COST FUNCTIONS FOR PREDICTING CAPITAL EXPENDITURE OF SMALL TOWN WATER SYSTEMS IN GHANA

Kostnadsfunktioner för investeringar i vattenförsörjningssystem i Ghanas mindre tätorter

by KRISTOFFER RISTINMAA¹, KWABENA BIRITWUM NYARKO² and BISMARK DWUMFOUR-ASARE³ 1 Sjöbjörnsvägen 9, 11767 Stockholm e-mail: kriris@kth.se 2 Environmental Quality Engineering Section, Department of Civil Engineering, Kwame Nkrumah University of Science and Technology e-mail: nyarko.k.b@gmail.com 3 Department of Environmental Health and Sanitation, College of Agriculture Education,

Asante-Mampong Campus, University of Education, Winneba, Ghana e-mail: dwumfourasare@gmail.com



Abstract

Using data gathered by the WASHCost project in Ghana, this study uses cost functions to predict capital expenditure for construction of piped water systems for small towns in Ghana. The data was collected from 45 small towns in peri-urban areas in the Volta, Ashanti and Northern regions of Ghana. Correlation and multiple regression analyses are used to derive cost functions for both the complete and components of water systems. The aim is to derive functions for the total capital expenditure of a small town water system and for the fixed assets components: mechanised borehole, water reservoir, pipe work and stand post. The results are presented as eight functions, representing the total capital expenditure for a water system and capital expenditure for the infrastructure components water reservoirs, pipework and stand posts. Explaining variables identified are design population, length of pipeline, volume of reservoir and dummy variables for region (location). The functions vary in accuracy with an explanation coefficient R² varying from 0.42 to 0.88 with limited data, but the functions are indicative for the cost of water systems with mechanised boreholes which are increasingly being used as small town water supply systems in Ghana.

Key words - Ghana, small town water system, cost function, capital expenditure, multiple regression analysis

Sammanfattning

Genom att använda data, insamlad av WASHCost i Ghana, undersöks vilka kostnadsdrivare som kan användas för att förutsäga kostnaderna vid nybyggnation av ett dricksvattensystem. Vattensystemen som undersöks använder grundvattenteknik och är avsedda för mindre orter med populationer under 20000 invånare. WASHCost är ett pågående internationellt forskningsprojekt med holländska IRC som initiativtagare. Databasen täcker 45 tätorter ifrån tre ghananska regioner; Volta, Ashanti och Norra regionen. Huvudmålet med studien har varit att skapa en kostnadsfunktion för att i ett tidigt skede kunna få en uppfattning om de totala investeringskostnaderna vid nybyggnation av ett drickvattensystem. Sekundärt har funktioner skapats för att förutsäga kostnaderna för följande ingående komponenter i ett ghananskt vattennät: borrhål med automatisk pump, vattenreservoarer, rörledningsnät och vattenposter. För att skapa relevanta funktioner har sambandsanalys och multipel regression analys med olika modellalternativ använts. Dataunderlaget innehöll inte tillräckligt med information för att kunna skapa någon funktion för borrhålen. I resultaten redovisas åtta olika funktioner, varav två beskriver den totala investeringskostnaden. De vanligaste förekommande förklarande variablerna är: i vilken region vattensystemet är lokaliserat, den dimensionerande populationen, den totala rörlängden och vattenreservoarens volym. Förklaringsgraden, R², för funktionerna varierar mellan 0.42–0.88, vilket ger en stor variation i tillförlitlighet.

Introduction

Ghana, according to WHO/UNICEF, has already achieved the water target of the Millennium Development Goals, among very few developing countries considered to be on track in terms of water coverage. This progress covers all areas rural, small towns and urban in the country. In 2010, according to WHO's definition of improved water, 86% of Ghana's population is supplied by improved water (WHO/UNICEF, 2012).

Although the majority of Ghana's population is provided with water from improved systems, there are still 14% who are not served (Ibid). In rural areas the relative number without improved water is 20% in 2010, according to WHO/UNICEF (2012), and 40% in 2009 according to Moriarty et al. (2010). The difference in relative number depends on how you define improved water in terms of quality, quantity, fetching time etc. as service delivery indicators. In absolute numbers, the above mentioned figures represent between two to four million people without access to improved water systems. The country is still highly dependent on foreign donors. Up to 95% of the new capital investments in the water sector comes from developing partners (mostly through loans) (Moriarty et al. 2010).

One common type of improved water system, so called rural piped systems, for small towns and communities in Ghana, is a water system consisting of a secured mechanized borehole with connecting pipelines to storage tanks and stand pipes (Nyarko 2007). It is popular in the sector that the use of these piped schemes could accelerate and widen access to water service towards the 2015 target of 76% coverage. While the usage of this water scheme is extensive, 451 systems as at 2009, the costs vary a lot from one system to another (Moriarty et al. 2010). The wide range of cost makes it difficult for proper planning. Thus the benefit of cost functions will be to improve financial planning for the provision of water facilities in Ghana.

Several studies have been made in the field of cost estimation and water supply including those from Antoniolo and Filippini (2002), Tsegai et al. (2009), WRc (1977), Clark and Stevie (1981), Eilers (1984), Kirshen et al. (2004) and OECD (2005). Antoniolo and Filippini (2002) and Tsegai et al. (2009) targeted the maintenance and annual cost, whereas the others focus on the capital expenditure (CapEx) for water supply. This paper investigates only the capital expenditure for small town water systems in Ghana. The estimation modelling presented in this study is based on the use of multiple regression analysis.

Nkrumah et al. (2011) presented the results of an investigation of the cost drivers for small towns' water systems in Ghana and identified the main cost drivers for the evaluated water systems as technology, population density, hydrogeology of area and contract packaging. This paper similarly focuses on same small towns from the three regions in Ghana using the same database generated by WASHCost project hosted by the Kwame Nkrumah University of Science and Technology (KNUST) in Kumasi, Ghana, together with The International Water and Sanitation Centre (IRC). WASH-Cost is an international action research project and aims to increase the availability and use of cost information for both rural areas and small towns. WASHCost's database is used in this paper to derive cost functions to en-



Figure 1. Ghana in the world. Source: Wikimedia Commons.

able estimations of CapEx of small town water systems. The main target is a function to predict the total capital expenditure (TotCapEx), in other words the total cost for a small town water system and, secondly, functions for parts of the water system will be derived. The parts that are analysed are mechanised boreholes, water reservoirs, pipework and stand posts. The model will contribute to sector knowledge as a key decision making and planning tool for water service delivery in Ghana.

Method

Data

The main data source has been a database from WASH-Cost project in Ghana. Other sources of data include published reports, mainly from Ghana's Community Water and Sanitation Agency (CWSA). According to Nkrumah et al. (2011), the selection of regions was conducted so the data would fulfil the following criteria:

- A diversity of donors and development partners with sufficient information regarding capital cost
- A diversity of hydrogeological and hydro-climatic zones
- Different approaches of how to implement a water system

The regions selected were Northern region, Ashanti region and Volta region, see figure 2. Northern region has a relatively dry tropical climate and is located in a zone of savannah land. Ashanti region is located in the middle of Ghana and is mostly covered by forest with a hot and humid tropical climate. Volta region is located in the eastern parts and has Volta Lake as its western border.

The database comprises of 45 water systems and all quantities and cost information have been collected from "contract documents, bills of quantities, payment certifications and completion reports" (Nkrumah et al. 2011). Collected cost data are presented in the local currency, Ghana Cedi (GHs) and are divided into the six



Figure 2. Map of Ghana with the three selected regions marked in grey. Source: Wikimedia Commons (modified by the article's author).

categories (dependent variables): "borehole site works", "borehole", "pipework", "storage tank", "standpost" and "total water system". Furthermore 13 attributes or quantities are collected as variables for all the data points. The numerical data is presented with minimum and maximum values in Table 1 below.

Remaining data tells more about the characteristics of water systems (see Table 2 below).

Of the 45 systems, 16 are located in Ashanti, 16 in Volta and 13 in the Northern region. The systems were constructed and paid between 1998 and 2010. Steel tanks are only used in the Northern region and are also the only reservoir option used in the region. Three water

Table 1. Minimum and maximum values of collected numerical variables.

Variable	Min	Max	Unit
Number of communities	1	14	_
Design population	1 533	19 477	persons
Number of mechanized boreholes per system	0	4	_
Length of pipeline	1 323	35 040	m
Number of storage tanks per system	1	3	-
Volume of water reservoir per system	12	400	m ³
Number of standposts per system	4	29	_

Table 2. Presentation and brief explanation of the collected quantitative data.

Variable	Description
Regions	Where the system is located. The regions are Volta, Ashanti or Northern region.
Year of construction	When the system is constructed.
Transmission and distribution pipe length	Percentage of the total length of pipelines in a water system that operates for transmission versus distribution of water.
Type of reservoir	If a system uses ground concrete reservoir, high level concrete reservoir, steel or plastic tank. One water system can use multiple options.
Technical option	Sorts out if mechanized borehole is used or the system is provided with bulk water from Ghana Water Company Limited's (GWCL)
Power source	Sorts out if power comes from the national grid or from a solar system
Contract packaging	Depending on the source of funding and could either be International Competitive Bidding (ICB) or National Competitive Bidding (NCB).

systems use bulk water from GWCL and are the only systems using plastic tank. Only four water systems use solar system as power source and they are all located in Northern region. Contract packaging follows the variation of region. The only contract packaging presented in Volta region is National Competitive Bidding (NCB), whereas International Competitive Bidding (ICB) is the only contract bidding system used in Ashanti and Northern region. This study uses data from towns with populations between two thousand and twenty thousand, which makes this study investigating the costs for both peri-urban and small urban areas. Although the sizes of the observed towns divide the data into two categories of settlement, the characteristics of the water systems used are similar for the settlements. The majority of the water systems in small towns or smaller urban areas observed in this article comprise of a mechanized borehole, pipework for transmission and distribution, some kind of reservoir and several stand posts.

Deflated costs

Analytical tools

The data points in WASHCost's database are from water systems built between 1998 and 2010. Small town water systems infrastructure provision contracts normally allow for payments of fluctuations due to inflation. Because of this, deflators are used to adjust past cost to current values. The cost adjuster used due to the inflation is the annual growth of Consumer Price Index (CPI) provided by Ghana Statistical Service. CPI measures measures the annual changes in price of final goods and services used by households in the country.Together with the annual growth of Gross Domestic Product (GDP) deflator, the index is an established way to adjust for inflation. These two follow each other rather well, while one can also use another more deviating index called the Prime Building Cost Index (PBCI). The PBCI is a measure of inflation in the construction industry while the CPI gives a measure for the whole economy. WASHCost has previously used both the PBCI deflator in Nkrumah (2010) and GDP deflator in other publications. From WASHCost database, the deflator based on the CPI gives a lower inflation rate over time. The two indexes (PBCI and CPI) are presented accumulated from 1997 to 2011 in figure 3. The difference will give lower costs if the CPI is used, which is done in present article. The costs will be particularly lower further away from year 2011, which is used as base year. Both CPI and PBCI are provided by Ghana's Statistical Service.

Correlation analysis

In this study, an approximation of the correlation coefficient, ρ_{xy} , was estimated. Pairs of variables are taken from n data points and are placed in a series of pairs. Then the estimation, here called r_{xy} , can be calculated

$$\rho_{xy}^{*} = r_{xy} = \frac{c_{xy}}{s_{x} \cdot s_{y}} = \frac{\sum_{i=1}^{n} (x_{i} - \bar{x}) (y_{i} - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_{i} - \bar{x})^{2}} \sqrt{\sum_{i=1}^{n} (y_{i} - \bar{y})^{2}}}$$
(1)

If selection of interesting independent variables is the first use of a correlation analysis; the second use for this study is to find out which variables really are independent. If the correlation is very high between two independent variables, then probably they are not really independent and indicates that both cannot be used in the same multiple regression analysis. In other word, the correlation analysis helps to indicate problems of multicollinearity (Montgomery et al. 2006).

Multiple regression analysis

Through a regression analysis, the variety of an independent variable are sought to be described by explaining variables. According to previous studies (WRc (1977), Clark and Stevie (1981), Eilers (1984), Kirshen



Figure 3. Comparison of the two deflators based on PBCI (grey line) and CPI (black line). The black lower line (CPI deflator) is used in present study. 1997 = GHS 100, base year. Source: Ghana Statistical Service (2011) and World Bank (2013).

et al. (2004) and OECD (2005), the most common function used to derive functions for capital costs is the so-called log-log model. Beside the log-log model, a regular linear model is tested and in the cases where one explaining variable seems to be highly correlated to the capital expenditure in a polynomial way, a polynomial model is tested.

Linear model:

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_n x_{ni} + \varepsilon_i$$
(2)

Log – Log model:

$$y_i = \alpha \cdot x_{1i}^{\beta_1} \cdot x_{2i}^{\beta_2} \cdot \ldots \cdot x_{ni}^{\beta_n} \cdot \varepsilon_i$$
(3)

Polynomial model:

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_i^2 + \dots + \beta_n x_i^n + \varepsilon_i$$
(4)

Where

 y_i = dependent variable from data point *i*, *i* = 1,2...., *n* x_{1i} , x_{2i} , ..., x_{ni} = explaining variables from data point *i* α = constant to be estimated

 $\beta_1, \beta_2, \dots, \beta_n$ = coefficients to be estimated, relating to explaining variables x_1, x_2, \dots, x_n

 ε_i = random error for data point *i*, $\varepsilon_i \in N(0,\sigma)$

The key parameters used to analyse the accuracy and precision of a model are as follows:

- Significance level Describes at what level all the computed factors (β_i) and constants (α) in the model are significant apart from zero. Lower than 0.05 is considered as a decent significance.
- Residual Analysis To ensure that the assumption of error, , is correct. Namely the error should be random, no patterns, and normal distributed (plots are not presented).
- Coefficient of determination (R^2) To compare and valuate how well a function describes the cost. 0–1 where 1 means that the cost is fully described by the function.

VATTEN $\cdot 1 \cdot 13$

• Root Mean Square Error (RMSE) = $\sqrt{\frac{\sum_{i=1}^{n} \Delta y^{2}}{n}}$. To describe the variation around the function. If RMSE is large, then the outcome of a function has to be considered with a large interval. The CV(RMSE) is also calculated which simply is the RMSE divided by the mean CapEx.

Almost half of the variables in the study are of a qualitative nature and dummy variables have been used as numeric stand-in for these qualitative data. As an example, the "region"-variable has three qualitative options. If data point X is located in Ashanti region, the dummy variable "Ashanti region" is 1 and the variables "Northern region" and "Volta region" are 0. For mathematical reasons, 0 and 1 have to be changed to 1 and 2 when a log-log model is applied.

Results and discussion

This section describes the estimated results from the analysis. Since the analysis failed to derive a function for predicting CapEx for mechanised borehole, there is no such function presented in this section.

Almost all the results are presented with two recommended cost functions for each analysed object. The *post-design* function has a better accuracy, but has such variables that a first design investigation must take place before the function can be used. The *initial* function is therefore derived to get an estimate of the cost before even an investigation of the location for the forthcoming water system has taken place. All the functions are presented in thousand (1000) Ghana Cedis (GHs, 2011). All the used parameters to evaluate the accuracy and precision of the function are listed in Table 8.

Table 3. Data summary for the variables used in the functions to predict TotCapEx.

Variable	Label	Unit	Min.	Max.	Mean	Median	Std.Dev.
Total Capital expenditure for a Water system	COST	GHS	58 109	960 748	494 150	499 217	235 352
Dummy variable for Volta region. 1 if Volta, 0 if other	REG	_	0	1	0,33	0	0,48
Dummy variable for Northern region. 1 if Norther	n,						
0 if other	REG	_	0	1	0,31	0	0,47
Design population	POP	Persons	1 533	11 493	5 288	4 7 4 6	2 738
Total length of pipelines	PIPE	m	1 323	17 574	6 328	5 188	4 192
Total volume of storage tanks in a water system	VOL	m ³	12.0	200.0	77.5	75.0	39.2

Small town water system

The data used for deriving the two cost functions to explain TotCapEx is summarized and presented in Table 3.

Post-design function

The function to explain TotCapEx with best accuracy contains the explaining variables length of pipeline, volume of tank and dummy variables for Volta region and Northern region. The recommended function to estimate TotCapEx for a small town water system is:

$$TotCap \ Ex (GHS '000) = 338 + REG + 0.0178 \cdot PIPE + 1.26 \cdot VOL$$
(5)

Where REG: If Volta region, then REG = -243, if Northern region, then REG = 86., if other region, then REG = 0, PIPE = Total length of pipelines (m), and VOL= Total volume of storage tanks (m³).

As presented in Table 8, the level of significance does not reach the commonly used level of 0.05. In fact, 70% of TotCapEx are explained by the first function with above mentioned input parameters. The CV (RMSE) is 25.7% of GHS 494 150, the mean of observed TotCapEx. The values are, compared with the other functions in the study, in the upper range. The predicted costs are slightly overestimated for most of the data points from Ashanti region, but the deviation where not significant enough to explain the predicted cost in the function. No patterns were detected by analysing the residual plots.

Initial function

The second cost function depends on the variables for Volta region and design population:

$$TotCap \ Ex \ (GHS \ '000) = 420 - 345 \cdot REG + 0.0359 \cdot POP$$
(6)

Where REG: If Volta region, then REG = 1, If other region, then REG = 0, and POP = Design population in persons.

Compared to the post-design function, this function has the advantage to be useful before even starting to design a water system. On the other hand, the initial function uses explaining variables which give less accuracy as showed in Table 8. R^2 decreases while the RMSE increases. The only favourable parameter is the lower significance level.

Water reservoirs

The results from analysing the costs of water reservoirs are split up and presented separately for steel tanks and for reservoirs made of concrete. No function is presented for plastic tanks due to the lack of (only three) data points with plastic tanks. For water reservoirs made of concrete only one function is presented due to the inferior reliability even of the post-design function. There is no reason to present an initial function which would have been even less accurate.

Steel tank

The data used for deriving the two cost functions to explain CapEx for building steel tanks is summarized and presented in Table 4 below.

Table 4. Data summary for the variables used in the functions predicting CapEx for steel tank.

Variable	Label	Unit	Min.	Max.	Mean	Median	Std.Dev.
Capital expenditure for steel tank	COST	GHS	150 872	226 180	176 522	161 482	27 943
Volume of steel tank	VOL	m ³	50	150	85	80	34
Design population for the water system	POP	person	2 207	11 493	5 482	3 098	3 772

 Table 5. Data summary for the variables used in the function predicting CapEx for reservoirs made of concrete.

 Variable
 Label
 Unit
 Min.
 Max.
 Mean
 Media

Variable	Label	Unit	Min.	Max.	Mean	Median	Std.Dev.
Capital expenditure for a reservoir made of concrete Size of the reservoir	COST VOL	GHS m ³	15 829 12.0	148 222 200.0	74 289 78.8	65 865 75.0	32 746 41.9
Dummy variable for type of reservoir. 2 if CT, 1 if GCT	TYPE	_	1.00	2.00	1.89	2.00	0.32
Dummy variable for region. 2 if Volta region, 1 if other region	REG	_	1.00	2.00	1.44	1.00	0.51

Post-design function

The recommended function uses the designed total volume of a steel tank and is written:

$$CapEx (GHS'000) = 110.28 + 0.78 \cdot VOL$$
 (7)

Where VOL = Designed volume of steel tank in m^3 .

Apart from the small number of observations, which is a weakness, this function has the strongest values of all functions in the article. This is especially observable by looking at the CV(RMSE) and the R², see Table 8. The residuals do not have any specific pattern.

Initial function

To initially estimate the cost for a steel tank, the recommended function is:

$$CapEx(GHS'000) = 145.57 + 0.00564 \cdot POP$$
 (8)

Where POP = Design population in persons.

The function, with only the design population as explaining variable, is far from the post-design function regarding accuracy and explanation degree. Table 8 shows that the R^2 drops to 0.58, but using design population as the only variable still gives a very low significance level. The user of the derived functions should consider possibilities to estimate the total need of storage volume in the study area to enable usage of the post-design rather than the initial function.

Reservoirs made of concrete

The data used for deriving the two cost functions to explain CapEx for building reservoirs made of concrete is summarized and presented in Table 5.

Post-design function

The final recommended function for deriving cost for a concrete reservoir is built on the explaining variables: dummy for Volta region (reduces the cost), the volume of the reservoir and whether it is a normally elevated concrete tank (CT) or it is a ground concrete tank (GCT). The function is:

$$CapEx (GHS'000) = 8519 \cdot VOL^{0.37} \cdot TYPE^{0.05} \cdot REG^{-0.66}$$
(9)

Where VOL = Designed volume of concrete reservoir in m^3 , TYPE = 2 if concrete tank, 1 if concrete ground reservoir, and REG = 2 if Volta region, 1 if other region.

The function to derive cost for a concrete reservoir has the poorest key values of all investigated assets, see Table 8. The function needs a pre-investigation of the study area to be useful and still, the output from the function is very uncertain. Therefore the function is only recommended in the sense of being the best possible from the input data. It does not mean that we recommend using the function to determine the potential costs. A comparison between observed and predicted costs reveals the lack of coherence between them and underlines the uncertainty of the function.

Pipework

The data used for deriving the two cost functions to explain CapEx for building a pipeline network for water supply in a small town is summarized and presented in Table 6 below.

Table 6. Data summary for the variables used in the functions predicting CapEx for pipeworks.

Variable	Label	Unit	Min.	Max.	Mean	Median	Std.Dev.
Capital expenditure for pipework Dummy variable for region, 1 if Volta region,	COST	GHS	14 870	455 360	156 740	137 880	103 060
0 if other region	REG	_	0	1	0.36	0	0.48
Total length of pipelines	PIPE	km	1.32	17.57	6.25	5.19	4.16
Design population	POP	persons	1 533	15 942	5 469	4 746	3 162

VATTEN · 1 · 13

Post-design function

CapEx for pipework in a small town water system is best explained by the total length of pipelines in the system and if the study area is located in Volta region. The function is written:

$$CapEx (GHS'000) =$$

61.21 + 19.5 · PIPE - 74.50 · REG (10)

Where PIPE= Total length of pipeline in km and REG= 1 if Volta region, 0 if other region.

The number of observations is similar to all the other functions, except for the water reservoirs. The significance level is sufficiently low and the cost is well explained by the two input variables with R^2 =0.80. An objection to use the function is the scattered residuals, especially some of the outliers that bring the values down for all the key parameters, e.g. CV(RMSE) = 29%, see Table 8.

Initial function

The initial function is written:

 $CapEx (GHS'000) = 472.15 \cdot POP^{0.70} \cdot REG^{-1.20}$ (11)

Where POP = Design population in person and REG = 2 if Volta region, 1 if other region.

As for all system characteristics above, the initial function for pipework does not reach the same accuracy as the post-design function does, see Table 8. Although the significance level is decent and the R^2 might be considered acceptable, the RMSE shows that the residuals are very scattered. This will give a very wide interval for the predicted cost. The real prediction interval would be larger than the derived cost plus/minus the RMSE and already the RMSE interval will give a multiplicative interval between 0.59 and 1.41 of the predicted cost.

Stand post

Surprisingly, the design population gave better values than the number of stand posts in the multiple regression analysis and because of the high correlation between them the model just got significant with one at a time. Therefore, the final cost function depends on the variables number of communities and design population. As the function is the best possible, considering the data given and the models tested, and at the same time uses variables which do not depend on a pilot study, it is presented as the only function in *Initial function*. The data used for deriving the cost function to explain CapEx for building the stand posts within a small town water system is summarized and presented in Table 7 below.

Initial function

The recommended function can be used without knowing more than the design population and the number of communities included in the water system. The function is written

$$Cap Ex (GHS'000) = 33.85 \cdot COM^{0.20} \cdot POP^{0.74}$$
 (12)

Where COM = Number of communities and POP = Design population in person.

The key parameters in Table 8 reveal that the function estimates the CapEx in a reasonable way. The values are not optimal but as a function for prediction, the function may give a first indication of the cost range. Again, the CV(RMSE) indicates quite scattered values.

Final discussion and conclusions

Most of the finally derived functions used data that fits best into a normal linear model instead of the multiplicative log-log model. This deviates from most of the previous studies, where the log-log model or versions of the log-log model are used, e.g. WRc (1977), Clark and Stevie (1981), Kirshen et al. (2004) and Eilers (1984). Three out of eight functions derived are best fit in a loglog model. The coefficient of determination, R², ranges from 0.42 to 0.88. This is in accordance with the results in similar studies by WRc (1977) and Clark and Stevie (1981). Kirshen et al. (2004) had over 1300 data points in their study and got an R^2 above 0.90 for all their functions. Meanwhile, no R² was reported in Eilers' (1984) or OECD's (2005) reports, which makes it difficult for any meaningful comparison with current findings. The level of significance is consistently lower than

Table 7. Data summary for the variables used in the functions predicting CapEx for stand posts.

Variable	Label	Unit	Min.	Max.	Mean	Median	Std.Dev.
Capital expenditure for stand posts Number of communities included in the water	COST	GHS	7 215	49 700	21 457	17 677	12 386
system Design population	COM POP	_ persons	1 1 533	14 15 942	1.75 5 577	1 4 882	2.28 3 130

Table 8. Key parameters for validation of the resultant cost functions.

Element	Function	Equ.nr.	Obser- vations	Significance level	R ²	R	RMSE (GHs)	CV (RMSE)
Small town water system	Post-design	5	42	< 0.10	0.70	0.84	127 067	25.7 %
	Initial	6	42	< 0.01	0.65	0.81	137 111	27.7 %
Steel tank	Post-design	7	13	< 0.01	0.88	0.94	9 123	5.1 %
	Initial	8	13	< 0.01	0.58	0.76	17 382	9.8 %
Reservoirs of concrete	Post-design	9	27	< 0.10	0.43	0.65	26 750	36.0 %
Pipe work	Post-design	10	42	< 0.01	0.80	0.89	45 455	29.0 %
	Initial	11	42	< 0.01	0.54	0.74	63 812	40.7 %
Stand post	Initial	12	44	< 0.05	0.64	0.80	6 991	32.6 %

0.10 (often <0.01) in present study which is reasonable. Although the conventional level of significance to form firm arguments is 0.05, this is still arbitrary (Stigler 2008), which leads to the conclusion that a function with level 0.10 should not be immediately rejected. Instead it is important to evaluate the significance level together with all other key parameters, such as the RMSE and the R^2 . The residual analysis is used to reject function where the errors did not fulfil the criteria to be independent each other and normal distributed. All the functions presented are considered to fulfil the criteria of independence and normally distributed errors.

Overall, it has been possible to derive cost functions for a small town water system in Ghana especially the post-design functions for TotCapEx and other CapEx which may be interesting to use, due to their better key values. These functions normally require a first investigation and design of the intended building area, but will on the other hand give more reliable cost predictions. However, in two cases it has not been possible to create functions adequate to predict the cost. In one of the cases, the key values of the function for concrete reservoirs are not good enough and the other case is the lack of a function for mechanized boreholes. The data was insufficient to derive a function to predict the cost for mechanized boreholes. The initial functions, that are intended to operate as a forecaster in the early stage of a new construction project, should be used sparsely and critically. Arguing from the key parameters, the initial function for TotCapEx, CapEx for steel tank and the function for stand posts are the functions with enough accuracy to use as proxy predictions. One might also question if 13 observations are enough for the functions to predict CapEx for steel tank, but then again the number of observations has taken into account when the models has been derived and when calculating the key parameters.

The design population is used in all the initial functions to determine CapEx. Population is one of few variables that are available before planning a new water system and is normally, together with region, the one with most correlation to capital cost. As could have been presumed, the design population does not explain all the variability of CapEx. Nkrumah et al. (2011) suggests that the population density of an intended construction area probably could explain the cost in a good way, but data for population density is not yet collected for the observed data points. Moreover, the authors suggested pipe length/capita as a proxy for the population density, but this requires knowing the length of the designed pipework and cannot be used before planning.

As the results shows, the regions often occur as explaining variables. Particularly the Volta region is part of most functions and consequently reduces the cost prediction. Here it is important to stress that explaining versus responding variables do not imply a clear causal relationship. It is easy to understand that a region in itself cannot drive the cost; instead it is the properties of a region which drive the cost. The problem is to determine what properties that might inflate CapEx. The differences between the regions, based on the criteria for collecting the data, are possible answers to the question of what drives the cost. However, this cannot be concluded in this paper. Contract packaging is an example of a possible cost driver, already concluded by Nkrumah et al. (2011). For example, NCB is the only contract packaging used in Volta region, which means that it would have worked identical with a dummy variable for Volta region in a function. Still, we chose Volta region as explaining variable due to the more general character of a region. NCB as a contract package is not represented in the other two regions, Northern and Ashanti, which prevents a comparison across the regions which implies that other factors cannot be excluded as cost drivers.

Moreover, there is one more problem with using region as an explaining variable; it gets more complicated to predict the cost for a water system in another region than the observed. If the functions are used for another region, the circumstances have to be well examined to ensure that the region has similar conditions as one of the regions in this study.

The choice of index for cost adjustment over time has an impact on the results as well. Due to the used index, CPI, CapEx for older construction is estimated lower than if PBCI deflator from Ghana's statistical service would have been used.

Suggestions for further work

Based on the study, the following recommendations are made for further research:

- 1. Cost drivers of mechanised boreholes to determine the relationship between cost, borehole depth, diameter and casing length are used.
- Conduct more qualitative analyses to determine what really causes an inflation of CapEx instead of the regions.
- 3. Conduct a life-cycle cost analysis where functions are derived for both CapEx and recurrent cost (operational and capital maintenance costs).

Acknowledgement

The Authors wish to acknowledge the support of the WASHCost Project in Ghana for providing the database for analysis and also the Swedish International Development Cooperation Agency (Sida) who sponsored the lead author Kristoffer Ristinmaa with "Minor Field Study"-scholarship to carry out field work in Ghana.

The WASHCost Ghana is a five-year action research project investigating the cost of providing water, sanitation and hygiene services to rural and peri-urban communities in Ghana, Burkina Faso, Mozambique and India (Andra prdesh) (see www.washcost.info).

References

Antonioli, B. and Filippini, M. (2002) The use of a variable cost function in the regulation of the Italian water industry. University of Lugano, Switzerland. Link 2012-06-07: http://doc.rero.ch/lm.php?url=1000%2C42%2C6%2C2 0051021110309-QZ%2F1_wp0201.pdf

- Clark, R.M. and Stevie, R.G. (1981) A Water Supply Cost Model Incorporating Spatial Variables. Land Economics. Vol. 57. No. 1, pp. 18–32. University of Wisconsin.
- Eilers, R.G. (1984) Cost equations for small drinking water systems. US Environmental Protection Agency, EPA.
- Ghana Statistical Service (GSS) (2011) Consumer Price Index (CPI). Link 2013-02-02: http://www.statsghana.gov.gh/ docfiles/CPI%20Release_pdf/cpi_national_time_series_ jan1970-sept2012.pdf
- Kirshen, P.H., Larsen, A.L., Vogel, R.M., Moomaw, W. (2004) Lack of influence of climate on present cost of water supply in the USA. Water Policy 6 ss 269–279.
- Montgomery, D.C., Peck, E.A., Vining, G.G. (2006) Introduction to linear regression analysis. (4th ed.) USA: Wileyinterscience.
- Moriarty, Naafs, Pezon, Fonesca, Uandela, Potter, Batchelor, Reddy, Mekala. (2010) Working paper 1 – WASHCost's theory of change: reforms in the water sector and what they mean for the use of unit costs. IRC – International Water and Sanitation Centre. Link 2012-06-07: http://www. washcost.info/page/753
- Nkrumah, E., Nyarko, K.B., Dwumfour-Asare, B., Oduro-Kwarteng, S. and Moriarty, P. (2011) Drivers of capital expenditure of rural piped water systems in Ghana: The Volta, Ashanti and Northern region, Link 2012-06-07: http://rwsnforum.files.wordpress.com/2011/11/194dwumfour-ghana-long-paper-docx.pdf
- Nyarko, K.B. (2007) Drinking water sector in Ghana Drivers for performance. Netherlands: Taylor & Francis Group.
- OECD EAP Task Force Secretariat (2005) Rural cost functions for water supply and sanitation. Link 2012-06-07: http:// www.oecd.org/dataoecd/18/12/36228167.pdf
- Stigler, S. (2008) Fisher and the 5% level. Chance 21 (4): 12. Link 2012-06-05: http://www.springerlink.com/content/ p546581236kw3g67/
- Tsegai, D.W., Linz, T., Kloos, J. (2009) Economic analysis of water supply cost structure in the Middle Olifants sub-basin of South Africa. Center for Development Research ZEF Bonn, Germany. Link 2012-06-07: http://www.zef.de/fileadmin/webfiles/downloads/zef_dp/zef_dp_129.pdf
- WHO/UNICEF World Health Organization & UNICEF (2010). Progress on drinking water and sanitation: 2012 update. Link 2013-01-21: http://www.who.int/water_sanitation_health/publications/2012/jmp_report/en/index. html
- World Bank (2013) World DataBank. Link 2013-02-02: http://databank.worldbank.org/data/
- WRc Water Research Centre (1977) TR61 Cost information of water supply and sewage disposal.