

Die herleiding van 'n datapunkkonsentrasie-afhanklike gewigsfunksie vir interpolasie na reënval-roosterpunte

C. J. de W. Rautenbach

Leerstoel in Weerkunde, Departement Siviele Ingenieurswese, Universiteit van Pretoria, Pretoria, 0002

Onvang 5 Februarie 1996; aanvaar 13 Februarie 1996

UITTREKSEL

Verskeie interpolasiealgoritmes is reeds opgestel om homogene roosterpunte uit heterogeen verspreide datawaardes te skep. 'n Tekortkoming in baie van hierdie algoritmes is dat daar aangeneem word dat datapunkkonsentrasies oor die hele roostergebied redelik homogeen is. Die metode vir die herleiding van gewigsfunksies, soos hier bespreek, oorbrug die probleem deur 'n veranderlike in die gewigsfunksie te beskou wat afhanklik is van datapunkkonsentrasies. 'n Reënval-dataveld oor Suid-Afrika word deur die bogenoemde algoritme saamgestel. Waarnemingstasies vir klimatologiese veranderlikes oor Suid-Afrika varieer noemenswaardig in konsentrasie. Daar word aangetoon dat die invloedssfeer van 'n reënval-datawaarde minder gelokaliseerd raak namate die datapunkkonsentrasie afneem. Voorbeelde van hierdie analise word grafies geïllustreer.

ABSTRACT

The construction of a data point concentration dependent weight function for interpolation to rainfall grid fields

Several interpolation algorithms have been developed to create homogeneous grid fields from heterogeneously distributed data. A shortcoming in most of these algorithms is the assumption that data field concentrations are fairly homogeneous over the grid field area. The method for constructing the weight functions as discussed here overcomes this problem by including a variable dependent on the concentration of data points in the weight function. A rainfall data field for South Africa is compiled by this algorithm. The concentration of climate stations in South Africa varies significantly. It is shown that the sphere of influence of a rainfall data value becomes less localized as datapoint concentrations decline. Graphic examples of this analysis are given.

INLEIDING

Wanneer 'n diskreet waargenome dataveld soos 'n reënvalveld geanaliseer word, is dit soms nodig dat die veld in terme van 'n homogeen verspreide roosterveld voorgestel word. Die posisies van roosterpunte oor 'n area verskil meestal van die posisies van die datapunte. Daarom moet 'n geïnterpoleerde waarde toegeken word aan elke roosterpunt, en is elke roosterpuntwaarde dus afhanklik van die omliggende datawaardes. Omdat datapunte om 'n roosterpunt nie altyd ewe ver van die roosterpunt geleë is nie, moet elke datawaarde tydens die interpolasieproses 'n afstandafhanklike invloed op 'n spesifieke roosterpuntwaarde uitoefen. Hierdie invloed word bepaal deur aan elke datawaarde 'n afstandafhanklike gewig toe te ken. Verskeie gewigsalgoritmes is reeds opgestel om bogenoemde te illustreer.^{1, 2, 3} Topografiese effekte op reënval is ook al ingesluit by interpolasieprosesse.^{4, 5}

Die oneweredige konsentrasieverspreiding van reënvalstasies oor Suid-Afrika gee aanleiding daartoe dat datapunkkonsentrasies in ag geneem moet word tydens 'n interpolasieproses.

Die doel is nou om 'n eenvoudige maar tog prakties bruikbare interpolasiemetode daar te stel. Direkte interpolasie word beskou en topografiese effekte word buite rekening gelaat.

GEWIGSFUNKSIE

Die hieropvolgende herleiding van 'n datapunkkonsentrasie-afhanklike gewigsfunksie $W(r_{kij})$ is 'n uitbreiding op 'n Gauss gewigsalgoritme,^{3, 6} naamlik

$$W(r_{kij}) = e^{-a \cdot (r_{kij})^2} \quad (1)$$

waar $W(r_{kij})$ die gewigsfunksie voorstel wat toegeken word aan die k -de diskrete datawaarde naamlik C_k vir $k=1, \dots, n$ by die roosterpunt (i, j) . Hier is r_{kij} die afstand volgens die roosterveldskaal tussen die posisie van enige datawaarde C_k en die posisie van 'n roosterpuntwaarde C_{ij} . Die voetskrif (i, j) verwys na die i^e en j^e roosterpunt in die x en y rigting onderskeidelik, wat geleë is op 'n punt met koördinate $((i-1) \cdot \Delta x, (j-1) \cdot \Delta y)$ waar Δx en Δy die afstand tussen roosterpunte in die x - en y -rigtings onderskeidelik voorstel. Die roosterpuntwaarde C_{ij} lê dus op posisie $(0,0)$. Die waarde van die parameter a kan willekeurig bepaal word.

Die roosterpuntwaarde vir elke roosterpunt kan nou bereken word deur

$$C_{ij} = \frac{\sum_{k=1}^n W(r_{kij}) \cdot C_k}{\sum_{k=1}^n W(r_{kij})} \quad (2)$$

Vergelyking (2) gee die gewegde gemiddeld van datawaardes, om 'n roosterpunt, wat begrens word deur 'n invloedssfeer soos gedefinieer in $W(r_{kij})$.

Kenmerkend van gewigsfunksie (1) is dat die invloedssfeer van 'n datawaarde meer gelokaliseerd raak met 'n toename in die gekose waarde van die parameter a . Hierdie aspek word in figuur 1 grafies uitgebeeld deur funksies A ($a = 0,5$) en B ($a = 0,0003$). Aangesien die beskouing van die kwadraat $(r_{kij})^2$ in vergelyking (1) slegs bydra tot groter gelokaliseerd-

heid van die invloedseer van 'n datawaarde, kan $(r_{kij})^2$ met (r_{kij}) vervang word. Met laasgenoemde doel voor oë word die volgende gewysigde gewigsfunksie voorgestel:

$$W(r_{kij}) = e^{-a \cdot r_{kij}} \tag{3}$$

Vereistes ten opsigte van die invloedseer vir 'n datawaarde word vervolgens onder die loep geneem.

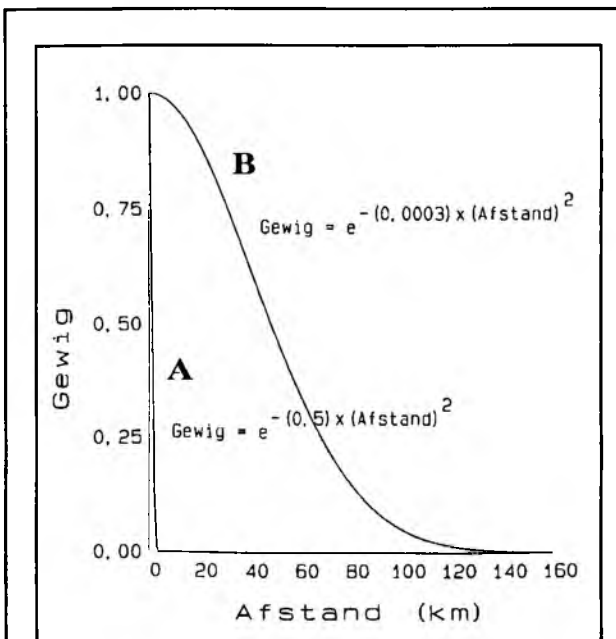
'n Relatiewe groot keuse vir die waarde van a gee daartoe aanleiding dat 'n datawaarde se invloedseer sterk gelokaliseerd raak. Oor gebiede waar datapunte yl verspreid is en r_{kij} relatief groot raak, kan gewigsfunksie (3) nadelig veroorsaak dat roosterpuntwaardes onrealisties klein voorgestel word.

'n Tipiese voorbeeld van so 'n dataveld is die waargenome reënvalveld oor Suid-Afrika. Figuur 2 gee die verspreiding van 2 100 reënvalstasies oor Suid-Afrika. Oor die oostelike deel van die land kom normaalweg 'n hoë konsentrasie van reënvalstasies voor. Daarom sal dit wenslik wees dat die invloedseer van datawaardes hier meer gelokaliseerd is. Oor die noord-westelike deel van die land (gebied A in figuur 1) is die afstand tussen stasies heelwat groter. Hier sal geïnterpoleerde roosterpuntwaardes, waar r_{kij} relatief groot is, baie klein raak indien dieselfte vereistes ten opsigte van die invloedseer van datawaardes geld as oor die oostelike deel.

Die doel is om die gelokaliseerdheid van die invloedseer van 'n datawaarde C_k vir $k=1, \dots, n$ by enige roosterpunt met posisie (i,j) groter te maak namate die afstand na die naaste datapunt kleiner raak en andersom. Definieer vervolgens:

$$r_{(min)ij} = \frac{\min}{k} (r_{kij})$$

Omdat 'n verandering in die waarde van a in die gewigsfunksie bydra tot 'n verandering in die gelokaliseerdheid van die invloedseer, word in die plek van a 'n funksie $a(r_{(min)ij}) = a_{ij}$ voorgestel wat sal varieer van roosterpunt tot roosterpunt soos $r_{(min)ij}$ varieer.



FIGUUR 1: Die grafiese voorstelling van gewigsfunksies A ($a = 0,5$) en B ($a = 0,0003$) dui aan dat die invloedseer van 'n datawaarde toeneem namate die waarde van die parameter a afneem.

Uit vergelyking (3) is dit duidelik dat die grootste gewig ($W(r_{kij}) = 1$ in hierdie geval) aan die datawaarde toegeken word wat op die roosterpunt self geleë is. Aangesien geen datapunt op 'n afstand kleiner as $r_{(min)ij}$ van enige roosterpunt voorkom nie, word daar vervolgens aangeneem dat die grootste gewig toegeken word aan die datawaarde wat op afstand $r_{(min)ij}$ van die roosterpunt af geleë is.

In besonder word 'n verdere gewysigde gewigsfunksie uit vergelyking (3) voorgestel:

$$W(r_{kij}) = e^{-a_{ij} \cdot (r_{kij} - r_{(min)ij})} \tag{4}$$

Die vorm van die funksie a_{ij} word nou bepaal. As voorwaarde word gewigsfunksie (4) so gekies dat $W(r_{kij}) = b$ waar $b > 0$ vir enige datawaarde C_k wat op afstand $2 \times (r_{(min)ij})$ van 'n roosterpunt C_{ij} geleë is. Dit sal tot gevolg hê dat die gewig stadiger na nul neig namate $r_{(min)ij}$ toeneem soos grafies voorgestel deur figuur 3 vir $b = 0,2$.

Die instel van bogenoemde voorwaarde in vergelyking (4) lewer:

$$b = e^{-a_{ij} \cdot (r_{(min)ij})}$$

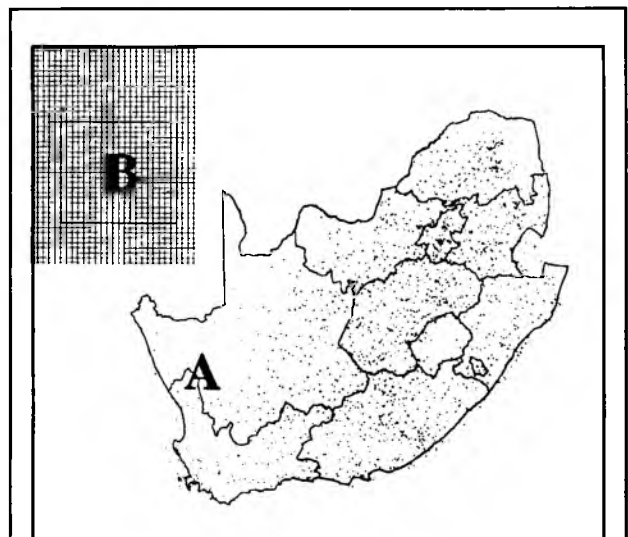
waaruit volg dat:

$$a_{ij} = -\ln(b) \cdot \frac{1}{r_{(min)ij}}$$

Vervang a_{ij} in vergelyking (4) wat lei tot 'n algemene vorm van die herleide gewigsfunksie:

$$W(r_{kij}) = b^{\frac{r_{kij} - r_{(min)ij}}{r_{(min)ij}}}; \quad r_{(min)ij} \neq 0 \tag{5}$$

Die parameter b kan volgens keuse aangepas word afhange van die afstandskaal, verspreiding van waargenome waardes en roostervelddimensies. Gewigsfunksie (5) is nie gedefinieer vir $r_{(min)ij} = 0$ nie. In so 'n geval word aangeneem dat $C_{ij} = C_k$.



FIGUUR 2: Die gebied (20° - 36° suid, 13° - 34° oos) waarvoor geïnterpoleer word. Die konsentrasieverspreiding van 2100 reënvaldatapunte word aangetoon. Laer konsentrasies kom voor oor gebied A. 'n Aanduiding van die roosterveldresolusie word deur B gegee.

RESULTATE

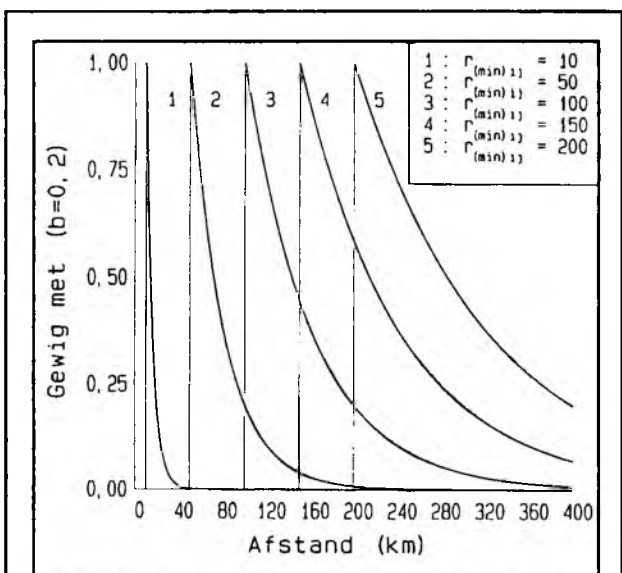
Die gewigsfunksie (5) is toegepas deur 'n heterogeen verspreide reënvalveld oor Suid-Afrika te interpoleer na 'n roosterveld met resolusie soos aangedui deur gebied **B** in figuur 2. Die afstand tussen roosterpunte in die x- en y-rigtings is na beraming 20 km. Die keuse van die roosterveldresolusie is belangrik en die digtheid van die reënval-datapunte speel ook hier 'n rol.^{7,8}

Om die invloed van gewigsfunksie (5) te ondersoek, word 'n 400 x 400 km-gebied gedefinieer met geselekteerde datawaardes op posisies soos aangedui met blokkies in figure 4, 5, 6. Uit hierdie figure is dit duidelik dat 'n groter konsentrasie datawaardes regs bo oor die gebied voorkom. Die relatiewe grootte van die area word deur die donker omlindende blok om **B** in figuur 2 voorgestel. Deur gebruik te maak van gewigsfunksies (1) en (5) is daar na 'n roosterveld

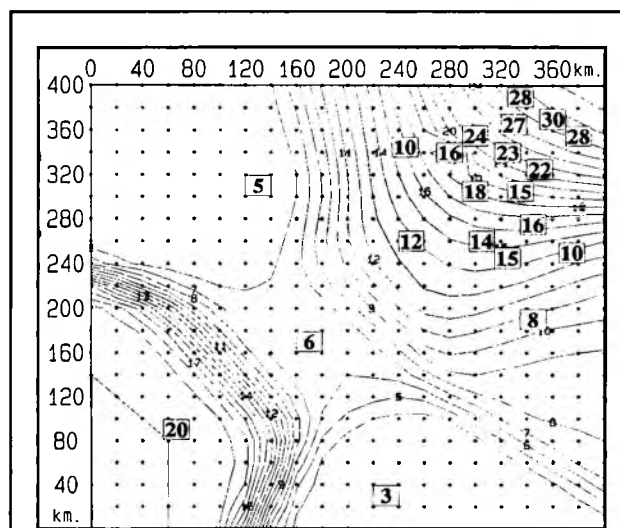
soos aangedui deur figure 4, 5, 6 geïnterpoleer.

Figuur 4 gee ruimtelike kontoere van 'n roosterveld soos geïnterpoleer deur gewigsfunksie (1) vir $a = 0,5$. Dit is duidelik dat die konstante keuse vir die waarde van a daartoe lei dat die invloedssfeer van 'n datawaarde te sterk gelokaliseerd is oor die gebied waar datapunte ver van mekaar geleë is. Hierdie probleem kan moontlik hanteer word deur $a = 0,0003$ as keuse te neem soos voorgestel in figuur 5. Dit gee egter aanleiding daartoe dat die invloedssfeer van datawaardes te groot is oor gebiede waar datapuntkonsentrasies hoër is. Die gewigsfunksies vir bogenoemde twee keuses vir die waarde van a word grafies in figuur 1 uitgebeeld.

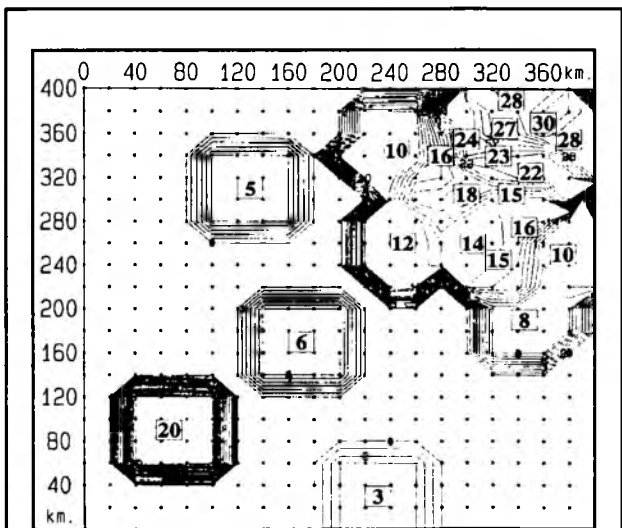
Figuur 6 gee ruimtelike kontoere van 'n roosterveld soos geïnterpoleer deur gewigsfunksie (5) vir $b = 0,2$. Oor die gebied (regs bo) waar hoër konsentrasies datawaardes voorkom, is die invloedssfeer van datawaardes merkbaar



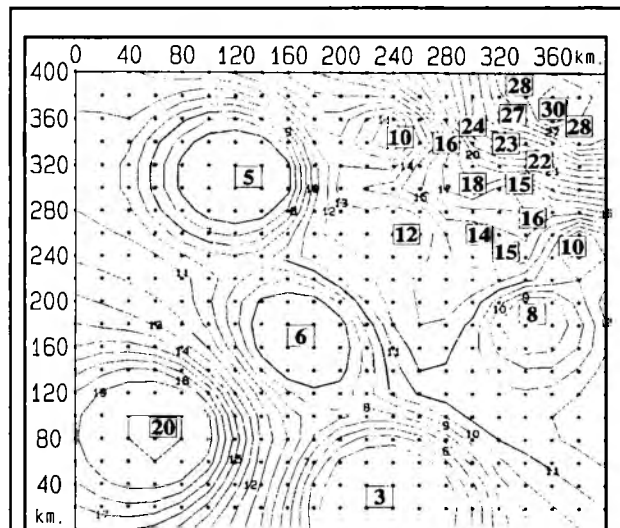
FIGUUR 3: Die grafiese voorstelling van gewigsfunksies vir geselekteerde waardes van $r_{(min)j}$ waar $b = 0,2$.



FIGUUR 5: Die ruimtelike kontoere van 'n roosterveld oor 'n 400 x 400 km-gebied soos geïnterpoleer deur gewigsfunksie (1) waar $a = 0,0003$.



FIGUUR 4: Die ruimtelike kontoere van 'n roosterveld oor 'n 400 x 400 km-gebied soos geïnterpoleer deur gewigsfunksie (1) waar $a = 0,5$. Die reënval-datapunte word in blokkies aangedui.



FIGUUR 6: Die ruimtelike kontoere van 'n roosterveld oor 'n 400 x 400 km-gebied soos geïnterpoleer deur gewigsfunksie (5) waar $b = 0,0003$.

kleiner as die gebied waar datawaardes verder van mekaar geleë is. Hoe groter die keuse vir die waarde van b , hoe meer gelokaliseerd word die invloedseer van 'n spesifieke datawaarde. Die gewigsfunksies vir verskillende waardes van $r_{(min)ij}$ met $b = 0,2$ word grafies in figuur 3 uitgebeeld.

GEVOLGTREKKING

Daar is in die doel geslaag om 'n datapuntkonsentrasie-afhanklike gewigsfunksie (5) op te stel waar daar volgens behoefte besluit kan word oor die gelokaliseerdheid van die invloedseer van 'n datawaarde. Die invloedseer van 'n datawaarde raak meer gelokaliseerd namate die waarde van die parameter b in vergelyking (5) toeneem, terwyl daar ook gladder geïnterpoleer word na roosterpuntwaardes wat relatief ver van datawaardes geleë is. Daar moet egter altyd in gedagte gehou word dat die roosterveld uit geïnterpoleerde roosterpuntwaardes bestaan, en dat realisties gemete reënvalwaardes slegs voorkom by die datapunte self.

SUMMARY

The proposed data point concentration dependent weight function is an extension of a Gaussian weighting scheme,^{3,6} i.e.,

$$W(r_{kij}) = e^{-a \cdot (r_{kij})^2} \tag{1}$$

where $W(r_{kij})$ is the weight function of the k^{th} data value C_k for $k=1, \dots, n$ at grid point (i,j) . The distance between the position of any data point value C_k and a grid point value C_{ij} is denoted by r_{kij} . The subscript (i,j) refers to the i^{th} and j^{th} grid point in the x and y direction, located at a point with coordinates $((i-1) \cdot \Delta x, (j-1) \cdot \Delta y)$ where Δx and Δy are the grid point interspace values in the x and y directions respectively. The value of parameter a is chosen arbitrarily.

The grid point value for each grid point can be determined by a method of weight interpolation defined by

$$C_{ij} = \frac{\sum_{k=1}^n W(r_{kij}) \cdot C_k}{\sum_{k=1}^n W(r_{kij})} \tag{2}$$

The influence of a data point value becomes more localized as the value of a in equation (1) increases. Seeing that the square $(r_{kij})^2$ in equation (1) only contributes to the radius of influence of a data point value, it can be replaced by r_{kij} . This results in the following modified weight function. A disadvantage of weight function (3) is that the computed grid point value C_{ij} at

$$W(r_{kij}) = e^{-a \cdot r_{kij}} \tag{3}$$

position (i,j) can become unrealistically small if a , as well as the smallest value of r_{kij} , become relatively large (widely distributed data points as shown over area A in figure 2). It is, therefore, appropriate to enlarge the radius of influence of a data point value C_k at any grid point (i,j) as the distance from the grid point to the nearest data point increases. Define

$$r_{(min)ij} = \frac{\min}{k} (r_{kij})$$

The value of a strongly contributes to the radius of influence of any data point value C_k . Replace a by the function $a(r_{(min)ij}) = a_{ij}$. Let the maximum weight ($W(r_{kij}) = 1$ for this case) be allocated to a data point at distance $r_{(min)ij}$ from a grid point. Thus

$$W(r_{kij}) = e^{-a_{ij} \cdot (r_{kij} - r_{(min)ij})} \tag{4}$$

Define a parameter b by letting $W(r_{kij}) = b$ for $b > 0$ at any data point at distance $2 \times (r_{(min)ij})$ from the position of grid point C_{ij} . Applying this to equation (4) results in the following weight function

$$W(r_{kij}) = b^{\frac{(r_{kij} - r_{(min)ij})}{r_{(min)ij}}}; \quad r_{(min)ij} \neq 0 \tag{5}$$

The weight function (5) was evaluated by interpolating 2 100 rainfall stations throughout South Africa to a grid field as indicated by B in figure 2. To investigate the influence of weight functions (5) and (1), a 400 x 400 km domain with data values as indicated in figures 4,5,6 was selected. A higher concentration of data values occurs in the top right section of the domain.

Figure 4 and figure 5 give the spacial contours of grid fields interpolated by function (1) for $a = 0,5$ and $a = 0,0003$. The weight functions are graphically displayed in figure 1. Figure 6 gives the spacial contours of a grid field interpolated by the constructed weight function (5) for $b = 0,2$. The weight functions for different values of $r_{(min)ij}$ are graphically displayed in figure 3.

A data point concentration dependent weight function (5) was constructed where the radius of influence of a data value decreases with an increase in the value of the parameter b , and increases with an increase in the distance $r_{(min)ij}$.

LITERATUURVERWYSINGS

1. Cressman, G.P. (1959). An operational objective analysis system, *Mon. Wea. Rev.*, 87, 367-374.
2. Glahn, H.R. (1980). Comments on A Comparison of Interpolation Methods for Sparse Data: Application to wind and concentration fields, *J. Appl. Meteor.*, 20, 88-91.
3. Goodin, W.R., McRae, G.J. & Seinfeld, J.H. (1979). A comparison of interpolation methods for sparse data: Application to wind and concentration fields, *J. Appl. Meteor.*, 18, 761-771.
4. Dent, M.C., Lynch, S.D. & Schulze, R.E. (1989). Mapping mean annual and other rainfall statistics over Southern Africa, Dept. of Ag. Engineering, University of Natal. *WRC Report No 109/1/89*.
5. Schäfer, N.W., Dent, M.C. & Schulze, R.E. (1991). A model for areal distribution of daily rainfall. Fith South African National Hydrological Symposium, University of Stellenbosch. 4A-2-1 - 4A-2-10.
6. MacCracken, M.C. & Sauter, G.D. (1975). Development of an air pollution model for the San Francisco Bay Area, Vol. 2, Appendices. *Lawrence Livermore Laboratory*, UCRL-51920, 229-230 [NTIS UCRL-51920].
7. McKay, G.A. (1976). Hydrological mapping. In *Facets of Hydrology*, Rodda, J.C. (ed.) (John Wiley and Sons, London) pp. 1-35.
8. Pawling, J.M. (1973). Trend surface analysis of local relief in the southern peninsula of Michigan, *Prof. Geogr.*, 25, 226-232.