

Assessing the performance and robustness of the UNICEF model for groundwater exploration in Ethiopia through application of the analytic hierarchy process, logistic regression and artificial neural networks

David Josephs-Afoko^{1*}, Samuel Godfrey² and Luiza C Campos¹

¹University College London, Gower Street, London WC1E 6BT, United Kingdom

²Chief of WASH, UNICEF Ethiopia, UNECA Compound, NOF-Building #020 3rd floor, Box 1169 Addis Ababa, Ethiopia

ABSTRACT

This study assesses the performance and robustness of the groundwater potential (GWP) maps produced by the UNICEF model for deep groundwater exploration in Ethiopia. The UNICEF model is a weighted linear combination of hydrogeological parameters including permeability, slope, recharge, and lineament density, which has been calibrated using the expert judgements of local hydrogeologists. In order to assess the performance and robustness of the model, three techniques were employed: the analytic hierarchy process (AHP), logistic regression (LR), and artificial neural networks (ANNs). Three study areas (Dallol, Halaba and Shinelle) were selected on the basis of climatic and geological variation, in addition to the availability of well data pertaining to depth and yield. The performance of the UNICEF model in predicting outcomes of the well data included in the study was assessed by computing the receiver operating characteristic (ROC) curve. The solutions produced by the AHP and ANN were more accurate than the UNICEF model in determining the productivity of deep wells in the study data, whilst the LR model was less accurate than the UNICEF model. The groundwater productivity maps produced by the AHP and ANNs showed clear correlation with the maps produced by the UNICEF model, despite moderate (AHP) and severe (ANN) parameter perturbation, demonstrating the robustness of the UNICEF model. Whilst the AHP and ANN models demonstrated higher accuracy than the UNICEF model, this must be considered against the well data used to assess accuracy, which were drawn from a small sample of non-ideal distribution. Although this study focuses on case studies in Ethiopia the key findings are applicable internationally, namely, that the use of the AHP in data-scarce environments provides robust models, and that with the addition of easily obtainable well data the accuracy of modelling can be significantly increased through the application of ANNs.

Keywords: groundwater exploration, hydrogeology, artificial neural networks, Ethiopia, WaSH

INTRODUCTION

In 2016, Ethiopia experienced the worst water drought in 30 years (Godfrey, 2016). Almost 10 million people in 6 regions of the country were affected by high temperatures and low rainfall which left the phreatic and shallow aquifers with limited recharge and potable water storage. Consequently, the majority of the population became reliant on tanked water that was extracted from resilient deep groundwater sources. A total of 3 million people survived solely on tanked water supplied by UNICEF for 6 months in 2016. To reduce the dependence on tanked water and to ensure more 'value for money' (VfM) solutions, the UNICEF programme invested in the exploration, identification and exploitation of deep groundwater resources.

UNICEF implemented a 3-phase approach to locating and providing groundwater, as detailed in the paper by Godfrey and Hailemichael (2016). Phase 1 of the approach generated remote-sensing data which was combined in Phase 2 with ground measurements of key hydrogeological parameters to produce the UNICEF model. The model is a weighted linear combination of hydrogeological parameters including permeability, slope, recharge, and lineament density, which has been calibrated using the expert judgements of local hydrogeologists. Phase 3 comprised the validation of the model

through the drilling of production boreholes in sites identified by the UNICEF model (Godfrey and Hailemichael, 2016). The 3-phase approach was developed and implemented in 2015 and 2016 in 12 districts of Ethiopia and resulted in the drilling of 13 productive boreholes (Godfrey, 2016). Before further expanding the approach to other hydrogeological zones of Ethiopia, it is essential to ascertain the performance and robustness of the methodology.

In several studies the weightings of parameters indicating GWP are assigned by expert judgements (Kumar et al., 2009; Madrucci et al., 2008; Magesh et al., 2012). More rigorous studies derive weightings using the analytic hierarchy process (AHP) (Agarwal, 2016; Dhar et al., 2015; Jha et al., 2010; Mandal et al., 2016; Rahmati et al., 2015; Srivastava & Bhattacharya, 2006). The groundwater potential index (GWPI) produced by such overlay methods lacks inherent meaning and requires interpretation; hence such studies delineate zones on qualitative scales typically ranging from 'high' to 'low' GWP. In some cases, these classifications are arbitrary as the study areas are divided into a predefined number of classes, regardless of the overlay results (Adji, 2014; Agarwal, 2016), or similarly when a predefined percentage of the study area will fall into a given class (Oh et al., 2011). These approaches mean that, in any non-homogenous study area, GWPIs are classified in all classes from 'high' to 'low', potentially leading to uneconomical decision making.

Several studies do not validate their models using knowledge of aquifer conditions or well outcomes, despite having this information available (Jasrotia et al., 2013; Jasrotia

* To whom all correspondence should be addressed.
e-mail: djafoko@live.co.uk

Received 16 May 2017; accepted in revised form 11 June 2018

et al., 2016; Kumar et al., 2009). Conversely, more rigorous studies such as those by Oh et al. (2011), Lee et al. (2012) and Nampak et al. (2014) have trained their models using yield data, then performed validation.

Many probabilistic models have been applied to GWP mapping, including frequency ratio models (Oh et al., 2011; Razandi et al., 2015), weight of evidence models (Corsini et al., 2009), evidential belief functions (Nampak et al., 2014), logistic regression (Nampak et al., 2014; Ozdemir, 2011) and artificial neural networks (Corsini et al., 2009; Lee et al., 2012). The value in probabilistic models is that the solutions have inherent meaning – the probability of groundwater occurrence. Whilst all of the aforementioned methods are capable of producing probabilistic maps, logistic regression (LR) and artificial neural networks (ANNs) both provide models which denote the relative importance or weight of the parameters included and find optimal solutions based on training data. These aspects facilitate an assessment of robustness and can readily be applied to multiple study areas.

This study aimed to assess the performance of the UNICEF model by considering its accuracy in determining well productivity and to assess the robustness of the solutions by comparing them to the solutions produced by numerical methods that consider either expert judgments (i.e. AHP) or a-posteriori knowledge of well outcomes (i.e. LR and ANN). Whilst this study has a focus on Ethiopia, it is intended that the findings will have international implications in groundwater exploration and particularly in the WaSH sector.

MATERIALS AND METHODS

Conventions

The following definitions were provided by UNICEF:

- A deep well is ≥ 150 m below ground level
- A productive well has yield ≥ 1 L/s

The minimum acceptable yield for rural drinking water supply is ≥ 1 L/s so as to maximize the multiple use of water for the drinking water and irrigation targets of the Ethiopia Growth and Transformation Plan (National Planning Commission, 2016).

Study areas

Three *woredas* (districts) are considered in this study; Dallol, Halaba and Shinelle (Fig. 1). The *woredas* have differing geological characteristics and are located in different climatic zones, allowing the UNICEF model to be tested in varying hydrogeological conditions. Each study area comprised a rectangular area centred on the *woreda* administrative boundary.

Dallol

Dallol is located in the Afar region of Ethiopia, with a 2007 population of approximately 84 000 (Central Statistics Agency, 2007). Dallol has the highest recorded annual mean temperature of any inhabited settlement on Earth of 34.4°C and peak temperatures in excess of 47°C (Burt & Stroud, 2007). Mean annual rainfall is 40 mm, as determined from the spatial database. 40% of the *woreda* is covered by alluvial fans composed of silt, sand, gravel and salt crust. The remainder of the *woreda* is composed of Tsaliyet group meta-volcanics, Tambien group meta-sediments, granitic intrusions of Precambrian age, and Mesozoic sedimentary units (UNICEF, 2016a). The study area is approximately 1 730 km².

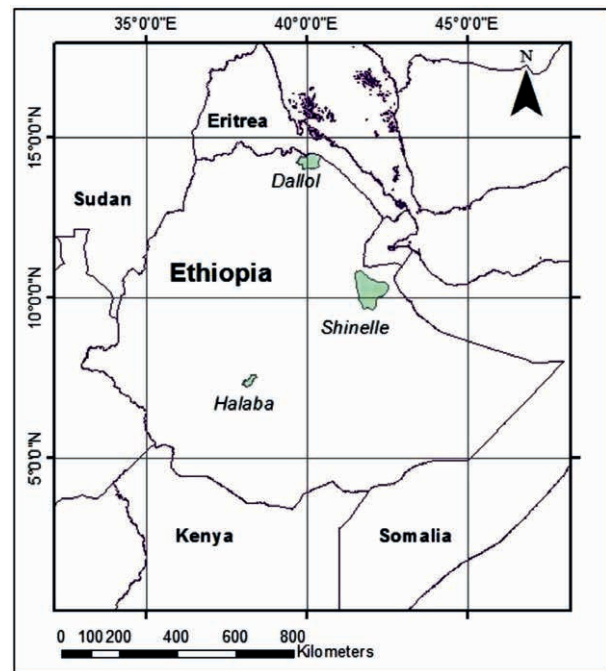


Figure 1
Overview of the study areas

Halaba

Halaba is located in the Southern Nations, Nationalities, and Peoples' Region (SNNPR) of Ethiopia, with a 2007 population of approximately 230 000 (Central Statistics Agency). The climate is temperate, with a mean maximum temperature between 19 and 22°C and mean annual rainfall of 940 mm, as determined from the spatial database. The *woreda* is covered by unwelded pumiceous pyroclastics and ignimbrites, tuffs, water lain pyroclastics and occasional lacustrine beds (UNICEF, 2016a). The study area is approximately 1 045 km². The *woreda* has a high population density and much of the existing groundwater supply contains high levels of fluoride – up to 13 mg/L (UNICEF, 2016b).

Shinelle

Shinelle is located in the Somali Region of Ethiopia, with a 2007 population of approximately 102 000 (Central Statistics Agency). The geology is dominated by alluvial deposits composing sand, silt and clay with gravel. The mean high temperature is between 20 and 28°C (Ethiopian Government, 2016) and mean annual rainfall is 196 mm, as determined from the spatial database. The study area is approximately 7 427 km². The area receives significant regional recharge from surrounding highlands (UNICEF, 2016b).

Spatial database

The spatial database was assembled jointly by the European Union Joint Research Centre (EU-JRC) and UNICEF hydrogeologists. The EU-JRC provided interpreted satellite data for slope, topographic wetness index (TWI), drainage networks and lineaments from Shuttle Radar Topography Mission (SRTM) data; precipitation data from Tropical Applications of Meteorology using Satellite data and

ground-based observations (TAMSAT) and Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) and evapotranspiration and normalised difference vegetation index (NDVI) from Moderate-Resolution Imaging Spectroradiometer (MODIS) (UNICEF, 2016). UNICEF assembled primary and secondary data on lithology and water points, such as depth, yield obtained from pumping tests, and location; along with non-hydrogeological data including settlement locations and roads. The database is constructed to a spatial resolution of 30 m by 30 m, based on the digital elevation model. The thematic layers were then converted to ArcMap grid format.

Well data

Initially, data were available for 40 productive deep wells across the study areas from the Government and UNICEF well database, with the distribution summarised in Table 1. The positions and some water quality data were available for an additional 187 wells in the study areas; however data on depth and yield were not both available for the majority of these wells, hence they were omitted from the study.

UNICEF deliver water by truck to drought-afflicted areas. The wells with depth data, but missing yield data, were superimposed on the locations of *kebeles* (villages) receiving emergency water to determine if wells were productive – a *kebele* with a productive well would not receive emergency water. A *kebele* more than 3 km from another productive water source receiving emergency supply from UNICEF was deemed to have unproductive wells. Seven productive deep wells in Dallol were inferred with this approach. No deep unproductive wells were identified. Hence, the depth criterion was relaxed

Study area	Number of productive deep wells	Number of unproductive deep wells	Number of unproductive wells at 90–150 m depth
Dallol	13a	0	6b
Halaba	23	0	8c
Shinelle	11	0	0

^a9 inferred, ^binferred, ^cknown

to 90 m for unproductive wells, which inferred 6 unproductive wells in Dallol, and for a further 8 known unproductive wells in Halaba to be included, bringing the final number of wells in the study to 61.

The yield in the dataset ranges from 0 to 50 L/s. The mean yield for productive wells is 10.5 L/s; 12 of the wells have yield ≥ 20 L/s.

The well data were partitioned into training and validation sets, with a target ratio of 70:30. Only the ANN and LR models required training; as the UNICEF model and the AHP models do not consider well outcomes, the entire dataset was used for validation. The partitioning was random with constraints to ensure the sets were representative:

- Each study area must contain approximately a 70:30 ratio of training to validation wells
- Training and validation sets must contain approximately a 70:30 ratio of productive and unproductive wells

The dimensions of the training and validation sets are displayed in Table 2. The training set contains 42 wells; there are 19 wells in the validation set. The training and validation sets of wells by study area are presented in Figs 2–4.

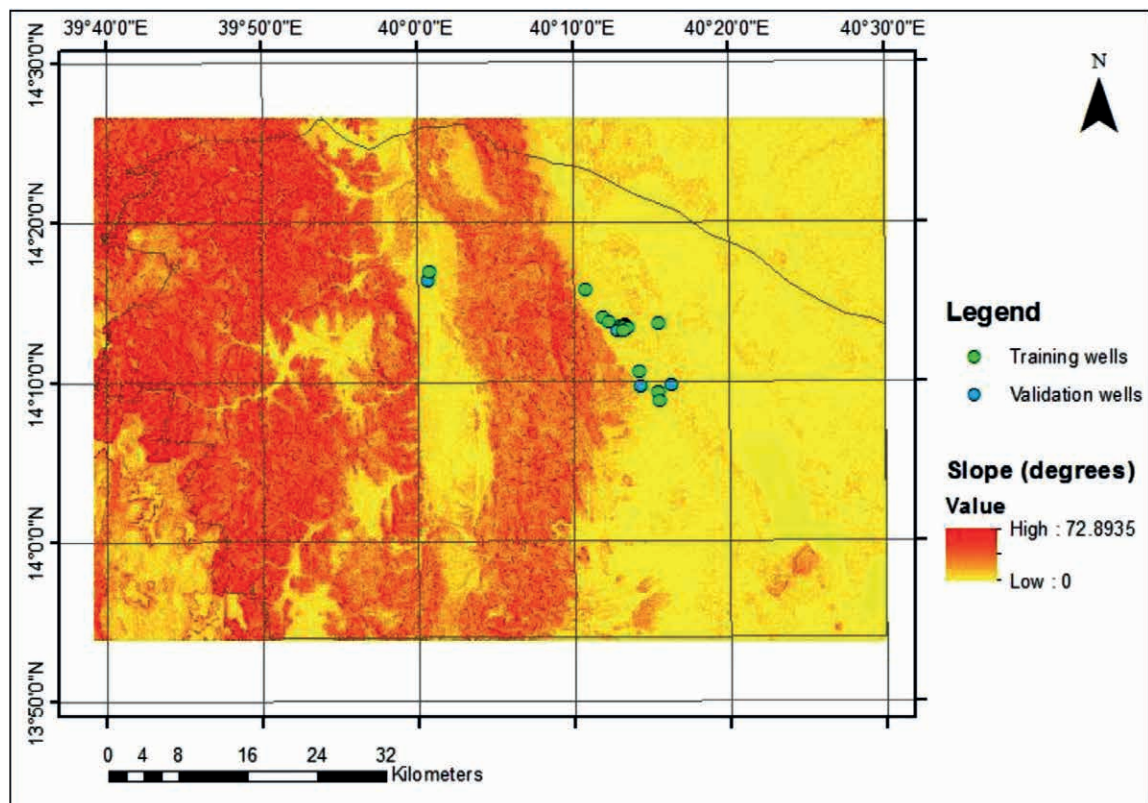


Figure 2
Slope map of Dallol study area

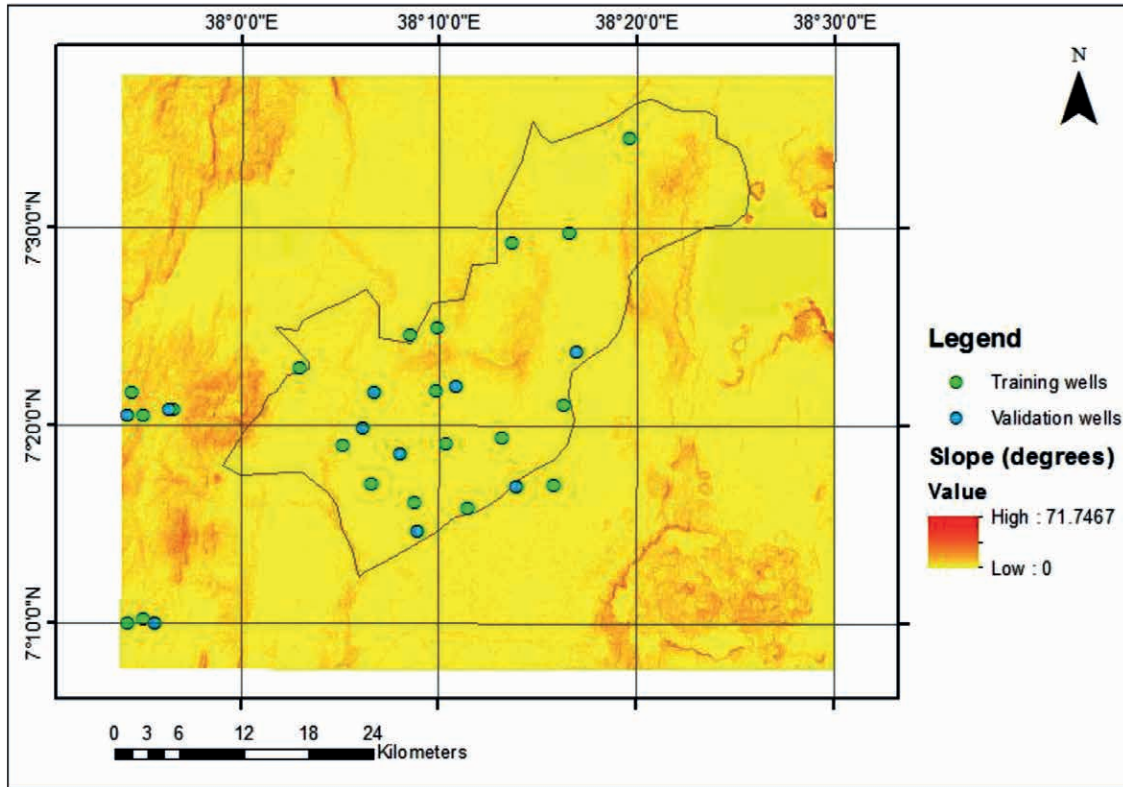


Figure 3
Slope map of Halaba study area

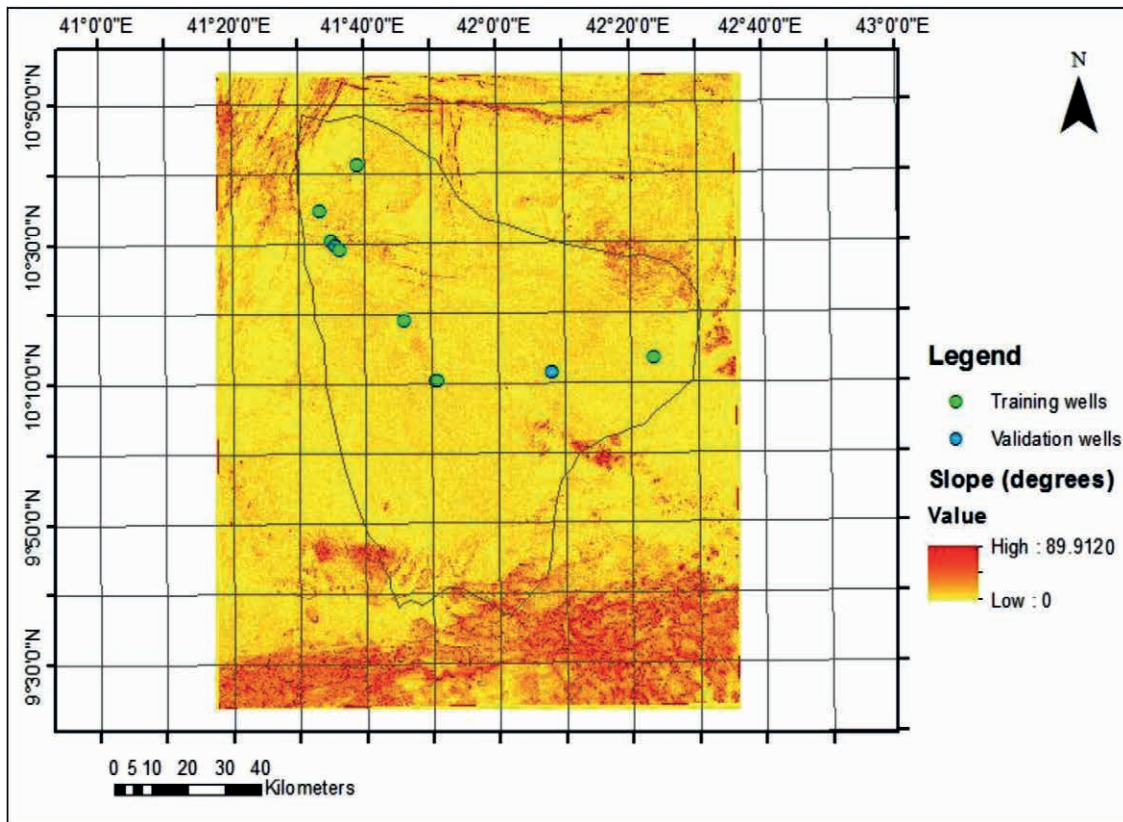


Figure 4
Slope map of Shinelle study area

Analytic hierarchy process

The AHP was applied to determine whether or not additional parameters should be included in the model and to capture the expert judgements of hydrogeologists on the relative importance of parameters. A survey was designed and distributed to elicit judgements from hydrogeologists in the Water, Sanitation and Hygiene (WaSH) team of UNICEF Ethiopia, as well as a hydrogeologist from Uganda. The UNICEF hydrogeologists are familiar with the study areas and with the UNICEF model, whilst the Ugandan hydrogeologist is unfamiliar with both the study areas and the UNICEF model. The survey required 21 pairwise comparisons to be made by the experts, on the 9-point scale shown in Table 3. Intermediate values were available to the respondents to represent a compromise.

A systematic review was conducted on the subject of GWP mapping, in order to establish the most frequently used parameters. This list was refined by considering the data available for this study. The remaining parameters were land use/land cover (LULC), elevation, and proximity to surface water bodies. These parameters were included in the AHP questionnaire to determine whether additional parameters should be included in the model.

The survey results were processed using the method outlined by Saaty (1980) to form a total of 6 pairwise comparison matrices (PCMs). Two PCMs were formed from each respondent, the first a 7×7 including the 4 parameters of the UNICEF model and the additional parameters of LULC, elevation and proximity to surface water bodies resulting from the literature review. The second PCM from each respondent is a 4×4 consisting of the original parameters. The priority vector, representing the relative contributions of each parameter, was obtained by calculating the row geometric mean of the PCM as outlined by Saaty (1980, pg. 19).

Consistency was calculated using the method outlined by Saaty (1980, pg. 19). All responses were revised due to inconsistency, with the sources of inconsistency highlighted, using the technique outlined by Harker (1987). One PCM for the UNICEF respondents was obtained by consensus, with consistency ratio (CR) calculated as 0.084 and hence deemed consistent. As consensus was reached, the UNICEF respondents are considered in the singular hereinafter, as Respondent 1. The final PCM from Respondent 2 had CR calculated as 0.082, hence was also consistent. The resultant priority vectors were normalised by dividing each component by the sum of the priority vector.

The normalised priority vectors were mapped using the weighted sum function of ArcMap. The GWPIs were classified qualitatively ranging from 'very high' to 'very low', using the natural breaks method which was applied by UNICEF in mapping the UNICEF model. Sensitivity (True Positive Rate), was plotted against $1 - \text{Specificity}$ (False Positive Rate) to produce the ROC curve. The area beneath the ROC curve, known as area under curve (AUC), represents the probability that a random positive event has a higher value of the metric calculated by the model (e.g. GWPI), than a random negative event (Altman and Bland, 1994). When expressed as a percentage, AUC is equivalent to accuracy.

Logistic regression

The dependent variable in the LR model was well productivity. Productive and unproductive wells were assigned values 1 and 0, respectively. Slope, recharge and lineament density were modelled as continuous variables. Permeability is a categorical variable – rock formations were assigned relative permeability values by UNICEF hydrogeologists which corresponded to hydraulic conductivity coefficients as displayed in Table 4. UNICEF hydrogeologists also considered surface permeability in assigning relative permeability values.

The logistic regression model is defined as:

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \quad (1)$$

where $g(x)$ is a weighted linear combination of all explanatory variables. Permeability, slope, recharge, and lineament density are represented by $x_{1,2,3,4}$, respectively, and $\beta_{1,2,3,4}$ are coefficients representing the contribution of each term to $g(x)$. β_0 is a constant. The parameter data were normalised to place the β coefficients on the same scale. Maximum likelihood estimation was then applied in the software SPSS to calculate an estimate

Relative permeability	Hydraulic conductivity coefficient
Very low	0.03
Low	0.05
Moderate	0.10
High	0.15
Very high	0.20

Study area	Productive wells in training set	Unproductive wells in training set	Productive wells in validation set	Unproductive wells in validation set
Dallol	9	4	4	2
Halaba	15	6	7	3
Shinelle	8	0	3	0

Less important				Equally important	More important			
Extremely	Very strongly	Strongly	Moderately		Moderately	Strongly	Very strongly	Extremely
1/9	1/7	1/5	1/3	1	3	5	7	9

of the vector β . The probability of well productivity was calculated by applying the method outlined by Hosmer et al. (2013). The ROC curve was then plotted.

Artificial neural networks

The parameters of the UNICEF model were input to the network, with the same conventions applied as in the LR model.

The network architecture comprised the 4 input signals (model parameters), 4 neurons in the hidden layer and 2 output neurons representing the binary response. The solutions produced by ANNs varied due to the randomisation of initial synaptic weights. Hence, multiple networks were computed as practiced by Lee et al. (2012), with the most accurate selected. The scaled conjugate gradient algorithm was applied to minimise back-propagated errors. A minimum relative change in the training set errors of 0.001 was applied, along with a constraint of no more than 100 iterations without error reduction before termination.

The ANN results were mapped by extracting parameter data from a grid of 500 m by 500 m intervals by study area and inputting the data to the trained network. Extracting parameter data at this resolution provided higher resolution than the current and AHP maps which use the weighted sum function in ArcMap that extracts data at the lowest resolution of all layers, in this instance recharge.

RESULTS AND DISCUSSION

Figures 5 to 7 show the GWP maps produced by each model grouped by study area.

UNICEF model

The UNICEF model had satisfactory accuracy (as represented by AUC Fig. 8a) of 62% across all three study areas. The accuracy in Dallol was satisfactory at 62.2% (Fig. 8f); however, the accuracy in Halaba was poor at only 56.1% (Fig. 8k). The ROC cannot be computed for Shinelle as there are no unproductive wells in the available data.

Approximately 54% of the Dallol study area was classified as low or very low potential, with 28% classified as moderate, and 18% as high or very high (Fig. 5a). Approximately 16% of the Halaba study area was classified as low or very low potential, with 43% classified as moderate potential, and 38% classified as high or very high potential (Fig. 6a). Approximately 13% of the Shinelle study area was classified as low or very low potential, with only 6% classified as moderate, and 81% classified as high or very high potential (Fig. 7a).

The large difference between the accuracy in Dallol and Halaba suggests that for improved performance the model should be calibrated for the study area to which it is applied.

AHP

Reviewing the priority vectors computed from the 7 x 7 PCMs (displayed in Table 5) shows the respondents generally agree that the parameters included in the UNICEF model are the most important in determining GWP. This demonstrates that the UNICEF model is robust to the addition of parameters. The difference in opinion between the importance of LULC as a parameter is explained by Respondent 2 being unfamiliar with the permeability term in the UNICEF model accounting for surface and sub-surface permeability.

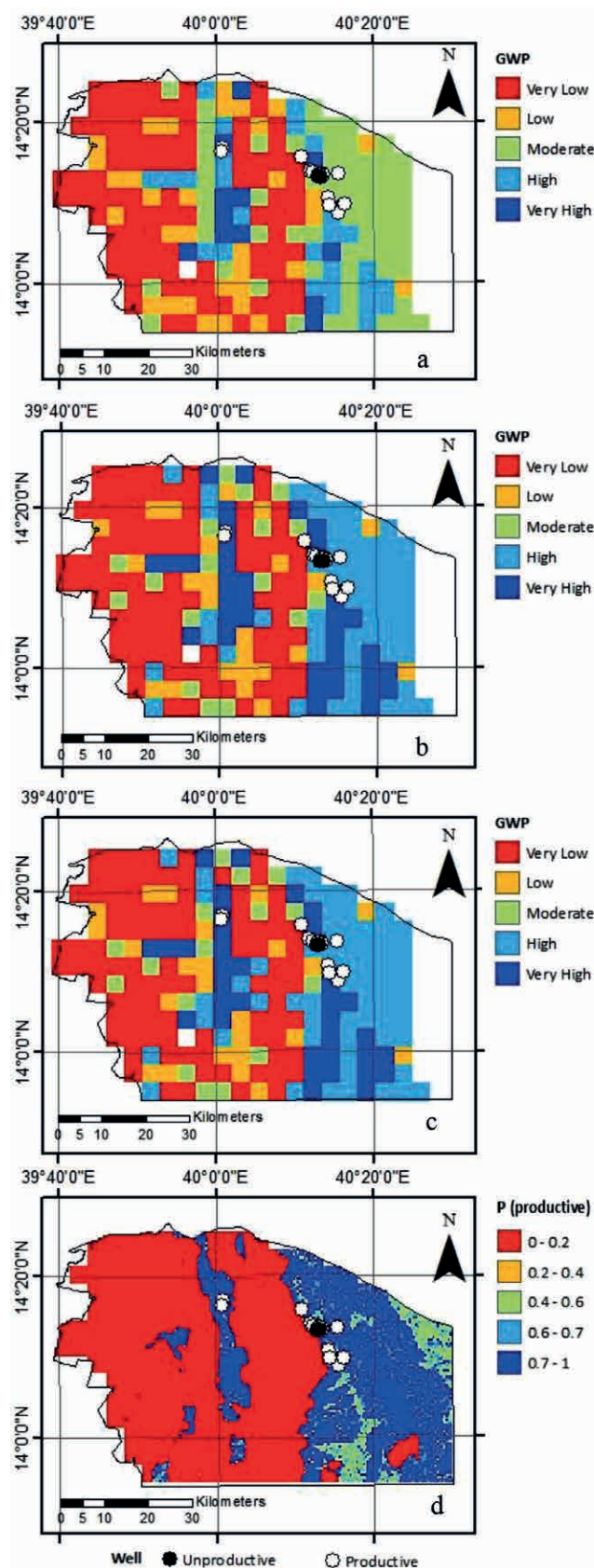


Figure 5
GWP maps for Dallol, a – UNICEF model, b – AHP1, c – AHP2, d – ANN (probabilistic map)

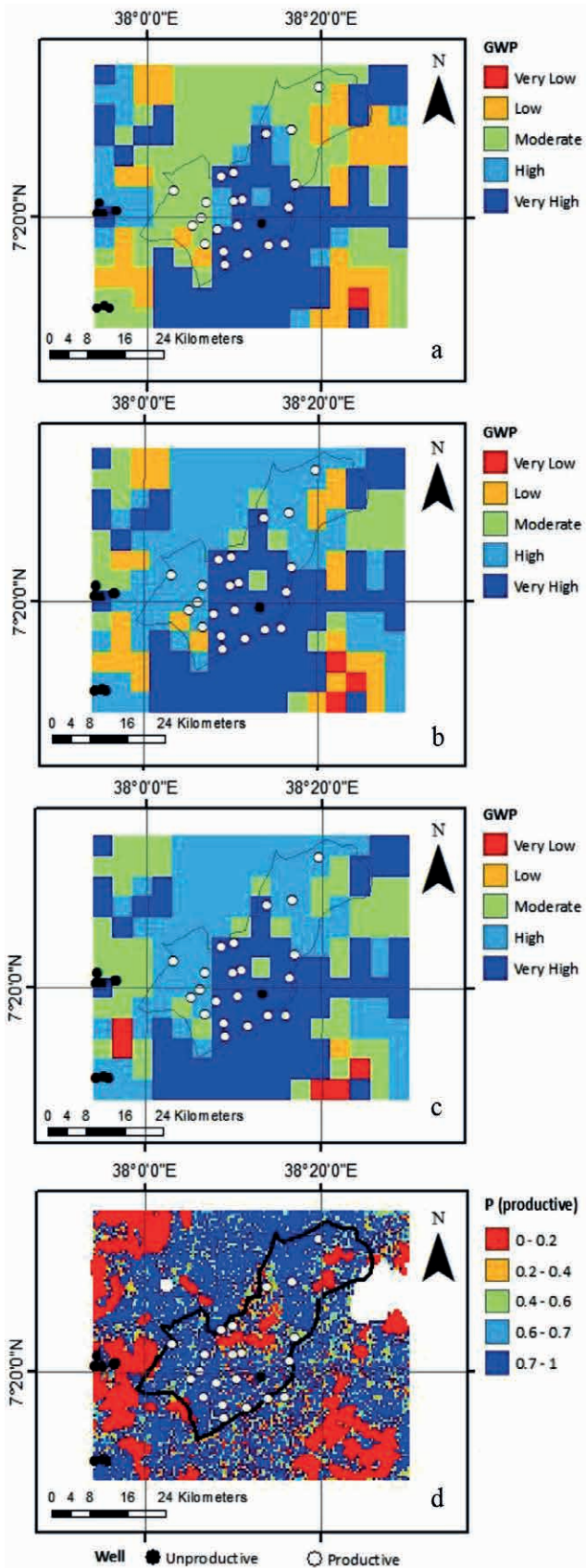


Figure 6

GWP maps for Halaba, a – UNICEF model, b – AHP1, c – AHP2, d – ANN (probabilistic map)

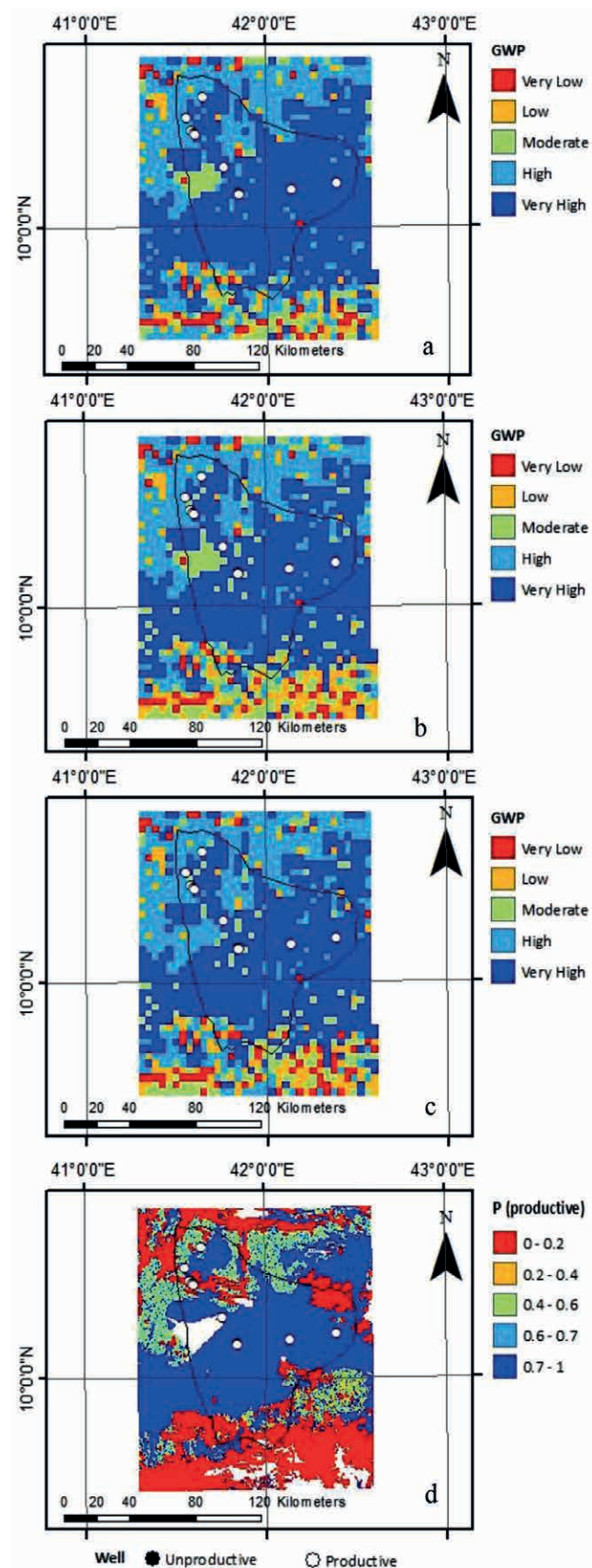


Figure 7

GWP maps for Shinelle, a – UNICEF model, b – AHP1, c – AHP2, d – ANN (probabilistic map)

The 4 × 4 priority vectors calculated from the PCMs and used to produce the AHP models are displayed in Table 6.

The GWP maps produced by the AHP models are shown in Fig. 5b and 5c, Fig. 6b and 6c and Fig. 7b and 7c. Both AHP models were equally accurate in predicting well outcomes, with equal accuracy to the UNICEF model in Dallol (Figs 8g and 8h) and greater accuracy in Halaba. The 69.7% accuracy of the AHP models in Halaba (Figs 8l and 8m) is comparable to that found by Rahmati et al. (2015), who also applied the AHP, with area under curve (AUC) = 72.7%. However, the accuracy of 62.2% (Figs 8g and 8h) observed in Dallol is considerably lower. The study by Rahmati was focused on a single area although it is unclear whether the experts in the study were instructed to base their judgements on the specific area. All models identified the same trends in each study area, indicating robustness of the UNICEF model as the solutions remain stable despite parameter perturbation. Whilst the UNICEF model is robust in this regard, the large difference in accuracy between the AHP models and the UNICEF model applied to Halaba suggests the UNICEF model is not optimal and should be calibrated by area. As the AHP1 and AHP2 models have equal accuracy in Dallol and Halaba, the difference in the overall accuracy of the AHP1 and AHP2 models (Figs. 8b and 8c) occurs due to results in Shinelle. Due to the small magnitude of the difference (1.1%) and the lack of unproductive wells in the available data for Shinelle, this difference is not considered significant.

LR

The accuracy of the LR model across all study areas is poor at 54.1% (Fig. 8d), performing worse than the UNICEF model. The maximum likelihood estimation used to calculate the β vector failed to converge with all 4 parameters included, due to insufficient variation of the lineament density values in the training set. As such, lineament density was removed from the explanatory variables and the regression was recomputed. The removal of the lineament density parameter contributes to the poor fit of the LR model by reducing the available information. Furthermore, due to the absence of unproductive wells a LR could not be computed for Shinelle as both positive and negative outcomes are required. The β vector for the LR is displayed in Table 7, where it can be seen from the negative value of the recharge coefficient that the LR directly contradicts the assumption that recharge should contribute to GWP.

The GWP maps produced using such a poor fitting model would be only marginally better than randomly assigning productivity outcomes, hence were not plotted. Although the fit of the model was very good in Dallol (Fig. 8i), the fact that none of the variables were significant at the 0.1 level means that this performance can be explained by chance and the model cannot be relied upon to provide suitable prediction elsewhere, evidenced by the worse than random fit in Halaba (Fig. 8n).

ANN

The relative parameter weightings of the ANNs are displayed in Table 8. The high predictive accuracies of the networks in spite of varying relative parameter importance indicate the solutions produced by the networks are robust. Whilst the relative importance values of the networks differed from the UNICEF model, due to the small sample size and distribution,

TABLE 5
Priority vectors calculated from 7 x 7 PCM

	Weights (Respondent 1)	Weights (Respondent 2)
Slope	0.32	0.34
LULC	0.05	0.22
Lineament density	0.12	0.11
Recharge	0.21	0.11
Proximity to surface water bodies	0.02	0.07
Permeability	0.25	0.12
Elevation	0.02	0.04

TABLE 6
Priority vectors calculated from 4 x 4 PCM

	Weights (Respondent 1)	Weights (Respondent 2)
Permeability	0.28	0.17
Slope	0.43	0.49
Recharge	0.20	0.17
Lineament density	0.09	0.17

TABLE 7
 β vector of LR coefficients

Parameter	β
Infiltration	0.071
Slope	-0.046
Recharge	-0.002
Constant	-0.118

combined with the unknown functions in the hidden layer, there was insufficient evidence to conclude that this was due to incorrect assumptions of the UNICEF model.

Network 3 had consistently high accuracy across all the study areas, outperforming the UNICEF model in Dallol and Halaba, and at least matching performance in Shinelle. Network 3 was used to produce the GWP maps shown in Figs 5d, 6d and 7d. The network performed similarly in Dallol and Halaba, with AUCs of 76.9% (Fig. 8j) and 77.8% (Fig. 8o), respectively. Whilst every prediction in Shinelle was correct there were no unproductive wells to predict, increasing the average accuracy across all study areas. The small variation in predictive accuracy between Dallol and Halaba suggests that the network may be applicable to new study areas without recalibration. When performance on the validation set across all areas was considered, the network had AUC of 85.5% (Fig. 8e), which is greater than the maximum accuracy of 81% found in the study by Lee et al. (2012) which focused on a single area.

The high predictive accuracies observed in Network 3 further validate the choice of parameters as GWP indicators, and the similarity between the solutions produced by the network and the UNICEF model indicate the current GWP maps are robust.

Network	Accuracy on training sample (%)	Accuracy on testing sample (%)	Infiltration	Slope	Recharge	Lineament density
1	81.0	78.9	0.428	0.159	0.084	0.330
2	76.2	89.5	0.383	0.242	0.073	0.302
3	78.6	89.5	0.324	0.209	0.122	0.346
4	78.6	78.9	0.310	0.221	0.075	0.394
5	76.2	78.9	0.349	0.136	0.051	0.464
6	76.2	78.9	0.365	0.240	0.066	0.329
7	76.2	84.2	0.315	0.112	0.072	0.500
8	76.2	78.9	0.315	0.164	0.106	0.415
9	78.6	84.2	0.381	0.185	0.074	0.360
10	78.6	78.9	0.265	0.208	0.080	0.447
Mean	0.784	0.821	0.343	0.188	0.080	0.389
Standard deviation	0.024	0.045	0.047	0.044	0.020	0.066

Comparison of models

The performance of the UNICEF model was satisfactory, with 62% accuracy over all three study areas (Fig. 8a). However, the accuracy in Halaba was poor, at only 56.1% (Fig. 8k). Both models derived from the AHP respondents performed satisfactorily, with total accuracies of 68.3% and 69.4% (Figs. 8b and 8c). Both AHP models matched the performance of the UNICEF model in Dallol, and substantially outperformed the UNICEF model in Halaba. The LR model had the worst performance, with accuracy of only 54.1% across all study areas (Fig. 8d), and is unstable with accuracies varying between 82.7% (Fig. 8i) in Dallol and 33.3% in Halaba (Fig. 8n). The best performing model in terms of overall accuracy and accuracy in each study area was the ANN, which represents the optimal solution with the available data. The ANN model had accuracy of 85.5% over all three study areas (Fig. 8e), with the accuracy remaining stable in Dallol and Halaba at 76.9% (Fig. 8j) and 77.8% (Fig. 8o), respectively. The ANN includes non-linear functions in the hidden layer, which likely account for the superior accuracy over the linear relationships modelled in the current, AHP and LR models, suggesting that GWP is governed by a non-linear relationship between the parameters which should be investigated further.

Robustness of UNICEF model

The AHP process determined that no additional parameters should be included in the model, indicating robustness. Clear correlation was observed between the AHP models and the UNICEF model, further indicating robustness. However, the best evidence of robustness is the correlation between the optimal solution of the ANN, and the UNICEF model. Despite severe parameter perturbations, with the inclusion of non-linear functions and a fundamentally different ranking of parameter importance, the correlation between the solutions of the ANN and the UNICEF model was clear across all study areas, providing strong evidence that the solutions provided by the UNICEF model are robust. Whilst the accuracy of the ANN was greater than

the UNICEF model, this must be considered against the limitations of this study.

The accuracy of the UNICEF model and the AHP models varied significantly between study areas, suggesting calibration by study area may be required for maximum accuracy. This finding is in agreement with the report produced by UNICEF (UNICEF, 2016a) on their exploration and drilling programme. However, the performance of the ANN model was stable across study areas suggesting that using such an approach may remove the requirement for calibration by study area.

Limitations

Well data were limited, despite knowledge of 227 wells only 61 could be used resulting in a small sample size of non-ideal distribution compared with similar studies which focussed on a single area.

The lack of unproductive wells in Shinelle is another limitation. Whilst the current and AHP models predicted all wells in Shinelle to be in areas of high or very high potential, and the ANN predicted all wells to be productive, the lack of negative wells prevents analysis of the specificity of the models. Furthermore, the absence of unproductive wells influences the overall predictive accuracy of all models.

The relaxation of the depth criterion from 150 m to 90 m limits the validity of the AUC interpretations as it is unknown whether wells unproductive at 90 m depth would be productive at 150 m depth. This relaxation influenced the training of the LR and ANN models, making the predictions more conservative as wells unproductive at 90 m were assumed unproductive at 150 m, despite the possibility they would be productive at greater depth.

A further limitation of the models considered is the inability to account for water quality. UNICEF are aware of high salinity in the Dallol and Shinelle regions, and the presence of fluoride in Halaba. Further research to develop models should seek to incorporate water quality as a parameter.

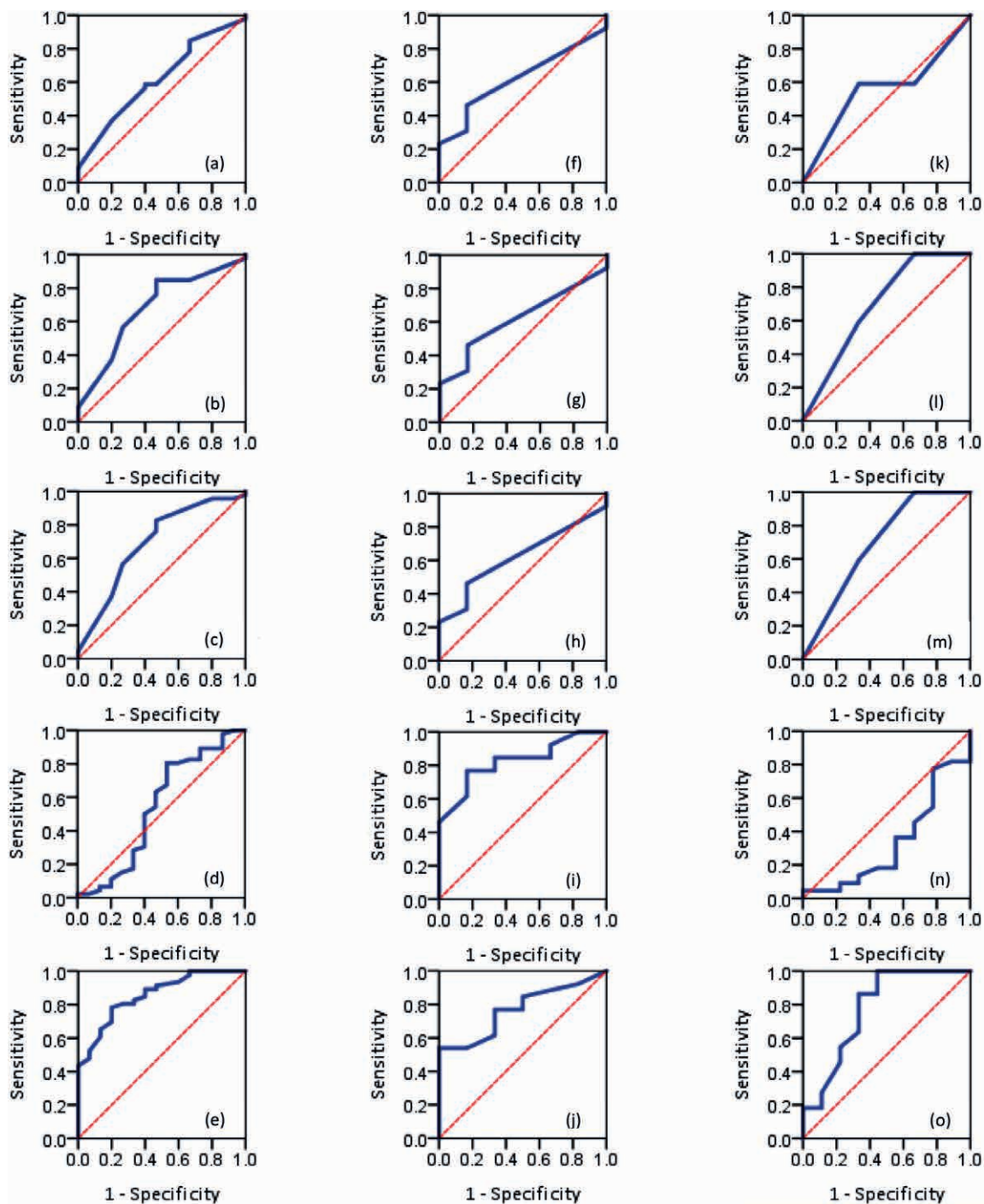


Figure 8

ROCs for (a) UNICEF model average over all study areas (AUC = 62.0%), (b) AHP1 average (AUC = 68.3%), (c) AHP2 average (AUC = 69.4%), (d) LR average (AUC = 54.1%), (e) ANN average (AUC = 85.5%), (f) UNICEF Dallol (AUC = 62.2%), (g) AHP1 Dallol (AUC = 62.2%), (h) AHP2 Dallol (AUC = 62.2%), (i) LR Dallol (AUC = 82.7%), (j) ANN Dallol (AUC = 76.9%), (k) UNICEF Halaba (AUC = 56.1%), (l) AHP1 Halaba (AUC = 69.7%), (m) AHP2 Halaba (AUC = 69.7%), (n) LR Halaba (AUC = 33.3%), (o) ANN Halaba (AUC = 77.8%).

CONCLUSIONS

The performance and robustness of the UNICEF model for groundwater exploration were assessed using several methods. Firstly, the ROC curve and AUC were calculated for the UNICEF model in order to determine predictive performance on the well data. Secondly, the AHP was applied to quantify the

expert judgement of hydrogeologists, in order to assess whether additional parameters should be included in the model, and to determine how the solutions varied under parameter perturbation to measure robustness. Thirdly, two numerical methods, LR and ANNs, were applied to determine optimal solutions for the available well data and to further assess the robustness of the current solutions.

The UNICEF model was found to be a satisfactory fit, with AUC = 62% (Fig. 8a) across all study areas. However, the accuracy varied between Dallol and Halaba with poor accuracy of 56.1% (Fig. 8k) observed in Halaba.

The AHP results suggest that the additional parameters 'elevation' and 'proximity to surface water bodies' were not important in determining the occurrence of deep groundwater, indicating robustness. Whilst both respondents' opinions differed over the importance of LULC, this was explained by the unfamiliarity of Respondent 2 with the assumptions of the UNICEF model. The AHP models were satisfactory fits, with AUC = 68.3% (Fig. 8b) and 69.4% (Fig. 8c) for the models from Respondents 1 and 2, respectively.

The LR model was a poor fit, due to failure of the maximum likelihood estimation to converge with all four parameters included, with AUC = 54.1% (Fig. 8d) across all study areas. Whilst high accuracy was observed in Dallol, this can be attributed to chance as all of the predictors were statistically insignificant at the 0.1 level.

The ANN model presented the optimum fit to the well data, displaying very good overall accuracy with AUC = 85.5% (Fig. 8e). The maps produced by the ANN and the UNICEF model correlate well, with both solutions identifying the same zones of relative high and low GWP. The correlation of the UNICEF model with the optimal solution is strong evidence for the robustness of the solutions produced by the UNICEF model.

Whilst the performance of models is useful in comparison, the GWP maps are the most important assessment of model performance and robustness, due to the limitations of the study.

Overall, the solutions produced by the UNICEF model have proven robust to moderate (AHP) and severe (ANN) parameter perturbation, and robust to the inclusion of additional parameters which were determined by experts to have low importance in determining the occurrence of deep groundwater.

Whilst this study has focused on data from Ethiopia, the findings are equally applicable to groundwater exploration internationally. The confirmation that the AHP can produce robust solutions in a data-scarce environment with complex hydrogeology should provide confidence to other WaSH programmes in Africa that combining data with the judgement of experts can improve drilling success rates. Furthermore, with the addition of readily attainable well data, including location, depth and yield, ANNs can be utilised to further increase drilling success rates. By improving drilling success rates and hence reducing costs, WaSH programmes will be able to provide potable water to more communities, ultimately increasing their impact.

ACKNOWLEDGEMENTS

This paper is dedicated to the memory of Dr Jeroen Ensink. The fieldwork was funded by The Chadwick Trust and UNICEF Ethiopia. The data for the study were provided by UNICEF Ethiopia and can be obtained from co-author Dr Samuel Godfrey subject to agreement.

REFERENCES

ADJI TN (2014) Identification of groundwater potential zones within an area with various geomorphological units by using several field parameters and a GIS approach in Kulon Progo Regency, Java, Indonesia. *Arab. J. Geosci.* 7 (1) 161–172. <https://doi.org/10.1007/s12517-012-0779-z>

ALTMAN DG and BLAND JM (1994) Diagnostic tests 3: receiver operating characteristic plots. *Brit. Med.J.* 309 188. <https://doi.org/10.1136/bmj.309.6948.188>

AGARWAL R (2016) Remote sensing and GIS based groundwater potential & recharge zones mapping using multi-criteria decision making technique. *Water Resour. Manage.* 30 (1) 243–260. <https://doi.org/10.1007/s11269-015-1159-8>

BURT CC and STROUD M (2007) *Extreme Weather: A Guide & Record Book*. W. W. Norton & Company, New York.

CENTRAL STATISTICS AGENCY (2007) *Population and Housing Census Report*. URL: <http://www.csa.gov.et/index.php/2013-02-20-14-51-51/2013-04-01-11-53-00/census-2007>

CORSINI A, CERVI F and RONCHETTI F (2009) Weight of evidence and artificial neural networks for potential groundwater spring mapping: an application to the Mt. Modino area (Northern Apennines, Italy). *Geomorphology* 111 (1–2) 79–87. <https://doi.org/10.1016/j.geomorph.2008.03.015>

DHAR A, SAHOO S and SAHOO M (2015) Identification of groundwater potential zones considering water quality aspect. *Environ. Earth Sci.* 74 (7) 5663–5675. <https://doi.org/10.1007/s12665-015-4580-7>

ETHIOPIAN GOVERNMENT (2016) The Somali National Regional State. URL: http://www.ethiopia.gov.et/en_GB/statesomali (Accessed August 2016).

GODFREY (2016) UNOCHA review of the El Nino induced drought. URL: <http://www.un.org/apps/news/story.asp?NewsID=52382> (Accessed April 2016).

GODFREY S and HAILEMICHAEL G (2016) Three-phase approach to improve deep groundwater supply availability in the Elidar District of Afar region of Ethiopia. *J. Water Sanit. Hyg. Dev.* <https://doi.org/10.2166/washdev.2016.021>

HARKER PT (1987) Derivatives of the Perron root of a positive reciprocal matrix – with application to the analytic hierarchy process. *Appl. Math. Comput.* 22 (2–3) 217–232. [https://doi.org/10.1016/0096-3003\(87\)90043-9](https://doi.org/10.1016/0096-3003(87)90043-9)

JASROTHIA AS, BHAGAT BD, KUMAR A and KUMAR R (2013) Remote sensing and GIS approach for delineation of groundwater potential and groundwater quality zones of western Doon Valley, Uttarakhand, India. *J. Indian Soc. Remote Sens.* 41 (2) 365–377. <https://doi.org/10.1007/s12524-012-0220-9>

JASROTHIA AS, KUMAR A and SINGH R (2016) Integrated remote sensing and GIS approach for delineation of groundwater potential zones using aquifer parameters in Devak and Rui watershed of Jammu and Kashmir, India. *Arab. J. Geosci.* 9 (4). <https://doi.org/10.1007/s12517-016-2326-9>

JHA MK, CHOWDARY VM and CHOWDHURY A (2010) Groundwater assessment in Salboni Block, West Bengal (India) using remote sensing, geographical information system and multi-criteria decision analysis techniques. *Hydrogeol. J.* 18 (7) 1713–1728. <https://doi.org/10.1007/s10040-010-0631-z>

KUMAR MG, BALI R and AGARWAL AK (2009) Integration of remote sensing and electrical sounding data for hydrogeological exploration—a case study of Bakhar watershed, India. *Hydrol. Sci. J.* 54 (5) 949–960. <https://doi.org/10.1623/hysj.54.5.949>

LEE S, SONG KY, KIM Y and PARK I (2012) Regional groundwater productivity potential mapping using a geographic information system (GIS) based artificial neural network model. *Hydrogeol. J.* 20 (8) 1511–1527. <https://doi.org/10.1007/s10040-012-0894-7>

MADRUCCI V, TAIOLI F and DE ARAUJO CC (2008) Groundwater favorability map using GIS multicriteria data analysis on crystalline terrain, Sao Paulo State, Brazil. *J. Hydrol.* 357 (3–4) 153–173. <https://doi.org/10.1016/j.jhydrol.2008.03.026>

MAGESH NS, CHANDRASEKAR N and SOUNDANAYAGAM JP (2012) Delineation of groundwater potential zones in Theni district, Tamil Nadu, using remote sensing, GIS and MIF techniques. *Geosci. Front.* 3 (2) 189–196. <https://doi.org/10.1016/j.gsf.2011.10.007>

MANDAL U, SAHOO S, MUNUSAMY SB, DHAR A, PANDA SN, KAR A and MISHRA PK (2016) Delineation of groundwater potential zones of coastal groundwater basin using multi-criteria decision making technique. *Water Resour. Manage.* 30 12 1–18.

NAMPAK H, PRADHAN B and ABD MANAP M (2014) Application of GIS based data driven evidential belief function model to

- predict groundwater potential zonation. *J. Hydrol.* **513** 283–300. <https://doi.org/10.1016/j.jhydrol.2014.02.053>
- NATIONAL PLANNING COMMISSION (2016) Growth and Transformation Plan II (2015/16 – 2019/20), Federal Democratic Republic of Ethiopia, Addis Ababa.
- OH HJ, KIM YS, CHOI JK, PARK E and LEE S (2011) GIS mapping of regional probabilistic groundwater potential in the area of Pohang City, Korea. *J. Hydrol.* **399** (3–4) 158–172. <https://doi.org/10.1016/j.jhydrol.2010.12.027>
- OZDEMIR A (2011) GIS-based groundwater spring potential mapping in the Sultan Mountains (Konya, Turkey) using frequency ratio, weights of evidence and logistic regression methods and their comparison. *J. Hydrol.* **411** (3) 290–308. <https://doi.org/10.1016/j.jhydrol.2011.10.010>
- RAHMATI O, SAMANI AN, MAHDAVI M, POURGHASEMI HR and ZEINIVAND H (2015) Groundwater potential mapping at Kurdistan region of Iran using analytic hierarchy process and GIS. *Arab. J. Geosci.* **8** (9) 7059–7071. <https://doi.org/10.1007/s12517-014-1668-4>
- RAZANDI Y, POURGHASEMI HR, NEISANI NS and RAHMATI O (2015) Application of analytical hierarchy process, frequency ratio, and certainty factor models for groundwater potential mapping using GIS. *Earth Sci. Inf.* **8** (4) 867–883. <https://doi.org/10.1007/s12145-015-0220-8>
- SAATY TL (1980) *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. McGraw-Hill, New York
- SRIVASTAVA PK and BHATTACHARYA AK (2006) Groundwater assessment through an integrated approach using remote sensing, GIS and resistivity techniques: a case study from a hard rock terrain. *Int. J. Remote Sens.* **27** (20) 4599–4620. <https://doi.org/10.1080/01431160600554983>
- UNICEF (2016a) *Geology of the project woredas*. WaSH Team UNICEF Ethiopia, Addis Ababa.
- UNICEF (2016b) *Hydrogeological study in Nine Drought Affected Woredas of Somali, Afar, Oromia and SNNP Regions*. WaSH Team UNICEF Ethiopia, Addis Ababa.
-