www. bdswd.com) and SD and ND tile drainage records are publicly available from the USGS Science Base website (www. sciencebase.gov).

All satellite and input data used are available through the Google Earth Engine code editor (code.earthengine.google.com) and the awesome-gee-community-catalog (gee-community-catalog.org/), except for soil drainage class, HLR, and distance to the nearest canals and ditches, which are freely available from SSURGO, USGS HLR, and NHD datasets. The PAW layers are derived from POLARIS dataset. The Python scripts for generating PAW layers and the R scripts for selecting variables, implementing the random forest model and visualizing the figures, are available on HydroShare. The SEETileDrain product is also available on HydroShare, along with the probability map. https://doi.org/10.4211/hs.4c8af0c89d2b4686808bcf9c6f7c0da3

References

Abatzoglou, J.T., Dobrowski, S.Z., Parks, S.A., Hegewisch, K.C., 2018. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. Sci. Data 5, 170191. https://doi.org/10.1038/sdata.2017.191.

Adelsperger, S.R., Ficklin, D.L., Robeson, S.M., 2023. Tile drainage as a driver of streamflow flashiness in agricultural areas of the Midwest, USA. Hydrol. Process. 37. https://doi.org/10.1002/hyp.15021.

Apley, D.W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. J. R. Stat. Soc. Series B Stat. Methodology 82(4), 1059-1086. doi:10.1111/rssb.12377.

Archer, K.J., & Kimes, R.V. (2008). Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis*, 52(4), 2249-2260. https://doi:10.1016/j.csda.2007.08.015.

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: a review of applications and future directions. Isprs J. Photogramm 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/a:1010933404324.

Breitkopf, T.-L., Hackel, L.W., Ravanbakhsh, M., Cooke, A.-K., Willkommen, S., Broda, S., Demir, B., 2022. Advanced Deep Learning Architectures for Accurate Detection of Subsurface Tile Drainage Pipes from Remote Sensing Images. arXiv. https://doi.org/10.48550/arxiv.2210.02071.

Cho, E., Jacobs, J.M., Jia, X., Kraatz, S., 2019. Identifying subsurface drainage using satellite big data and machine learning via google earth engine. Water Resour. Res. 55, 8028–8045. https://doi.org/10.1029/2019wr024892.

Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.-C., Skakun, S.V., Justice, C., 2018. The harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sens. Environ. 219, 145–161. https://doi.org/10.1016/j.rse.2018.09.002.

Congalton, R.G., Green, K., 2019. Assessing the Accuracy of Remotely Sensed Data, Principles and Practices, Third edition. https://doi.org/10.1201/9780429052729.

Deines, J.M., Kendall, A.D., Hyndman, D.W., 2017. Annual irrigation dynamics in the U.S. Northern High Plains derived from Landsat satellite data. Geophys. Res. Lett. 44, 9350–9360. https://doi.org/10.1002/2017gl074071.

Deines, J.M., Kendall, A.D., Crowley, M.A., Rapp, J., Cardille, J.A., Hyndman, D.W., 2019. Mapping three decades of annual irrigation across the US High Plains aquifer using Landsat and Google Earth Engine. Remote Sens. Environ. 233, 111400. https://doi.org/10.1016/j.rse.2019.111400.

FAO, 2017. FAO Statistical Databases.

Fausey, N.R., Brown, L.C., Belcher, H.W., Kanwar, R.S., 1995. Drainage and water quality in Great Lakes and Cornbelt states. J. Irrig. Drain. Eng. 121, 283–288. https://doi.org/10.1061/(asce)0733-9437(1995)121:4(283).

Finocchiaro, R.G., 2016. Agricultural Subsurface Drainage Tile Locations by Permits in North Dakota.

Finocchiaro, R.G., 2014. Agricultural Subsurface Drainage Tile Locations by Permits in South Dakota.

Gökkaya, K., Budhathoki, M., Christopher, S.F., Hanrahan, B.R., Tank, J.L., 2017. Subsurface tile drained area detection using GIS and remote sensing in an agricultural watershed. Ecol. Eng. 108, 370–379. https://doi.org/10.1016/j.ecoleng.2017.06.048.

Gupta, A., Hantush, M.M., Govindaraju, R.S., 2023. Sub-monthly time scale forecasting of harmful algal blooms intensity in Lake Erie using remote sensing and machine learning. Sci. Total Environ. 900, 165781. https://doi.org/10.1016/j.scitotenv.2023.165781.

Hirt, U., Volk, M., 2011. Quantifying the proportion of tile-drained land in large river basins. Phys Chem Earth Parts B C 36, 591–598. https://doi.org/10.1016/j.pce.2011.05.004.

ICID, 2018. World Drained Area.

Ikenberry, C.D., Soupir, M.L., Schilling, K.E., Jones, C.S., Seeman, A., 2014. Nitrate-nitrogen export: magnitude and patterns from drainage districts to downstream river basins. J. Environ. Qual. 43, 2024–2033. https://doi.org/10.2134/jeq2014.05.0242.

Jame, S.A., Frankenberger, J., Reinhart, B.D., Bowling, L., 2022. Mapping agricultural drainage extent in the U.S. Corn Belt: the value of multiple methods. Appl. Eng. Agric. 38, 917–930. https://doi.org/10.13031/aea.15226.

King, K.W., Williams, M.R., Macrae, M.L., Fausey, N.R., Frankenberger, J., Smith, D.R., Kleinman, P.J.A., Brown, L.C., 2015. Phosphorus transport in agricultural subsurface drainage: a review. J Environ. Qual. 44, 467–485. https://doi.org/10.2134/jeq2014.04.0163.

Kokulan, V., 2019. Environmental and economic consequences of tile drainage systems in Canada, Canadian Agri-food Policy Institute.

Lawrence, R.L., Wood, S.D., Sheley, R.L., 2006. Mapping invasive plants using hyperspectral imagery and Breiman cutler classifications (randomForest). Remote Sens. Environ. 100, 356–362. https://doi.org/10.1016/j.rse.2005.10.014.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2(3), 18–22.

Lundberg, S.M., & Lee, S.I., 2017. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30* (pp. 4765-4774).

Ma, Z., Guan, K., Peng, B., Sivapalan, M., Li, L., Pan, M., Zhou, W., Warner, R., Zhang, J., 2023. Agricultural nitrate export patterns shaped by crop rotation and tile drainage. Water Res. 229, 119468. https://doi.org/10.1016/j.watres.2022.119468.

Michaud, A.R., Poirier, S., Whalen, J.K., 2019. Tile drainage as a hydrologic pathway for phosphorus export from an agricultural subwatershed. J. Environ. Qual. 48, 64–72. https://doi.org/10.2134/jeq2018.03.0104.

Miller, S.A., Lyon, S.W., 2021. Tile drainage causes flashy streamflow response in Ohio watersheds. Hydrol. Process. 35. https://doi.org/10.1002/hyp.14326.

Møller, A.B., Beucher, A., Iversen, B.V., Greve, M.H., 2018. Predicting artificially drained areas by means of a selective model ensemble. Geoderma 320, 30–42. https://doi.org/10.1016/j.geoderma.2018.01.018.

Molnar, C., 2018. iml: an R package for interpretable machine learning. J. Open Source Softw. 3, 786. https://doi.org/10.21105/joss.00786.

Nakagaki, N., Wieczorek, M.E., 2016. Estimates of Subsurface Tile Drainage Extent for 12 Midwest States, 2012.

Nakagaki, N., Wieczorek, M.E., Qi, S.L., 2016. Estimates of Subsurface Tile Drainage Extent for the Conterminous United States, Early 1990s.

Naz, B.S., Bowling, L.C., 2008. Automated identification of tile lines from remotely sensed data. Trans. ASABE. 51(6), 1937-1950.

Naz, B.S., Ale, S., Bowling, L.C., 2009. Detecting subsurface drainage systems and estimating drain spacing in intensively managed agricultural landscapes. Agr. Water Manage. 96, 627–637. https://doi.org/10.1016/j.agwat.2008.10.002.

Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. Hydrol. Earth Syst. Sc. 11, 1633–1644. https://doi.org/10.5194/hess-11-1633-2007.

Prinds, C., Petersen, R.J., Greve, M.H., Iversen, B.V., 2019. Locating tile drainage outlets and surface flow in riparian lowlands using thermal infrared and RGB-NIR remote sensing. Geogr Tidsskr-den 119, 1–12. https://doi.org/10.1080/00167223.2019.1573408.

Rabalais, N.N., Turner, R.E., 2019. Gulf of Mexico hypoxia: past, present, and future. Limnol. Oceanogr. Bull. 28, 117–124. https://doi.org/10.1002/lob.10351.

Redoloza, F.S., Williamson, T.N., Headman, A.O., Allred, B.J., 2023. Machine-learning model to delineate sub-surface agricultural drainage from satellite imagery. J. Environ. Qual. 52, 907–921. https://doi.org/10.1002/jeq2.20493.

Ren, D., Engel, B., Mercado, J.A.V., Guo, T., Liu, Y., Huang, G., 2022. Modeling and assessing water and nutrient balances in a tile-drained agricultural watershed in the U.S. Corn Belt. Water Res. 210, 117976. https://doi.org/10.1016/j.watres.2021.117976.

Roy, S., Swetnam, T., Trochim, E., Schwehr, K., Pasquarella, V., 2023. samapriya/awesome-gee-community-datasets: Community Catalog (1.0.4). Zenodo.

Sasaki, Y., 2007. The truth of the F-measure. Teach tutor mater 1(5), 1–5.

Schilling, K.E., Helmers, M., 2008. Effects of subsurface drainage tiles on streamflow in Iowa agricultural watersheds: exploratory hydrograph analysis. Hydrol. Process. 22, 4497–4506. https://doi.org/10.1002/hyp.7052.

Schober, P., Boer, C., Schwarte, L.A., 2018. Correlation Coefficients. Anesthesia Analg. 126, 1763–1768. https://doi.org/10.1213/ane.0000000000002864.

Skaggs, R.W., Brevé, M.A., Gilliam, J.W., 1994. Hydrologic and water quality impacts of agricultural drainage*. Crit. Rev. Environ. Sci. Technol. 24, 1–32. https://doi.org/10.1080/10643389409388459.

Smith, D., King, K., Johnson, L., Francesconi, W., Richards, P., Baker, D., Sharpley, A.N., 2015. Surface runoff and tile drainage transport of phosphorus in the Midwestern United States. J. Environ. Qual. 44, 495–502. https://doi.org/10.2134/jeq2014.04.0176.

Song, H., Woo, D.K., Yan, Q., 2021. Detecting subsurface drainage pipes using a fully convolutional network with optical images. Agr Water Manage 249, 106791. https://doi.org/10.1016/j.agwat.2021.106791.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. Bmc Bioinformatics 9, 307. https://doi.org/10.1186/1471-2105-9-307.

Sugg, Z., 2007. Assessing US Farm Drainage: Can GIS Lead to Better Estimates of Subsurface Drainage Extent? World Resources Institute.

Tilahun, T., Seyoum, W.M., 2020. High-resolution mapping of tile drainage in agricultural fields using unmanned aerial system (UAS)-based radiometric thermal and optical sensors. Hydrology 8, 2. https://doi.org/10.3390/hydrology8010002.

USDA, 2013. Soil Survey Geographic (SSURGO) database - Drainage Class.

USDA, 2017. Soil Survey Manual (USDA Handbook No. 18). US Government Printing Office.

USDA, 2023. World Agricultural Production (WAP27-34), Foreign Agricultural Services Circular Series.

USDA-NASS, 2012. Agricultural Census data for county-level tile drainage area, 2012.

USDA-NASS, 2017. Agricultural Census data for county-level tile drainage area.

Valayamkunnath, P., Barlage, M., Chen, F., Gochis, D.J., Franz, K.J., 2020. Mapping of 30-meter resolution tile-drained croplands using a geospatial modeling approach. Sci Data 7, 257. https://doi.org/10.1038/s41597-020-00596-x.

Valayamkunnath, P., Gochis, D.J., Chen, F., Barlage, M., Franz, K.J., 2022. Modeling the Hydrologic Influence of Subsurface Tile Drainage Using the National Water Model. Water Resour. Res. 58. https://doi.org/10.1029/2021wr031242.

Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., Bai, X., 2015. Flood hazard risk assessment model based on random forest. J. Hydrol. 527, 1130–1141. https://doi.org/10.1016/j.jhydrol.2015.06.008.

Webber, J.J., Williamson, T.N., 2021. Workflow for Using Unmanned Aircraft Systems and Traditional Geospatial Data to Delineate Agricultural Drainage Tiles at Edge-of-Field Sites. U.S. Geological Survey Scientific Investigations Report 2021–5013.

White, M.J., Arnold, J.G., Bieger, K., Allen, P.M., Gao, J., Čerkasova, N., Gambone, M., Park, S., Bosch, D.D., Yen, H., Osorio, J.M., 2022. Development of a field scale SWAT+ modeling framework for the contiguous U.S. JAWRA J. Am. Water Resour. Assoc. 58, 1545–1560. https://doi.org/10.1111/1752-1688.13056.

Woo, D.K., Ji, J., Song, H., 2023. Subsurface drainage pipe detection using an ensemble learning approach and aerial images. Agric. Water Manag. 287, 108455. https://doi.org/10.1016/j.agwat.2023.108455.

Woo, D.K., Song, H., Kumar, P., 2019. Mapping subsurface tile drainage systems with thermal images. Agr Water Manag 287, 108455. doi.org/10.1016/j.agwat.2023.108455.

Xie, Y., Lark, T.J., 2021. Mapping annual irrigation from Landsat imagery and environmental variables across the conterminous United States. Remote Sens. Environ. 260, 112445. https://doi.org/10.1016/j.rse.2021.112445.

Yan, L., Roy, D.P., 2016. Conterminous United States crop field size quantification from multi-temporal Landsat data. Remote Sens. Environ. 172, 67–86. https://doi.org/10.1016/j.rse.2015.10.034.

Zhang, C., Walters, D., Kovacs, J.M., 2014. Applications of low altitude remote sensing in agriculture upon farmers' requests– A case study in northeastern Ontario, Canada. PLoS ONE 9, e112894. https://doi.org/10.1371/journal.pone.0112894.