Local random errors in drop-counting rain gauges

P.W. Chan * and C.M. Li
Hong Kong Observatory, Hong Kong, China

Grzegorz J. Ciach
The University of Iowa, U.S.A.

1. INTRODUCTION

Drop-counting rain gauges (model: Ogawa 7182R) are used at the Hong Kong International Airport (HKIA) by the Hong Kong Observatory (HKO) for operational rainfall reporting. Three identical gauges are set up at the meteorological garden at HKIA. They fulfill the exposure requirements for rainfall measurement as laid down in World Meteorological Organization (WMO), No. 8 – "Guide to Meteorological Instruments and Methods of Observation". Regular maintenance and calibration of the gauges are carried out. In particular, calibration is performed with a generally constant flow of water at different flow rates, and the WMO accuracy requirement of 5% was found to be fulfilled up to a rainfall rate of about 100 mm/hour for the dataset under study in the present paper.

With data collected by three gauges at the same time, the local random error of the type of drop-counting gauge in use has been analyzed in a preliminary study by Chan and Li (2009). There are two major limitations with this previous study:

(a) The rainfall data have gone through minimal quality control procedure only. As such, it may not be concluded from the previous dataset whether the analytical model for standard error (Equation (5) in Chan and Li (2009)) could be employed for the type of rain gauge under consideration;
(b) The value of the parameter *a* in the Nadaraya-Watson kernel regression estimator for the standard error is taken to be 0.2 following Ciach (2003). It is not certain if another, more suitable value of the parameter should be chosen.

The above two issues are studied in the present paper.

2. QUALITY CONTROL OF THE RAINFALL DATA

An iterative procedure has been adopted in the more vigourous quality control of the rainfall data. In general, it is expected that the standard error σ_K should be a monotonic decreasing function of the local rainfall rate R_T. However, in the previous analysis of the rainfall dataset (a sample chart of Chan and LI (2009) is reproduced in Figure 1(a), with T = 50 minutes), there are many "bumps" in the plot of σ_K against R_T. Such "bumps" may be related to potentially erroneous rainfall record of one or more of the rain gauges. For the range of R_T in a selected "bump", the corresponding rainfall data of the gauges with such values of local rainfall rate are examined carefully through two methods: (a) checking of the maintenance logs of the rain gauges to see if there were any reported problems/maintenance in the period, and (b) the rainfall data of all the Ogawa gauges are compared with those from other gauges at the same time, including a 0.5-mm tipping bucket rain gauge, a tilting-siphon rain gauge, and manual rainfall measurement. If there are sufficient reasons to suspect that the data from one or more of the Ogawa rain gauges in the period may be erroneous, for instance, much different (say, > 20%) from the rainfall record of the other gauges, the rainfall data from the three gauges are not considered. With the removal of the rainfall data in the period, the resulting σ_K is plotted against R_T again to check the shape of the curve. If there are still bumps present, the above process is repeated.

Following the above steps, the rainfall data in the periods as given in Table 1 have been removed. The resulting "clean" dataset is used to plot σ_K against R_T. For instance, the plot for T = 50 minutes is given in Figure 1(b), together with the variation of the number of samples with the local rainfall rate. It could be seen that the standard error generally drops with R_T, and the curve is smooth without significant "bumps" in comparison with that in Figure 1(a). It is noted that it may not be practical to prepare an "ideally clean" rainfall dataset. For instance, in the iterative quality control method as described above, one criterion being used to assess the quality of the rainfall data is to compare with the rainfall records of the other independent gauges with different operation mechanism at the meteorological garden at HKIA. It is not trivial to set up an objective guideline to determine if one or more of the Ogawa rain gauges is/are not functioning well through such comparisons because of the natural variability of the rainfall (though all the gauges are close to each other, with a maximum separation of at most a couple of metres or so) and the different working principles and reporting resolutions of the various gauges. The best attempt has been made by the authors to "clean" the Ogawa rainfall dataset in order to produce a plot of σ_K against R_T that looks reasonable and much better than the one obtained before the quality control procedure.

For completeness, the plot of σ_K against R_T for other values of T are given in Figure 2, namely, T = 10 seconds, 1 minute, 3 minutes and 10 minutes. With the "clean" Ogawa dataset, the standard error generally falls with the local rainfall rate for the various values of period T. This is another indication that the dataset is relatively "clean", and the new plots are much better than those in the previous study, e.g. Figures 4(a) and (c) in Chan and Li (2009).

* *Corresponding author address:* P.W. Chan, Hong Kong Observatory, 134A Nathan Road, Hong Kong
email: pwchan@hko.gov.hk

3. PRELIMINARY STUDY OF THE CHOICE OF THE PARAMETER *a*

In both Ciach (2003) and Chan and Li (2009), the parameter *a* in the computation kernel estimator is taken to be 0.2. A preliminary study has been conducted to find out the suitable range of value for *a* based on the Ogawa dataset. The initial results are presented here.

As a start, the period *T* is taken to be 3 minutes. The Ogawa dataset is arranged in a list of increasing average rainfall from the three gauges. The odd and even samples of this list are grouped together to form two subsamples S1 and S2. If there are a number of samples with the same average rainfall, it does not matter which one goes to S1 or S2.

For subsample S1, the first nonparametric sigma square function σ_1^2 is calculated with the application of the kernel regression. The first prediction mean square difference, $PMSD_1$, between the error square values $e^2(R_2)$ and $\sigma_1^2(R_2)$ is then computed, where R_2 are the given values of R_T in subsample S2 and $R_2 > R_0$, a threshold value:

$$PMSD_1 = \sum\left[e^2(R_2) - \sigma_1^2(R_2)\right]^2 \qquad (1)$$

where the summation is made over all data fulfilling $R_2 > R_0$. Note that in general there may not be data points in subsample S1 having R_T at exactly the value R_2 given in subsample S2. In that situation, linear interpolation is performed on σ_1^2 to obtain $\sigma_1^2(R_2)$. Similar computation is carried out for subsample S2 to obtain $PMSD_2$. The average of $PMSD_1$, and $PMSD_2$ gives the final *PMSD* for a particular value of parameter *a*. The process is repeated for each value of *a*.

A plot of *PMSD* against *a* is made for each value of R_0 in order to find out the range of the parameter *a* that minimizes the *PMSD*. Figure 3 shows some plots with R_0 = 0.5, 2 and 5. It appears that the optimum range of *a* does not depend very much on R_0, and it could be taken as 0.1 – 0.4. Thus the previous choice of 0.2 is a reasonable value.

As an illustration, the sigma plots for the various values of *a* (with *T* taken to be 3 minutes) are given in Figure 4. As could be expected, with an increasing value of *a*, the sigma plot becomes smoother. Starting from *a* = 0.1 onwards, the shape of the sigma plot does not change significantly.

The above analysis is based on a particular value of *T* only. Other values of *T* would be considered in future studies. Moreover, a lower bound R_0 is adopted in the present analysis, and an upper bound may be considered as well. Preliminary results (not shown here) seem to suggest that the optimum range of parameter *a* does not depend very much on this upper bound.

4. ANALYTICAL MODEL FOR STANDARD ERROR

In Chan and Li (2009), attempt was made to use an analytical model to fit σ_K against $1/R_T$. It turned out that, for larger value of *T*, the model did not fit so well with the data and the correlation coefficient of the fit decreased. It is not sure if the problem is related to the model itself or the quality of the dataset. Based on the rainfall dataset with better quality in the present paper, the use of the analytical model is revisited.

Figure 5 shows the following for *T* = 10 seconds, 3 minutes and 50 minutes: the fit of the analytical model in the plot of σ_K against R_T for the whole range of R_T, the corresponding fit for small local rainfall rates, and the fit in the plot σ_K against $1/R_T$ with the equation and correlation coefficient of the fit. In the last plot, the data points in general fluctuate less rapidly with $1/R_T$ and get closer to a straight line in comparison with the results in Chan and Li (2009) (e.g. Figure 4(f) of that paper for *T* = 50 minutes). However, the correlation coefficient R^2 of the fit still decreases quite dramatically with *T*, from 0.96 for *T* = 10 seconds to 0.44 for *T* = 50 minutes. As such, based on the present dataset, it may be concluded that the analytical model in Ciach (2003) does not seem to work well, at least it is not universal to be applicable for different values of *T*. Another model of standard error may need to be established.

5. CONCLUSIONS

A more vigourous control quality procedure is adopted in the present study in order to obtain a relatively "clean" dataset of Ogawa rainfall as recorded at the meteorological garden at HKIA.

In the previous studies, the parameter *a* in the calculation of the kernel estimator is taken to be 0.2. The optimum range of value of this parameter is studied in more detail using prediction mean square difference in statistical analysis. Preliminary results are presented here, namely, a period *T* = 3 minutes is adopted, and only a lower bound R_0 is applied to the dataset. It turns out that the optimum range of *a* is between around 0.1 and 0.4. As such, the previous choice of 0.2 appears to be reasonable. More in-depth study of the choice of *a* would be carried out in the future, for instance, for different values period *T*.

Moreover, using the "clean" dataset from Ogawa rain gauges, it looks like the linear analytical model as adopted in Ciach (2003) and Chan and Li (2009) is not universal for the various values of the period *T*. Another model would need to be set up.

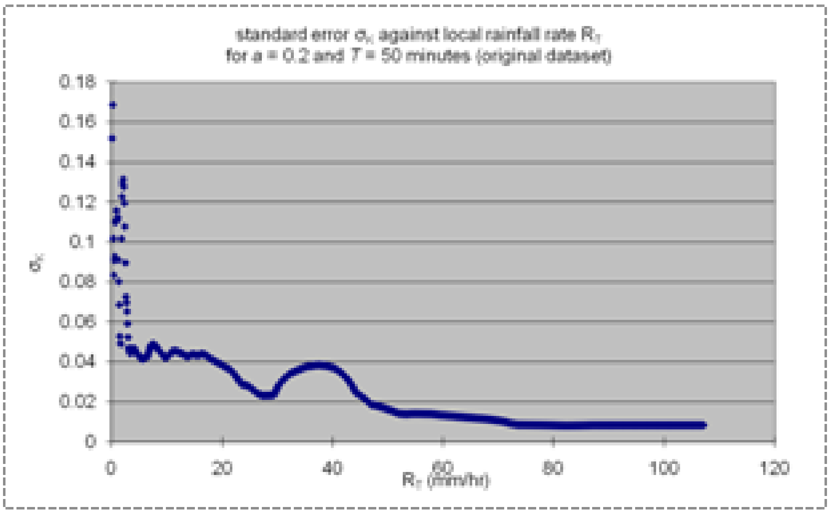
The three Ogawa rain gauges have been installed at HKIA again since September 2009, though there are some changes in the characteristics of the gauges based on the new calibration results. The new dataset obtained since September 2009 would also be analyzed separately to see if there is a corresponding change of the behaviour of the local random error.

References

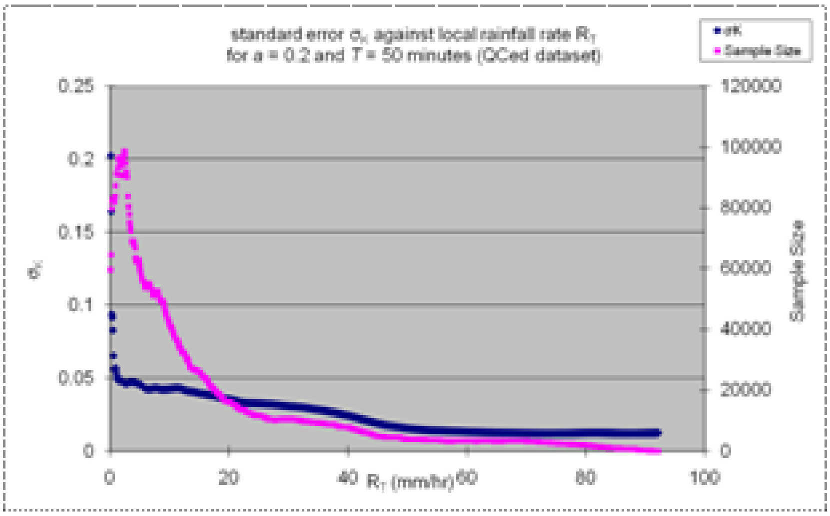
Chan P.W., and C.M. Li, 2009: Performance of drop-counting rain gauges in an operational environment. *13th Conference on Integrated*

Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface, American Meteorological Society, Phoenix, U.S.A.

Ciach, G.J., 2003: Local random errors in tipping-bucket rain gauge measurements. *J. Atmos. Oceanic Technol.*, **20**, 752 – 759.



(a)



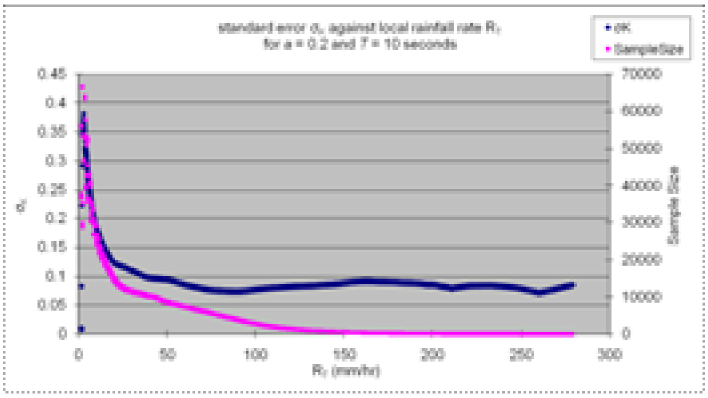
(b)

Figure 1 The plot of standard error against local rainfall rate for T = 50 minutes: (a) results in Chan and Li (2009) and (b) results based on the dataset in the present paper.

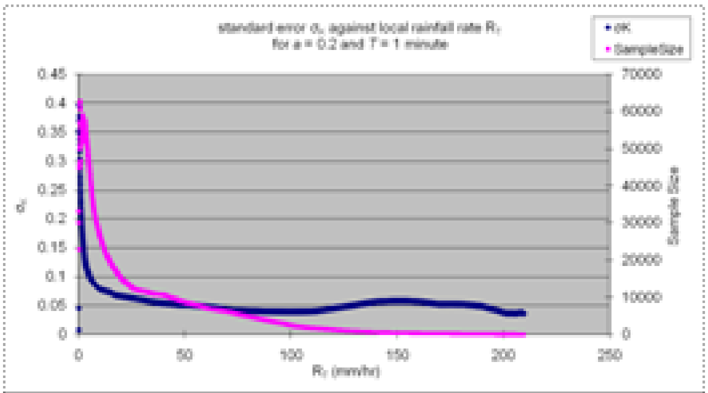
Table 1 The periods (in Hong Kong time) in which the Ogawa rainfall data are removed.

2007/06/28 0000-2359
2007/08/08 1640-1700
2008/01/25 0125-1014
2008/04/19 0000-2359
2008/05/10 1940-2325
2008/06/03 1755-1900
2008/06/06 0425-2359
2008/06/07 0000-0035
2008/06/18 0816-2347
2008/08/22 0000-2359
2008/08/23 0000-2359
2008/09/24 1000-1115

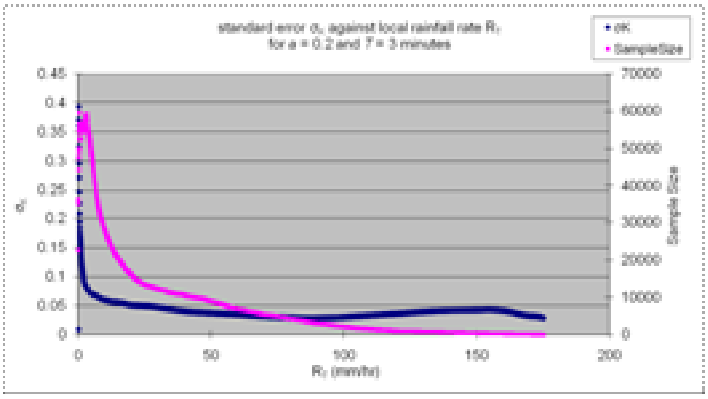
(a) 10 seconds



(b) 1 minute



(c) 3 minutes



(d) 10 minutes

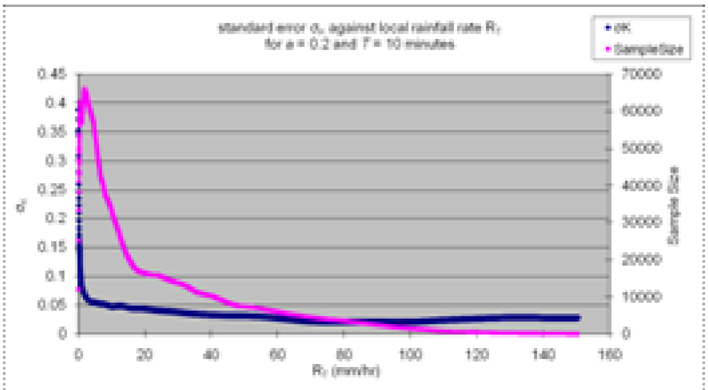
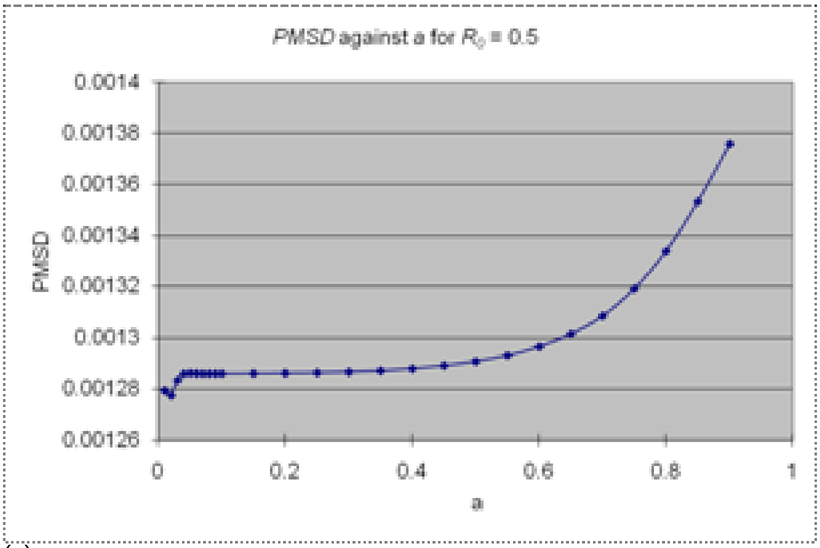
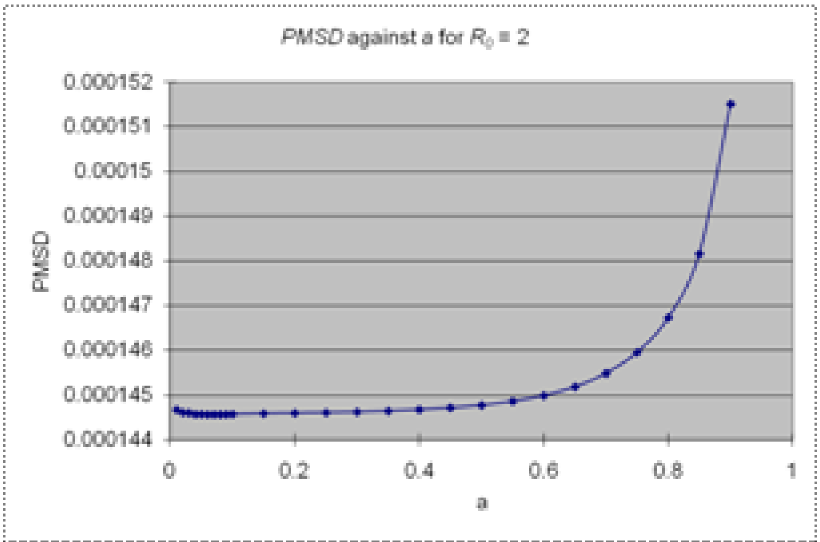


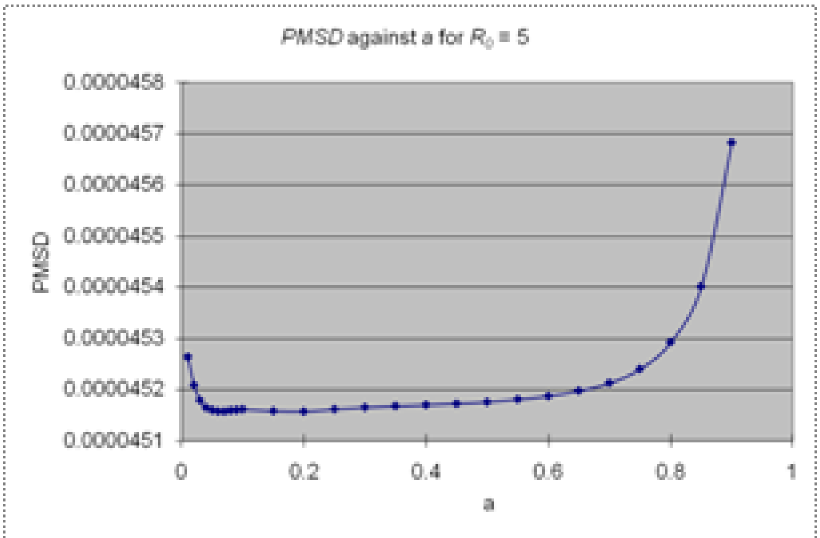
Figure 2 The plot of standard error against local rainfall rate for T = 10 seconds, 1 minute, 3 minutes and 10 minutes.



(a)

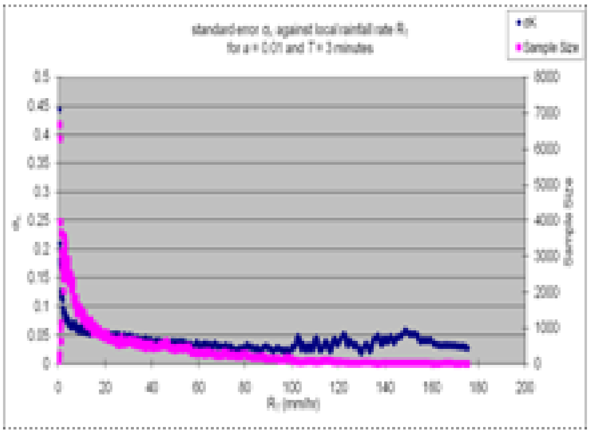


(b)

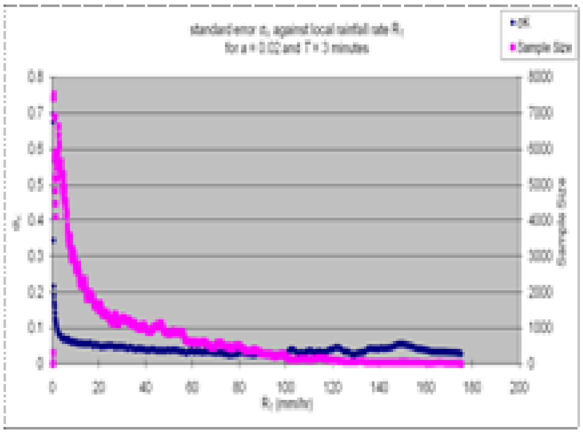


(c)

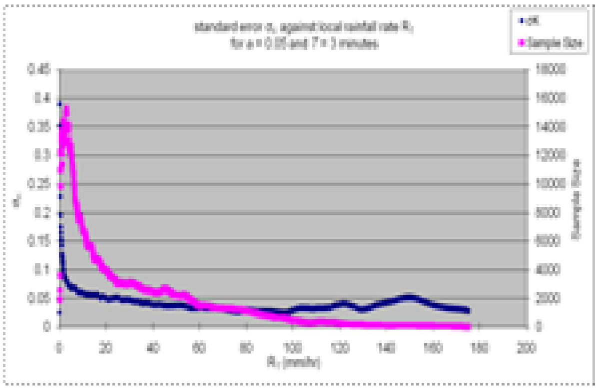
Figure 3 Plots of *PMSD* against *a* for R_0 = (a) 0.5, (b) 2, and (c) 5.



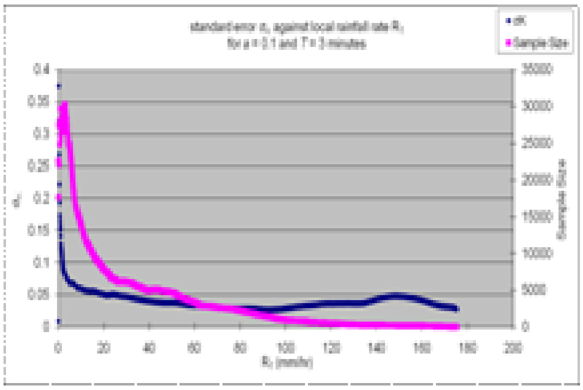
(a)



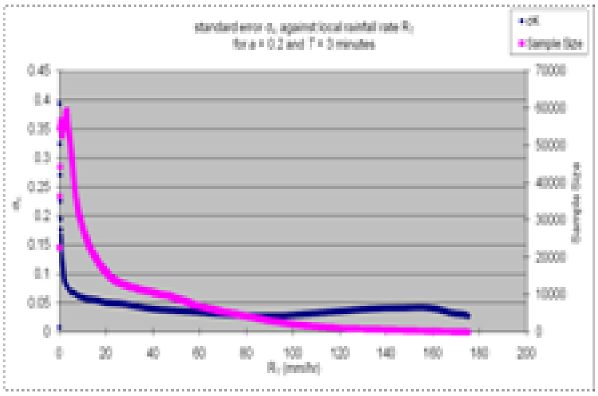
(b)



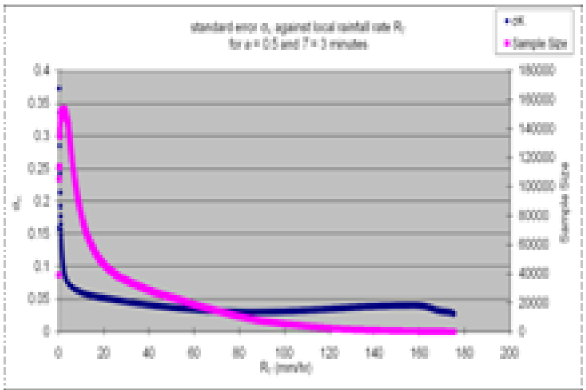
(c)



(d)



(e)



(f)

Figure 4 Sigma plots for the various values of parameter *a*. *T* is taken to be 3 minutes.

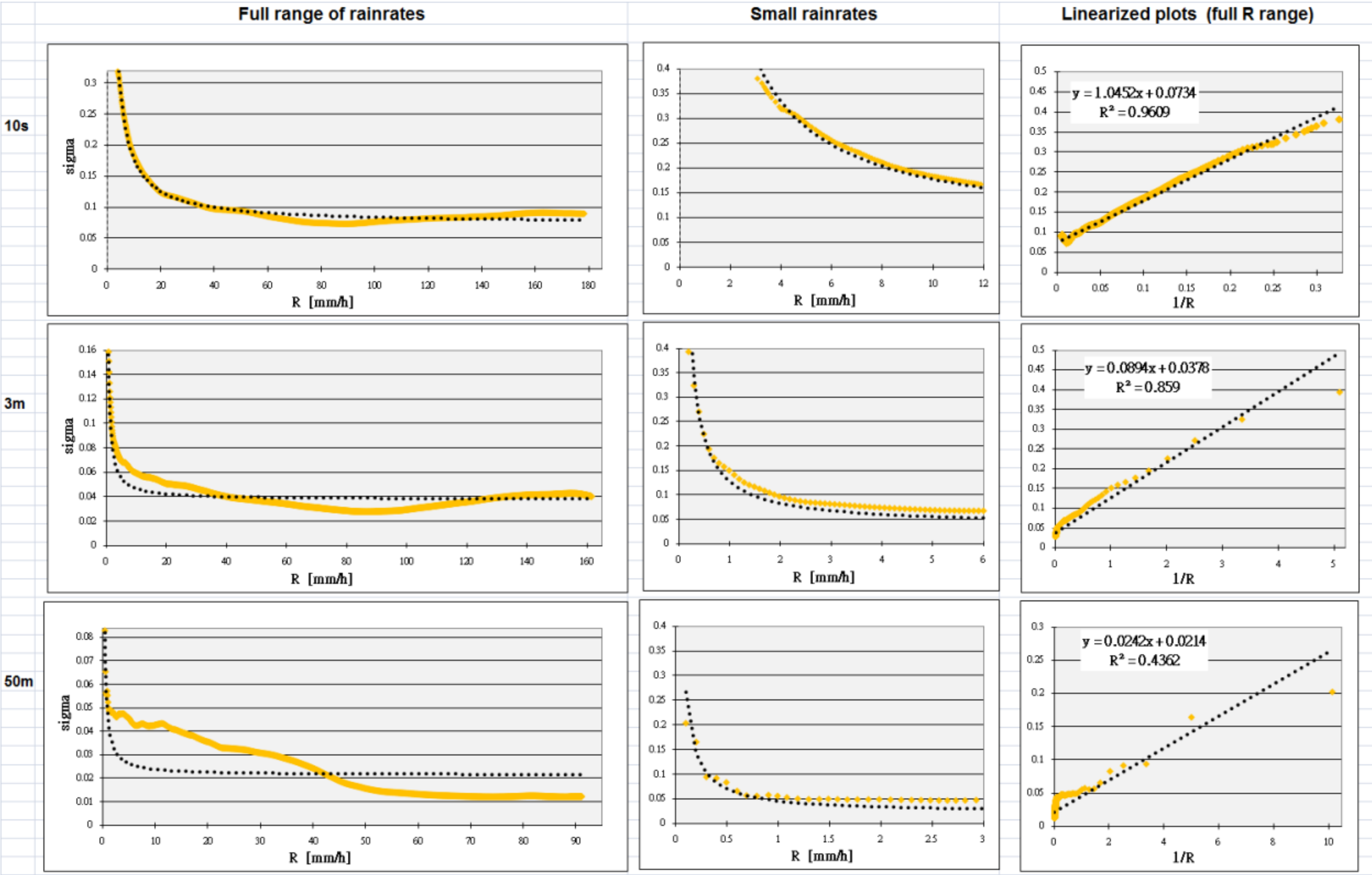


Figure 5 Fitting of standard error data (black dots) using an analytical model (yellow curves). In the plots, R means R_T for T = 10 seconds (10s), 3 minutes (3m) and 50 minutes (50m).