Environmental Modelling & Software 91 (2017) 241–250

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Accidental infrastructure for groundwater monitoring in Africa

Farah E. Colchester [a], Heloise G. Marais [a], Patrick Thomson [b], Robert Hope [b], David A. Clifton [a, *]

[a] Department of Engineering Science, University of Oxford, UK
[b] School of Geography and the Environment, University of Oxford, UK

A R T I C L E I N F O

Article history:
Received 21 October 2016
Received in revised form
16 January 2017
Accepted 27 January 2017
Available online 20 February 2017

A B S T R A C T

A data deficit in shallow groundwater monitoring in Africa exists despite one million handpumps being used by 200 million people every day. Recent advances with "smart handpumps" have provided accelerometry data sent automatically by SMS from transmitters inserted in handles to estimate hourly water usage. Exploiting the high-frequency "noise" in handpump accelerometry data, we model high-rate wave forms using robust machine learning techniques sensitive to the subtle interaction between pumping action and groundwater depth. We compare three methods for representing accelerometry data (wavelets, splines, Gaussian processes) with two systems for estimating groundwater depth (support vector regression, Gaussian process regression), and apply three systems to evaluate the results (held-out periods, held-out recordings, balanced datasets). Results indicate that the method using splines and support vector regression provides the lowest overall errors. We discuss further testing and the potential of using Africa's accidental infrastructure to harmonise groundwater monitoring systems with rural water-security goals.

1. Introduction

Timely and cost-effective groundwater monitoring is a global challenge in both industrialised and developing countries (Gorelick and Zheng, 2015; Giordano, 2009; Shah, 2010; Llamas and Martínez-Santos, 2005; Nelson, 2012; Foster and Garduño, 2013). Increasing urgency for policy action is driven by global groundwater depletion rates doubling between 1960-2000 and 2000–2009 from 56 km^3 per year to 113 km^3 per year (Doell et al., 2014). However, global groundwater data vary in extent and quality (Giordano, 2009; Wada et al., 2010; Mulligan et al., 2014). Africa is the most data-poor region with limited records (<0.001%) of global shallow groundwater records (Fan et al., 2013). Though high-yielding groundwater sites (>5 l s^{-1}) are limited and unevenly-distributed, groundwater is a strategic resource for Africa's growth and development, with groundwater storage estimated to be over 100 times greater than annual renewable freshwater sources (MacDonald et al., 2012). Africa's systematic data deficit in shallow groundwater monitoring is juxtaposed by rapid and often competing demands from domestic, industrial, and agricultural sectors with regulatory and enforcement systems either weak or absent. An unpredictable future climate will place new pressures on managing and allocating groundwater, thereby increasing the need for high-quality, low-cost shallow groundwater data in a distributed monitoring system.

Africa's shallow groundwater systems (<80 m depth) supply domestic water for around 200 million rural Africans lifted by one million handpumps distributed across rural areas (Foster and Garduño, 2013). Handpumps emerged as a low-cost, durable technology in the 1980s to supply drinking water to rural communities (Hope, 2015). Shallow groundwater accessed by handpumps operating throughout the year provides generally good quality water to buffer dry periods. With ongoing challenges of repairing broken handpumps in remote rural areas, a transmitter was designed, tested, and successfully deployed in handpump handles to automatically send data on pump usage via the GSM network. Volumetric abstraction is calculated from accelerometry data generated by the movement of the pump handle (Thomson et al., 2012). Since 2012, these data have provided information on hourly pump use, and have allowed local mechanics to be alerted when failure events occur, thus reducing the down-time following such events from over a month to several days (University of

* Corresponding author. Institute of Biomedical Engineering, Department of Engineering Science, Old Road Campus Research Building, University of Oxford, Headington, Oxford, OX3 7DQ, UK.
E-mail address: david.clifton@eng.ox.ac.uk (D.A. Clifton).

http://dx.doi.org/10.1016/j.envsoft.2017.01.026

Oxford/RFL, 2015). Methods using pressure sensors and water detection have also been developed to monitor pump usage remotely, similarly aimed at reducing pump downtimes (Nagel et al., 2015).

Further analysis of the accelerometry data revealed that elements of the high-frequency components appeared to correspond to groundwater depth. This study provides proof-of-concept analysis of novel methods based on machine learning to predict aquifer depth from the high-frequency signal. The implications of the findings present a potentially scalable approach to address Africa's groundwater monitoring deficit by harnessing handpumps as accidental infrastructure (Frischmann, 2012) to improve groundwater resource management; regulate and monitor irrigation, mining, or other commercial groundwater users; and provide early-warning systems for vulnerable populations dependent on shallow groundwater resources.

In this article, we model the high-rate waveforms from the accelerometry data using robust machine learning techniques that are sensitive to the subtle interaction between the dynamics of the handpump and the depth of the aquifer beneath the pump. We compare the ability of various candidate machine-learning models for the purposes of estimating aquifer depth.

2. Materials and methods

2.1. Study site and data description

The work described considers two datasets of accelerometery recordings, collected from two different models of handpumps: the Afridev and the India MK II. The first set of recordings, referred to as the "Oxford" dataset, was collected from an India MK II handpump installed at the University of Oxford, UK, between April and November, 2014. The second set of recordings, referred to as the "Kenya" dataset was collected from 11 Afridev handpumps installed in Kwale County, located between Mombasa and Tanzania's northern border, over a two-week period in April, 2014.

Each dataset consists of recordings taken at the pump location. To obtain recordings during our experiments, a consumer-grade accelerometer was mounted to the handle of each pump, and connected to a nearby laptop via a Bluetooth data connection. Each recording comprises a single person pumping for 20 s–120 s. The resulting accelerometry measurements in three orthogonal ("triaxial") dimensions are recorded at 96 Hz. The signal recorded by the accelerometer is proportional to the force applied to the handle during the pumping motion. As the angle of the handle changes, the axis along which the acceleration is sensed changes. The lateral movement of the handle results in the presence of additional acceleration components; however, the accelerometer is mounted close to the fulcrum of the motion, and the angular velocity of the handle is low, and so these additional components are small compared to the effect of the applied force (Thomson et al., 2012).

Depth measurement at the Oxford site was performed using a manual "dipper", which is lowered into the borehole and which sounds on contact with water. Measurements were made before each recording to the nearest 1 cm, a level of precision deemed to be appropriate with respect to the measurement error (as shown later). The depths for the Kenya dataset are estimates based on the known depth of the pump's rods. The volume of water abstracted in Oxford was very low, being tens of litres using a pump that is capable of pumping over a thousand litres per hour. Combined with the properties of the shallow aquifer in Oxford, this level of pumping would have no impact on water level, meaning that aquifer level could be viewed as being constant over the period of each recording. This is addressed further in section 5.

Of the three measurement dimensions recorded by the device, we use the dimension perpendicular to the pump handle, associated with the waveform of the largest amplitude for our analysis. Intervals of 5 s of accelerometry data collected using this method are shown in Fig. 1.

These time-series accelerometry signals show the fundamental pumping motion, similar to the motion of the handle. The increasing parts of each waveform in Fig. 1 correspond to the handle being pushed downwards to lift water and the decreasing parts to the handle lifting to reset the pump. We refer to each of these cycles as being a *period*.

The example waveforms in the figure also show the noise present in the data, which is mostly caused by the mechanical vibration in the pump due to the motion of the handle. The figure shows that this noise is of larger amplitude on the increasing part each period than on the decreasing part; this effect is anticipated, because the increasing parts of each period correspond to mechanical loading of the handpump, while the weight of the water is being lifted, as described above. The examples in the figure have different levels of noise; it may be seen that the India MK II pump at Oxford (shown in the lowermost plot in the figure) has the noise with the highest amplitude - the handle rubs against the body of the pump which causes substantial vibration levels when water is being lifted. Infrequent use of this pump (because it is a prototype in the university setting) means that the pump has not yet reached a dynamic equilibrium by being "worn in".

We separate the time-series accelerometry data at the troughs (after smoothing using a low-pass filter) to divide the recordings into individual periods. Thus, each recording for each pump yields a series of periods of accelerometry data. The latter are generally between 0.8 s and 1.2 s in length, and hence contain approximately 80 and 120 data points (sampled at 96 Hz).

2.2. Representing each period of accelerometry data

The next step of our analysis aims to reduce the high-rate (96 Hz) time-series accelerometry data contained within each period into quantities that capture their dynamical characteristics in a parsimonious manner, suitable for modelling. We consider two main characteristics for each period: the shape (representing the pumping movement) and the high-frequency vibrations in the handle during the movement.

We chose to investigate three methods for representing each period. For each method, we summarise each period of waveform data from each recording using (i) a *feature vector* representing the shape of that period and (ii) a feature vector representing the vibration levels observed during that period. These feature vectors are sets of scalar variables (defined below), and labelled $\mathbf{s}$ and $\mathbf{v}$, for shape and vibration, respectively.

2.2.1. Representation method I: wavelets

The wavelet transform (Torrence and Compo, 1998) provides information about the magnitude of different frequency components present in a time-series, and how these change over time. The wavelet transform should reveal both the underlying shape of the waveform corresponding to the gross pumping motion (which corresponds to relatively low frequencies in the signal), in addition to components describing the vibration (which correspond to relatively high frequencies in the signal).

Fig. 2 shows the wavelet transform applied to an example waveform of accelerometry data. Fig. 2a shows the original 96 Hz waveform Fig. 2b shows its wavelet transform. The figure shows a time vs. frequency plot for this wavelet transform, which is a heatmap corresponding to the strength of frequency content of the signal, through all points in time. (Higher frequencies correspond to