Using AI to Identify Optimal Drilling Locations for Sustainable Irrigation for Subsistence Agriculture

Wanru Li, [1] Kathryn B. Laskey, [1] Mekuanent Muluneh, [2] Rupert Douglas-Bate, [3] Hemant Purohit, [1] Paul Houser [1]

George Mason University, United States;[1] Arba Minch University, Ethiopia;[2] Global MapAid, United Kingdom;[3]

[wli15, klaskey, hpurohit, phouser] @gmu.edu,[1] mulunehmekuanent@gmail.com,[2] rupertdouglas@gmail.com,[3]

Abstract

In East Africa, many drought events have occurred over the past few decades. Droughts have resulted in severe food crises, especially for countries relying heavily on agriculture. From the perspective of sustainability, utilizing groundwater for crop irrigation could be an avenue toward resilience to drought. In this study, we aim to use AI to identify optimal drilling locations for sustainable irrigation for subsistence agriculture. Our initial focus is the Hare watershed in southern Ethiopia. To identify suitable drilling locations, a hydrogeological model (TOPMODEL) for estimation of discharge and depth to water table will be implemented first; machine learning models will be constructed to estimate the probability of finding groundwater at a particular location; and finally these will be provided as inputs to an optimization model. Since this study is in progress, preliminary intermediate results are presented in this paper. A topographic wetness index (TWI) map was developed. TWI captures topographic features related to groundwater potential and will be an important input to our drilling location model.

Keywords

AI; Groundwater potential; Topographic wetness index; TOPMODEL; Machine learning; Optimal drilling locations

Introduction

In Ethiopia, small farmers comprise 95% of all farmers, and about 80% of the population (Douglas-Bate et al. 2019. Thus, the population is heavily reliant on agriculture. A decreasing of water supply has affected the yield of crops and increased vulnerability to hunger. In 2017, about 20.6% of Ethiopians suffer from hunger (Ethiopia Hunger Statistics, n.d). At present in 2020, Ethiopia has faced a large outbreak of desert grasshoppers which results in loss of food and income (ActionAid UK. 2017). To mitigate the impact of food shortages, drilling wells for irrigation could support sustainability for subsistence agriculture.

Understanding the factors that affect groundwater availability is important for estimating the probability of finding water at a location. Previous research has shown that lithology, geological structures, drainage density, soils, lineament density, geomorphology, slope and land cover land use are the main factors that have an impact on the occurrence and movement of groundwater in an area (Jaiswal et al. 2003, Greenbaum 1985, Jha et al. 2010, Andualem and Demeke 2019). A recent groundwater drought study has shown that the evapotranspiration rate, precipitation, and soil moisture are significant factors affecting groundwater drought propagation (Han et al. 2019).

To estimate groundwater potential, we used TOPMODEL (a TOPography based hydrological MODEL) proposed by Beven and Kirkby (1979). TOPMODEL simulates hydrological processes and has been used in a variety of applications. The topographic wetness index (TWI), one of the TOPMODEL outputs, uses elevation data to estimate places where water tends to accumulate. Moreover, previous studies have shown that TOPMODEL has successfully predicted streamflow (Ambroise et al. 1996, Ibbitt and Woods 2004, Nourani et al. 2011, Andualem and Demeke 2019).

Optimization approaches have been widely applied in optimal well placement problems. Ma et al. (2018) developed a mixed-integer linear programming model to identify the optimal layout of wells with minimizing the total irrigation costs in an oasis area in Northwest China. A nonlinear programming model has been constructed by Liu et al. (2019) to find the optimal well layouts with minimized pumping costs in another oasis area in Northwest China. Yin et al. (2020) focused on developing a nonlinear multi-objective model to explore optimal freshwater pumping strategies and optimal pumping locations. The multi-objective setup ensures groundwater sustainability. However, these well placement studies do not include uncertainty in the optimization models. Researchers in a previous study modeled the optimization problem using an infinite aquifer assumption, that is, it is assumed there are no constraints on the amount of water that can be pumped out from the wells (Ma et al. 2018). In fact, this is a very strong assumption which may

AAAI Fall 2020 Symposium on AI for Social Good.

not be appropriate for areas that are facing severe water scarcity. The optimization model of this study will enable decision making under uncertainty, incorporate a sustainable irrigation objective, and will relax the infinite aquifer assumption. The overall objective of this study is to identify optimal drilling locations for sustainable irrigation. To achieve this objective, we first model the hydrogeological processes by implementing TOPMODEL to estimate the discharge and depth to water table. The outputs of TOPMODEL will be incorporated into machine learning algorithms to estimate the probability of finding water at a particular location, which will then be used as input in an optimization model for identifying optimal drilling locations. We will demonstrate our model with a prototype in the Hare watershed in Ethiopia.

Applying machine learning requires data on wells. Data availability will be a challenging problem. A recent study has addressed challenges in collecting groundwater data (Lall et al. 2020). The researchers compared the number of well data points that were used in two studies. One study examined global water table depths based on 1.4 million well data points in North America and hundreds of wells in Africa. The other study focused on groundwater age estimates using 6455 wells around the globe. The comparison highlights the extreme paucity of wells information in global, especially in Africa. This finding is consistent with the extreme limitations in available data on wells in the study area in the present paper.

The rest of the paper is organized as follows. Section 2 presents the research questions for this study; Section 3 describes the study area and data; Section 4 discusses the methodologies used in this study; Section 5 shows preliminary results and discussion; Section 6 concludes the paper.

Research Questions

This study reports on research questions related to groundwater recharge, groundwater potential, probability of finding shallow groundwater, and optimal drilling locations. Although this paper focuses on the Hare region, the questions can be generalized to other agricultural regions with similar characteristics such as dry seasons and irrigation difficulties.

Questions Related to Groundwater Recharge

- What is the groundwater recharge within the study area?

Questions Related to Groundwater Potential

- What is the depth to the water table for different sites in the study area?
- What factors are considered to generate a groundwater potential map?

Questions Related to Probability of Finding Shallow Groundwater

- What are the factors that affect shallow groundwater availability?
- What is the estimated probability of drilling out water at a specific location?

Questions Related to Optimal Drilling Locations

- Without depleting the water table, how many wells can be drilled to help satisfy the irrigation need?
- Where are optimal drilling locations that could yield water with an acceptable distance to the crop fields?
- What are the optimal distances between wells?

Description of Study Area and Data

Study Area

The Hare region is located near the Abaya Lake in southern Ethiopia with latitude 6˚1′ to 6˚17′ N and longitude 37˚27′ to 37˚36′ E. The total area is 195.43 km^2. Elevation of the region ranges from 1161 m to 3465 m.

Data Collection and Preparation

In this study, observed daily discharge data for the period 1987 to 2006 was collected from the Ministry of Water, Irrigation and Energy of Ethiopia. Units were converted from cubic meters to cubic millimeters. To be consistent with the meteorological data, the plan was to collect data from 1987 to 2016. However, discharge data could be obtained only from 1987 to 2006.

Digital Elevation Model (DEM) data were obtained from Alaska satellite services with a resolution of 25 m by 25 m. For convenience of data analysis, an elevation matrix was created representing a digital elevation model with equally sized pixels and equal NS and EW resolution.

Meteorological data including precipitation and temperature was collected to estimate streamflow and groundwater recharge. Daily precipitation data was retrieved from three different meteorological stations including Arba Minch, Chencha, and Dorze stations for the period 1987 to 2016. Since precipitation data in 1987 is missing for the Dorze station and temperature data from 2006 to 2016 are missing in both Dorze and Chencha stations, filling in the missing values is necessary. All the missing data was downloaded from NASA Power Single Point Data Access (power.larc.nasa.gov, n.d.) Since we have multiple precipitation and temperature measurements, measurements from the three stations were integrated. The Thiessen Polygon approach (Rhynsburger 1973) was used to determine the average precipitation and average temperature in Hare. The basic concept of this approach is be summarized as follows. First, we divide the watershed into three polygons (Figure 1) namely Arba Minch (34.47 km^2), Chencha (64.79 km^2) and Dorze (96.17 km^2). Each contains a measurement point

(Figure 1). The coordinates for the measurement points are shown in Table 1. Second, we take a weighted average of the measurements based on the size of each polygon. The formula is:

$$\bar{P} = \frac{\sum_i^n P_i A_i}{\sum_i^n A_i}$$

where $\bar{P}$ is the weighted average; P_i is the measurement at polygon i; A_i is the area of polygon i; n is the total number of measurement points. After performing the above steps, we have the finalized weighted average precipitation and temperature data. To be consistent with the other data, the period 1987 to 2006 was used.

Table 1: Coordinates for each meteorological station near Hare

Station Name	X_UTM (m)	Y_UTM (m)	Longitude (degree)	Latitude (degree)
Arba Minch	339823.781	666130.500	37.553	6.025
Chencha	342243.250	691186.313	37.574	6.251
Dorze	341939.290	683857.503	37.571	6.185

Potential evapotranspiration (ETp) is an important input for estimating groundwater recharge. Global ETp data was downloaded from the NASA FLDAS site. The ETP data units and range are the same as the observed discharge and precipitation: millimeters per day from Jan 01, 1987 to Dec 31, 2006.

Soil parameters including texture, moisture, porosity, and hydraulic conductivity are related to the groundwater recharge potential. Previous soil data are not satisfactory, and an intensive field campaign over 195 square kilometers will be required to collect the soil parameters. This is a time-consuming and costly project.

Existing local well information is urgently needed for this study. Such information includes whether the well is working (dry or not), what type the well is (hand dug, borehole or deep wells), how much water the well yields, and what is the depth of the well (depth-to-water table). With support from the Czech Geological Survey, we have obtained locations of only four existing wells in the Hare watershed in Ethiopia. Information other than the location these wells is unknown. Since local well data are not available online or from any local organizations, field work is required to collect the data we need. This is another labor-intensive and costly task.

As the project progresses, other geology parameters, such as lithology, geological structures, drainage density, lineament density, and land use land cover, will be collected and processed.

Methodology

Hydrogeology Model

In this study, TOPMODEL was used to estimate discharge and depth to water table. The inputs to TOPMODEL include the topographic wetness index computed from the digital elevation data, a delay function derived by DEM and outlet data, a set of parameters that need to be calibrated (Table 2), and hydrometeorological and geological variables including precipitation data, potential evapotranspiration, and observed discharge.

Figure 1. The divided watershed by Thiessen Polygon approach

Table 2: Parameter set for TOPMODEL (Buytaert, 2011)

Parameter	Description [Possible unit]
Q_{s0}	Initial subsurface flow per unit area [m]
T_0	Transmissivity of the soil profile at full saturation [m^2/h]
lnTe	Log of the areal average of T_0 [m^2/h]
m	Model parameter controlling the rate of decline of transmissivity in the soil profile
$S_{r\,0}$	Initial root zone storage deficit [m]
$S_{r\,max}$	Maximum root zone storage deficit [m]
t_d	Unsaturated zone time delay per unit storage deficit [h/m]
V_{ch}	Channel flow outside the catchment [m/h]
V_r	Channel flow inside catchment [m/h]
K_0	Surface hydraulic conductivity [m/h]
CD	Capillary drive [m]
dt	The timestep [h]

The optimal parameters are chosen by matching as closely as possible the simulated discharge from TOPMODEL to observed discharge in the training period. To do this, input parameters including $m, T_0, S_{r\,max}$ are adjusted to obtain the best match between model results and training data. After calibration, model validation is performed on the validation data set to evaluate the goodness of the calibrated parameters. The calibration metric is the Nash-Sutcliffe efficiency criterion. Values close to 1 indicate a good fit; a value of 1 indicates a perfect match (Nash and Sutcliffe 1970). The formula for Nash-Sutcliffe efficiency is:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(Q_{obs} - Q_{sim})^2}{\sum_{i=1}^{N}(Q_{obs} - \bar{Q}_{obs})^2}$$

where Q_{obs} is the observed discharge; Q_{sim} is the simulated discharge; $\bar{Q}_{obs}$ is the mean of the observed discharge; and N is the total number of time steps.

The depth to water table is simulated based on the saturation deficit, which is simulated using TOPMODEL. To evaluate the goodness of the simulation result of the depth to water table, information on the depth of the existing wells should be collected, which requires field work.

Machine Learning Algorithms

As mentioned in previous section, to find optimal drilling locations, we need to make predictions on the probability of drilling water out of a well in Hare region. This would be an input parameter for the optimization model. We divide the Hare region into small pixels with equal area. A machine learning model, such as logistic regression, could be constructed to predict the probability of water availability for each pixel in Hare. The binary dependent variable is whether the well at the location yield water. The independent variables may include precipitation, elevation, ETP, land cover land use, soil texture, and percentage of topsoil moisture, the data will be collected along with the dependent variable by launching a field work.

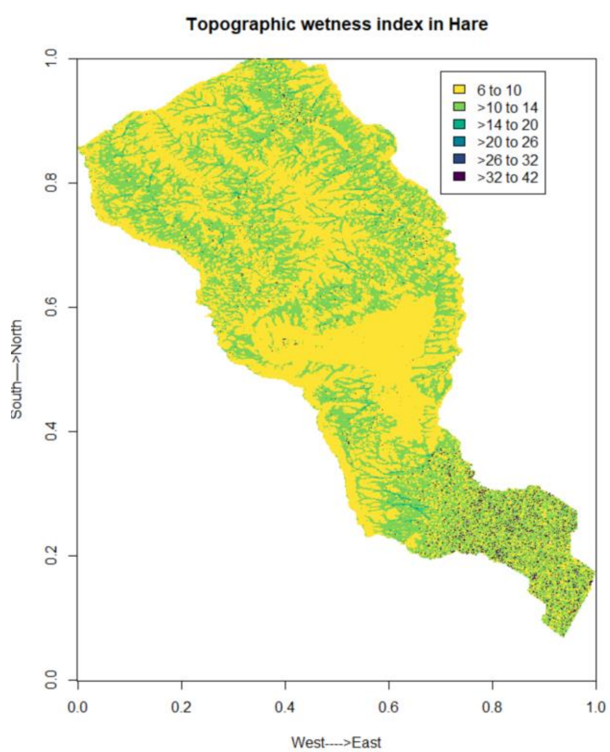

Figure 2. Topographic Wetness Index Map for Hare, Ethiopia

Optimization Approaches

To find the optimal drilling locations, a two-stage stochastic mixed integer programming (SMIP) problem could be formulated. The two-stage SMIP approach allows users to make decisions under uncertainty with two decision variables, one in the first stage and the other in the second stage (Küçükyavuz 2017). In this study, we plan to formulate our problem with binary first stage and continuous second stage variables. The objective functions for the two stages should be defined with respect to the two decision variables. Uncertainty only exists in the second stage.

The general idea of the two-stage SMIP optimization will be described from the initial formulation including the objective functions, decision variables for each stage, uncertainty and possible constraints, reformulation of the problem, and how to solve the problem.

For the initial formulation, the first stage objective function could be minimizing the total construction cost with decision variable x_i denoting whether there is a well ($x_i = 0\ or\ 1$) at location i. The second stage objective function could be minimizing the pumping cost with decision variable x_i denoting the pumping hours. Uncertainty could be the yield of water which has a distribution that should be determined prior the optimization model. A set of constraints (e.g. restriction on the pumping hours and total amount of water withdrawn) will be added to fulfill the groundwater sustainability considerations. Reformulation will be generated based on the initial formulation to make the problem tractable. Gurobi, an optimization solver, will be used to solve this optimization problem.

Preliminary Results and Discussion

Groundwater Potential

The topographic wetness index map of Hare Ethiopia derived from digital elevation data shows the potential for where water may tend to accumulate (Figure 2). Areas with higher values of topographic index indicate large contributing areas and low slopes. Higher topographic indices (darker green to purple) are mainly found in the southern part of the watershed, and a little in the central and northern parts. These regions have greater potential to become saturated with rainfall. Higher TWI values are found in the areas with surface water, such as streams and wetlands. Lower TWI values indicate the area has small contributing areas and high slope. In our study, lower TWI values (yellow) are found in the central and northern parts of the watershed. Since lower TWI indicates lower moisture storage in the soils, there may be little accumulation in many parts of the Hare watershed. As such, it could be challenging to find shallow good drilling locations for drawing groundwater.

Optimal Drilling Locations

Before finalizing the formulation of the optimization problem, we need to estimate the parameters for an initial formulation from the collected data in our study area. As mentioned, data collection is the most challenging task in this study. If the collection for some data items requires too much effort and cost to be practical, the formulation would be modified to adjust. The research questions related to optimal drilling locations would be answered after completing the data collection, parameter estimation and optimization.

Conclusion

This study focuses on using AI to identify optimal drilling locations for sustainable irrigation for subsistence farmers in Hare Ethiopia. We have found that collecting hydrogeological data has become the main challenge to develop an AI model. After data items are collected, we will first construct the TOPMODEL to estimate discharge and depth to water table, which will be used as inputs in machine learning models for an estimation of the probability of finding water at a particular location. With the probabilities as input, an optimization model for identifying optimal drilling locations for sustainable irrigation for subsistence agriculture will be constructed. Our preliminary intermediate result is the topographic wetness index map of Hare Ethiopia. The TWI map indicates that southern part of the watershed has greater potential to accumulate water; central and northern parts of the watershed show lower moisture storage in soils, which make it challenging to identify shallow groundwater. As the study moves forward, more results will be provided.

Acknowledgements

This research is performed by the MODL (Modeling Optimal Drilling Locations) team, which comprises researchers from the George Mason University Center for Resilient and Sustainable Communities (C-RASC), the Arba Minch University Water Technology Institute, and Global MapAid, with support from Czech Geological Survey. Wanru Li gratefully acknowledges support from a C-RASC fellowship for her efforts on this project.

References

ActionAid UK. 2017. Food crisis in East Africa 2017-2019. Available at: https://www.actionaid.org.uk/about-us/what-we-do/emergencies-disasters-humanitarian-response/east-africa-crisis-facts-and-figures.

Ambroise, B., Beven, K. and Freer, J., 1996. Toward a gen-eralization of the TOPMODEL concepts: Topographic indices of hydrological similarity. Water Resources Research, 32(7), pp.2135-2145.

Andualem, T.G. and Demeke, G.G., 2019. Groundwater potential assessment using GIS and remote sensing: A case study of Guna tana landscape, upper blue Nile Basin, Ethiopia. Journal of Hydrology: Regional Studies, 24, p.100610.

Beven, K.J. and Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrolog-ical Sciences Journal, 24(1), pp.43-69.

Buytaert, W., 2011. topmodel: Implementation of the Hy-drological Model TOPMODEL in R. Global Change Biology, pp.679-706.

Douglas-Bate, R., Pascual, P., Prakash, H., Kemal, A., and Mohammed, K., 2019. AI scoping mission, Ethiopia, to enhance sustainable irrigation for food supply. Paper presented at Asso-ciation for Advancement of Artificial Intelligence.

Ethiopia Hunger Statistics, n.d., Available at: https://www.macrotrends.net/countries/ETH/ethiopia/hungerstatistics#:~:text=Ethiopia%20hunger%20statistics%20for%202017,a%202.2%25%20decline%20from%202013

Greenbaum, D., 1985. Review of remote sensing applications to groundwater exploration in basement and regolith.

Han, Z., Huang, S., Huang, Q., Leng, G., Wang, H., Bai, Q., Zhao, J., Ma, L., Wang, L. and Du, M., 2019. Propagation dynamics from meteorological to groundwater drought and their possible influence factors. Journal of Hydrology, 578, p.124102.

Ibbitt, R. and Woods, R., 2004. Re-scaling the topographic index to improve the representation of physical processes in catchment models. Journal of Hydrology, 293(1-4), pp.205-218.

Jaiswal, R.K., Mukherjee, S., Krishnamurthy, J. and Saxena, R., 2003. Role of remote sensing and GIS techniques for genera-tion of groundwater prospect zones towards rural development--an approach. International Journal of Remote Sensing, 24(5), pp.993-1008.

Jha, M.K., Chowdary, V.M. and Chowdhury, A., 2010. Groundwater assessment in Salboni Block, West Bengal (India) using remote sensing, geographical information system and multi-criteria decision analysis techniques. *Hydrogeology journal*, 18(7), pp.1713-1728.

Küçükyavuz, S. and Sen, S., 2017. An introduction to two-stage stochastic mixed-integer programming. In *Leading Developments from INFORMS Communities* (pp. 1-27). INFORMS.

Lall, U., Josset, L. and Russo, T., 2020. A Snapshot of the World's Groundwater Challenges. *Annual Review of Environ-ment and Resources*, 45.

Ma, T., Wang, J., Liu, Y., Sun, H., Gui, D. and Xue, J., 2019. A Mixed Integer Linear Programming Method for Opti-mizing Layout of Irrigated Pumping Well in Oasis. *Water*, 11(6), p.1185.

Nash, J.E. and Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology*, 10(3), pp.282-290.

Nourani, V., Roughani, A. and Gebremichael, M., 2011. TOPMODEL capability for rainfall-runoff modeling of the Ammameh watershed at different time scales using different terrain algorithms. *Journal of Urban and Environmental Engi-neering*, 5(1), pp.1-14.

power.larc.nasa.gov, n.d. NASA POWER Data Access Viewer. Available at: https://power.larc.nasa.gov/data-access-viewer/.

Rhynsburger, D., 1973. Analytic delineation of Thiessen polygons. *Geographical Analysis*, 5(2), pp.133-144.

Yin, J., Pham, H.V. and Tsai, F.T.C., 2020. Multiobjective Spatial Pumping Optimization for Groundwater Management in a Multiaquifer System. *Journal of Water Resources Planning and Management*, 146(4), p.04020013.