# Modelling the Instantaneous Spatial Variability of Rainfall

Kin C. Luk and James E. Ball
Water Research Laboratory, School of Civil Engineering, The University of New South Wales

ABSTRACT: Simulation of catchment processes which influence the quantity and quality of stormwater runoff is employed in many circumstances. Notably, when it is proposed to modify the existing catchment management practices, it is infeasible to physically trial the proposed management practice. An essential and necessary component of any catchment simulation is the model by which the rainfall over the catchment is estimated. Most of these models, such as Thiessen polygons, were developed prior to the development of the digital computer and the subsequent development of Geographic Information Systems (GIS). Presented herein are the results of a study investigating the accuracy and reliability of GIS for modelling the spatial and temporal distribution of rainfall over a catchment. It was found that using spline surfaces within a GIS produced robust and accurate estimates of rainfall and enabled real-time estimation of spatially distributed patterns. Furthermore, use of the GIS enabled the estimation of alternative hyetographs for different locations within the catchment.

## 1 INTRODUCTION

Management of the quantity and quality of water in urban drainage systems is a complex task which, over the last few years, has become increasingly important to the community. This community awareness has increased the need for managers of these systems to obtain information relevant to the response of the systems invested in their control. Two methodologies by which the desired system information can be obtained are, firstly, through monitoring of the system for stormwater quantity and quality, and, secondly, by mathematical simulation of the system, or systems, through quantity and quality models. Where management changes to the drainage system are proposed and it is desired that the impacts of these changes be predicted, it is necessary to use the second methodology; the first methodology can only provide historical information after implementation of the proposed management changes.

Ideally, models which fully replicate the processes and their spatial and temporal variability should be used. In practice, however, this does not occur because many processes are so complicated and interrelated that a full description may be impossibly complex. Even when a process can be described concisely and completely, the volume of calculations involved may be prohibitive. Also, the data available to define the model control parameter values are limited in both spatial and temporal dimensions. As a result, simplying assumptions are made and the real situation is idealised. Alternative idealisations result in the emphasis of different processes and require different magnitudes of computational effort. Consequently, instead of one model of reality, alternative models with differing degrees of complexity and computational effort may be developed.

Models may be defined as a collection of principles set out in a mathematical formulation. In the case of deterministic catchment models, a reductionist approach is adopted whereby the pertinent hydrologic and hydraulic processes are simulated by individual models with the response of the catchment being the summation of the individual process models. It is convenient, therefore, to arbitrarily subdivide the total model into a number of conceptual components with each of these conceptual components consisting of models simulating pertinent hydrologic and hydraulic processes. One such conceptual subdivision of a catchment model is that presented by Ball [1992] who proposed the following four components.

| | |
|---|---|
| • *Generation* | That component of the model primarily concerned with the estimation of the available quantity of water and pollutant constituents |
| • *Collection* | That component of the model primarily concerned with the accurate prediction of the quantity and quality of flow at the entry point to the transport component of the model. Generally this is the hydrologic component of the model. |
| • *Transport* | That component of the model where the water and pollutant constituents are routed along the channels of the catchment drainage system. Generally, this is referred to as the hydraulic component of the model. |
| • *Disposal* | That component of the model where the runoff and pollutant constituents are discharged into receiving waters. |

A schematic arrangement of these four conceptual components is presented in Figure 1. Also shown in this figure is the direction of information flow. It is important to note that this flow is unidirectional. If each conceptual component of the system is considered as a mapping of the input data and information to output data and information, then the transformation results in a nonlinear mapping due to the number of alternative transformations which arise from alternative idealisations of the catchment processes. There are, therefore, a substantial number of possible combinations of input data, information and transformations which produce the same output information. Consequently, a successful reproduction of the hydrograph or pollutograph at the catchment outlet does not imply that individual steps in the simulation system are correctly reproducing the relevant processes within the catchment nor that the predicted internal catchment information and data are accurate.

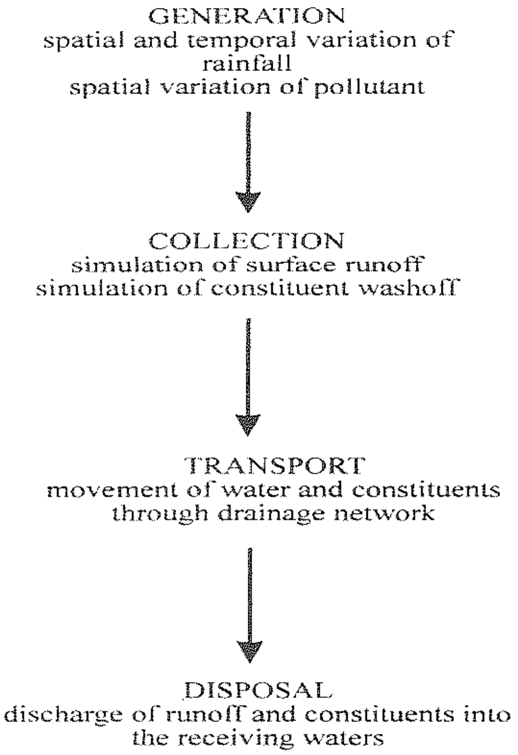

Figure 1 - Conceptual Components of a Catchment Model

The first conceptual component of a catchment model is the *generation* component where, from a water quantity viewpoint, the primary concern is the model used to estimate the spatial and temporal distribution of rainfall over the catchment. Most of the models used for the estimation of the spatial and temporal distribution of rainfall, such as Theissen polygons, were proposed prior to the development of the digital computer and the subsequent development of GIS. Presented herein are the results of a study investigating the accuracy and reliability of GIS for modelling of the spatial and temporal distribution of rainfall over a catchment.

## 2 SPATIAL AND TEMPORAL RAINFALL DISTRIBUTION

### 2.1 Alternative Rainfall Models

The measurement of rainfall during a storm event consists of determining the time over which an increment of rainfall depth occurs at a defined location. Consequently, the measurement of rainfall is a point measurement of a spatially variable parameter. Information, such as the rainfall intensity, at locations other than the measurement location is not defined by the measurement process and must be inferred from known information.

As discussed earlier one item of information required for the *generation* component of a catchment model is the rainfall distribution in space and time. Many alternative techniques have been developed for the inference of the spatial distribution of rainfall from the measured rainfall at a specific location; these alternative techniques are in effect alternative models of the spatial distribution of rainfall at an instance in time. For the temporal variability of rainfall at a specific location, it can be obtained by interpolation over the time increments used for assessment of the spatial variation in rainfall.

### 2.2 Theissen Polygons

Theissen polygons are probably the most common approach for modelling the spatial distribution of rainfall. As presented by Theissen [1911], the approach is based on defining the area closer to a particular gauge than any other gauges and the assumption that the best estimate of rainfall on that area is represented by the point measurement at the gauge. As the basis of the model is the geometry of the catchment, implementation of Theissen polygons in a GIS is not difficult.

An impact of the use of Theissen polygons, however, is the development of discontinuous functions defining the rainfall depth over the catchment. This effect arises at the boundaries of the polygons where a discrete change in rainfall depth, or intensity, occurs.

### 2.3 Inverse Distance-Weights

The inverse distance-weighted method as presented by Watson and Philip [1985] estimates the rainfall at a desired location, j, by a weighted interpolation based on the distance between each rain gauge and the desired location. Interpolation weights ($w_{ij}$) for each gauge are determined from

$$w_{ij} = \frac{d_{ij}^{-r}}{\sum_{i=1}^{n} d_{ij}^{r}} \qquad (1)$$

where $d_{ij}$ is the distance from the desired location j to gauge i and r is an exponent applied to the distance; typical values for r are 0.5 to 3. The rainfall at the desired location ($P_j$), therefore, is given by

$$P_j = \sum_{i=1}^{n} w_{ij} P_i \qquad (2)$$

Where $P_i$ are measured rainfall. Similar to the Theissen polygons, the inverse distance-weights are based on the geometry of the catchment and, hence, can easily be implemented within a GIS environment. In contrast to the Theissen polygon methods, however, this approach results in a smooth transition between rain gauge locations.

### 2.4 Kriging

Use of kriging for interpolation of rainfall at locations not monitored is based on the assumption of spatial homogenuity which is to say that the same pattern of variation can be observed at all locations within the catchment. Furthermore, it is assumed that the spatial variability of the rainfall is a function of a structural component which is associated with a constant trend, a spatially correlated component, and a random residual error term.

Using this assumption, rainfall at a point ($P_j$) can be estimated by

$$P_j = m(x) + \epsilon_1(x) + \epsilon_2 \tag{3}$$

where $m(x)$ is the structural component, $\epsilon_1(x)$ is the spatially correlated component and $\epsilon_2$ is the random residual error term having a mean of zero and variance of $\sigma^2$.

The variance of the spatially correlated component is assumed to be dependent only on distance so that

$$\begin{gathered} E\left[\{P(x) - P(x + \Delta x)\}^2\right] = \\ E\left[\{\epsilon_1(x) - \epsilon_1(x + \Delta x)\}^2\right] = \\ 2\gamma(\Delta x) \end{gathered} \tag{4}$$

where $\gamma(\Delta x)$ is known as the semi-variance which can be estimated from observed rainfall using

$$\gamma(\Delta x) = \frac{\sum_{i=1}^{n} \{P(x_i) - P(x_i + \Delta x)\}^2}{2n} \tag{5}$$

where n is the number of gauges separated by a distance of $\Delta x$. The relationship between $\gamma(\Delta x)$ and $\Delta x$ is known as the semi-variogram. An important element of kriging is the fitting of a mathematical function to the semi-variogram. Once this function is determined, the rainfall depth at any location can be estimated using equation 3.

Similar to the previous two approaches discussed, a GIS can be used to implement kriging for modelling the spatial distribution of rainfall.

### 2.5 Fitting a Continuous Surface

The basis of these methods is the fitting of a continuous surface function to the rainfall measurement points. There are two generic methods by which a surface function can be developed; these are

- polynomial surfaces
- spline surfaces

Other surface fitting methods are available but these two were considered the most appropriate for fitting of rainfall surfaces.

Polynomial surfaces are based on constructing a surface to fit the input data with a least-squares minimisation of the errors. The form of the polynomial function used is

$$f(x,y) = \sum b_{rs} x^r y^s \tag{6}$$

where the order of the polynomial surface is p which is given by

$$p \geq r + s \tag{7}$$

In equation 6 the coefficients ($b_{rs}$) are chosen to minimise

$$\sum_{i=1}^{n} \{P(x_i,y_i) - f(x_i,y_i)\}^2 \tag{8}$$

where n is the number of rain gauges, $P(x_i,y_i)$ are measured rainfall.

An important point with a polynomial surface is that, since the fitting of the polynomial function is based on a best fit for the complete catchment, the generated function may not pass through all the measurement points.

Spline surfaces in contrast is a local fitting technique which is based on interpolation between several adjacent measurement points using some low-order polynomials. This approach avoids overfitting the measurement points by high-order polynomials. In addition, an estimation of the order of polynomials is not required. Surfaces of this type have been found to be a robust spatial interpolation for many meteorological problems; for example, Hutchinson [1991] applied spline surfaces to fit long term monthly mean values of daily maximum and minimum temperature across Tasmania, Australia.

As implemented in this project, the spline surfaces have the following conditions imposed

- The surface must pass through all the known data points, and
- The surface must have the minimum curvature which is defined as the minimum of the cumulative sum of the squares of the second derivatives of the surface taken at each point on the surface.

The resulting spline interpolation method is referred to as *thin plate spline interpolation*. It ensures a smooth surface which is continuous and differentiable. Also provided are continuous first derivatives or slopes of the surface. Mathematically, spline surfaces are defined by

$$f(x,y) = T(x,y) + \sum_{i=1}^{n} \lambda_i R(r_i) \tag{9}$$

where n is the number of data points, $\lambda_i$ are weighting coefficients, $r_i$ is the distance from the desired point (x,y) to the measurement point i, and T(x,y) and $R(r_i)$ are defined by

$$T(x,y) = a_1 + a_2 x + a_3 y \tag{10}$$

and

$$\begin{gathered} R(r_i) = \frac{1}{2\pi}\left\{\frac{r^\phi}{4}\left[\ln\left(\frac{r}{2\tau}\right) + c - 1\right] + \right. \\ \left. \tau^\phi\left[K_0\left(\frac{r}{\tau}\right) + c + \ln\left(\frac{r}{2\pi}\right)\right]\right\} \end{gathered} \tag{11}$$

where $a_i$ are coefficients, $\tau$ and $\phi$ define the weight of the third derivative of the surface, r is the distance between the desired location and the measurement point, $K_0$ is a modified Bessel function, and c is a constant equal to 0.5772.

## 3 THE UPPER PARRAMATTA RIVER CATCHMENT

### 3.1 Catchment Details

As shown in Figure 2, the Upper Parramatta River Catchment is located in the western suburbs of Sydney, Australia. The Parramatta River drains into Sydney Harbour and is tidal to the Charles Street Weir in Parramatta; this weir is considered to be the downstream extent of the Upper Parramatta River Catchment Area. Total area of the catchment above this tidal limit of the Parramatta River is approximately 112km$^2$. Within the upper catchment area, the dominant land use is typical of urban environments with a mix of residential, commercial, industrial and open space (parkland) areas.

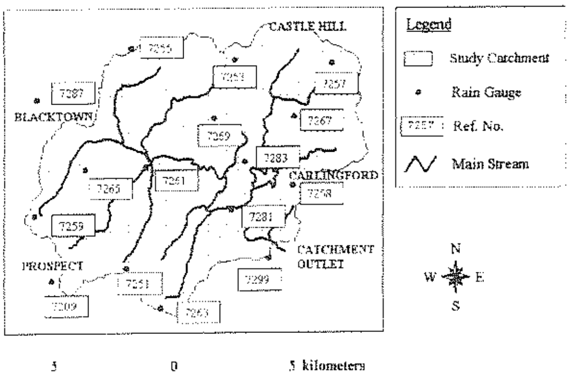

Figure 2 - The Upper Parramatta River Catchment

Considerable development has occurred within the catchment over the past two decades which has resulted in an increase in the frequency of recorded flood levels. To mitigate the social and economic losses associated with flood events in this catchment, the Upper Parramatta River Catchment Trust (UPRCT) was instituted in 1989 with the role of managing flood mitigation measures within the catchment area.

### 3.2 Available Data within the Upper Parramatta River Catchment

There are sixteen (16) rain gauges within the catchment; locations of these gauges are shown in Figure 2. The reason for installation of these gauges is for improving the flood forecasting for the catchment area and improving the accuracy of the catchment modelling system used to assess the impacts of alternative management practices. On average, one point rainfall sample is being obtained for every seven (7) km$^2$ of catchment. While this is a high density of information from rain gauge for most catchments, Urbonas et al. [1992] suggest that an even higher density of spatial information is required if accurate predictions of catchment response are to be obtained for convective events.

### 3.3 Rainfall-Runoff Model

For the catchment management purposes and, particularly, the flood management aspects of the catchment, the UPRCT has been implementing a rainfall-runoff model for the catchment. As part of the implemention of this model, the catchment has been subdivided into a number of subcatchments. The subcatchments considered in the analysis of the alternative rainfall models were based on the UPRCT's subcatchment delinearation. As expected, the area of each of these subcatchments was not constant but rather differed according to the catchment characteristics. The largest subcatchment was approximately 9km$^2$ while the smallest was approximately 1km$^2$. The remaining subcatchment areas were evenly distributed between these limits.

## 4 COMPARISON OF ALTERNATIVE RAINFALL MODELS

### 4.1 Methodology

This investigation of the spatial and temporal variability of rainfall over a catchment is aimed at improving the modelling of rainfall and, hence, the information and data transferred between the *generation* and *collection* components of a catchment model. In this investigation, the five (5) alternative rainfall models are implemented under a geographic information system (GIS).

A GIS is described by the Environmental Systems Research Institute [1995] as a collection of computer hardware, software, and spatial data designed to efficiently capture, store, manipulate, analyse, and update geographically referenced information. With the recent development of pixel-based modelling facilities, the ability of GIS to model the spatial variability of rainfall has been significantly enhanced.

For comparison between the alternative rainfall models, the ARC/INFO GIS was used as the software base for implementation of the alternatives. As well as providing basic modelling facilities, ARC/INFO provided a facility for programming the models through a macro programming language AML (Arc Macro Lanuage). Use of this language permits sequencing of ARC/INFO commands, and this enables repetition of operations and, in consequence, real-time operation of the system.

Within a GIS there are two main forms by which data can be stored; these forms are

- vector data which consists of points, lines and polygons; and
- raster data which is sometimes referred to as grid data and consists of information stored as an array of rows and columns.

There are merits in both forms of data storage; in fact, these two systems should be considered as complementary. Under certain conditions, such as for the study discussed herein, there is a need to convert between the two forms of data so as to maximise the utility of the available information and data. Within the scope of this investigation, vector data was used to define the catchment, subcatchments and drainage channels while the primary use of the raster data was the representation of the rainfall intensity, or depth, over individual grids within the catchment area.

In the comparison of the predicted rainfall over the catchment obtained from each of the alternative rainfall models, the rainfall intensity for each individual grid cell was obtained from the rainfall model. The spatial analysis tools of the GIS were then used to convert the grid cell data to mean rainfall depths (rainfall intensities integrated over a specified time interval) for individual subcatchments.

### 4.2 Storm Events

In comparing the alternative rainfall models, both real and artificial storm events were considered. The artificial storms were used to ascertain the accuracy of rainfall estimated by the alternative models under ideal conditions while the real events were used to assess the accuracy of rainfall estimated by the alternative models under conditions more likely to be found during real events. In addition, tests were undertaken with the assumption that some of the gauges were malfunctioning and, hence, were not recording.

The artificial storm event was a five (5) hour storm event which moved from east to west over the catchment at a speed of 3km/h. At all times during this event, the spatial distribution of rainfall had a Gaussian pattern which was defined by

$$p(x,y) = p_{max} e^{-az^2} \qquad (12)$$

where $p_{max}$ is the rainfall intensity at the centre of the storm, a is a coefficient which controls the spread of the rainfall and z is a position variable defined by

$$z = \sqrt{x^2 + y^2} \qquad (13)$$

Generation of the spatial rainfall distribution for this artificial storm event was based on the rainfall information at the sixteen gauge locations as if recorded during an actual event.

The two real events used in the comparison were extracted from the UPRCT's database for the dates of 5 Nov 1992 and 6 Nov 1992. Rainfall over these two days was of different magnitudes; this enabled an assessment of the alternative models for different magnitudes of rainfall.

### 4.3 Results

In comparing the predicted spatial rainfall distributions obtained from the alternative models, both visual and arithmetic comparisons were established. The arithmetic comparisons were based on the predicted rainfall occurring on each of the twenty-four (24) subcatchments into which the Upper Parramatta River Catchment had been divided for modelling purposes. In addition, for the artificial storm event, the comparisons were undertaken on hourly rainfall totals to consider motion of the storm centre.

For the artificial storm event, the following general characteristics of the alternative rainfall models were noted

- The Theissen polygon model resulted in a constant rainfall intensity within individual polygons which, in general, did not represent the actual pattern of rainfall.
- The Inverse Distance Weighted Model produced isolated peaks and troughs at the gauge locations. Additionally, the pattern of rainfall produced tended to differ significantly from the theoretical pattern.
- The model based on kriging was able to identify the centre of the storm but failed to recognise the storm extent and, consequently, tended to overestimate rainfall at the extremities of the catchment. It is suggested that this result is due to the lack of information at the extremity of the catchment.
- The polynomial surface was able to reproduce the Gaussian pattern of the rainfall when used to interpolate data. However, when extrapolation of data beyond the available information was required, the polynomial surface tended to predict negative rainfall depths.
- The model based on the spline surface resulted in the best arithmetic fit to the theoretical rainfall depths and, also, did not tend towards unrealistic estimates of rainfall depth when extrapolation was required.

The relative errors in the resultant areal average subcatchment hourly rainfall depth estimated by each of the alternative models were obtained. Ranges of these relative errors are presented in Table 1 while shown in Figure 3 are the estimation errors for one time interval as a function of the rainfall intensity. A measure of the error can be obtained from the vertical distance of the plotted estimations from the line of exact agreement.

Table 1 - Ranges of Relative Errors for Alternative Rainfall Models with the Artificial Storm Event

| | **RELATIVE ERROR (%)** | |
|---|---|---|
| | **Low** | **High** |
| **Theissen** | -76 | 488 |
| **Kriging** | -23 | 146 |
| **Inverse Distance Weighted** | -24 | 114 |
| **Polynomial** | -139 | 27 |
| **Spline** | -22 | 22 |

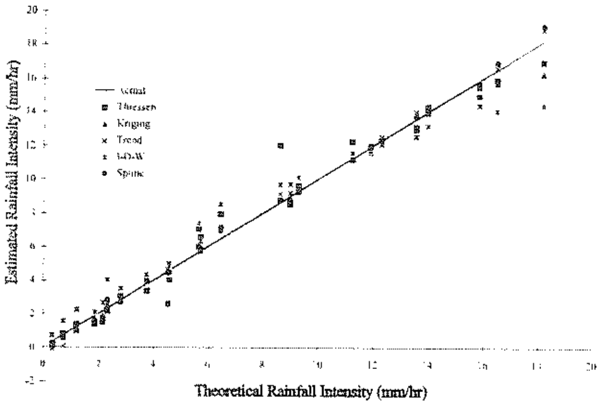

Figure 3 - Comparison of Estimated and Theoretical Rainfall Intensities

An additional aspect of the artificial storm event was the storm centre did not occur over any of the rainfall gauges. Of the five alternative models, only the kriging, polynomial and spline surfaces can predict peak rainfall intensities at locations remote from the gauges. However, the kriging model can predict peak intensities only slightly higher than those recorded. Both the polynomial surface and the spline surface provided reasonable estimates of the peak rainfall intensity with both tending to slightly overestimate the peak intensity.

Comparisons of the estimations from the alternative rainfall models were obtained also for real storm events that occurred on 5 November and 6 November 1992. In this case, the estimations could not be compared with theoretical values. Nonetheless, some valuable comparisons could still be made.

Presented in Table 2 are the volumes of rainfall over the catchment estimated by the alternative rainfall models.

Table 2 - Estimated Rainfall Volume

| | **Rainfall Volume (mm-km$^2$)** | |
|---|---|---|
| | **5 Nov 1992** | **6 Nov 1992** |
| **Theissen** | 1381 (0.6%) | 4582 (2.6%) |
| **Kriging** | 1492 (3.8%) | 4523 (1.3%) |
| **Inverse Distance Weighted** | 1382 (0.6%) | 4602 (3.1%) |
| **Polynomial** | 1398 (1.8%) | 4597 (3.0%) |
| **Spline** | 1280 (-6.8%) | 4016 (-10.0%) |

Note: A figure in bracket shows the relative variation from the mean volume.

As indicated by the volumes shown in this table, all five models estimated rainfall volumes that were within 10% of the mean volume; this mean volume of rainfall was determined from the average of the volumes estimated by each of the five alternative models. These variations in rainfall volume are less than the current accuracy with which lumped loss models, such as models of infiltration and interception, can be evaluated. Also shown in this table are the relative variations from the mean volume of rainfall. Due to these variations being from the average of the estimations, the sum of the variations must be and is equal to zero.

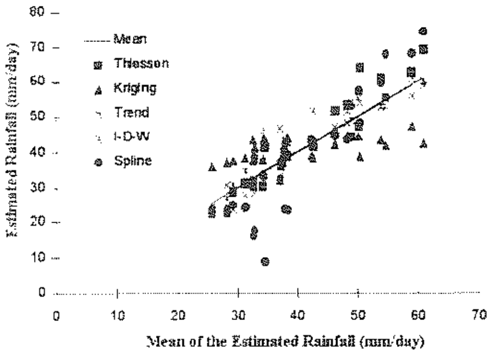

Figure 4 - Variations of Estimated Rainfall Depths form the Mean Rainfall Depths

Shown in Figure 4 is the variation in estimated rainfall depth on individual subcatchments compared to the mean rainfall depth for that subcatchment.

From an inspection of the spatial distribution of rainfall estimated by the alternative models, the following trends were noted

- Estimated rainfall depths obtained from the spline surface model tended to show the greatest variation from the average of all the models. Associated with this was a tendency for the spline model to produce a high estimate for situations where the average rainfall depth was high and to produce a low estimate when the average rainfall depth was low. This trend is shown in Figure 4.
- The model based on kriging produced the smoothest rainfall surface.
- The model using Theissen polygons showed the most abrupt changes in subcatchment rainfall.
- For individual subcatchments, the estimated rainfall depth obtained from each of the models displayed substantial variation; relative variations of 50% from the mean estimation (average of all estimations) were obtained.

## 5 CONCLUSION

Presented herein has been a discussion of the implementation of a number of models for the description of the spatial distribution of rainfall using a Geographic Information System. A comparison of the estimated distribution obtained from the alternative models was discussed also. Among the five alternative rainfall models, the Spline surfaces showed the best estimation of the patterns of rainfall for both artificial and real storm events.

## 6 REFERENCES

Ball, JE, (1992), A Review of Numerical Models for Prediction of Catchment Water Quantity and Quality, *Research Report No. 180*, Water Research Laboratory, Dept. of Water Engineering, School of Civil Engineering, University of NSW, 38 pages.

Environmental Systems Research Institute, (1995), *Understanding GIS - The ARC/INFO Method*, GeoInformation International (co-publisher John Wiley & Sons), Cambridge, UK.

Hutchinson, MF, (1991), The Application of Thin Plate Smoothing Splines to Continent-Wide Data Assimilation, in Jasper, JD (editor), Data Assimilation Systems, *Research Report No. 27*, Bureau of Meteorology, Australia, pp 104 - 113.

Theissen, AH, (1911), Precipitation Averages for Large Areas, *Monthly Weather Review*, 39(7):1082-1084.

Urbonas, BR, Guo, JCY, and Janesekok, MP, (1992), Hyetograph Density Effects on Urban Runoff Modelling, ***Proc. International Conf. On Computer Applications in Water Resources***, Tamkang University, Tamsui, Taiwan, pp 32 - 37.

Watson, DF, and Philip, GM, (1985), A Refinement of Inverse Distance Weighted Interpolation, *Geo-Processing*, 2:315-327.