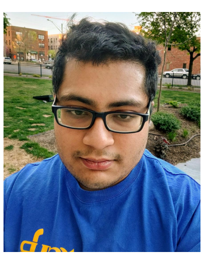
Integrating Remote Sensing and Machine Learning for Groundwater Withdrawal Estimation in Arizona

Sayantan Majumdar [1], Ryan Smith [1], Brian D Conway [2], James J Butler Jr [3], Venkataraman Lakshmi [4]

[1] Missouri University of Science and Technology, Rolla, MO, USA [2] Arizona Department of Water Resources, Phoenix, AZ. USA [3] University of Kansas, Kansas Geological Survey, Lawrence, KS, USA [4] University of Virginia, Charlottesville, VA, USA





PRESENTED AT:

AGU FALL MEETING
Online Everywhere | 1–17 December 2020

WATER RESOURCES | GROUNDWATER

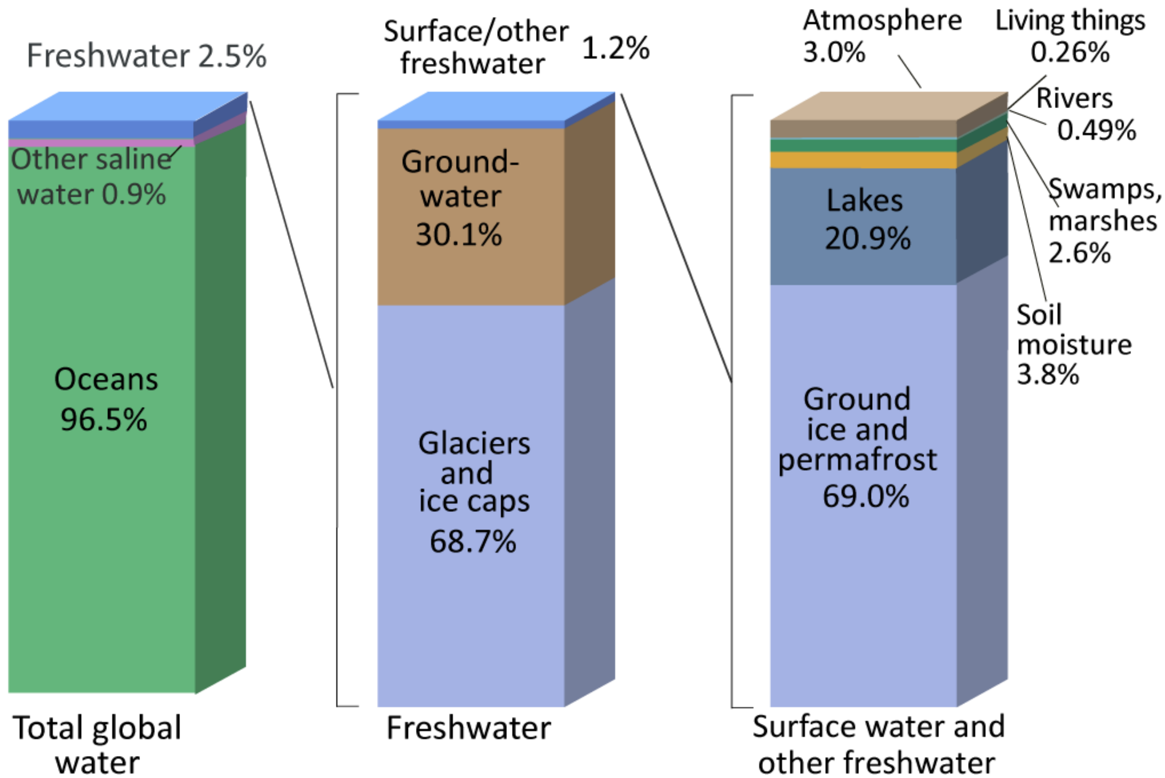


Figure 1: Distribution of Earth's water (USGS **(https://www.usgs.gov/special-topic/water-science-school/science/groundwater-storage-and-water-cycle)**).

- Groundwater is the largest source of Earth's liquid freshwater and plays a critical role in global food security. Hence, overuse of groundwater resources is a major concern.
- It is hard to estimate groundwater use or storage at local scales. Existing satellite methods for estimating groundwater storage change involve using GRACE/GRACE-FO (https://en.wikipedia.org/wiki/GRACE_and_GRACE-FO) data at a coarse resolution (~ 400 km).
- In this study, we combine publicly available datasets into a machine learning framework for estimating groundwater withdrawals (which are related to change in groundwater storage) at very high resolution (5 km) over the state of Arizona.

REMOTE SENSING & MACHINE LEARNING | WORKFLOW

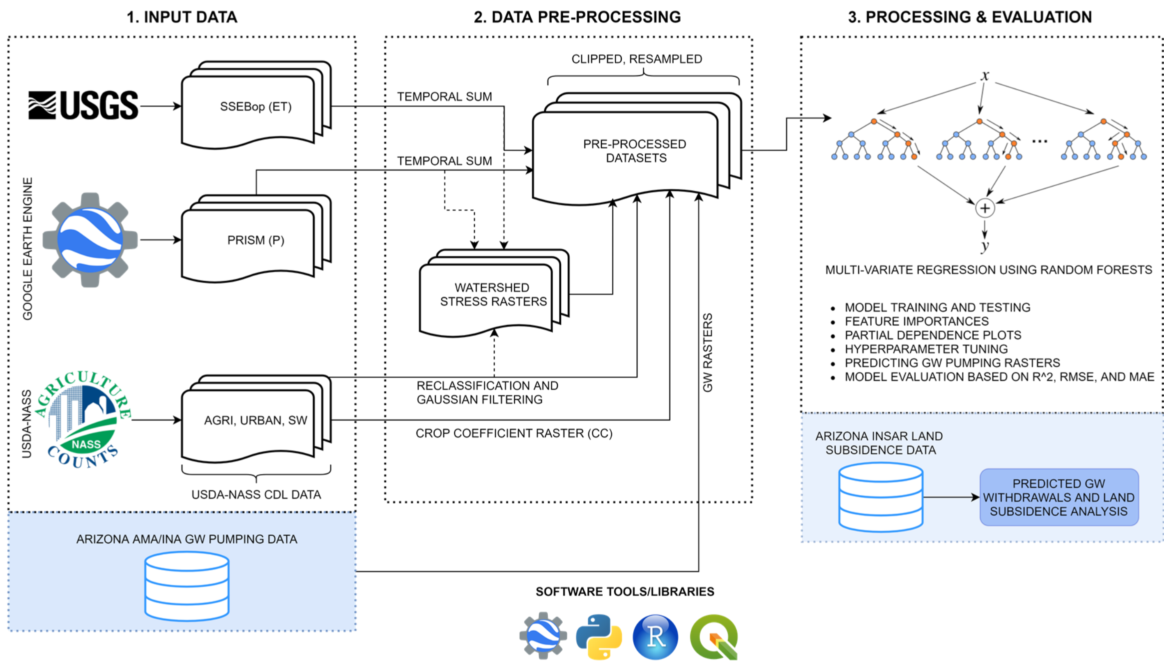


Figure 2: Overall workflow (Logos: official websites, RF figure: HARP/DSC-SPIDAL (https://dsc-spidal.github.io/harp/docs/examples/rf/)).

- Here, we use data from various sensors that measure different components of the water balance for monitoring groundwater withdrawal (SSEBop (https://earlywarning.usgs.gov/ssebop/modis), PRISM (https://prism.oregonstate.edu/), USDA-NASS CDL (https://www.nass.usda.gov/Research_and_Science/Cropland/SARS1a.php)).
- This extends a previous study (Majumdar et al., 2020) in which we estimated groundwater withdrawals in Kansas, where the climatic conditions and aquifer characteristics are significantly different.
- Random forests (https://en.wikipedia.org/wiki/Random_forest) (RF), a widely popular machine learning algorithm, are employed for predicting groundwater withdrawals from 2002-2018 at 5 km spatial resolution.
- We used in-situ groundwater withdrawals available over the Arizona Active Management Area (AMA (https://new.azwater.gov/sites/default/files/media/AMAFACTSHEET2016%20%281%29_0.pdf)) and Irrigation Non-Expansion Area (INA (http://infoshare.azwater.gov/docushare/dsweb/Get/Document-10190/Irrigation%20Non-Expansion%20Areas%20(INAs).pdf)) from 2002-2010 for training and 2011-2018 for validating the model respectively.
- The *predictors* for the RF model include ***SSEBop evapotranspiration (ET)***, ***PRISM precipitation (P)***, ***agriculture (AGRI)***, ***surface water (SW)***, ***urban (URBAN) densities, crop coefficient (CC),*** and ***watershed stress (WS) metrics*** (Smith & Majumdar, 2020)***, WS_PA*** (WS calculated using averaged P) and ***WS_PA_EA*** (WS calculated using averaged P adjusted with averaged ET), with the *response* variable being ***groundwater (GW) withdrawal***.

GROUNDWATER WITHDRAWALS | STUDY AREA, RESULTS, AND ANALYSIS

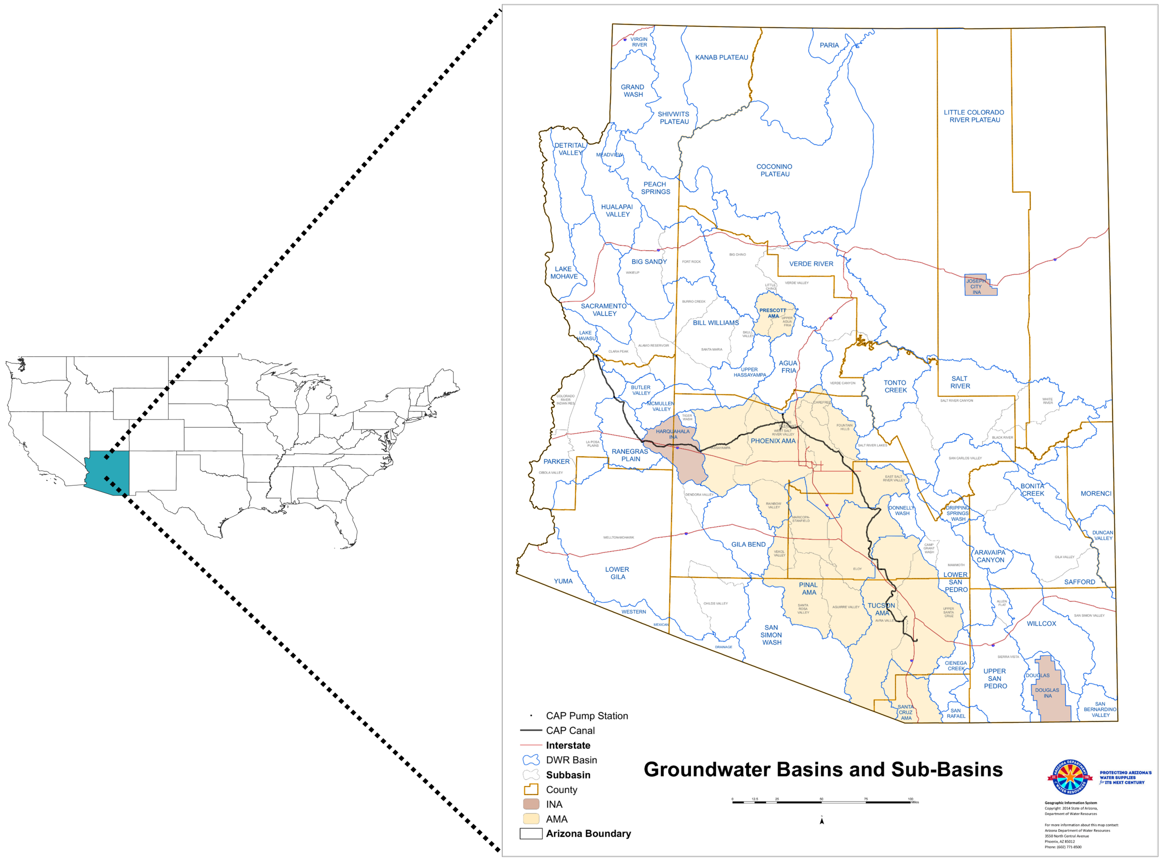


Figure 3: Groundwater basins and sub-basins in Arizona highlighting the AMA/INA regions. This map has been downloaded from the Arizona Department of Water Resources (ADWR) portal (https://new.azwater.gov/sites/default/files/GWBasin_ShowingCAP_0.pdf).

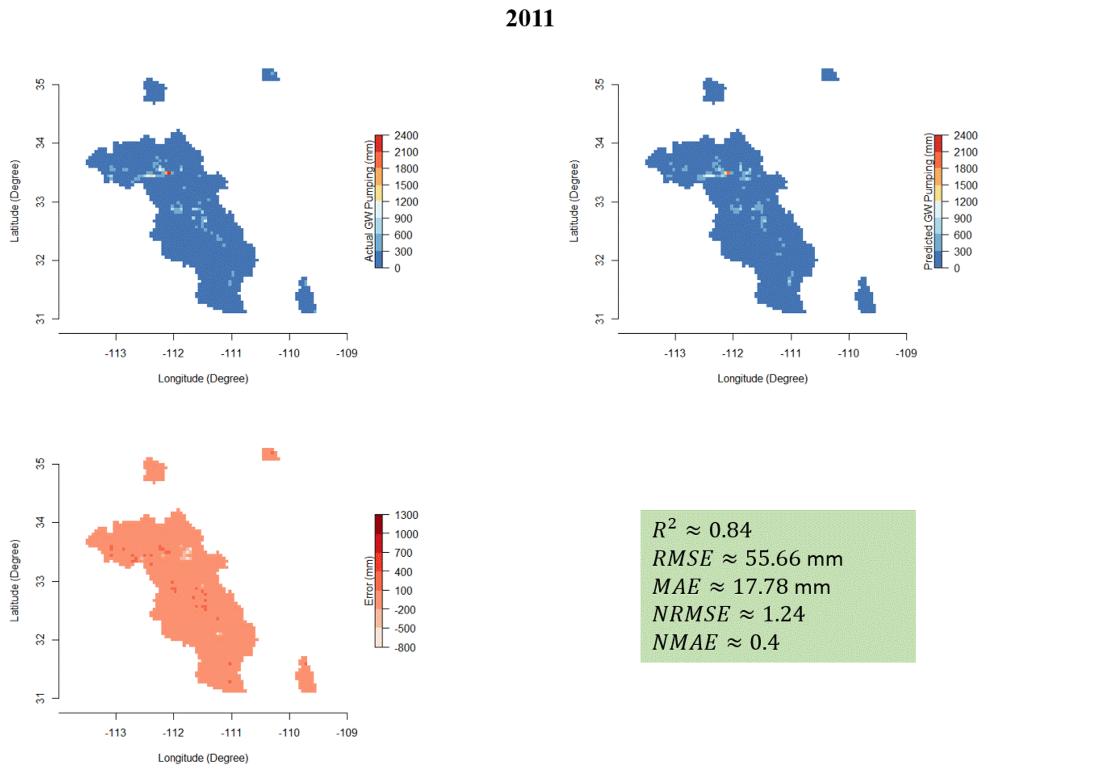


Figure 4: Groundwater (GW) pumping analysis for the test data (2011-2018). Here, the actual and predicted GW pumping rasters are shown along with the error raster. The residual error analysis plots are provided in the 4K resolution video of this GIF which can be viewed here (https://mailmissouri-my.sharepoint.com/:v:/g/personal/smxnv_umsystem_edu/ESy-R-N7s4BHoYjdkl3FPeYBiMWBIAAJg2HIs7Np05n-tw?e=QmFZge).

- For the overall test data (2011-2018), the error metrics include coefficient of determination (R2) ≈ 0.82, mean absolute error (MAE) ≈ 19.57 mm, root mean square error (RMSE) ≈ 60.85 mm, normalized MAE (NMAE) ≈ 0.41, and normalized RMSE (NRMSE) ≈ 1.29. The training R2 is 0.98.
- Here, the actual groundwater withdrawals outside the AMA/INA region are unknown. A total of 28,203 samples are present in this region spanning all the years from 2002-2019. The training (2002-2010) and testing (2011-2018) data split is 53%-47%.

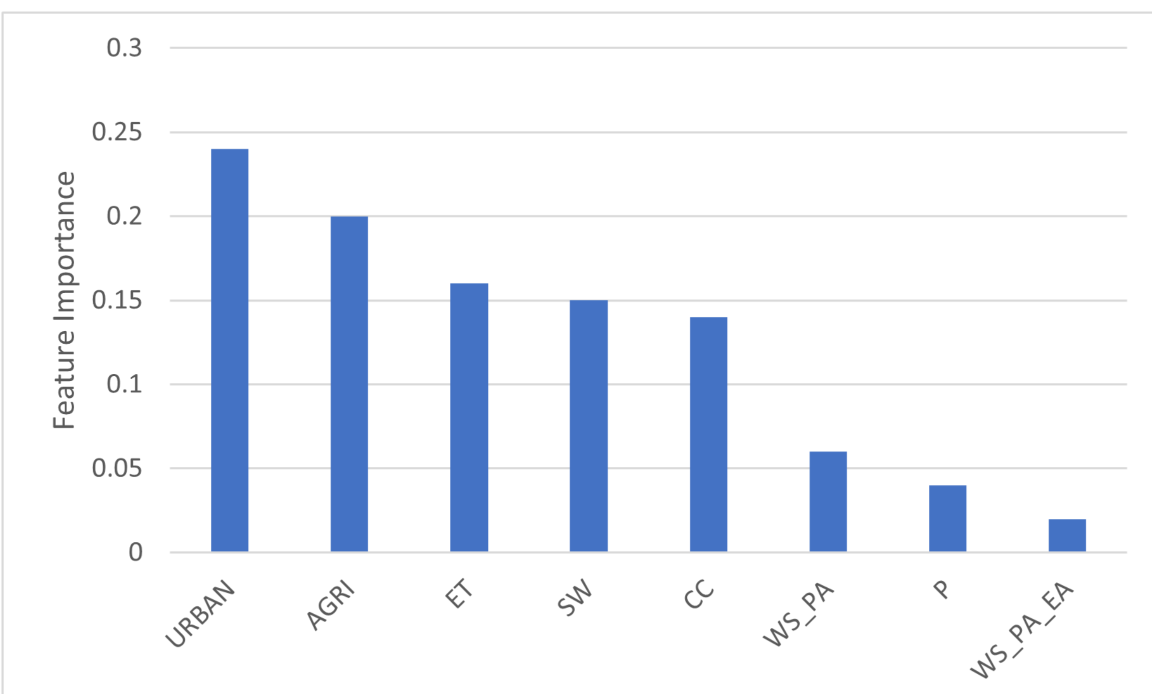


Figure 5: The feature or variable importances are values (sum up to 1) signifying the impact of each variable (the higher the value, the more important the feature).

- Accordingly, URBAN is the most important predictor followed by AGRI, ET, and SW.
- For the random forest model, we kept the number of trees as 500, and maximum number of featrues as 8.

GROUNDWATER WITHDRAWALS | MORE ANALYSIS, LAND SUBSIDENCE

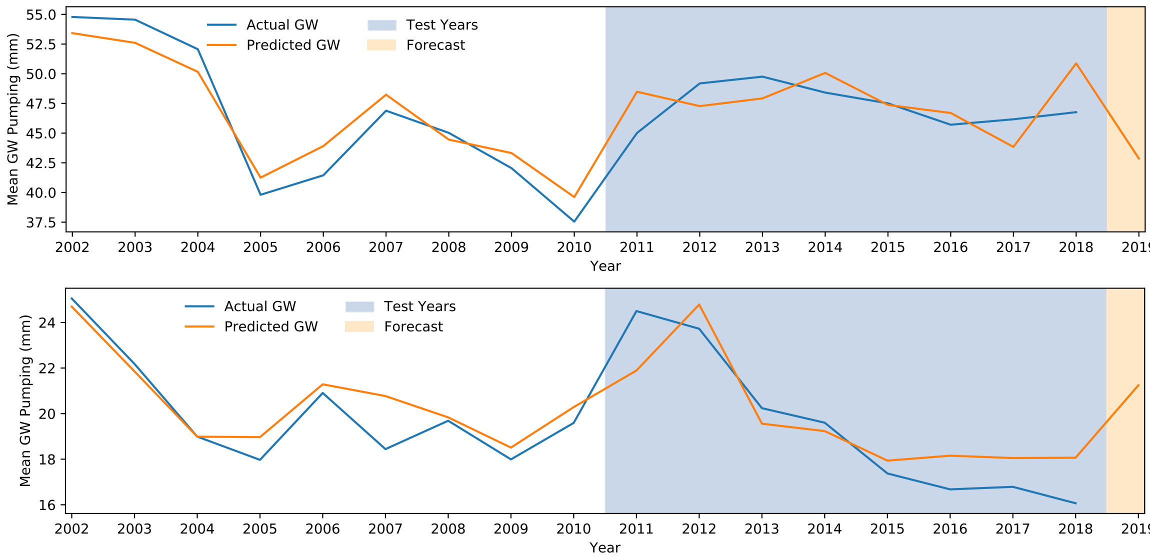


Figure 6: Mean Groundwater (GW) withdrawals for (a) the AMA/INA region in Arizona and (b) the entire state of Kansas (Majumdar et al., 2020).

- The predicted groundwater withdrawals at 5 km spatial resolution for both Arizona and Kansas show good accuracy.

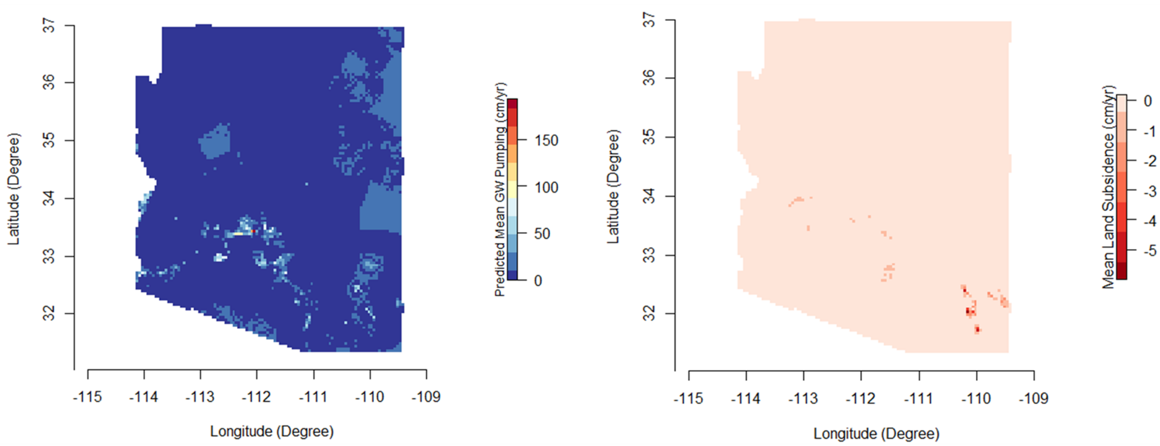


Figure 7: Predicted mean groundwater withdrawals and mean land subsidence maps for the 2010-2018 period.

- Most of the land subsidence is occurring in southern and south-eastern Arizona.
- All subsiding areas have high or moderate predicted groundwater pumping.
- Groundwater withdrawals are slightly correlated with land subsidence. Subsidence is a function of withdrawals, clay content and aquifer confinement. The mismatch between predicted pumping and subsidence can provide clues to these properties.

CONCLUSION AND FUTURE WORK

- Our machine learning model shows promising results in sub-humid and semi-arid (Kansas) and arid regions (Arizona) at very high resolution (5 km), which proves the robustness and extensibility of our integrated approach combining remote sensing and machine learning into a holistic, automated, and fully-reproducible workflow.
- The success of this method indicates that it could be extended to areas with more limited groundwater withdrawal data under different climatic conditions and aquifer properties.

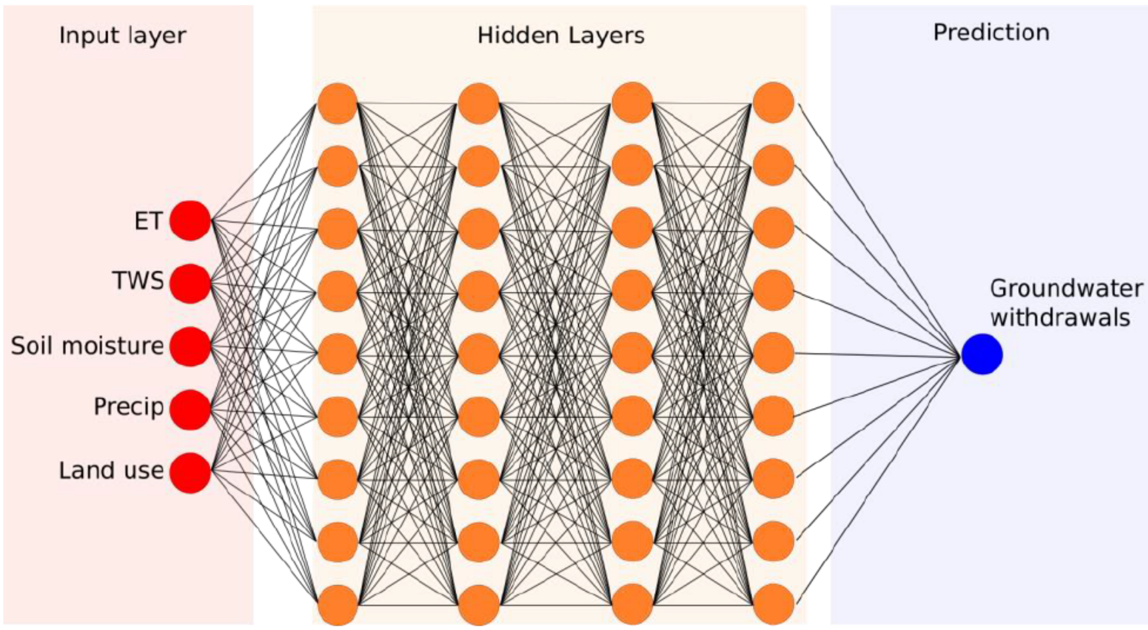


Figure 8: Proposed Deep Learning Framework.

- As part of future work, we plan to introduce additional datasets such as soil moisture (SMAP (https://smap.jpl.nasa.gov/)/SMOS (http://www.esa.int/Applications/Observing_the_Earth/SMOS)), sediment thickness, and other ET data sets (ECOSTRESS (https://ecostress.jpl.nasa.gov/)). Furthermore, we are considering applying deep learning based models to potentially improve our predictions.

ACKNOWLEDGMENT

We would like to acknowledge all the open-source software and data communities for making their resources publicly available. Particularly, we would like to thank NASA/JPL, USGS, USDA, the PRISM group, and the ADWR for providing all the necessary remote sensing and in-situ data sets.

Moreover we are extremely grateful to the AGU Fall Meeting 2020 scientific committee for accepting this work. The first author would also like to thank the AGU Virtual Student Travel Grant committee for awarding him a travel grant and enabling him to present this poster.

Finally, we are grateful to our colleagues and families for their continuous motivation and support. Currently, a journal paper based on this work is in progress. The entire source code used for implementing the proposed workflow will be made publicly available at the time of paper submission.

DISCLOSURES

We confirm that there is no conflict of interest among the authors of this manuscript

AUTHOR INFORMATION

Sayantan Majumdar is currently working as a Graduate Teaching Assistant at the Geosciences and Geological and Petroleum Engineering (GGPE (https://ggpe.mst.edu/)) department, Missouri University of Science and Technology, USA. His PhD research under the supervision of Dr. Ryan Smith (https://people.mst.edu/faculty/smithryang/) is focused on groundwater withdrawal estimation using integrated remote sensing datasets (satellite, airborne, modeled) and machine learning (deep).

Research Group: https://sites.google.com/mst.edu/hydro/ (https://sites.google.com/mst.edu/hydro/)

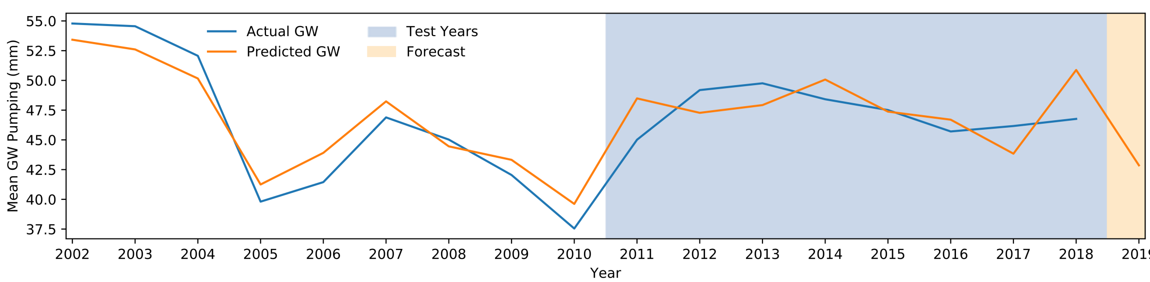
Contact information: s.majumdar@mst.edu, https://www.linkedin.com/in/sayantanmajumdar/ (https://www.linkedin.com/in/sayantanmajumdar/), https://github.com/montimaj

ABSTRACT

Groundwater is the largest source of Earth's liquid freshwater and plays a critical role in global food security. With the rising global demand for drinking water and increased agricultural production, overuse of groundwater resources is a major concern. Because groundwater withdrawals are not monitored in most regions with the highest use, methods are needed to monitor withdrawals at a scale suitable for implementing sustainable management practices. In this study, we combine publicly available datasets into a machine learning framework for estimating groundwater withdrawals over the state of Arizona. This extends a previous study in which we estimated groundwater withdrawals in Kansas, where the climatic conditions and aquifer characteristics are significantly different.

Datasets used in our model include energy-balance (SSEBop) and crop coefficient evapotranspiration estimates, precipitation (PRISM), and land-use (USDA-NASS Cropland Data Layer), and a watershed stress metric. Random forests, a widely popular machine learning algorithm, are employed for predicting groundwater withdrawals from 2002-2018 at 5 km spatial resolution. We used in-situ groundwater withdrawals available over the Arizona Active Management Area (AMA) and Irrigation Non-Expansion Area (INA) from 2002-2010 for training and 2011-2018 for validating the model respectively. The results show high training ($R^2 \approx 0.98$) and good testing ($R^2 \approx 0.82$) scores with low normalized mean absolute error ≈ 0.42 and root mean square error ≈ 1.29 for the AMA/INA region. Using this method, we are able to spatially extend estimates of groundwater withdrawals to the whole state of Arizona.

We also observed that land subsidence in Arizona is predominantly occurring in areas having high yearly groundwater withdrawals of at least 100 mm per unit area. Our model shows promising results in sub-humid and semi-arid (Kansas) and arid regions (Arizona), which proves the robustness and extensibility of our integrated approach combining remote sensing and machine learning into a holistic, automated, and fully-reproducible workflow. The success of this method indicates that it could be extended to areas with more limited groundwater withdrawal data under different climatic conditions and aquifer properties.



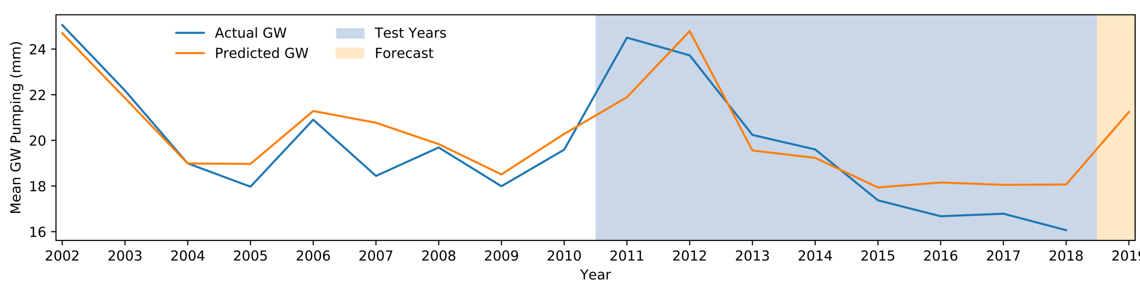


Figure 1. Mean Groundwater (GW) withdrawals for (a) the AMA/INA region in Arizona and (b) the entire state of Kansas.

REFERENCES

Majumdar, S., Smith, R., Butler, J. J., & Lakshmi, V. (2020). Groundwater Withdrawal Prediction Using Integrated Multi-Temporal Remote Sensing Datasets and Machine Learning. Water Resources Research, 56. https://doi.org/10.1029/2020WR028059 (https://doi.org/10.1029/2020WR028059)

Smith, R., & Majumdar, S. (2020). Groundwater Storage Loss Associated with Land Subsidence in Western US Mapped Using Machine Learning. Water Resources Research. https://doi.org/10.1029/2019WR026621 (https://doi.org/10.1029/2019WR026621)