Domestic Roofwater Harvesting
Research Programme

DTU

RN – RWH04

RWH Performance Predictor for Use with Coarse (i.e. Monthly) Rainfall Data

by

Terry Thomas

This document is an output from a project funded by the UK Department for International Development (DFID) for the benefit of developing countries. The views expressed are not necessarily those of the DFID.

CONTENTS

1. THE NEED TO MODEL RWH SYSTEM PERFORMANCE

In every roofwater harvesting system there is a trade-off obtainable between increasing system performance and increasing system cost. The medium of this trade-off is normally tank size and hence at the centre of RWH design is the optimisation of that size. Many methodologies for tank sizing have been published, several are reviewed by (Gould & Nissen-Petersen, 1999). Fortunately, as with all optimisation, the plot of performance (e.g. cost:benefit) against tank size is 'flat-topped' in the area of interest, so that a ±10% variation in size there has little influence on economic performance. The cost of a tank of given size can usually be readily assessed, but the performance of the system containing that tank cannot. We therefore seek a methodology for predicting performance over a system's expected life: one that suits both the surrounding constraints (e.g. of data availability and access to computing facilities) and the RWH system use envisaged.

It is only possible to *roughly* predict the performance of a RWH system. Performance depends upon many factors, so its prediction can only be as good as knowledge of those driving factors. Some of these like tank size and roof size, once selected, remain conveniently constant. By contrast future user-demand behaviour and climate are uncertain. Water demand may vary widely with house occupancy, social calendar and season. Rainfall is hard to forecast more than two days ahead. To model system performance we therefore essentially average such variables in some way, assuming for example uniform water demand or a climatically 'typical' year. Normally we use the past as a template for the future. Since the critical factors can be estimated but crudely, we will do well to predict performance measures within 5% accuracy.

There are many such measures, including *reliability* (fraction of days that demand is met), *satisfaction* (fraction of demand volume that is met), *efficiency* (fraction of run-off water that is used) and *water value*. These measures can be applied to a typical year, to a typical wet or dry season or to an exceptional year/season, such as the driest in the last decade. They can be expressed for a representative location or for a particular one – for which meteorological data must then be available.

Suitable meteorological information is actually, at least in developing countries, rarely available and affordable in the right form and for the exact location of interest. This note addresses the specific problem of inadequately rainfall data.

RWH performance models are generally based on 'mass' balances. At each time-step, the roof run-off belonging to that step is added to the volume (mass) in the tank and the user's draw-off is subtracted. Tests and corrections are applied to cover the three cases 'tank overflows', 'tank runs dry' and 'demand exceeds the water available'. The time-step may be 1 day or 1 month: which of these is appropriate is discussed below. There are further fine modelling details to be decided, for example whether inflow is assumed to *precede* draw-off within any one time-step or to follow it.

A distinction needs to be made between RWH systems with respectively *no* storage, *some* storage and *very large* storage, since modelling them has quite different data needs. Three flows are to be modelled - input, usage and overflow - normal interest is in usage and occasionally in overflow. Modelling a 'no storage' system is simple: all daily inputs in excess of daily demand are spilled. Both *volumetric demand satisfaction* and *reliability* can be simply calculated directly from the rainfall record. Modelling a 'very-large-storage' system is even simpler. If annual roof run-off R_a exceeds annual water demand D_a then *satisfaction* and *reliability* will be 100% and *efficiency* will equal D_a/R_a. Conversely if R_a is less than D_a then *satisfaction* and *reliability* will fall to R_a/D_a , while *efficiency* will rise to 100%.

The situation however becomes more complex when the storage volume can be approximated neither to zero nor to infinity. This situation is addressed by the rest of this Note.

2. DATA NEEDS OF RWH PERFORMANCE MODELS

The list of data required to run a performance-prediction model comprises

- roof area and a roofwater run-off coefficient (for the latter, a constant such as 0.85 is often used to approximate the very variable behaviour found in practice)
- nominal ('standard') daily water demand D
- the management strategy the user proposes to use for selecting demand on a particular day as a multiple, greater or less than 1, of nominal demand
- a past rainfall record long enough to act as a reliable guide to future precipitation patterns
- proposed tank size V

We may express the tank size via an associated water residence time T, defined as tank volume divided by nominal daily demand ($=V/D$). Generally the rainfall data to accurately drive any model of that tank needs to be expressed in time steps shorter than T. Low-cost (and thus low-security) RWH systems having tanks of under say 15 day's capacity (T<15) need daily data sets, whereas high-security RWH systems with '3-month' tanks may be modelled with monthly data.

The length of the requisite rainfall record depends mainly on the level of supply reliability sought. Large RWH systems that constitute a supply of last resort in arid areas must be modelled (Vyas V., 1999)with long data sequences (say 25 years) in order to pick up extreme climatic events. Low-cost systems can usefully be modelled with 5 or 10-year sequences or with incomplete records where corresponding portions of previous years' data may have to be pasted into gaps.

The desirability of using *daily* rainfall data in RWH modelling has been noted (Heggen, 1993). Table 1 explores the error introduced by employing monthly instead of daily data to drive a RWH performance model. For four climate types the model was driven firstly with daily rainfall and then with rainfall produced by uniformly spreading each month's rainfall uniformly across the days of that month. In both cases the average *reliability* over 10 years was computed. The 'error' tabulated is the difference between the two reliability estimates.

For the large tank, both with a low demand (D = 0.6 x mean daily roof run-off) and an unmeetably high demand (D = 1.2 x mean runoff), the use of monthly data introduces negligible error. An even larger tank would raise the *reliability* with the high demand towards its theoretical maximum of 83% and get an even close agreement between forecasts using actual and 'uniform' daily data. For the medium and small tank sizes however, there is a significant error introduced by using synthetic uniform ('monthly') data instead of actual daily rainfall.

Table 1. Computed *reliability* of RWH supply as a function of rainfall data used in model

Mean run-off = constant 100 l/day.
Demand = constant.
Measures are averaged over 10 years.
Forecasts using *actual* daily data compared with those using *uniform* daily rainfall = monthly / 30

Location, annual rainfall and climate type	Rainfall data used for model	Small tank, 700 l		Medium tank, 3000 l		Large tank, 12000 l	
		Nominal daily demand as percentage of mean daily roof run-off					
		60%	120%	60%	120%	60%	120%
Saiya, W Kenya	actual daily	0.79	0.51	0.97	0.69	1.00	0.76
1507 mm/year	uniform	0.83	0.39	0.99	0.51	1.00	0.76
Double rains	*error %*	*4%*	*-12%*	*2%*	*-18%*	*0%*	*0%*
Bangkok	actual daily	0.58	0.42	0.73	0.57	1.00	0.76
1500 mm/year	uniform	0.58	0.45	0.71	0.56	1.00	0.76
Monsoon climate	*error %*	*0%*	*3%*	*-1%*	*-2%*	*0%*	*0%*
Panama	actual daily	0.71	0.50	0.83	0.66	1.00	0.76
1500 mm/year	uniform	0.71	0.47	0.83	0.59	1.00	0.76
Long single rains	*error %*	*0%*	*-3%*	*0%*	*-7%*	*0%*	*0%*
Petrolina, Brazil	actual daily	0.45	0.27	0.65	0.47	0.96	0.76
mm/year	uniform	0.55	0.38	0.71	0.48	0.98	0.76
Semi-arid	*error %*	*11%*	*11%*	*6%*	*0%*	*2%*	*0%*

Notes: (a) Reliability is the fraction of days in 10 years that demand was met.

(b) The error measures the bias in the estimates of supply reliability introduced by using coarse (monthly) rainfall data instead of fine (actual daily) data.

(c) Shaded cells indicate reliability less than 50% and therefore not very suitable for RWH

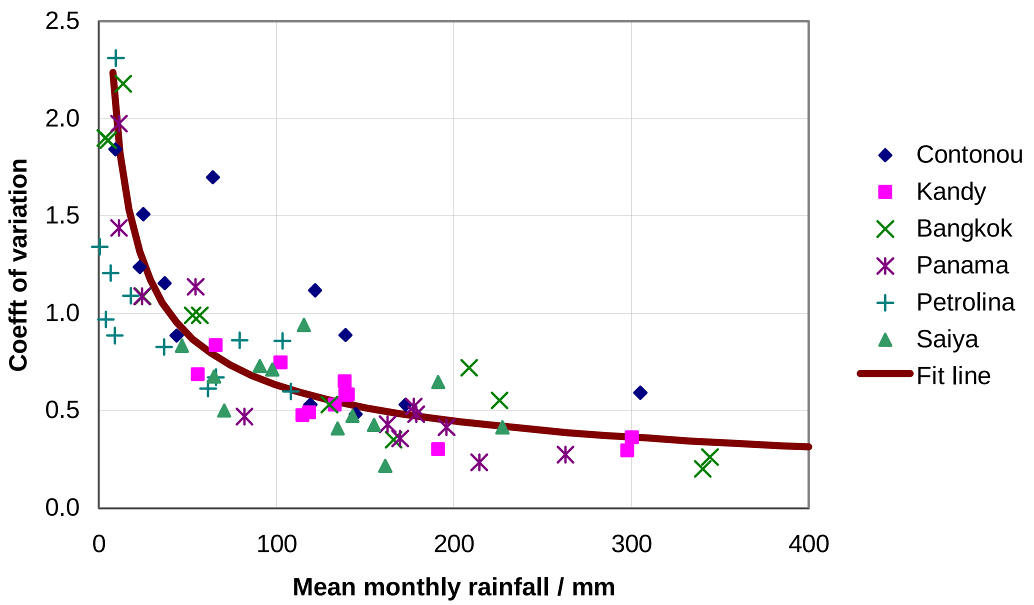
Tables similar to Table1 have also been produced to show not *reliability* but *satisfaction*. The values in such tables are of the order of 10% bigger than in Table 1 but the error introduced by employing only monthly data is similar. Other management strategies, such as rationing when the tank content is low, have also been modelled and show similar levels of error.

3. AVAILABILITY OF SUITABLE DATA

Meteorological data is rarely detailed, reliable and free. In many countries a 10-year daily rainfall record costs more than a small RWH system to buy. Rainfall varies with location, season and year. Its spatial variability is strongly influenced by topology and factors like distance from a coast. Its temporal variability becomes proportionally greater the lower the mean rainfall. The rough 'Fit line' in Figure 1 below shows that the variability, as measured by a coefficient of variation, may be expressed by $CoV = (R_m/40)^{-0.5}$ where R_m is the mean rainfall for that month in that location. The data is for two tropical sites from each of the Asian, African and American Continents.

In terms of geographical location, it may be possible to interpolate between neighbouring meteorological stations, or to use annual records to obtain a rainfall multiplication factor to apply to data from one place to make it better suit another.

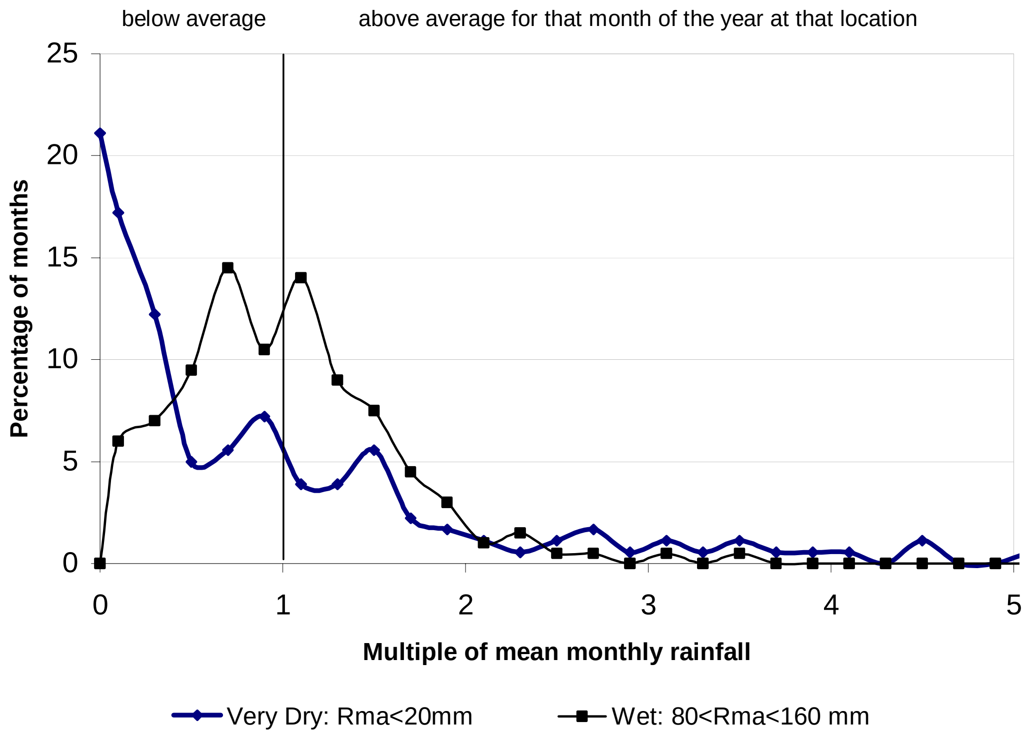
Figure 1. Variation in monthly rainfall (over 10 years & 6 sites)



In terms of time resolution, rainfall data is most readily available in ‘mean annual’ form, then as ‘mean monthly’, then as ‘actual monthly’ and least readily as ‘daily’. Even half-hourly data is available in richer countries with automatic recording equipment. Unfortunately ‘mean annual’ rainfall is little use for RWH modelling unless the modeller also possesses a library of representative seasonal distributions for the relevant region. ‘Mean monthly’ data is both too coarse and too time-averaged, but it is of some use and it is quite widely available (Pearce E & Smith C G, 1998). (Figure 2 shows amplitude distributions for 180 ‘dry-month’ tropical rainfalls and 200 ‘wet-month’ rainfalls normalised to the mean rainfall for the relevant calendar month and location. It confirms the greater variability and much higher chance of zero rain in months with a low average rainfall.) ‘Actual monthly’ data may be sufficient for quite accurate modelling as will be discussed below. Daily rainfall records may be regarded as effectively unavailable or unaffordable for most tropical sites.

Figure 2. Amplitude probability distributions

Monthly rainfalls placed in bands of width 20% mean monthly, or as zero rain



4. GENERATING PSEUDO DAILY RAINFALL DATA FROM ACTUAL MONTHLY RAINFALL DATA

In using monthly data directly we are effectively assuming constant daily conditions throughout that month – this is a very poor approximation and we could do better. From daily records for a climatic zone we could statistically characterise the rainfall distribution (in amplitude and sequence). We can then use these statistics to generate random rainfall data with the correct distribution and monthly totals.

In the normal collection and generation of rainfall data, daily records are condensed into monthly or annual records by the process below ($\rightarrow^S$ denotes summation and $\rightarrow^A$ denotes averaging).

$$\textit{Actual daily rainfall } (r) \rightarrow^S \textit{Actual monthly } (R_m) \rightarrow^S \textit{Actual annual}(R_a) \rightarrow^A \textit{Mean annual } (R_{aa})$$

$$\downarrow^A$$

$$\textit{Mean monthly } (R_{ma})$$

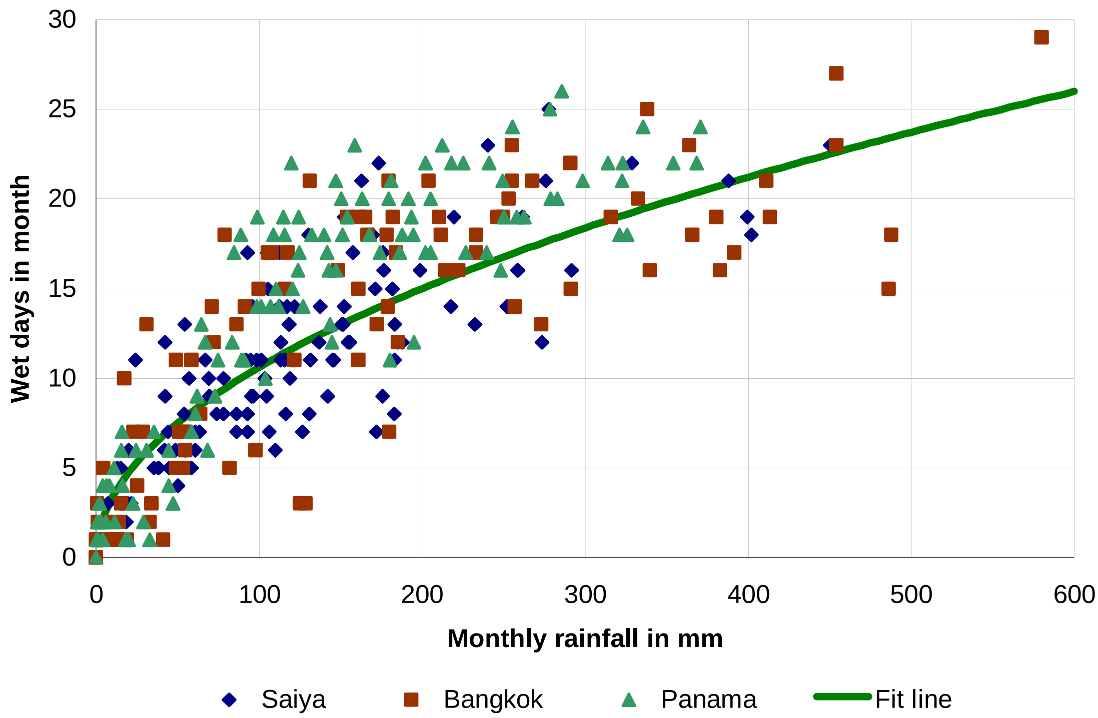
In generating pseudo data we reverse the process ($\rightarrow^R$ denotes randomisation)

$$\textit{Mean monthly } (R_{ma}) \rightarrow^R \textit{Pseudo monthly } (R'_m)$$

$$\textit{Actual } (R_m) \textit{ or pseudo } (R'_m) \textit{ monthly} \rightarrow \textit{Probability of wet day } (P) \rightarrow^R \textit{Pseudo daily } (r')$$

The amplitude distribution for daily rainfall we can treat as having two parts, an impulse at the origin corresponding to no-rain days and a generally falling curve away from the origin (reflecting that the lighter the rainfall the higher its probability of occurring). It is practical to replace these parts by a probability P (of a day being wet) and an amplitude distribution applicable only if the day is wet. The latter should give the correct wet day average R_w which in the humid tropics varies surprisingly little with variation in monthly rainfall (i.e. in wetter months there are primarily more wet days rather than wetter wet days).

Figure 3. Wet days per month (i.e. 30 P) versus monthly rainfall



The number of wet days in a standard month will be $n_w = 30 \; x \; P$. Figure 3 shows a relationship between wet days and monthly rainfall and a crude fit line suggesting the number of them (and hence P) is proportional to square root R_m. In subsequent modelling we have therefore used $P = \mathrm{SQRT}(R_m/A)$ where A = 800 mm generally gives a good fit in the tropics. To maintain the correct monthly totals we have a mean wet-day rainfall $R_w = R_m / n_w$ and this ranges from around 6mm in a dry month to around 16mm in a very wet one.

The distribution of these wet days is open to debate. Table 2 (column 4) compares the actual probability of getting rain on successive days with that derived from a random (Poisson) rain-event sequence having the same mean wet-day probability P. 'Rain yesterday' clearly raises the probability of 'rain today' above its long-term value, indicating some bunching of rainy days. The table suggests that there is a higher than expectedly high chance of rain on successive days and a lower than expected chance of rain on alternate days. The auto-correlation of a long record of daily rainfall (with a 1-day shift) is also typically positive and of size 0.1 to 0.2. Seasonality explains some, but not all, of this apparent bunching

Table 2. Spacing between rainy days, actual *v* random (same average spacing)

Location	Basis of prob-abilities	Probability of rain today			
		In General $P = 1/ d_a$	**yesterday was wet**	**last rained 2 days ago**	**last rained >14 days ago**
Saiya, W Kenya	Actual	0.353	0.509	0.338	0.145
	Poisson	*0.353*	*0.353*	*0.353*	*0.353*
Bangkok	Actual	0.338	0.566	0.414	0.043
	Poisson	*0.338*	*0.338*	*0.338*	*0.338*
Petrolina NE Brazil	Actual	0.157	0.407	0.178	0.054
	Poisson	*0.157*	*0.157*	*0.157*	*0.157*

In computational terms, producing a suitable Markov process, whereby yesterday's turn-out affects today's probability, is complex and therefore to be avoided unless proven really beneficial. Initially, therefore, all modelling will be done assuming no such preceding-day influence – i.e. each day will be considered independent and only influenced by monthly rainfall. Thus wet days are taken to be randomly spaced and can be assigned using a random number generator ("wet if $X<P$" where X is a random number that is uniformly distributed between 0 and 1).

We next meet the task of assigning pseudo rainfall to each wet day. Examination of graphs suggest a roughly falling exponential amplitude distribution for wet day rainfall: $p_r = e^{-r/R}/R$ where p_r is the probability density in units of mm^{-1}. However it can be shown that the probability of very low and of very high rainfall is higher than this distribution would suggest. The modeller has therefore the options of using a straight exponential amplitude distribution – which is fairly easy to implement, or of 'tweaking' such a model with an additional tuning adjustment or of using some other distribution difficult to mimic. The second option was employed in this study. It can be shown that transforming a uniformly distributed random variable X (range 0 to 1) by the equation $r' = -R \ln(X)$ gives the generated rainfall r' the desired distribution, provided that the constant R is the mean wet-day rainfall R_w for the current month. The probability of low and high rainfall can be increased by processing X before taking its logarithm: a suitable algorithm is:

$$r' = -R \ln(\mathrm{Z}) \quad \text{and} \quad Z = \mathrm{a}X + \mathrm{b}X^2 + \mathrm{c}X^3 \quad \text{where} \quad \mathrm{b} = 3(1\text{-a});\ \mathrm{c} = \text{-}2(1\text{-a})$$

Table 3. Procedure for generating pseudo-daily rainfall *r"* from monthly rainfall R_m

Step	Procedure
(i)	Estimate the wet-day probability *P* from the monthly rainfall R_m and hence also obtain mean wet-day rainfall R_w for that month,
(ii)	Use a random number generator in conjunction with the threshold *P* to decide if today is a wet day,
(iii)	If it is a wet day, generate a random rainfall *r'* for today which has a suitable amplitude distribution (assuming no influence from the previous day's rainfall)
(iv)	Re-scale all the daily *r'* values for the month to generate new values *r"* that add correctly to the given monthly total R_m

Using this procedure, pseudo rainfall sequences were generated for three tropical locations with respectively double rains, Monsoon rains and low erratic rains. The agreement between actual and pseudo daily rainfall – in terms of totals, rainy days and distributions as shown in

Table 4 – is quite good, and of course much superior to the agreement between actual and *uniform* simulated rainfall as shown italicised.

Table 4. Comparison of characteristics of actual, pseudo and 'uniform' daily rainfall

Location	Data type	Annual rainfall	Wet days per year	Wet days/year >30mm rain	Wet days/yr <10mm rain	Auto-correlation
Saiya W Kenya	Actual	1482	131	11	84	0.144
	Pseudo	1486	133	11	82	0.025
	Uniform	*1486*	*354*	*0*	*339*	
Bangkok	Actual	1546	121	15	75	0.250
	Pseudo	1551	113	16	65	0.149
	Uniform	*1551*	*318*	*0*	*270*	
Panama	Actual	1528	154	13	108	0.107
	Pseudo	1530	127	13	76	0.083
	Uniform	*1531*	*345*	*0*	*318*	
Petrolina NE Brazil	Actual	507	56	3	41	0.221
	Pseudo	506	65	3	46	0.094
	Uniform	*507*	*312*	*0*	*312*	

On the strength of this comparison we may now test the accuracy of employing synthetic daily rainfall data where we have no actual such data.

5. PERFORMANCE OF MODELS USING PSEUDO DAILY DATA

Table 5 mimics Table 1 except that now we are using more carefully generated pseudo daily rainfall. As before we compare predicted *reliability* for a range of RWH scenarios using respectively actual and pseudo rainfall data. The comparison is fairly good but not excellent. In the range of practical interest (*reliability* > 50%) the error does not exceed 5% and averages about 2%. As the error is always positive (pseudo data over-estimates performance) it might be prudent to subtract 2% from all *reliability* forecasts.

The source of what error there is may lie in the randomisation process itself – the fluctuations it produces may need longer than 10 years modelling to cancel out. More likely the practical decision to treat successive days as statistically independent, ignoring the small influence of one day's rain on the next day's likelihood of rain, has made the pseudo data slightly too favourable. Extensive natural clumping of rainy days is likely to increase the probability of tank overflow and hence degrade performance: such clumping is absent in the pseudo data.

The comparison was deemed sufficiently close to justify offering an open-access RWH system modelling service on the web at www.eng.warwick.ac.uk/dtu/rwh/model which is driven by user's monthly rainfall data and proposed system details and which yields *reliability, satisfaction* and *efficiency* estimates.

Table 5. Computed reliability of RWH supply as a function of rainfall data used in model

(Constant demand, actual daily data compared with pseudo daily rainfall)

Location, annual rainfall and climate type	**Rainfall data used for model**	**Small tank, 700 l**		**Medium tank, 3000 l**		**Large tank, 12000 l**	
		Nominal daily demand as percentage of mean daily roof run-off					
		60%	**120%**	**60%**	**120%**	**60%**	**120%**
Saiya, W Kenya	actual daily	0.79	0.51	0.97	0.69	1.00	0.76
1507 mm/year	pseudo daily	0.84	0.53	0.99	0.70	1.00	0.76
Double rains	*error %*	5%	2%	2%	1%	0%	0%
Bangkok, Thai	actual daily	0.58	0.42	0.73	0.57	1.00	0.76
1500 mm/year	pseudo daily	0.62	0.44	0.76	0.59	1.00	0.76
Monsoon	*error %*	4%	2%	3%	2%	0%	0%
Panama	actual daily	0.71	0.50	0.83	0.66	1.00	0.76
1500 mm/year	pseudo daily	0.72	0.51	0.84	0.66	1.00	0.76
Long single rains	*error %*	1%	1%	1%	0%	0%	0%
Petrolina, Brazil	actual daily	0.45	0.27	0.65	0.47	0.96	0.76
mm/year	pseudo daily	0.53	0.33	0.70	0.52	0.98	0.76
Semi-arid	*error %*	8%	7%	4%	5%	2%	0%

6. CONCLUSIONS

In the humid tropics, where economy rather than extreme water reliability is the normal design objective, RWH tank volumes may be as small as 7 days' mean roof run-off. Under these circumstances direct use of monthly rainfall data in RWH system modelling will lead to considerable bias in performance predictions.

The conflict, between wanting to use *daily* rainfall data to achieve unbiased predictions and having to make do with coarser available data like actual monthly or even mean monthly rainfall data, can be resolved by generating pseudo daily data. A RWH model that employs suitably generated pseudo daily data as its input gives performance predictions quite similar to one using actual daily rainfall. A procedure for generating such pseudo data, valid for tropical sites, has been identified and employed in a web-based modelling service.

The procedure assumes that 10 years *actual monthly* rainfall data is available for the proposed RWH system site. This condition can often not be met. Since *mean monthly* data is more often available, the user of the performance forecasts can use that instead and hope that an average year is a typical year. This hope is slightly optimistic but acceptable in most circumstances.

Meanwhile studies continue into whether *pseudo monthly* rainfall data, generated by randomising *mean monthly* data can reliably be used for RWH performance forecasting. Other modelling refinements, including using conditional probabilities in lieu of a fixed local value of wet-day probability P, may reduce prediction errors further.

7. BIBLIOGRAPHY

Gould, John and Nissen-Petersen, Erik (1999) *Rainwater catchment systems for domestic supply,* IT Publications London.

Heggen, R J, (1993) "Value of daily data for rainwater catchment", in Bambrah, G. K. (ed.), *Proc 6th Int Conf Rainwater Catchment Systems,* Nairobi.

Pearce E A & Smith C G (1998) *Hutchinson World Weather Guide,* Helicon Oxford.

Vyas V (1999). *Modelling water sources for rural communities* (SimTanka). Ajit Foundation.