A new dataset of global irrigation areas from 2001 to 2015

Deepak Nagaraj[1], Eleanor Proust[1], Alberto Todeschini[1], Maria Cristina Rulli[3], Paolo D'Odorico[2]

[1] *School of Information, University of California, Berkeley, USA*

[2] *Department of Environmental Science, Policy, and Management, University of California, Berkeley, USA*

[3] *Department of Civil and Environmental Engineering, Politecnico di Milano, Italy*

Abstract

About 40% of global crop production takes place on irrigated land, which accounts for approximately 20% of the global farmland. The great majority of freshwater consumption by human societies is associated with irrigation, which contributes to a major modification of the global water cycle by enhancing evapotranspiration and reducing surface and groundwater runoff. In many regions of the world irrigation contributes to streamflow and groundwater depletion, soil salinization, cooler microclimate conditions, and altered land-atmosphere interactions. Despite the important role played by irrigation in food security, water cycle, soil productivity, and near-surface atmospheric conditions, its global extent remains poorly quantified. To date global maps of irrigated land are often outdated and based on estimates from circa year 2000. Here we apply artificial intelligence methods based on machine learning algorithms to satellite remote sensing and monthly climate data to map the spatial extent of irrigated areas between 2001 and 2015. We provide global annual maps of irrigated land at ≈9km resolution for the 2001-2015 and we make this dataset available online.

Introduction

The global demand for agricultural products is increasing as a result of demographic growth, shifts to resource-intensive diets, and increasing reliance on biofuels (Godfray et al., 2010; Foley et al., 2011; Cassidy et al., 2013). To sustain these ongoing trends, global crop production will have to more than double by 2100 (Beltran-Pena et al., 2020), thereby dramatically increasing human pressure on the limited land and water resources of the planet (e.g., Falkenmark et al., 2006; Ramankutty et al., 2008; Rockstrom et al., 2009; Cassidy et al., 2013). Despite the big push for food security pathways that rely on more efficient use of resources, reduction of food waste, and moderation of consumption (Kummu et al., 2012; Davis et al., 2014; Springmann et al., 2018), the demand for increased agricultural production will be unavoidable. It will require either the expansion of agriculture at the expenses of natural ecosystems such as forests, savannas, and grasslands, or the increase in crop yields in the land that is currently cultivated (Foley et al., 2011). Known as "agricultural intensification" the latter approach would prevent additional losses of natural habitat and biodiversity and avoid the greenhouse gas emissions associated with land conversions (Runyan and D'Odorico, 2016). At the same time, intensification will require the provision of additional inputs that are needed to improve yields through an adequate supply of fertilizers and water (Erisman et al., 2012; Rosa et al., 2018). In many regions of the world the closure of the gap between actual and maximum potential yields requires irrigation (Mueller et al., 2012). Previous studies have mapped the rainfed croplands where the local water resources are sufficient to sustainably meet the local irrigation water requirements (Jägermeyr et al 2017; Rosa et al., 2018; 2020a). A major limitation in this line of research is the lack of knowledge of the extent and distribution of irrigated land. Most studies rely on a reconstruction of the areas equipped for agriculture around year 2000 (Portmann et al., 2010). A recent extension of these analyses has reconstructed the global history of irrigated areas between 1900 and 2005 (Siebert et al., 2015), while of more regional studies have mapped irrigated areas at 30m resolution for South Asia and Australia (GFSAD30) (Salmon et al., 2015; Meier et al., 2018). Therefore, it is often reported that roughly 20% of the cultivated land is irrigated and accounts for 40% of the global crop production (e.g., Molden, 2010). These estimates, however, are outdated, as the spatial extent of irrigation has likely changed over the past two decades. With agriculture contributing to 90% of human water consumption, lack of knowledge of irrigated areas prevents an analysis of the extent to which water resources around the world are used sustainably (i.e., without depleting local groundwater stocks and environmental flows). It also limits our ability to investigate ongoing changes in agricultural practices around the world (Davis et al., 2017; Jägermeyr et al 2017; D'Odorico et al., 2018; Rosa et al., 2019).

Satellite remote sensing combined with modern machine learning techniques provides unprecedented opportunities to identify areas equipped for irrigation, where irrigation may therefore take place for at least part of the year. Irrigated areas tend to exhibit higher productivity and "greenness" than their rainfed counterparts in the same region. As a result, irrigation is expected to be detectable with greenness indices such as EVI and NDVI (e.g., Kotsuki and Tanaka, 2015). Moreover, there is compelling evidence (Muller et al., 2016; 2017; Thiery et al., 2020) that irrigated areas tend to be cooler during the day than adjacent rainfed land as a result of the different partitioning of the incoming solar radiation into sensible and latent heat fluxes, with irrigated soil exhibiting greater evapotranspiration and associated latent heat fluxes (Kueppers et al., 2007; Puma et al., 2010; Lobell et al., 2008; 2009; Bonfils

and Lobell, 2007). Therefore, surface temperature, which can be detected from space, is expected to be a good proxy for irrigation (Wei et al., 2013; Cook et al., 2020).

In this study, we develop a global assessment of irrigation and provide global annual maps of irrigated land at ≈9km resolution for the 2001-2015 period. To that end, we apply advanced machine learning methods to data available from satellite remote sensing, using as "label" (to train the algorithm) the known distributions of irrigated land from Siebert et al. (2015). We make this dataset available on the Zenodo repository https://zenodo.org/deposit/4392826 .

2. Data and Methods

Irrigation maps are produced by two machine learning models: (A) a time-series model, and (B) a point-in-time (time-stationary) model, as explained below. The results from these two models are then combined for better overall performance.

2.1 Data sets

Both models are trained on a global dataset of irrigation extent (the time-series model is trained only on areas identified as croplands), originating from Siebert et al. (2015). This dataset is available at a spatial resolution of 5 arc minutes, i.e. 1 pixel of the map corresponds to an area of 86 km^2 (or 8604 ha) at the equator, which corresponds to a pixel side of ≈9.276 km. We use "area equipped for irrigation" data from the Siebert et al. (2015) dataset, specifically the one listed as "HYDE Final" as labels to train the algorithm.

2.1.1 Time-Series Model

The four available "HYDE final" datasets from Siebert et al. (2015) for 1985-2000 were used as "labels". Because satellite data used in this study were not consistently available pre-1981, labels from previous years were discarded. Given the restraints on satellite and geographic data available from 1985-2015, we used as features the following datasets: (a) NDVI Data sourced from the AVHRR 15 day data set and compiled in the Global Inventory Monitoring and Modeling System, GIMMS (NCAR, 2018). These data were aggregated biannually to take the mean, variance and max of all inputs over a 6 month period; (b) TerraClimate data were taken from the University of Idaho online archives and re-projected to the same resolution as the NDVI data and labels (Abatzoglou, et al., 2018). For each year, monthly values of climate variables were used to estimate annual mean, maximum, variance and minimum. The long-term averages aggregated by TerraClimate for the years 1981-2010 were also fed to the model as a stand-in for long term suitability for agriculture. The datasets that proved most useful to the purpose of this study were maximum temperature, minimum temperature, vapor pressure, downward surface shortwave radiation, wind-speed, actual evapotranspiration, climate water deficit (i.e., the difference between potential and actual evapotranspiration), and soil moisture. These hydroclimatic variables are known to control evapotranspiration rates, and crop water demand (e.g., Katul et al., 2012).

Given the relatively high computing requirements to train the model globally over two decades, a decision was made to concentrate on cultivated land by filtering the data based on the 2010 GFSAD

(Global Food Security-support Analysis Data) croplands extent, including irrigated and rainfed croplands (Massey et al., 2017). This mask was selected because it contained most of the irrigated land and was more extensive than the MODIS land cover data set. Applying this mask reduced the required computing time by a factor of ten and allowed for the inclusion of many more training features.

The time series model was performed in two stages. Firstly, a binary test was used to determine whether at least 1% of land was irrigated. Secondly, any land that was detected to be irrigated was then run through another binary test to detect whether the land was highly irrigated (defined as a pixel with at least 20% of the area or 2000 hectares irrigated). The result was a three class feature irrigated at 1%, irrigated between 1 to 20%, and irrigated 20% or more.

Within the GFSAD cropland area the non-irrigated land is more than four times greater than the irrigated land (i.e., about 20% of the cultivated land is irrigated (e.g., Molden, 2010)). Decision tree models (including 'random forests' and 'extra trees') tend to not perform well with such an imbalance between classes. Therefore, we under-sampled the non-irrigated land at a rate of 60%. This significantly improved the overall performance of the model.

The F1 and kappa metrics were used to evaluate the performance of the model. Given the class imbalance, a model could deliver relatively high accuracy by only predicting non-irrigated land. The kappa metric represents how much better the model performs with respect to a baseline model that simply predicts based on class frequencies (or "expected accuracy"). Thus, kappa is expressed as (model accuracy – expected accuracy) / (1 – expected accuracy). Kappa adjusts the accuracy score to reflect class imbalance and therefore Kappa is a better metric when there is class imbalance (i.e., there is no even split between classes as in our case, as we have only 11% irrigated land). The F1 score is a measure that incorporates both recall and precision. The kappa metric of the first stage processing was 78%, with an F1 score of 92%. On the second layer of learning, differentiating between "highly irrigated" and other irrigation classes kappa was 76% and F1 again was 92%; this led to an overall model performance of kappa=0.76 and accuracy (a measure of how often the model classifies correctly) of 0.89, and F1 of 0.88.

2.1.2 Time-Stationary Model

For the time-stationary model, we used labels from Siebert et al. (2015) for the year 2005 and the same spatial resolution as in the time series model (i.e., with pixels of 8604 ha, which corresponds to a pixel side of about 9.276 km). We ignored pixels with less than 25 hectares of land equipped for irrigation and classified them as not irrigated. Pixels with 25 to 2000 hectares equipped for irrigation were classified as "low or medium" irrigation. Pixels with more than 2000 hectares of land equipped for irrigation were classified as "high" irrigation. Because the distribution is highly skewed, it drops rapidly as we go to higher values of irrigated areas -- there is only a relatively limited amount of data at higher levels of irrigation.

We took a random sample of 20,000 points worldwide to develop a machine learning model. To train the model, we used features pertaining to climate, soil, vegetation and land-cover from TERRACLIMATE (Abatzoglou et al., 2018), the Global Land Data Assimilation System, GLDAS (Rodell et al., 2004), and MODIS datasets. Our final choice was a random forest model with 1000 trees, a bagging fraction of 0.63, and 10 variables per split. We also used a random forest algorithm to select features based on their importance. Our final model used 11 features, including latitude and longitude (X and Y), annual average maximum temperature (tmmx), potential evapotranspiration (pet), downward surface shortwave radiation (srad), and wind-speed (vs) from TERRACLIMATE; annual average albedo

(Albedo_inst), direct evaporation from bare soil (ESoil_tavg), atmospheric pressure (Psurf_f_inst) from GLDAS; land cover type (MCD12Q1 for UMD) and annual maximum EVI from MODIS. The relative importance of these predictors is shown in Figure 1. Again, the microclimate variables chosen as "features" are known for playing a role in determining the rates of evapotranspiration and the crop water requirements (e.g., Katul et al, 2012).

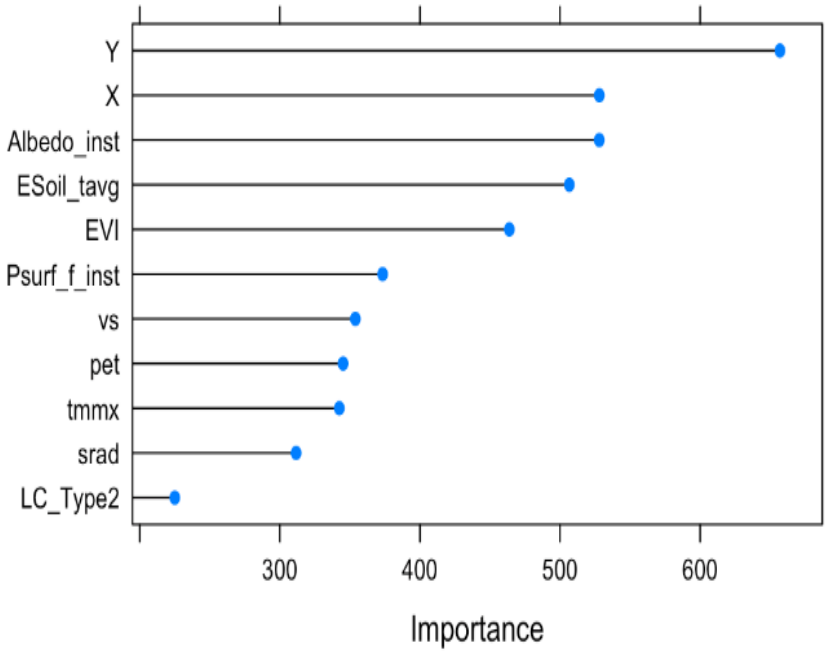

Figure 1. Relative importance of the features used for the time-stationary model, which included including latitude and longitude (X and Y), annual average maximum temperature (tmmx), potential evapotranspiration (pet), downward surface shortwave radiation (srad), and wind-speed (vs) from TERRACLIMATE; annual average albedo (Albedo_inst), direct evaporation from bare soil (ESoil_tavg), atmospheric pressure (Psurf_f_inst) from GLDAS; land cover type (MCD12Q1 for UMD) and annual maximum EVI from MODIS.

We used the R statistical environment for model development and tuning. We then exported this model to Google Earth Engine® and ran it there for the years 2001 through 2015. This platform provided us with a repository to access many geospatial datasets, as well as a distributed computing infrastructure that allowed us to process large amounts of feature data and classify the cultivated land based on the occurrence and intensity of irrigation. The time-stationary model has accuracy of 0.89, and a Kappa value of 0.56.

2.2 Combining the Two Models

We then combined predictions from the above two models. First, we partitioned the world's land area into croplands and non-croplands using the GFSAD crop dominance dataset to determine the cropland boundaries (considering all crop classes in GFSAD). For croplands, we predict irrigation using the time-series model. For non-croplands, we predict with the time-stationary model. Our combined map has accuracy of 0.94, and a kappa value of 0.73. The confusion matrix and summary statistics are shown in Tables 1 and 2.

Table 1. Confusion matrix of the combined model

↓ Predicted | Actual →	Not irrigated	Low to medium	High	TOTAL
Not irrigated	1958518	41151	2150	2001819
Low to medium	59439	136056	7626	203121
High	1455	9313	38468	49236
TOTAL	2019412	186520	48244	2254176

Table 2. Summary statistics or accuracy, precision, recall, and F1-score

Class	n(truth)	n(classified)	Accuracy	Precision	Recall	**F1-score**
Not irrigated	2019412	2001819	95.38%	0.98	0.97	**0.97**
Low to medium irrigated	186520	203121	94.79%	0.67	0.73	**0.70**
Highly irrigated	48244	49236	99.09%	0.78	0.80	**0.79**

2.3 Validation

After running our model, we validated the false positives and false negatives produced by the model, by taking a random sample of 100 points in each class. We downloaded LANDSAT 7 TOA (Top of the Atmosphere) imagery for these points, for a buffer of 9 km and a buffer of further 9 km. We then visually inspected each picture to validate the model classification.

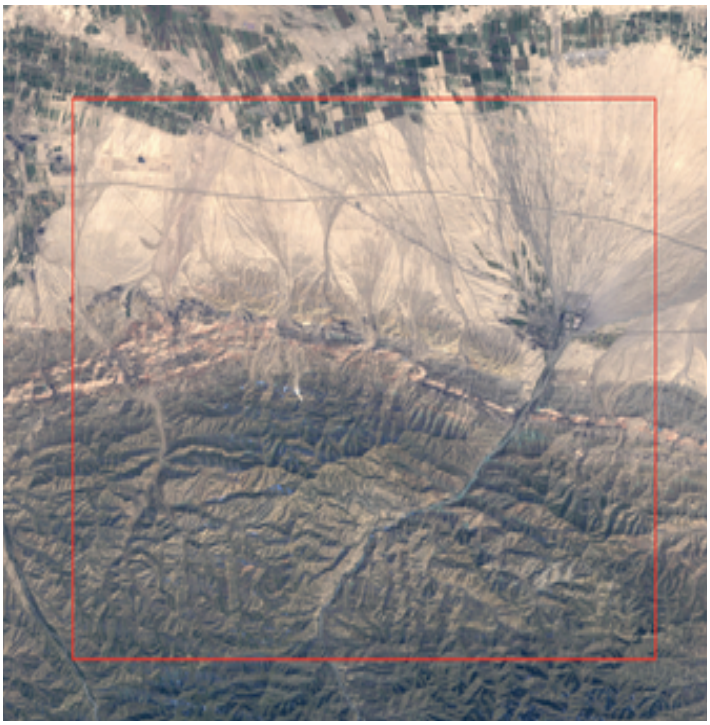

Figure 2. An example of pixels exhibiting a mosaic of croplands and non-croplands (the red square wraps a circle of 9.3km diameter)

3. Results

We developed annual global maps of irrigation (Figure 3), encompassing all continents except Antarctica and the Pacific islands of Oceania. These maps are at a spatial resolution of 5 arc minutes (9276 m at the equator) and are available for every year from 2001 through 2015. The maps show the extent of 3 classes of irrigation: (i) no irrigation, (ii) low-concentration irrigation (1% to 20% irrigated), and (iii) high-concentration irrigation (>20% irrigated). The model predicts that as of 2005 11.2% of the world's land was used as irrigated cropland. Of this, 80.5% is in the low irrigation class, and 19.5% is in the high irrigation class. The confusion matrix (Table 1) shows that 97% of the non-irrigated areas were correctly predicted as non-irrigated; 73% of low-concentration irrigated areas and 80% of high concentration irrigated areas were correctly predicted as irrigated at low or high concentrations, respectively. These discrepancies between predictions and actual irrigation are for most part due to 22% of low irrigation areas being mistakenly classified as non-irrigated and 16% of high-concentration irrigation areas being mistakenly classified as low-concentration (Table 1). We use a validation process to evaluate to what extent these discrepancies are due to limitations in our predictions or are the result of inconsistencies in our label and do not correspond to actual misclassifications in our predictions.

The results of the validation process show that model performance is better than what is reported in Sections 2.2 and 2.3. In fact, for points that are false positives, we found that 41 out of 100 points were croplands (27) or a mosaic of croplands and non-croplands such as uncultivated hills, water bodies or forests (14). This means that the model was correct in its classification for these points. The remaining 59 points were non-croplands. For points that are false negatives, we found that 47 out of 100 points were indeed non-croplands. This means the model was correct in its classification for these points. 33 points were a mosaic, and 20 points were croplands. These numbers show some of the limitations in the

use of Siebert et al., (2015) as training labels, bring uncertainty to the model assessment metrics reported above based solely on the training labels. These results show that our models fare poorly in areas which have a mosaic of croplands and non-croplands. This problem occurs in the Brazilian highlands, northern Germany, and southern China, where the terrain is a mix of hilly areas and croplands. Our spatial resolution assigns 86 square km to each pixel, and this can be too coarse to identify irrigation in such mosaics (Figure 2). Additionally, one of the feature datasets, GLDAS, is at even coarser resolution, which can cause predictions to mis-classify adjoining pixels. Moreover, mosaics have the effect of confusing our model in the low class of irrigation (Figure 2).

Globally, most irrigation occurs at low density (i.e., with less than 20% of the area being irrigated), particularly in sub-Saharan Africa, Oceania, and South America. South Asia and East Asia, however, exhibit a relatively large fraction of their irrigated areas at high-density irrigation (51% and 32%, respectively), pointing to regions of the world of particularly intensified crop production (Figure 3).

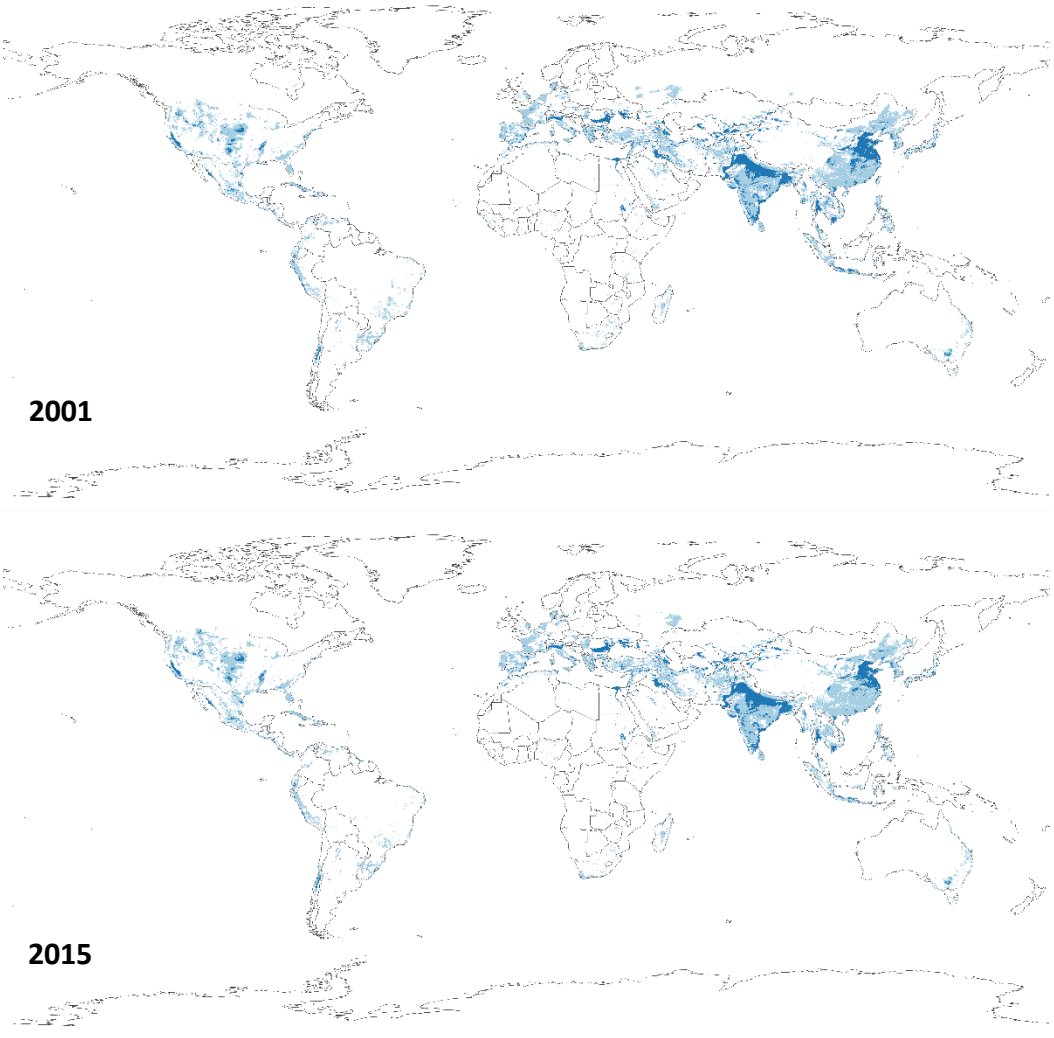

Figure 3 shows a map of the irrigated areas in 2001 (top) and 2015 (bottom). Dark blue corresponds to high density (i.e., >20%) irrigation and lighter blue to areas with low density (1-20%) irrigation.

Our predictions of irrigated and non-irrigated areas between 2001 and 2015 show some interesting trends (Figure 4). There is a slight decreasing trend in irrigation in high levels of irrigation. But there is an increasing trend in irrigation in low levels of irrigation (1-20% irrigated land). This may indicate increased irrigation because of drought or reduced rainfall in regions traditionally reliant on rainfall.

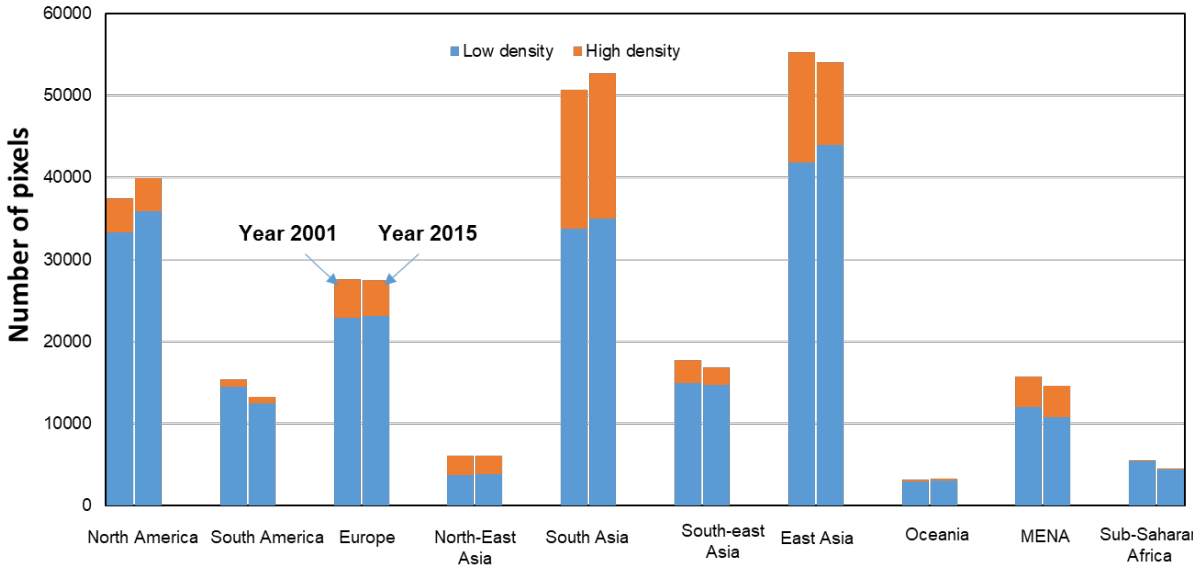

Figure 4. Number of pixels with low and high-density irrigation between year 2001 and 2015 (see figure S1 in the Appendix for the definition of these regions).

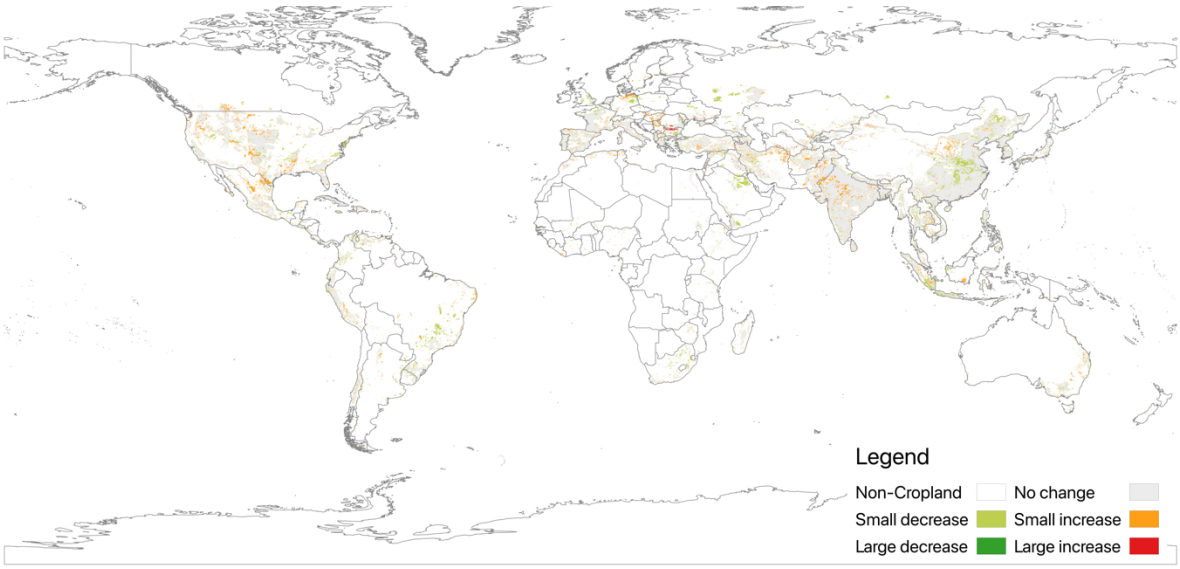

Figure 5. Difference in irrigation between 2015 and 2001. Green areas exhibited a decrease in irrigation; in orange areas an increase in irrigation. Darker colors correspond to bigger changes (i.e. from high density irrigation to no irrigation or vice versa); Lighter colors correspond to smaller changes in irrigation (from high to low density or from low density to no irrigation).

Interestingly, we see that irrigation has decreased in central parts of eastern China, Thailand, Saudi Arabia, Russia, and the Southeastern USA, Syria, Albania, Eastern Germany, Northern Italy, and certain parts of Brazilian highlands (Figure 5). We see an increase in irrigation in central parts of the United States (Figure 6), along the mid-low course of the Danube (Hungary, Serbia, Romania, and Bulgaria), Northern India, and in drier regions of Central Asia. Links to these global maps and the validation areas are provided in the online materials.

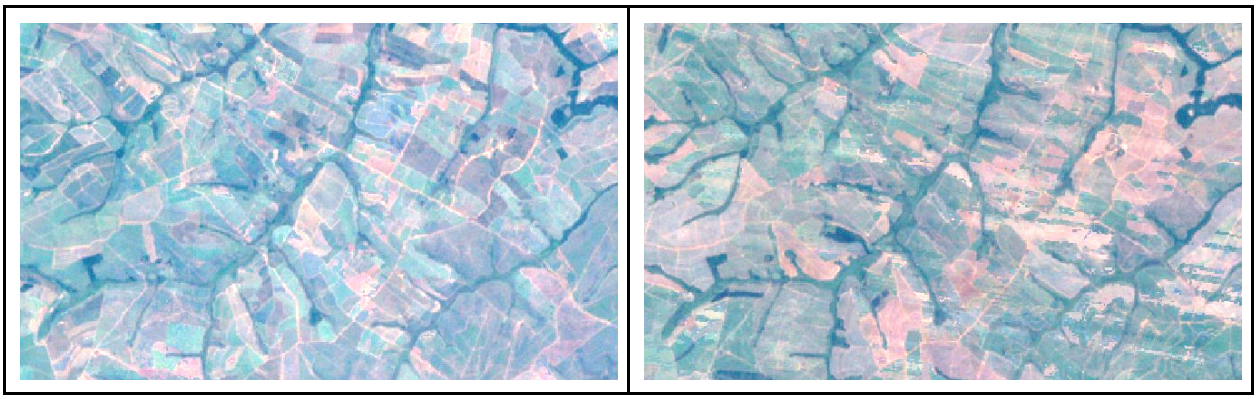

Figure 5. Example of detected decrease in Brazilian highlands: 2001 (left) vs. 2015 (right).

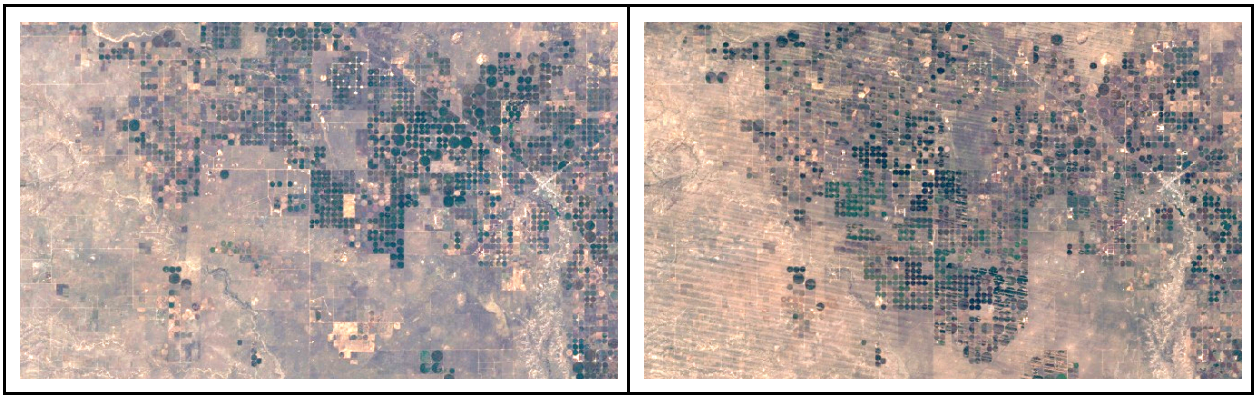

Figure 6. Example of detected increase of irrigation in central plains of North America: 2001 (left) vs. 2015 (right).

Discussion

This study differs from previous efforts based on statistical methods (Portmann et al., 2010), in that it uses machine learning algorithms to map irrigated areas worldwide. The dataset developed in this research provides a global scale mapping of low and high intensity/density irrigated areas with about 9 km resolution. This analysis allows us to investigate spatiotemporal patterns of irrigation worldwide. Specifically, we find that high density irrigated areas are found mostly in South Asia and East Asia, followed by North America, Europe and the Middle-East/North Africa (MENA) region. Between 2001 and 2015 irrigated areas have increased across North America and South Asia, and the increase was contributed by an expansion of low density irrigation areas. Conversely, South America and the MENA

region saw a decrease in irrigated areas as a result of a decrease in areas with low-density irrigation. The case of East Asia is different because, while low density irrigation areas have increased, low density irrigated areas have shrunk, leading to an overall decrease of irrigated areas. These patterns are expected to change as an effect of climate change with an expansion of irrigated agriculture at the mid-high latitude and a loss of irrigation suitability or the need for seasonal reservoirs in breadbasket regions of Eurasia (Rosa et al., 2020b).

A comparison between our results for 2015 and those obtained by Meier et al (2018) for the same years shows an overall agreement in the classification or irrigated and non-irrigated areas ranging between 81%-96% - depending on the region - with an average of 91%. When the comparison is limited to the areas classified as irrigated in our study the agreement ranges between 59%-81% with an average of 73%. Alternatively, we can measure the agreement as a percentage of the irrigated area in Meier et al (2018); in that case the agreement drops to 24%-77% with an average of 48%. The differences are likely a consequence of the different methods used by the two studies. We are unable to establish, however, which one of the two methods provides the correct classification. Most likely, both methods produce correct or wrong results in different regions. The validation results presented in the previous section for a sample of 100 pixels shows that for the year 2001 our classification tends to be in stronger agreement than the "label" (Siebert et al., 2015) with the signs of irrigation detectable with Google Earth in areas classified in this study as irrigated. Of course, a more extensive analysis would be needed to generalize these conclusions to the entire world.

Table 2. Comparison between the irrigated areas in this study and in Meier et al. (2018) for the year 2015.

Country	Total # Pixels	Irrigated in this study but not in Meier et al. (2018)	Irrigated for both	Irrigated in Meier et al., (2018) but not in this study	Agreement (% of total area)	Agreement (% of irrigated area in Meier et al., 2018)	Agreement (% of irrigated area in this study)
N. America	533,902	11,339	22,383	17,980	95	55	66
S. America	221,636	4,973	8,719	14,056	91	38	64
Europe	207,960	3,235	14,042	16,602	90	46	81
Russia	400,742	1,553	2,240	5,221	98	30	59
India	89,279	8,711	27,429	8,387	81	77	76
South East Asia	53,226	3,140	11,328	5,885	83	66	78
China	172,649	9,027	33,161	12,865	87	72	79
Oceania	111,050	1,109	1,608	3,997	95	29	59
UAE	167,892	4,392	7,088	8,153	93	47	62

Africa	262,258	1,677	2,801	9,107	96	24	63
Total	2,220,594	49,156	130,799	102,253	91	48	73

Acknowledgments

The authors thank Lorenzo Rosa (ETHZ) for providing comments and suggestions.

References

Abatzoglou, J.T., S.Z. Dobrowski, S.A. Parks, K.C. Hegewisch, (2018), Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015, Scientific Data.

Beltran-Peña, A. A., L. Rosa, and P. D'Odorico, (2020). "Global food self-sufficiency in the 21st century under sustainable intensification of agriculture", Environm. Res. Lett., 15(9) 095004.

Bonfils, C. and Lobell, D., 2007. Empirical evidence for a recent slowdown in irrigation-induced cooling. *Proceedings of the National Academy of Sciences*, *104*(34), pp.13582-13587.

Cassidy, E. S., West, P. C., Gerber, J. S., and Foley, J. A. (2013). Redefining agricultural yields: from tonnes to people nourished per hectare. Environmental Research Letters, 8(3), 034015. https://doi.org/10.1088/1748-9326/8/3/034015

Cook, B. I., et al., 2020. Regional climate consequences of maintaining current irrigation rates in the 21st century. Journal of Geophysical Research: Atmospheres, e2019JD031814

Davis, K.F., M.C. Rulli, and P. D'Odorico, "Moderating diets to feed the future", Earth's Future, 2, doi:10.1002/2014EF000254, 2014.

Davis, K.F., Rulli, M.C., Seveso, A. and D'Odorico, P., 2017. Increased food production and reduced water use through optimized crop distribution. Nature Geoscience, 10(12), pp.919-924.

D'Odorico, P., K.F. Davis, L. Rosa, J.A. Carr, D. Chiarelli, J. Dell'Angelo, J.A. Gephart, G.K. MacDonald, D.A. Seekell, S Suweis, M.C. Rulli, 2018. "The global food-energy-water nexus", Reviews of Geophysics, 56, 456–531, https://doi.org/10.1029/2017RG000591

Erisman, J. W., Sutton, M. A., Galloway, J., Klimont, Z., and Winiwarter, W. (2008). How a century of ammonia synthesis changed the world. Nature Geoscience, 1(10), 636. https://doi.org/10.1038/ngeo325

Falkenmark, M., and Rockström, J. (2006). The new blue and green water paradigm: Breaking new ground for water resources planning and management. https://doi.org/10.1061/(ASCE)0733-9496(2006)132:3(129)

Foley, J. A., et al. (2011). Solutions for a cultivated planet. Nature, 478(7369), 337-342. https://doi.org/10.1038/nature10452

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S.M., and Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. Science, 327(5967), 812-818. https://doi.org/10.1126/science.1185383

Katul, G. G., R. Oren, S. Manzoni, C. Higgins, and M. B. Parlange (2012), Evapotranspiration: A process driving mass transport and energy exchange in the soil-plant-atmosphere-climate system, Rev. Geophys., 50, RG3002, doi:10.1029/2011RG000366

Kotsuki, S. and Tanaka, K., 2015. SACRA--global data sets of satellite-derived crop calendars for agricultural simulations: an estimation of a high-resolution crop calendar using satellite-sensed NDVI. Hydrology and Earth System Sciences Discussions, 12(1).

Kueppers, L. M., M. A. Snyder, and L. C. Sloan, 2007. Irrigation cooling effect: Regional climate forcing by land-use change. Geophys. Res. Lett., 34, L03703.

Kummu, M., de Moel, H., Porkka, M., Siebert, S., Varis, O., and Ward, P. J. (2012). Lost food, wasted resources: Global food supply chain losses and their impacts on freshwater, cropland, and fertiliser use. Science of the total environment, 438, 477-489. https://doi.org/10.1016/j.scitotenv.2012.08.092

Jägermeyr, J., Pastor, A., Biemans, H., and Gerten, D. (2017). Reconciling irrigated food production with environmental flows for Sustainable Development Goals implementation. Nature Communications, 8.

Lobell, D., Bala, G., Mirin, A., Phillips, T., Maxwell, R. and Rotman, D., 2009. Regional differences in the influence of irrigation on climate. *Journal of Climate*, *22*(8), pp.2248-2255.

Massey, R., Sankey, T.T., Yadav, K., Congalton, R.G., Tilton, J.C., Thenkabail, P.S. (2017). NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) Global Food Security-support Analysis Data (GFSAD) @ 30m for North America: Cropland Extent Product (GFSAD30NACE). NASA EOSDIS Land Processes DAAC.

Meier, J., Zabel, F. and Mauser, W., 2018. A global approach to estimate irrigated areas–a comparison between different data and statistics. *Hydrology and Earth System Sciences*, *22*(2), p.1119.

Molden, D., Oweis, T., Steduto, P., Bindraban, P., Hanjra, M. A., and Kijne, J. (2010). Improving agricultural water productivity: between optimism and caution. Agricultural Water Management, 97(4), 528-535. https://doi.org/10.1016/j.agwat.2009.03.023

Mueller, N. D., Gerber, J. S., Johnston, M., Ray, D. K., Ramankutty, N., and Foley, J. A. (2012). Closing yield gaps through nutrient and water management. Nature, 490(7419), 254-257. https://doi.org/10.1038/nature11420

Mueller, N.D., Butler, E.E., McKinnon, K.A., Rhines, A., Tingley, M., Holbrook, N.M. and Huybers, P., 2016. Cooling of US Midwest summer temperature extremes from cropland intensification. *Nature Climate Change*, *6*(3), pp.317-322.

Mueller, N.D., Rhines, A., Butler, E.E., Ray, D.K., Siebert, S., Holbrook, N.M. and Huybers, P., 2017. Global relationships between cropland intensification and summer temperature extremes over the last 50 years. *Journal of Climate*, *30*(18), pp.7505-7528.

NCAR, National Center for Atmospheric Research Staff, 2018. "The Climate Data Guide: NDVI: Normalized Difference Vegetation Index-3rd generation: NASA/GFSC GIMMS." Retrieved from https://climatedataguide.ucar.edu/climate-data/ndvi-normalized-difference-vegetation-index-3rd-generation-nasagfsc-gimms. (last accessed 8/16/2020).

Portmann F T, Siebert S and Döll P 2010 MIRCA2000—global monthly irrigated and rainfed crop areas around the year 2000: a new high-resolution data set for agricultural and hydrological modeling Global Biogeochem. Cycles 24 GB1001

Puma, M. J., and B. I. Cook, 2010. Effects of irrigation on global climate during the 20th century. J. Geophys. Res., 115, D16120.

Ramankutty, N., Evan, A. T., Monfreda, C., and Foley, J. A. (2008). Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. Global Biogeochemical Cycles, 22(1). https://doi.org/10.1029/2007GB002952

Rockström, J., et al. (2009). A safe operating space for humanity. Nature, 461(7263), 472-475. https://doi.org/10.1038/461472a

Rodell, M., P.R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J.K. Entin, J.P. Walker, D. Lohmann, and D. Toll, The Global Land Data Assimilation System, Bull. Amer. Meteor. Soc., 85(3), 381-394, 2004.

Rosa., L., M.C. Rulli, K.F. Davis, D. Chiarelli, C. Passera, P. D'Odorico, 2018. “Closing the yield gap while ensuring water sustainability", Environm. Res. Lett., 13 104002.

Rosa, L., D. Chiarelli, C. Tu, M.C. Rulli, and P. D'Odorico, (2019). “Global unsustainable virtual water flows in agricultural trade”, Environm. Res. Lett., 14, 114001.

Rosa, L., D.D. Danilo Chiarelli, M.C. Rulli, J. Dell’Angelo, and P. D’Odorico (2020a). “Global agricultural economic water scarcity”, Science Advances, 6: eaaz6031.

Rosa, L., D.D. Chiarelli, M. Sangiorgio, A.A. Beltran-Peña, M.C. Rulli, P. D’Odorico, and I. Fung (2020b). “Potential for sustainable irrigation expansion in a 3C warmer climate”, Proc. Natnl. Acad. Sci. USA, https://doi.org/10.1073/pnas.2017796117

Runyan, C.W. and P. D’Odorico. Global Deforestation, Cambridge University Press, New York, 248 pp., 2016.

Salmon, J.M., Friedl, M.A., Frolking, S., Wisser, D. and Douglas, E.M., 2015. Global rain-fed, irrigated, and paddy croplands: A new high resolution map derived from remote sensing, crop inventories and climate data. *International Journal of Applied Earth Observation and Geoinformation*, *38*, pp.321-334.

Siebert, S., Kummu, M., Porkka, M., Döll, P., Ramankutty, N. and Scanlon, B.R., 2015. A global data set of the extent of irrigated land from 1900 to 2005. Hydrology and Earth System Sciences, 19(3), pp.1521-1545.

Springmann, M., Clark, M., Mason-D’Croz, D., Wiebe, K., Bodirsky, B.L., Lassaletta, L., De Vries, W., Vermeulen, S.J., Herrero, M., Carlson, K.M. and Jonell, M., 2018. Options for keeping the food system within environmental limits. *Nature*, *562*(7728), pp.519-525.

Thiery, W., Visser, A.J., Fischer, E.M., Hauser, M., Hirsch, A.L., Lawrence, D.M., Lejeune, Q., Davin, E.L. and Seneviratne, S.I., 2020. Warming of hot extremes alleviated by expanding irrigation. *Nature communications*, *11*(1), pp.1-7.

Wei, J., Dirmeyer, P.A., Wisser, D., Bosilovich, M.G. and Mocko, D.M., 2013. Where does the irrigation water go? An estimate of the contribution of irrigation to precipitation using MERRA. Journal of Hydrometeorology, 14(1), pp.275-289.

Appendix

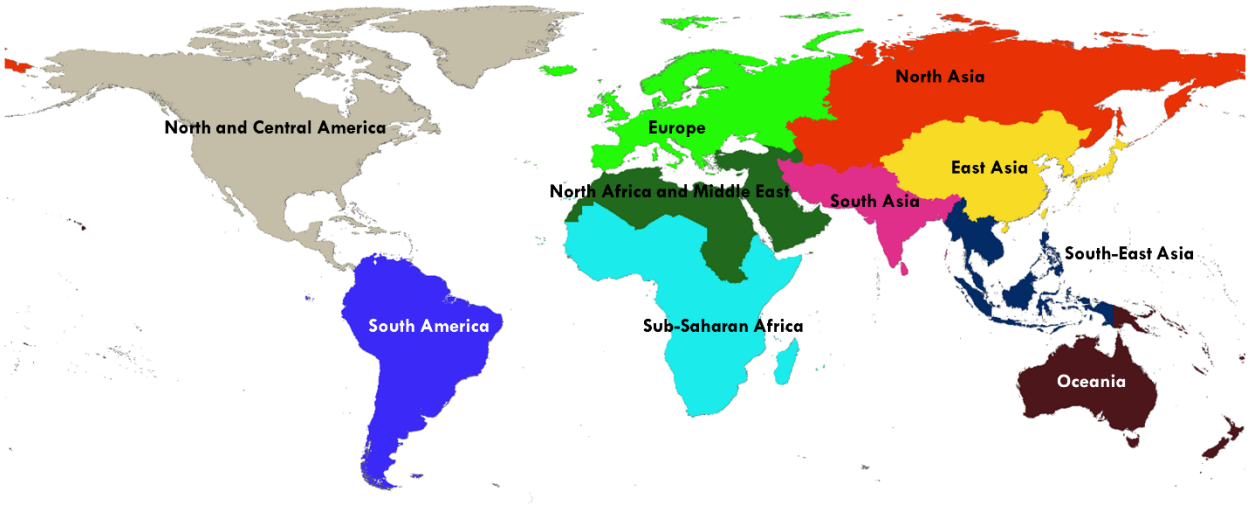

Figure S1. Regions used in figure 1 and 2.

In the Zenodo archive (https://zenodo.org/deposit/4392826) the reader can find:

Prediction maps: Prediction maps for every year from 2001 to 2015, in GeoTIFF format. Class 0 represents no irrigation, class 1 is low to medium irrigation, and class 2 is high irrigation.

Assessment map: Model assessment map for the year 2005.

Difference map: Map showing cropland differences between 2001 and 2015.