

## **Exhibit S**

Prasanth Valayamkunnath et al., *Mapping of 30-Meter Resolution Tile-Drained Croplands Using Geospatial Modeling Approach*, 7 SCI. DATA 1 (2020)



OPEN

DATA DESCRIPTOR

# Mapping of 30-meter resolution tile-drained croplands using a geospatial modeling approach

Prasanth Valayamkunnath<sup>1</sup>✉, Michael Barlage<sup>1</sup>, Fei Chen<sup>1</sup>, David J. Gochis<sup>1</sup> & Kristie J. Franz<sup>2</sup>

Tile drainage is one of the dominant agricultural management practices in the United States and has greatly expanded since the late 1990s. It has proven effects on land surface water balance and quantity and quality of streamflow at the local scale. The effect of tile drainage on crop production, hydrology, and the environment on a regional scale is elusive due to lack of high-resolution, spatially-explicit tile drainage area information for the Contiguous United States (CONUS). We developed a 30-m resolution tile drainage map of the most-likely tile-drained area of the CONUS (AgTile-US) from county-level tile drainage census using a geospatial model that uses soil drainage information and topographic slope as inputs. Validation of AgTile-US with 16000 ground truth points indicated 86.03% accuracy at the CONUS-scale. Over the heavily tile-drained midwestern regions of the U.S., the accuracy ranges from 82.7% to 93.6%. These data can be used to study and model the hydrologic and water quality responses of tile drainage and to enhance streamflow forecasting in tile drainage dominant regions.

## Background & Summary

Subsurface tile drainage is one of the most widely-used agriculture management practices to enhance crop yield in regions with high water tables or poorly drained soils. Based on the U.S Department of Agriculture (USDA) National Agriculture Statistics Service (NASS) 2017 Census of Agriculture<sup>1</sup>, the spatial extent of tile-drained croplands in the United States is around 22.48 million ha, out of which 18.79 million ha (83.8%) are mainly in six Midwestern states. Tile drains, generally installed below the crop rootzone to drain excess water from the soil, improve soil aeration and soil quality in the rootzone<sup>2</sup>, and enhance crop yield<sup>2,3</sup>. Compared to undrained cropland, tile drainage also causes significant changes in watershed hydrology<sup>4,5</sup>. Studies have shown subsurface drainage removes excess water from the rootzone<sup>6</sup>, results in higher infiltration and lower surface runoff<sup>7</sup>, peak flows<sup>8</sup> and flooding<sup>9</sup>. Tile drainage tends to increase the watershed baseflow<sup>10,11</sup>, decrease groundwater travel times<sup>10,12</sup>, increase annual runoff volume<sup>4,13–16</sup>, and increase instream nitrate concentrations<sup>4,8,17,18</sup>. Subsurface tile drainage can influence the runoff volume, the timing and shape of the hydrograph<sup>4,5</sup>, and the local and regional climate by reducing ET from croplands<sup>19,20</sup>.

Correctly modeling the tile-drainage impacts on the hydrologic cycle is a daunting challenge due to the lack of continental-scale high-resolution tile-drainage data<sup>21</sup>, because farm-, field-, and subcounty-scale information on tile drainage are not available. USDA-NASS has well documented the county-level tile-drained area for the U.S., however, the geographical information on the spatial distribution of tile-drained croplands within each county are not provided. As subsurface tile drainage has significant effects on land surface hydrology<sup>4,5</sup>, water quality<sup>22</sup>, local and regional climate<sup>19,20</sup>, accurate, detailed geographical information on the sub-county spatial distribution of tile drained croplands is indispensable to answer a wide variety of questions related to hydrology, climate, and earth system science<sup>21</sup>.

During the past two decades, several spatial and gridded data of tile drainage have been developed for the U.S. For instance, Jaynes and James<sup>23</sup> developed gridded (30-m resolution) tile drainage fraction data for the 23 most intensively drained states in the U.S. with a Geographic Information System (GIS) approach using the 1992 National Land Cover Database (NLCD) croplands and Natural Resources Conservation Service (NRCS) State Soil Geographic (STATSGO) poorly drained soil dataset. By applying a similar approach, Sugg<sup>21</sup> developed tile-drainage data for the Midwestern corn belt of the U.S. using 1992 NLCD row crops information and STATSGO drainage class and soil hydrologic soil groups; however, the result is likely to be an overestimate compared to census data because

<sup>1</sup>National Center for Atmospheric Research (NCAR), Boulder, Colorado, 80301, USA. <sup>2</sup>Geological and Atmospheric Sciences, Iowa State University, Ames, Iowa, 50011, USA. ✉e-mail: [prasanth@ucar.edu](mailto:prasanth@ucar.edu)

their method classified all the row crops with poorly drained soils as tile trained. Using a GIS-based analysis, Sui<sup>24</sup> generated a tile drainage map for Indiana considering land cover, soil, and slope data where the slope is less than 2%, and soils are poorly drained. By employing aerial imagery, Naz *et al.*<sup>25</sup> identified tile drainage locations in Hoagland watershed of Indiana. Moreover, Brown<sup>26</sup> mapped tile drainage locations in three counties of Minnesota with GIS by using croplands from NLCD - 2001 and NASS - 2008, soil drainage class, hydrologic group, and land capability class from SSURGO data, and topographic slope (less than 2%) from National Elevation Data (NED). Nakagaki *et al.*<sup>27</sup> developed 30-m gridded tile-drainage fraction data for the conterminous U.S. for the year 1992 using NLCD 2011 and STATSGO version 2.0 data using the Sugg<sup>21</sup> methodology. These data are updated using the 2012 USDA-NASS census but only for 12 Midwestern states<sup>28</sup>. Recently, Cho *et al.*<sup>29</sup>, identified tile-drained croplands of the Red River basin covering portions of North Dakota, South Dakota, and Minnesota using satellite big-data and random forest machine learning approaches. Their data-intensive machine learning approach indicated that soil properties and land surface temperatures are the strongest predictors of tile drainage. However, this approach is data-, time-, and computationally-intensive, and constrained to the quantity, quality, and resolution of satellite data or predictors. Also, the quality of the tile drainage map generated using this approach relies on the training samples or the quantity and quality of field-level ground-truth information (or tile drainage permit records). Thus, extending this tile drainage mapping approach to other regions or the CONUS scale is a cumbersome task.

Remote sensing-based (e.g. microwave brightness signatures) tile-drainage mapping approaches primarily depend upon the soil reflectance. Soil above the tile drainage tends to dry faster (higher reflectance) compared to the soil midway (low reflectance) between two tile drainage lines. Previous studies have indicated that the ideal time to acquire aerial imagery for tile drainage mapping is within three days after a 25 mm or greater rainfall event<sup>30,31</sup>. That is, the degree of accuracy of a remote sensing approach to identify tile drainage over a large area mainly depends upon the time of acquiring aerial imagery after a rainfall event and presence of vegetation cover<sup>32</sup>. Despite the high performance of remote sensing approaches to delineate tile drainage over a field or watershed scale, mapping tile drainage for regional or national level is expensive and constrained by weather and availability of resources.

To date, comprehensive well-validated, gridded, high-resolution tile drainage data are unavailable for the U.S. Here, we develop a present-day, 30-m resolution tile drainage map for the CONUS using USDA Census of Agriculture 2017 county-level tile drainage information, NLCD-2016 cropland, SRTM-DEM derived slope, SSURGO version 2.0 soil drainage information and by employing a GIS model. We present verification of the resulting tile drainage map with 16000 field-level ground-truth points.

## Methods

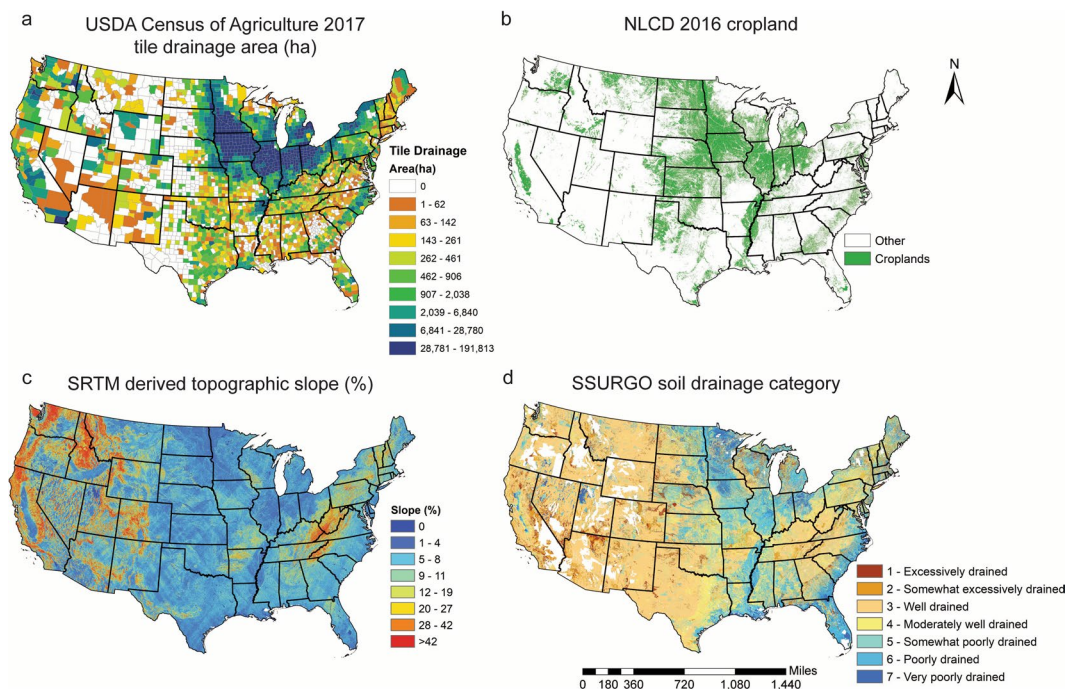
This section first discusses the list and descriptions of data required for mapping tile drainage croplands of the U.S. Next, it describes details of a strong yet simple and easy to implement GIS decision tree model to delineate tile drainage areas, hereafter referred as SSURGO SRTM-slope defined Agriculture Tile Drainage Dataset for the U.S. (AgTile-US).

**County tile drainage statistics.** The most critical input into the AgTile-US geospatial model is the county-level tile-drainage census estimates from USDA, which is conducted once every five years. Here, we use the most recent census data (Census of Agriculture, 2017)<sup>1</sup>. It is the best tile drainage area estimate available for the entire U.S. Using the USDA-NASS Quick Stats tool<sup>33</sup>, the tile drainage area information and error estimate were downloaded in spreadsheet format. It was then joined to the U.S. county shapefile in ESRI ArcMap 10.7<sup>34</sup> before feeding into the AgTile-US geospatial model. The spatial distribution of county-level tile drainage is presented in Fig. 1a. The north-central U.S. shows the greatest areal extent of tile drainage, especially in Iowa (5.7 Mha), Illinois (3.83 Mha), Minnesota (3.27 Mha), Indiana (2.57 Mha), Ohio (2.18 Mha), and Michigan (1.22 Mha). These states are in the headwater regions of the Mississippi River and contribute a significant amount of runoff, sediment, and nutrients to the Mississippi River. Parts of the north-eastern U.S., coastal areas of south-eastern U.S., and the south-central U.S. also show a localized high-intensity tile drainage.

**National land cover database (NLCD) 2016.** To constrain the mapping of tile drained areas to cropland, we ingested a cropland mask (Fig. 1b) derived from NLCD 2016<sup>35</sup> into the geospatial model. Since tile drainage mapping by the geospatial model is mainly based on the county-level census estimate, soil drainage categories, and topographic slope, a cropland mask is necessary to mask out non-agricultural lands. The “cultivated cropland” category from NLCD 2016 was used to define the cropland mask, and cells or grids that were not cropland were masked out of the drainage category raster and slope raster by the geospatial model before mapping the tile drainage areas.

**Slope generated from Shuttle Radar Topography Mission (SRTM) derived DEM.** The SRTM<sup>36</sup> is an international research effort intended to generate a high-resolution (30-m to 90-m) digital topographic database on the near-global scale (56°S to 60°N). This project was a combined effort of the National Aeronautics and Space Administration (NASA) and the National Geospatial-Intelligence Agency (NGA). The data were collected by interferometric radar, which compares two radar signals captured at slightly different angles by two different antennas. Surface elevation was estimated using the differences between the two radar signals. In this study, we used 30-m resolution, void-filled DEM data from SRTM to estimate topographic slope for the CONUS. The data were downloaded as GeoTIFF format from the Opentopography website (<https://opentopography.org/>; accessed in October 2019). The topographic slope (Fig. 1c) was estimated using ESRI ArcMap10.7 at 30-m resolution.

**Soil Survey Geographic (SSURGO) data from USDA NRCS.** The SSURGO database (<https://websoilsurvey.nrcs.usda.gov/>; accessed on 15 October 2019) contains soil information collected by the National



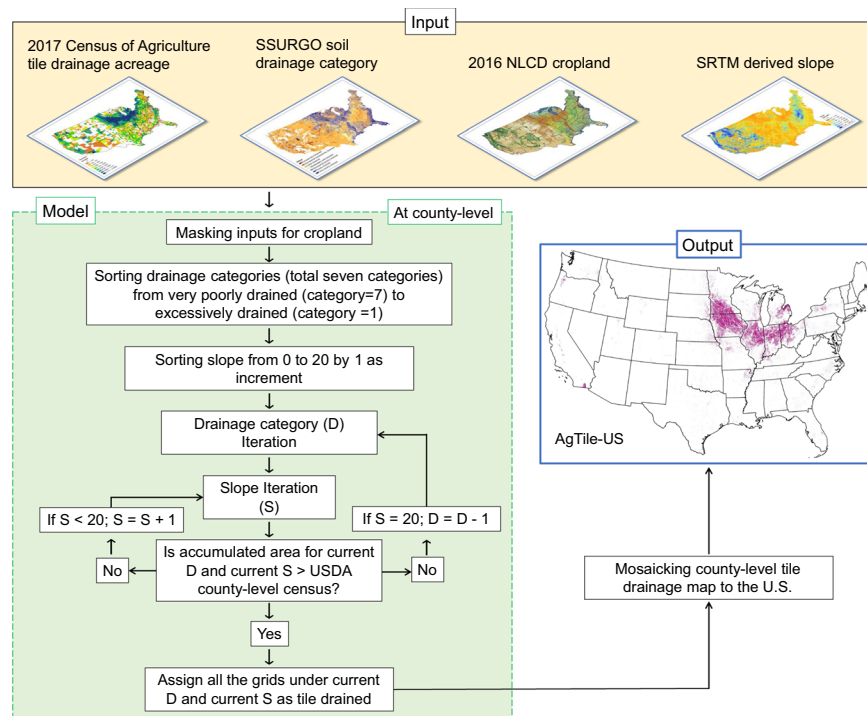
**Fig. 1** Input data to AgTile-US GIS model for generating gridded tile drainage data for the U.S. **(a)** County-level tile drainage area (ha) from USDA Census of Agriculture 2017, **(b)** National Land Cover Database (NLCD) 2016 cropland mask at 30 m resolution, **(c)** Shuttle Radar Topography Mission Digital Elevation Model derived slope (%) at 30 m, and **(d)** Spatial pattern of soil drainage characteristics based on the SSURGO database at 30 m.

Cooperative Soil Survey of the U.S. The soil information was obtained by field visits and laboratory analysis. The collected soil information is available at scales ranging from 1:12,000 to 1:63,360 in database tables and maps. The SSURGO database contains various information including available water capacity, soil reaction, electrical conductivity, drainage characteristics, soil texture, and frequency of flooding; yields for cropland, woodland, rangeland, and pastureland; and limitations affecting recreational development, building site development, and other engineering uses. Map units are the map outline areas, and within a map unit, soil and other components have unique properties, interpretations, and productivity. Maps can be linked to database tables using the unique map unit number to extract information about component soils and their properties. Information in SSURGO can be viewed in the Web Soil Survey (WSS) or obtained in ESRI shapefile format for the study area.

In this study, we used 10-m resolution, gridded SSURGO (gSSURGO) data obtained for the CONUS from the USDA Geospatial Data Gateway ([https://datagateway.nrcs.usda.gov/GDGHome\\_DirectDownload.aspx](https://datagateway.nrcs.usda.gov/GDGHome_DirectDownload.aspx); accessed on 15 October 2019). The gSSURGO database consists of a 10-m resolution map unit raster in GeoTIFF format (MapunitRaster\_10m.tif) along with a component database table. Soil drainage classes are prepared by joining the map unit raster to its component database table in ESRI ArcMap using the map unit key (MUKEY) and the resulting raster consisting of seven drainage classes was then resampled to a 30-m resolution raster. This raster helped the AgTile-US GIS model to identify tile drainage by distinguishing soil with different drainage categories. The spatial pattern of soil drainage information is depicted in Fig. 1d. Most of the poorly drained soil are found in the north-central U.S., parts of the south-central U.S., especially surrounding the Mississippi River, and coastal states of the south-eastern U.S.

**The AgTile-US geospatial tile drainage model.** A geospatial model was developed to map tile drainage areas based on soil drainage information and a topographic slope threshold within agricultural land cover that constitute an equivalent tile drainage area by county estimated by the Census of Agriculture 2017 (Fig. 2). The geospatial framework was based on the idea that subsurface tile-drainage systems are needed for flat (or low slope) croplands with poorly drained soil. The model was implemented at county-level and initiated by generating an ordered list of drainage category (D) (see legends of Fig. 1d for seven drainage categories and corresponding numbers) and slope (S). To validate the AgTile-US and to set the upper limit of slope in the geospatial model, we have manually identified 16000 ground-truth points of tile drainage across the U.S. using ESRI multiresolution aerial imagery. Slope values are extracted to these point locations from the SRTM slope raster. Using the extracted slope values, we have identified the upper limit of slope ( $=20\%$ ) in the geospatial model. The details are presented in the results section.

In the first iteration, the geospatial model chose grids with the highest drainage category ( $D=7$ ) (which corresponds to very poorly drained soil) and zero slope ( $S=0$ ) and accumulated area covered by those grids was estimated. The accumulated area was then compared with the target tile drainage area from the USDA census. If the accumulated area of selected grids was less than the census, the geospatial model chose grids based on the highest



**Fig. 2** GIS Workflow diagram of the SSURGO SRTM-slope defined Agriculture Tile Drainage Dataset (AgTile-US) methodology. The output map has a spatial resolution of 30 m.

drainage category ( $D = 7$ ) and slope less than or equal to the next lowest slope value ( $S + 1$ ). That is, an increment to the slope was applied by 1%. If the geospatial model reached the upper limit of slope (i.e.,  $S = 20$ ) and the accumulated area of selected grids was less than the census, then grids with drainage category greater than or equal to the next highest drainage category ( $D = 6$ ) were selected. In that case, the model set the slope back to zero ( $S = 0$ ). An increment to the slope was made if the accumulated area was less than the target tile drainage area from the census. The model repeated these steps until the estimated tile drainage area exceeded the USDA census tile drainage area. When the estimated tile drainage area exceeded the USDA census, the model reinitialized with the current drainage category and previous slope. With a slope increment of 0.1, the model estimated absolute error between the estimated tile drainage area and the USDA census to the current slope. A final county-level tile drainage map was generated by the model with current drainage category, and the slope corresponds to the least error (Supplementary Fig. S1). When the model finished creating tile drainage maps for all counties, county maps were mosaicked together to generate a seamless 30-m U.S. tile drainage map or AgTile-US using ESRI ArcMap.

The generated AgTile-US map has a spatial resolution of 30-m, which is equivalent to the NLCD land cover. Based on the USDA farms and land in farms summary (2018), the average farm size is 443 acres. Thus, the AgTile-US tile drainage map provides adequate accuracy to identify field-level tile drainage over the U.S. However, the model presented here highly depends upon tile drainage area estimates from the USDA census and SSURGO soil data. Any uncertainty in these datasets will lead to inaccuracies in the model estimated tile drainage map.

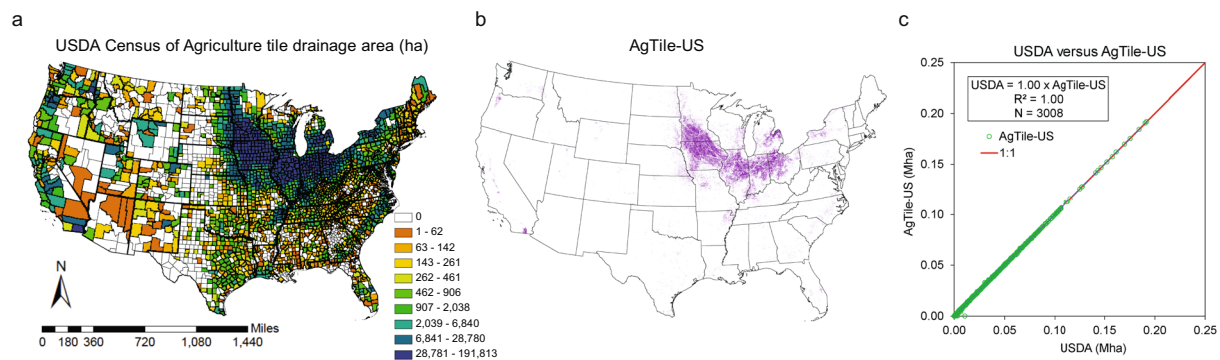
### Data Records

The generated tile drainage spatial data (AgTile-US<sup>37</sup>) extends from  $65^{\circ} 20' 46.49''$  W to  $127^{\circ} 53' 13.65''$  W and  $22^{\circ} 51' 41.18''$  N to  $51^{\circ} 36' 16.74''$  N, which covers the entire CONUS. The tile drainage data and its quality details are freely available to the public in GeoTIFF format and NetCDF format through an unrestricted, public repository (Figshare<sup>37</sup>). The data are available at 30-m resolution in binary format, where 0 represents undrained agricultural land and 1 represents tile-drained agricultural land. In the AgTile-US data, each 30-m grid identified as tile drained is entirely tile drained. The data deliver a static representation of U.S. tile drainage for the year 2017. The data quality information is an estimate of the accuracy of the AgTile-US based on 16000 ground-truth point and is presented in next section. Upon the availability of new data, the repository will be updated with a newer version of the tile drainage map.

### Technical Validation

The spatial distribution of the tile drained area in the AgTile-US data is consistent with USDA 2017 county-level census because the geospatial model was constrained by county-level census estimates (Fig. 3). Based on the AgTile-US spatial map, most of the major tile-drained croplands across the CONUS are located in the north-central U.S., especially in Iowa, Illinois, Minnesota, Indiana, Ohio, and Michigan. Regions, including southern California, eastern Wisconsin, north-western New York, and parts of Arkansas, Missouri, and Oregon, are also identified as tile drained croplands. Even though the USDA county-level estimate was used to constrain the



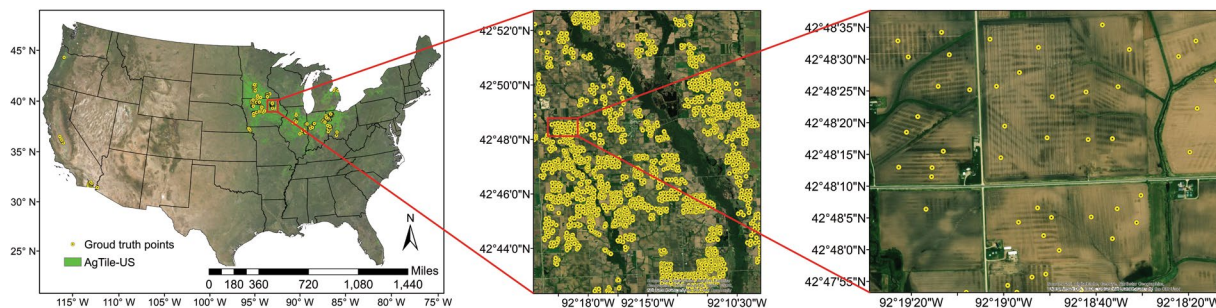


**Fig. 3** Comparison of USDA Census of Agriculture tile drainage area and AgTile-US estimates. **(a)** USDA county-level tile drainage area (ha), **(b)** the AgTile-US tile drainage data at 30 m, In **(b)**, magenta color indicates grids with tile drainage, **(c)** scatterplot of USDA census and AgTile-US tile drainage area (Mha).

State	Tile drained area (ha)		Difference (%)	State	Tile drained area (ha)		Difference (%)
	USDA	AgTile-US			USDA	AgTile-US	
Iowa	5,708,945	5,706,831	-0.04	Oklahoma	43,031	36,983	-14.06
Illinois	3,833,811	3,836,371	0.07	Virginia	42,571	40,408	-5.08
Minnesota	3,266,234	3,264,744	-0.05	Georgia	40,196	38,747	-3.6
Indiana	2,572,515	2,572,162	-0.01	Wyoming	35,090	35,657	1.62
Ohio	2,183,253	2,178,405	-0.22	Colorado	31,355	30,418	-2.99
Michigan	1,225,734	1,225,235	-0.04	Utah	31,175	29,803	-4.4
Missouri	453,346	450,114	-0.71	Idaho	29,592	29,930	1.14
Wisconsin	397,952	396,948	-0.25	Louisiana	26,518	26,245	-1.03
New York	348,537	346,233	-0.66	Alabama	24,840	24,759	-0.33
South Dakota	264,949	265,357	0.15	Maryland	18,370	16,973	-7.6
California	253,260	254,011	0.3	West Virginia	16,487	9,500	-42.38
Nebraska	223,968	224,583	0.27	New Mexico	13,943	2,919	-79.06
Kentucky	184,419	176,125	-4.5	Vermont	13,096	13,104	0.06
Arkansas	171,879	168,358	-2.05	New Jersey	10,998	10,968	-0.28
Oregon	127,125	124,760	-1.86	Mississippi	10,025	8,431	-15.9
Pennsylvania	125,337	123,859	-1.18	Maine	5,435	5,359	-1.41
Kansas	118,707	118,433	-0.23	Montana	4,919	4,887	-0.66
North Dakota	112,531	116,146	3.21	Delaware	4,619	4,610	-0.19
Texas	110,788	98,018	-11.53	Nevada	3,502	3,495	-0.2
North Carolina	110,363	107,704	-2.41	Massachusetts	1,461	1,442	-1.31
Washington	76,758	73,498	-4.25	New Hampshire	1,119	1,123	0.34
Florida	51,186	51,120	-0.13	Arizona	1,104	1,224	10.85
Tennessee	43,558	40,574	-6.85	Connecticut	929	946	1.86
South Carolina	43,434	42,926	-1.17				
Total U.S.					22,418,966	22,340,442	-0.35

**Table 1.** State wise tile-drained area estimates from USDA and AgTile-US for the year 2017.

AgTile-US tile drainage mapping by the geospatial model, the actual tile drained area identified in the AgTile-US data is  $-0.35\%$  lower for the CONUS (Table 1). The AgTile-US shows higher underestimation of tile drainage over the croplands of New Mexico ( $-79.06\%$ ), West Virginia ( $-42.38\%$ ), Mississippi ( $-15.90\%$ ), Oklahoma ( $-14.06\%$ ), and Texas ( $-11.53\%$ ) (Table 1). However, the total tile drained area of these states is less than 0.11 Mha. The underestimation of tile drainage by the AgTile-US is mainly due to underrepresentation of croplands in NLCD land cover 2016. Also, the USDA census reported tile drainage area for non-croplands. For instance, the tile drainage estimate from the USDA 2017 census for the Guadalupe county of New Mexico is 10602.8 ha, whereas the total cropland area based on the census is only 5386.8 ha. For the same county, the cropland area based on NLCD 2016 is 796.5 ha. A comparison of cropland from USDA county-level census 2017 and NLCD 2016 indicate an underrepresentation of croplands by  $-92.00\%$ ,  $-27.45\%$ , and  $-18.15\%$ , for West Virginia, New Mexico, and the CONUS, respectively. Since the geospatial model identified tile drained area within NLCD cropland, the underestimation of tile drainage area by the AgTile-US can be attributed to the underrepresentation of croplands in the NLCD 2016. However, for the heavily tile drained north-central U.S., the percentage difference in tile drained area between the model estimate and USDA census is less than  $-0.25\%$  (Table 1).



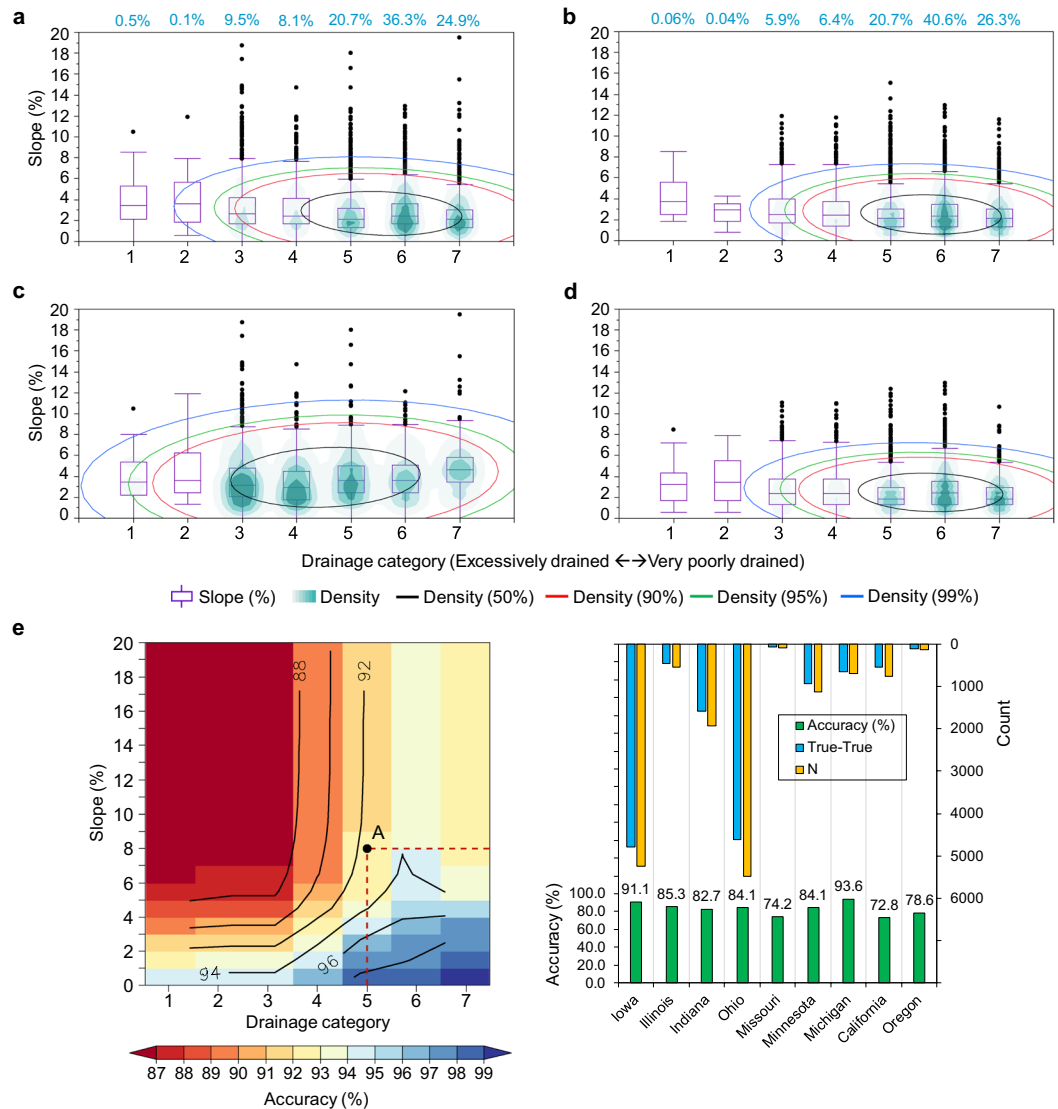
**Fig. 4** Spatial distribution of tile drainage ground-truth data (16000 points) used for the accuracy assessment. These point data were manually identified from ESRI multi-resolution (15 m, 1 m, 0.6 m, and 0.3 m) aerial imagery basemap (sources include Esri, DigitalGlobe, Earthstar Geographics, CNES/Airbus DS, GeoEye, USDA FSA, USGS, Aerogrid, IGN, IGP, and the GIS User Community; Acquisition dates of imagery ranges from 10/2/2012 to 4/23/2017 which is identified using ESRI ArcMap identification tool from the imagery basemap).

Generally, the usefulness of scientific data mainly depends upon the level of accuracy and the quality of validation data. Most of the previous studies identified tile drained croplands at a smaller scale (field, catchment, or watershed scale) than the present study and employed GIS-based, remote-sensing-based or data-intensive methods like machine learning approaches<sup>25,29,32</sup> to achieve an overall accuracy of 79% to 86%. The availability and the spatial extent of validation data is the major constraint to these studies. Thus, the present study manually identified 16000 tile drainage ground-truth points across ten states of the CONUS from the ESRI multi-resolution (15 m, 1 m, 0.6 m, and 0.3 m) aerial imagery basemap (Fig. 4). The number of tile drainage ground-truth points identified depends upon the presence of a clear tile drainage signature at field-level. As indicated by *Tetzlaff et al.*<sup>32</sup>, the presence of vegetation and the acquisition time of aerial imagery after a rainfall event make this task difficult. So, we limited the tile drainage ground-truth identification mostly to fallow croplands where the tile drainage signature is easily detectable. The points are randomly placed in a field when a tile drainage signature is detected (Fig. 4). Simple random sampling method is used as it is the best method of sampling and each member of the population has an equal probability of being selected<sup>38</sup>. Slope and soil drainage categories were extracted for these 16000 points, and based on that we constrained the upper limit of slope in the model before mapping tile drainage.

To validate the AgTile-US tile drainage data with ground truth points, we plotted ground-truth points on drainage category-topographic slope two-dimensional distribution space (Fig. 5a–d) and its statistics are presented in Table 2. Figure 5a–d and Table 2 provides the information on the distribution of slope for each drainage category. For instance, the distribution of 16000 tile drainage ground-truth points on drainage category-topographic slope two-dimensional distribution space is presented in Fig. 5a, and it is clear from the figure that most (36.3%) of the tile drained croplands have poorly drained soil (category 6). Moreover, the data presented in Fig. 5a indicates a very small presence (0.5%) of tile drainage in excessively drained croplands (category = 1) and croplands with up to 10% slope. Over 50 percent of the ground-truth points are located over drainage categories greater than or equal to four and within 4.0% slope, and a slope of 8.76% or less covers 99 percent of the ground-truth points. Also, for the poorly drained category, 25% of the tile drained croplands are within 1.67% slope and a slope within 3.58% covers 75% of tile drained croplands (Table 2).

To compare the AgTile-US with ground-truth points, we used 21582 random samples from AgTile-US (Fig. S2). We plotted values of slope and soil drainage category identified for 21582 random samples (used simple random sampling technique in ESRI ArcMap) on drainage category-topographic slope two-dimensional distribution space (Fig. 5b). The geospatial model captured the tile drainage ground-truth distribution characteristics (compare Fig. 5a,b, and Table 2). Identical to ground-truth points, most (40.60%) of the grids in the AgTile-US that identified as tile drained are in poorly drained croplands, and about 99 percent of the tile drained grids are within 7.77% slope. For the poorly drained category, 25% of the tile drained croplands are within 1.67% slope and a slope of 3.44% covers 75% of tile drained croplands (Table 2). The density ellipses in Fig. 5b are slightly contracted along the major and minor axis. This is because in the geospatial model, we set higher priority for poorly drained soil with low slopes. We have also extracted the AgTile-US grid values to the ground-truth points and separated the points into True-True (drained in AgTile-US and in ground-truth) and False-True (undrained in AgTile-US and drained in ground-truth), and plotted each distribution on drainage category-topographic slope two-dimensional distribution space (Fig. 5c,d). The AgTile-US failed to identify tile drainage at about 13.97% of ground-truth locations, and most of these points are in marginal drainage categories (category = 3–5). However, the True-True points (Fig. 5d), which sum up to 86.03% of total points, show a strong agreement to the ground-truth distribution (Fig. 5a).

To examine the influence of soil drainage category and topographic slope on the AgTile-US accuracy, we extended our analysis using 16000 ground-truth points. The accuracy distribution of AgTile-US on drainage category-topographic slope two-dimensional space is depicted in Fig. 5e. The AgTile-US tile drainage data indicated higher accuracy (greater than 98%) over croplands with very poorly drained and poorly drained soils and slope less than or equal to one. The accuracy of AgTile-US consistently decreases as we include grids with higher slope and increasing soil drainage. For instance, point ‘A’ in Fig. 5e indicates an accuracy of 92 to 93% (area



**Fig. 5** Verification of AgTile-US tile drainage data with 16000 ground truth points. **(a)** Distribution of 16000 tile drainage ground-truth points on drainage category-topographic slope two-dimensional distribution space, **(b)** As in **(a)**, but for AgTile-US tile drainage data, **(c)** As in **(b)**, but for ‘false’ AgTile-US grids when compared to ground truths (False-True), **(d)** As in **(b)**, but for valid or ‘true’ AgTile-US (True-True). In **(a–d)**, turquoise color grading over the box and whisker plot and number in turquoise color over the panel **(a,b)** indicate the data density in each drainage category and colored ellipse indicate the data distribution density (%), **(e)** AgTile-US accuracy distribution on drainage category-topographic slope two-dimensional distribution space based on ground truth points. In **(e)**, black line indicates isoline of accuracy or accuracy contour, **(f)** State-wise summary of accuracy assessment for selected nine states. In **(f)**, N indicates the total number of ground truth identified. In **(a–e)** x-axis indicate soil drainage categories, 1-Excessively drained, 2-Somewhat excessively drained, 3-Well drained, 4-Moderately well-drained, 5-Somewhat poorly drained, 6-Poorly drained, and 7-Very poorly drained.

under red dashed lines). It includes grids with drainage category from seven to five and slope from 0 to 8%. Our geospatial approach to identify tile drainage has higher accuracy (greater than 90%) over croplands with any poorly drained soils (category  $\geq 5$ ) and slope up to 20%, or croplands with any well-drained soils (category  $< 5$ ) and less than 2% slope. Also, the accuracy distribution presented in the Fig. 5e can be used to extract AgTile-US grids of required accuracy. The state-level summary of AgTile-US validation with ground-truth points is presented in Fig. 5f. The AgTile-US indicated higher accuracy (72.8% to 100%) over most of the heavily tile drained states of the north-central U.S. The state with the most tile drained croplands is Iowa (5.7 Mha), where we have identified 5253 ground truth points, and the AgTile-US showed 91.1% accuracy. To explore this under-performance, a comparison of the ground-truth points and NLCD 2016 cropland showed that 91 out of 16000 (0.57%) points are not in croplands. A comparison of NLCD 2016 cropland area with county-level cropland area estimate from the Census of Agriculture 2017<sup>1</sup> indicated a good agreement with  $R^2 = 0.93$  and RMSE 1.82 Mha. Overall, NLCD 2016 has 18% less cropland when compared to the USDA census estimate. This indicates an added uncertainty due to using NLCD 2016 to define the crop mask. To address this uncertainty we repeated the



Drainage Category	25th Quantile slope (%)		75th Quantile slope (%)		Median slope (%)		Data Density (%)	
	GT	AgTile-US	GT	AgTile-US	GT	AgTile-US	GT	AgTile-US
1	2.12	2.63	5.17	5.07	3.44	3.72	0.5	0.06
2	1.86	1.86	5.43	3.53	3.58	2.96	0.1	0.04
3	1.67	1.67	4.17	3.95	2.64	2.46	9.5	5.9
4	1.67	1.58	4.12	3.77	2.42	2.42	8.1	6.4
5	1.32	1.32	3.17	3	2.12	2.12	20.7	20.7
6	1.67	1.32	3.58	3.44	2.42	2.35	36.3	40.6
7	1.32	1.32	3	3	2.12	2.12	24.9	26.3

**Table 2.** Validation statistics for the AgTile-US. In Table 2, GT represents statistics for ground-truth points based on data presented in Fig. 5a and AgTile-US represents statistics for the AgTile-US based on data presented in Fig. 5b.

AgTile-US methodology with the USDA CropScape crop data layer<sup>39</sup>. However, the overall accuracy of mapped tile drainage area using the USDA CropScape was the same as with the NLCD crop layer. Additionally, we performed a sensitivity analysis to evaluate the effect of DEM uncertainty on AgTile-US accuracy. Analysis indicated nearly similar accuracy (only ~2% lower) of AgTile-US with new DEM (a 30-m resolution DEM resampled from 90-m SRTM-DEM). The SRTM derived DEM used in this study has good accuracy (with a CONUS scale mean error of 0.16 m) to produce quality topographic slope<sup>40,41</sup>.

Furthermore, we constrained our geospatial model with the USDA 2017 county-level tile drainage estimate. The most significant sources of uncertainty in our tile drainage area estimate can be attributed to uncertainty in USDA tile drainage census estimates and NLCD 2016 cropland. According to the Census of Agriculture, 2017 final report<sup>1</sup>, there is the potential for error within the census estimate due to the adjustments for nonresponse, undercoverage, misclassification, calibration, and integerization. The average error in tile drainage estimates based on the Census of Agriculture 2017 is 19.8%. However, we consider the USDA tile drainage census estimate to be the best reliable tile drainage data since it comes from a federally published national agriculture database. Unavailability of tile drainage ground truth points over vegetated croplands and uncertainty in croplands without tile drainage constitutes the primary limitations of the validation approach we adopted.

The AgTile-US agreed well with ground-truth points and indicated an overall accuracy of 86.03% for the CONUS with an omission error of 13.97%. The amount of ground-truth points used in this study were sufficient to conclude the accuracy of AgTile-US to 86.03% (Fig. S3). Despite some remaining uncertainty, a 30-m resolution tile drainage data generated using the geospatial model and input datasets outlined in this paper are of sufficient quality to be used to answer a wide variety of questions related to tile drainage and its impacts on hydrology and environment, climate, and agriculture in the U.S.

### Usage Notes

This paper presented a robust but easy-to-implement decision-tree-based geospatial model for mapping tile drainage for the U.S. using SSURGO soil drainage information and topographic slope as inputs. The tile drainage dataset was prepared by constraining the geospatial model with USDA county-level tile drainage estimates, and thresholds of slope and soil drainage properties were estimated separately for each county. The dataset was successfully analyzed and validated with 16000 ground-truth points across the CONUS. Overall, the dataset indicated 86.03% accuracy for the CONUS, which was comparable with the accuracy of tile drainage mapping at the watershed-scale reported by other studies.

In general, the spatial information of tile drainage is necessary to study the influence of subsurface drainage on plant growth, soil quality, in-stream nutrient load, and streamflow; for applications such as hydrological modeling, flood forecasting, water balance analysis, climate modeling, spatial and temporal changes in crop yield and cropping pattern, and spatial and temporal changes in evapotranspiration. Even though subsurface tile drainage in the Midwestern U.S. causes severe problems of nutrient pollution and hypoxia in the surface water, especially in the Gulf of Mexico (the “dead zone”), those tile-drainage data limitation constrains researchers to scale-up impact assessment studies from watershed scale to regional scale<sup>21</sup>. For instance, the tile drainage spatial map is supporting an effort to improve the prediction of streamflow over the Midwestern U.S. within the NOAA National Water Model of the U.S. Other potential uses enabled by this data set include the estimation of the contribution of subsurface tile discharge to streamflow, and the modeling of agriculture nutrient transport to the Mississippi River. This dataset can also be used to estimate the effect of tile drainage on regional climate with the help of a regional climate model.

The AgTile-US tile drainage maps are presented in georeferenced GeoTIFF, and NetCDF-4 format files along with its quality levels prepared using accuracy information presented in Fig. 5e. The user has the freedom to choose the grids of required accuracy from the AgTile-US based on the quality information provided with the dataset. This dataset can be easily imported to ESRI ArcMap<sup>34</sup> and any other geospatial software.

### Code availability

The AgTile-US model R code and example R code to read dataset are provided in GitHub (<https://github.com/NCAR/AgTile-US>) and Figshare<sup>37</sup>.

Received: 12 February 2020; Accepted: 13 July 2020;

Published online: 05 August 2020

## References

1. USDA NASS. *2017 Census of Agriculture* (2017).
2. Fausey, N. R. Drainage management for humid regions. *Int. Agric. Eng. J.* **14**, 209–214 (2005).
3. Du, B., Arnold, J. G., Saleh, A. & Jaynes, D. B. Development and application of SWAT to landscapes with tiles and potholes. *Trans. ASAE* **48**, 1121–1133 (2005).
4. Blann, K. L., Anderson, J. L., Sands, G. R. & Vondracek, B. Effects of agricultural drainage on aquatic ecosystems: a review. *Crit. Rev. Environ. Sci. Technol.* **39**, 909–1001 (2009).
5. King, K. W., Fausey, N. R. & Williams, M. R. Effect of subsurface drainage on streamflow in an agricultural headwater watershed. *J. Hydro.* **519**, 438–445 (2014).
6. Darzi-Naftchali, A., Mirlatifi, S. M., Shahnazari, A., Ejlali, F. & Mahdian, M. H. Effect of subsurface drainage on water balance and water table in poorly drained paddy fields. *Agric. Water Manage.* **130**, 61–68 (2013).
7. Robinson, M. & Rycroft, D. W. The impact of drainage on streamflow. *Agric. Drain.* **38**, 767–800 (1999).
8. Skaggs, R. W., Breve, M. A. & Gilliam, J. W. Hydrologic and water quality impacts of agricultural drainage. *Crit. Rev. Environ. Sci. Technol.* **24**, 1–32 (1994).
9. Henine, H. *et al.* Effect of pipe pressurization on the discharge of a tile drainage system. *Vadose Zone J.* **9**, 36–42 (2010).
10. Schilling, K. E., Jindal, P., Basu, N. B. & Helmers, M. J. Impact of artificial subsurface drainage on groundwater travel times and baseflow discharge in an agricultural watershed, Iowa (USA). *Hydrol. Processes* **26**, 3092–3100 (2012).
11. Schilling, K. E. & Libra, R. D. Increased baseflow in Iowa over the second half of the 20th Century. *J. Am. Water Resour. Assoc.* **39**, 851–860 (2003).
12. Schilling, K. E. & Helmers, M. Effects of subsurface drainage tiles on streamflow in Iowa agricultural watersheds: Exploratory hydrograph analysis. *Hydrol. Processes* **22**, 4497–4506 (2008).
13. Haitjema, H. M. On the residence time distribution in idealized groundwatersheds. *J. Hydrol.* **172**, 127–146 (1995).
14. Magner, J. A., Payne, G. A. & Steffen, L. J. Drainage effects on stream nitrate-N and hydrology in south-central Minnesota (USA). *Environ. Model. Assess.* **91**, 183–198 (2004).
15. Eastman, M., Gollamudi, A., Stämpfli, N., Madramootoo, C. A. & Sarangi, A. Comparative evaluation of phosphorus losses from subsurface and naturally drained agricultural fields in the Pike River watershed of Quebec, Canada. *Agric. Water Manage.* **97**, 596–604 (2010).
16. Schottler, S. P. *et al.* Twentieth century agricultural drainage creates more erosive rivers. *Hydrol. Processes* **28**, 1951–1961 (2014).
17. Rozemeijer, J. C. *et al.* Integrated modeling of groundwater–surface water interactions in a tile-drained agricultural field: The importance of directly measured flow route contributions. *Water Resour. Res.* **46** (2010).
18. Van den Eertwegh, G. A. P. H., Nieber, J. L., De Louw, P. G. B., Van Hardeveld, H. A. & Bakkum, R. Impacts of drainage activities for clay soils on hydrology and solute loads to surface water. *Irrig. Drain.* **55**, 235–245 (2006).
19. Khand, K., Kjaersgaard, J., Hay, C. & Jia, X. Estimating impacts of agricultural subsurface drainage on evapotranspiration using the Landsat imagery-based METRIC model. *Hydrology* **4**, 49 (2017).
20. Yang, Y. *et al.* Impact of tile drainage on evapotranspiration in South Dakota, USA, based on high spatiotemporal resolution evapotranspiration time series from a multisatellite data fusion system. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **10**, 2550–2564 (2017).
21. Sugg, Z. Assessing US farm drainage: Can GIS lead to better estimates of subsurface drainage extent. *World Resour. Institute, Washington, DC* **20002** (2007).
22. Tomer, M. D., Meek, D. W., Jaynes, D. B. & Hatfield, J. L. Evaluation of nitrate nitrogen fluxes from a tile-drained watershed in central Iowa. *J. Environ. Qual.* **32**, 642–653 (2003).
23. Jaynes, D. B. & James, D. E. The extent of farm drainage in the United States. *US Dep. Agric.* (2007).
24. Sui, Y. Potential impact of controlled drainage in Indiana watersheds (Doctoral dissertation, Purdue University). Preprint at <https://docs.lib.purdue.edu/dissertations/AAI3307491/> (2007).
25. Naz, B. S., Ale, S. & Bowling, L. C. Detecting subsurface drainage systems and estimating drain spacing in intensively managed agricultural landscapes. *Agric. Water Manage.* **96**, 627–637 (2009).
26. Brown, D. C. Applying a Model to Predict the Location of Land Drained by Subsurface Drainage Systems in Central Minnesota. *Pap. Resour. Anal. Saint Mary's Univ. Minnesota Univ. Serv.* (2013).
27. Nakagaki, N., Wieczorek, M. E. & Qi, S. L. Estimates of Subsurface Tile Drainage Extent for the Conterminous United States, early 1990s. *US Geological Survey*, <https://doi.org/10.5066/F7RB72QS> (2016).
28. Nakagaki, N. & Wieczorek, M. E. Estimates of subsurface tile drainage extent for 12 Midwest states. *US Geological Survey*, <https://doi.org/10.5066/F7W37TDP> (2016).
29. Cho, E., Jacobs, J. M., Jia, X. & Kraatz, S. Identifying Subsurface Drainage using Satellite Big Data and Machine Learning via Google Earth Engine. *Water Resour. Res.* **55**, 8028–8045 (2019).
30. Northcott, W. J., Verma, A. K. & Cooke, R. A. Mapping subsurface drainage systems using remote sensing and GIS. *ASAE Ann. Int. Meeting*. 2625–2634 (2000).
31. Varner, B. L., Gress, T., Copenhaver, K. & White, S. The effectiveness and economic feasibility of image based agricultural tile maps. *Final Rep. NASA ESAD.* (2002).
32. Tetzlaff, B., Kuhr, P. & Wendland, F. A new method for creating maps of artificially drained areas in large river basins based on aerial photographs and geodata. *Irrig. Drain.* **58**, 569–585 (2009).
33. NASS, USDA. USDA/NASS QuickStats Ad-hoc query tool. *US Dep. Agric.* (2014).
34. ESRI. ArcMap 10.7. *Environ. Syst. Res. Institute, Redlands, California, USA* (2019).
35. Jin, S., Homer, C., Dewitz, J., Danielson, P. & Howard, D. National Land Cover Database (NLCD) 2016 Science Research Products. *AGUFM* **2019**, B111–2301 (2019).
36. USGS. USGS EROS Archive - Digital Elevation - Shuttle Radar Topography Mission (SRTM) Void Filled 1 Arc-Second Global. *Earth Resour. Obs. Sci. Cent.* 1–7, <https://doi.org/10.5066/F7F76B1X> (2018).
37. Valayamkunnath, P., Barlage, M., Chen, F., Gochis, D. & Franz, K. AgTile-US. *figshare* <https://doi.org/10.6084/m9.figshare.11825742> (2020).
38. Olofsson, P. *et al.* Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* **148**, 42–57 (2014).
39. Boryan, C., Yang, Z., Mueller, R. & Craig, M. Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto Int.* **26**, 341–358, <https://doi.org/10.1080/10106049.2011.562309> (2011).
40. Gesch, D., Oimoen, M., Zhang, Z., Meyer, D. & Danielson, J. Validation of the ASTER global digital elevation model version 2 over the conterminous United States. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **39**, B4 (2012).
41. Gesch, D. B., Oimoen, M. J. & Evans, G. A. Accuracy assessment of the US Geological Survey National Elevation Dataset, and comparison with other large-area elevation datasets: SRTM and ASTER. Report No. 2014-1008 (US Geological Survey, 2014).

### Acknowledgements

We would like to acknowledge the support from the NOAA OAR grant NA18OAR4590381, NCAR Water System, USDA NIFA grants 2015-67003-23508 and 2015-67003-23460, and the NSF grant 1739705. We are grateful to three anonymous reviewers for their valuable comments.

### Author contributions

P.V. and M.B. conceived the idea. P.V. performed the analysis and developed the data. P.V., M.B., F.C., D.G. and K.F. wrote the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41597-020-00596-x>.

**Correspondence** and requests for materials should be addressed to P.V.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2020