


A data-driven approach to assessing the impact of water harvesting structures on regional water storage in East Africa

J. A. Eisma  and V. Merwade

ABSTRACT

A small-scale water harvesting structure known as a sand dam has gained popularity across East Africa due to the efforts of non-governmental organizations. A sand dam is a subsurface water reservoir that stores water between sand grains. Stored thus, the water is filtered and protected from evaporation. This study uses remotely sensed data to investigate the impact of these structures on water storage and vegetative growth. The relationship between sand dams and water storage was modeled using a binary sand dam factor, climate data from the Famine Early Warning Systems Network Land Data Assimilation System (FLDAS), and water storage data measured by the Gravity Recovery and Climate Experiment (GRACE) twin satellites. The analysis revealed that GRACE largely fails to detect a statistically significant impact of sand dams on regional water storage. However, analysis of the Normalized Difference Vegetation Index (NDVI) indicated that sand dams have a significant impact on regional vegetation. Vegetative growth is correlated with groundwater levels, indicating that sand dams have a positive impact on water storage albeit on a smaller scale than GRACE can regularly detect. Significantly, this study shows that NDVI data can be used effectively to study small-scale, regional changes in vegetation and water storage.

Key words | generalized linear model, principal component analysis, remote sensing, sand dam, spatial autocorrelation

J. A. Eisma  (corresponding author)
Department of Civil Engineering,
University of Texas at Arlington,
Arlington, TX,
USA
E-mail: jessica.eisma@uta.edu

V. Merwade
Lyles School of Civil Engineering,
Purdue University,
West Lafayette, IN,
USA

HIGHLIGHTS

- The impact of small-scale water harvesting structures on regional water resources is explored through spatial generalized linear models of remote sensing products.
- Sand dams increase regional vegetation, which suggests groundwater levels are also positively impacted.
- Sand dams may have a smaller impact on groundwater levels than believed.
- NDVI and, to a lesser extent, GRACE record the impact of sand dams.

INTRODUCTION

One of the world's grand challenges, addressing the problem of water insecurity, is particularly problematic in developing

nations where resources and infrastructure are limited. This is especially true in poorly connected rural communities, where women and children might spend 3–4 h a day collecting water (Henry *et al.* 2015). A single solution to water security in the developing world does not exist, due to variability in climate, hydrology, and geology. Each solution must

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

doi: 10.2166/hydro.2021.115

be adapted to the unique requirements of the land and local community. Such solutions are johads, or small earthen dams, in India (Gupta 2011), fog-harvesting meshes in Peru and Chile (Qadir *et al.* 2018), and sand dams in sub-Saharan Africa (Borst & de Haas 2006). Despite the wide-ranging implementation of small-scale solutions to water security, very few studies have examined their effectiveness and ability to increase water availability.

Effective small-scale solutions to water security are essential to the maintenance of rural communities in developing countries. Rural communities are particularly vulnerable to climate change, and current climate projections indicate that rainfall will either decrease or be concentrated in fewer, higher intensity events in the developing world (Reyer *et al.* 2017). Small-scale water harvesting technologies are increasingly recognized as a viable solution to rural water security and climate change. However, most efforts to spread or intensify water harvesting technologies have limited success (Bouma *et al.* 2016), despite the great deal of time and money spent bringing water harvesting projects to fruition. For example, one sand dam constructed with volunteer labor takes 3 weeks to build and costs an average of 12,000 USD (Lasage *et al.* 2008; Lasage & Verburg 2015). Despite this infusion of resources, approximately 50% of sand are not functioning properly (Viducich 2015; de Trincheria *et al.* 2018). The widespread issues with sand dam effectiveness are a relatively new revelation (de Trincheria *et al.* 2015, 2018; Viducich 2015; Eisma & Merwade 2020), and similar issues may pervade other well-known water harvesting technologies.

Few studies have assessed the effectiveness of small-scale water harvesting technologies in rural communities. Thus far, evaluation of water harvesting has primarily been performed via simulation (Hut *et al.* 2008; Quilis *et al.* 2009; Lasage *et al.* 2015), field studies and surveys (Ngigi *et al.* 2005; Previati *et al.* 2010; Quinn *et al.* 2019; Eisma & Merwade 2020), meta-analysis of the existing literature (Lasage & Verburg 2015; Bouma *et al.* 2016), or multi-criteria analysis of potential projects (Jaber & Mohsen 2001; Garfi *et al.* 2011). While such techniques can offer valuable information regarding the effectiveness of specific water harvesting projects, their ability to provide a broad perspective is limited. Conversely, an assessment methodology based on remotely sensed data has the potential to provide insight unhindered by site accessibility, cost, and selection

bias. Such a methodology must match a quantifiable expected impact of water harvesting structures with a remote sensing product that records that same impact. For example, sand dams supposedly raise the groundwater table (Borst & de Haas 2006; Excellent Development 2019). If this claim is true, a high density of sand dams should have a measurable impact on regional groundwater levels.

No studies assess the impact of water harvesting structures on regional water availability. This study aims to develop a methodology to assess the ability of water harvesting structures to improve regional water availability. The methodology is developed and applied to study whether sand dams in south-eastern Kenya are functioning as reported. Considering that most community-based water harvesting structures exist in developing nations with limited, publicly available local data, the methodology in this study is developed using globally available remotely sensed data to enable broader applicability in other regions. While this study uses sand dams in Kenya as a test case, the methodology presented here can be applied globally to assess the impact of local water harvesting schemes on regional water resources. For example, sand dams function as a type of managed aquifer recharge, and this methodology could be applied to study the effectiveness of various managed aquifer recharge schemes (Standen *et al.* 2020). Coastal urban centers and reclaimed islands have recently begun using sandy deposits to develop fresh groundwater lenses in a process similar to that of sand dams (Yao *et al.* 2019), which may also be examined by the methodology outlined here.

Sand dams

Sand dams are small dams built atop impermeable riverbeds in arid regions with infrequent, high-intensity rainfall (see Figure 1). The high-intensity rain washes sand overland where it builds up behind the dam. After a few rainy seasons, the sand begins storing water, where it is naturally filtered, protected from evaporation, and helps raise the groundwater level due to recharge from increased subsurface storage (Borst & de Haas 2006; Hut *et al.* 2008; Quilis *et al.* 2009).

Intensified sand dam construction began in Kenya in 1979, and estimates today suggest Kenya is home to over 3,000 sand dams (Viducich 2015; de Trincheria *et al.* 2018). Myriad NGOs have largely constructed sand dams with private funds and volunteer labor from local communities. The

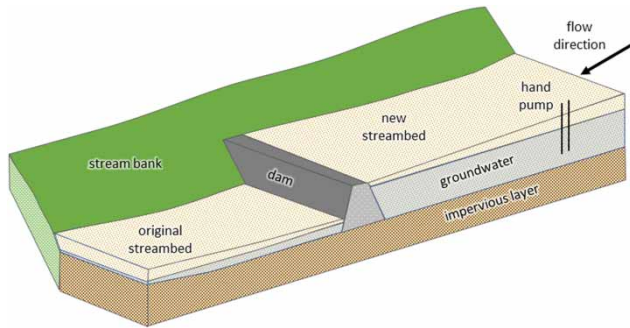


Figure 1 | Schematic of a sand dam (adapted from Borst & de Haas (2006)).

de-centralized sand dam construction effort has led to a dearth of publicly available information, with few NGOs providing details of their sand dam projects. Most sand dam information comes from a handful of researchers investigating one or two ‘ideal’ sand dams (Borst & de Haas 2006; Aerts *et al.* 2007; Hut *et al.* 2008; Quilis *et al.* 2009). Researchers have not yet examined whether sand dams are impacting the environment at a regional scale.

Most of Kenya’s 3,000+ sand dams are in three counties: Kitui, Makueni, and Machakos (de Trincheria *et al.* 2015), hereafter referred to as the sand dam counties (SDCs). This study hypothesizes that the additional water storage provided by the high density of sand dams in the SDC leads to a higher regional groundwater table and increased vegetative cover. One study finds that a sand dam can influence the groundwater level up to 350 m upstream and downstream of the dam (Quilis *et al.* 2009); however, a sand dam’s area of influence is also reported to be up to 2 km upstream of the dam and 500 m to each side of the stream (Ryan & Elsner 2016). Given the 3,000+ sand dams in the SDC, anywhere from 3.4 to 17.2% of the surface area in the SDC may be impacted. The sand dam’s large area of influence should be detectable via remote sensing.

STUDY AREA AND DATA

Study area

The SDC in south-eastern Kenya (see Figure 2) cover 34,670 km². This study compares water storage changes in the SDC to those in a buffer area of nearly equal size (34,537 km²). Annual rainfall within the SDC is

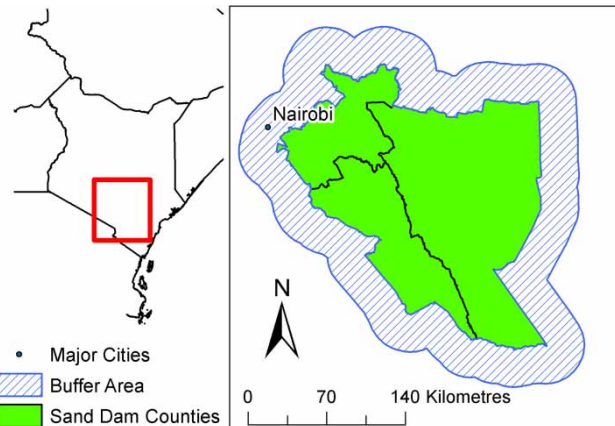


Figure 2 | Study area in Kenya.

approximately 700 mm compared to 750 mm in the buffer area (Fick & Hijmans 2017). Elevation in the SDC ranges from 246 to 1,803 m above sea level with a mean elevation of 830 m (Fischer *et al.* 2008). Similarly, elevation in the buffer area ranges from 180 to 1,991 m with an average of 853 m (Fischer *et al.* 2008). Most SDC soil (92.4%) is well-drained sandy clay loam (Ulsaker & Kilewe 1983; Hengl *et al.* 2017). The buffer area is also primarily sandy clay loam but has 27.0% sandy loam (Hengl *et al.* 2017). The SDC and buffer areas are mainly covered with natural and semi-natural terrestrial vegetation, 77.0 and 72.8%, respectively. Both areas are 21.1% cultivated terrestrial (FAO 2002).

Data

Gravity Recovery and Climate Experiment

Gravity Recovery and Climate Experiment (GRACE) measures Earth’s gravity field via twin satellites (Tapley *et al.* 2004). Changes in water storage can be derived from changes in the Earth’s gravity field, because most gravity field fluctuations result from shifting water volumes. Therefore, the GRACE data provides an estimate of the total water stored from the Earth’s core to the Earth’s surface. The satellites record the gravity field at a resolution ranging from 400 to 40,000 km every 30 days (Tapley *et al.* 2004). The mission, started in March 2002, exceeded its planned lifetime of 5 years and remained operational until October 2017. The GRACE dataset is nearly continuous with some months missing from 2011 onwards due to battery management efforts

(Herman *et al.* 2012). The raw data from GRACE is processed by three research centers: Center for Space Research (CSR) at the University of Texas at Austin, Jet Propulsion Laboratory (JPL) at California Institute of Technology, and GeoforschungsZentrum (GFZ) Potsdam in Germany (Tapley *et al.* 2004). The centers each produce monthly spherical harmonic gravity models that JPL converts to gridded data of the change in total water storage relative to a time-mean baseline from 2004 to 2009 at a resolution of 1° , or approximately 111 km. For the study area, the GRACE data has an estimated error of approximately 3.8 cm (Landerer & Swenson 2012).

GRACE data has been used successfully in many studies to explore general trends in water storage at a variety of scales: 45,000 km² (Hachborn *et al.* 2017), 55,000 km² (Henry *et al.* 2011), 766,000 km² (Huang *et al.* 2012), and 2,200,000 km² (Cao *et al.* 2015) up to the global scale. GRACE data may contain signal leakage errors, but leakage errors are assumed to be homogenous at the small scale of this study (Chinnasamy & Ganapathy 2018). A GRACE Follow-On mission began in May 2018, providing an opportunity for future studies to apply the methodology developed here with updated and improved data. This study attempts to discern the consequences of man-made structures on water storage by analyzing trends in the GRACE total water storage data.

Normalized Difference Vegetation Index

Normalized Difference Vegetation Index (NDVI) is calculated from spectral bands recorded by the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument every 1–2 days. Vegetation absorbs most of the blue (470 nm) and red (670 nm) wavelengths while reflecting most of the near-infrared radiation. The contrast between the absorbed and reflected wavelengths is manipulated to calculate a grid of NDVI values for Earth's entire surface (Didan *et al.* 2015). NDVI, ranging from negative one to one, serves as an indicator of the vegetation present on the land surface. Values close to one indicate the presence of thriving vegetation, whereas values from negative one to zero indicate the presence of bare ground or dead vegetation. MODIS NDVI grids are available in 16-day and monthly increments at resolutions ranging from 250 m to 1 km. For the purposes of this study, monthly MODIS NDVI at a resolution of 1 km is used (Didan 2015).

Famine Early Warning System Network Land Data Assimilation System

Famine Early Warning System Network Land Data Assimilation System (FLDAS) is a set of models designed to provide accurate climate estimates for the purpose of drought monitoring in data-sparse regions susceptible to food and water security issues (McNally *et al.* 2017). FLDAS currently provides daily and monthly estimates consisting of 25 different variables for sub-Saharan Africa. Research-quality FLDAS simulation C results are used for this study (Anderson *et al.* 2012; Yilmaz *et al.* 2014; McNally *et al.* 2017). Simulation C uses the Climate Hazards group Infrared Precipitation with Stations (CHIRPS) and the Modern Era Reanalysis for Research and Applications version 2 (MERRA-2) as forcing data (McNally *et al.* 2017). FLDAS provides simulation C outputs for the Noah Land Surface Model (Noah) at a resolution of 0.1° and Variable Infiltration Capacity (VIC) at a resolution of 0.25° (NASA/GSFC/HSL 2016).

METHODS

The methodology included here outlines two techniques for quantifying the impact of sand dams: one using GRACE total water storage and one using NDVI. Both techniques are based on developing a generalized linear model that seeks to define the relationship between the presence of sand dams and a magnitude of change in GRACE data or NDVI data. In both instances, covariates such as weather, time of year, and population are included to ensure that the observed changes in GRACE and NDVI can be attributed to sand dams. A significant ($p < 0.05$) coefficient for the sand dam indicator variable indicates that the sand dam does impact total water storage or vegetation.

Relating GRACE water storage to the presence of sand dams

Processing GRACE

GRACE monthly mass grids are produced by JPL from monthly GRACE spherical harmonic gravity models and

independently generated by CSR, JPL, and GFZ. The monthly mass grids have been processed and are provided relative to the time-mean baseline of the gravity field from 2004 to 2009, inclusive (Swenson & Wahr 2006; Swenson *et al.* 2008; Cheng *et al.* 2011; Geruo *et al.* 2013). The resultant product is a monthly grid of the equivalent water depth of total water storage relative to a time-mean baseline. Variations in the monthly mass grids are attenuated at small scales due to filtering that occurs during processing. To restore the lost energy, a grid of multiplicative scale factors is applied to the GRACE data. The scale factors minimize the difference between the filtered and unfiltered variations in total water storage (Landerer & Swenson 2012).

The three available GRACE solutions (CSR, JPL, and GFZ) have slightly different error structures. Rather than select one solution to analyze, the ensemble mean of the three solutions was used to minimize noise in the scatter of the solutions by 5–10 mm root mean square (Sakumura *et al.* 2014).

The GRACE solutions were obtained for months available from the beginning of the GRACE mission to the end of 2016: April 2002 to December 2016 (TELLUS 2012). During this time, GRACE data is unavailable for 21 of the potential 177 months, leaving 156 months available for this analysis. Eighteen of the 21 months of missing data occurred during battery management efforts from 2011 to 2017 (Herman *et al.* 2012).

The Level-3 GRACE data has a resolution of 1° , while the FLDAS dataset, employed as a proxy for climate data in this research, has a resolution of 0.1° . The GRACE data was resampled using bilinear interpolation to a resolution of 0.1° , or approximately 11 km by 11 km, to allow for a one-to-one relationship between the two datasets (Cao *et al.* 2015).

Processing FLDAS for GRACE

As mentioned above, the GRACE data provides total water storage values relative to a time-mean baseline from 2004 to 2009, inclusive. To create a comparable FLDAS dataset, the gridded average value from 2004 to 2009 was calculated for each climate variable and subtracted from the raw data.

Validating FLDAS for the study area

The terrestrial water balance dictates that the change in water storage is equal to the difference between the inflows and outflows of the system, which can be written as:

$$\frac{dW(t)}{dt} = P(t) - R(t) - E(t), \quad (1)$$

where W is water storage, P is precipitation, R is runoff, E is evapotranspiration, and t is time. Each component of the right-hand side (RHS) of Equation (1) can then be separated into constant, linear, and time-variable sinusoidal terms (Crowley *et al.* 2006). For example, the precipitation term can be written as:

$$P(t) = P_0 + P_1 t + P^*(t) \quad (2)$$

The same is true of the runoff and evaporation terms. Re-writing Equation (1) to include the terms shown in Equation (2) and integrating produces:

$$\begin{aligned} W(t) - (P_0 - R_0 - E_0)t - \frac{1}{2}(P_1 - R_1 - E_1)t^2 + C \\ = \int_{t_0}^t [P^*(t') - R^*(t') - E^*(t')] dt', \end{aligned} \quad (3)$$

where C is the constant of integration (Crowley *et al.* 2006).

Equation (3) can be used to test the validity of substituting FLDAS data for climate data in the study area. FLDAS includes the data necessary to solve Equation (3) for total water storage, which can then be compared with the GRACE-provided total water storage in the region. FLDAS data includes evapotranspiration, storm surface runoff (Q_s), baseflow-groundwater runoff (Q_{sb}), and precipitation.

Generalized linear model for GRACE

To determine whether the 3,000+ sand dams built in the SDC have a collective impact on groundwater storage at a regional scale, monthly generalized linear models (GLMs) were developed using the GRACE data as the dependent variable. As predictor variables, the GLM includes initial water storage, sand dam presence, and a suite of climate variables via

principal components. The principal component analysis (PCA) was performed on 15 FLDAS climate variables: evapotranspiration, specific humidity, storm surface runoff, baseflow-groundwater runoff, total precipitation rate, soil moisture 0–10, 10–40, 40–100, and 100–200 cm underground, soil temperature 0–10, 10–40, 40–100, and 100–200 cm underground, near-surface air temperature, and near-surface wind speed. A sensitivity analysis conducted via stepwise regression indicated the contribution of each predictor variable to the overall adjusted R^2 (Wang *et al.* 2016).

Principal component analysis

Principal component analysis (PCA) is a statistical method employed to capture most of a multivariate dataset's variance in a few orthogonal components. Each component is a linear combination of the standardized multivariate dataset, such that a principal component takes the form:

$$PC_i = \varepsilon_{i,1}X_1^S + \varepsilon_{i,2}X_2^S + \varepsilon_{i,3}X_3^S + \dots + \varepsilon_{i,p}X_p^S, \quad (4)$$

where PC_i is the i th principal component and $\varepsilon_{i,1}$ is the coefficient for the first variable in the standardized multivariate dataset, X_1^S . The first component describes most of the dataset's variance with each successive component describing most of the variance remaining after the previous component is determined. Therefore, PCA is essentially a dimension-reduction technique utilized when the original multivariate dataset contains an inordinate number of variables. PCA is conducted via eigenanalysis, with the resulting eigenvalues representative of the amount of variance captured by each principal component. For the purposes of this study, the singular value decomposition PCA was employed. A complete description of the PCA computation can be found in Shlens (2014).

The Kaiser–Guttman criterion was used to determine the appropriate number of principal components to retain such that information-loss is minimized. The Kaiser–Guttman criterion dictates that every principal component with an eigenvalue greater than unity is retained. While this is not the most sophisticated technique available, the Kaiser–Guttman criterion is sufficient for datasets with a high communality (Yeomans & Golder 1982).

Finalized GRACE model

The monthly GLM model developed to determine whether the presence of sand dams has a statistically significant effect on total water storage is:

$$\text{GRACE}_t = \alpha + \beta \text{GRACE}_{t-1}^S + \gamma \text{PC}_1 + \delta \text{PC}_2 + \varepsilon \text{PC}_3 + \eta \text{PC}_4 + \kappa \text{SD} + \epsilon, \quad (5)$$

where GRACE_t is the $\log(\text{GRACE}_t + \text{constant})$ total water storage in month t , GRACE_{t-1} is the standardized GRACE total water storage in month $t-1$, PC_i is the i th principal component, and SD is the sand dam indicator variable. GRACE data is provided as the monthly change in total water storage, necessitating the inclusion of GRACE_{t-1} in the GLM to supply the initial condition from which the change is calculated. The superscripted S indicates that the variable is standardized to make it comparable in scale to the other predictors, here, the principal components. The SD variable was included as a binary, categorical variable, indicating whether the point is in the SDC or buffer area.

To develop a set of training and validation data for each monthly GLM, 43 or 44 points were randomly sampled from each available month (e.g. 43 or 44 points from each December for a total of 564 points). The sampled points were randomly split: 50% for a training set and 50% for validation. The coefficients of Equation (5) were determined through statistical analysis for every GLM. One GLM model was determined for every month of the year and validated via cross-validation.

Relating MODIS NDVI to the presence of sand dams

Processing MODIS NDVI and FLDAS

MODIS NDVI, hereafter NDVI, was obtained for the same period as the GRACE total water storage data, April 2002 to December 2016. Invalid NDVI data were omitted, and the data were scaled by 0.0001 (Didan *et al.* 2015). NDVI has a resolution of 1 km, so the FLDAS dataset was resampled using bilinear interpolation to a resolution of 1 km to allow for a one-to-one relationship with NDVI.

Generalized linear model for NDVI

To determine whether the 3,000+ sand dams built in the SDC have a collective impact on vegetation, a GLM was developed using NDVI data as the dependent variable. As predictor variables, the GLM includes sand dam presence, month of year, standardized population count, a suite of climate variables via principal components, and an autocovariate to account for spatial autocorrelation. Like the GRACE model, a sensitivity analysis of predictor variables will be conducted via stepwise regression.

Spatial autocorrelation

Analyzing spatial data introduces the issue of spatial autocorrelation. Spatial autocorrelation occurs, because a point is likely more similar to points nearby than distant points (Tobler 1970). This phenomenon produces residuals of statistical analyses that violate the assumption of independently and identically distributed errors, which may cause incorrect rejection of a true null hypothesis (Anselin 2002). Spatial autocorrelation can be detected via the calculation of Moran's I (Dormann *et al.* 2007). If Moran's I is statistically significant and/or if the Moran's I plot depicts a strong trend, then spatial autocorrelation may be an issue.

To address the issue of spatial autocorrelation in the NDVI residuals, an autocovariate was added to the GLM. The autocovariate is the result of a distance-weighted function that assesses how closely the response variable at one point is related to the response variable of the surrounding points (Dormann *et al.* 2007). The autocovariate, A_i , for point i is calculated as:

$$A_i = \sum_{j \in k_i} w_{ij} y_j, \quad (6)$$

where y_j is the response variable at point j , a neighbor of point i within the area k_i , and w_{ij} is the weight assigned based on a predetermined relationship between points j and i (Dormann *et al.* 2007). For this study, the weight is based on the inverse distance. Area k_i was selected via trial and error to minimize the spatial autocorrelation present in the residuals. Area k_i was determined to be 3 km for the NDVI data.

Finalized NDVI model

The GLM model developed to determine whether the presence of sand dams has a statistically significant effect on standardized vegetative growth is:

$$\text{NDVI}^S = \alpha + \beta \text{PC}_1 + \gamma \text{PC}_2 + \delta \text{PC}_3 + \varepsilon \text{PC}_4 + \eta A + \kappa_{1-12} \text{Month} + \lambda \text{Pop}^S + \mu \text{SD} + \varepsilon, \quad (7)$$

where PC_i is the i th principal component, A is the autocovariate, Month is the month of year, Pop^S is the standardized population count, and SD is the sand dam indicator variable. The coefficients were determined using a set of training data and validated using a separate dataset.

The training and validation datasets for the NDVI analysis were developed similarly to the GRACE training and validation datasets. However, the higher resolution of the NDVI data provides for a much denser network of points. 457 or 458 points were randomly sampled from each month, and the sampled points were randomly split in half into training and validation sets.

RESULTS

Sand dam impact on total water storage

GRACE water storage anomalies in the SDC and buffer area are similar. The mean ($p = 0.92$) and variance ($p = 0.49$) of the water storage anomalies in the two areas are not statistically different. However, GRACE water storage anomalies exhibit a significant difference between the SDC and buffer area stored water for individual months (see Figure 3). Overall, 60 of the 156 available months display a significant difference in the water storage anomalies of the two regions. Interestingly, a significant difference is most often observed as the region shifts from the dry season into the wet season. This may be a signature of the water retained in sand dams at the end of the dry season, but climatic differences must also be considered.

The validity of using FLDAS data as a proxy for climate data is determined by comparing FLDAS water storage anomalies calculated via Equation (3) with GRACE water storage anomalies in the SDC. The smoothed results (see

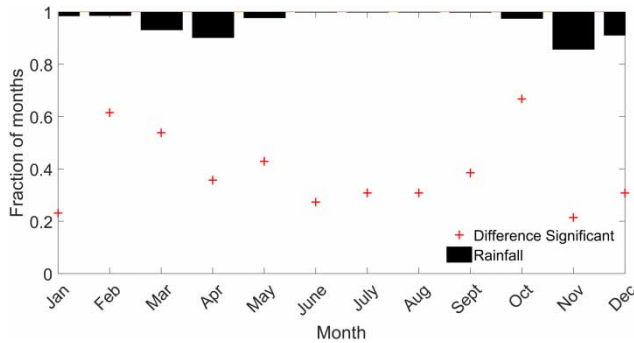


Figure 3 | Fraction of available months indicating that there is a significant difference between the water storage anomalies observed in the SDC and buffer area.

Figure 4) indicate that the two datasets provide similar information regarding the magnitude and direction of water storage anomalies in the SDC. The FLDAS water storage anomalies largely fall within or near the GRACE error estimated by Landerer & Swenson (2012). The most glaring exception occurs from 2013 to 2014, likely caused by the battery management data gaps in GRACE.

The GLM designed in Equation (5) indicates that the SDC store a significantly different amount of water than the buffer area during 3 months of the year, and that this difference can likely be attributed to the presence of sand dams (see Table 1). The sand dam indicator variable is a significant factor for the months of January ($p = 0.025$), April ($p = 0.037$), and November ($p = 0.015$) while considering the confounding factors of climate and initial water storage.

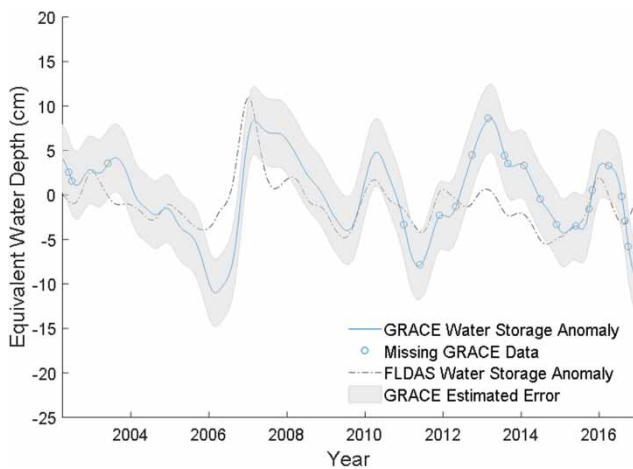


Figure 4 | GRACE water storage anomalies compared with FLDAS-calculated water storage anomalies in the SDC. Both were smoothed with a 12-month Gaussian filter.

Table 1 | Water storage anomaly GLM summary

Model	Sand dam indicator variable (SD)				Model adjusted R^2
	Coefficient	Standard error	t-statistic	p-value	
January	-0.043	0.02	-2.26	0.025*	0.68
February	-0.003	0.02	-0.18	0.858	0.43
March	0.019	0.01	1.50	0.134	0.61
April	0.046	0.02	2.10	0.037*	0.76
May	0.027	0.01	1.85	0.066	0.79
June	-0.005	0.01	-0.31	0.754	0.75
July	0.005	0.01	0.34	0.731	0.73
August	-0.020	0.01	-1.37	0.171	0.69
September	-0.001	0.01	-0.08	0.937	0.76
October	0.006	0.01	0.57	0.570	0.79
November	0.036	0.01	2.45	0.015*	0.83
December	-0.005	0.02	-0.21	0.836	0.62

*Significance at $p < 0.05$.

Interestingly, these 3 months showed relatively low fractions of significance in Figure 3, indicating that the difference between total water storage in the SDC and buffer area cannot be attributed solely to climate and initial conditions. In January, sand dams collectively decrease the amount of total water storage by 4.3%, while in April and November, sand dams increase the amount of total water storage by 4.6 and 3.6%, respectively. January, the first month of a dry season, may see lower levels of total water storage in the SDC than in the buffer area due to increased evapotranspiration potential resulting from added water storage during the preceding rainy season. This increased rate of evapotranspiration may reduce the amount of water stored in the SDC to levels below those of the buffer area. April and November, the wettest months of the two rainy seasons, see increased levels of water storage in the SDC.

The sensitivity analysis revealed that $GRACE_{t-1}^S$ had the largest impact on the monthly adjusted R^2 , contributing an average of 0.61. The remaining five predictors have a much smaller average influence on adjusted R^2 , listed from highest to lowest: PC_1 (0.045), PC_2 (0.013), PC_3 (0.013), PC_4 (0.006), and SD (0.001). While the categorical sand dam indicator variable contributes little to overall model adjusted R^2 , the variable is sometimes a significant predictor (see Table 1).

Overall, the monthly GLM models are a good fit, with adjusted R^2 values ranging from 0.43 to 0.83 and averaging 0.70. In the SDC, the inclusion of the sand dam indicator variable generally results in an improved normalized mean square error (NMSE, see Figure 5). However, the inclusion of the sand dam indicator variable results in a slightly higher NMSE for the southern and northeastern portions of the SDC. These areas likely have fewer functioning sand dams or a lower density of sand dams, reducing the water storage signal. Alternatively, the southern and northeastern regions of the SDC may not show improvement due to aliasing of the GRACE signal (Winsemius *et al.* 2006) or possibly signal attenuation during GRACE pre-processing (Landerer & Swenson 2012).

Sand dam impact on vegetation

NDVI in the SDC and buffer area is similar in magnitude, but the mean ($p = 0.001$) and variance ($p = 0.001$) of NDVI in the two areas are statistically different. Sand dams may collectively have a significant impact on

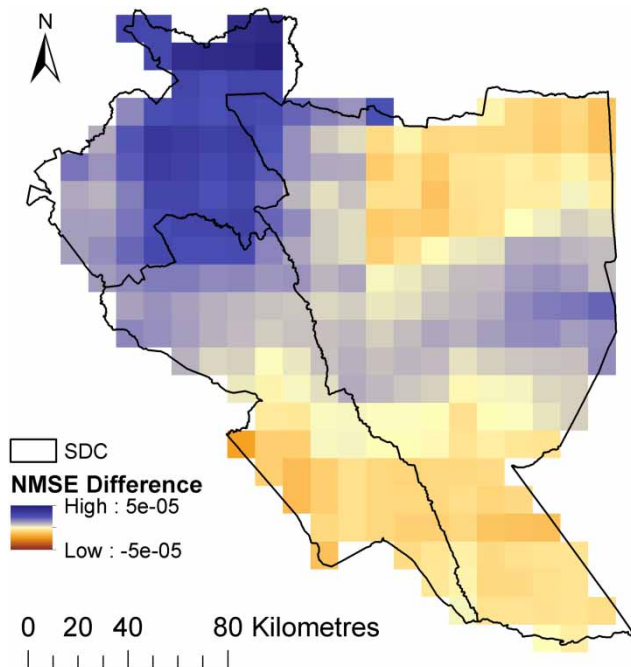


Figure 5 | Model improvement after adding the sand dam indicator variable to the November GRACE GLM. NMSE difference is the average reduction in NMSE within a five-cell range after adding the sand dam indicator variable to the model.

vegetation, but the relationship must be examined in conjunction with confounding factors.

The spatial autocorrelation in NDVI could cause issues with model validity and interpretability. The autocovariate function included in Equation (6) accounts for most of the spatial autocorrelation in the NDVI data (see Figure 6). Including an autocovariate in the model reduces Moran's I of the residuals from 0.4559 to 0.0419, more than a 90% decrease in spatial autocorrelation. A Monte Carlo simulation of Moran's I for the model with an autocovariate indicated that spatial autocorrelation is not an issue ($p = 0.01$).

The GLM designed in Equation (7) indicates that the SDC has significantly more vegetation than the buffer area, and that this difference can likely be attributed to sand dams. The sand dam indicator variable is significant ($p < 0.001$) when the GLM accounts for variations in climate, time of year, and population (see Table 2). The presence of sand dams increases NDVI by 0.07 standard deviations. The confounding factors included in the GLM also significantly impact NDVI. The amount of water present in the atmosphere, represented by PC_2 , has the greatest positive impact on NDVI; for every unit increase in PC_2 , NDVI increases by 0.19 standard deviations. The month of October has the greatest negative impact on NDVI—NDVI is 0.87 standard deviations lower in October than in January. October falls right at the end of the dry season and the beginning of the rainy season, so the vegetation is expected to be in poor condition in October.

The sensitivity analysis revealed that the Month categorical factor had the largest impact on the model adjusted R^2 , contributing 0.392, followed closely by the autocovariate, A , contributing 0.303. The remaining six predictors have a much smaller influence on adjusted R^2 , listed from highest to lowest: PC_2 (0.095), PC_1 (0.066), PC_4 (0.031), PC_3 (0.017), SD (0.001), and Pop (0.000). While the categorical sand dam indicator variable contributes little to overall model adjusted R^2 , the variable is a significant predictor (see Table 2).

The vegetation GLM simulates the standardized NDVI based on climate, time of year, population, spatial autocorrelation, and sand dam presence very well, resulting in an adjusted R^2 of 0.69. Throughout the SDC, inclusion of the sand dam indicator variable results in an improved NMSE (see Figure 7). The overall model fit is good and establishes

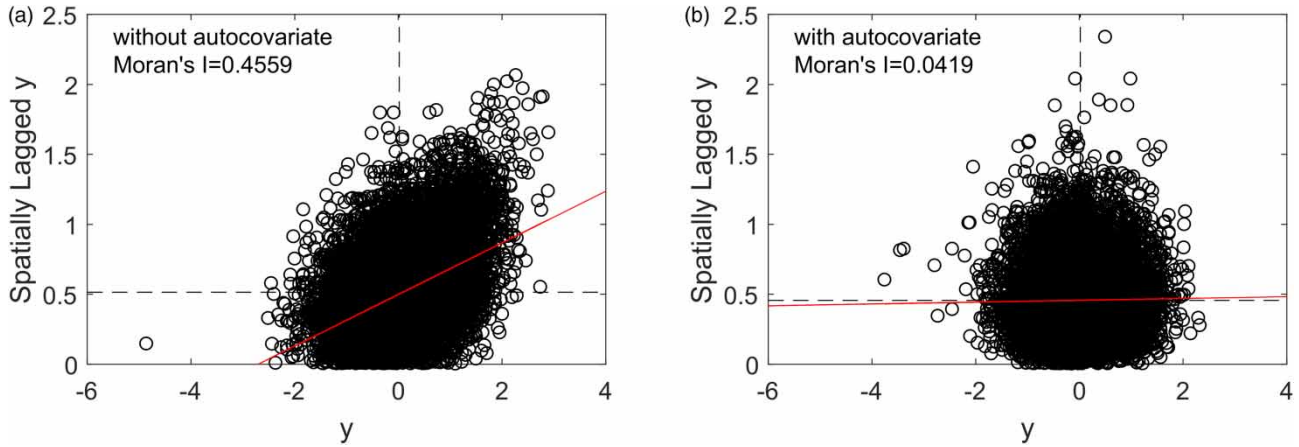


Figure 6 | Moran scatter plot for GLM validation residuals (a) without an autocovariate function in the GLM and (b) with an autocovariate function. The red line is the line of best fit through the points. Please refer to the online version of this paper to see this figure in colour: <https://doi.org/10.2166/hydro.2021.115>.

confidence that the model structure can provide valuable insight into the relationship between the collective impact of sand dams and NDVI.

Table 2 | Vegetation GLM summary

	Coefficient	Standard error	t-statistic	p-value
Intercept	0.32	0.01	30.45	<0.001*
PC ₁	-0.11	0.00	-77.66	<0.001*
PC ₂	0.19	0.00	83.64	<0.001*
PC ₃	-0.13	0.00	-41.02	<0.001*
PC ₄	-0.14	0.00	-48.55	<0.001*
A	0.08	0.00	138.13	<0.001*
January	0.00	-	-	-
February	-0.28	0.01	-20.16	<0.001*
March	-0.55	0.01	-38.94	<0.001*
April	-0.42	0.01	-28.92	<0.001*
May	-0.06	0.01	-4.08	<0.001*
June	-0.11	0.01	-7.63	<0.001*
July	-0.27	0.02	-17.44	<0.001*
August	-0.46	0.02	-29.10	<0.001*
September	-0.57	0.02	-36.99	<0.001*
October	-0.87	0.01	-61.57	<0.001*
November	-0.61	0.01	-41.03	<0.001*
December	-0.03	0.01	-2.21	0.027*
Pop	-0.01	0.00	-2.03	0.042*
SD no (0)	0.00	-	-	-
SD yes (1)	0.07	0.01	12.18	<0.001*

*Significance at $p < 0.05$.

DISCUSSION

Role of GRACE data in recording sand dam impact

The authors hypothesized that the sand dam's large area of influence coupled with the 3,000+ sand dams in the SDC

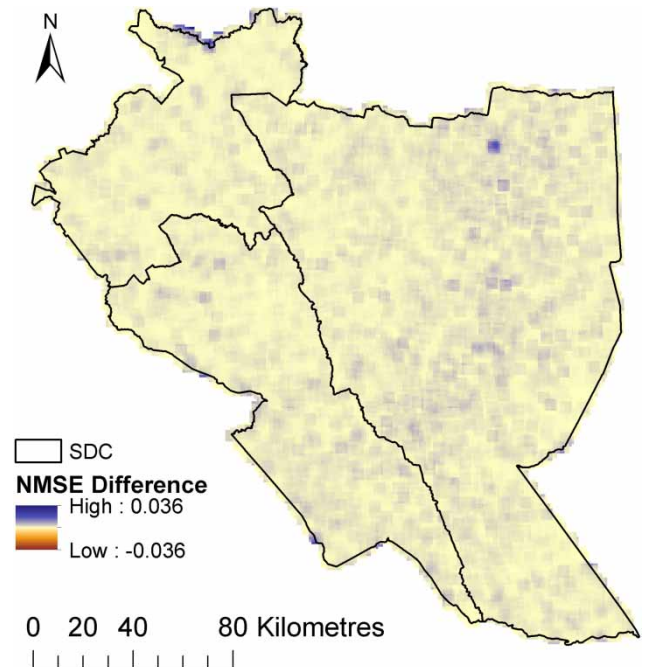


Figure 7 | Model improvement after adding the sand dam indicator variable to the NDVI GLM. NMSE difference is the average reduction in NMSE within a five-cell range after adding the sand dam indicator variable to the model. Adjusted R^2 is 0.70.

would lead to substantial groundwater recharge that GRACE could detect. A monthly test of means revealed that some months of the year are more likely to produce differences in water storage in the SDC and buffer area. When accounting for climate factors and initial conditions, the GRACE GLMs indicated that GRACE total water storage data is only sufficient for detecting a significant regional impact of sand dams on water storage during 3 months of the year. The sand dams have a positive impact on total water storage during the months that experience the highest rainfall during each rainy season, April and November. However, the months following April and November do not exhibit a significant increase in SDC total water storage compared to the buffer area. This pattern indicates that sand dams capture more water, but they do not necessarily store significantly more water in their watershed for long periods. The inability to store water long-term can likely be attributed to seepage under the sand dam and high evapotranspiration (de Trincheria *et al.* 2018; Eisma & Merwade 2020). A study of 30 sand dams in Kenya found that up to 37% of sand dams have severe seepage issues, and up to 87% of sand dams lose over half of their stored water to evaporation (de Trincheria *et al.* 2015).

Most rainfall in the SDC is lost either to baseflow-groundwater runoff or evapotranspiration (Figure 8). A properly sited and constructed sand dam will minimize the amount of stored water lost to baseflow-groundwater runoff, because the dam will be built atop a near-

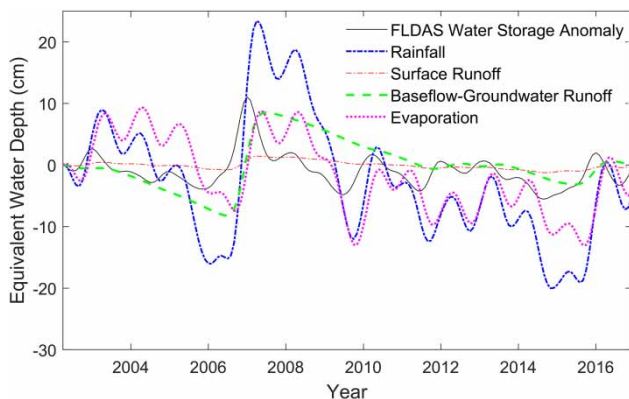


Figure 8 | Influence of FLDAS climate factors on the calculation of FLDAS water storage anomaly. Precipitation is a forcing variable for the FLDAS models; surface runoff, baseflow-groundwater runoff, and evapotranspiration are simulated.

impermeable streambed such as bedrock or consolidated clay. Nevertheless, sand dams are most often located in rural communities with limited or no access to subsurface surveying technology. Many sand dams may be constructed upon fractured bedrock, which could lead to significant water loss from the sand dam via rainwater infiltration into the bedrock (Kosugi *et al.* 2006). Furthermore, farmers cultivate the riparian zone and use water from the sand dam to irrigate their crops, increasing the amount of water lost from the sand dam via evapotranspiration. Water loss through baseflow-groundwater runoff and/or evapotranspiration could be further contributing to the failure of GRACE total water storage data to confirm the positive, collective impact of sand dams on regional water storage throughout the year.

Further complicating the ability of GRACE to detect a year-round, significant impact of sand dams on total water storage is the 1° resolution of the GRACE data. Sand dams have a positive impact on their local environment (Eisma & Merwade 2020), but these impacts may be occurring at too small a scale to be regularly observed via GRACE. The sand dams in the SDC may be constructed too far from each other to have overlapping areas of influence, decreasing the magnitude of their water storage signal. If approximately 50% of sand dams are not functioning properly due to issues of siltation and/or seepage (Viducich 2015; de Trincheria *et al.* 2018), the issue of GRACE's low resolution is exacerbated.

Link between increased vegetation and sand dams in the SDC

While the water storage GLM was unable to discern a significant impact of sand dams on regional water storage throughout the year, the vegetation GLM indicated that sand dams have a small, positive impact on the health and density of vegetative cover. Interestingly, vegetative cover and groundwater are linked, allowing conclusions to be drawn about the small-scale impact of sand dams on water storage (Le Maitre *et al.* 1999; Aguilar *et al.* 2012; Fu & Burgher 2015; Shemsanga *et al.* 2018; Eisma & Merwade 2020). This impact is restricted to the influence area of the sand

dams, and thus occurs at a smaller scale than can be detected by GRACE each month.

NDVI is linked to groundwater depth, particularly in arid regions (Fu & Burgher 2015). Vegetation undergoes a marked decrease in height and structural complexity as groundwater retreats from the land surface (Le Maitre *et al.* 1999), leading to a decrease in NDVI (Fu & Burgher 2015). Sand dams store water just below the land surface, resulting in pseudo-perched groundwater. The pseudo-perched groundwater seeps through the stream banks and has the potential to raise the local groundwater table, supporting increased vegetative growth (Manzi & Kuria 2011; Quinn *et al.* 2019; Eisma & Merwade 2020). Farmers also grow crops irrigated with sand dam water on the stream-banks, further increasing NDVI. The increased natural vegetation and cultivated crops result in increased transpiration, thereby lowering the groundwater table and reducing the duration of stored water in the dry season.

The interaction between the sand dam's influence on the local groundwater table and local vegetation creates a boom–bust dynamic for the vegetation (see Figure 9). During the rainy season, rainfall is stored in the sand dam and raises the local groundwater table; vegetation springs up as a result. The vegetation transpires water from the sand dam, drawing down the levels of stored water. When the rainy season ends and the sand dam dries up, the vegetation senesces and NDVI decreases. The cycle is repeated each rainy season. Figure 9 depicts this cycle very well, with the GLM coefficient being the lowest for the month right at the beginning of the rainy season and the highest at the end of a rainy season.

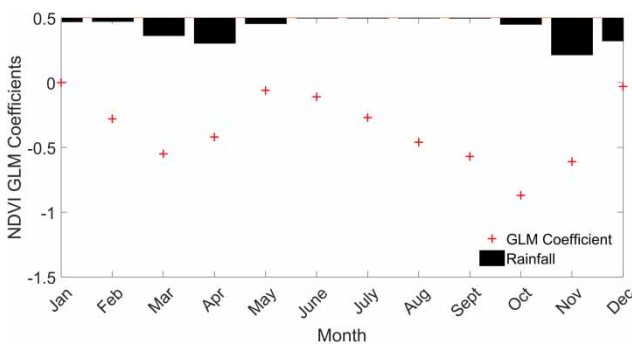


Figure 9 | Coefficients for the Month categorical variable in the NDVI GLM are directly correlated to the rainy and dry seasons in the SDC.

The NDVI GLM revealed that sand dams collectively impact NDVI and, therefore, the amount of available water in the region. The methodology was applied regionally in this study but can likely be applied to a different water harvesting technology on an individual basis. NDVI or GRACE, however, will not always be the appropriate remote sensing product to apply in this methodology. A dataset that is relevant to the function and/or impacts of the technology must be selected. For example, earthen dams create surface water storage. The Normalized Difference Water Index calculated from Landsat infrared radiation could be applied with this methodology to determine whether earthen dams are significantly contributing to available surface water (McFeeters 1996). High-resolution remotely sensed data can be an effective tool for exploring and quantifying impacts of small-scale water harvesting structures.

Future opportunities in the application of methodology

While every effort was made to develop a generalizable methodology that leveraged widely available datasets and well-documented techniques, future applications may benefit from an alternative approach. For example, GRACE data may not be an ideal dataset for small, regional studies. Future applications should consider using available local groundwater data to represent changes in total water storage, assuming that surface storage changes are negligible (Oiro *et al.* 2020). Furthermore, if GRACE data is appropriate for the selected study area and local groundwater data is available, the GRACE data should be downscaled to improve its resolution and accuracy (Miro & Famiglietti 2018; Seyoum *et al.* 2019). In addition, principal component analysis is inherently linear and may fail to represent trends in select GLM predictors. In such cases, alternative machine learning techniques designed to reduce the complexity of model predictors without losing information, like genetic programming, should be explored (Meshgi *et al.* 2014, 2015). Genetic programming is also capable of inducing an acceptable model given available data and model building blocks, which may be more effective than a PCA-style approach to representing climate variability (Chadalawada *et al.* 2020).

CONCLUSIONS

A methodology was developed and tested using two different remote sensing products. The methodology detected and quantified the effectiveness of a water harvesting technology common in south-eastern Kenya. The study revealed that sand dams likely do not significantly affect regional groundwater levels beyond a few months of the year. The GRACE total water storage data indicates that sand dams capture and store additional water during the rainy season, but that this additional storage is not significantly greater than water storage in an area without sand dams beyond the 2 months of the year with the most rainfall. GRACE's spatial resolution may limit the detection of increased water storage provided by sand dams, even when there are thousands of said structures in an area. MODIS NDVI, however, can detect the impact of sand dams on vegetative cover throughout the year. NDVI indicates that there is increased vegetation in an area with a high density of sand dams. This increased vegetation suggests higher groundwater levels resulting from the water storage added by sand dams. While GRACE and NDVI datasets were leveraged to quantify the regional impact of sand dams on water storage in this study, application of this methodology is not limited to GRACE and NDVI. Given the myriad remotely sensed data freely available, spatial GLMs show potential as a low-cost, adaptable technique for assessing water harvesting structures.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1333468.

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories (GRACE: <https://podaac.jpl.nasa.gov/>; MODIS: <https://lpdaac.usgs.gov/>; FLDAS: <https://disc.sci.gsfc.nasa.gov/>).

REFERENCES

- Aerts, J., Lasage, R., Beets, W., de Moel, H., Mutiso, G., Mutiso, S. & de Vries, A. 2007 *Robustness of sand storage dams under climate change*. *Vadose Zone J.* **6** (3), 572. doi:10.2136/vzj2006.0097.
- Aguilar, C., Zinnert, J. C., Polo, M. J. & Young, D. R. 2012 *NDVI as an indicator for changes in water availability to woody vegetation*. *Ecol. Indic.* **23**, 290–300. doi:10.1016/j.ecolind.2012.04.008.
- Anderson, W. B., Zaitchik, B. F., Hain, C. R., Anderson, M. C., Yilmaz, M. T., Mecikalski, J. & Schultz, L. 2012 *Towards an integrated soil moisture drought monitor for East Africa*. *Hydrol. Earth Syst. Sci.* **16** (8), 2893–2913. doi:10.5194/hess-16-2893-2012.
- Anselin, L. 2002 *Under the hood issues in the specification and interpretation of spatial regression models*. *Agr. Econ.* **27** (3), 247–267. doi:10.1016/S0169-5150(02)00077-4.
- Borst, L. & de Haas, S. 2006 *Hydrology of Sand Storage Dams: A Case Study in the Kiindu Catchment, Kitui District, Kenya*. Master's thesis, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands.
- Bouma, J. A., Hegde, S. S. & Lasage, R. 2016 *Assessing the returns to water harvesting: a meta-analysis*. *Agr. Water Manage.* **163**, 100–109. doi:10.1016/j.agwat.2015.08.012.
- Cao, Y., Nan, Z. & Cheng, G. 2015 *GRACE gravity satellite observations of terrestrial water storage changes for drought characterization in the arid land of northwestern China*. *Remote Sens.* **7** (1), 1021–1047. doi:10.3390/rs70101021.
- Chadalawada, J., Herath, H. M. V. V. & Babovic, V. 2020 *Hydrologically informed machine learning for rainfall-runoff modeling: a genetic programming-based toolkit for automatic model induction*. *Water Resour. Res.* **56** (4), e2019WR026933. doi:10.1029/2019WR026933.
- Cheng, M., Ries, J. C. & Tapley, B. D. 2011 *Variations of the Earth's figure axis from satellite laser ranging and GRACE*. *J. Geophys. Res.* **116** (B1). doi:10.1029/2010JB000850.
- Chinnasamy, P. & Ganapathy, R. 2018 *Long-term variations in water storage in Peninsular Malaysia*. *J. Hydroinform.* **20** (5), 1180–1190. doi:10.2166/hydro.2017.043.
- Crowley, J. W., Mitrovica, J. X., Bailey, R. C., Tamisiea, M. E. & Davis, J. L. 2006 *Land water storage within the Congo Basin inferred from GRACE satellite gravity data*. *Geophys. Res. Lett.* **33** (19). doi:10.1029/2006GL027070.
- de Trincheria, J., Nissen-Petersen, E., Filho, W. & Otterphol, R. 2015 *Factors affecting the performance and cost-efficiency of sand storage dams in south-eastern Kenya*. Presented at the 36th IAHR World Congress, The Hague, The Netherlands.
- de Trincheria, J., Filho Leal, W. & Otterpohl, R. 2018 *Towards a universal optimization of the performance of sand storage dams in arid and semi-arid areas by systematically minimizing vulnerability to siltation: a case study in Makueni, Kenya*. *Int. J. Sediment Res.* **33**, 221–233. doi:10.1016/j.ijsrc.2018.05.002.

- Didan, K. 2015 *MOD13A3 MODIS/Terra Vegetation Indices Monthly L3 Global 1 km SIN Grid V006*. NASA EOSDIS Land Processes DAAC. doi:10.5067/MODIS/MOD13A3.006 (accessed 18 August 2018).
- Didan, K., Barreto-Munoz, A., Solano, R. & Huete, A. 2015 *MODIS Vegetation Index User's Guide: MOD 13 Series (Version 3.00)*. Available from: https://lpdaac.usgs.gov/sites/default/files/public/product_documentation/mod13_user_guide.pdf (accessed 18 August 2018).
- Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., Davies, R. G., Hirzel, A., Jetz, W., Kissling, W. D., Kühn, I., Ohlemüller, R., Peres-Neto, P. R., Reineking, B., Schröder, B., Schurr, F. M. & Wilson, R. 2007 *Methods to account for spatial autocorrelation in the analysis of species distributional data: a review*. *Ecography* **30** (5), 609–628. doi:10.1111/j.2007.0906-7590.05171.x.
- Eisma, J. A. & Merwade, V. M. 2020 *Investigating the environmental response to water harvesting structures: a field study in Tanzania*. *Hydrol. Earth Syst. Sci.* **24** (4), 1891–1906. doi:10.5194/hess-24-1891-2020.
- Excellent Development 2019 *Reversing Land Degradation and Desertification*. Available from: <https://www.excellentdevelopment.com/reversing-land-degradation-and-desertification> (accessed 1 July 2019).
- FAO 2002 *Multipurpose Landcover Database for Kenya – AFRICOVER*. Food and Agriculture Organization of the United Nations, Digital Map, Rome, Italy.
- Fick, S. E. & Hijmans, R. J. 2017 *Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas*. *Int. J. Climatol.* **37** (12), 4302–4315. doi:10.1002/joc.5086.
- Fischer, G., Nachtergaele, F., Prieler, S., van Velthuisen, H., Verelst, L. & Wiberg, D. 2008 *Global Agro-Ecological Zones Assessment for Agriculture*. International Institute for Applied Systems Analysis, Digital Map, Laxenburg, Austria & Rome, Italy: Food and Agricultural Organization of the United Nations.
- Fu, B. & Burgher, I. 2015 *Riparian vegetation NDVI dynamics and its relationship with climate, surface water and groundwater*. *J. Arid Environ.* **113**, 59–68. doi:10.1016/j.jaridenv.2014.09.010.
- Garfi, M., Ferrer-Martí, L., Bonoli, A. & Tondelli, S. 2011 *Multi-criteria analysis for improving strategic environmental assessment of water programmes. A case study in semi-arid region of Brazil*. *J. Environ. Manage.* **92** (3), 665–675. doi:10.1016/j.jenvman.2010.10.007.
- Gero, A., Wahr, J. & Zhong, S. 2013 *Computations of the viscoelastic response of a 3-D compressible Earth to surface loading: an application to Glacial Isostatic Adjustment in Antarctica and Canada*. *Geophys. J. Int.* **192** (2), 557–572. doi:10.1093/gji/ggs030.
- Gupta, S. 2011 *Demystifying 'tradition': the politics of rainwater harvesting in rural Rajasthan, India*. *Water Altern.* **4** (3), 347–364.
- Hachborn, E., Berg, A., Levison, J. & Ambadan, J. T. 2017 *Sensitivity of GRACE-derived estimates of groundwater-level changes in southern Ontario, Canada*. *Hydrogeol. J.* **25** (8), 2391–2402. doi:10.1007/s10040-017-1612-2.
- Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S. & Kempen, B. 2017 *Soilgrids250 m: global gridded soil information based on machine learning*. *PLoS ONE* **12** (2), e0169748. doi:10.1371/journal.pone.0169748.
- Henry, C. M., Allen, D. M. & Huang, J. 2011 *Groundwater storage variability and annual recharge using well-hydrograph and GRACE satellite data*. *Hydrogeol. J.* **19** (4), 741–755. doi:10.1007/s10040-011-0724-3.
- Henry, M., Baldwin, G. & Quathamer, G. 2015 *Designing a community-based water harvesting system: understanding water use in Endallah, Tanzania*. *J. Purdue Undergrad. Res.* **5** (1), 38–47. doi:10.5703/jpur.05.1.05.
- Herman, J., Davis, A., Chin, K., Kinzler, M., Scholz, S. & Steinhoff, M. 2012 *Life with a weak heart: Prolonging the GRACE mission despite degraded batteries*. Presented at the 12th International Conference on Space Operations, Stockholm, Sweden.
- Huang, J., Halpenny, J., van der Wal, W., Klatt, C., James, T. S. & Rivera, A. 2012 *Detectability of groundwater storage change within the Great Lakes Water Basin using GRACE*. *J. Geophys. Res.-Sol. Ea.* **117** (B8). doi:10.1029/2011JB008876.
- Hut, R., Ertsen, M., Joeman, N., Vergeer, N., Winsemius, H. & van de Giesen, N. 2008 *Effects of sand storage dams on groundwater levels with examples from Kenya*. *Phys. Chem. Earth* **33** (1–2), 56–66. doi:10.1016/j.pce.2007.04.006.
- Jaber, J. O. & Mohsen, M. S. 2001 *Evaluation of non-conventional water resources supply in Jordan*. *Desalination* **136** (1–3), 83–92. doi:10.1016/S0011-9164(01)00168-0.
- Kosugi, K., Katsura, S., Katsuyama, M. & Mizuyama, T. 2006 *Water flow processes in weathered granitic bedrock and their effects on runoff generation in a small headwater catchment*. *Water Resour. Res.* **42** (2). doi:10.1029/2005WR004275.
- Landerer, F. W. & Swenson, S. C. 2012 *Accuracy of scaled GRACE terrestrial water storage estimates*. *Water Resour. Res.* **48** (4). doi:10.1029/2011WR011453.
- Lasage, R. & Verburg, P. H. 2015 *Evaluation of small scale water harvesting techniques for semi-arid environments*. *J. Arid Environ.* **118**, 48–57. doi:10.1016/j.jaridenv.2015.02.019.
- Lasage, R., Aerts, J., Mutiso, G.-C. M. & de Vries, A. 2008 *Potential for community based adaptation to droughts: sand dams in Kitui, Kenya*. *Phys. Chem. Earth* **33** (1–2), 67–73. doi:10.1016/j.pce.2007.04.009.
- Lasage, R., Aerts, J. C. J. H., Verburg, P. H. & Sileshi, A. S. 2015 *The role of small scale sand dams in securing water supply under climate change in Ethiopia*. *Mitig. Adapt. Strat. Gl.* **20** (2), 317–339. doi:10.1007/s11027-013-9493-8.
- Le Maitre, D., Scott, D. & Colvin, C. 1999 *A review of information on interactions between vegetation and groundwater*. *Water SA* **25** (2), 137–152.

- Manzi, H. & Kuria, D. 2011 The use of satellite images to monitor the effect of sand dams on stream bank land cover changes in Kitui District. *J. Agr. Sci. Technol.* **13** (2), 133–150.
- McFeeters, S. K. 1996 The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* **17** (7), 1425–1432. doi:10.1080/01431169608948714.
- McNally, A., Arsenault, K., Kumar, S., Shukla, S., Peterson, P., Wang, S., Funk, C., Peters-Lidard, C. D. & Verdin, J. P. 2017 A land data assimilation system for sub-Saharan Africa food and water security applications. *Sci. Data* **4**, 170012. doi:10.1038/sdata.2017.12.
- Meshgi, A., Schmitter, P., Babovic, V. & Chui, T. F. M. 2014 An empirical method for approximating stream baseflow time series using groundwater table fluctuations. *J. Hydrol.* **519**, 1031–1041. doi:10.1016/j.jhydrol.2014.08.033.
- Meshgi, A., Schmitter, P., Chui, T. F. M. & Babovic, V. 2015 Development of a modular streamflow model to quantify runoff contributions from different land uses in tropical urban environments using Genetic Programming. *J. Hydrol.* **525**, 711–723. doi:10.1016/j.jhydrol.2015.04.032.
- Miro, M. E. & Famiglietti, J. S. 2018 Downscaling GRACE remote sensing datasets to high-Resolution groundwater storage change maps of California's Central Valley. *Remote Sens.* **10** (1), 143. doi:10.3390/rs10010143.
- NASA/GSFC/HSL, A. M. 2016 *FLDAS Noah Land Surface Model L4 Monthly 0.1 × 0.1 Degree for Southern Africa (MERRA-2 and CHIRPS), Version 001 [Data set]*. NASA Goddard Earth Sciences Data and Information Services Center. doi:10.5067/8LPWNKCBUDA6.
- Ngigi, S. N., Savenije, H. H. G., Thome, J. N., Rockström, J. & de Vries, F. W. T. P. 2005 Agro-hydrological evaluation of on-farm rainwater storage systems for supplemental irrigation in Laikipia district, Kenya. *Agr. Water Manage.* **73** (1), 21–41. doi:10.1016/j.agwat.2004.09.021.
- Oiro, S., Comte, J.-C., Soulsby, C., MacDonald, A. & Mwakamba, C. 2020 Depletion of groundwater resources under rapid urbanisation in Africa: recent and future trends in the Nairobi Aquifer System, Kenya. *Hydrogeol. J.* doi:10.1007/s10040-020-02236-5.
- Previati, M., Bevilacqua, I., Canone, D., Ferraris, S. & Haverkamp, R. 2010 Evaluation of soil water storage efficiency for rainfall harvesting on hillslope micro-basins built using time domain reflectometry measurements. *Agr. Water Manage.* **97** (3), 449–456. doi:10.1016/j.agwat.2009.11.004.
- Qadir, M., Jiménez, G., Farnum, R., Dodson, L. & Smakhtin, V. 2018 Fog water collection: challenges beyond technology. *Water* **10** (4), 372. doi:10.3390/w10040372.
- Quilis, R. O., Hoogmoed, M., Ertsen, M., Foppen, J. W., Hut, R. & Vries, A. d. 2009 Measuring and modeling hydrological processes of sand-storage dams on different spatial scales. *Phys. Chem. Earth* **34** (4–5), 289–298. doi:10.1016/j.pce.2008.06.057.
- Quinn, R., Rushton, K. & Parker, A. 2019 An examination of the hydrological system of a sand dam during the dry season leading to water balances. *J. Hydrol. X* **4**, 100035. doi:10.1016/j.hydroa.2019.100035.
- Reyer, C. P. O., Rigaud, K. K., Fernandes, E., Hare, W., Serdeczny, O. & Schellnhuber, H. J. 2017 Turn down the heat: regional climate change impacts on development. *Reg. Environ. Change* **17** (6), 1563–1568. doi:10.1007/s10113-017-1187-4.
- Ryan, C. & Elsner, P. 2016 The potential for sand dams to increase the adaptive capacity of Easican drylands to climate change. *Reg. Environ. Change* **16** (7), 2087–2096. doi:10.1007/s10113-016-0938-y.
- Sakumura, C., Bettadpur, S. & Bruinsma, S. 2014 Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models. *Geophys. Res. Lett.* **41** (5), 1389–1397. doi:10.1002/2013GL058632.
- Seyoum, W. M., Kwon, D. & Milewski, A. M. 2019 Downscaling GRACE TWSA data into high-resolution groundwater level anomaly using machine learning-based models in a glacial aquifer system. *Remote Sens.* **11** (7), 824. doi:10.3390/rs11070824.
- Shemsanga, C., Muzuka, A. N. N., Martz, L., Komakech, H. & Mcharo, E. 2018 Indigenous knowledge on development and management of shallow dug wells of Dodoma Municipality in Tanzania. *Appl. Water Sci.* **8** (2). doi:10.1007/s13201-018-0697-7.
- Shlens, J. 2014 *A Tutorial on Principal Component Analysis*. Cornell University, Ithaca, NY.
- Standen, K., Costa, L. R. D. & Monteiro, J.-P. 2020 In-channel managed aquifer recharge: a review of current development worldwide and future potential in Europe. *Water* **12**, 3099. doi:10.3390/w12113099.
- Swenson, S. & Wahr, J. 2006 Post-processing removal of correlated errors in GRACE data. *Geophys. Res. Lett.* **33** (8). doi:10.1029/2005GL025285.
- Swenson, S., Chambers, D. & Wahr, J. 2008 Estimating geocenter variations from a combination of GRACE and ocean model output. *J. Geophys. Res.-Sol. Ea.* **113** (B8). doi:10.1029/2007JB005338.
- Tapley, B. D., Bettadpur, S., Watkins, M. & Reigber, C. 2004 The Gravity Recovery and Climate Experiment: mission overview and early results. *Geophys. Res. Lett.* **31** (9). doi:10.1029/2004GL019920.
- TELLUS 2012 *Grace Monthly Land Water Mass Grids NETCDF Release 5.0*. NASA PO.DAAC. doi:10.5067/TELND-NC005.
- Tobler, W. R. 1970 A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* **46**, 234. doi:10.2307/143141.
- Ulsaker, L. & Kilewe, A. 1983 Runoff and soil erosion for an Alfisol in Kenya. *E. Afr. Agr. Forestry J.* **44**, 210–241.
- Vidulich, J. 2015 *Spillway Staging and Selective Sediment Deposition in Sand Storage Dams*. Master's thesis, Oregon State University, Corvallis, Oregon, USA.
- Wang, M., Wright, J., Brownlee, A. & Buswell, R. 2016 A comparison of approaches to stepwise regression on variables sensitivities in building simulation and analysis. *Energ. Buildings* **127**, 313–326. doi:10.1016/j.enbuild.2016.05.065.

- Winsemius, H. C., Savenije, H. H. G., van de Giesen, N. C., van den Hurk, B. J. J. M., Zapreeva, E. A. & Klees, R. 2006 [Assessment of Gravity Recovery and Climate Experiment \(GRACE\) temporal signature over the upper Zambezi: GRACE temporal signature assessment](#). *Water Resour. Res.* **42** (12). doi:10.1029/2006WR005192.
- Yao, Y., Andrews, C., Zheng, Y., He, X., Babovic, V. & Zheng, C. 2019 [Development of fresh groundwater lens in coastal reclaimed islands](#). *J. Hydrol.* **573**, 365–375. doi:10.1016/j.jhydrol.2019.03.062.
- Yeomans, K. A. & Golder, P. A. 1982 [The Guttman-Kaiser criterion as a predictor of the number of common factors](#). *Statistician* **31** (3), 221. doi:10.2307/2987988.
- Yilmaz, M. T., Anderson, M. C., Zaitchik, B., Hain, C. R., Crow, W. T., Ozdogan, M., Chun, J. A. & Evans, J. 2014 [Comparison of prognostic and diagnostic surface flux modeling approaches over the Nile River basin: prognostic and diagnostic modeling over Nile](#). *Water Resour. Res.* **50** (1), 386–408. doi:10.1002/2013WR014194.

First received 9 July 2020; accepted in revised form 18 December 2020. Available online 19 January 2021