Statistical Problems in Measuring Convective Rainfall

A Thesis

Submitted to

The Faculty of Graduate Studies and Research

of

McGill University

by

(C) Alan William Seed

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

Department of Agricultural Engineering

Macdonald College of McGill University

Sainte-Anne-de-Bellevue

Quebec,Canada

H9X 1C0

June 1989

Abstract

Ph.D. Alan W. Seed Agricultural Engineering

Statistical Problems in Measuring Convective Rainfall

Simulations based on a month of radar data from Florida, and a summer of radar data from Nelspruit, South Africa, were used to quantify the errors in the measurement of mean areal rainfall which arise simply as a result of the extreme variability of convective rainfall, even with perfect remote sensing instruments. The raingauge network measurement errors were established for random and regular network configurations using daily and monthly radar-rainfall accumulations over large areas. A relationship to predict the measurement error for mean areal rainfall using sparse networks as a function of raining area, number of gauges, and the variability of the rainfield was developed and tested. The manner in which the rainfield probability distribution is transformed under increasing spatial and temporal averaging was investigated from two perspectives. Firstly, an empirical relationship was developed to transform the probability distribution based on some measurement scale, into a distribution based on a standard measurement length. Secondly, a conceptual model based on multiplicative cascades was used to derive a scale independent probability distribution.

Résumé

Ph.D. Alan W. Seed Génie agricole

Problémes statistiques pour la mesure des pluies convectives

Des simulations numériques basées sur un mois de données radar de Floride et un été de données de Nelspruit, Afrique du Sud, ont été utilisées pour quantifier les erreurs dûes a l'extréme variabilité des pluies convectives dans la mesure de la pluie surfacique moyenne en supposant des instruments de télédétection parfaits. Les erreurs de messures ont été etablies pour des réseaux pluviometriques à configuration aléatoire ou régulière en utilisant les accumulations journalières et mensuelles de pluie estimées par radar sur de grandes surfaces. Une méthode pour prévoir les erreurs sur les mesures sur la pluie surfacique moyenne utilisant des réseaux épars en fonction de la surface où il pleut, du nombre de pluviomètres et de la variabilité du champ de pluie a été developpée et testée. La facon dont laquelle la distribution de probabilité du champ de pluie est transformée en augmentant l'étendue spatiale et temporelle de la moyenne a été etudiée de deux perspectives: Premièrement, une relation empirique a été développée pour transformer la distribution de probabilité basée sur une echelle de mesures quelconque en une distribution basée sur une autre échelle de mesures. Deuxièmement, un model conceptual basée sur des cascades multiples a été utilisé pour déterminer une distribution de probabilité indépendante de l'échelle.

Acknowledgements

The author wishes to extend his appreciation to Dr. G.L. Austin, Professor of Physics, friend and mentor, for his supervision of this thesis. The author also acknowledges Dr. R.S. Broughton, Professor of Agricultural Engineering, and Dr. V.T.V. Nguyen, Professor of Civil Engineering and Applied Mechanics for their continued guidance during the course of this study. The time and effort freely given to the author by Dr. S. Lovejoy, Professor of Physics, during the research reported in Chapter 5 is gratefully acknowledged.

This thesis would not have been possible without generous financial support from the author's employer, the Department of Water Affairs, South Africa. In particular, the support of the erstwhile Director: Hydrological Research Institute, Mr. E. Braune is acknowledged.

This work, also would not have been possible without the love and patience of the author's wife, Sharon Seed, who has shared her husband with a computer for the past few years.

Table of Contents

Chapter		Page

List of Figures

Figure		Page

Figure		Page

List of Tables

Table		Page

Contributions to Knowledge

Many studies have been undertaken to estimate the sampling errors in the use of raingauge, radar, and satellite measurement of mean areal rainfall. However, what has not been addressed to date, is the lower limit of the measurement error that exists as a result of the extreme variability in the rainfields, even with a perfect instrument. The following contributions have been made for the three instruments:

a) Radar

1. The effect of the various digitizing algorithms currently in use to estimate the mean reflectivity from a number of fluctuating signals has been re-evaluated. It was found that the exponential averaging algorithm, although one of the most common in commercial applications, was the least effective in reducing the estimation variance.

2. The reduction in the estimation variance after averaging in space and accumulating in time was assessed. Geostatistical techniques were used to calculate the estimation variance for mean areal rainfall. It was found that it is more efficient to remove the measurement noise during the polar to

Cartesian transformation and accumulation in time, than to use a large number of samples during signal digitization.

b) Raingauges

3. For mean areal rainfall, over large areas, it was found that the mean standard errors for regular gauge networks were approximately the same as those for random gauge networks. The difference between the two configurations was that the errors for the random network had a larger spread about the mean standard error than the regular networks, particularly for sparse networks.

4. A relationship to predict the mean measurement error for sparse gauge networks was developed using the number of gauges, raining fraction of the area, and the rainfield variability as predictors.

c) Satellite

5. Measurement errors as a function of the number of satellite over-passes per day were estimated for the Florida area. The large sampling errors associated with orbiting satellites having two passes per day were quantified.

The transformation of the probability distribution of a rainfield under increasing averaging is an efficient characterization of the spatial organization and variability of the rainfield. Furthermore, a great many practical hydrological problems depend on the transformation of point rainfall statistics to mean areal rainfall statistics. The second section of the thesis contributed the following in this topic:

6. It was found that it is possible to characterize the way the rainfall probability distributions are transformed under spatial averaging by means of a two parameter power law. This can be applied to the problem of comparing rainfall estimates from techniques with different space-time sampling characteristics.

7. A resolution independent probability distribution was developed using the concept of multiplicative cascades and was fitted to space and time averaged data.

8. Parameter estimation techniques were developed for the estimation of the parameters used in the conceptual model.

Chapter 1

Introduction

The hydrological cycle makes it possible for man to exist on this planet, controlling the availability of water and vegetation as well as driving the global atmospheric circulation that distributes the heat from the tropics to the higher latitudes. Water vapour is able to transport the energy received during evaporation over large distances and then to release the latent heat into the atmosphere in local regions of intense precipitation. This implies that a knowledge of mean areal rainfall over large areas is a prerequisite to modelling global climate. It is for this purpose that the joint U.S.A. - Japan Tropical Rainfall Measuring Mission satellite (TRMM) is scheduled to be launched during 1994.

Concerns over the impact of possible changes in climate due to increased concentrations of carbon-dioxide in the atmosphere have increased over the past few years. One possible, although by no means certain, scenario is that the increasing concentrations of carbon-dioxide will result in global warming, the so-called green house effect. While this would not necessarily be a bad thing for all regions, Quebec winters could usefully be warmed resulting in a longer growing season for example, increased seasonal variability in the rainfall could seriously affect agriculture in regions where the water supply is marginal, even if the long term mean rainfall remained constant. The total annual runoff out of a large river catchment in a semi-arid region depends critically on a handful of days of widespread rain over the catchment. Therefore, water supply in semi-arid regions would also be seriously affected by even small changes in the number of widespread rain days per rainy season.

The rapidly expanding populations in semi-arid regions are placing increasingly heavy demands on the scarce water resources in these areas. This in turn implies that Water Resource Engineers need to be more certain about both the long-term mean areal rainfall, which determines the mean water supply, and the temporal and spatial variability of the rainfields which determines the reliability of the water supply. The optimal operation of increasingly complex water resource systems is being achieved through the use of real-time rainfall estimates and short-term stream flow forecasts.

Man's activities in a river catchment can have a significant impact on the rainfall-runoff relationship in the catchment. For example, deforestation of a catchment's uplands could increase the variability of the stream flow, increasing the probability of both floods and drought. Hydrological models are being used to predict the impact of various land-use changes on the stream flow. Although research into the next generation of detailed, spatially distributed hydrological models is currently underway, historical rainfall data with a commensurate level of detail are scarce.

Agriculture of course, is another major user of rainfall data. Crop yield prediction models are routinely used to forecast harvests and in semi-arid regions depend on the total moisture budget for the growing season. Satellite rainfall estimation schemes are used to monitor the famine potential and locust population in the Sahel region of Africa, for example.

Rainfall provides a key input into a large, diverse set of applications that directly or indirectly affect the majority of the Earth's population. It is interesting to note therefore, that despite seven centuries of rainfall measurement, we are still not able to measure mean areal rainfall with good accuracy. This is due to the highly intermittent, variable nature of the rainfields, and instruments that either measure the rainfall at a point with good accuracy or over an area with mediocre accuracy.

Hydrologists have been aware of the importance of the spatial and temporal variability of rainfall in hydrology for some time. Thiessen (1911) was possibly the first to explicitly account for the spatial variability of rainfall when estimating areal rainfall amounts. The usual measure of the spatial variability of rainfall is to express the correlation between two gauges as a function of the distance separating the gauges, for example Sharon (1972), Zawadzki (1973a), and is used by virtually all the current models used to generate synthetic rainfields, Eagleson, Fennessey, Qinliang and Rodriguez-Iturbe (1987) for example.

The minimum correlation between neighbouring raingauges has been used extensively in determining the spacing between gauges in a gauge network. Hershfeld (1965) was probably the first to propose that the minimum correlation between adjacent raingauges in a network should be 0.9. A typical study of a network spacing requirement for daily rainfall was made by Hendrick and Comer (1970). The error of the raingauge estimation of areal daily rainfall was obtained by comparing the arithmetic mean derived from a network of 23 gauges with the means derived from sub-networks. They concluded that if the minimum allowable error for the particular day or event must be known, then some criteria other that the correlations would be required.

Hutchinson (1972) pointed out that the use of a correlation criteria in its simplest form could only give a relative assessment of the standard error. Zawadzki (1973b) assumed an exponential decorrelation with distance and derived a formula to estimate the error in the area-averaged rainfall as estimated by a regular grid of rain gauges

The error in estimating the mean areal rainfall is not only a function of the gauge network, but also depends to some extent on the method used to estimate the areal rainfall from a set of gauge rainfall depths. Excellent comparative studies may be found in Creutin and Obled (1982) and Tabios and Salas (1985). Since measuring areal mean rainfall is rarely an end in itself, but rather to provide input into hydrological models, the impact of the areal rainfall measurement error on the model output is of concern. Eagleson (1967) assumed that the catchment behaves as a low-pass filter thus reducing the effect of the input measurement errors. Indeed, Eagleson (1967) concluded that only four gauges (correctly positioned) would be sufficient to estimate peak runoff from a catchment area of 3200 km^2 Wilson, Valdes and Rodriguez - Iturbe (1979) used a storm model to generate synthetic rainfall and simulated the effect of increased measurement error by sampling the rainfield in two different manners. The ground-truth was established by sampling the rainfield with 20 gauges. These rainfall data were than used to generate the true runoff from a catchment by means of a rainfall runoff model. The second data set was generated using only one gauge to sample the rainfield, which was then passed through the catchment model. The catchment was found to amplify the input errors, for example a 12% mean depth over-estimate resulted in a 24% runoff volume over-estimate and a 9% mean depth over-estimate resulted in a 52% peak flow over-estimate. More recently, Milly and Eagleson (1988) showed that hydrological models of catchments that are potentially larger than the storm size must give explicit consideration to the effect of the raining area on the runoff process.

In general, the rainfall data at the disposal of hydrologists are derived from rain gauge measurements which are essentially point measurements. Therefore, all estimates of rainfall exceedance probabilities or return periods are only valid for point rainfall. However, when sizing a culvert or drain, the engineer needs the return period for areal rain over some area and for some period of accumulation So the transformation of point-rainfall statistics into equivalent areal rainfall statistics has great practical importance. Correction factor to transform point rainfall into area averaged precipitation were first developed by the U S Weather Bureau in 1957 and updated in 1980 by Myers and Zehr (1980). Rodriguez-Iturbe and Mejia (1974) developed an alternative method to transform point rainfall into mean areal rainfall that used an exponential correlation structure together with an estimate of the mean distance separating two point within the area. Other methods have been proposed by Nguyen (1984) and Niemczynowicz (1984). However, Arnell et al (1984) in their review of these methods for urban hydrology concluded that there was a lack of suitable, experimentally justified models to transform point-rainfall into area averaged rainfall.

Throughout this thesis the terms "convective" and "widespread" in descrbing rainfall events in this thesis. Convective is used to mean areas of rainfall of small horizontal estent and large cloud height, generally synonymous with showers (American Meteorological Society, 1959) widespread rain on the other hand is taken to mean areas of rainfall of large horizontal extent and shallow depth with low rainfall intensities and weak gradients. For a particular case, both of these rainfall types may exist simultaneously, but the predominate type is usually clearly evident.

The thesis has been written as a collection of manuscripts of papers that have been, or will shortly be submitted for publication in a Journal. The contributions by the various co-authors has been to provide guidance during the research, and to give editorial comment during the preparation of the manuscripts. The specific regulations governing such a thesis format are reproduced below.

"The candidate has the option, subject to the approval of the Department, of including as part of the thesis the text, or duplicated published text (see below), of an original paper, or papers. In this case the thesis must still conform to all other requirements explained in Guidelines Concerning Thesis Preparation. Additional material (procedural and design data as well as descriptions of equipment) must be provided in sufficient detail (e g. in appendices) to allow a clear and precise judgement to be made of the importance and originality of the research reported. The thesis should be more than a mere collection of manuscripts published or to be published It must include a general abstract, a full introduction and literature review and final overall conclusion Connecting texts which provide logical bridges between different manuscripts are usually desirable in the interests of cohesion.

It is acceptable for theses to include as chapters authentic copies of papers already published, provided these are duplicated clearly on regulation thesis stationery and bound as an integral part of the thesis. Photographs or other materials which do not duplicate well must be included in their original publishable form In such instances, connecting texts are mandatory and supplementary explanatory material is almost always necessary

The inclusion of manuscripts co-authored by the candidate and others is acceptable but, the candidate is required to make an explicit statement on who contributed to such work and to what extent, and supervisors must attest to the accuracy of the claims before the Oral Committee. Since the task of the Examiners is made more difficult in these cases, it is in the Candidate's interest to make the responsibilities of authors perfectly clear Candidates following this option must inform the Department before it submits the thesis for review."

This thesis will attempt to evaluate the limitations that the highly variable nature of convective rainfall places on the measurement of areal rainfall even with perfect instruments. Three measurement devices, viz. the raingauge, radar and satellite will be examined using simulation techniques and a data base comprising of 21 days of continuous radar data of tropical rainfall recorded during August 1987 at the Kennedy Space Center located in Florida. The second chapter of the thesis is a paper co-authored with Prof Austin that has been accepted by the Journal of Geophysical Research, and was jointly presented by the two authors at a conference at M.I.T. during 1988. This paper evaluated the measurement error for the radar in some detail, and estimated measurement errors in estimating mean areal rainfall over large areas using gauge networks and satellites. The third chapter returns to the use of raingauge networks to measure daily mean areal rainfall over various size areas. The effects of network geometry, random or regular are considered. A new method to predict the mean standard estimation error for large area mean areal daily rainfall is presented and evaluated. The text in this chapter substantially represents a paper co-authored with Profs. Austin and Broughton to be submitted to the Journal of Hydrology.

Stochastic simulation of rainfields is an important tool for hydrological studies evaluating the effects of changing land-use and climate. Unfortunately, the characteristics of the rainfields change under increasing spatial and temporal smoothing. The fourth and fifth chapters therefore deal with the dependence of the rainfall probability distribution on the sensor resolution. Chapter four, co-authored with Prof. Nguyen, is a paper that explores the transformation of the probability distribution under spatial smoothing. It presents an empirical method to transform the probability distribution representing the rainfield at one scale into the distribution at a different scale. The recent advances in the theory of fractals, a theory that explicitly models the effects of changes in scale, raise interesting possibilities in their use as rainfield simulation models. However, more empirical evidence that rainfields do indeed behave in a fractal-like fashion is needed. Chapter five, a paper co-authored with Profs. Lovejoy and Austin, therefore investigates a method to derive a resolution independent probability distribution that is based on the theory of multifractal multiplicative cascades. It also develops a new method to estimate the parameters needed by a multiplicative cascade model for rainfield simulation, and as such is a first step in the application of fractal theory to rainfield simulation.

References

American Meteorological Society, 1959: Glossary of Meteorology. American Meteorological Society, Boston, Mass.

Arnell, V., P. Harremoes, M. Jensen, N.B. Joansen, and J. Niemczynowicz, 1984: Review of rainfall data application for design and analysis. Water Sci. Tech., Vol. 16, Copenhagen, 1-45.

Creutin, J.D., and C. Obled, 1982: Objective analysis and mapping techniques for rainfall fields: An objective comparison. Water Resources Res., 413-431

Eagleson, P.S., 1967: Optimum density of rainfall networks. Water Resources Res., 3(4): 1021-1033.

Eagleson, P.S., N.M. Fennessey, Wang Qinliang, and I. Rodriguez-Iturbe, 1987: Application of spatial Poisson models to air mass thunderstorm rainfall J. of Geophysical Research 92 (D8): 9661-9678.

Hendrick, R.C., and G.M. Comer, 1970. Space variations of precipitation and implications for raingauge network design J. of Hydrology 10: 151-163.

Hershfield, D.M , 1965: On the spacing of raingauges. In 'Design of Hydrological Networks'. Int. Assoc. Hydrol. Sci Pub No 67, 72-79

Hutchinson, P, 1972: Comment on "The relative efficiency of the density of rain-gauge networks" by PH-TH Stol. J. of Hydrology 17: 243-246.

Milly, P C.D , and P S. Eagleson, 1988: Effect of storm scale on runoff volume. Water Resources Res. 24(4): 620-624.

Myers, V.A., and R. M. Zehr, 1980: A methodology for point-to-area rainfall frequency ratios. NOAA Technical Report NWS24, NOAA, U.S. Dept. Commerce.

Niemczynowicz, J., 1984: An investigation of the areal and dynamic properties of rainfall and it's influence on runoff generating processes. Report No. 1005. Department of Water Resources Engineering, Lunds Institute of Technology, University of Lund.

Nguyen, V-T-V., 1984: Short time increment rainfall analyses for urgan drainage studies. Proc. of the Fourth Congress-Asian and Pacific Division, International Association for Hydraulic Research, Chiang Mai-Tailand, 11-13 September, 1984: 1265-1275.

Sharon, D., 1972: The spottiness of rainfall in a desert area. J. of Hydrology 17: 161-175.

Tabios, G.Q., and T.D. Salas, 1985: A comparitive analysis of techiques for spatial interpolation of precipitation. Water Resources Bull., 21(3): 365-380.

Thiessen, A.H., 1911: Precipitation averages for large areas. Monthly Weather Review, July 1911, 1082-1084.

U.S. Weather Bureau, 1957-60: Rainfall intensity-frequency regime, Part 1; the Ohio Valley, Part 2, southeastern United States, Tech. Paper No 29, U.S. Department of Commerce, Washington, D.C

Wilson, C.B., J.B. Valdes, and I. Rodriguez-Iturbe, 1979: On the influence of the spatial distribution of rainfall on storm runoff. Water Resources Res. 15 (2): 321-238.

Zawadzki, I.I., 1973a: Statistical porperties of precipitation patterns. J. of Applied Meteorology 12: 459-472.

Zawadzki, I.I., 1973b: Errors and fluctuations of raingauge estimates of areal rainfall. J of Hydrology 18: 243-255.

Prolegomenon to Chapter 2.

An estimate of the mean areal rainfall in the tropics is an important input into numerical weather models particularly of climate change. Unfortunately, much of this rainfall occurs over the ocean and therefore can only be measured from space platforms. The major objective of the joint U.S.A. - Japan Tropical Rainfall Measuring Mission Satellite is to provide estimates of mean monthly rainfall over large areas. Earlier studies into the likely error in such estimates are based largely on statistical rainfield models [1,2]. These models were calibrated on data collected during the 1974 GATE experiment where the Inter-Tropical Convergence Zone off the coast of Senegal was being observed. It was therefore appropriate to conduct a somewhat similar study based on radar data taken from other tropical regions and compare the results. To this end, a month of continuous radar data was collected at the Patrick Airforce Base, Florida during August 1987 Chapter 2 reports on the findings in this investigation and attempts to explain the differences between the results based on the data from GATE and Florida. The serious impact of these findings on the satellite temporal sampling error is also discussed.

[1] MacConnell, A., and G.R. North, 1987: Sampling errors in satellite estimates of tropical rain. J. of Geophysical Research, 92(D8): 9567-9570.

[2] Shin, K-S, and G.R. North, 1988: Sampling error study for rainfall estimate by satellite using a stochastic model. J of Applied Meteorology, 27 (4): 1218-1231.

List of Symbols in Chapter 2

Z	Radar reflectivity (mm^6/m^3)
R	Rainfall intensity (mm/hr)
z	Measured radar reflectivity (mm^6/m^3)
μ	Mean radar reflectivity in a radar bin (mm^6/m^3)
k	Number of measurements of radar reflectivity in a radar bin
$\gamma(x,h)$	Semi-variogram (mm^2)
Z(x)	Point value of the rainfield at the position x (mm)
σ^2	Point variance of the field Z(x) (mm^2)
μ	Point mean of the Z(x) (mm)
Z(A)	True areal mean of Z(x) for the region A (mm)
Z(A)	Estimated areal mean of Z(x) for the region A (mm)
C(X,Y)	Covariance between two points X and Y on the field Z(x) in the region A (mm^2)
h	Mean distance seperating two points in a region

Chapter 2

On the Variability of Summer Florida Rainfall and it's Significance for the Estimation of Rainfall by Gauges, Radar and Satellite[1]

2.1 Abstract

Simulations based on a month of radar data from the Patrick Airforce Base radar in Florida give the following results for the estimation of daily rainfall amounts over the 124,000 km2 area covered by the radar:

1) 30-minute sampling by a perfect satellite sensor will increase the rainfall measurement error by 7% compared with 5-minute radar sampling

2) If the instrument is only able to determine the raining area and a good climatological rainfall rate is available, then the measurement error is increased to 35%.

---

1 By A.W. Seed and G.L. Austin. Accepted for publication by the Journal of Geophysical Research.

3) Exceptionally dense gauge networks are needed to estimate daily convective rainfall. For example, a network with 625 km^2 per gauge would be required to equal a 30-minute rain-area only instrument.

For monthly areal mean rainfall, the proposed Tropical Rain Measuring Mission (TRMM) sampling strategy with a perfect satellite sensor gives an error of 22%. A 30-minute rain-area only technique combined with a good climatological rainfall rate yields an error of at least 3%. It would seem then that the main contribution of TRMM could be to provide good estimates of the mean climatological rainfall rate (given that it is raining) which could then be used with the geostationary weather satellite to provide the required monthly area rainfall estimates.

2.2 Introduction

There have been many attempts to estimate the accuracy with which daily and monthly areal averaged rainfall amounts can be estimated by various satellite techniques (see Barrett and Martin, 1981, for example). These techniques may be broadly divided into two classes - those based on geostationary satellites which typically have a 1/2-hour time resolution and a few kilometers spatial resolution, and those which operate from low orbiting platforms with a high spatial resolution, but have a temporal resolution of a few overpasses per day. The major thrust of this paper is to investigate the statistical properties of convective rainfields in Florida, and to use these statistics to simulate the sampling problems for raingauge, radar, and satellite rainfall measurement systems. This has been done partly to compliment the earlier studies of tropical rainfall undertaken for the Tropical Rain Measuring Mission (TRMM) satellite, since they were largely based on data collected during the 1974 GATE experiment where the Inter-Tropical Convergence Zone was being observed.

Many authors have recognized over the years that the major problem to be faced in the estimation of area-averaged rainfall amounts is the extreme variability and intermittency of the rainfield in both space and time. Systematic study of this variability has been advocated by Schertzer and Lovejoy (1987) within the theoretical framework of fractals. Many empirical studies have been described in the hydrological literature (e.g. Damant et al. (1983) and the references therein). Bellon and Austin (1986) estimated mean point total storm accumulation differences from raingauges at different spacings, and obtained 60% differences at a distance of 10 km, and 100% differences at a separation of 100 km for rainfields in the mid-latitudes.

It is precisely this variability that limits the accuracy of the rainfall measuring systems, even assuming that the measurements, when and where they are taken, have no error. This paper will attempt to evaluate the lower bound of the measurement error for various rain measuring systems resulting from the inherent variability of the rainfield being measured.

2.3 Analysis Scheme

If a rainfield in space and time which COULD HAVE existed were simulated stochastically, then simulated rainfall estimates based on the known physics of the measuring system could be generated. This idea accounts, in part at least, for the great deal of activity in recent years directed toward the stochastic modelling of rainfall patterns. An alternative strategy, which to some extent avoids the difficulties associated with producing stochastic rainfall models with not only the correct mean and variance, but also higher order statistical properties, is to take a series of 3-D radar data sets, and argue that although they DO NOT represent the actual rainfield at the indicated time, they DO represent a rainfield that COULD HAVE existed.

It is clear that no further assumptions need to be made about the statistical properties of the rainfields except at space and time scales that are smaller than the resolution of the measuring system, perhaps 1 km and 5 minutes. Some would argue, however, that much of the observed discrepancy between radar and gauge data is due to real sub-resolution variability.

Radar data archived at the Patrick Airforce Base (PAFB), Florida, U.S.A., during the period 8th to 30th August, 1987, have been processed to form a single data base of approximately 1000 CAPPIS at 3 km altitude. The raw data tapes were first read to produce a data base consisting of raw digitized radar reflectivities for each 5-minute CAPPI in polar coordinates. These data were then transformed into rainfall intensities by means of the Marshall-Palmer relationship

$$Z = 200R^{1.6} \tag{2.1}$$

The rainfall amounts were then mapped onto a Cartesian coordinate system with a grid spacing of 2 km.

Each polar data point falling within a Cartesian pixel was first converted into the equivalent rain rate before the mean rain rate over the pixel was calculated. The resulting maps were interpreted as if they were true rainfall rates based on the electrical calibration of the radar without being calibrated against the existing gauge network. This was mainly due to the difficulty in obtaining historical raingauge data for that period. For the present study we will argue that while the radar data may not represent the actual rainfall that fell during August 1987, it is a plausible realization of the same process, and therefore has the same statistical structure.

An interactive editing program was written so as to enable strict quality control on the 5-minute CAPPI data. After inspecting the CAPPI's at higher altitudes and watching time-lapse sequences on a graphics system, a significant amount of echo was determined to be ground echo, anomalous propagation and interference. During this phase of the work, it became apparent that an average of 15 minutes of data were missing per hour recorded. This turned out to be due to a routine Airforce procedure which has been eliminated for later data sets. The approximate translation of the rainfield during the period of missing data was calculated by locating the position of the maximum cross-correlation coefficient between the images at either end of the missing period. This translation was evenly distributed over the period, with the image at the start of the missing period being offset by an appropriate amount for each of the missing images.

2.4 Accuracy of the Radar Rainfall Data

There can be no doubt that a weather radar does not measure the rainfall rate with perfect accuracy. The literature is replete with examples of radar-rainfall calibration exercises designed to find the optimum Z/R relationship for particular weather radars in different climatic conditions throughout the world (e.g. Battan, 1973). Since radar measured rainfall will form the basis for this research, it is appropriate to first investigate the effect that the radar induced measurement error will have on the derived statistics.

The causes of radar measurement error may be broadly categorized into rain process dependent errors and radar sampling errors. These sources of error have received wide attention in the literature over the years (e.g Zawadski, 1984 and Austin, 1987). Briefly, the sources of error include:

1. Uncertainty in the Z-R relationship
2. Accretion or evaporation of rain droplets at low altitude
3. Variability in the drop-size distribution
4. Hail
5. Vertical air motions in convective cells
6. Gradients in the rainfields within the sampled volume

Austin (1987) investigated the effect of various physical factors that influence the relationship between measured radar reflectivity and surface raingauge measurements. She found that the natural variability in drop-size spectra was not a major factor once the storms had been classified into intense, moderate and stratiform events, each with a separate Z-R relationship.

After comparing raingauges with the radar for 20 storms, Austin concluded that the two major causes of radar-raingauge measurement errors were:

1. Down-drafts reducing the radar reflectivity for a given rainfall rate and

2. The presence of hail in the intense convective cells.

In her conclusion, Austin states

" We conclude that for research and operational applications where there is a need to know the spatial distribution of surface rainfall or areal rainfall amounts, these quantities can often be measured by radar with sufficient accuracy. In highly convective storms, radar may well be the most (or the only) reliable data source for the desired information".

While this conclusion is encouraging, Austin assumed that the accuracy in the measured reflectance value returned by the radar was of the order of 1.0 to 1.5 dBZ. This accuracy is only attainable if other radar variables are compromised, for example bin length, antenna rotation speed and beam width. It can also only be achieved by means of an intelligent digitizing and averaging scheme. If operational constraints are such that non-optimal radar parameters are chosen, or if an indifferent ditigization scheme is used, these errors may be considerably in excess of 1.5 dBZ.

The operational constraints were such that the radar parameters selected at the Patrick Airforce Base were believed to yield errors in excess of 1.5 dBZ due largely to unusual radio interference regulations enforced for safety reasons. Therefore, it is relevant to re-examine the impact of radar measurement error on accumulated statistics derived from radar data. This section will therefore investigate the impact of various radar digitization schemes on the resulting radar-measured probability distributions. Thereafter, the geostatistical technique of variograms will be used, after some assumptions, to directly estimate the measurement error for various rainfall accumulations. Finally, the estimation variance of mean areal rainfall after averaging in time and space will be estimated.

2.4.1 The Observer's Problem

The problem of inferring the mean reflectivity of a volume of space given a limited number of randomly fluctuating signals is known as the Observer's Problem. The intensity of a signal received by a radar is simply the sum of the signals received from each individual water droplet in the volume being sampled (Marshall and Hitschfeld, 1953). The signal therefore fluctuates as the droplets (or scatterers) move relative to each other in a random fashion. Marshall and Hitschfeld derived the probability distribution of independent echoes from a random array of scatterers and showed that :

$$Pr(Z \geq z) = \frac{1}{\mu} e^{-\frac{z}{\mu}} \tag{2.2}$$

where

z = the measured radar reflectivity, and

μ = mean radar reflectivity for that volume.

This exponential distribution of z implies that several independent measurements of z required before the mean reflectivity μ can be estimated with any degree of confidence. However, in practice, the measurements are not of the same volume of space, but from successive pulses down or cross range. It is obvious that steep gradients in the reflectivity field increases the measurement error. Smith (1964) was able to show from probabilistic considerations that the optimum method for obtaining the mean value of Z was to average the intensity values and not the logarithm of the intensity values.

In general, this method has not been implemented on radar digitizers in the past mainly because the dynamic range of Z extends over 6 orders of magnitude, therefore the amplifier used prior to the digitizer returns the logarithm of the returned signal and not the signal itself. The current popular methods for estimating the mean reflectivity from a number of logarithmic signals include averaging, peak reading, and exponential averaging where the next value is averaged with the mean of the previous values. Modern electronics have removed many of the constraints that restrict the choice of a digitizing algorithm so that a digitizer that averages in Z is now feasible.

Smith's 1964 work examined the measurement error in terms of dBZ and assumed that each dBZ intensity level had an equal probability of occurrence. It would be interesting therefore to repeat some of this early work using an assumed rainfall probability distribution to examine the relationship between measurement error and the number of samples (k) for the various digitizing algorithms.

"True rainfall rates" were generated assuming that instantaneous rainfall rates are drawn from a population of exponential independent random variables. Using the Marshall-Palmer Z-R relationship, 50,000 "true rainfall rates" were generated and converted into an equivalent Z.

Marshall and Hitschfeld (1953) showed that the intensity of independent echoes from a random array of scatterers fluctuates randomly with an exponential distribution. Therefore, each "true value" of Z was used to generate k observations, z_i, $i = 1,k$, drawn from an exponential distribution with a mean Z. In reality, the radar does not sample the same volume of space k times, but rather k pulses either down or cross range. Therefore, the results of this simulation will be somewhat optimistic since the small-scale gradients in the Z field have not been explicitly taken into account.

The "observed" z were then passed through the following algorithms

1) Mean z

2) Mean dBz

3) Mean rainfall

4) Exponential smoothing $\overline{dBZ}_{i+1} = 0.5(dBz_{i+1} + \overline{dBZ}_i)$

5) Peak reading $dBZ = \max(dBz_i) - 0.1(1 + \log_2 k), i = 1, k$

The mean and variance of the "observed" rainfall, the population distribution, the chi squared goodness of fit statistic and the mean standard error expressed as a percentage of the mean were calculated for k = 2 to 64 using the five algorithms.

Table 2.1 lists the mean standard error for the various k and algorithms. It is immediately apparent from this table that the exponential smoothing algorithm has the largest mean square error for k greater than two. The mean standard error for exponential smoothing does not decrease after k greater than four, and therefore if the radar achieves a large k through exponential azimuthal smearing, which is the most commonly used technique, the net result is a radar with both poor spatial resolution and accuracy. The second feature of this table is that the accuracy of instantaneous radar-rainfall measurements can be rather depressing when k is less than four. The optimum algorithm, as expected, was to average the intensity values themselves. The peak-reading technique was nearly as good for k less than four although the difference widens considerably for large k.

The chi squared goodness of fit statistic is tabulated in Table 2.2. This statistic was calculated to determine if the probability distribution of the input rainfall and the distribution of the measured rainfall were significantly different. The chi squared statistic for 17 degrees of freedom at the 95% level of significance is 27.6, and therefore the radar significantly modifies the rainfall probability distribution that it samples, even with a perfect Z-R relationship.

The ratios of input mean verses output mean and input variance verses output variance are found in Tables 2.3 and 2.4 respectively. These Tables show that the technique of averaging the dBz introduces a bias in both the mean and variance, that gets progressively larger as k increases. This would account for the fact that the mean standard error for this technique does not decrease rapidly with increasing k.

Table 2.1 Mean standard error expressed as a percentage of the mean rainfall rate using mean reflectivity (Z), mean dBZ, mean rainfall (R), exponential averaging (Exp), and peak reading (Pr) for various sample sizes (k).

k	Z	dbZ	R	Exp	Pr
2	57	61	56	60	58
4	42	51	42	54	47
8	30	46	31	53	40
16	21	44	24	53	32
32	14	43	20	53	26
64	10	42	17	54	22

Table 2.2 Chi squared goodness of fit statistic for mean reflectivity (Z), mean dBZ, mean rainfall (R), exponential averaging (Exp), and peak reading (Pr) for various sample sizes (k).

k	Z	dbZ	R	Exp	Pr
2	508	1573	627	1505	612
4	141	1435	327	1605	150
8	84	1585	235	1690	68
16	40	1615	190	1510	64
32	27	1694	172	1478	53
64	36	1608	149	1416	75

Table 2.3 Ratio of the estimated mean and actual mean rainfall for mean reflectivity (Z), mean dBZ, mean rainfall (R), exponential averaging (Exp), and peak reading (Pr) for various sample sizes (k).

k	Z	dbZ	R	Exp	Pr
2	0.94	0.80	1.16	0.80	0.91
4	0.97	0.75	0.90	0.76	0.99
8	0.99	0.70	0.90	0.76	1.02
16	0.99	0.70	0.90	0.76	1.02
32	1.00	0.70	0.90	0.76	1.00
64	1.00	0.70	0.90	0.76	0.96

Table 2.4 Ratio of the estimated standard deviation and actual deviation for mean reflectivity (Z), mean dBZ, mean rainfall (R), exponential averaging (Exp), and peak reading (Pr) for various sample sizes (k).

k	Z	dbZ	R	Exp	Pr
2	1.13	1.02	1.08	1.00	1.10
4	1.08	0.87	1.00	0.91	1.11
8	1.04	0.79	0.95	0.90	1.16
16	1.02	0.75	0.92	0.90	1.08
32	1.00	0.72	0.90	0.90	1.04
64	1.00	0.71	0.90	0.90	0.99

2.4.2 Empirical Determination of Measurement Error for the Polar Radar Data

The Observer's Problem is not the only source of measurement errors in the radar data as noted previously. This section will use the geostatistical technique of variograms to attempt to determine the approximate measurement error in the raw radar data. After examining the radar data available for 21 days in August 1987, the data extending from 1500h August. 21 to 2400h August 22, some 220 five-minute rainfall maps, were selected as being representative of heavy convective activity in that area.

A variogram is a geostatistical technique used to uncover the basic structure of a random spatial variable. Using the notation of Journell and Huijbregts (1978), the variogram is defined as

$$2\gamma(x,h) = E\{[Z(x)-Z(x+h)]^2\} \tag{2.3}$$

where $\gamma(x,h)$ is the semi-variogram and

$Z(x)$, $Z(x+h)$ are the values assumed by the random function at the points x and x+h.

Assuming the intrinsic hypothesis, viz. the semi-variogram depends only on the separation vector h and not on the location x, it is possible to estimate the semi-variogram by means of

$$2\gamma(h)=\frac{1}{n}\sum_{i=1}^{n}[z(x_i)-z(x_i+h)]^2 \qquad (2.4)$$

where $z(x_i), z(x_{i+h})$ are the experimental values separated by the vector h.

Figure 2.1 shows the variogram for the raw polar radar data. The variogram was calculated at 0.5km intervals from 0.5km to 10.0km using all pairs of data. The entire data set was used in the analysis since it was assumed that the mean and variance of the random spatial process was constant over the 220 maps. Only points that had non-zero rainfall at positions x and x+h were used in the analysis. The variogram was thus calculated using 130,000 data points. The sample mean for the data was 1.19mm and the sample variance was 3.4mm^2. A straight line was fitted to the first two points on the variogram intersects the origin at about 1.0mm^2 or 29% of the variance. The variogram reached a maximum value at about 4km, therefore the mean rainfall averaged over the area of a radar bin is independent after this distance.

In this study, it is assumed that the measured rainfields derived from the radar data represent fields that could have happened, thus removing from the discussion errors resulting from the very real sources arising from the Z-R conversion and other physical processes, for example evaporation and accretion. Therefore, it is not unreasonable to make the further assumption that the errors in the radar-rainfall measurement process are independent and additive. The measurement error predicted by the variogram could then be thought of as being comprised of errors inherent in the radar measurement of rainfall and would be the lower bound for the actual measurement error which includes all the sources of error.

From Table 2.1, the mean standard error for k = 2, using the peak reading algorithm is 58% of the mean. This implies a measurement error variance of $0.48mm^2$ or 48% of the error variance. Therefore, the Observer's problem alone could account for about half of the observed radar-dependent measurement error in the Patrick Airforce Base radar. Since the operating parameters for this radar are unusual, the error variance for other radars with a larger k would be expected to be less than the 29% of the total variance in this case.

Instantaneous rainfall rates are rarely used in hydrological modelling. Rather, rainfall accumulations of between one hour and one month are used. Therefore, the measurement error of radar-based hourly rainfall is probably of more interest than the instantaneous measurement error. The five-minute polar rainfall maps were accumulated to form hourly rainfall totals, and were similarly processed to form the hourly-rainfall variogram for polar data of Figure 2.2. The first two points in the variogram were used to extrapolate a straight line back to the origin. The radar-dependent measurement error was 10.5mm^2 or 15% of the measured variance of the polar data.

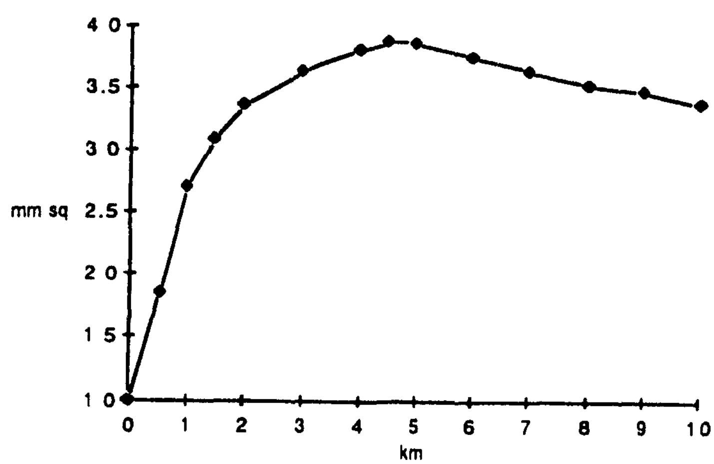

Figure 2.1 Variogram for five minute polar data

In conclusion then, the Patrick Airforce Base over-estimates the five-minute polar variance by 29% and the one-hour polar accumulation by 15%. These should be considered to be the lower bounds for the total measurement error since they are estimates for the radar-derived measurement error only, and do not include the other well known sources of error. It is clear however, that for the purposes of operational hydrology, climatology and as verifying data sets for satellite rainfall estimates, rainfall estimates over larger areas are required. The mapping of the polar data into the Cartesian grid involves further averaging and it is therefore fluctuations in these areal estimates that are of concern. The following section will investigate the effect of further averaging and accumulation on the radar-derived measurement errors.

2.4.3 Variance of the Estimate of Mean Areal Rainfall

In general, rainfall maps based on radar data have undergone two further stages in data processing. Firstly, the data have been transformed from polar to Cartesian coordinate space, and secondly, the data have been accumulated over a period of time. This section develops a technique to estimate the variance of the estimate of the mean areal rainfall rates for 5-minute and hourly accumulations and assesses the relative

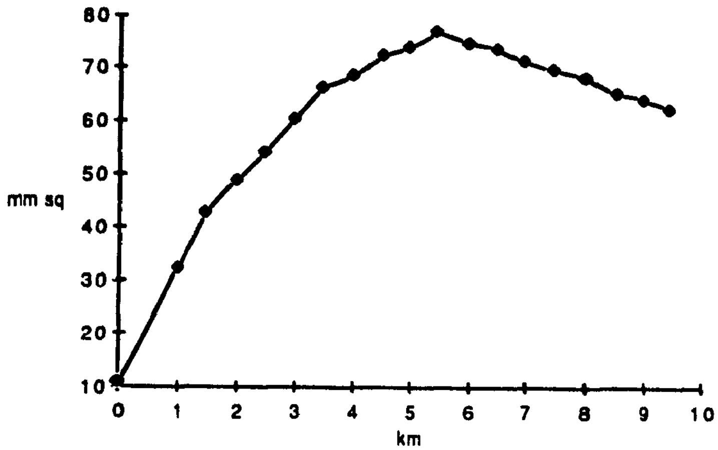

Figure 2.2 Variogram for hourly polar data

importance of the averaging during these two steps.

The variograms in the previous section show that the polar data are significantly correlated over distances of less than two kilometers. If Z(x) is a realization of a homogeneous stochastic process with covariance function C, with

$\mu = E\{Z(x)\}$ and $\sigma^2 = var[Z(x)]$

Ripley (1981) showed that

$\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z(x_i)$ could be used to estimate

$\bar{Z}(A) = \int_A Z(x)\frac{dx}{a}$

where a is the area of A.

The variance of this estimation was shown by Ripley to be

$$var\{\bar{Z}-\hat{Z}(A)\}=\frac{1}{n}[\sigma^2-E\{C(X,Y)\}] \tag{2.5}$$

where

X,Y are independent uniformly distributed points in A.

The variogram $2\gamma(h)=c(0)-c(h)=\sigma^2-c(h)$ for all points h apart. From Figure 2.1 it is apparent that the variogram is plausibly straight for $0 \le |h| \le 2km$. Therefore

$2\gamma(\bar{h})=\sigma^2-E\{C(X,Y)\}$

where $\bar{h}$ is the mean distance between the polar bins for all (X,Y) positions of the centers of the polar bins within the Cartesian pixel. The following steps were required to evaluate the mean estimation variance

1. The mean distance between bins in the same Cartesian pixel was calculated for the particular pixel size, bin length, beam width and number of bins per pixel (n).
2. This mean distance was then used in the polar variogram to estimate $2\gamma(\bar{h})$ and hence the estimation variance for each n.
3. The frequencies for the various n in the Cartesian map were calculated and hence the mean estimation variance over the entire map.

Using this method, it was found that the mean error variance for the estimation of the mean areal rainfall over 2 km pixels is approximately 75% of the variance of the polar data for 5-minute and 71% for the hourly rainfall. The main reason for the rather high error variance is that 82% of the pixels in the map have 1 or 2 polar bins per pixel. The analysis was repeated for 4 km pixels. The mean error variance for the mean areal rainfall in this case was 36% of the variance of the polar data for 5-minute and 26% for the hourly rainfall.

From the above analysis, it is evident that for the Patrick Airforce Base radar, there is not much to be gained by mapping the five-minute rainfall at 2km resolution out to 120km from the radar. At the hourly accumulation level, the radar is able to provide somewhat better rainfall measurements at the 2km resolution, but even so a 4km resolution would be more appropriate.

The mean error variance is inversely proportional to the mean number of polar bins per pixel, whereas there is only a 30% decrease in the error attributable to the Observer's Problem as the number of samples per volume is doubled from 4 to 8. Therefore, it is more effective to use the polar to

Cartesian transformation to smooth the data, rather than to select a large number of samples per volume with a commensurate increase in bin size.

The error variance estimated in this section is based on the assumption that the radar-rainfall maps are exactly correct, and thus once again represent the error inherent in sampling such a random field with a perfect Z-R conversion. The figures quoted here should therefore be considered as the lower bound for the error variance, a more complete analysis would have to include the Z-R conversion error structure. Given the non-trivial errors in the radar-rainfall measurement, even under optimistic assumptions, it would be worth while to use this data set to develop a feel for the magnitude of the errors arising simply as a result of the different space and time sampling strategies of raingauge networks and satellite rainfall estimation schemes. The following two sections briefly simulate the sampling strategies of the raingauge network and satellite measurement systems.

2.5 Raingauge Sampling Errors

It was thought a priori that some form of Kriging would be the most suitable method to interpolate the random scatter of rain gauges onto the Cartesian grid. The average variogram for the daily accumulations and the variogram for the monthly accumulation is shown in Figure 2.3. From this figure it is apparent that daily convective rainfall has a correlation length of 14 km and monthly rainfall decorrelates over a distance of approximately 20 km. Therefore Kriging, which relies heavily on inter-station correlations will not yield anything more obviously useful than say, Thiessen polygons, which at least have the virtue of being easy to compute.

Uniform random raingauge networks were generated for networks of 25 through to 1,000 gauges covering an area of 125,000 km^2. Twenty such networks were generated for each network size. The value of the 2km pixel that covered the gauge position was assigned to each gauge. The mean standard error was calculated for the areal mean, mean rain depth, rain area and rainfall variance for each network density using daily and monthly data. The rain variance was calculated using the "gauge" data only.

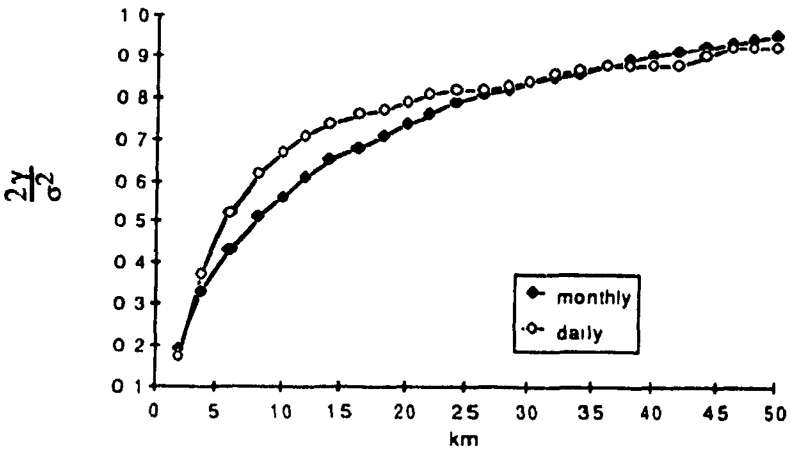

Figure 2.3 Average variogram for daily and monthly rainfall expressed as a fraction of the measured variance

The mean standard errors as a function of gauge density for daily and monthly rainfall are shown in Figures 2.4 and 2.5 respectively. Of the four statistics calculated, the rain variance had the largest standard error. For example, the error in the variance measurement of daily rainfall is 30% for the network of one gauge per 140 km^2. It would seem therefore, that raingauge networks have difficulty estimating the second moment, as predicted by Schertzer and Lovejoy (1987) who postulate that the second moment is not defined. The mean areal monthly rainfall estimated using only 100 gauges had an error of approximately 8%. Therefore, sparse raingauge networks are able to provide reasonable estimates of mean areal monthly rainfall, particularly in areas where the rainfall is not affected by strong local processes.

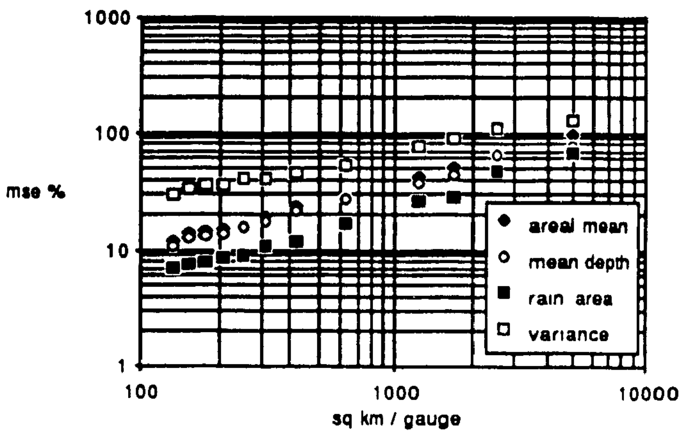

Figure 2.4 Mean standard error for daily mean areal rainfall over 125,000 km^2, mean rain depth, area, and variance vs raingauge network density

2.6 Satellite Sampling Errors

The 1000 radar rainfall maps in the data base were accumulated into hourly, daily and monthly totals of rainfall. There were two days in the data set with no radar data and these were assumed to have had no rain. This is probably not true, but in the spirit of this simulation clearly could have happened. For each number of overpasses per day, equally spaced

5-minute maps were selected, starting from midnight. These maps were then used to calculate the areal daily rainfall over the 480 km diameter area covered by the radar. The simulation was repeated by starting the sequence 5 minutes later until the second time-slot of the original sequence was reached. The daily totals for each simulation were accumulated over the 20 days in the set to yield a number of "monthly" rainfall totals for each sampling frequency. The mean standard monthly error was calculated using these monthly totals.

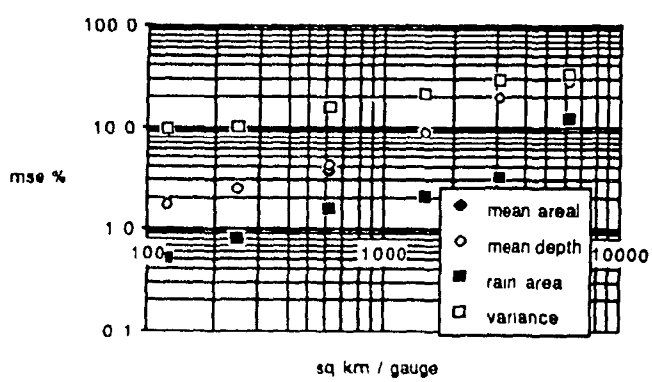

Figure 2.5 Mean standard error for monthly mean areal rainfall over 125,000 km^2, mean rain depth, area and variance vs raingauge network density

It is well known that existing satellite rainfall-measuring techniques have more success in delineating the rain area than estimating the instantaneous rain rates. The infra-red and visible data can estimate respectively cloud height and optical thickness, which can be used to predict whether the cloud is raining. Although there is some correlation between height and rainfall rate for raining clouds, it turns out to be insufficient to determine particularly high rainfall rates, since many clouds that are just as high as the raining cloud show either light or no rain (Lovejoy and Austin, 1979). Therefore, the simulation was repeated using the mean 5-minute rainrate for the 670,000 instantaneous radar rain measurements in the data set.

Figure 2.6 shows the 25, 50, and 75 percentiles for the daily errors using the 20 days. The proposed TRMM sampling frequency is two visits per day. For daily rainfall, the TRMM will have a 130% error. The mean standard error for 30 minute sampling increases from 7% to 32% if the satellite measures rain area only. It would seem then, that there is little to be gained by measuring the mean areal rainfall more frequently than at 90 minute intervals if the sensor can measure raining area only. However, for a perfect instrument, the error approximately halves when the frequency is increased from 16 to 32 visits per day.

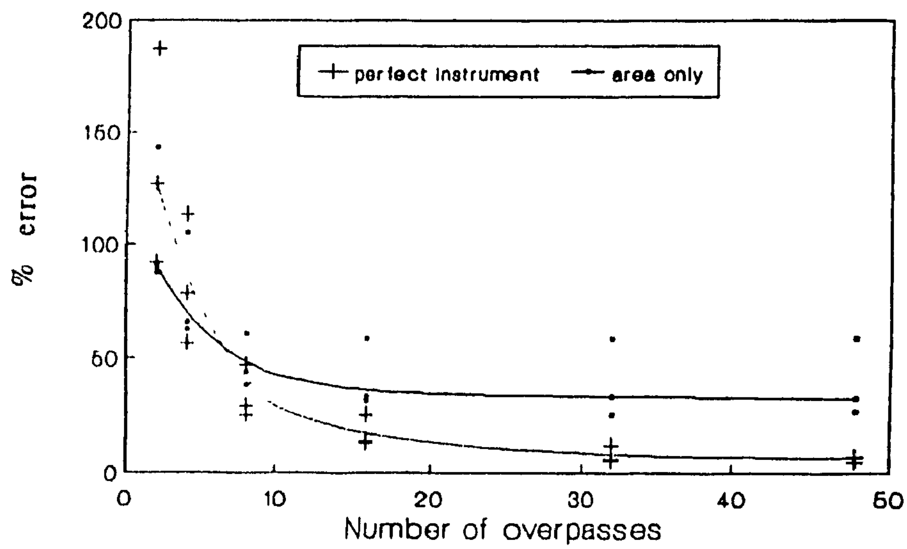

Figure 2.6 The 25, 50, and 75 percentiles of the errors for daily areal rainfall vs number of overpasses per day

Figure 2.7 shows the 25, 50, and 75 percentiles of the errors for monthly totals. The TRMM error for this data would be approximately 22% if the entire area was covered by the satellite at each overpass. It is interesting to note the large spread in the estimation error for fewer than eight overpasses per day. This compares rather unfavourably with the 8% reported by Shin and North (1988) using the GATE data set. It was thought that the difference was probably due to the fact that Florida rainfall is more intermittent than that observed during the GATE experiment, since the GATE experiment was conducted in the Inter-tropical Convergence Zone (ITCZ) which is known to have less intermittent rainfall than Florida. The auto-correlation function for mean areal rainfall over a (360 x 360) km area was calculated and is shown in Figure 2.8. It is evident from this Figure that the decorrelation time is of the order of 3 hours if the zero rain rates are included and 2 hours if they are excluded from the analysis. The mean areal rainfall for the period 8th August to 30th August was 0.1 mm/hr with a standard deviation of 0.3 mm/hr. The probability of zero 5-minute rainfall over the area was 0.86 for the period. These statistics are quite different from those calculated for the GATE data where the decorrelation time for a (280 x 280) km area was found to be 7.7 hours (Bell et al, 1988).

Decreasing sensor resolution was simulated using the 2 km resolution radar data to calculate the mean areal rainfall over 4, 8, 16, 32 and 64 km pixels. These hourly and daily rainfall maps were then resampled at the 2 km resolution, and used to calculate the rain variance, mean depth and area within a portion of the original rainfall map. The results for hourly and daily accumulations are shown in Figures 2.9 and 2.10 respectively. The 64 km pixel data introduces a serious bias in the rain area, 700% for hourly and 300% for daily data, with a commensurate drop in the mean areal rain rate.

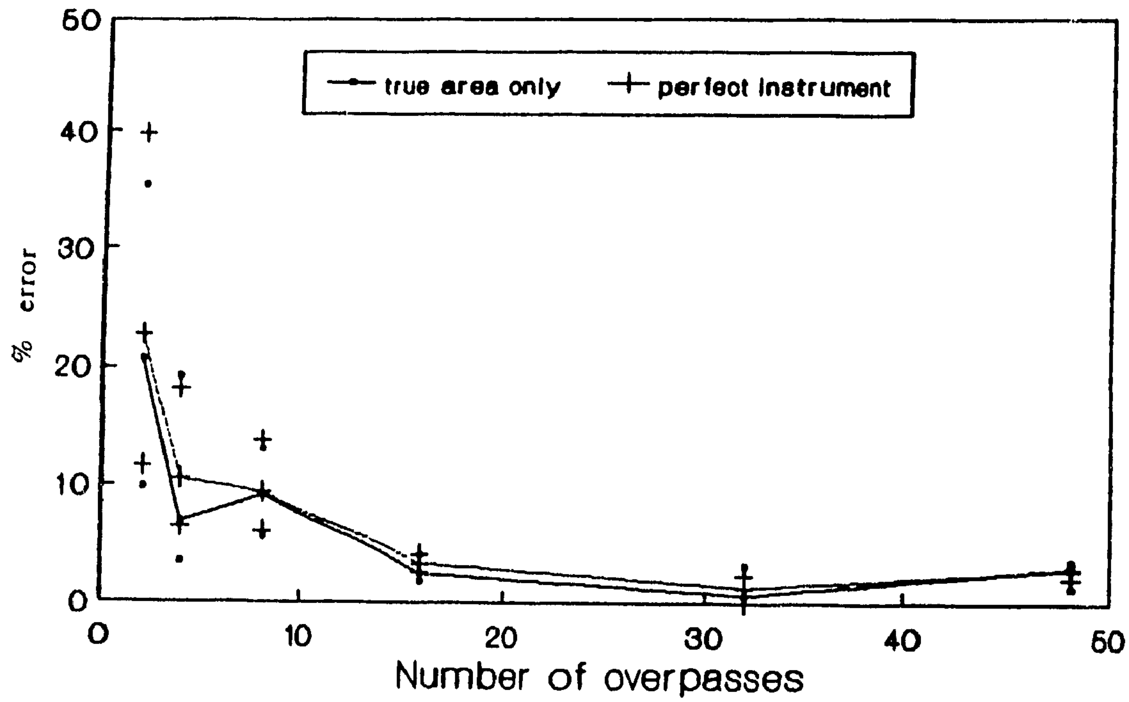

Figure 2.7 The 25, 50, and 75 percentiles of the errors for monthly areal rainfall vs number of overpasses per day

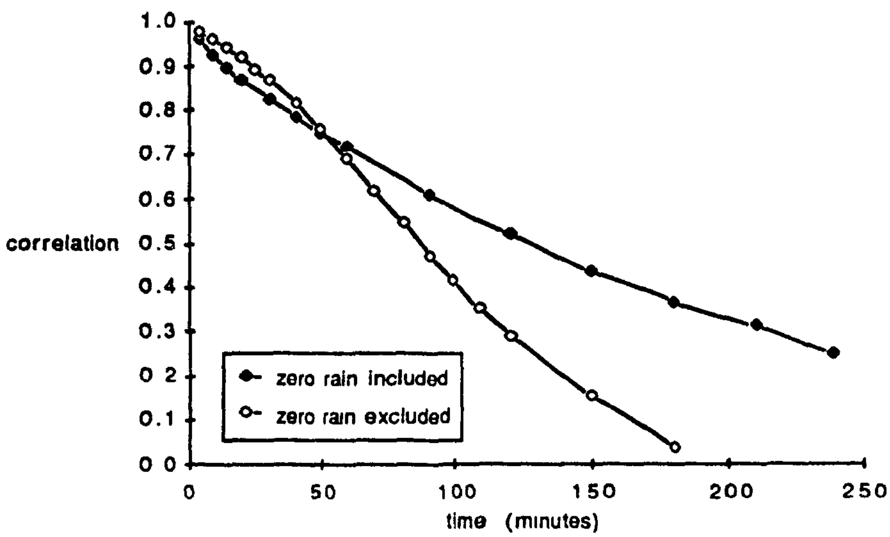

Figure 2.8 Auto-correlation vs time for mean areal rainfall over (367 x 367) km area

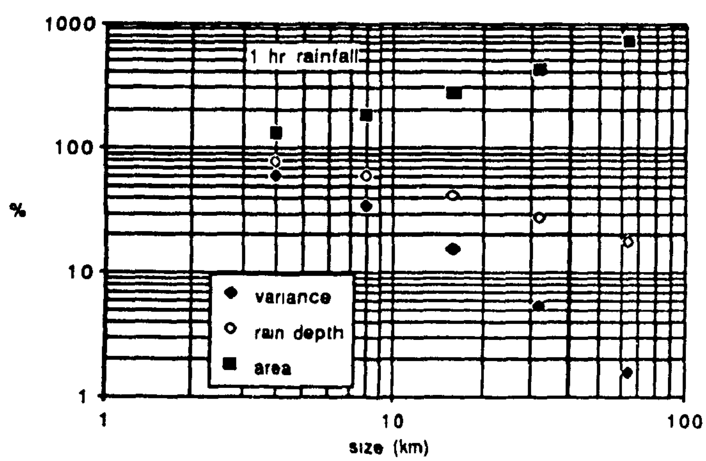

Figure 2.9 The variance, mean depth and rain area for hourly rainfall using a 4, 8, 16, 32, 64 km resolution sensor, expressed as a percentage of the statistic for the 2 km map

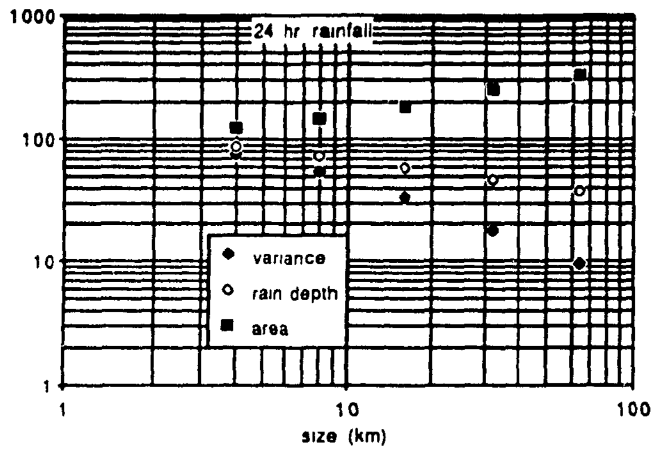

Figure 2.10 The variance, mean depth and rain area for daily rainfall using a 4, 8, 16, 32, 64 km resolution sensor, expressed as a percentage of the statistic for the 2 km map

2.7 Conclusions

Examination of the statistical and sampling problems associated with weather radars in general, leads us to believe that the data from the Patrick Airforce Base radar, when combined to give hourly accumulations on a Cartesian grid of 4 km, estimates the actual mean areal rainfall over a pixel with a variance equal to 26% of the variance of the hourly polar data. The analysis also suggests that the optimum recording strategy for radar data is to record at high spatial and temporal resolution, with few independent samples if necessary, and then average during the polar-to-Cartesian conversion and time integration using an appropriate pixel size. The usual strategy of combining a large number of polar data with exponential averaging not only produces low spatial resolution data, but also does not remove the majority of the statistical fluctuations.

The random sampling error of the radar is substantially reduced if the data are accumulated over periods as short as one hour provided the radar sampling period is set at five minutes. The systematic bias with range is potentially more serious and can be reduced by restricting the effective range of the radar.

Convective rainfall in Florida is intermittent even when averaged over 124,000 km^2. Therefore, the measurement error for perfect instruments increases sharply as the frequency of observation decreases. Exceptionally dense raingauge networks are required to measure daily convective rainfall. For example, a random network with a gauge density on one gauge per 625 km2 would yield estimates equivalent to the 30-minute area-only instrument. The 30-minute perfect instrument is equivalent to a gauge density as high as one gauge per 150 km^2.

The picture improves somewhat for monthly areal mean rainfall. The proposed TRMM sampling frequency of twice per day would be expected to give errors of the order of 22% if the satellite covered the entire area with each overpass. Rainfall patterns tend to be clustered in space and therefore the measurement errors for a system that only partially samples the area with each overpass are likely to be larger than this. A 30-minute area-only instrument would have an error equal to approximately 3% assuming that the climatological mean rainfall rate given that it is raining is known for that area.

The raingauge network with a density of one gauge per 2,500 km^2 estimated monthly mean areal rainfall with an error of only 8%. This could be partially due to the fact that the monthly rainfield is fairly smooth since there is no significant topography in the radar coverage area and apparently few local driving mechanisms. It is clear the between 8 and 16 visits per day are required to meet the stated TRMM objective of measuring monthly mean areal rainfall with better than 10% error over a (500 x 500) km area. The main contribution from the TRMM experiment would then be to estimate the climatological mean rainfall rate (given that it is raining) which could then be used in conjunction with the geostationary weather satellites to provide the required estimates.

2.8 Acknowledgements

The authors would like to thank the personnel at the Cape Canaveral Airforce Station Forecast Office for their assistance in the collection of the data presented here. Mr. David Tees' assistance in the preparation of the paper is acknowledged. Mr. Seed is indebted to the South African Department of Water Affairs for Educational Leave.

2.9 References

AUSTIN, P.M., 1987. Relationship between measured radar reflectivity and surface rainfall. Mon. Wea. Rev., 115, 1053-1070.

BARRETT, E.C., and D.W. Martin, 1981. The use of satellite data in rainfall monitoring. Academic Press, London, 352.

BATTAN, L.J., 1973. Radar observation of the atmosphere. Univ. of Chicago, 324.

BELL, T.L., A. Abdullah, R.L. Martin and G.R. North, 1988. Sampling errors for satellite-derived tropical rainfall: Monte Carlo study using space-time stochastic model. This issue.

BELLON, A. and G.L. Austin, 1986. On the relative accuracy of satellite and raingauge rainfall measurements over middle latitudes during daylight hours. J. Climate and Appl. Meteor., 25, 1712-1724.

DAMANT, C., G.L. Austin, A. Bellon, and R.S. Broughton, 1983. Errors in the Thiessen technique for estimating areal rain amounts using weather radar data. J. Hydrol., 62, 81-94.

JOURNEL, A.G., and Ch.J. Huijbregts, 1978. Mining Geostatistics, Academic Press, London, 26-57.

LOVEJOY, S., and G.L. Austin, 1979. The delineation of rain areas from visible and IR satellite data for GATE and mid-latitudes. Atmos.-Ocean, 17, 77-92.

MARSHALL, J.S., and W. Hitschfeld, 1953. Interpretation of the fluctuating echo from randomly distributed scatterers. Part 1. Canadian J. of Physics, 11, 962-993.

RIPLEY, B.D., 1981. Spatial Statistics. John Wiley and Sons, New York, 252.

SCHERTZER, D., and S. Lovejoy, 1987. Physical modelling and analysis of rain and clouds by anisotropic scaling multiplicative process. J. Geophys. Res., 92, 9636-9714.

SHIN, K., and G.R. North, 1987. Sampling error study for rainfall estimates by satellite using a stochastic model. Toyko TRMM Conf. 1987.

SMITH, P.L., 1964. Interpretation of the fluctuating echo from randomly distributed scatterers. Part 3. McGill Univ. Stormy Weather Group, Report MW-39, 70.

ZAWADSKI, I., 1984. Factors affecting the precision of radar measurements of rain. 22nd Conf. on Radar Meteor., 10-13 Sept. 1984, Zurich, Switerland.

Prolegomenon to Chapter 3

Estimates of mean areal rainfall are required as basic input into hydrological models used to model the transformation of rainfall into river flow. The bulk of these areal rainfall estimates are based on effectively point raingauge measurements and therefore the areal rainfall is only known to within some fairly large error margin. A major problem to be overcome when assessing the error in an estimate of areal rainfall, is that the true areal rainfall is not known. Two common approaches to this problem have been to either use a dense raingauge network[1] to estimate the true areal rainfall, or to use a stochastic model to generate a synthetic rainfield[2]. Damant, et al 1983[3] were possibly the first to use radar estimates of areal rainfall as input data to such a simulation. Chapter 3 will build on this work and will investigate the influence of the geometry of various raingauge networks and rainfield characteristics on the measurement error in areal rainfall estimates based on gauges over large areas.

[1] Hendrick, R C., and G.M Comer, 1970: Space variations of precipitation and implications for raingauge network design J of Hydrology 10· 151-163.

[2] Wilson, C B , J B Valdes, and I Rodriguez-Iturbe, 1979· On the influence of the spatial distribution of rainfall on storm runoff Water Resources Res 15 (2) 321-238.

[3] Damant, C , G L Austin, A Bellon, and R S Broughton 1983. Errors in the Thiessen techniques for estimating areal rain amounts using weather radar data J. of Hydrology, 62, 81-94.

List of symbols in Chapter 3

Z	Radar reflectivity (mm^6/m^3)
R	Rainfall intensity (mm/hr)
mse	Mean standard error (mm)
y	Estimated mean areal rainfall (mm)
y	Actual mean areal rainfall (mm)
ε	Normalized error for estimated mean areal rainfall
w(d)	Distance weighting function for a gauge at a distance d
α	Average nearest neighbour distance for the gauge network (km)
E	Mean standard error expressed as a percentage of the areal mean rainfall
V	Ratio of the standard deviation of the rainfield at a point over the mean rain depth at a point
A	Ratio of the area with non-zero rainfall over the total area covered by the gauge network
N	Number of gauges in the gauge network
a,b,c	Empirical constants

Chapter 3

Sampling Errors for Raingauge Derived Mean Areal Daily and Monthly Rainfall[1]

3.1 Abstract

Radar data from two geographical locations are used to simulate the mean standard error in using a sparse raingauge network to estimate daily and monthly mean areal convective rainfall over areas ranging from 45,000 to 180,000 km^2. It was found that a network with a regular configuration gave somewhat less variable errors than the uniform random raingauge network, although the mean errors were very similar. The difference became more pronounced for the very sparse networks. The mean standard error for a particular network and rainfield was found to be a function of the number of gauges in the network, the raining fraction of the area and the ratio of the standard deviation over the mean of the non-zero portion of the rainfield. A simple three parameter relationship was proposed to relate the mean standard error, expressed as a percentage of the mean areal rainfall, to these variables. It was found that

---

1 By A.W. Seed, G.L. Austin and R.S. Broughton

a single relationship was able to explain 63% of the variability in the estimated mean standard estimation error, combining data from both regions. Finally, the domain over which the relationship is able to make reasonable predictions is discussed, the principal constraint being that the raining fraction of the area should not exceed 0.5 for networks with more than 200 raingauges.

3.2 Introduction

In spite of the heavy emphasis that the modern literature places on various exotic rainfall measuring systems, the old fashioned raingauge still provides the bulk of the rainfall data to practising hydrologists and climatologists throughout the world, and will inevitably have a major role to play in calibrating the new generation of rainfall estimating satellites as well as weather radars. This is simply due to the length of record that exists for the raingauge as compared with the radar for example. The accuracy of raingauge derived mean areal rainfall needs to be understood before gauge data can be used as ground truth for satellite or radar rainfall measuring systems and operational stream flow forcasting. A fundamental question that has to be asked therefore, is how well can a raingauge network with a particular geometrical configuration measure mean areal rainfall over fairly large areas.

Given the long history of the raingauge, it is surprising that we are as yet unable to measure rainfall with high accuracy, even at a point. Rainfall measured by a raingauge is strongly affected by small scale wind effects and local turbulence around the lip of the gauge. The local topography surrounding the gauge, particularly the slope and aspect also affect the gauge measurement e.g.. Rodda (1971). Rodda compared the annual rainfall measured by pit gauges and standard gauges, and found that the standard gauge measured up to 30% less rainfall in some parts of Britain. It is interesting that this is the order of discrepancies between gauges and "well calibrated radars" (Bellon and Austin, 1984). Once the gauge design and siting guidelines have been established, the best we can hope for is some relative measure of rainfall.

The accuracy with which a gauge network can measure rainfall depends on the variability of the rainfield and the geometric organization of the network. In areas where physiographic factors, distance from the sea, altitude and rain-shadow effects, for example, influence the rainfield, the network configuration should be analyzed in the space of these variables rather than the more usual Cartesian space. Seed

(1987) used multidimensional cluster analysis to identify homogeneous physiographic regions when mapping convective rainfall in a physiographically complex region of Natal, South Africa. The gauge network as it exists in Natal has the classic problem that it is the mountainous areas that have the highest rainfall and river runoff, and therefore are hydrologically the most significant areas, but also have virtually no raingauges due to the practical problems involved in siting and maintaining the gauges.

This study uses a large quantity of radar data from both Florida and South Africa to estimate the measurement error for daily and monthly rainfall accumulations over large areas as a function of the network organization and density. In particular two questions will be addressed, viz.

a) To what extent do the errors in estimating mean areal rainfall over large areas depend on the network configuration as a random or rectangular array? and

b) What are the network and rainfield characteristics that influence the estimation error?

3.3 Basic Data Processing

Radar data continuously archived at the Patrick Air Force Base (PAFB), Florida, U.S.A. during the period from the 8th to 30th

August 1987 and a summer of radar data from Nelspruit, South Africa were processed into 5-minute rainfall maps. The raw data tapes were read and the digitized radar reflectivities were transformed into rainfall intensities by means of the Marshall-Palmer relationship

$$Z = 200R^{1.6}$$

The rainfall amounts were then mapped onto a Cartesian coordinate system with a 2km pixel.

Each polar data point falling within a Cartesian pixel was converted into the equivalent rain rate before the mean rain rate over the pixel was calculated. The resulting maps were interpreted as if they were true rainfall rates based only on the electrical calibration of the radar without being calibrated against the existing gauge network. For the present study we will argue that while the radar may not represent the actual rainfall that fell during the two periods, it is a plausible realization of the same random process, and therefore has the same statistical structure.

Two types of raingauge networks were generated for this study, regular, and uniform random networks. The uniform random networks were generated using a uniform random distribution for the two coordinates of the gauge location. The locations were constrained such that the distance to the nearest neighbour was at least 8km. Each gauge was assigned the value of the map pixel at the gauge position, with no attempt to simulate sub-resolution variability. These "gauge measurements" were then used to re-create the rainfields and the original and estimated fields were compared.

3.4 Selection of the Raingauge Interpolation Scheme

A large variety of methods are available when interpolating from a random scatter of data points in an area onto a regular grid. Perhaps the first of such methods was published by Thiessen (1911) and Thiessen polygons are still widely used in hydrology today. There are basically two types of interpolation schemes, local estimation where only the known points in a restricted neighbourhood are used to interpolate onto an unmeasured point, and global estimation techniques where the entire set of points are used, often by means of a least squares regression.

Trend surface and multiple regression techniques are commonly applied to annual, mean monthly, and mean annual rainfall where the rainfield is non-zero at all points in the map area. Examples of such methods include Hutchinson (1968) who mapped mean annual precipitation in New Zealand and Storr and Ferguson (1973) who mapped mean annual precipitation in British Columbia, Canada. All these regression methods are fairly sensitive to the spatial distribution of the rain gauges, particularly near the edges of the map area, see Whitten (1975) for a discussion on this problem.

Local estimation techniques use the known values within a small neighbourhood around the point of interpolation. There are a great number of operational schemes in use, particularly in the field of computer generated contour mapping. McLain (1974) lists a number of schemes that were common at that time. The underlying assumption of these techniques is that data are more likely to be useful if they were measured near the point of interpolation. Delfiner and Delhomme (1975, p96) made the following interesting comment with regard to distance weighting schemes:

> "Clearly, no general rule can be derived from experiment on particular data and point configurations. Consequently, the choice of a distance weighting function is more or less a matter of personal belief, of tradition or of confidence in the advice of 'influential authorities' ".

Although distance weighting schemes suffer from a certain arbitrariness in the selection of parameters, Ripley (1980) was able to show that in order for the interpolated surface to be smooth in the neighbourhood of the data points the derivative of the weighting function must tend to zero as the distance to the point tends to zero, and that the function should decay at a rate faster than the inverse square of the distance. Distance weighting schemes also do not cope well with clustered data although ad hoc solutions can be used to remedy the situation.

The so called optimal interpolation techniques which are designed to minimize the variance of the interpolation error are another major class of local interpolators. This class of interpolation techniques include the various flavours of Kriging, and methods proposed by Gandin (1965) and Ripley (1980). These techniques do in fact out-perform most of the other interpolation techniques, see Creutin and Obled (1982) and Tabios and Salas (1985) for comparative studies. However problems are experienced when the rainfield has zero rain rates, the "hole effect" in Kriging parlance, which requires special treatment e.g. Creutin (1988) and they are far more expensive in computer time than the other techniques.

The validity of any interpolation scheme has to be seen against the extreme variability in the short-duration accumulated rainfields. The observed existence of extreme gradients and a generally discontinuous behaviour with rain often falling over less than 10% of the area, leads to the conclusion that any interpolation technique will not show great accuracy in these cases. It is only recently that the Meteorological community has started to deal with the extreme intermittency of rainfall and cloud fields as compared with the more traditional variables of temperature, pressure and wind. The underlying cause for this extreme variability is the drastic non-linearity involved in cloud and rain formation. The response from a hydrological point of view is to exercise extreme caution about the likely accuracy of any interpolation scheme, including those of great mathematical complexity.

So then, a method must be selected out of this plethora of competing interpolation schemes. Since a large number of maps will be generated, over 100 for each network density and configuration, using up to 1000 gauges, the method must be above all fast. This more or less restricts the selection to fairly crude distance weighting and Thiessen polygons. However, it is interesting to note that both Tabios and Salas (1985) and Creutin and Obled (1982) found that these crude methods did not perform much worse than the more sophisticated methods, and gave fairly satisfactory results for interpolation accuracies at a point. More recently, Lebel, Bastin, Obled and Creutin (1987), estimated the accuracy of the various techniques in measuring mean areal rainfall over several hundreds of square kilometers for hourly data, and found that Kriging had approximately 25% less measurement error variance than Thiessen polygons for a catchment of 545 km^2 and a dense raingauge network.

A casual examination of a rainfield interpolated by means of Thiessen polygons will be sufficient to convince one that the technique does not produce aesthetically pleasing rainfields. The comparative analysis between regular and random gauge networks will be done using Thiessen polygons, in the interests of reducing the computer time. Thereafter, a distance weighting technique will be used to estimate the measurement error variance as a function of network density and averaging area. Finally, a relationship using raining area, number of gauges, and the variability of the rainfield will be used to predict the mean standard error in mean areal daily rainfall estimates.

3.5 Random vs Regular Raingauge Networks

Since this study was exploratory in nature, and the computer time requirements were substantial if the entire data set were to be analyzed, seven 24 hour accumulations were chosen as the "truth" for the study. The maps included one day of intense convective rainfall, but with a small rain area which was expected to provide the largest errors. The mean standard error

$$mse = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{3.1}$$

where

$\hat{y}_i$ is the estimated mean areal rainfall using the i th gauge network and

y_i is the actual mean areal rainfall for that day,

was calculated for each of the 7 days in the data set using nine raingauge networks. Figure 3.1 shows the maximum, median and minimum mse as a function of network density using random and regular network configurations. The random network gave a larger median and maximum mse for the 7 days analyzed. In particular, the maximum mse for the random network configuration increased more quickly than the maximum mse for the regular network as the network density was decreased.

The errors for each of the nine networks for each of the 7 days were normalized by means of

$$\epsilon = \left(\frac{\hat{y}_i - y_i}{y_i} \right) \times 100 \tag{3.2}$$

The cumulative distribution of ϵ for the 900 km per gauge density network is shown in Figure 3.2. From this figure it is clear that the random network configuration produces longer tails, particularly on the over-estimation tail. The under-estimation tail is constrained by -100 by construction, the worst underestimation possible is to measure none of the rain that fell on that day.

It is interesting to note that the rate at which the maximum mse decreases with increasing gauge density is noticeably faster after 400 km^2 per gauge, a mean spacing of 20 km. The probability of the raingauge network making a drastic error rapidly decreases once the network density exceeds 400 km^2. Therefore, although the rate at which the mse decreases with increasing gauge density is depressingly slow, the dense network at least has a fairly constant error from one day to the

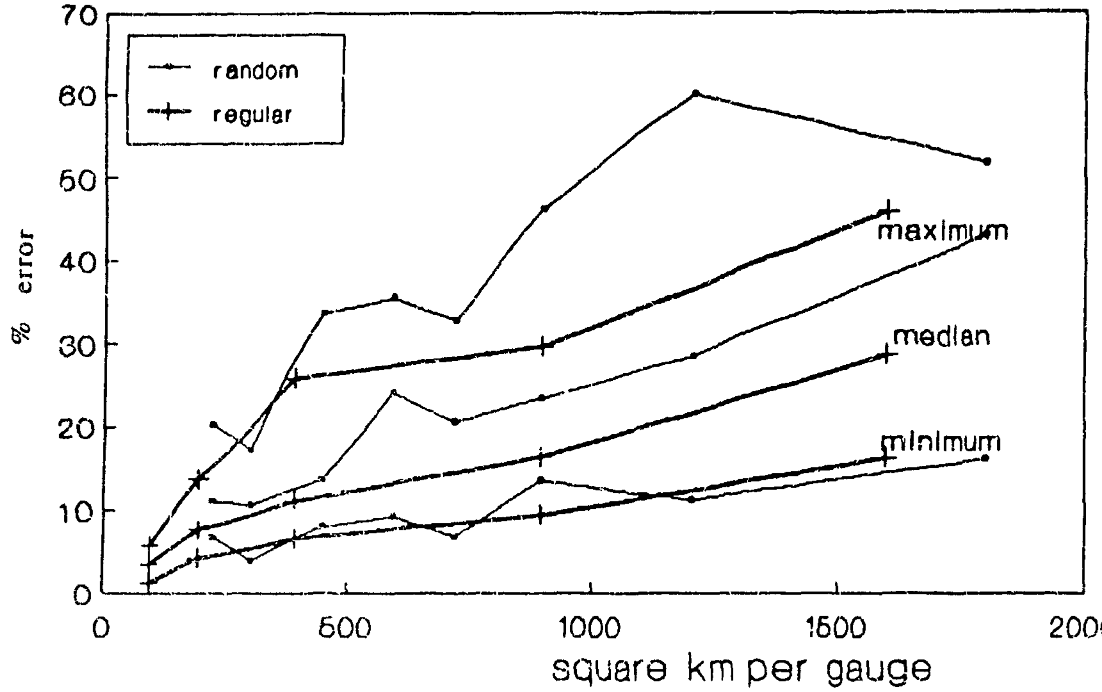

Figure 3.1 Errors in the estimation of mean areal daily rainfall over 180,000 km^2 using Thiessen polygons with regular and random raingauge networks

next, in contrast with the sparse network situation where the maximum error can easily be twice the mean error for a particular rainfield and network.

An examination of the mse for the various days and gauge densities revealed that the days with a large rain area in general had a low error and those with small rain areas had high errors. Figure 3.3 shows a plot of error versus rain area for network densities of 100, 200 and 900 km^2 per gauge using a regular network design and Thiessen polygon interpolation. The dependence of error on rain area is striking. Therefore, the rain gauge network would tend to have lower estimation errors on the days of heavy mean areal rainfall since rain area alone is able to account for most of the mean areal rainfall variance (Rosenfeld, Atlas and Short, 1988). The dense network was less sensitive to the rain area since the network was able to sample even very sparse rainfields with good probability.

3.6 Mean Standard Error vs Network Density

The distance weighting function

$$w(d_i) = e^{\left(-\frac{d_i}{\alpha}\right)} \tag{3.3}$$

where

$w(d_i)$ is the weight of the i th gauge a distance d_i away, and

α is the average nearest neighbour distance for the network

was used to generate daily and monthly rainfall maps for the Florida and Nelspruit data. The Nelspruit data had a maximum range of 120 km so the areal mean rainfall for 44,000 km^2 was calculated using 16 days for Nelspruit, and 7 days for Florida,

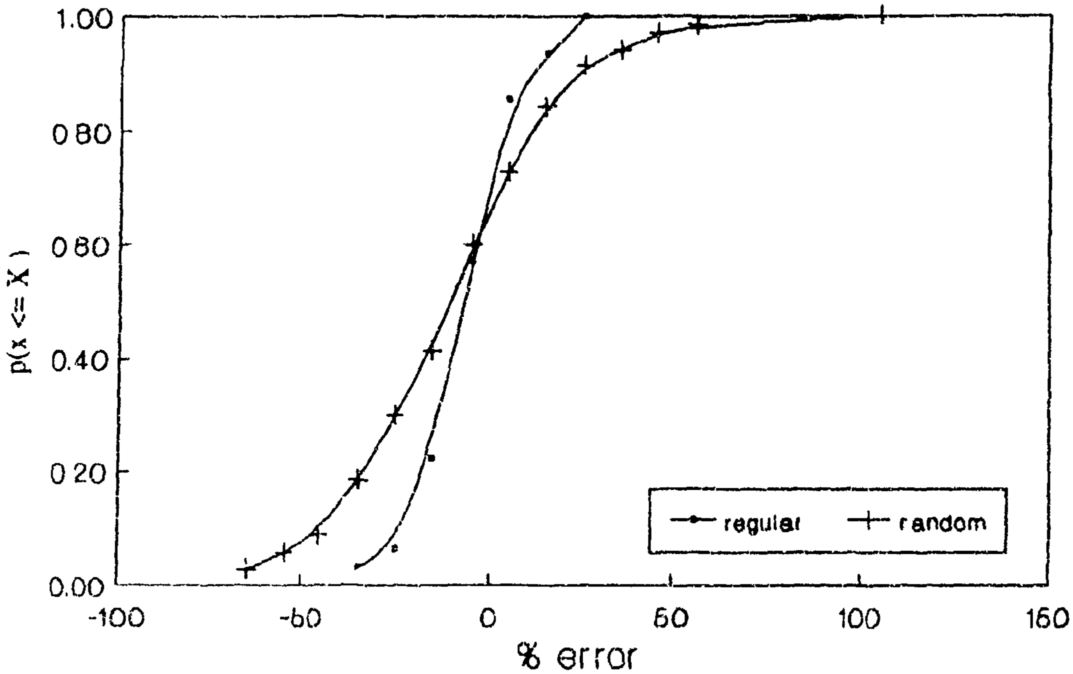

Figure 3.2 Cumulative distribution of normalized errors in the estimation of mean areal rainfall over 180,000 km^2 using Thiessen polygons with regular and random raingauge networks

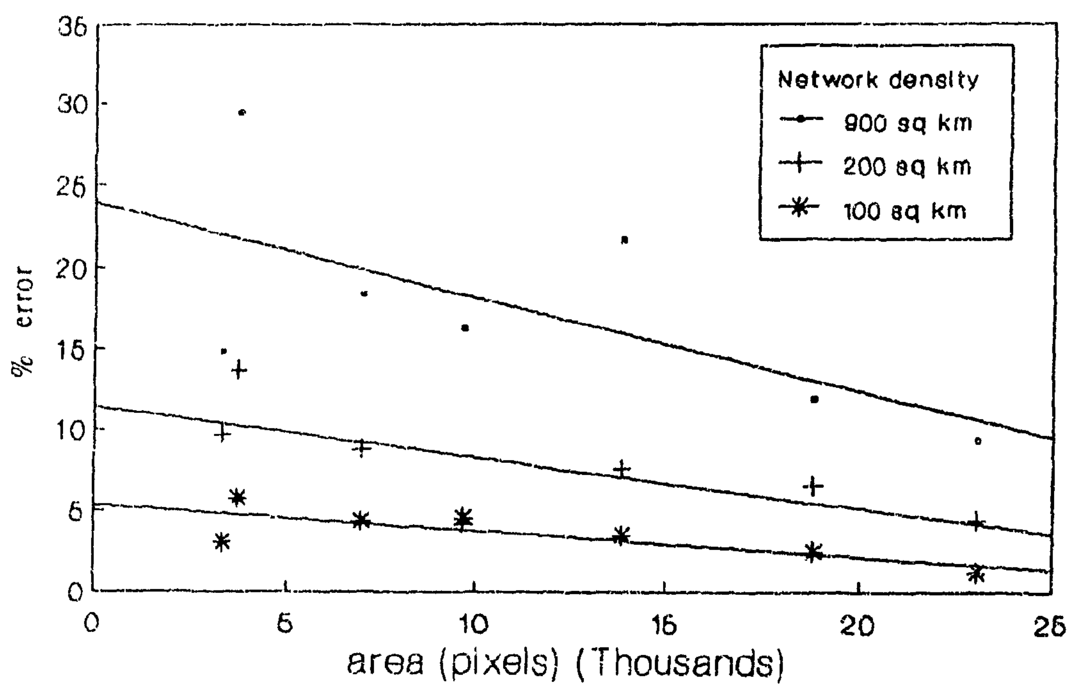

Figure 3.3 Estimation error as a function of rain area for various network densities

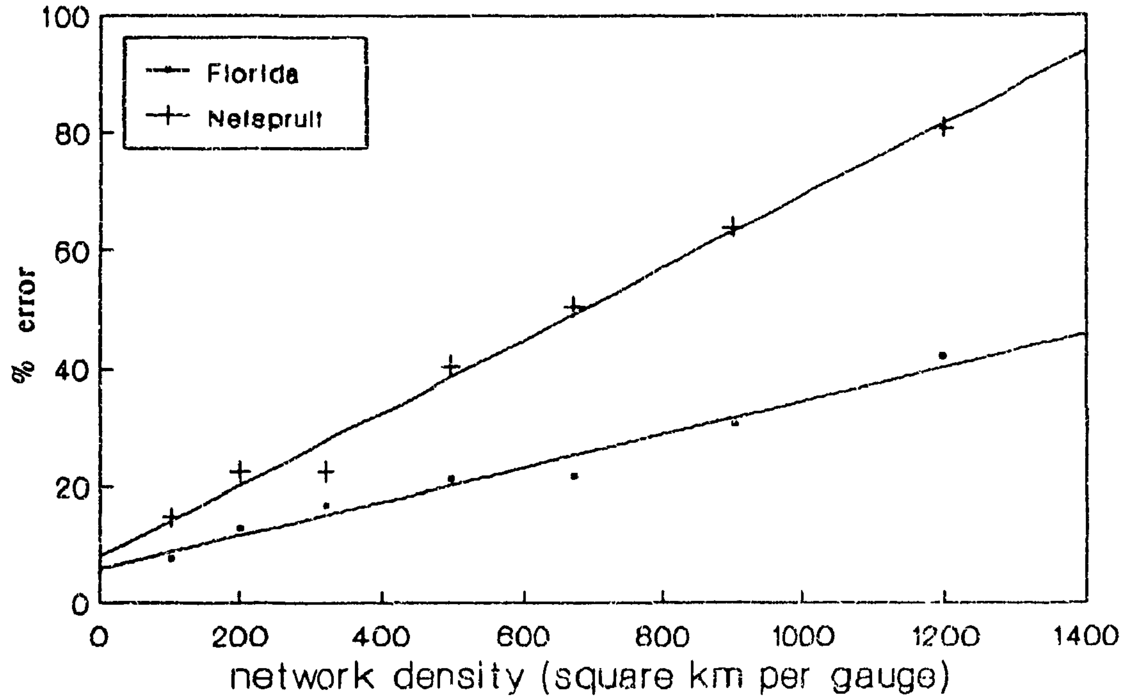

Figure 3.4 Errors in the estimation of mean areal daily rainfall over 45,000 km^2 using distance weighting interpolation

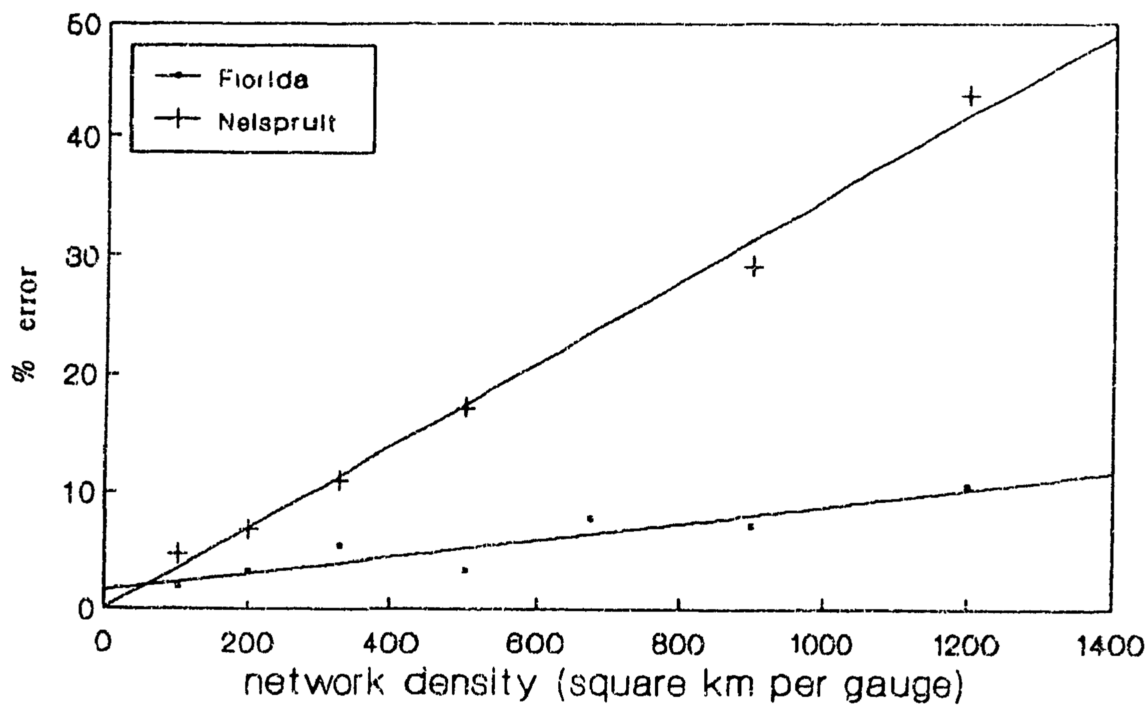

Figure 3.5 Errors in the estimation of mean areal monthly rainfall over 45,000 km^2 using distance weighting interpolation

and nine different gauge networks. The entire simulation took six hours of CPU time on a super-mini computer. Figures 3.4 and 3.5 show the error as a function of gauge network density for daily and monthly data respectively. From these Figures it is apparent that Nelspruit has interpolation errors that are approximately a factor of 2 higher for monthly rainfall and a factor 4 higher for daily rainfall. The ratio of standard deviation over the mean 2 km mean areal rainfall was 1.5 for Florida and 2.47 for Nelspruit, the mean number of raining pixels per day was 2500 for Florida and 3000 for Nelspruit. It seems then, that the raingauge network error is quite sensitive to the relative variability of the rainfield, not an unexpected result.

3.7 Mean Standard Error vs Averaging Area

The Florida data were used to determine the error for monthly mean areal rainfall over 45,000, 100,000, and 180,000 km^2 as a function of network density. For sparse networks, the correlation between any two raingauges is likely to be slight, and therefore the measurement error, expressed as the mean standard error, would be expected to be inversely proportional to the square-root of the number of gauges, independent of the averaging area. Figure 3.6 shows a log-log plot of measurement error verses number of gauges for the three averaging areas. Unfortunately, the maximum range of the radar at Nelspruit limited the coverage area to 44,000 km^2 and therefore it was not possible to repeat the analysis for those data. While the results from an analysis using only one monthly mean areal rainfall map can hardly be considered conclusive, Figure 3.6 suggests that error is indeed independent of averaging area, provided the network is sparse in the sense that the mean correlation between any two gauges is small. It should also be noted that the gauges in the various networks were deployed over the entire averaging area.

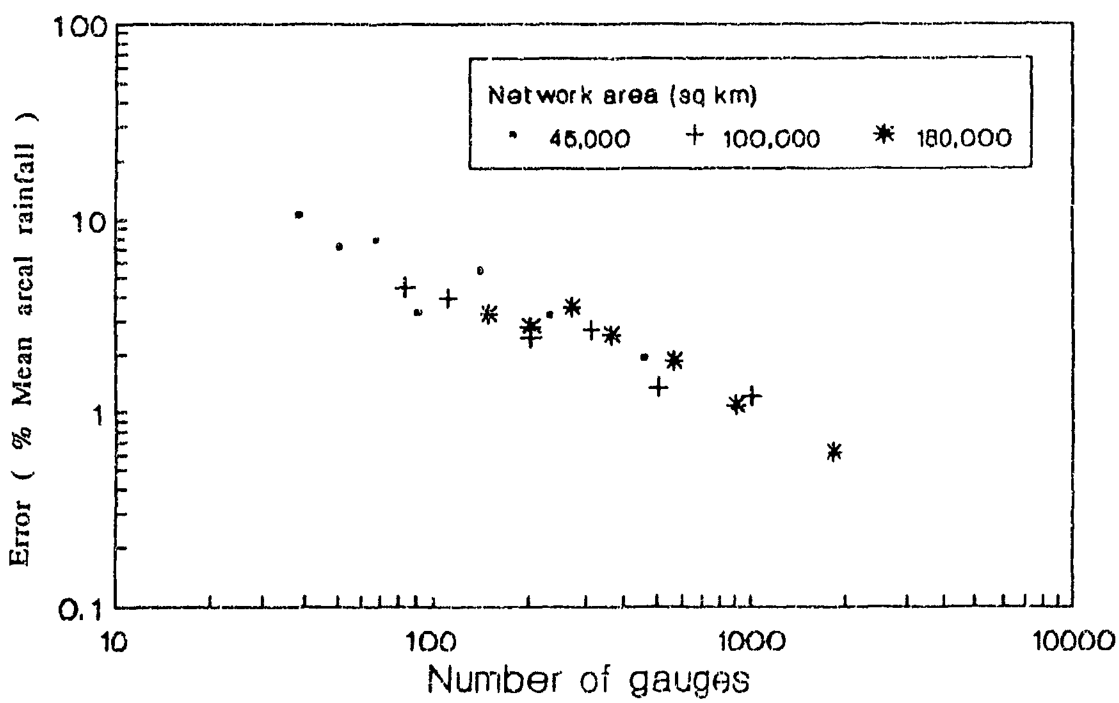

Figure 3.6 Estimation error for mean areal monthly rainfall over 45,000, 100,000, and 180,000 km^2 as a function of the number of gauges

3.8 Mean Standard Estimation Error for Daily Mean Areal Rainfall

From the above analysis, it is apparent that the estimation error for mean areal rainfall over large areas and sparse networks, the most common kind of network, is a function of the number of gauges, the raining area and the variability of the non-zero fraction of the rainfield. A function of the form

$$E = (a + bV)N^{-05} - cA \qquad (3.4)$$

where

E = mean standard error expressed as a percentage of the areal mean rainfall

V = standard deviation / mean

A = raining area / total area

N = the number of gauges in the network, and

a,b,c are empirical constants to be estimated from the data

would seem a reasonable first guess, assuming that N and V have no influence on c.

A least-squares fit was undertaken for the 16 days of Nelspruit data and 7 days of Florida data first separately and then using the combined data set. Table 3.1 gives a summary of the three sets of parameters and Figure 3.7 plots the estimated vs predicted E using the combined data set parameters.

From Table 3.1 it can be seen that the a and c parameters were very similar for all three data sets. The second parameter, b, was possibly badly estimated in the Florida case since the 7 days used in the analysis all had very similar, relatively low, variability. Since the regression was based on estimates of the mean standard error, which itself has a large variance, the regression was able to estimate E surprisingly well in explaining 63% of the variance of E. The model has the interesting property of predicting negative errors when the raining fraction exceeds 0.5, the relative variability of the field is low, 1.4 say, and N is of the order 200 or greater. Clearly, the error function is not able to make predictions in this domain. The lower limit for the raining fraction (A) appears to be somewhat less than 0.05, the lower limit for A in the combined data set. The data set to hand only had two days with A greater than 0.5, making it impossible to explore this domain more thoroughly.

Of course, the actual measurement error for a given rainfield and network can rather easily be very much higher than the mean measurement error, the variance of E increasing as the mean increases. A complete description of E must therefore include an estimate of the variance or preferably a description of the probability distribution.

Table 3.1 Summary of the model fit to the Nelspruit, Florida and combined data sets

	a	b	c	r2
Nelspruit	461.6	25.0	44.5	0.50
Florida	416.5	75.8	44.6	0.66
Combined	411.7	42.0[2]	39.1	0.63

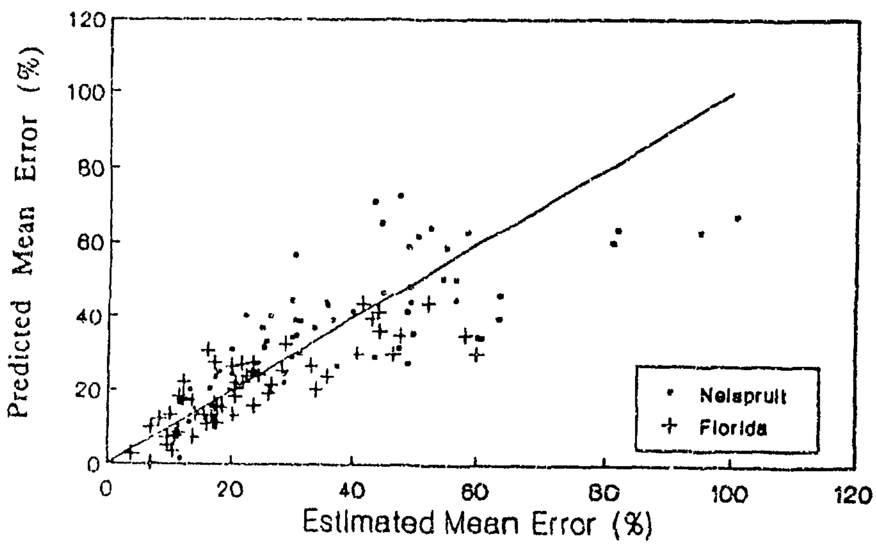

Figure 3.7 Predicted vs estimated mean standard error

---

2 Influential authorities, e.g. Adams (1979) claim that this value has a deep significance

3.9 Conclusions

In conclusion then, it has been found that the regular networks were somewhat better than the uniform random networks, in so far as the variance of the estimation error tended to be lower for regular networks, particularly for the more sparse networks. As the number of gauges increased, the difference between the two configurations, as expected, diminished.

The mean standard estimation error is independent of the averaging area if the gauge density is expressed as the number of gauges in the network and not in square kilometers per gauge. This only holds true when the density of the network is less than one gauge per 15 km, in which case the inter-station correlations can be ignored. Since this is of the order of the correlation length for convective rainfields, one could speculate that this result, suitably scaled by the correlation length, can be applied to shorter rainfall accumulations over smaller areas.

After the number of gauges in the network, the raining fraction of the area covered by the network was found to significantly affect the network measurement error. In general, the small rain areas gave the largest mean measurement error, once again this effect is greatest for the more sparse networks. A simple model to predict the mean standard measurement error given the variability of the rainfield, number of gauges in the network and, the raining fraction of the area covered by the network was proposed. The model was able to explain 63% of the variance in a combined data set using data from South Africa and Florida. Two of the three parameters were quite similar when estimated from the South African and Florida data separately. The nature of the model is such that it is unable to predict measurement errors for raining fractions that exceed about 0.5, except when the number of raingauges less than about 200.

3.10 References

Adams, D., 1979. The Hitch-hikers guide to the Galaxy. Pan Books.

Creutin, J.D., and C. Barancourt, 1988. Pattern and variability analysis of heavy rainfields in a mountainous Mediterranean Region. Conference on Mesoscale Precipitation: Analysis, simulation and forecasting. M.I.T. Sept 13-16 1988.

Creutin, J.D., and C. Obled, 1982. Objective analysis and mapping techniques for rainfall fields: An objective comparison. Water Resour. Res., 413-431.

Delfiner, P., and J.P. Delhomme. 1975. Optimum interpolation by Kriging. In: Davis, J.C. and McCullagh, M.J. (Editors). Display and analysis of spatial data. NATO Advanced Studies, Geneva, 96-112.

Gandin, J.S., 1965. Objective analysis of meteorological fields. Translated from Russian by R. Hardin, p 242, Israel Program for Scientific Translation, Jerusalem.

Hutchinson, F., 1968. An analysis of the effect of topography on rainfall in the Taieri catchment area, Ontago. Earth Science, 2, 51-68.

Lebel, T., G. Bastin, C. Obled, and J.D. Creutin, 1987. On the accuracy of areal rainfall estimation: A case study. Water Resources Research, Vol. 23, No 11, 2123-2134.

McLain, D.H., 1974. Drawing contours from arbitrary data points. Computer Journal, 17, 318-324.

Ripley, B., 1980. Spatial statistics. John Wiley and Sons, London, 44-75.

Rodda, J.C., 1971. The precipitation measurement paradox: The instrument accuracy problem. W.M.O. Report No. 316, Geneva, 42pp.

Rosenfeld, D., D. Atlas, and D.A. Short, 1988. The estimation of convective rainfall by area integrals part 2: The height area rainfall threshold (HART) method. Conference on Mesoscale Precipitation: Analysis, simulation and forecasting. M.I.T., Cambridge, Sept 13-17, 1988.

Seed, A.W., 1987. The techniques of rainfall mapping. Unpublished M.Sc. thesis, Dept Ag. Eng., University of Natal, South Africa.

Storr, and Ferguson, 1973. The distribution of precipitation in some mountainous Canadian watersheds. Proc. Symp. On mountainous areas. W.M.O., Report 326, Geneva: 244-263.

Tabios, G.Q., and T.D. Salas, 1985. A comparative analysis of techniques for spatial interpolation of precipitation. Water Resour. Bull., 21(3), 365-380.

Thiessen, A.H., 1911. Precipitation averages for large areas. Monthly Weather Review, July 1911, 1082-1084.

Whitten, E.H.T., 1975. Practical use of trend surface analysis in Geological Sciences. In: Davis, J.C., and M.J. McCullagh (Editors). Display and analysis of spatial data. NATO Advanced Studies, Geneva, 82-96.

Prolegomenon to Chapter 4

Since the conversion of rainfall into runoff by a river catchment is a spatially averaging process, and since point-rainfall measurements in general are the only data available to hydrologists, the transformation of point-rainfall statistics into equivalent areal-rainfall statistics is of fundamental concern to hydrologists. This concern is particularly important in the case of estimating exrtreme events for the design of drainage systems. The normal situation is to know for example the 50 year return period for 1 hour accumulation at a point from raingauge data. What the designer requires is the 50 year return period accumulation over the catchment area of the culvert. Empirical methods to effect such a transformation have in the past been based on raingauge network data usually from geographical regions located far from the location of the installation[1]. It would seem that radar data, with its excellent spatial resolution, is a rather natural data source for such a study Chapter 4 uses radar data as a basis for an empirical investigation into the sensitivity of the rainfall probability distribution to averaging area.

[1] U.S Weather Bureau, 1957-60: Rainfall intensity-frequency regime, Part 1; the Ohio Valley; Part 2; southeastern United States, Tech. Paper No 29, U.S. Department of Commerce, Washington, D.C.

List of symbols in Chapter 4

R_{2km}	Mean areal rainfall depth over (2 x 2) km area (mm)
$R_{n\ km}$	Mean areal rainfall depth over (n x n) km area (mm)
a,b	Empirical constants

Chapter 4

On the Sensitivity of the Rainfall Probability Distribution to Averaging Area[1]

4.1 Abstract

An empirical investigation on how the probability distribution of mean areal rainfall responds to varying degrees of spatial averaging was undertaken. Twenty one days of 5-minute radar rainfall data were collected at the Patrick Air Force Base, Florida, U.S.A. During August 1987. These data were processed into maps of 5-minute, hourly and daily rainfall. The maps were then successively averaged, and the probability distributions for the various scales of spatial smoothing were calculated. These distributions were then used to estimate the rain-rate for various levels of exceedance probability. A function was fitted to describe the mapping of mean areal rainfall from one scale to the next whilst keeping the exceedance probability constant over the scale change. The analysis was repeated for two meteorologically different

1 By A.W. Seed, V.T.V. Nguyen, and G.L. Austin

events, namely 12 consecutive hours of general rain and 12 consecutive hours of heavy convective rainfall. It was found that a power-law closely fits the mapping of the probability distribution over changes in the scale of measurement. The parameters for the transformation were found to be dependent on the characteristics of the meteorological system that produced the rainfields. This approach is believed to be more convenient for practical applications than previously available techniques.

4.2 Introduction

A knowledge of the point to area transformation of the rainfall distribution is essential to the hydrologist when attempting to apply statistics derived from point rainfall measurements to mean areal rainfall. This situation arises frequently since most rainfall estimates are made from sparse gauge networks. The question to be answered in point to areal rainfall transformations is: "Given the point rainfall for a certain level of probability at an arbitrary point on the area , what is the average rainfall over the area for the same level of probability?" (Raudkivi, 1979, and Nguyen, 1984). For many years areal reduction factors (ARFs) have been used to transform the point rainfall depth into an equivalent mean areal rainfall. Perhaps the most widely used of these have been the U.S. Weather Bureau (1957) curves, details of which can be found in Bras and Rodriguez-Iturbe (1985).

More recently , Rodriguez-Iturbe and Mejia (1974) developed a method that used either an exponential or Bessel correlation structure together with an estimate of the mean distance separating two random points within the area to calculate the ARF. Myers and Zehr (1980) fitted surfaces in area-duration space to five different gauge-pair statistics to estimate the upper and lower bounds of the first and second moments of the annual maximum mean areal rainfall series. These moments were then used to estimate the mean areal rainfall depth for various return periods, and hence the ARF for each return period used. Nguyen (1984) assumed that hourly rainfall accumulations had a mixed distribution with the non-zero rainfall values distributed as exponential, non-identical, non-independent random variables. Nguyen was able to derive an expression for the distribution of the areal mean rainfall and hence the ARF for any exceedance probability. Niemczynowicz (1984) developed ARFs for short duration, 1 to 40 minutes, and small areas, up to 25 km2 using 12 gauges to cover an area of 25 km^2 in the central part of Lund, Sweden , over a period of three years.

Previous studies (see e.g. Myers and Zehr, 1980; Niemczynowicz, 1984) have found that the ARF depended on the rainfall duration, exceedance probability and area.

Many of the current methods are not suited to short duration, small area rainfall typical of most urban hydrology problems since they are based on long duration, large area data. This lead Arnell et al (1984) in their review of rainfall data for urban hydrology to state that "there is a lack of fully developed and experimentally justified models for transferring point rainfall to areal rainfall". However, any method to derive the ARF for short durations and small areas using rain gauge data will first have to deal with the problem of estimating the mean areal rainfall using the point data. Radar data on the other hand, allows one to estimate the mean areal rainfall rather more easily, assuming a well behaved Z-R relationship. While radar data have excellent space-time coverage, it measures rainfall neither directly nor perfectly. Austin (1987) found the radar gave 15% errors for storm totals over a number of gauges, therefore the radar is expected to measure instantaneous rain rates over small areas with a far higher error variance. Rain gauges measure the rainfall at a point with a better, but not perfect, accuracy. Exceptionally dense networks, however, are required to measure short duration convective areal rainfall. Huff (1979), cited by Arnell et al (1984) for example found that a network with a mean spacing of 1.8 km was required to explain 90% of the rainfield variance.

None of the current techniques have taken the meteorology that produced the rainfields explicitly into account, although the importance of a synoptic meteorological appraisal was noted by Myers and Zehr, (1980). It seems quite plausible that rainfields produced by large scale general rain events would respond quite differently to spatial and temporal smoothing than would rainfields resulting from small scale isolated convective storms, for example. Furthermore, since the scale of the area over which the rainfield is averaged varies from hectares in the case of urban hydrology to thousands of square kilometers for the case of water resource hydrology, it seems likely that extreme run-off events could be caused by different types of meteorological events, depending on the scale of the catchment in question.

This paper will use radar rainfall data to provide some experimental insight on how the probability distributions of short duration rainfields are transformed under spatial averaging over the range 2 km to 64 km. An empirical method to model the transformation of the probability distribution under spatial averaging will be tested. The model will then be used to determine if the rainfields resulting from widespread general rain and convective rain respond differently to spatial averaging.

Weather radar data archived over 21 days during August 1987 at the Patrick Airforce Base, Florida, U.S.A. were processed to form a data base of some 1000 5-minute rain maps at 2 km resolution. These maps were then carefully inspected and edited to remove echoes resulting from anomalous propagation, electrical interference and ground clutter. The maximum range of the radar was restricted to 240 km so as to reduce radar range effects.

4.3 Method

The region covered by the radar was subdivided into non-overlapping square regions, 64 km on a side. The data within these squares were then successively averaged to produce distributions of 4 km through to 64 km mean areal rainfall. The distributions for 5-minute, hourly and daily mean non-zero areal rainfall are shown in Figures 4.1, 4.2, and 4.3 respectively. The rain rates for various levels of exceedance probability were then estimated from the empirical distributions and plotted against one another. Figures 4.4, 4.5, and 4.6 show some of these plots for 5-minute, hourly and daily rain rates respectively. Based on the straightness of the log-log plots the following relationship was assumed

$$R_{2km} = a R_{nkm}^{b} \tag{4.1}$$

where

R2 km and Rn km are the mean areal rainfall over 2 km and n km respectively, and

a,b are the fitted parameters for n = 4, 8, 16, 32, and 64 km.

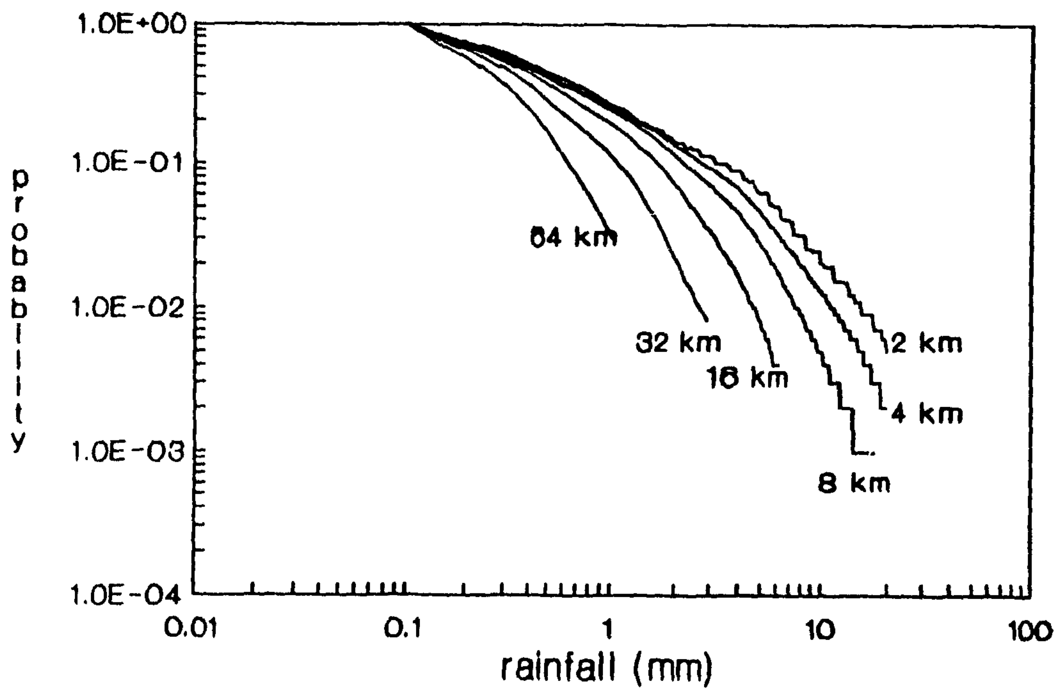

Figure 4.1 Probability distributions for 5-minute rainfall accumulations

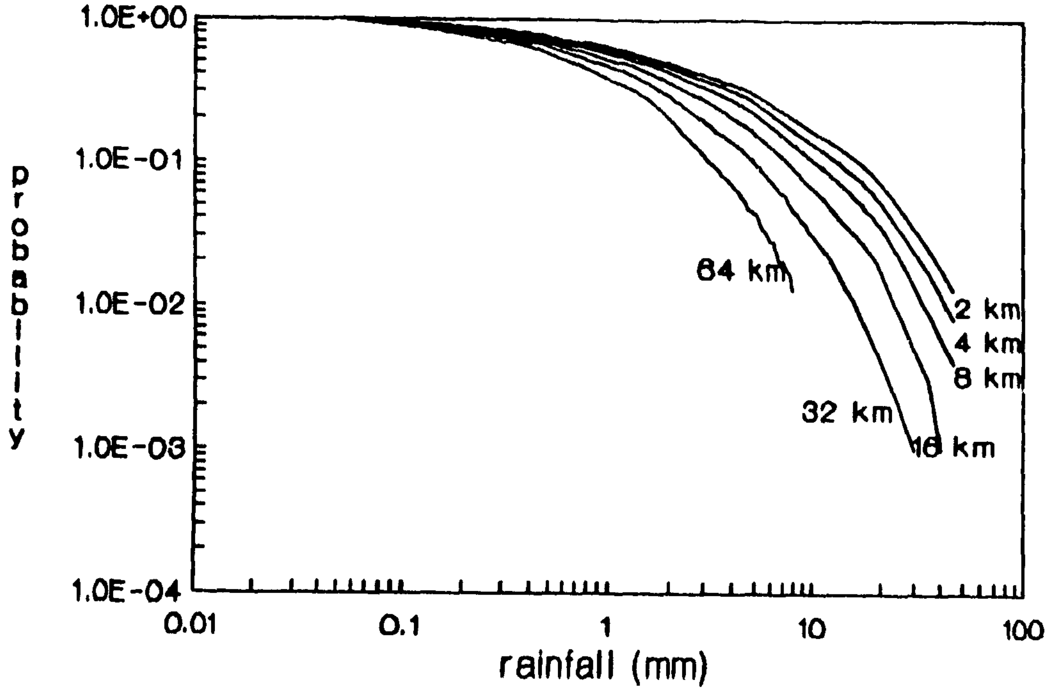

Figure 4.2 Probability distributions for 1-hour rainfall accumulations

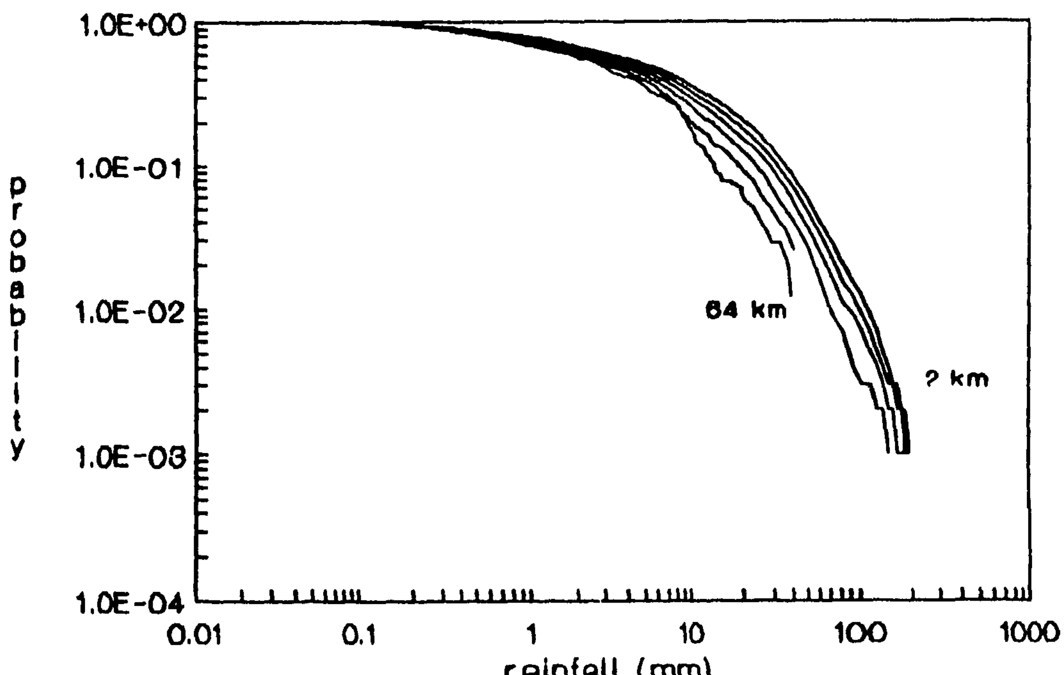

Figure 4.3 Probability distributions for 24-hr rainfall accumulations

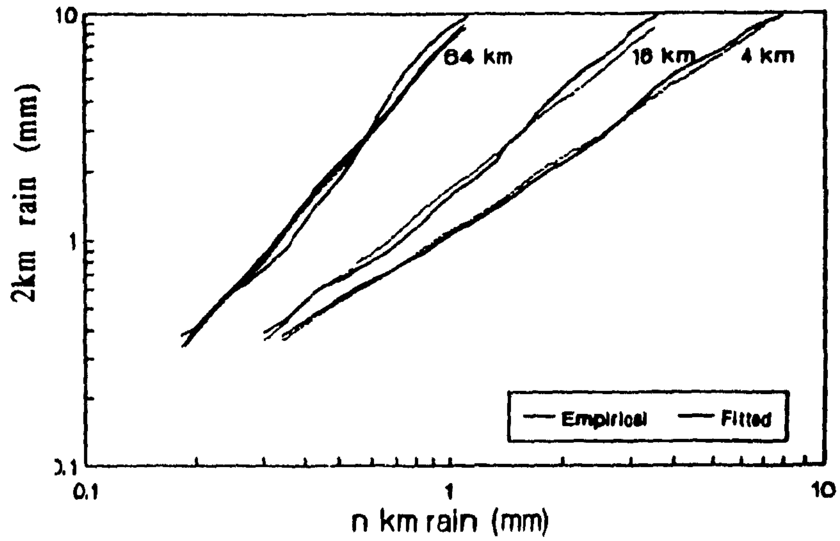

Figure 4.4 5-minute mean areal rainfall over 4 km, 16 km, and 64 km plotted against mean areal rainfall over 2 km at equal probability

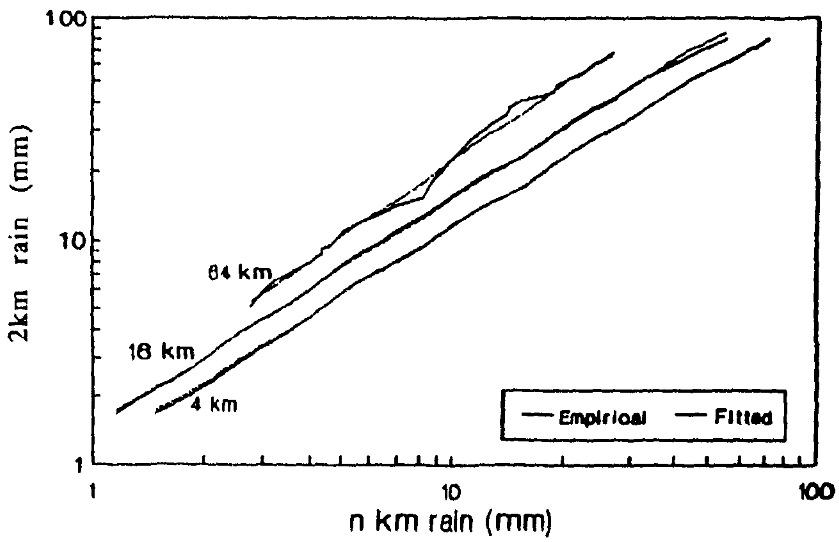

Figure 4.5 Hourly mean areal rainfall over 4 km, 16 km, and 64 km plotted against mean areal rainfall over 2 km at equal probability

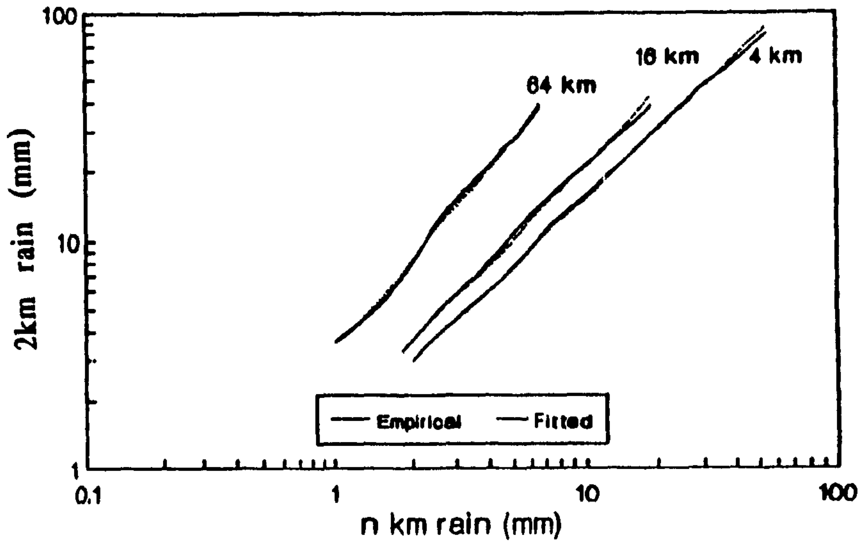

Figure 4.6 Daily mean areal rainfall over 4 km, 16 km, and 64 km plotted against mean areal rainfall over 2 km at equal probability

This equation is identical to that proposed by Lin (1976) to transform 60-minute rainfall probability distributions into shorter rainfall accumulation distributions.

Least squares regression was used to estimate the two parameters a and b in (4.1) for each transformation. Figs 4.4, 4.5 and 4.6 also show the best fit for the various transformations, supporting the power law assumption. Figures 4.7 and 4.8 show the "a" and "b" parameters as functions of the averaging area and rainfall accumulation period when transforming the

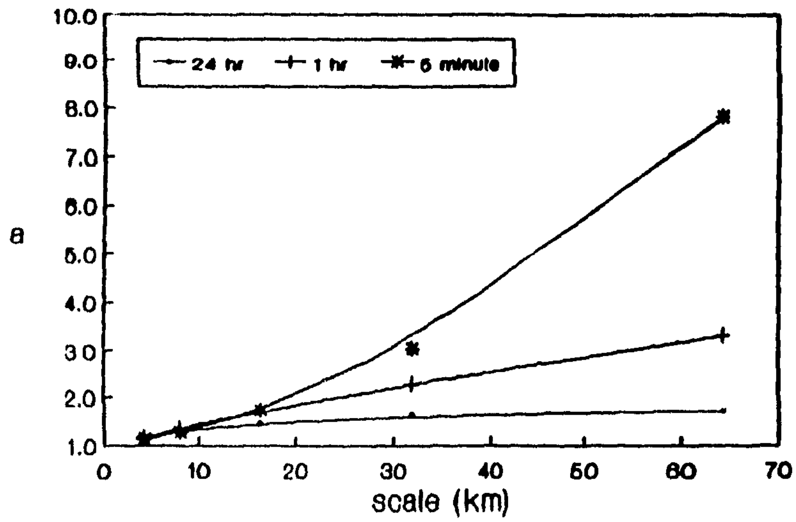

Figure 4.7 "a" coefficients vs scale for 5-minute, hourly, and daily rainfall accumulations

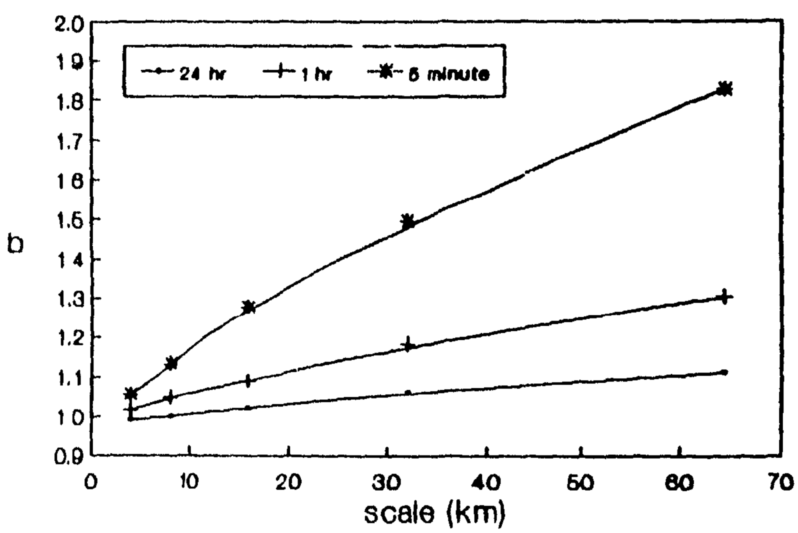

Figure 4.8 "b" coefficients vs scale for 5-minute, hourly, and daily rainfall accumulations

probability distributions to 2 km. Tables 4.1 and 4.2 show the "a" and "b" parameters for hourly rainfall under various spatial transformations.

It is interesting to note that the transformation from say 64 km to 32 km is more severe than the transformation from 4 km to 2 km. This fact is also apparent from the probability distributions in Figure 4.2, where the lines converge as the scale is systematically decreased. This would lead one to assume that the point distributions as derived from rain gauge data are not too different from the 2 km data, particularly for the hourly and daily rainfall accumulations.

To investigate the stability of the "a" and "b" parameters over various meteorological conditions, the analysis was repeated using hourly data for two 12 hour periods in the Florida record. The rainfall on the 14th August 1987 was light and widespread whereas the rainfall on the 22nd August 1987 was intense isolated thunderstorms. Figures 4.9 and 4.10 show "a" and "b" as functions of scale for both 12 hour periods. The curves are derived from the entire record of 21 days

Table 4.1: Table of "a" coefficients for mean areal hourly rainfall for various transformations.

Measured Rain / Transformed Rain	2 km	4 km	8 km	16 km	32 km	64 km
2 km	1.000					
4 km	1.198	1.000				
8 km	1.440	1.199	1.000			
16 km	1.862	1.546	1.281	1.000		
32 km	2.605	2.155	1.772	1.364	1.000	
64 km	3.757	3.097	2.527	1.915	1.374	1.000

is included for reference. From these figures it is quite apparent that the two sets of rainfields respond to spatial averaging in different ways. Not unexpectedly, the widespread rain day had coefficients which were close to unity over the range of scales investigated. The convective case, on the other hand showed coefficients that were strongly dependent on scale. Indeed, the significant gradient in the curves for the convective case would suggest that care should be taken in making the possibly incorrect assumption that the probability distributions for point raingauges are the same as areal averaged radar data.

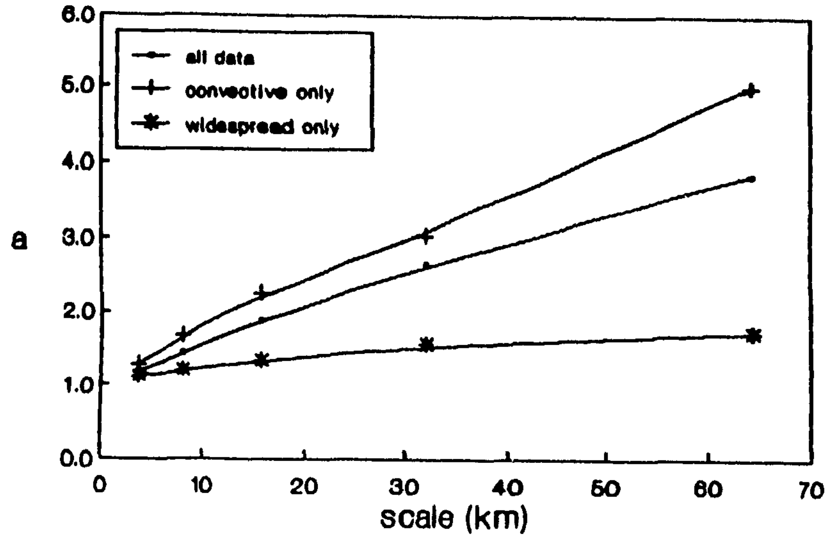

Figure 4.9 "a" coefficients vs scale using hourly rainfall accumulations for convective and widespread rainfall

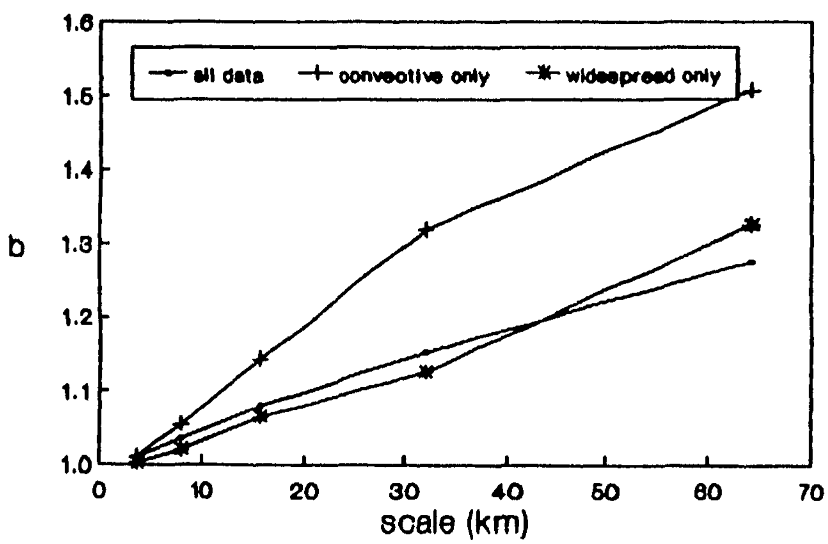

Figure 4.10 "b" coefficients vs scale using hourly rainfall accumulations for convective and widespread rainfall

Table 4.2: Table of b coefficients for mean areal hourly rainfall for various transformations.

Measured Rain / Transformed Rain	2 km	4 km	8 km	16 km	32 km	64 km
2 km	1.000					
4 km	1.012	1.000				
8 km	1.034	1.023	1.000			
16 km	1.080	1.069	1.044	1.000		
32 km	1.152	1.140	1.115	1.067	1.000	
64 km	1.277	1.263	1.234	1.182	1.107	1.000

4.4 Conclusions

It is possible to characterize the way the rainfall probability distributions are transformed under spatial averaging by means of a two parameter power law. The parameters are sensitive to the meteorology that produced the rainfields, the magnitude of the scale change, and the scale itself. It is clear that the spatial organization of the widespread rain type of rainfield is quite different from that of the convective rainfield, and responds differently to spatial averaging. It is therefore not possible to produce only one set of parameters for all weather types, but at least two; one for widespread rain and another for convective events. It is also possible that these transformations depend on the local climatology as well as rain type. This method is more convenient than previous methods because it relates the entire probability distribution at one scale to the distribution at some other scale, and not just particular, convenient, return periods. The method is also remarkably simple, in contrast with the Meyers and Zehr (1980) method, for example, and involves no assumptions about the underlying form of the probability distribution or the correlation structure of the field.

4.5 References

Arnell, V., P. Harremoes, M. Jensen, N.B. Joansen, and J. Niemczynowicz, 1984. Review of rainfall data application for design and analysis. Water Sci. Tech., Vol. 16, Copenhagen, 1-45.

Austin, P.M., 1987. Relationship between measured radar reflectivity and surface rainfall. Mon. Wea. Rev., 115, 1053-1070.

Bras, R. and I. Rodriguez-Iturbe, 1985. Random functions and hydrology. 559pp. Addison-Wesley Publishing Company, Reading.

Huff, F.A., 1979. Spatial and temporal correlation of precipitation in Illinois. Illinois State Water Survey, Circular No 141, ISWS/CIR-141/79, Urbana, U.S.A.

Lin, S.H., 1976. Dependence of rain-rate distribution on rain-gauge integration time. The Bell System Technical Journal, January 1976, 135-141.

Myers, V.A. And R.M. Zehr, 1980. A methodology for point-to-area rainfall frequency ratios. NOAA Technical Report NWS24, NOAA, U.S. Dept. Commerce.

Niemczynowicz, J., 1984. An investigation of the areal and dynamic properties of rainfall and it's influence on runoff generating processes. Report No 1005. Department of Water Resources Engineering, Lunds Institute of Technology, University of Lund.

Nguyen, V-T-V., 1984. Short time increment rainfall analyses for urban drainage studies. Proc. of the Fourth Congress - Asian and Pacific Division, International Association for Hydraulic Research, Chiang Mai-Tailand, 11-13 September, 1984, 1265-1275.

Raudkivi, A.J., 1979. Hydrology: An advanced introduction to hydrological processes and modelling. 479pp. Pergamon Press, Oxford.

Rodriguez-Iturbe, I., And J.M. Mejia, 1974. The design of rainfall networks in time and space. Water Resource Research, 10(4):729-735.

Prolegomenon to Chapter 5

A more appealing approach to the resolution dependence of rainfall statistics than the empirical approach of Chapter 4 would be to derive a scale-independent distribution which could then be used over a range of measurement scales. It is likely that rainfields respond differently to spatial and temporal averaging from one realization of the field to the next, highly intermittent rainfields, arising from air mass thunderstorms for example, would be expected to be quite different from a large scale general rain event. A parameterization of these differences could be quite useful in the machine classification of different rain events. Recently, the conceptual model of multiplicative cascades has been studied[1] and many of the properties of th cascade type models are now known to some extent. These models could be of great interest to hydrologists since they deal directly with the resolution dependence of the field being modeled. Chapter 5 introduces some of the concepts in the cascade type of models and uses radar data to test empirically some of the relationships derived from the theory.

[1] Schertzer, D., and S. Lovejoy, 1987: Physical modeling and analysis of rain and clouds by anisotropic scaling multiplcative processes. J. of Geophysical Research 92 (D8): 9693-9714.

List of symbols in Chapter 5

Z	Radar reflectivity (mm^6/m^3)
γ	Order of singularity
Z_L	Radar reflectivity averaged over an (L x L) km area (mm^6/m^3)
L_0	Effective outer length scale of the cascade process
Z_{L_0}	Ensamble average of Z over the scale L_0 (mm^6/m^3)
$c(\gamma)$	Co-dimension associated with the order of singularity γ
h	Order of the moment
$C(h)$	Co-dimension associated with the moment h
Z_{ob}	Observed radar reflectivity (mm^6/m^3)
MAD	Mean absolute difference
T_0	Effective outer time scale of the cascade process
R_{T_0}	Ensamble average of the rainfield R over the time scale T_0
$\alpha, \alpha', C_1, c_0, \gamma_0, \Delta\gamma, \gamma_c$	Empirical constants

Chapter 5

A Multifractal Approach to Scale Independent Rainfall Probability Distributions[1]

5.1 Abstract

A fundamental problem in exploiting both remotely sensed and in situ network rainfall measurements is that the statistical characteristics of the resulting rainfield depend strongly on the resolution of the measuring device. In this paper we discuss a method to produce resolution independent rainfall probability distributions. The model, conceptually based on multiplicative cascades, uses 6 parameters to describe the transformation of the probability distribution during averaging over various scales of measurement. Four periods of 12 hours of 5-minute data were used in the analysis, three of the periods represented periods of moderate to heavy convective rainfall whereas the fourth represented widespread rainfall. The model was able to predict the distribution of the radar reflectivity (Z) field for a resolution greater than the highest resolution used to estimate the parameters with good

---

1 By A.W. Seed, S. Lovejoy and G.L. Austin

accuracy. It was found that the four days had somewhat different parameters, particularly for the widespread rain day. The method successfully modelled the transformation of the probability distribution of mean rain rate under increased averaging in time whilst keeping the space resolution fixed at 2 km. Two simpler methods involving four and three parameters were tested against the six parameter model. It was found that the four parameter model was nearly as good as the full model, but was easier to calibrate. The three parameter model gave significantly less accurate predictions of the probability distributions over changes in the scale of measurement, even though the numerical values of the parameters did not vary greatly from one meteorological situation to another.

5.2 Introduction

A major problem with the use of rainfall statistics is that the probability distribution depends significantly on the space-time resolution of the data used to obtain the distribution. The problems are particularly acute when attempting to obtain the areal rainfall distribution using raingauge data. The way that a rainfield responds to spatial and temporal averaging has profound implications for the estimation of areal mean rainfall required for hydrological modelling, extreme event analysis and the calibration of rainfall remote sensing systems. It is also possible that rainfields respond differently to spatial and temporal averaging from one realization of the field to the next, highly intermittent rainfields, arising from airmass thunder storms for example, might be expected to be quite different from a large scale general rain event. The behaviour of the rainfield under various levels of temporal and spatial averaging limits the usefulness of many stochastic rainfield models which often ignore the resolution dependence problem, and are calibrated around a narrow band in space-time, see Rodriguez-Iturbe (1986) for a discussion on the limitations of various models.

This paper will use radar-derived rainfields to observe how the rainfall probability distribution is transformed under spatial and temporal averaging. A conceptual model based on multifractal multiplicative cascades will be used to develop a scale independent probability distribution. Models of this sort were originally developed to study the problem of intermittency in turbulence, particularly in an attempt to obtain log-normal distributions for turbulent energy fluxes. The generic form of the probability distributions resulting from multiplicative cascades, their so called "universality classes", can then be used to predict the probability distribution for a given spatial or temporal resolution. The paper has two major sections: 1) introduction to multifractal measures and, 2) application to spatial averaging at a fixed temporal resolution and an application to temporal averaging at a fixed spatial resolution.

5.3 Data

Weather radar data archived during August 1987 at the Patrick Airforce Base, Florida, U.S.A., were processed to form a data base of some 1000 5-minute rain maps at 2 km resolution. These maps were then carefully inspected and edited to remove echoes resulting from anomalous propagation, electrical interference and ground clutter. Four 12-hr periods of continuous 5-minute data were selected for this study. Three of the periods, taken from the 12th, 22nd and 29th August 1987 represented convective rainfall with varying degrees of intensity and spatial organization. The the fourth period, taken from the 14th August 1987, represented widespread, non-convective rainfall. The maximum range of the radar was restricted to 240 km. The area covered by the radar was further reduced to an area covered by non-overlapping 64 km pixels falling within the 240 km radius. Radar reflectivity maps at 2 km resolution were generated using the rainfall data, which were then successively averaged over 4, 8, 16, 32, and 64 km pixels. Histograms of the frequency of the dBZ values for each 12-hour period were built up for each resolution using all of the 144 maps in each sequence, a sample size of 4,320,000 per histogram for the 2 km data. The histograms were then converted into probability distributions. The minimum detectable signal for the data derived from the rainfall data base was 24 dBZ since the minimum rain rate recorded in the rainfall data base was 1.2 mm/hr. This is some 10 dBZ higher than the minimum detectable signal for the raw data. The 2km resolution rain rate data for the 22nd August were averaged over 5, 10, 20, 40, 80, and 160 minutes, and the probability distributions were derived for each averaging period.

5.4 Multifractal measures

The extreme variability and intermittency of the atmosphere results from the concentration of various conserved fluxes, energy for example, into smaller and smaller regions through the action of non-linear interactions and instabilities operating over a wide range of scales. Even when the exact dynamical equations, and corresponding conserved quantities, are not known, it is still likely that such cascades are responsible for much of the observed variability. Fairly recently, research has shown that cascades of this sort, where the large scale multiplicatively modulates the small, when carried out with a repeating scale invariant mechanism over a wide enough range of scales, generally leads to multifractal measures (Schertzer and Lovejoy, 1987). In this paper we will investigate empirically how well rainfields can be described by such multiplicative cascade processes. The analysis of the data presented in the paper will be seen to lend some credence to this hypothesis since the probability distributions over the entire range of scales available, fit into the theoretically predicted functional forms. Fields resulting from such cascade processes can be regarded as superpositions of singularities of order γ, each distributed over sets with fractal dimension $d(\gamma)$. One way of expressing this is by considering the probability distribution of a multifractal field Z_L, radar reflectivity for example, when averaged over some scale L, as a function of the scale

$$Pr\left(\frac{Z_L}{Z_{L_0}} > \left(\frac{L}{L_0}\right)^{-\gamma}\right) \sim \left(\frac{L}{L_0}\right)^{c(\gamma)} \tag{5.1}$$

where

Z_L is the value of the multifractal field, when averaged over the scale L,

Z_{L_0} is the ensamble average,

L_0 is the "effective" external scale of the data, and

$c(\gamma)=d-d(\gamma)$, d is the dimension of the space in which the process occurs.

The above equation shows that the basic scale invariant "co-dimension" function $c(\gamma)$ is really just an appropriately normalized probability distribution:

$$c(\gamma)=\frac{\log Pr\left\{(Z_L/Z_{L_0})>(L/L_0)^{-\gamma}\right\}}{\log(L/L_0)} \tag{5.2}$$

where the second order terms in Eqn (5.1) for equality have been ignored.

This formula has an equivalent statement in terms of the statistical moments of Z_L :

$$<Z_L^h>=L^{-K(h)}=\int L^{h\gamma}dPr=\int L^{(h\gamma-c(\gamma))}d\gamma \tag{5.3}$$

where

h = the order of the moment and

$\langle\ \rangle$ indicates ensemble statistical averaging

Hence, using the method of steepest descents (see Frisch and Parisi, 1985), we obtain:

$$K(h) = \max_{\gamma} (h\gamma - c(\gamma)) \qquad (5.4)$$

This Legendre transformation arises because in the limit $\lambda \to \infty$, for each moment h, there is a corresponding singularity $\gamma(h)$ which dominates the exponent in the integrand:

$$h = c'(\gamma_h)$$

Note that because the Legendre transform is equal to its inverse, we also obtain:

$$c(\gamma) = \max_{h} (h\gamma - K(h)) \qquad (5.5)$$

showing the complete equivalence between a description in terms of moments and probabilities. It is also possible to define another co-dimension function associated with moments of various orders:

$$C(h) = \frac{K(h)}{h-1} \qquad (5.6)$$

A great simplification in multifractal analysis and modelling occurs because for quantities conserved by the cascade, the $c(\gamma)$ function is characterized by the following two parameter functional form or universality class:

$$c(\gamma) = C_1\left(\frac{\gamma}{C_1\alpha'} + \frac{1}{\alpha}\right)^{\alpha'} \tag{5.7}$$

with

$C_1 \leq d$, the dimension of the space in which the process occurs,

$0 \leq \alpha \leq 2$

$$\frac{1}{\alpha} + \frac{1}{\alpha'} = 1$$

$(\Rightarrow \alpha' \geq 2 \; for \; 1 < \alpha \leq 2, \;\; \alpha' < 0 \; for \; 0 \leq \alpha < 1)$

The corresponding universal K(h) function is given by :

$$K(h) = \frac{C_1\alpha'}{\alpha}(h^{\alpha} - h) \tag{5.8}$$

(Schertzer and Lovejoy, 1987)

The above functions are for conserved, stationary quantities and are the multiplicative analogues of the standard central limit theorem for the addition of random variables, the case $\alpha = 2$ corresponding to (log) gaussian processes, and $\alpha < 2$ to (log) Levy processes. For other quantities, related to these by either dimensional and/or power law relations, the corresponding $c(\gamma)$ functions can be obtained by the linear transformation $\gamma \rightarrow a\gamma + b$. If we regard cascade processes as concrete implementations of the idea of proportional effects (e.g. Lopez, 1978), then we see that the latter generally do not yield log-normal distributions, but only approximately log-normal distributions.

For example, in turbulent cascades, the energy flux ϵ is conserved, and the fluctuations in components of the velocity field are obtained by $\Delta v = \epsilon^{1/3} l^{1/3}$ hence a = b = 1/3. For passive scalar clouds (see Schertzer and Lovejoy, 1987, and Wilson et al, 1989 for details on these multifractal cloud and rain models), the corresponding quantities are $\Delta\rho = \psi^{1/3} l^{1/3}$ where $\psi = \chi^{2/3}\epsilon^{-1/2}$ and χ is the variance flux of the passive scalar concentration ρ. Allowing for these linear transformations of γ, we obtain the following more general three parameter universality classes:

$$c(\gamma) = c_0\left(\frac{\gamma}{\gamma_0} + 1\right)^{\alpha'} \qquad (5.9)$$

An equivalent, more convenient form for the three parameter class is obtained by taking $\gamma=\gamma+\Delta\gamma$ in Eqn. (5.7):

$$c_0 = C_1\left(\frac{\Delta\gamma}{C_1\alpha'} + \frac{1}{\alpha}\right)^{\alpha'} \tag{5.10}$$

$$\gamma_0 = \Delta\gamma + \frac{C_1\alpha'}{\alpha}$$

These relationships are useful, since as will be seen, $c_0, \Delta\gamma$ and C_1 are much easier to estimate than α, α' which are essentially measures of the concavity of $c(\gamma)$ which will only be pronounced for large γ. The three parameter universality classes can then be written:

$$c(\gamma) = C_1\left(\frac{\gamma+\Delta\gamma}{C_1\alpha'} + \frac{1}{\alpha}\right)^{\alpha'} \tag{5.11}$$

$$K(h) = \frac{C_1\alpha'}{\alpha}(h^{\alpha} - h) - h\Delta\gamma$$

Before discussing the analysis of radar data using the above formalism, we must first discuss a complication which arises because of a basic distinction between "bare" and "dressed" cascade quantities. The "bare" quantities are essentially theoretical: they are obtained after a cascade process has proceeded only over a finite range of scales;

strictly speaking, Eqn. (5.11) applies only to these quantities. The experimentally accessible quantities are different; they are obtained by integrating the multifractal fields, usually by means of the measuring device, over scales much larger than the inner scale of the cascade, which in the atmosphere is of the order of 1mm. The properties of such spatial (and/or) temporal averages are approximated by those of the "dressed" cascades i.e. those in which the cascade has proceeded down to the small scale limit and then integrated over a finite scale. The small scale limit of these multiplicative processes is singular and is responsible for this basic distinction.

Unlike the bare cascade, the dressed cascade displays the interesting phenomenon of divergence of high order statistical moments, that is:

$$<Z_L^h> \rightarrow \infty \tag{5.12}$$

for all $h \geq h_c$

where h_c is the critical exponent for divergence. The precise condition for divergence is given by :

$$C(h_c) = d(A) \tag{5.13}$$

where

d(A) is the dimension of the averaging set (e.g. line, plane, or fractal in the case of measuring networks) over which the process is averaged (Schertzer and Lovejoy, 1987).

The phenomenon of divergence of high order statistical moments arises directly from the fact that C(h) is generally unbounded, and hence for any averaging set A, for large enough h, C(h) > d(A). In the universality classes above, the only exception occurs when $\alpha<1$, which yields

$$\max(C(h)) = \frac{C_1}{1-\alpha} - \Delta\gamma$$

which can be $\leq d(A)$ (Schertzer and Lovejoy, 1983, 1985 discuss another model , the "α" model in which this also occurs). Note that in the latter case, divergence will still occur if the set A is sufficiently sparse so that d(A) is small enough.

Rewriting the above, we obtain the following equation for h_c:

$$K(h_c) = d(\Lambda)(h_c - 1) \Rightarrow \tag{5.14}$$

$$\frac{C_1\alpha'}{\alpha}h_c^{\alpha} + h_c\left(-\Delta\gamma - d(\Lambda) - \frac{C_1\alpha'}{\alpha}\right) + d(\Lambda) = 0$$

Corresponding to h_c, there is a critical singularity γ_c such that $h_c = c'(\gamma_c)$. The functional forms for the three parameter universality classes are therefore valid for the observable (dressed) quantities only for $h \le h_c$. $\gamma \le \gamma_c$ with γ_c written explicitly as:

$$\gamma_c = C_1\alpha'\left(h_c^{\alpha-1} - \frac{1}{\alpha}\right) - \Delta\gamma \tag{5.15}$$

For $\gamma > \gamma_c$, $c(\gamma)$ is a straight line with slope h_c. For $h > h_c$, the moments $\langle Z^h \rangle \to \infty$ hence, strictly speaking, K(h) is no longer defined. However, experimentally, since finite sample sizes are used to estimate the moments, we obtain the phenomenon of "pseudo-scaling"- see Schertzer and Lovejoy, 1987 and Lavallée et al 1989. Like α, estimating γ_c from the data is difficult because it too is very sensitive to the low-probability, large γ, $c(\gamma)$ part of the function. It is therefore of interest to develop approximate graphical methods for it's estimation. See Fig. 5.1 for details of this construction.

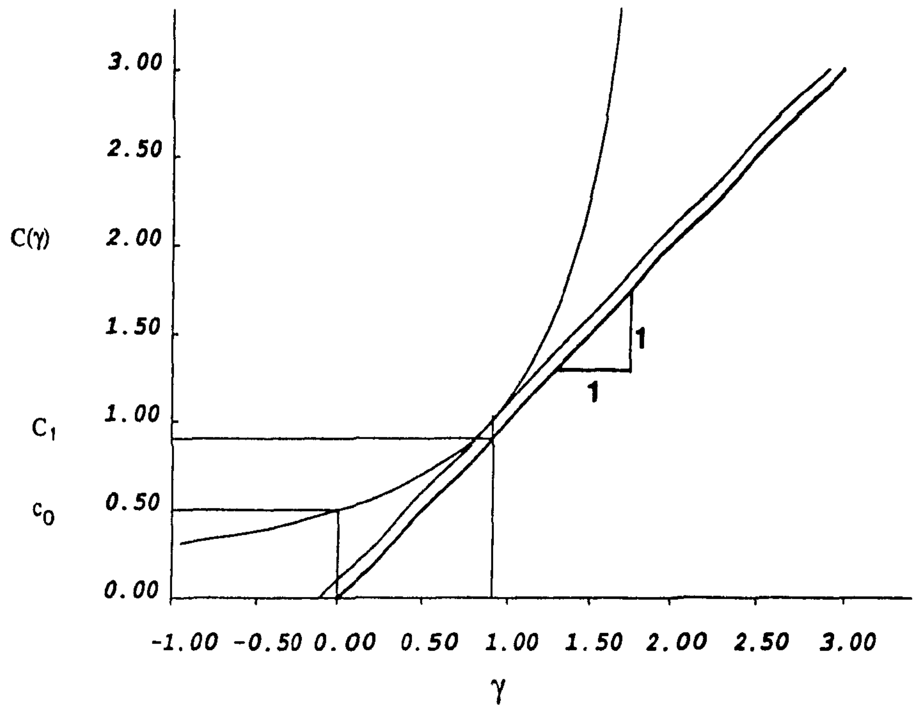

Figure 5.1 Graphical construction to estimate c_0 and C_1.

5.5 The Impact of Radar Induced Measurement Error

Since this study relies heavily on radar derived rain and reflectivity fields, and it is well known that the radar introduces a significant measurement error in the estimation of the mean Z over some small volume in space, the likely impact of such measurement error on any conclusions reached in this study must be assessed. As was pointed out by Zawadzki (1987) these statistical fluctuations added to the Z, caused by the shuffling of the raindrops, may serve to enhance the extreme values in the data sets, causing spurious hyperbolic distributions. The standard theory for radar measurement of rain seeks to relate the observed "effective radar reflectivity factor" Z_{*} with the "radar reflectivity factor Z", and from Z, via various assumptions, to the rain rate R. This step involves making assumptions about the probability distribution of drop volumes as well as their correlation structure. The usual approach assumes the drops to be independently distributed with finite variance distributions; this leads to incoherent scattering and the following conditional probability distribution for Z_{*} given Z:

$$p(Z_{*}\,|\,Z) \propto e^{-\frac{Z_{*}}{Z}} \qquad (5.16)$$

If we now assume that Z is the result of a cascade process, with an associated $c_Z(\gamma)$, we seek to know the relation between $c_Z(\gamma)$ and the measured $c_{*}(\gamma_{*})$ for the "effective radar reflectivity factor". This is a case of the "Observer's Problem" applied to the probability distributions. In terms of multifractals, the problem may be posed as follows:

$$Z_{ob} \propto L^{-\gamma_{ob}} \sim e^{\gamma_{ob}\zeta}$$

$$Z \propto L^{-\gamma} \sim e^{\gamma\zeta} \qquad (5.17)$$

where

$\zeta = -\ln L$

Since we are typically interested in small scales, we take $\zeta \gg 1$. We now obtain:

$$\begin{aligned} p(Z_{ob}) &\propto \int_{-\infty}^{\infty} p(Z_{ob} \mid Z) e^{-c(\gamma)\zeta} d\gamma \\ &\propto \int_{-\infty}^{\infty} e^{\left(-e^{\zeta(\gamma_{ob}-\gamma)} - c(\gamma)\zeta\right)} d\gamma \\ &\propto e^{-\max_{\gamma}\left(e^{\zeta(\gamma_{ob}-\gamma)} + c(\gamma)\zeta\right)} \end{aligned} \qquad (5.18)$$

where in the last step we have again used the method of steepest descents.

In the case of interest, where ζ is large, $c(\gamma)$ is algebraic, and we obtain to within high order corrections in ζ

$$\max_{\gamma}(e^{\zeta(\gamma_{ob}-\gamma)} + c(\gamma)\zeta) = c(\gamma_{ob})\zeta \qquad (5.19)$$

hence we obtain

$$c_{ob}(\gamma_{ob}) = c(\gamma)$$

i.e. the co-dimension function and probability distributions for effective reflectivity are the same as those of Z as long as $c(\gamma)$ is algebraic and L is small. The exponential relation between the Z and Z_* therefore will not affect the form of the probability distribution if the distribution is "long" or "fat" tailed. These results were tested by Monte Carlo simulation where it was found that the shuffling fluctuations, although they change the shape of the distribution for the extreme values, the effects are essentially exponential in form and are not capable of turning an exponential distribution into a hyperbolic distribution. Thus the concerns of Zawadski (1987) mentioned earlier do not appear to be well founded.

5.6 The PDMS Method of Estimating $c(\gamma)$

According to our universality formula for $c(\gamma)$, the latter are generally unbounded. However, in single realizations of multiplicative processes - no matter how much resolution is available - singularities with co-dimensions greater than d are not observable since they would have negative dimensions. The statistical properties of these rare extreme events can only be studied with large samples and with the help of the concept of sampling co-dimension (see Schertzer and Lovejoy, 1989).

Physically, it implies that with no matter how many data, individual realizations give only limited statistical information about the process. This is quite different from many standard stochastic processes, and is also different from "microcanonical" cascades (e.g. Screenivasan and Meneveau, 1987) for which, in principle, all the statistical properties can be obtained from a single realization.

That extremely large sample sizes are necessary has been stressed and studied in detail by Lavallée et al (1989). This fact alone makes experimental determination of $c(\gamma)$ quite difficult. An additional problem, mentioned earlier, is that the basic probability distributions will generally have log corrections which are difficult to estimate: $c(\gamma)$ is only the leading exponential part of the scaling. Early techniques for estimating the scaling exponents (see Schertzer and Lovejoy, 1985, and Halsey, 1986) worked directly with the moments - yielding $K(h)$, and in the latter case $c(\gamma)$ by Legendre transformation. More recent techniques estimate $c(\gamma)$ directly, either by "functional box-counting" (Lovejoy et al 1987, Gabriel et al 1988), or via the Probability Distribution / Multiple Scaling (PDMS) technique (Lavallée et al 1989). This technique directly exploits the fundamental equation (5.1) by systematically degrading the resolution of the data through averaging over increasing scales L. The field is normalized by dividing by an ensamble averaged (climatological) value Z_{L_0}, and the length scales are normalized by L0, the external scale of the data. When the data are correlated rather than independent samples, for example arising from time-series that have been degraded in space, but accumulated in time, L0 must be placed by an "effective" L0 which takes these correlations into account. Empirically, it is simplest to obtain both L0 and Z_{L_0} from regressions. The $c(\gamma)$ function can be estimated by means of

$$c(\gamma) = \frac{\log(Pr)}{\log(L/L_0)} \tag{5.16}$$

and

$$\gamma = -\frac{\log(Z_L/Z_{L_0})}{\log(L/L_0)}$$

5.7 Estimation of L_0 and Z_{L_0}

Since it was by no means clear that the Z and rainfields could in fact be conceptualized as multiplicative cascades, the first step in the analysis was to determine whether or not a Z_{L_0} and L_0 could be found for each data sequence. If it was found to be possible to select a Z_{L_0} and L_0 such that the $c(\gamma)$ function was invariant over changes in the measurement scale, the second step in the analysis would be to empirically determine the other parameters in Eqns. (5.9) and (5.10).

A measure of how closely two lines lie on top of each other is the mean minimum distance between any point on the first curve and some point on the second curve. Since there were six curves in this analysis representing the scales from 2 km up to 64 km, it was decided to measure the mean minimum distance between successive curves, the objective function was then to minimize the sum of the mean minimum distances between successive curves. Since both $c(\gamma)$ and γ have $\log\left(\frac{L}{L_0}\right)$ as a denominator, the mean distance between curves will always decrease with increasing L_0. Therefore the mean minimum distance was normalized by expressing it as a percentage of the mean length of the two lines. The downhill simplex method (see Press, Flannery, Teukolsky and Vetterling (1988) for details) was used to estimate the optimum Z_{L_0} and L_0. The results for the four periods analyzed, found in Table 5.1, and Figure 5.2 for the 14th August case, were much better than expected. The 2, and 64 km resolution distributions were not used in any of the parameter estimation procedures, but were retained as reference to determine whether the model could predict the distributions for scales larger and smaller than those used during the model calibration.

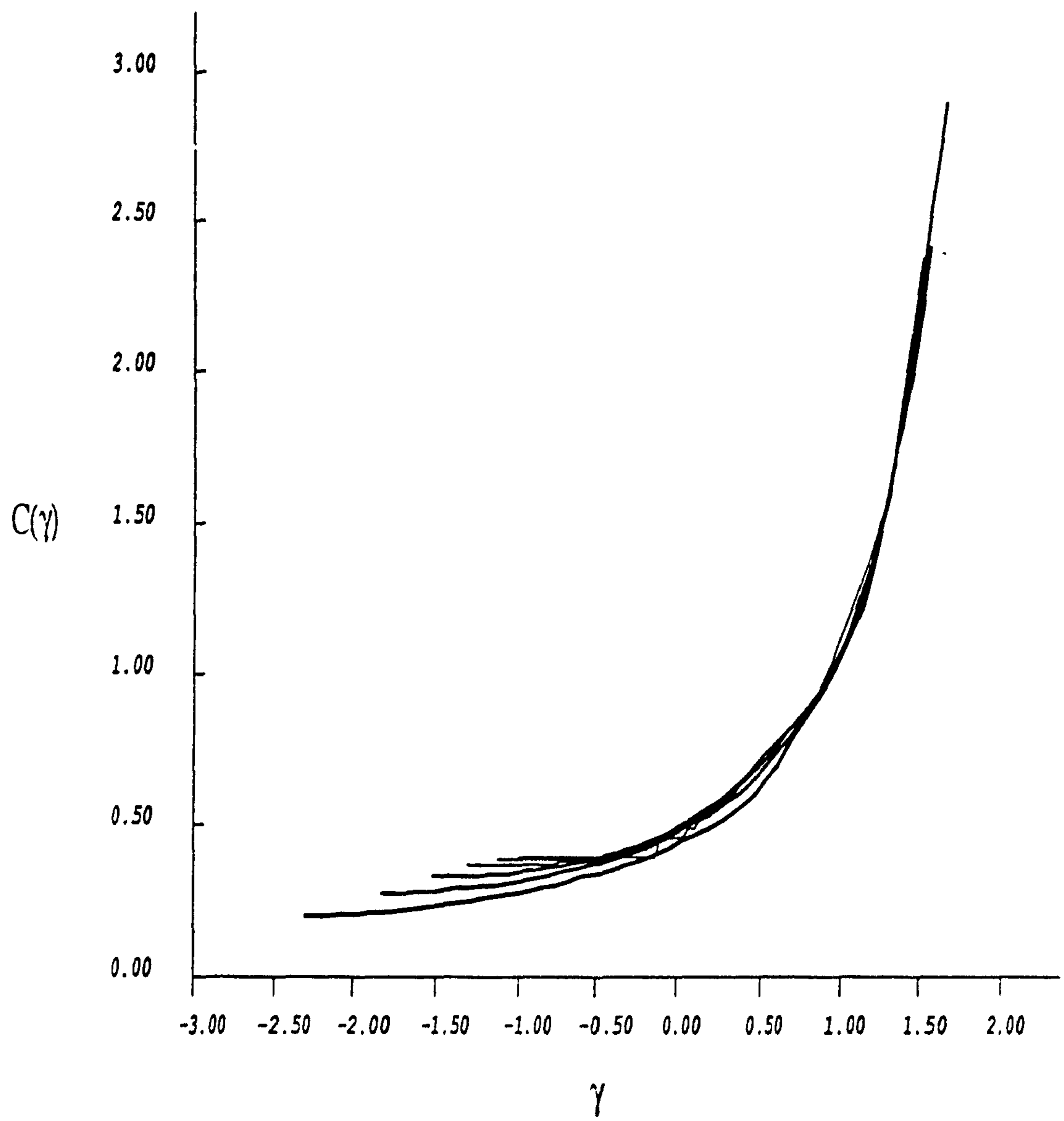

Figure 5.2 $c(\gamma)$ *vs* γ curves using L = 2, 4, 8, 16, 32, and 64 km for 14th August data

Table 5.1 Z_{L_0} and L_0 estimated using the mean minimum distance between $c(\gamma)$ for successive scales of measurement, expressed as a percentage of the mean length of the two $c(\gamma)$ curves

Date	$Z_{L_0}(mm^6m^3)$	L_0 (km)	Mean Min Distance (%)
29/8/87	225	312	4.9
22/8/87	423	273	4.8
14/8/87	510	929	4.6
12/8/87	644	1174	4.3

5.8 Parameter estimation

Since it was possible to find Z_{L_0} and L_0 such that the various $c(\gamma)$ curves were at least approximately scale invarient over the range of scales used in the analysis, it is reasonable to estimate the parameters for the proposed model. If the relations between the various parameters in Eqn. (5.9) are not assumed, and allowing for the extra parameter γ_c, we have at most six parameters that need to be estimated. However, if the relations in Eqns (5.9), (5.10), and (5.13) hold, and $\Delta\gamma$ is constant, which is plausible since it is a basic dimensionally determined quantity, the number of parameters is reduced to four. A further assumption could be that α is constant, which would be theoretically appealing since α is the fundamental parameter characterizing the generator of the process, the number of parameters is reduced to three. Each of these three models will be fitted to the data in the following sub-sections, starting with the 6 parameter model. Thereafter, the 4 and 3 parameter models will be fitted, and the theoretical values for the remaining parameters compared with the empirical values found for the six parameter model.

5.8.1 Six Parameter Model

At least two strategies are possible when attempting to determine the c_0, γ_0, γ_c, α' parameters, the first would be to use a least squares scheme to fit the parameters to the mean $c(\gamma)$ curve having first determined the Z_{L_0} and L0 parameters. The second is to use the probability distributions to estimate all six parameters directly. The objective function for the second method was chosen to be

$$MAD=\sum_{j=1}^{m}\frac{1}{n_j}\sum_{i=1}^{n_j}\left|\log(Pr)-\log(\hat{Pr})\right| \tag{5.17}$$

where

m = the number of probability distributions

n_j = the number of points on the j th probability distribution above the minimum detectable signal at that resolution,

$Pr = Pr(Z_L' > Z_L)$, and

$\hat{Pr}$ is the Pr predicted by the model for the same value of Z_L.

The minimum detectable signal at each scale was simply assumed to be the first class in the dBZ frequency histogram greater than zero that was non-zero. The $c(\gamma)$ for $\gamma>\gamma_c$ is a straight line with a slope m such that

$$m = \frac{c_0}{\gamma_0 \alpha'}\left(1+\frac{\gamma_c}{\gamma_0}\right)^{\alpha'-1} \tag{5.18}$$

It was decided to evaluate all six parameters simultaneously using the MAD statistic with the probability distributions and as a check, the c_0, γ_0, γ_c, α' parameters using the Z_{L_0} and L_0 evaluated earlier.

Two data sets were used for this analysis, radar reflectivity data for instantaneous radar images, averaged in space, and rainfall data averaged over 5, 10, 20, 40 and 80 minutes, but at a fixed spatial resolution of 2km.

5.8.1.1 Spatial Averaging

The results of the analysis of the four sets of Z data, using both techniques are summarized in Table 5.2 and the predicted and measured probability distributions are plotted in Figures 5.3 to 5.6. The figures also show the 95% confidence limits for the experimental probability distributions using the

Kolmogorov-Smirnov one sample two-sided statistic for samples sizes larger than 40 (Daniel 1978, Table A.17). The distribution at the 2 km scale of measurement has very narrow confidence limits since the sample size is very large. Never the less, the predicted probability distributions for scales larger than 2 km are all within the 95% confidence limits of the experimental distribution for intensities greater than the minimum detectable signal for that scale. The method was able to extrapolate down to the 2 km scale as well as up to the 64 km scale, see August 14th distributions in Figure 5.4 for example. The parameters estimated by the two methods proved to be within about 10% of each other for two of the four days. However, on the 29th and 22nd August the method using the probability distributions directly gave different results for L0. These two days also gave the smallest L0, and hence had the greatest sensitivity to spatial averaging. The analysis was repeated for the 29th August using only the 16 km to 64 km scales in the parameter estimation and the probability distribution method. The results for this analysis are found in Figure 5.7 and Table 5.3. The 2 km probability distribution still fits remarkably well considering the small sample size for the 64 km distribution and the extent of the extrapolation.

Once again, the parameters are within 10% of the value estimated using the 4 km to 32 km data, with the exception of the L0 parameter which is approximately 25% less for the 16 km to 64 km data.

Table 5.2 Best fit parameters for four 12 hour sequences using radar reflectivity (Z)

	29th August		22nd August		14th August		12th August	
	MAD	LS	MAD	LS	MAD	LS	MAD	LS
γ_c	-1.69	-1.73	-1.70	-2.15	-2.01	-2.02	-1.42	-1.38
γ_0	-2.02	-2.02	-1.98	-1.84	-2.18	-2.08	-1.87	-1.75
c_0	0.88	0.88	0.80	0.91	0.50	0.52	0.61	0.63
α'	-0.42	-0.36	-0.48	-0.52	-1.26	-1.17	-0.68	-0.60
Z_{L_0}	210	225	372	423	531	510	666	644
L_0	430	312	375	273	915	930	1180	1175

Table 5.3 Best fit parameters for the 29th August using the 16 km to 64 km scales only

Parameter	4 km - 32 km	16 km - 64 km
γ_c	-1.69	-1.91
γ_0	-2.02	-2.24
c_0	0.88	0.96
α'	-0.423	-0.457
L_0	210	234
Z_{L_0}	430	335

5.8.1.2 Temporal Averaging

Since the analysis of the Z data was successful, it was decided to attempt a similar analysis on short duration mean rain rates. To this end probability distributions for the mean rainfall intensities over 5, 10, 20, 40, 80 and 160 minutes were estimated using the 22nd August case with 2 km resolution. The 160-minute distribution was not included in the data set used to estimate the model parameters. The summary of the parameter values is found in Table 5.4 and the predicted and measured probability distributions are plotted in Figure 5.8. From this Figure it is apparent that the model is able to fit data for the 5 to 80 minute time resolution, as well as extrapolate outside the calibration range to give a prediction for the 160-minute distribution that is within the 95% confidence limits for the empirical distribution.

Table 5.4 Summary of two possible sets of parameter values for temporal averaging using 2 km resolution data

γ_c	γ_0	c_0	α'	R_{T_0}(mm/hr)	T_0(min)
-0.75	-0.85	0.37	-0.41	0.129	79373
-0.95	-1.05	0.36	-0.46	0.044	79640

The most striking difference between this set of parameters and those for the spatial smoothing case is the very large value for T_0 as compared with Z_0. This arises from the fact that the probability distributions resulting from temporal averaging are much closer to each other than is the case for the distributions resulting from spatial smoothing. It is clear that much of the temporal variability has already been smoothed out after spatial averaging over a few kilometers. The closeness of the probability distributions made the estimation of the various parameters less precise. A second, nearly optimal local minimum is also listed as the second set of parameters in Table 5.4. This solution somewhat over-estimated the increase in the rainfield variability with decreasing time resolution although it fitted the lower temporal resolutions rather well.

5.8.2 Four and Three Parameter Models

Using the Z_{L_0} and L0 values found previously for the four series of radar reflectivity data, the graphical analysis discussed earlier was used to estimate $\Delta\gamma = 0.1$, and $C_1 = 1.2$, 1.2, 1.3, and 1.4 for the data from August 12, 14, 22, and 29 respectively. This simple graphical technique indicates that the parameter $\Delta\gamma$ is nearly independent of the meteorological situation, as we had hoped. If we allow and c_0 to vary, estimating them graphically, the other parameters can be calculated using Eqn. (5.10). Table 5.5 shows a comparison of these parameters with those obtained from the full multi-parameter regression used for the six parameter model. There is no divergence of moments for the 12th August since

$$\frac{C_1}{1-\alpha}-\Delta\gamma<2$$

Figures 5.9 and 5.10 show the empirical probability distributions and the distributions predicted by the four parameter model for the 14th August and 22nd August respectively.

Table 5.5 A comparison between the parameters estimated by the four parameter model and those estimated by the six parameter model

Date	12	14	22	29
C_1	1.2	1.2	1.4	1.4
c_0	0.61	0.50	0.80	0.88
$\Delta\alpha$	0.02	0.01	0.03	0.02
$\Delta\gamma_0$	0.07	0.02	0.02	0.04
$\Delta\gamma_s$	-	0.07	0.30	0.35

The two parameter model was tested by assuming that the α parameter was equal to 0.4 over the four days in the data set. Once again Eqn (5.10) was used to estimate the other parameters. Table 5.6 lists a comparison between these parameters and those obtained for the six parameter model. Figures 5.11 and 5.12 show a comparison between the empirical distributions and those predicted by the model using the 14th and 22nd August cases. It is readily apparent that the two parameter model does not fit the probability distributions over the range of scales for the 14th August case, although the fit is somewhat better for the 22nd August case. Therefore a single α is not able to reproduce the richness and variety found in real rainrate or Z field probability distributions.

Table 5.6 A comparison between the parameters estimated by the three parameter model compared and those estimated by the six parameter model

Date	12	14	22	29
C_1	1.2	1.2	1.4	1.4
Δc_0	0.06	0.17	0.07	0.10
$\Delta \gamma_0$	0.03	0.18	0.09	0.21
$\Delta \gamma_c$	-	-	0.34	0.42

5.9 Conclusions

A method to produce resolution independent short duration rainfall accumulations and instantaneous radar reflectivity probability distributions has been developed and tested. The conceptual model using multiplicative cascades to describe the concentration of flux into smaller, more intense regions was found to provide good predictions of the probability distributions extrapolated outside the range of data used to calibrate the model. This lends some credibility to the initial hypothesis that the rainfields may usefully be described by multiplicative cascade type of models. Three flavours of the model were tested, having six, four and three parameters. The six parameter model was calibrated from the probability distributions directly, using the downhill simplex method in the parameter estimation. The four and three parameter models were calibrated in two steps, the first step was to estimate the values for Z_{L_0}, L_0 that minimized the normalized mean distance between successive co-dimension curves. Thereafter, a graphical construction was used on the resulting curve to estimate the other parameters. It was found that the four parameter model was nearly as good as the full six parameter model and was easier to calibrate, and therefore should be used in preference to the full model in practical applications. The three parameter model was unable to produce a good fit to the probability distributions. We believe that we have demonstrated a general method for obtaining resolution independent probability distributions, thus allowing the construction of probability distributions at other time and space scales. Furthermore, these results could be the first step towards the use of multiplicative cascade type of models in stochastic rainfield simulations.

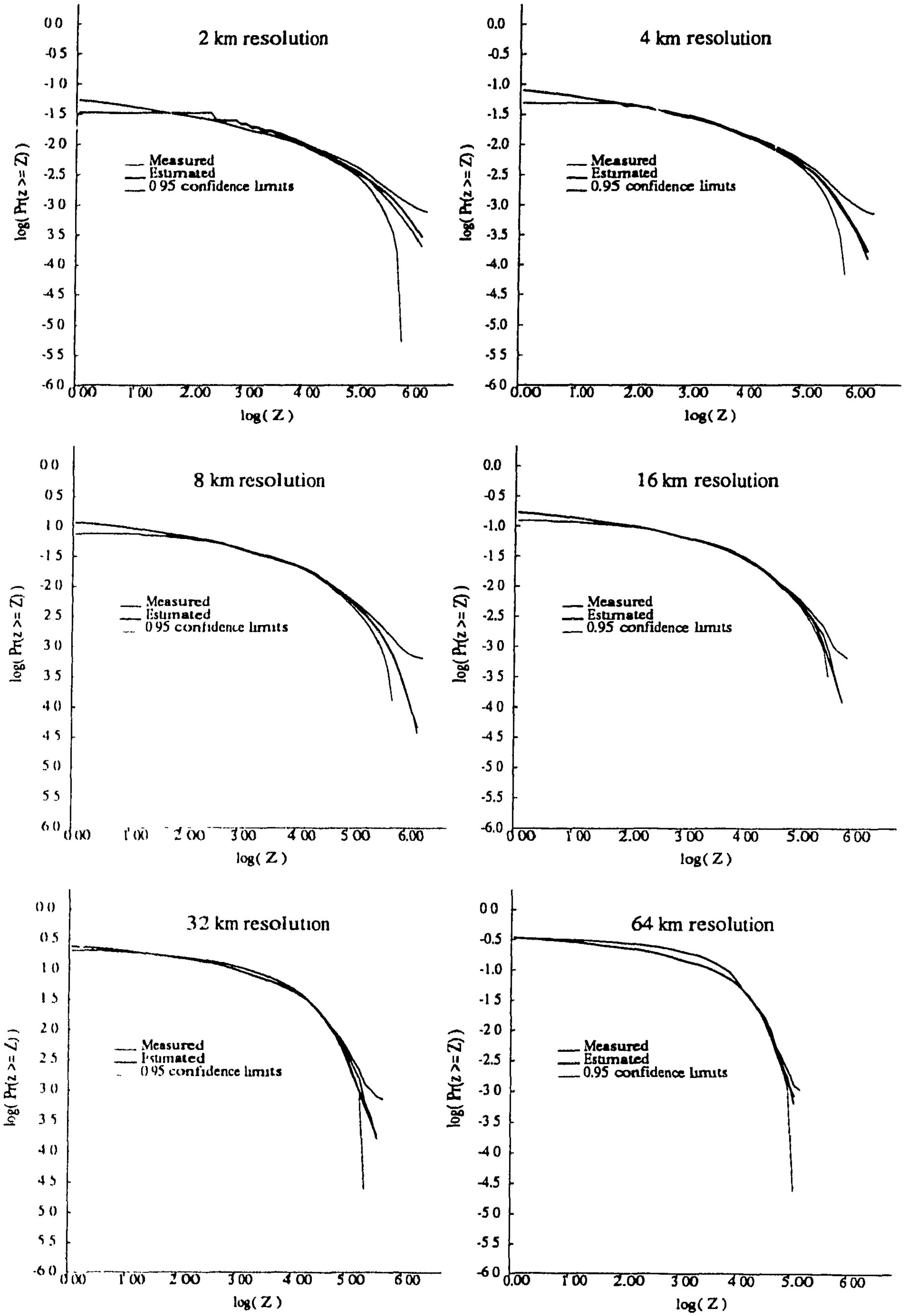

Figure 5.3 Estimated and predicted probability distributions for instantaneous 12th August Z fields averaged over 2, 4, 8, 16, 32, and 64 km, 6 parameter model calibrated using 4 - 32 km data

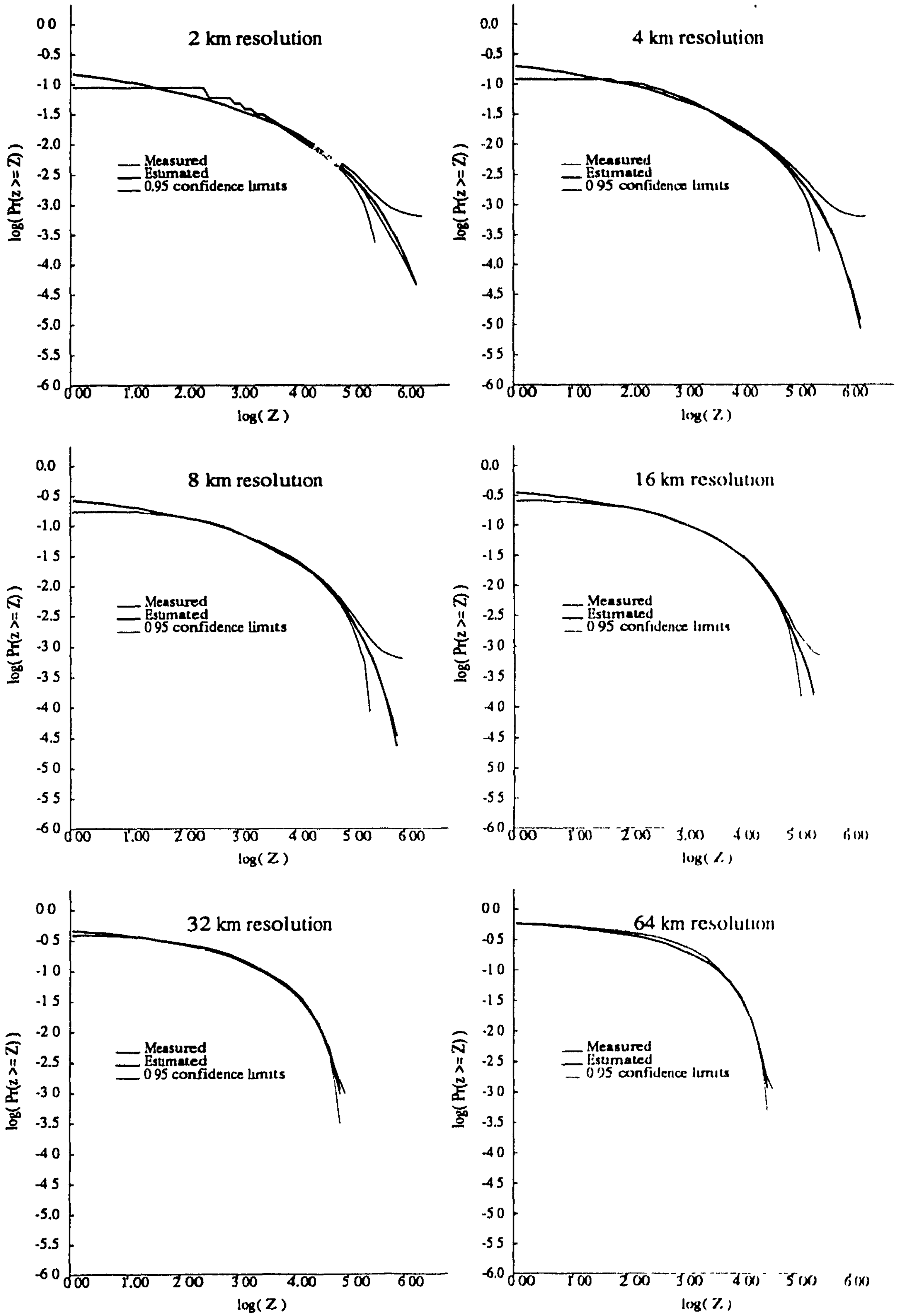

Figure 5.4 Estimated and predicted probability distributions for instantaneous 14th August Z fields averaged over 2, 4, 8, 16, 32, and 64 km, 6 parameter model calibrated using 4 - 32 km data

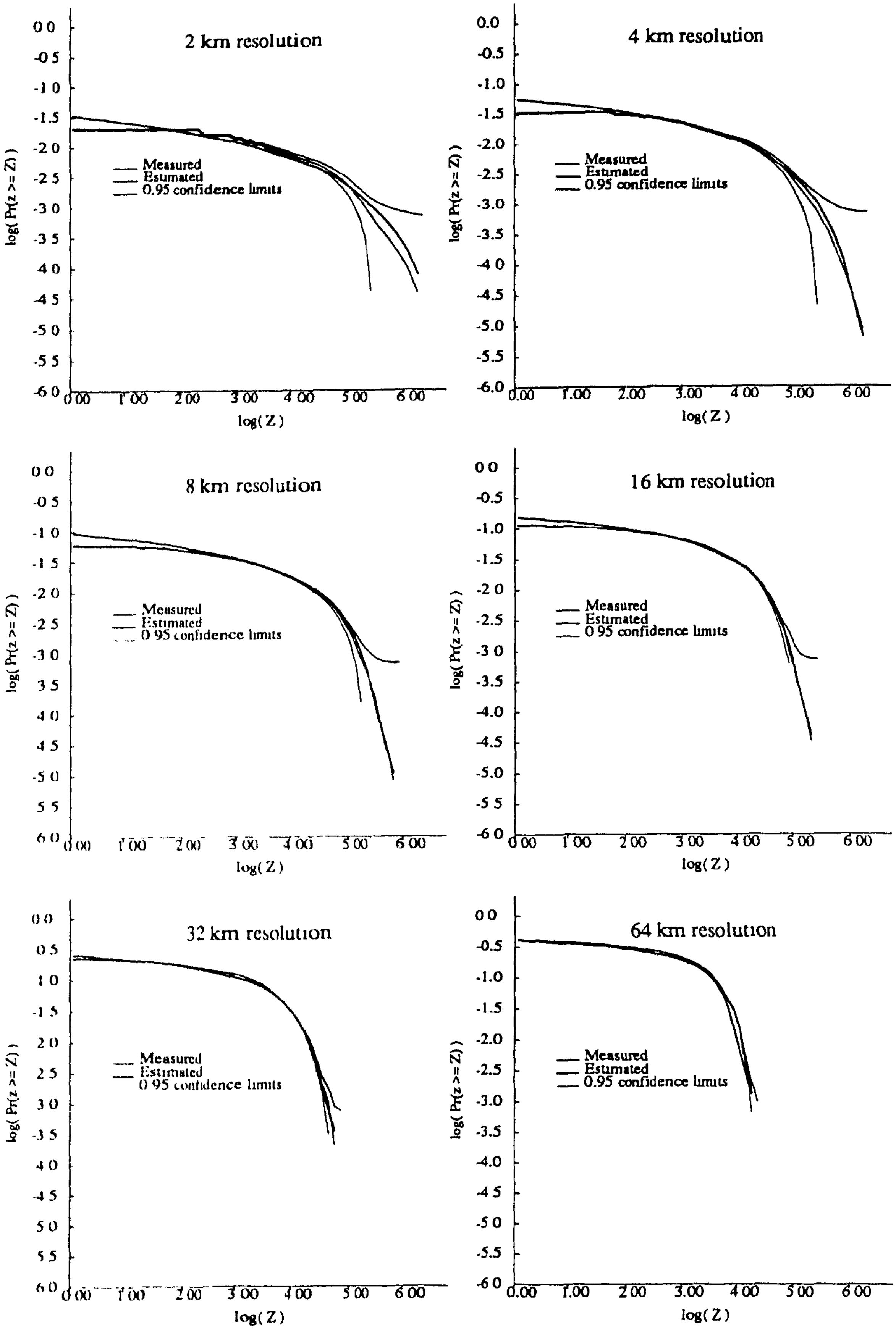

Figure 5.5 Estimated and predicted probability distributions for instantaneous 22nd August Z fields averaged over 2, 4, 8, 16, 32, and 64 km, 6 parameter model calibrated using 4 - 32 km data

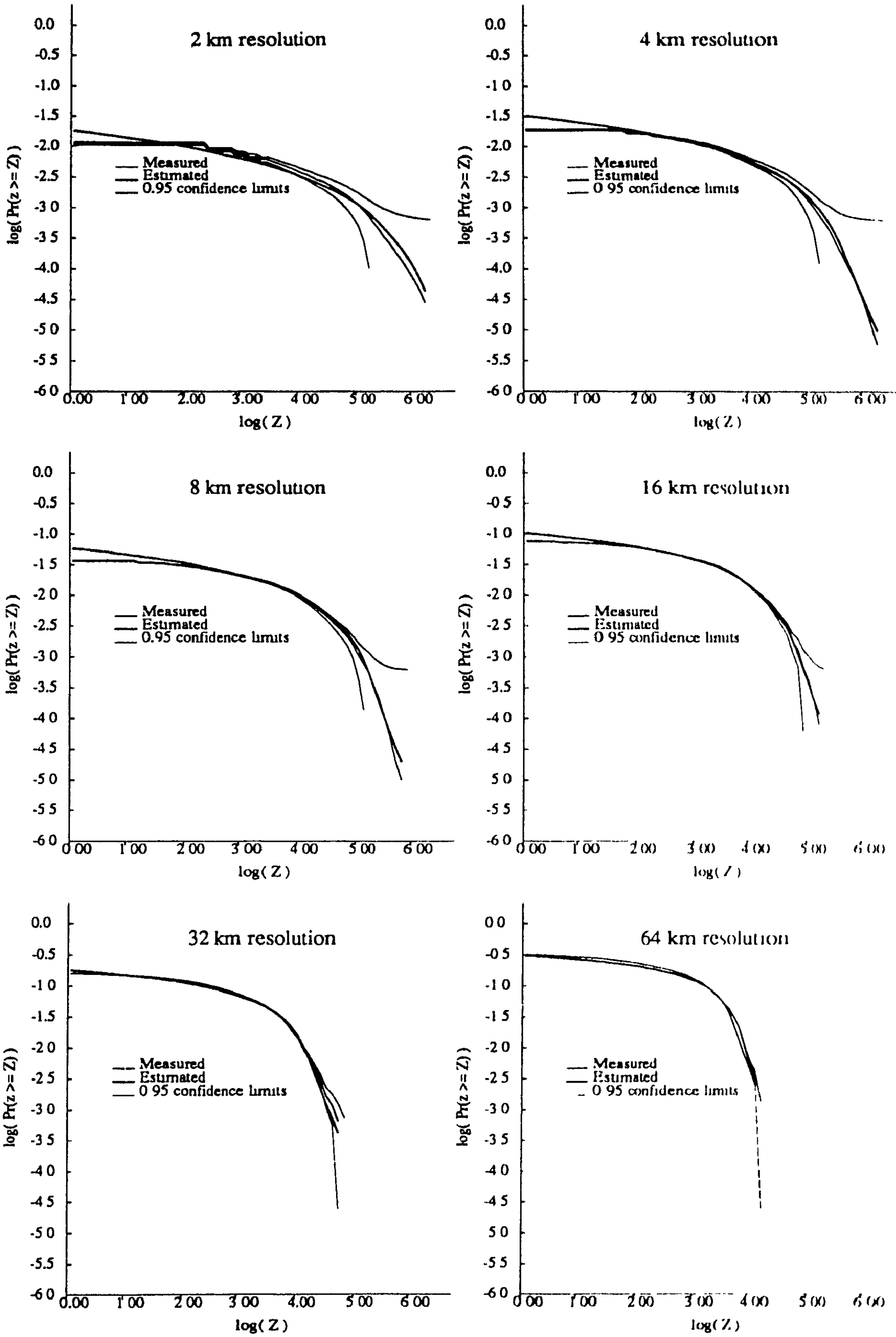

Figure 5.6 Estimated and predicted probability distributions for instantaneous 29th August Z fields averaged over 2, 4, 8, 16, 32, and 64 km, 6 parameter model calibrated using 4 - 32 km data

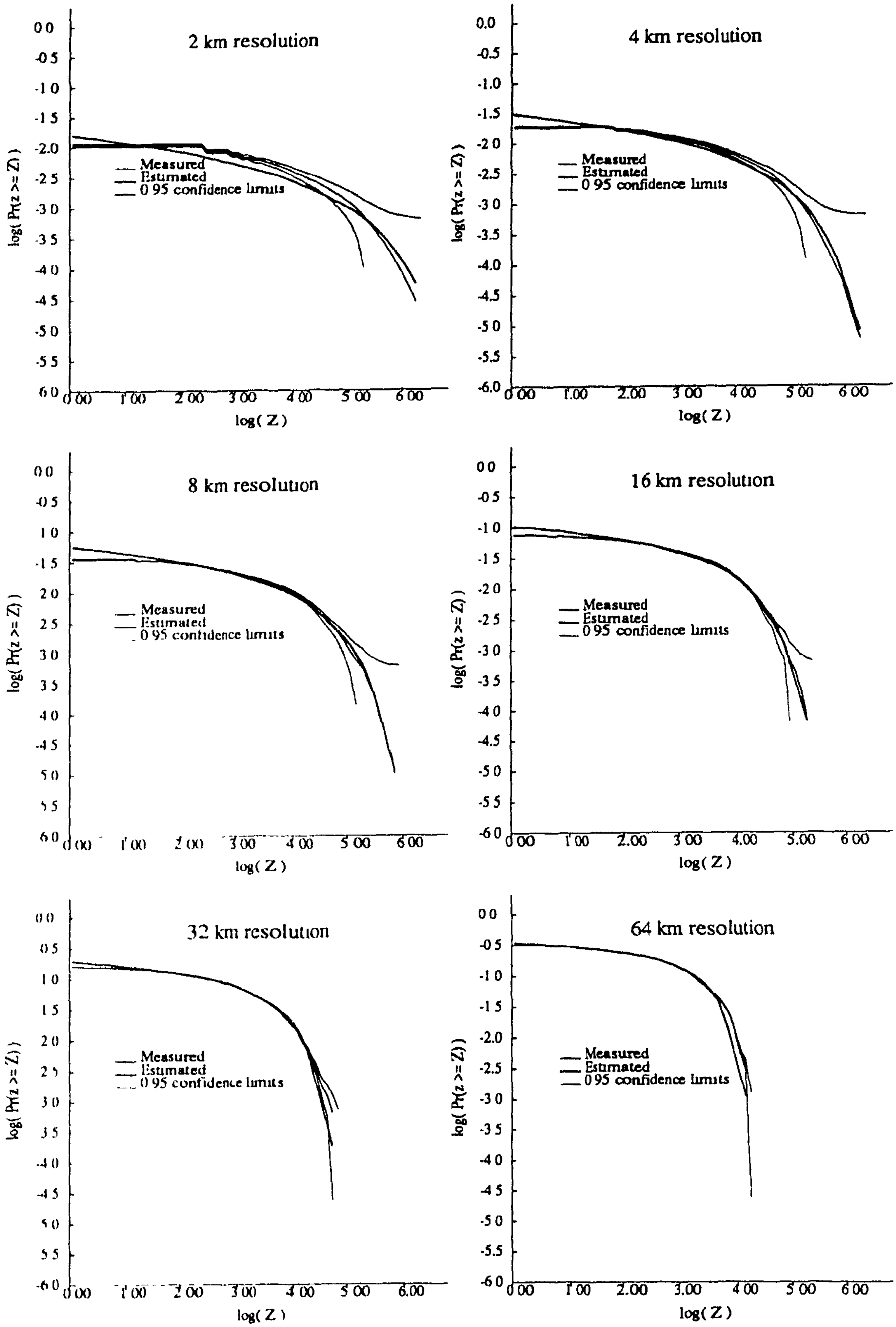

Figure 5.7 Estimated and predicted probability distributions for instantaneous 29th August Z fields averaged over 2, 4, 8, 16, 32, and 64 km, 6 parameter model calibrated using 16 - 64 km data

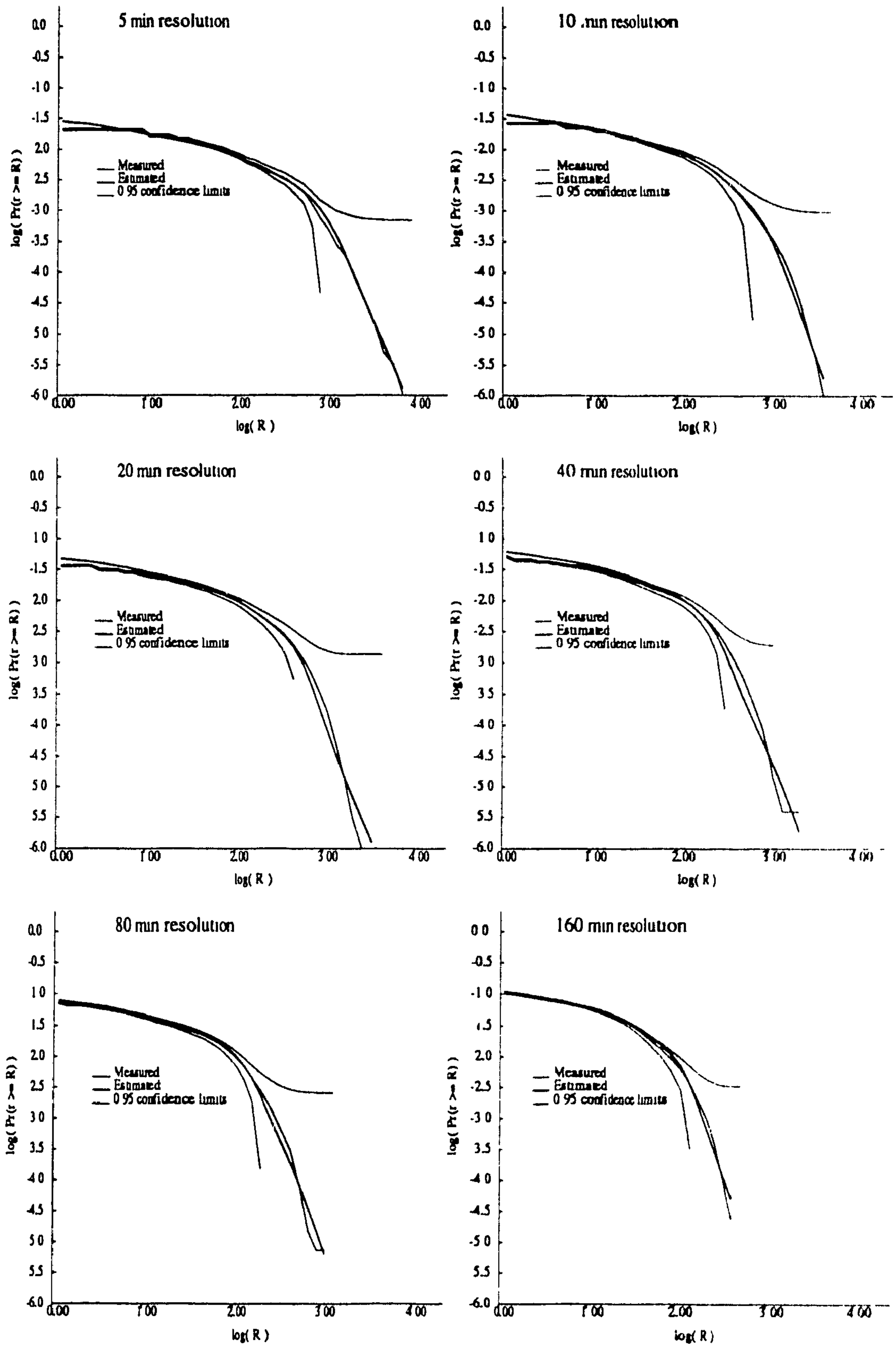

Figure 5.8 Estimated and predicted probability distributions for 22nd August 2km rainfields averaged over 5, 10, 20, 40, 80, 160 minutes, 6 parameter model calibrated using 5 - 80 minute data

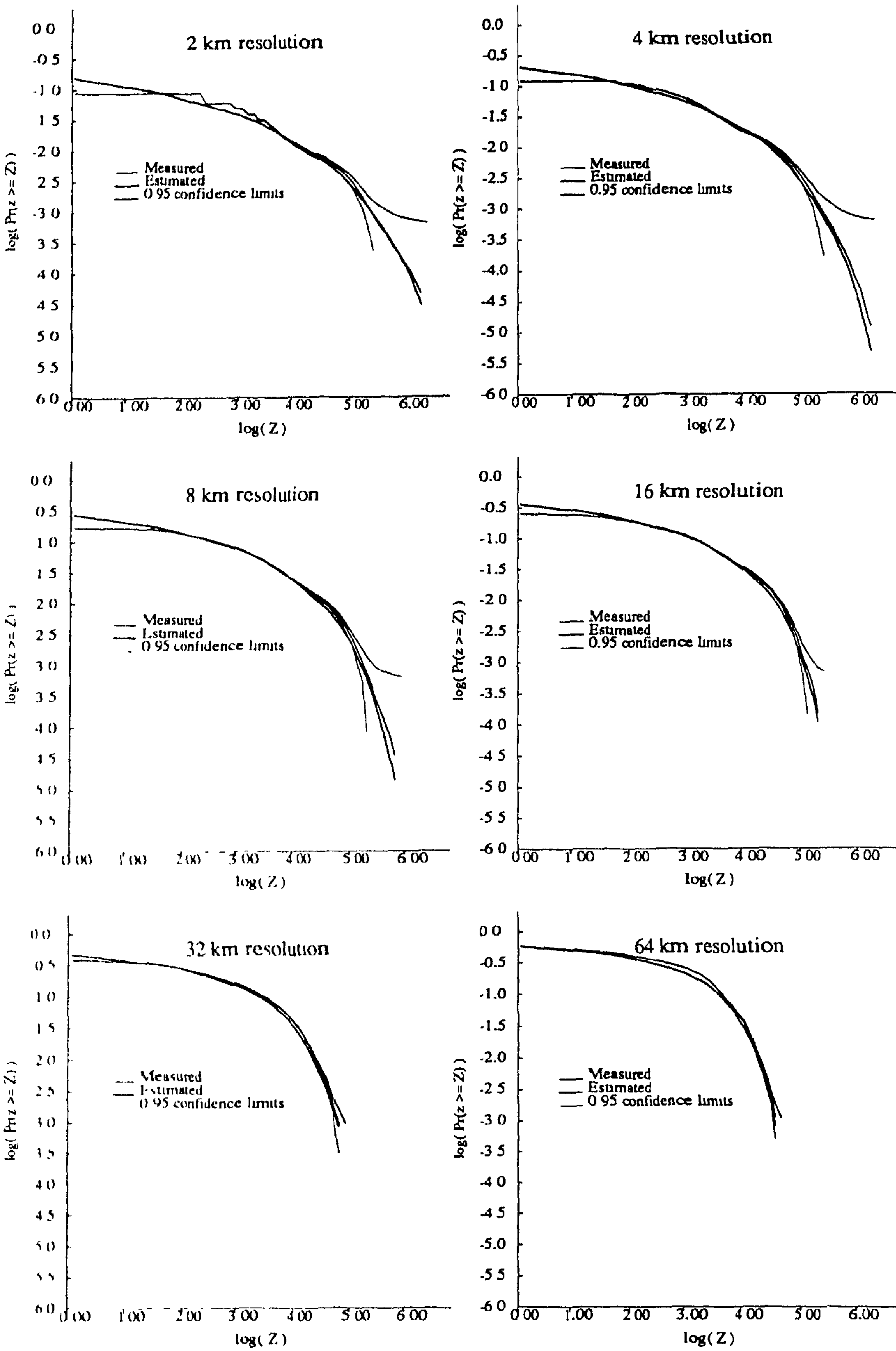

Figure 5.9 Estimated and predicted probability distributions for instantaneous 14th August Z fields averaged over 2, 4, 8, 16, 32, and 64 km, 4 parameter model

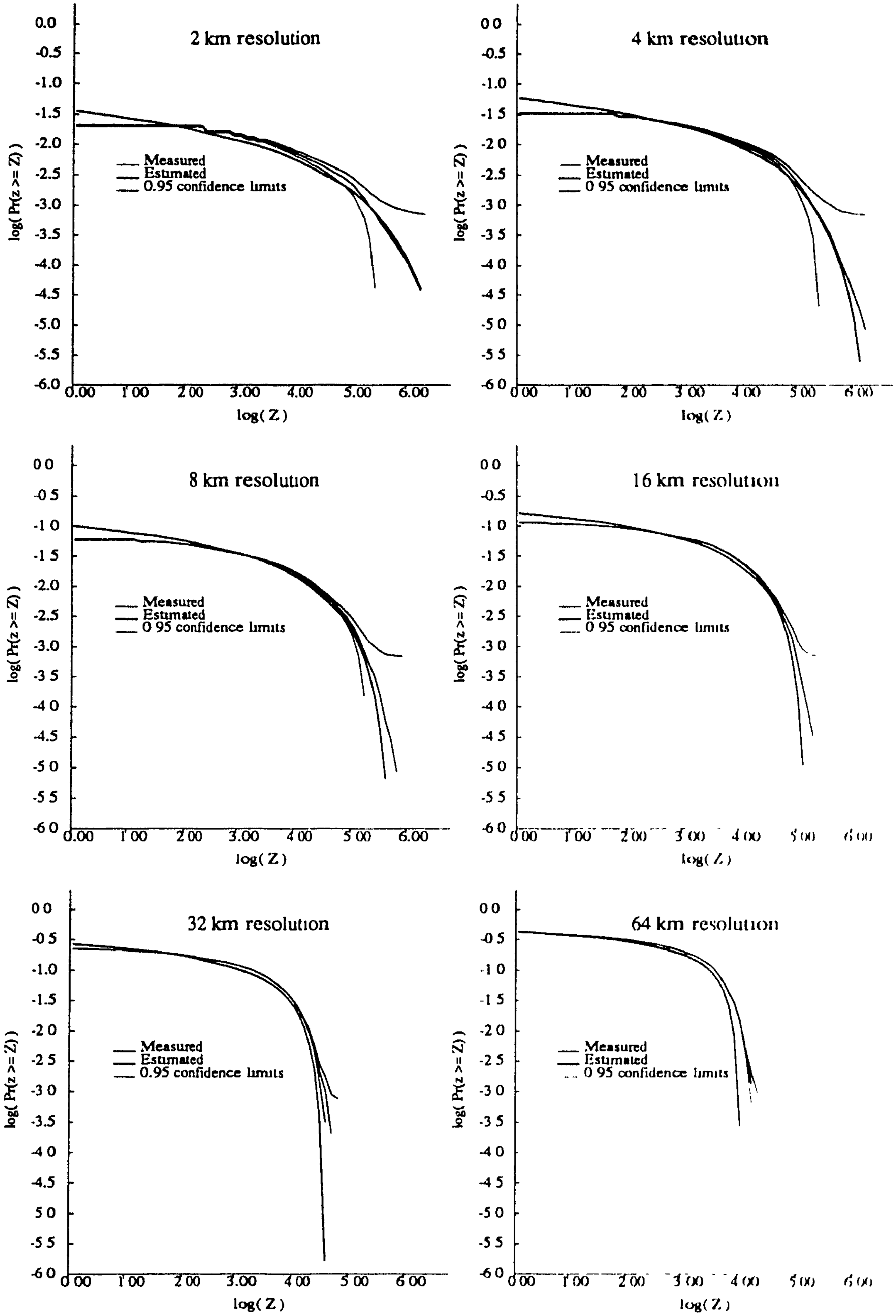

Figure 5.10 Estimated and predicted probability distributions for instantaneous 22nd August Z fields averaged over 2, 4, 8, 16, 32, and 64 km, 4 parameter model

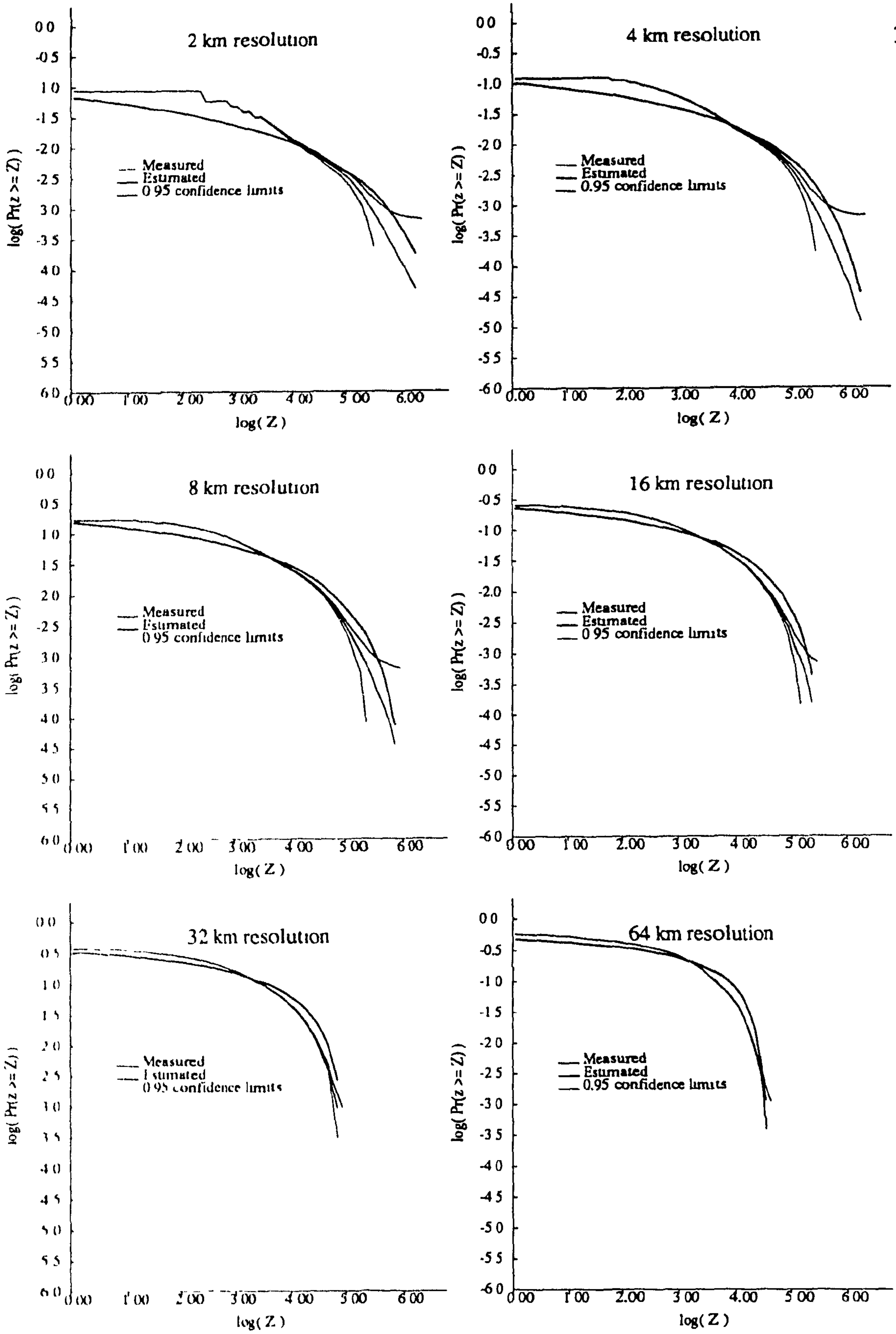

Figure 5.11 Estimated and predicted probability distributions for instantaneous 14th August Z fields averaged over 2, 4, 8, 16, 32, and 64 km, 3 parameter model

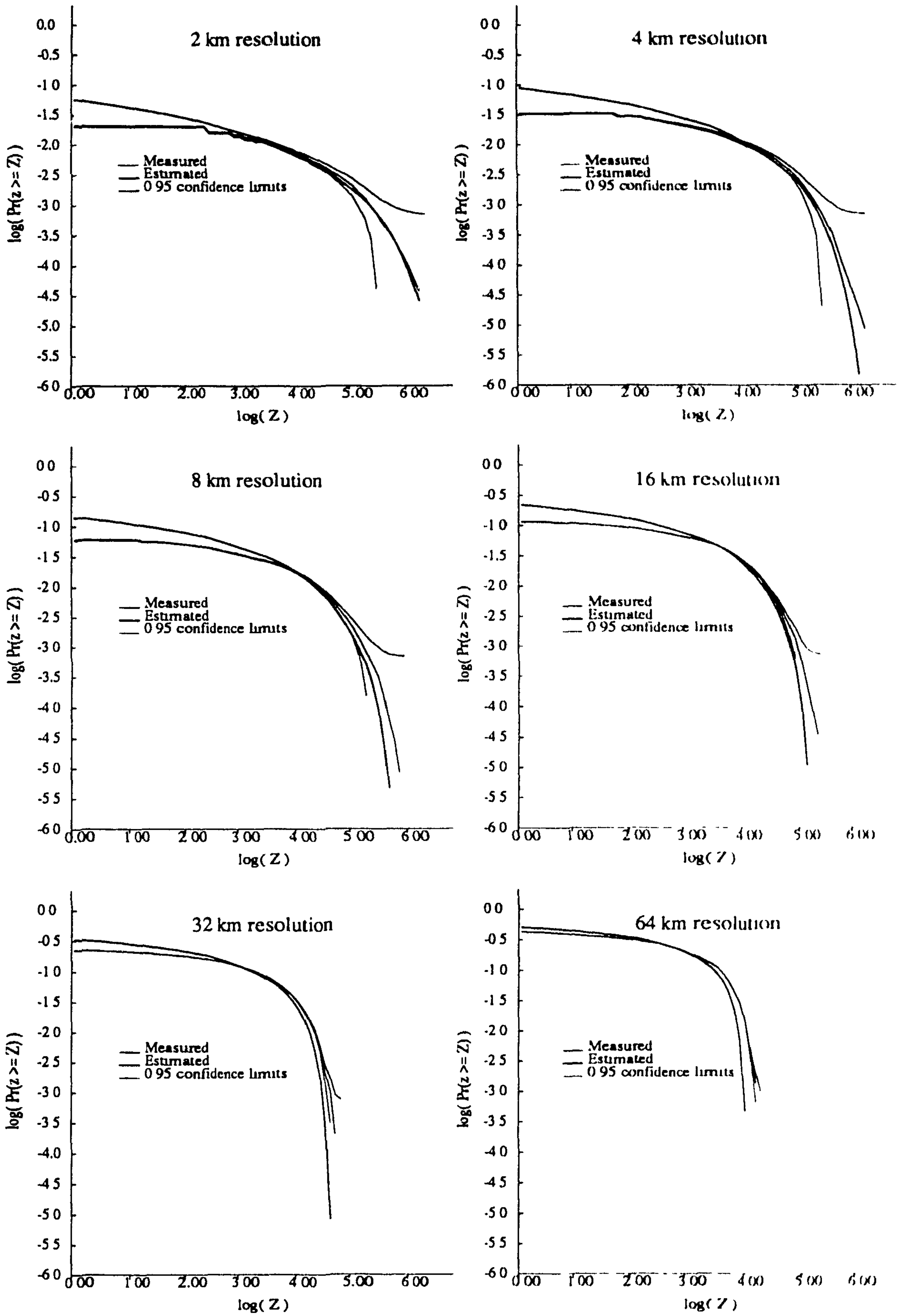

Figure 5.12 Estimated and predicted probability distributions for instantaneous 22nd August Z fields averaged over 2, 4, 8, 16, 32, and 64 km, 3 parameter model

5.10 References

Daniel, W.W., 1978. Applied nonparametric statistics. Houghton Mifflin Company, Boston. 503pp.

Frisch, U., and G. Parisi, 1985. A multifractal model of intermittency. In M.Ghil, R. Bensi, and G. Parisi (Editors). Turbulence and predictability in geophysical fluid dynamics and climate dynamics. Elsevier North-Holland, New York, 84-88.

Gabriel, P., S. Lovejoy, and D. Schertzer, 1988. Resolution dependence in satellite imagery: multifractal analysis. J. Geophysical Research Letters.

Hasley, T.C., M.H. Jensen, L.P. Kadanoff, I. Procaccia, and B. Shraiman, 1986. Fractal measures and their singularities: the characterization of strange sets. Phys. Rev., A33:1141-1151.

Lavallée, D., D. Schertzer, and S. Lovejoy, 1989. On the determination of the co-dimension function. In S. Lovejoy, and D. Schertzer (Editors). Scaling, fractals and non-linear variability in geophysics, in press.

Lopez, R.E., 1978. The lognormal distribution and cumulus cloud populations. Mon. Wea. Rev., 105: 865-872.

Lovejoy, S., D. Schertzer, and A.A. Tsonis, 1987. Functional box-counting and multiple elliptical dimensions in rain. Science, 235:1036-1038.

Press, W.H., B.P. Flannery, S.A Teukolsky, and W.T. Vetterling, 1988. Numerical recipes in C. Cambridge University Press, Cambridge. 735pp.

Rodriguez-Iturbe, I., 1986. Scale of fluctuation of rainfall models. Water Resources Research, 22(9): 15s-37s.

Schertzer, D., and S. Lovejoy, 1983. The dimension and intermittency of atmospheric dynamics. In Bradbury, L.J.S., F. Durst, B.E. Launder, and J.H. Whitelaw (Editors). Turbulent Shear Flows 4: Selected Papers from the Fourth International Symposium on Turbulent Shear Flows, University of Karisruhe, Karisrue, FRG, Sept. 12-14, 1983. Springer-Verlag, New York, 7-33.

Schertzer, D., and S. Lovejoy, 1987. Physical modelling and analysis of rain and clouds by anisotropic scaling and multiplicative processes. J. Geophysical Res., 92(D8): 9693-9714.

Schertzer, D., and S. Lovejoy, 1988. Multifractal simulations and analysis of clouds by multiplicative processes. Atmospheric Research, 21: 337-361.

Schertzer, D., and S. Lovejoy, 1989. Observables and universality classes for multiplicative fields. In S. Lovejoy, and D. Schertzer (Editors). Scaling, fractals and non-linear variability in geophysics, in press.

Wilson, J., S. Lovejoy, and D. Schertzer, 1989. Physically based cloud modelling by scaling multiplicative cascade processes. In S. Lovejoy, and D. Schertzer (Editors). Scaling, fractals and non-linear variability in geophysics, in press.

Zawadzki, I., 1987. Fractal structure and exponential decorrelation in rain. J. Geophysical Res., 92(D8): 9586-9590.

Chapter 6

Conclusions

An investigation into the extent to which the variability of observed rainfields limits the accuracy with which gauge networks, radar, and satellite rainfall measurements can estimate mean areal rainfall over large areas was undertaken. The radar introduced further measurement errors resulting from the fact that the measured radar reflectivity fluctuates randomly as the rain droplets move relative to each other in the volume of space being sampled. Monte Carlo simulations of this fluctuation, assuming no sub-resolution variability, together with an analysis of one month of radar data from Florida, showed that it was possible to minimize the combined effects of these errors and those caused by the variability of the rainfield by accumulating the rainfield in time and averaging in space. For example, a choice of 4 km pixel size for hourly rainfall accumulations would result in estimates of mean areal rainfall over a pixel with a variance equal to 26% of the variance of the hourly polar data. It was found that it was more efficient to deal with the radar measurement errors by accumulating in time and averaging over larger areas than to increase the number of samples used to estimate the mean reflectivity of a single radar bin.

The measurement errors for raingauge network estimates of mean areal rainfall over large areas were simulated using daily and monthly radar rainfall accumulations of the Florida data and a summer of radar rainfall from Nelspruit, South Africa. It was found that the difference between the mean standard error from regular and random gauge networks was slight, particularly for the more dense networks. However, the random networks gave errors that were more widely distributed about the mean standard error, becoming more apparent with decreasing network density. A relationship was found to predict the mean standard error given the number of gauges, the raining fraction of the area, and the variability of the rainfield being sampled. This relationship was able to explain 63% of the variance in the data set comprising of both Nelspruit and Florida data.

Satellite measurement errors were simulated using the Florida data. Two types of sensors were used, a perfect instrument, and an instrument that could measure rain area only. It was found that there was little to be gained using an area-only instrument with a sampling frequency greater than 16 visits a day when measuring mean areal daily rainfall. However, if the instrument has skill in measuring the rainrates, the mean standard error is approximately halved when the frequency is doubled from 16 to 32 visits per day. It was also found that estimates of the raining area were very sensitive to the resolution of the instrument. For example, a 64 km resolution sensor over-estimated the area by 300% of the area estimated by the 2 km sensor. The monthly areal mean rainfall was much easier to estimate, with a two visits-per-day sampling giving 22% errors.

The second part of the thesis investigated the transformation of the rainfall probability distribution under spatial and temporal averaging. This has immediate applications in a number of hydrological problems, for example using gauge data to estimate the statistics for mean areal rainfall. It also is relevant when attempting to compare statistics from rainfields with different spatial resolutions. An empirical method was developed to transform the entire probability distribution based on some measurement scale into a distribution for another measurement scale. It was found that the transformation of a rain rate at some probability level and scale, into a rate with an equal probability but over a different scale was well described by a power law. The two parameters for the transformation were dependent on the type of rainfield, the magnitude of the change in scale, and the scale itself.

A more general approach to the resolution dependence of the probability distribution would be to develop a resolution independent distribution. A conceptual model based on the theory of multiplicative cascades to describe the concentration of rain flux into smaller, more intense regions, was used to derive a resolution independent distribution. The good fit obtained by the model when used to predict the distribution outside the range of scales used to calibrate the model provides empirical evidence that rainfields may be usefully described by such a class of models.

During the latter part of the research it became apparent how sensitive rainfall statistics are to the space-time resolution of the data used in their estimation. It is imperative therefore to account for differences in resolution when comparing sets of rainfall fields based on different resolutions. It is also evident that the extreme intermittency and variability of short accumulation rainfields precludes accurate estimates of mean areal rainfall, even over large areas, using sparse networks. This emphasizes the important role of remote sensing in the estimating of rainfall amounts.