SPATIAL INTERPOLATION AND MAPPING OF RAINFALL (SIMAR)

VOLUME 3. DATA MERGING FOR RAINFALL MAP PRODUCTION

GEOFFREY G.S. PEGRAM

Civil Engineering Programme
University of Natal, Durban

Final Report to the Water Research Commission

by the

University of Natal

December 2003

WRC Report No 1153/1/04
ISBN 1-77005-159-7

EXECUTIVE SUMMARY

The programme ***Spatial Interpolation and Mapping of Rainfall (SIMAR)*** was a three-year initiative encompassing three component projects, viz:

- Maintenance and upgrading of radar and raingauge infrastructure
- Radar and satellite products
- Optimal integration of raingauge, radar and satellite derived data in the production of daily rainfall maps

The final report on this programme, which was undertaken by scientists, researchers and engineers of the METSYS group of the South African Weather Service (SAWS) and the School of Civil Engineering of the University of Natal, in collaboration with the Department of Water Affairs (DWAF) and ESKOM, is contained in three volumes. The volumes are:

VOLUME 1. ***Maintenance and Upgrading of Radar and Raingauge Infrastructure***

VOLUME 2. ***Radar and Satellite Products***

VOLUME 3. ***Data Merging for Rainfall Map Production***

Rationale for SIMAR

Water resources in South Africa are not well buffered against natural rainfall variability. Rainfall deficits and excesses readily translate to droughts and floods, respectively. Well-developed water resource infrastructure, which South Africa is fortunate to possess, has to be managed extremely skillfully to successfully balance water surpluses and deficits at an inter-catchment level, as well as to achieve best trade-offs between flood mitigation and storage maximisation at basin level.

The concept of water resource management is no longer restricted to regulating flows, storage and abstractions in and from rivers, dams and aquifers. Water resource management is increasingly becoming concerned with applying measures to ensure resource (ecosystem) sustainability and also with activities in the catchment which impact on both sustainability and availability of water for abstraction and use. A particular focus in the 1998 Water Act is on activities, termed streamflow reduction activities, the licensing and regulation of which are provided for in the Act. To do this objectively requires defendable information on water usage associated with entities such as forests, agricultural lands, natural veld, farm dams, soil conservation schemes, etc. Such water usage is tightly linked, through the catchment water balance, to catchment water availability and thus rainfall, a link which imposes an obligation on catchment management agencies to obtain detailed and accurate rainfall measurements.

Since natural disasters and fluctuations in agricultural production are also closely linked to rainfall, these sectors have a similar need for such detailed rainfall information.

Raingauges have traditionally provided the rainfall measurements required for water resource management purposes. Because the national raingauge network is rapidly becoming too sparse to meet existing and anticipated management requirements, a new rainfall monitoring/information system, incorporating the optimal use of remote sensing, has become necessary to satisfy the needs of South Africa.

This need was envisaged to be best satisfied through an umbrella research and development programme (SIMAR) having the ultimate goal of merging satellite/radar/gauge data to produce **one field** that is acceptable to the water-resources (and hence also agricultural and disaster-management) users.

The specific aim was to produce a daily rainfall map of 24 hour accumulated rainfall to a resolution of 2 km, over the whole subcontinent, accessible on the Internet. This has been accomplished. Furthermore, this primary product can and will be refined where needed to finer time scales over selected areas. This refinement will be of particular interest to the disaster-management users and those involved with the mapping of the tendency for, or forecasting of, flash floods. As improvements to the data streams and modelling techniques become available, they will be incorporated into the products emanating from the research. Follow-on research projects supported by the Water Research Commission are designed to bring about such improvements.

Results of SIMAR component projects

Details of the results deriving from the three component projects included under the SIMAR umbrella are contained in the three volumes which make up the SIMAR final report. Some background information pertaining to these results, and the results *per se*, are summarised below.

*VOLUME 1. **Maintenance and Upgrading of Radar and Raingauge Infrastructure***

Since conventional meteorological infrastructure is dwindling at an alarming rate in South Africa, it became necessary to investigate the complementary use of conventional and less conventional infrastructure in sourcing rainfall data. The complementary sources here considered are surface networks and remote sensing sources, namely radar and satellite. The focus has fallen on maintaining current systems as well as using new technologies and techniques to upgrade systems, where necessary, with a view to securing and sustaining a reliable data flow from the above-mentioned data sources. The specific objectives of this part of the programme were as follows:

- Maintain automated surface gauge networks (Durban and Liebenbergsvlei) as well as investigate the application of new technology to ensure real time availability of data.
- Maintain and upgrade the National Weather Radar Network (NWRN) by improving the monitoring capabilities at each radar system, investigating the possible expansion of the network coverage and pursuing an improved funding base for the radar network.
- Utilise the latest remote sensing technology in order to improve the quality of products to be generated.
- In cooperation with other stakeholders, assist in establishing a real time precipitation database for South /southern Africa.

- Actively seek and promote collaboration with stakeholders, institutions and organisations, on a formal and informal basis, to expand and enhance surface coverage of precipitation measurements.
- Equip and train individuals, especially those from previously disadvantaged communities, with a view to their acquisition of remote sensing and electronic maintenance skills.

The raingauges of the Liebenbergsvlei and Durban networks played a vital role in investigations regarding elimination of ground clutter and in validation of radar-estimated rainfall on the ground. An investigation into the feasibility of using cell phone communication technology and infrastructure resulted in such technology being implemented in the Liebenbergsvlei and Durban networks, and gave rise to the vision of also implementing the technology at all the SA Weather Service second order stations.

Improvements and upgrades were introduced at the majority of radar installations within the network, while also ensuring a reliable power supply to the systems. A remote control and monitoring system, whereby the functioning of individual radar systems can be monitored from a central point, was implemented. This new capability has, for the first time, allowed objective assessments of the reliability of individual radar systems within the NWRN to be made. During the course of this project, the SAWS drastically increased its funding for maintenance and upgrading of the NWRN.

Unfortunately, the Meteosat Second Generation (MSG) satellite did not become operational during the lifetime of the SIMAR project as anticipated, as the MSG programme suffered lengthy delays prior to, and further problems after, the launch of the satellite. Nevertheless, research into utilising Meteosat 7 to fulfill the SIMAR project's objectives, continued. New techniques were developed which will be applied to MSG when it eventually becomes operational.

Close cooperation with the SAWS database developers led to provision being made for the archiving of real-time data generated by the SIMAR programme. The development of the new database has proceeded through different phases and will continue to be developed to accommodate such data products for routine applications and research purposes.

This component project succeeded in promoting data sharing between institutions, albeit on a small scale, and initially limited to operational exchange of data through collaboration with DWAF and Suidwes Agriculture. The pursuit of this objective of institutional collaboration will not end with SIMAR but will be carried on and expanded.

The training of personnel in the maintenance and upgrading of observational systems received high priority in the SIMAR project. The training initiative was also expanded to the international arena with training on SIMAR related subjects presented to students from other African countries such as Botswana and Tanzania.

*VOLUME 2. **Radar and Satellite Products***

The focus of this component project was to provide the two remote sensing-based rainfall fields – the one derived from radar and the other from satellite – to be merged with yet another rainfall field, derived from daily reporting raingauges.

The specific objectives were as follows:

Radar products

- Provide a Meteorological Data Volume (MDV)-based, real-time, radar rainfall map of the radar covered area of South Africa
- Optimize merging in areas of radar overlap and utilize the reflectivity measurements in these areas for additional performance testing
- Improve radar-rainfall algorithms which address the outstanding issues of data quality and integrity (hail, bright band, ground clutter and coastal/orographic rain)
- Include additional radar information as it becomes available and identify the most serious gaps in South Africa's weather radar coverage.

Satellite products

- Investigate and develop suitable rainfall estimation algorithms from satellite data for South Africa, which also address known problems related to coastal/orographic rain
- Incorporate the latest satellite (Meteosat Second Generation, which should have become available during the project period) in order to address issues related to temporal resolution
- Provide an MDV-based, real-time, satellite rainfall map of South Africa
- Integrate raingauge, radar and satellite rain fields to provide the desired daily rainfall maps, making use of the modelling component (reported in *VOLUME 3* of the SIMAR final report).

Despite the fact that the MSG Satellite did not become operational during this project as originally anticipated, SIMAR accomplished major advances in the estimation of rainfall using remote sensing techniques and the integrated mapping of rainfall over South Africa.

Radar rainfall estimation

South Africa's NWRN represents a unique system based on a local solution to the complex problem of networking of several individual radars and merging of their individual data fields. This system, developed in-house, is a combination of South African innovation and shareware/freeware available from various sources in the world. Very few countries in the world operate successful weather radar networks and even fewer a network as elegant and modular as the one in South Africa. The report gives a summary of how this was achieved and highlights the data flow and product generation from the eleven radars within the network.

A major advance in radar rainfall estimation is the unique methodology that was developed to filter the negative impact of ground clutter. This technique, that uses the scan-to-scan coherence in the echo field to dynamically build up a slowly evolving clutter mask, has all but solved the problems of ground clutter contamination. This advance has had a positive impact on the resolution of the rain fields that are archived, displayed and used as input to create the integrated satellite-radar-raingauge rain fields.

The verification of radar performance, and independent procedures and tests to investigate the inter-calibration of network radars, have been fully documented. Of significance is the success with which the sun has been used as an independent calibration source.

Conversion of radar reflectivity into rain rate and the computation of rainfall depth accumulations have been refined considerably. Using the dense Liebenbergsvlei raingauge network as a basis for comparison, the mean reflectivity in the vertical column was found to be a more appropriate input to the Z-R relationship than the customarily-used maximum reflectivity in the vertical column. The mean reflectivity has a smoothing effect on the enhanced or erroneous reflectivity caused by the occurrence of hail, or by the so-called *Bright-Band* and *Anomalous Propagation* phenomena.

Furthermore, methods have also been developed to generate a merged rain field from rain fields generated at the individual radar sites instead of from a merged reflectivity field. This allows the use of different (and more appropriate) Z-R relationships for different regions and also allows use of data with the finest temporal resolution. Rainfall estimates over South Africa obtained in this way has been evaluated using rainfall data from 60 automatic weather stations across South Africa.

The above-mentioned studies and advances all provided a sound foundation for improvements already made to the NWRN rainfall estimation techniques, or for improvements due to be implemented in the near future.

Satellite rainfall estimation
Building on a review of literature on past South African and international experience, a technique (probably the most sophisticated satellite-rainfall estimation technique yet available for South Africa) that makes optimal use of all three channels (IR, Visible, Water Vapour) of the current Meteosat 7 satellite was developed and implemented operationally. Particular attention was also given to the characteristics of the MSG Satellite. Although it did not become operational during the project as originally anticipated, some of the first data examples from this satellite are shown.

The first stage in the evolutionary development of the MSRR (Multi-Spectral Rain Rate technique) was the ITR (Infra-red Power Law Rain Rate technique), which gave rise to the intermediate BSRR (Bi-Spectral Rain Rate technique). The use of all three channels leads to improved methods for filtering out non-precipitating clouds and enhances the estimation of rainfall from maritime clouds. Image processing techniques (including edge detection and speckle removal techniques) were also introduced to better identify rainy pixels from those that are cloudy but not rainy. Systems to reformat and communicate the satellite data in the same MDV format being used for radar data were developed. A novel development in the satellite rainfall estimation was the use of topographical slope to enhance estimated rainfall over mountainous regions. The advantages of using a Geographical Information System to display and process the data and products from the various sources was clearly demonstrated.

Methods to verify the satellite rainfall estimates in terms of their spatial extent and quantitative values through comparisons with radar and raingauge estimates were developed and applied. It became clear that the satellite rainfall estimation technique which was developed achieves the objective of providing useful, large-scale rainfall fields for the southern Africa region.

Rainfall data integration and product distribution

An analysis of the strengths and weaknesses of raingauge-, radar- and satellite-derived rainfall information provided the basis for additional measures to address the major weaknesses in each source. The accuracy (including human-introduced error factors) and coverage of daily raingauge data, in particular, has become a matter that needs to be addressed as a national priority.

The generation of the merged satellite-radar-raingauge field is a stepwise process, starting with the merging of the radar and raingauge fields. Thereafter the satellite and raingauge fields are merged before the two resultant fields are combined.

SIMAR has a dedicated section on the South African Weather Service (METSYS) web page (http://metsys.weathersa.co.za) which displays the various individual daily rainfall fields (radar, gauge and satellite) together with the integrated fields. Archived data relating to these fields are also presented.

VOLUME 3. ***Data Merging for Rainfall Map Production***

This component project was initiated against the background of the following premises and objectives:

- There existed a large collection of daily-read rainfall data in the country, which were (and still are) continuously being added to. The first aim of the research was to be able to interpolate optimal rainfields between raingauges at individual locations and also to suggest the best estimates of catchment (areal) totals of rainfall, both historically and currently.
- The accuracy of radar in pinpointing, in considerable detail, where rain falls was not in question, but some difficulties still existed in terms of estimating rain rates from radar data. The second aim was to combine raingauge and radar data into a meaningful composite to provide an optimal rainfield acceptable to users. However, radars did and continue to cover only part of the country. There are less densely-populated areas with sparse raingauge coverage and no radars, but where satellite surveillance information could be accessed. It was therefore considered important to link satellite, radar and gauge data together to obtain the best estimate of rainfall in these remote areas.
- The third and main aim of the research was to devise a product to enable the publication, on a daily basis, of the 24-hour rainfall over the country. The means of achieving this aim was seen to be optimal integration of gauge, radar and satellite estimated data, which would of necessity improve with commissioning of more radars, upgrading of software and introduction of a new generation of satellites.

Theoretical development

Combining the precision of raingauge data with the coverage of satellite data and the detail of radar data was, in effect, an important objective of this research. The techniques initially envisaged as a means of achieving this were optimal spatial interpolation using a technique called Kriging and an associated one called co-Kriging.

It turned out that co-Kriging was not a good option because of the large computational load. This load comes from the fact that there are of the order of one million small areas (pixels) approximately 1.5 kilometers square (the typical spatial resolution of a weather radar) covering the subcontinent and the surrounding oceans. The challenge was to be able to map the country's rainfall routinely to that detail. Even the quadrupling of computer speed, between the year 2000 (when the project was proposed) and its end in 2003, did not diminish the need to find a better way to process data, which would be easy to automate. A method of Kriging, exploiting the efficiency of the Fast Fourier Transform, was consequently developed. The process of development necessitated having to deal with a highly technical subject, involving some difficult and advanced mathematical ideas and theory.

Outcome and Technology Transfer
The techniques developed for optimal integration (merging) of data fields and their implementation to date have been most fruitful. The daily rainfall maps on the SAWS:METSYS website bear testimony to this successful outcome.

The Fast Fourier Transform approach to Kriging provided the basis for the coding of an algorithm to accomplish the massive computing task efficiently and speedily. Speed is of the essence in the delivery of the daily rainfall maps in real time. Information on the accumulated rainfall for the 24 hours until 8:00 am SA time, derived from the recording raingauges around the country, arrives at METSYS (Bethlehem) by 9:00 am daily. By that time, the previous 24 hours' satellite and radar images will have been used to produce the best estimates, respectively, of the rainfall totals per pixel over the whole area. The merging of the three fields: gauge, radar and satellite is then done and the result posted on the METSYS web-site by 11:30 am. A thorough description of the practical implementation of this methodology is presented in the body of *VOLUME 2* of the SIMAR final report. Some examples of the website output are reproduced therein.

Conclusions and recommendations

SIMAR has successfully met its objectives and laid the foundation for a national (and potentially regional) rainfall observing system which promises to meet all reasonable requirements regarding spatial and temporal resolution and real-time availability of data.

There are several areas in which the current SIMAR system, with further attention to data availability, research and development, can be improved. These are:

- *Radar inter-calibration* – improved techniques to constantly monitor the complex radar calibrations within the NWRN.
- *Modern electronic techniques* – ongoing development and use of modern electronic technology to improve data collection, communication, processing and storage.
- *Radar and satellite product research* – ongoing research to improve the quality of remote sensed data and derived products.

- *Data exchange and availability* - It is evident that much more work is required on the issues related to institutional willingness to collaborate and share information. The process should occur at institutional level and be formalized through Memorandums of Understanding or other binding means. In addition government agencies should consider making remotely sensed data available free of charge and without any restriction or accessibility issues to researchers. This can only occur if government sponsors the necessary infrastructure to obtain remotely sensed data from an array of platforms to be utilised for research and training purposes.
- *Speeding up and refinement of merging algorithms* – the currently used method of combining gauge, radar and satellite measurements of rainfall can be refined using variants of Kriging which exploit the clustering of measured data and regions to be infilled.
- *Repair of weather radar images of rainfall* – ground clutter and anomalous propagation are nuisance contaminants of images of rainfall estimated by radar. These can be infilled with good estimates of rainfall if they have been identified correctly.
- *Accumulation of rainfall from radar and satellite images* – because the radar and satellite images are instantaneous snapshots of rainfields, the naïve superposition of the images gives a false accumulation field when total depths of rain are required. A method of morphing based on the calculated advection field will overcome this present deficiency.

The SIMAR system would also benefit from certain infrastructural improvements, the most crucial of these being:

- *Raingauge network* – incorporation of all qualifying raingauges in South Africa and the region, irrespective of institutional ownership. The modernization of the raingauge infrastructure through the use of modern electronics and communication systems to provide better temporal resolution on a real-time basis.
- *Radar network* – incorporation of all operational radars and standardisation of operations and data acquisition systems in the region. The expansion of the NWRN to fill areas not covered.
- *Satellite* – immediate exploitation of opportunities presented by the deployment of the MSG satellite.

The above envisaged improvements build on the existing platform of work developed under SIMAR and will make a considerably more acceptable product. Follow-on projects already under way are addressing these issues with energy.

CAPACITY DEVELOPMENT

The capacity developed at both technical and professional levels through SIMAR has provided a sound foundation upon which further capacity can be built.

The training of technical personnel in the maintenance and upgrading of observational systems received high priority in the SIMAR programme. Four individuals from the previously disadvantaged groups were trained in maintaining the electronic observational infrastructure thus ensuring the long-term sustainability of the observing systems. Training was conducted in-house, through courses as well as self-development. The training initiative was also expanded to the international arena with training on SIMAR related subjects presented to students from other African countries such as Botswana and Tanzania. As SIMAR products are used routinely by institutions, training will continue well beyond the lifetime of this project. It is especially training in the utilisation and interpretation of SIMAR products where a strong need exists. Users should also be trained and educated in the use of remotely sensed data, its advantages as well as its limitations.

There are two aspects to professional capacity building which were achieved here – indirect (people being exposed to the ideas and concepts but not working on the project) and direct (those people personally involved with aspects of the project). In addition, there was a strong component of Competency Development as a direct result of the project.

Indirect Capacity Development

In the Hydrology Section of Umgeni Water, where one of the researchers (Scott Sinclair) worked in 2001 and 2002, two PDIs were kept abreast of the developments of the project in both informal and formal (reports, presentations) ways. The 2002 final year class of 28 Civil Engineering Students in Hydrology at the University of Natal, Durban, contained 16 PDIs (of whom 5 were women) and 2 white women. The project co-leader (Geoff Pegram) made frequent reference to the SIMAR in class and repeated the oral presentations given in this regard at the European Geophysical Society in Nice in April 2002. These presentations tempted two students from previously disadvantaged backgrounds to undertake dissertations under the project co-leader's supervision during the second semester of 2001

Direct Capacity Development

In the second semester of 2001, a female final year student, Deanne Everitt, undertook a dissertation study under the supervision of the project leader entitled "Flood Impacts: Planning and Management". This was an overview study making use of the output from SIMAR, with special focus on the Umlazi catchment in Durban. A later addition to the team was Nokuphumula (Phums) Mkwananzi, a practising Engineer, who registered for an MScEng at Natal University under the supervision of the project co-leader in 2002, worked on the WRC project "Extension of Research on River Flow Nowcasting to include Levels of Inundation" which depended on SIMAR input of rainfields, and completed his Masters in September 2003.

Competency Development

Because of the nature of the Research, a number of people in Umgeni Water, Durban Metro/eThekwini Municipality, SAWS: METSYS and the University of Natal have been exposed to new ideas and potentials for ameliorating flood damages using the ideas that are direct spinoffs from SIMAR; new technology has been developed and existing technology has been improved and refined. Every individual involved has grown in competence and benefited from the project; in the long run the wider community in the region will be beneficiaries.

KNOWLEDGE DISSEMINATION

Knowledge generated by SIMAR has been disseminated through peer-reviewed articles, conference presentations, workshops and during international visits. These include the annual South African Society for Atmospheric Sciences (SASAS) conferences. The SASAS conference that coincided with the World Summit on Sustainable Development (WSSD) in 2002 provided an international platform for three SIMAR presentations. Members of the SIMAR team also used opportunities during visits to Lesotho, Botswana, Mozambique, Burkina Faso and the Kenya Institute for Meteorological Training and Research as well as the Drought Monitoring Centre to present the progress within SIMAR. The potential agricultural applications of SIMAR products were presented by members of the research team at a workshop organised by an agricultural service provider.

An important means of relatively quick dissemination of the ideas that are the outcomes of research are via presentations at conferences and Symposia. Such presentations at National and International Fora include the following:

National:

1. Burger R.P., P.J.M. Visser, K.P.J. de Waal and D.E. Terblanche (2002). Convective Storm Climatology over the South African Interior. Annual Conference of the South African Society for Atmospheric Sciences. Pretoria, 2002.
2. Deyzel I.T.H. (2002). Application of Satellite Data in Estimating Surface Rainfall. Annual Conference of the South African Society for Atmospheric Sciences. Pretoria, 2002.
3. Kroese N.J., J.N.G. Swart and A.J. Lourens (2002). The Implementation of a real-time reporting raingauge network in South Africa. Annual Conference of the South African Society for Atmospheric Sciences. Pretoria, 2002.
4. Visser P.J.M, J.A. Blackie and S. Boersma (2002). Quality Control and Product Development for the National Weather Radar Network. Annual Conference of the South African Society for Atmospheric Sciences. Pretoria, 2002.
5. Fernandes L. and L. Dyson (2003) Comparison between SIMAR Rainfall and MM5 Rainfall Prognosis for the Rainfall of March 2003. Annual Conference of the South African Society for Atmospheric Sciences. Pretoria, 2003.
6. Kroese N.J. (2003). Meteosat Second Generation (MSG) and its application in South Africa. Annual Conference of the South African Society for Atmospheric Sciences. Pretoria, 2003.
7. Visser P.J.M. (2003). The Detection and Removal of Ground Clutter by Auto-Correlating Volume Scanned Radar Reflectivity Fields. Annual Conference of the South African Society for Atmospheric Sciences. Pretoria, 2003.
8. Nhlapo A.L. (2003). Weather Radar Reliability (*Poster presentation).* Annual Conference of the South African Society for Atmospheric Sciences. Pretoria, 2003.

International:

1. Seed A.W. and G.G.S. Pegram (2001). Using Kriging to Infill Gaps in Radar Data due to Ground Clutter in Real-Time. *Fifth International Symposium on Hydrologic Applications of Weather Radar - Radar Hydrology,* Kyoto, Japan, November.

2. Pegram, G.G.S.and Seed, A.W., (2002). 3-Dimensional Kriging using FFT to Infill Radar Data. Presentation at *27th EGS Assembly*, Nice, France. April.
3. Pegram, G.G.S., Seed, A.W. and Sinclair, D.S. (2002). Comparison of Methods of Short-Term Rainfield Nowcasting. Presentation at *27th EGS Assembly*, Nice, France. April.
4. Sinclair, D.S., Ehret, U., Bardossy, A and Pegram, G.G.S., (2003). Comparison of Conditional and Bayesian Methods of Merging Radar & Raingauge Estimates of Rainfields, Presentation at *EGS - AGU - EUG Joint Assembly*, Nice, France, April.

In yet other ways, SIMAR benefited substantially from international exchanges of knowledge. Initiatives to present data and results led to fruitful discussions and the pursuit of new ideas. In particular, Professor Geoff Pegram was active in fostering Australian and European links, as marked by the following personal invitations:

- 1999 - present : Invited to collaborate with the Australian Cooperative Research Centre for Catchment Hydrology
- 2001 - Mieyegunyah Distinguished Fellow Awardee, Melbourne University - Visiting Research Fellow (12 weeks)
- 2002, 2003 & 2004 - Visiting Research Fellow - Civil and Environmental Engineering Department - University of Melbourne - (8 weeks)
- 2002 - Keynote Speaker: 27th Hydrology and Water Resources Symposium, Melbourne, 20-23 May.
- 2003 - Invited to participate as rapporteur (and future full member of Steering committee) in European Union project: MUSIC / CARPE DIEM Joint Workshop with End Users, at Düsseldorf-Neuss, Germany, May 27 and 28, 2003: “Current Flood Forecasting Practices In Europe”

The knowledge gained by these interactions has benefited not only the participants in SIMAR but has already realized its potential to benefit the post-graduate students working on on-going projects which are out-growths of the Water Research Commission’s investment in SIMAR.

ACKNOWLEDGEMENTS

The members of the Steering Committee for this project were:

Dr GC Green	:	Water Research Commission - Chairman
Dr DE Terblanche	:	SA Weather Services
Mr E Poolman	:	SA Weather Services
Mr K Estié	:	SA Weather Services
Mr S van Biljon	:	Department of Water Affairs and Forestry
Mr DB du Plessis	:	Department of Water Affairs and Forestry
Mr JC Perkins	:	Department of Water Affairs and Forestry
Prof GGS Pegram	:	University of Natal
Dr J C Smithers	:	University of Natal
Dr C Turner	:	Eskom
Prof S Walker	:	University of the Orange Free State
Dr D Sakulski	:	National Disaster Management Committee
Ms B van Wyk	:	Rand Water
Mr M Summerton	:	Umgeni Water

I am deeply indebted to the Steering Committee for making this study so successful. In particular, I want to single out the chairman, Dr George Green, for his contribution. The Water Research Commission's vision in appointing experts and interested parties to the steering committees, which guide, advise and monitor the researchers in their endeavours, has borne excellent fruit in this research. In addition, the interest shown by the potential end-users (here represented by DWAF) has sharpened the focus of the research in providing an end-product which will be useful, not just another academic curiosity.

The team of SAWS:METSYS in Bethlehem was always available for discussion (some of it heated) which ensured that the esoterics did not become stratospheric, but were rooted in practical applications of Hydrometeorology.

My thanks go to Dr Alan Seed of the Australian Bureau of Meteorology, whose continued collaboration in the applications of weather radar spanning 14 years has been a fruitful inspiration.

My thanks go to the Board of the Water Research Commission for providing the funding that made this applied research possible. It has enabled considerable collaboration between those who are normally disparate individuals in traditionally compartmentalised organisations.

Geoff Pegram 3 December 2003

TABLE OF CONTENTSTHE INTERNATIONAL START SECRETARIAT
2000 FLORIDA AVENUE, NW SUITE 200
WASHINGTON, DC 20009

1. INTRODUCTION

Motivation for the research project

Water resources in South Africa are not well buffered against natural rainfall variability. Rainfall deficits and excesses readily translate to droughts and floods, respectively. Well-developed water resource infrastructure, which South Africa is fortunate to possess, has to be managed extremely skilfully to successfully balance water surpluses and deficits at an inter-catchment level, as well as to achieve best trade-offs between flood mitigation and storage maximisation at basin level.

The concept of water resource management is no longer restricted to regulating flows, storage and abstractions in and from rivers, dams and aquifers. Water resource management is increasingly becoming concerned with activities in the catchment which impact on water availability. Such activities are termed streamflow reduction activities in the new Water Act, which also makes provision for their licensing and regulation. To do this objectively will require defendable information on water usage associated with entities such as forests, agricultural lands, natural veld, farm dams, soil conservation schemes, etc. Such water usage is tightly linked, through the catchment water balance, to catchment water availability and thus rainfall, a link which imposes an obligation on catchment management agencies to obtain detailed and accurate rainfall measurements.

Raingauges have traditionally provided the rainfall measurements required for water resource management purposes. Because the national raingauge network is rapidly becoming too sparse to meet existing and anticipated management requirements, a new rainfall monitoring/information system, incorporating the optimal use of remote sensing, needs to be designed and implemented in South Africa.

This need was envisaged to be best satisfied through an umbrella research and development programme of which this proposed project constituted a component. The fundamental aim of this research was to merge the satellite/radar/gauge data to produce ONE field that is acceptable to the Water Resources (and hence Agriculture and Disaster Management) users.

The primary aim was to produce a daily rainfall map of 24 hour accumulated rainfall to a resolution of one minute of arc, over the whole subcontinent, accessible on the Internet, which has been accomplished; this can and will be refined where needed to finer time scales over selected areas. This refinement will be of particular interest to the Disaster Management users and those involved with the mapping of the tendency to, or forecasting of, flash floods. As improvements to the data stream and the modelling techniques become available, they will be incorporated into the products emanating from the research. A new Water Research Commission project in 2003 (K5/1245/1) is designed to do just that.

The aims of this project as set out in the original research proposal were:

1. There is a large collection of daily read rainfall data in the country. These data continue to be read and recorded. The first aim of the research is to be able to interpolate optimal rainfields between raingauges at individual locations and also to suggest the best estimates of catchment totals of rainfall, both historically and currently. This will be complimented by work being done using radar and satellite data, but is a very important element of the interpolation work.

2. The accuracy of radar in pinpointing where rain is falling in considerable detail is without question. There are some difficulties that still exist in terms of estimating rain rates from radar. The second aim is to combine raingauge and radar data into a meaningful composite to provide an optimal rainfield acceptable to users. However, radars only cover part of the country. There are less densely populated areas where there is sparse raingauge coverage and no radars but where we have satellite surveillance. For the purpose of monitoring rainfall in these areas it is important to link satellite, radar and gauge data together to obtain the best estimate of rainfall in these remote areas.

3. The third and main aim of the research is to devise a product to enable the publication, on a daily basis, of the 24 hour rainfall over the country. This can be achieved by optimal use of gauge radar and satellite estimated data, which will of necessity be improved as more radars are commissioned with the updated BPRP devised software and as the new generation of satellites becomes more informative.

Envisaged Methodology

Raingauges measure rainfall effectively at a point. They are good at measuring accumulations of rain but are poor at giving an idea of the spatial distribution of rain, especially over short durations (from an hour to a day). By contrast, weather radars are very good at indicating where it is raining but there are some difficulties in getting the amount of rain correct. Satellites give wide coverage where there are sparse raingauges and no radars which makes their use in estimating rainfall a tempting option. Currently, METEOSAT-7 scans South Africa at 30 minute intervals and the information (visible and infrared) is weakly linked to rainfall amounts - the frequency and band-width is due to increase shortly. To get the precision of raingauge data with the coverage of satellite data and the detail of radar data is the aim of this research. The techniques to be used to achieve this are optimal spatial interpolation using a technique called Kriging and an associated one called co-Kriging. The mathematics of Kriging are well known; it is the application that is difficult.

A brief outline of the theoretical development

It turned out that co-Kriging was not a good option because of the large computational load. This load comes from the fact that there are of the order of one million small areas approximately 1.5 kilometers square covering the subcontinent and the surrounding oceans. The challenge was to be able to map the country's rainfall to that detail routinely. Even with the quadrupling of computer speed between 2000 (when the project was mooted) and 2003 at its end, there had to be a better way which was easy to automate. A method of Kriging exploiting the efficiency of the Fast Fourier Transform was developed by the project leader and is outlined in this report.

This technique enabled Izak Deysel of METSYS to code an efficient algorithm to accomplish the computing efficiently and speedily. Speed is of the essence in the delivery of the daily rainfall maps in real time. Information from the recording raingauges, of the accumulated rainfall for 24 hours until 8:00 am SA time around the country, arrives at METSYS by 9:00 am. By that time, the previous 24 hours' satellite and radar images have been used to produce their best respective estimates of the rainfall totals in detail for the whole area. The merging of the three fields: gauge, radar and satellite is then done and the result posted on the METSYS web-site by 11:30 am. A thorough description of the practical implementation of this methodology is presented in the companion report K5/1152b. Some examples of the website output will be reproduced herein.

Structure of the report

The report comprises 3 sections after this introduction:

- theoretical development of the Kriging methodology and merging algorithm
- some practical implementations in 2-dimensional and 3-dimensional situations
- the way forward - although not part of the original brief, these ideas are to be carried into the next project

2 KRIGING AND VARIANCE FIELD ESTIMATION USING THE FAST FOURIER TRANSFORM – THE LARGE DATA MAPPING PROBLEM

Preamble

In conventional applications of Kriging there are a manageable number of points where information is available (usually between 100 and 2000 say) and the solution of the sets of linear equations with this number of variables is quite feasible.

In the applications envisaged here, the number of points where information is available can vary between 10 000 and 100 000. Solution of sets of equations with this number of variables was thought to be beyond current computer ability, so an alternative was sought. The FFT was suggested and tried out on a few theoretical examples with success. The practical trials will be described in Section 3. The remainder of this section will outline the new developments.

2.1 INTRODUCTION

Among the many candidate methods for fitting surfaces to rainfield data the most promising are Kriging and co-Kriging. The primary objective of Kriging as in other techniques of surface fitting (see Cressie, 1991:356-379, for a comparative summary in a unified notation) is to optimally extend information from "known" values observed in a random field to regions where observations were not made. This operation can be thought of as interpolation (smoothing) between data points scattered in the region of interest or as extrapolation (forecasting) from clustered observations into the rest of the field. Kriging brings two added bonuses: first, the estimated fields are bounded and converge to the mean in extrapolation (unlike polynomial and multiquadric surfaces) and second, a measure of the accuracy of the estimate of the field is provided by the Kriging variance, which is routinely calculated with the surface fitting.

Kriging has become a routinely used tool in the analysis of spatial data and literature has accumulated over the last three decades. The work by Journel and Huijbregts, (1978) is one of the standard geostatistical reference works. In hydro-geological applications, some informative articles stand out. Ahmed and de Marsily, (1987) present a useful development of the subject in ground-water applications where they compare four Kriging methods: (ordinary) Kriging combined with linear regression, external drift and a guess field, and co-Kriging. Krajewski (1987) used co-Kriging to merge a radar and a rain gage field generated by contaminating GATE data, to provide an improved estimate of the underlying rainfield (the GATE data were set up as the "truth"). Seo et al. (1990) compared variants of co-Kriging (simple, ordinary, universal and disjunctive) in two simulation experiments designed to generate Kriged fields to compare them to generated fields with a view to assessing their effectiveness; in their conclusions they noted that the simulations and comparisons were computationally burdensome. This was possibly due to the large number of repetitive calculations required to perform the numerical convolution in the space domain, a problem to which a solution is presented herein.

When manipulating large spatial data sets, it is often the case that Fourier based methods have application, however it appears that there has been very little use made of Fourier transform techniques in the hydrological literature when applied to surface fitting, Kriging in particular. Applications include a paper by Mueller (1986) who examined the spacing of data in a one-dimensional application of Kriging and computed the covariogram using Fourier methods. Schertzer and Lovejoy (1987) used Fourier transforms to generate sample rainfields with covariance structures estimated from radar data, but there was no hint of Kriging there; the article is cited because it reports a useful method for simulating random fields. Robin et al. (1993) introduced a random field simulation method similar to that of Schertzer and Lovejoy (1987) to groundwater applications for the purpose of co-generating cross-correlated random fields. Menabde et al. (1997) used Fourier methods to generate self-similar rainfields by power-law smoothing white noise fields in wave-number (Fourier) space before inverse transformation. Lanza (2000) used Fourier methods to convolve synthetic rainfields conditionally with point rainfall observations but did not use Kriging. In none of the above cited applications was Kriging performed by Fourier methods, however Zimmerman et al. (1998) compared seven geostatistical methods, including the Fast Fourier Transform (FFT), in solving the inverse problem as presented in four simulated data sets. The FFT method used was an outgrowth of the work reported by Robin et al. (1993) but appears to have been used in conditional simulation mode for confidence limit estimations.

Kriging can be viewed as being composed of two separate operations, Constrained Deconvolution and Direct Convolution. This article demonstrates how these calculations can be achieved using Fast Fourier Transform (FFT) methods, which become an attractive alternative once the data sets become large ($> 10^4$). An approximate (but good, fast and economical) method of obtaining the Explained Variance (the complement of the Kriging Variance to the Field Variance) is presented, which gives an objective assessment of the accuracy of the estimated field. The Explained variance also provides a plausible way of merging two or more Kriged fields; an example is presented.

The crux of the FFT-based methods of obtaining the Kriged field and the Explained Variance field is the Iterative Constrained Deconvolution algorithm (ICD). ICD is used to obtain what are herein called *reciprocal* data, which are the linearly transformed (deconvoluted) values of the data. Reciprocal data are used as the raw material for Direct Convolution to obtain the Kriged field, as well as recovering (collocating) the observed values. Thus, once the reciprocal data have been derived, the optimal field can be obtained easily by conventional numerical convolution or alternatively by FFT filtering. The links between the matrix and FFT based methods are developed and presented in Section 4.1. An algorithmic summary of the methods is presented in Section 5.

2.2 DATA REPRESENTATION

In a region D, the n_1 observations on the underlying random field are labeled z_{1i} : i = 1, 2, …, n_1, collected into a vector z_1. These data may be scattered sparsely or clumped together compactly in a data region D_1 within the larger region D as shown in Figure 2.1. It is desired to estimate the unobserved random field z_2 in the complementary field D_2 = D-D_1 at n_2 locations. The observations in z_1 are assumed to have been made at points in the 2-dimensional region D_1; the coordinates of the observations are $s_i = (x_i,y_i)$. Thus, in various equivalent notations:

$z_1 = \{z_{1i}\} = \{z_1(s_i)\} = \{z_1(x_i,y_i)\}$, $\{s_i \in D_1: i = 1, 2, ... , n_1\}$.

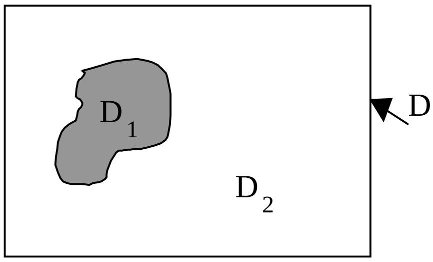


Figure 2.1. Partition of region D into subregions: D_1 in which data were observed, D_2 in which data were not observed.

The expected value of the random field (in Kriging, the drift) at the location s is m(s) which is assumed to be known (Simple Kriging) or to be unknown (Ordinary Kriging).

In both Simple and Ordinary Kriging it is conventional to calculate weights λ_{ji} by which to multiply the z_{1i} in D_1 in order to give optimal (linear, minimum variance) estimates z_{2j} at $s_j \in D_2$: j = 1, 2, ... , n_2. Thus in Simple Kriging:

$$z_{2j} = m_2(s_j) + \sum_{i=1}^{n_1} \lambda_{ji}[z_{1i} - m_1(s_i)] \ , \quad j = 1, 2, \ldots, n_2. \qquad (2.1a)$$

or in vector notation:

$$z_2 = m_2 + \Lambda_{12}^T[z_1 - m_1] \qquad (2.1b)$$

where Λ_{12} is a matrix (n_1 x n_2) of weights. In Ordinary Kriging, m_1 and m_2 are omitted in equation (2.1) and the resulting weights are numerically different from those calculated in Simple Kriging, compensating for the unknown mean. In this article (unlike much of the literature on Kriging) the weights λ_{ji} are explicitly linked to the target location s_j.

Covariogram

In this article, the spatial dependence within the random field (and hence between the data) will be assumed to be described by an isotropic covariogram (correlation or covariance function), rather than variogram function. This choice is a matter of taste, not of necessity if the field can be assumed to be stationary. In Fourier Transform based methods of analysis of dense data-sets, the two-dimensional spectrum (which is the Fourier transform of the 2-D covariogram) is relatively easily determined from the data, when for example the data are spatial rainfall fields sampled by radar. In sparse data cases, such as where sets of rainfall data are measured from a network of rain gages, a covariogram model is usually fitted to, or estimated from, the point data directly. In the application of the methods appearing in this article, the covariogram of the rain gage data is assumed to be that estimated from the dense data-set obtained from the radar measurements. The technical details of obtaining the 2-D covariogram using FFT methods appears in the Press et al. (1992), where the necessary conditions for obtaining good covariogram estimates are given.

In any case, the covariogram has to be positive definite to ensure that the variance of the modeled field is non-negative (Cressie, 1991:84). Where spatial rainfall is measured by radar, the way the covariogram is estimated herein guarantees its positive-definiteness. This follows because it is usual, in spatial rainfall studies, to assume that the power spectrum is an isotropic function which can be radially averaged and modeled by a power-law curve (Menabde et al., 1997). This power spectrum is bounded and positive at all frequencies; as a result, its Fourier transform, the covariogram, is guaranteed to be positive-definite, by Bochner's theorem (Cressie, loc.cit.).

In summary, the covariogram is here assumed to be univariate, isotropic, homogeneous and positive-definite:

$\mathrm{cov}[z_k, z_l] = \mathrm{cov}[z(s_k), z(s_l)] = g(|s_k - s_l|)$

where $|s_k - s_l|$ is the distance s between the points s_k and s_l.

Estimation of 2-dimensional Covariograms by Fourier Transform.

There are practical watch-points to be kept in mind when computing a covariance function from data using transform methods. This sub-section sets out some of the ways to be sure that one gets what is needed from the computation.

The theory relating the power spectrum and the covariance function is well understood (see, for example, Jenkins and Watts (1968) in time series applications and Press et al. (1992) in both statistical and engineering applications), so will be stated where required. The practical implications of the theory are not always understood, therefore some definitions of terms are required as a preamble. The treatment will start in one dimension before moving on to two dimensions with an obvious generalization of notation.

Commencing in one dimension, given a discrete sample $x(s) = \{x(s_j) = x_j : j = 1, 2, \ldots, n\}$ from a stochastic process $\{X(t)\}$, the temporal covariance (defining the covariogram) is given by:

$$c(l) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_{i+l} - \bar{x})$$

where $\bar{x}$ is the sample mean of x(s) as usual. c(0) is then the sample variance, and immediately, the sample correlation coefficients (defining the correlogram) are obtained as: $r(l) = c(l)/c(0)$, where it is conventional to limit l to n/3 or at most n/2.

It is mathematically correct to derive the variance as $c(0) = \frac{1}{n}\sum_{i=1}^{n} x_i^{\,2} - \bar{x}^2$ (the mean of the squares minus the square of the means) but for numerical reasons it is preferred to use $c(0) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$. The former is often used because it is less trouble to compute and herein lies a potential problem with deriving the correlogram by Fourier Transforms. There is another hidden problem - aliasing - which requires padding the sample with zeros to obtain a correct result. The solution to these problems will be presented in the development which follows.

Mathematically, one typically writes down that the Fourier Transform of x(s) is

$X(u) = F[x(s)]$ (where u is the frequency or wave number) and then obtains the sample power spectrum as $p(u) = |X(u)|^2$. The literature typically gives the following relationship for deriving the covariance function from the spectrum: $c(l)^* = F^{-1}[p(u)]$, (where $F^{-1}[.]$ denotes the inverse Fourier transform) implying that the covariance function is the inverse Fourier transform of the power spectrum. What is often not stressed in the discussion of this operation is that $c(l)^*$ is not equal to $c(l)$ unless x(s) has been pretreated in a specific way.

First, x(s) must have the sample mean subtracted from each element to yield $y(s) = x(s) - \bar{x}$ and then second, it should be padded with a string of zeros to increase the sample's length by at least double to N, which for compatibility with most FFT algorithms, should be a power of 2. Only if these steps are taken will $c(l)^*$ be equal to $c(l)$. There is one other numerical problem to attend to. Form $y(s) = x(s) - \bar{x}$, padded with zeros to length N as recommended, then $p(u) = | F[y(s)]|^2$. The result is that the variance is (theoretically) equal to the integral of p(u), however many programs for computing the FFT do not scale by N, leaving the value $p(0) = Y(0)^2$ equal to the sum of squares of the values of y(s) multiplied by a factor equal (in this case) to nN.

This number nN may be a nuisance to keep track of, so the most convenient work-around, which is not computationally costly, is to scale y(s) by the sample standard deviation computed from x(s) to yield: $z(s) = y(s)/c(0)^{1/2}$. The vector resulting from the calculation $c(s)^* = F^{-1}\{|F[z(s)]|^2\}$ should then be scaled by the value $c(0)^*$, to yield the correlation function r(s), which folds about the Nyquist frequency N/2. The covariance can immediately be obtained from r(s) by rescaling using c(0), if desired.

Deriving the two dimensional covariogram of a set of spatial data in a compact region D_1 containing n points x(s) follows the same procedure as for one dimension, but with two caveats. The first is that the minimum dimension of D_1 should be greater than the distance at which the correlation is negligible: l_{crit}. Second, the sample mean and variance of the data x(s) are obtained and the data are scaled to $z(s) = [x(s)-\bar{x}]/c(0)^{1/2}$. The region D_1 is padded with zeros by at least a distance l_{crit} so that the dimensions of the augmented rectangle D are powers of two. The two dimensional spectrum p(u) is then calculated by FFT. It should be noted that this procedure is inappropriate for use with sparse data like rain gauges because the FFT will treat the pixels in the region $D-D_1$ (where there are no data) as zeros. These zeros contaminate and downgrade the FFT calculation and although not impossible to remove, make it almost mandatory that the covariance function of sparse data be computed in the usual manner by direct enumeration.

The two dimensional spectrum p(u) is then calculated from z(s) by FFT. At this stage there are three paths to deriving the 2-D correlogram, and the decision as to which path to follow hinges on the assumption of isotropy; (homogeneity is tacitly assumed in Fourier transform methods.) For the purpose of this discussion isotropy will be assumed. This is reasonable in most instantaneous images of rainfields such as those obtained by satellite and radar, but is violated in accumulations of these fields into hourly or daily estimates of rain depth, because of the effects of advection.

In the isotropic field case the 1-D correlogram can be obtained by two nearly equivalent but computationally different means. The first is to obtain the column and row averages of the 2-D spectrum to give the best estimate (in the sense of accuracy and speed) of the 1-D spectrum. If the domain D is not square, then these should be treated separately to derive the correlogram, alternatively the column and row totals can themselves be averaged to one composite 1-D spectrum. This average composite 1-D spectrum is inverse transformed to yield the correlogram, as explained in the discussion above - an example appears in Figure 6(a) later in the text. If the domain D is not square, the row and column sums of the 2-D spectrum are individually inverse transformed to correlograms which are finally averaged out to a reasonable distance s.

The second way to estimate the 1-D correlogram via the 2-D spectrum is from the inverse of the latter. After obtaining the 2-D correlogram it can be radially averaged to obtain the 1-D correlogram under the assumption of isotropy - an example appears in Figure 2.6b in the text. The difference between the 1-D averaged spectrum inverse and the 2-D radial correlogram estimates is that the radial estimate has few data points near the origin where an accurate estimate of the correlation function is required and a large number of data for large lags where the estimate is relatively unimportant. By contrast the 1-D correlations obtained from the 1-D spectra averaged from the 2-D spectrum possess equal amounts of data at each lag. (A warning: it is not correct to radially average the spectrum and transform the result. The dimension of the radially averaged spectrum is too large by 1 Vanmarcke (1983)).

A final polishing step in the estimation of the 1-D correlation is to fit a model to the 1-D correlogram to remove the noise and to ensure monotonicity and non-negativity. A useful candidate is the generalized exponential function $r(s) = \exp[-(s/L)^a]$ with a and L > 0, where L is the correlation length and the exponent a = 1 for exponential, or 2 for Gausian models of correlation.

2.3 Obtaining Reciprocal Data from the Observations Using the Matrix Methods of Conventional Kriging

Two types of Kriging will be considered, Simple and Ordinary Kriging, whose equations are well known and will be stated for notational completeness. Reciprocal data is defined and applied to both types.

2.3.1. Simple Kriging

In Simple Kriging it is assumed that the mean surface of z(s) is known and equal to m(s). The Simple Kriging weights in equation (2.1) are derived from:

$$\Gamma_{11}\Lambda_{12} = \Gamma_{12} \qquad (2.2)$$

where Γ_{11} is the (n_1 x n_1) covariance matrix of the observations z_1, Λ_{12} is a matrix of Kriging weights and Γ_{12} is the (n_1 by n_2) covariance matrix between the variables listed in the vectors z_1 and z_2. Thus the weights are:

$$\Lambda_{12} = \Gamma_{11}^{-1}\Gamma_{12} \qquad (2.3)$$

and the vector of estimates z_2 is obtained from equation (2.1b) which can alternatively be written as:

$$z_2 = m_2 + \Gamma_{21}\Gamma_{11}^{-1}[z_1-m_1] \text{ , from equation (2.3)}$$
$$= m_2 + \Gamma_{21}\beta_1 \qquad (2.4a)$$

Reciprocal Data defined. Equation (2.4a) is identically satisfied when substituting z_1 (and m_1) in place of z_2 (and m_2). β_1 is then the vector of *reciprocal data* calculated from an inverse (reciprocal matrix) transform or deconvolution of the observed data as

$$\beta_1 = \Gamma_{11}^{-1}[z_1 - m_1] \,. \qquad (2.4b)$$

β_1 remains constant through the calculation of z_2 in D_2. Equation (2.4a) can be augmented, once β_1 has been determined from (2.4b), to give:

$$[z - m] = \Gamma\beta \qquad (2.5a)$$

where

$$\Gamma = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix} \qquad (2.5b)$$

and $\beta = [\beta^T \; 0^T]^T$.

Equation (2.5a) can be viewed as a discrete data convolution over D.

2.3.2 Ordinary Kriging

In Ordinary Kriging, in contrast to Simple Kriging, it is assumed that the mean m(s) of the random field z(s) is unknown, but constant. The Simple Kriging equations (2.1) and (2.2) are modified by setting m_1 and m_2 to zero and adding the restriction that the weights λ_{ji} for a given s_j in D_2 must sum to unity:

$$\lambda_j^T u = 1 \qquad (2.6a)$$

where u is a vector of ones, so:

$$z_{2j} = \lambda_j^T z_1 \quad , \qquad j = 1, 2, \ldots, n_2 \,. \qquad (2.6b)$$

The set of equations to derive the weights in Ordinary Kriging is an augmented version of equation (2.2):

$$\begin{bmatrix} \Gamma_{11} & u \\ u^T & 0 \end{bmatrix} \begin{bmatrix} \lambda_j \\ \mu_j \end{bmatrix} = \begin{bmatrix} g_j \\ 1 \end{bmatrix} \qquad j = 1, 2, \ldots, n_2 \qquad (2.7)$$

where μ_j is a Lagrange multiplier introduced to satisfy equation (2.6a). The consequences are that in general the Ordinary Kriging weights calculated in equation (2.7) will be different from those for Simple Kriging derived from equation (2.2).

Using the inverse of a partitioned matrix, the vector of Kriging weights from equation (2.7) are:

$$\lambda_j = [I - uu^T\Gamma_{11}^{-1}/(u^T\Gamma_{11}^{-1}u)]g_j + \Gamma_{11}^{-1}u/(u^T\Gamma_{11}^{-1}u) \qquad (2.8a)$$

and

$$\mu_j = (1 - u^T\Gamma_{11}^{-1}g_j)/(u^T\Gamma_{11}^{-1}u) \qquad (2.8b)$$

An associated quantity which will be used later is the Kriging (prediction) variance:

$$\sigma_j^2 = g(0) - g_j^T\Gamma_{11}^{-1}g_j + (1 - u^T\Gamma_{11}^{-1}g_j)^2/(u^T\Gamma_{11}^{-1}u) \qquad (2.9)$$

Equations (2.6) to (2.9) are standard (e.g. Cressie, 1991:122-3); when the estimates of the missing data over large areas are to be calculated, the treatment presented here departs from the conventional.

First, equation (2.8a) can be rewritten as:

$$\lambda_j = Hg_j + f \qquad (2.10a)$$

where H and f are independent of j, a point which turns out to be valuable in the following development. Second, to obtain the estimates of the unobserved data z_2, equations (2.6b), (2.8a) and (2.10a) are combined to give:

$$z_{2j} = g_j^THz_1 + f^Tz_1$$

$$= g_j^T\beta_1 + \delta \ , \qquad j = 1, 2, \dots , n_2 \qquad (2.10b)$$

where the vector β_1 of reciprocal data (different from β_1 in equation (2.4b)) and the scalar δ are independent of j. From equations (2.8a), (2.10a) and (2.10b), δ is given by:

$$\delta = f^Tz_1 = (u^T\Gamma_{11}^{-1}z_1)/(u^T\Gamma_{11}^{-1}u) \qquad (2.10c)$$

Equation (2.10b) must be satisfied when it is written in terms of the observed data in D_1:
$z_1 - \delta u = \Gamma_{11}\beta_1$,
which yields the equation for the calculation of the *Reciprocal Data* in Ordinary Kriging:
$$\beta_1 = \Gamma_{11}^{-1}[z_1 - \delta u] \qquad (2.10d)$$

This result yields the unifying simplification that the reciprocal data in the Ordinary Kriging case are calculated using the inverse of the covariance matrix and the observations, by the same formulation as the Simple Kriging case given by equation (2.4b), with the difference that m_1 (the vector of the means in D_1) in that equation is replaced by δu. It turns out that δ is very nearly equal to (and converges to) $(z_1^Tu)/n_1$, the sample mean of the observations, when n_1 is moderately large. This assertion is justified in the following argument.

The constant shift δ in Ordinary Kriging converges to the sample mean. Write $\Gamma_{11} = \sigma_z^2R$ where R is the correlation matrix. An autoregressive relationship between the data (Kriging in one dimension with equally spaced data) is based on a geometric variogram $g(s) = r^s$, so:

$$R = \begin{bmatrix} 1 & r & r^2 & .. & r^{n_1-1} \\ r & 1 & r & .. & .. \\ r^2 & r & 1 & .. & .. \\ .. & .. & .. & .. & .. \\ r^{n_1-1} & .. & .. & .. & 1 \end{bmatrix} \quad , \text{ a Toeplitz matrix.}$$

R will describe mutually independent data if $r \to 0$ (if the distances between the data points are large relative to the correlation length L : $r^L = e^{-1}$) and, conversely, near complete dependence as $r \to 1$. R's inverse is a tridiagonal matrix:

$$R^{-1} = \begin{bmatrix} 1 & -r & 0 & .. & 0 \\ -r & 1+r^2 & -r & .. & .. \\ 0 & -r & 1+r^2 & .. & .. \\ .. & .. & .. & .. & .. \\ 0 & .. & .. & .. & 1 \end{bmatrix} 1/(1-r^2)$$

whose columns (except for the first and last) are composed of elements which sum to a constant, $(1-r)/(1+r)$, so that $u^T R^{-1} u \to n_1(1-r)/(1+r)$ as n_1 becomes large and $u^T R^{-1} z_1 \to u^T z_1 (1-r)/(1+r)$. Thus from equation (2.10c), $\delta \to (u^T R^{-1} z_1)/(u^T R u) = (u^T z_1)/n_1$ for this inverse matrix, for large n_1, independent of r. This result will not be materially affected by random scattering of the data points in two dimensions. In Section 3.3, an example where n_1 is small shows that $\delta = (u^T z_1)/n_1$ with good approximation, as asserted.

Substituting $(u^T z_1)u/n_1$ for δu in equation (2.10d) yields the formula for obtaining the approximate reciprocal data set in Ordinary Kriging:

$$\beta_1 = \Gamma_{11}^{-1}[z_1 - (u^T z_1)u/n_1] \qquad (2.11)$$

This equation is the same as equation (2.4b) except that the known mean m_1 is replaced, not by δu as in equation (2.10d), but by the sample mean. Note that the intermediate step of computing the weights λ_j has disappeared in this scheme - the emphasis is now on the computation of the reciprocal data set β_1. This reduces complexity of the calculations and equations (2.4b), (2.10d) and (2.11) unify the notation of Simple and Ordinary Kriging. If the data sets are "small", then there is no need for the approximation of equation (2.11) and equation (2.10d) can be used to give the exact result. On a practical note, if m_1 in Simple Kriging is estimated by the sample mean, then it yields an estimated field identical to that calculated by Ordinary Kriging based on the approximate reciprocal data calculated from equation (2.11).

2.3.3 Example comparing approximate to exact Ordinary Kriging:

A small, one-dimensional (n_1=3, n_2=6) comparison between the exact and approximate equations in Ordinary Kriging is offered. The data are $z_1 = (8, 6, 1)^T$ at the points s = (0, 1, 2) and the covariogram is g(s) = exp(-s). The correct value of δ is 4.800; the approximation to δ is $u^T z_1/n_1 = 5$. Using equations (2.10b), (2.10d) and (2.11), the comparison is presented in Figure 2.2 for the case where data are grouped and estimates are extrapolated (forecast) using numerical convolution in both calculations.

This comparison shows how closely equation (2.10d) is approximated by equation (2.11), even for small n_1, providing a visual/numerical corroboration of the analysis leading to the latter equation, incidentally highlighting the fact that Kriging can be used as a forecaster/extrapolator. An attractive advantage that Kriging has over multiquadric and polynomial fitting procedures (which grow without bound in extrapolation) is that Kriging produces "safe" extrapolations, asymptotic to m_1 (or δ) in Simple (or Ordinary) Kriging.

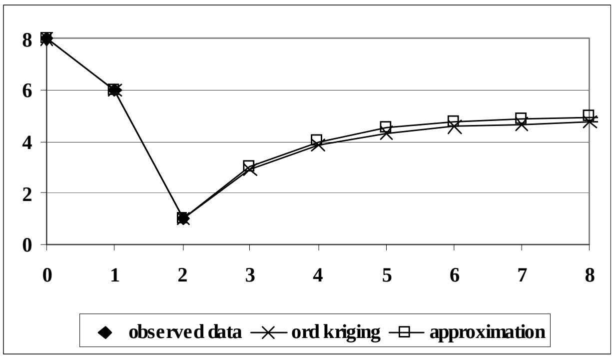


Figure **2.2**. Comparison between Ordinary Kriging using equations (2.10) and the approximation using equation (2.11) for a small-dimensioned (n_1=3), 1-dimensional example. The "observed" data, $z_1 = (8, 6, 1)^T$ in D_1, are at s = 0, 1and 2.

2.3.4 The Explained Variance of the estimates

A by-product of Kriging is the predictive (or Kriging) variance given by equation (2.9). Obtaining these estimates with large data sets is even more computationally demanding than numerical convolution. It requires the computation of a quadratic form of dimension n_1 at each of n_2 points in D_2. This section presents the derivation of an approximation to the variance field that is relatively good and economical and can be exploited in FFT methods.

Equation (2.9), evaluated at each point s_j in D_2, is repeated here in modified form:

$$\sigma_j^2 = \sigma_z^2 - \Delta_j + \varepsilon_j \qquad \text{for } j = 1, 2, \ldots, n_2 \,; \qquad (2.12)$$

where

- the first term on the right, the field variance σ_z^2 , equals g(0)
- the second term, $\Delta_j = g_j^T\Gamma_{11}^{-1}g_j$, is a quadratic form
- the third term, ε_j, equals $(1\text{-}u^T\Gamma_{11}^{-1}g_j)^2/(u^T\Gamma_{11}^{-1}u)$.

The *Explained Variance* $\sigma_{\Delta j}^2$ at s_j is defined as the complement of the Kriging variance σ_j^2 to the Field variance σ_z^2, i.e.

$$\sigma_{\Delta j}^2 = \sigma_z^2 - \sigma_j^2 = \Delta_j - \varepsilon_j, \qquad (2.13)$$

which tends to Δ_j , as $n_1 \to\infty$, as the following development shows.

Equation (2.12) must apply for the set of observations at s_i in D_1 as well as s_j in D_2. In that case the Kriging variance σ_i^2 equals zero identically. This is because,

- $\Delta_i = g_i^T\Gamma_{11}^{-1}g_i = g_i^Te_i$ (noting that g_i is the i^{th} vector in Γ_{11} and e_i is a vector with a one at the i^{th} position, otherwise null) so $\Delta_i \equiv \sigma_z^2$
- the numerator of $\varepsilon_i = (1 - u^T\Gamma_{11}^{-1}g_i) = (1 - u^Te_i) = (1\text{-}1) = 0$, because in D_1, $\Gamma_{11}^{-1}g_i = e_{i,}$ so that $\varepsilon_i \equiv 0$.

This proves the assertion. It remains to examine the behavior of the second and third terms of equation (2.12) when s is not in D_1 but in D_2. The third term of equation (2.12), which is $\varepsilon_j = (1-u^T\Gamma_{11}^{-1}g_j)^2/ (u^T\Gamma_{11}^{-1}u)$, will be addressed first.

Γ_{11} and g_j can be rewritten as $R\sigma_z^2$ and $r_j\sigma_z^2$ respectively, where R is the correlation matrix and r_j is the correlation vector at s_j obtained from g_j, the covariance vector, by scaling by the field variance σ_z^2; ε_j can thus be written alternatively as

$\varepsilon_j = \sigma_z^2(1-u^TR^{-1}r_j)^2/(u^TR^{-1}u)$.

There are two extremes that ε_j can obtain for a given correlation structure, depending on the distance of the point s_j from the data in D_1. If s_j approaches a point s_i in D_1, then r_j will converge to r_i in which case $R^{-1}r_j \to e_j$ so that $\varepsilon_j \to 0$. If s_j is distant from D_1, r_j will tend to a zero vector, so that ε_j will attain its maximum value, $\sigma_z^2/(u^TR^{-1}u)$. Assuming that is Toeplitz as in the discussion in Section 3.2.1, then in this case, where r is a scalar,

$\max[\varepsilon_j]/\sigma_z^2 = (1+r)/\{(1-r)n_1\}$

so that for high correlations between neighboring points (r near 1) this tends to $2/\{(1-r)n_1\}$ which becomes negligible, for large r, if n_1 is large. e.g. if n_1=20000 and at a given location r = 0.99, then the error in neglecting ε_j relative to σ_z^2 is 1% and can be ignored. Thus, for n_1 large, ε_j can be neglected in equation (2.12) in most of D_2.

It remains to obtain a useful means of computing the second term on the right hand side of equation (2.12), the quadratic term $\Delta_j = g_j^T\Gamma_{11}^{-1}g_j$, which for large n_1 becomes the Explained variance, because then $\varepsilon_j \to 0$, as shown above. What is proposed here is an approximation to Δ_j which (when computed from covariance functions like the exponential and Gaussian) appears to have smaller errors due to the suggested approximation than due to the difference between the covariance models.

The idea is based on the following observation: the elements Δ_j are properly obtained from the calculation of the trace of the matrix product developed from the partitioned matrix defined in equation (2.5b):

$$\Delta = \{\Delta_j\} = \text{trace}\begin{bmatrix}\Gamma_{11} \\ \Gamma_{21}\end{bmatrix}\Gamma_{11}^{-1}\begin{bmatrix}\Gamma_{11} & \Gamma_{12}\end{bmatrix} \qquad (2.14)$$

yielding $\Delta_i = \sigma_z^2$ for s_i in D_1 (as required), and the correct values for Δ_j when s_j is in D_2. In searching for a simplifying relationship, it was discovered that numerical values of Δ_j obtained from "typical" covariograms were approximately equal to the square of g_j, which prompted the following development.

If (trivially) there is only one point in D_1, then equation (2.14) reduces to
$\Delta =$ trace $\{[g_1](\sigma_z^2)^{-1}[g_1]^T\} = \{g_{1i}^2\}/\sigma_z^2$, so that in this special case, Δ_i is in fact the square of the corresponding covariance. The naïve approximation exploits this simple relationship and defines the (approximate) explained variance filter as

$g^{\$}(s) = g^2(s)/\sigma_z^{\ 2}$ (2.15)

The proposal is to use this as a filter to obtain an approximation to Δ_j by convolution:

$\Delta^{\$}(s) = g^{\$}(s){*}\beta^{\$}{}_1\,(s)$ (2.16)

where $\beta^{\$}{}_1$ is a new vector/set of reciprocal variables that ensures that $\Delta^{\$}{}_i = \Delta_i \equiv \sigma_z^{\ 2}$ in D_1 (because there, $\sigma_i^{\ 2} = \varepsilon_i \equiv 0$), or in vector notation $\Delta^{\$}{}_1 = (\sigma_z^{\ 2})u$.

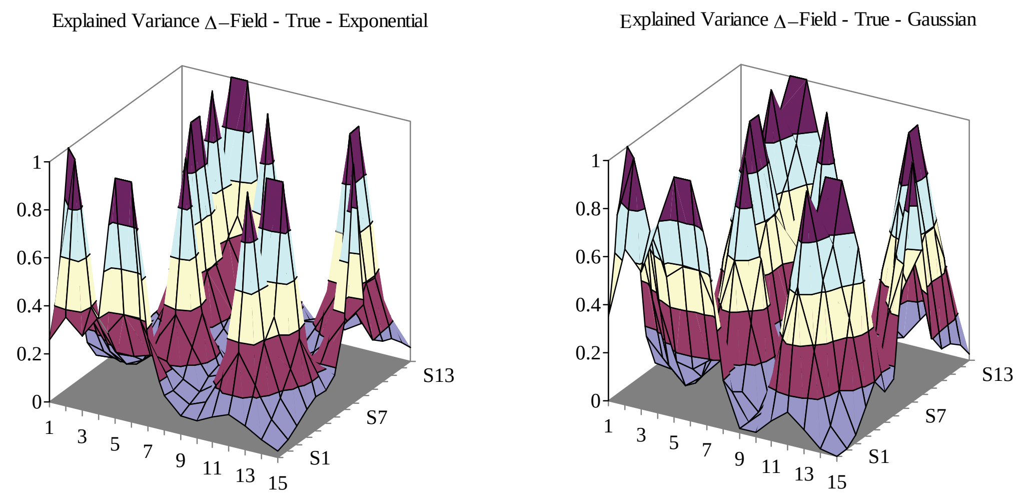
In matrix terms, in D_1, equation (2.16) becomes:

$\Delta^{\$}{}_1 = (\sigma_z^{\ 2})u = G^{\$}{}_{11}\beta^{\$}{}_1$
yielding
$\beta^{\$}{}_1 = \{G^{\$}{}_{11}\}^{-1}u(\sigma_z^{\ 2}),$ (2.17)

where it should be noted that the calculation depends only on the spatial distribution (topology) of the data in D_1 and not on the data values z_1. The notation $\beta^{\$}{}_1$ is chosen deliberately, because equation (2.17) computes the reciprocal vector to the constant vector $(\sigma_z^{\ 2})u$. Methods for solving for the reciprocal data in equations (2.4a), (2.10d) and (2.11) can be exploited to solve equation (2.17) in the large data case.

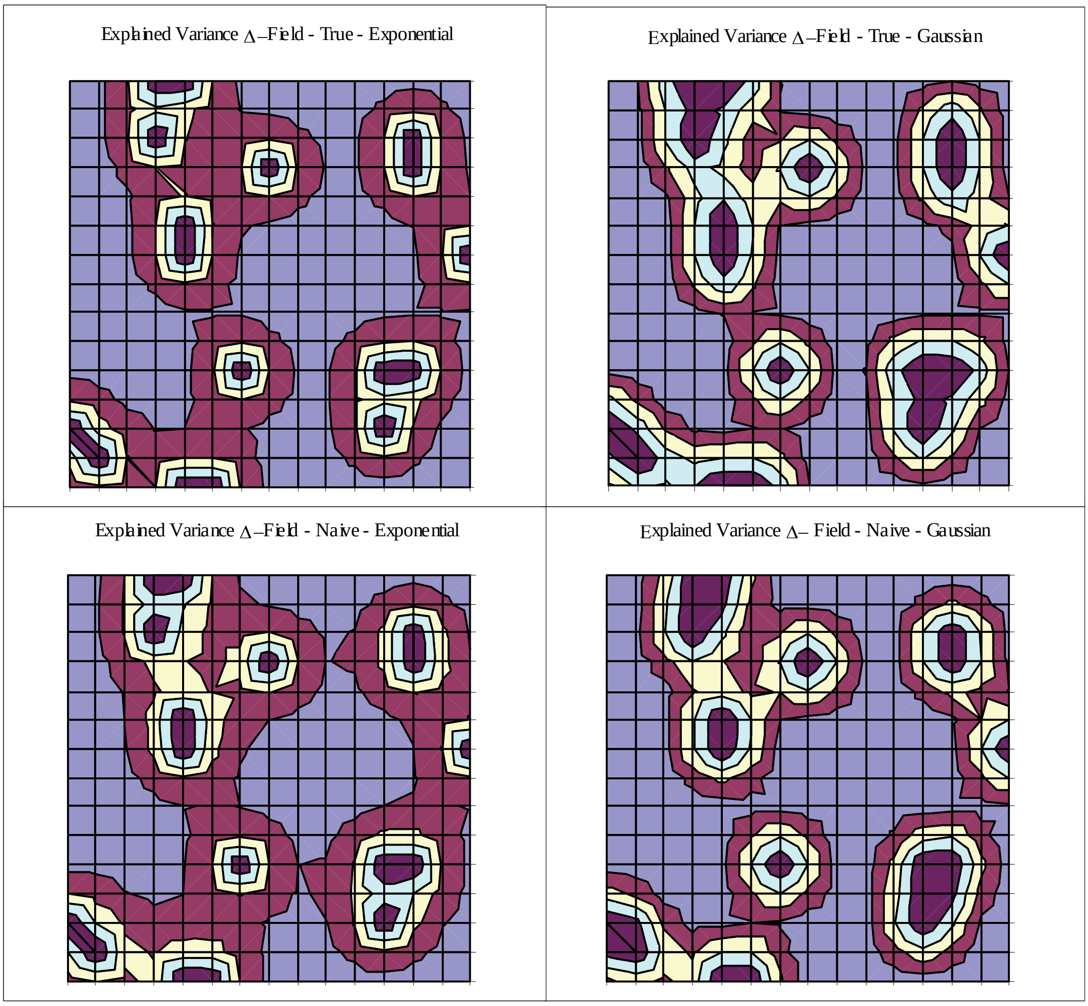
2.3.5 Example to compare methods of computing Explained variance

To visualise the approximation to the Explained variance Δ of equation (2.14) by the naïve approximation $\Delta^{\$}$ using equations (2.15), (2.16) and (2.17), and to be able to distinguish the detail, a small-dimensional numerical experiment was conducted. A few points on a square grid of 15 units were randomly selected. Exponential and Gaussian covariance functions with the same correlation length were used to compare the naïve approximation with the exact calculation. The results of these calculations are presented in Figures 2.3a to 2.3f.



Figures 2.3a and b: True Explained Variance Field computed by equation (2.13).
(a) Exponential Model of covariance (left).
(b) Gaussian Model of covariance (right).

Figures 2.3a and b are the "true" Explained variance fields displayed in relief, calculated using equation (2.13) with the Exponential and Gaussian models of the covariance, respectively. Figures 2.3c and d show these in plan for better comparison. By contrast, Figures 2.3e and f display the Explained variance fields in plan, calculated using the "naïve" approximation given by equation (2.16) with the Exponential and Gaussian models of the covariance, respectively. It is clear that the "naïve" approximation for calculating the Explained variance differs less from the outcome of the "true" calculation procedure than the outcomes of using the two models.



Figures 2.3c to f: Explained Variance Field - plan - two different models of the covariance function computed by equations (2.13) and (2.16).
(c): top left - True - Exponential Model.
(d): top right - True - Gaussian Model.
(e): bottom left - Naïve - Exponential Model.
(f): bottom right - Naïve - Gaussian Model.

This comparison suggests that the naïve approximation is a useful alternative to the true Explained variance field in situations where it would be computationally intractable to calculate the latter.

2.4 ITERATIVE DECONVOLUTION TO OBTAIN THE RECIPROCAL DATA SET

2.4.1 The link between Matrix and FFT methods

To obtain a set of reciprocal data (β_1 in Simple or Ordinary Kriging, or $\beta^{\$}{}_1$ in the calculation of the Explained variance) equations (2.4b), (2.11) or (2.17) need to be solved. These all have the same form. However, the problem of inverting an n_1-square matrix and then calculating values of either z_2 or Δ_2 when n_1 and n_2 are both large can be computationally daunting. The solution of a set of linear equations like (2.17) typically requires of the order of n_1^3 multiplications and a convolution like equation (2.16) requires of the order of $n_2 n_1^2$ multiplications. By contrast, using Fast Fourier Transform (FFT) based methods on D requires of the order of $(n_1+n_2)\log(n_1+n_2)$ multiplications, making the FFT-based methods an attractive option when n_1 and n_2 are both large.

To develop the ideas for obtaining β_1 by Iterative Constrained Deconvolution using FFT filtering, it is helpful to re-pose the problem of Kriging in a unified notation. Recalling equation (2.4b), the reciprocal data are calculated from $\beta_1 = \Gamma_{11}^{-1}[z_1-m_1]$ and from the point of view of the whole region D, the Kriged surface is given by equation (2.5a): $[z-m] = \Gamma\beta$. Note that β_1 cannot be obtained by solving this last equation (2.5a) by inverting Γ because z_2 is as yet unknown and setting it to zero will yield the wrong values in β. The solution needs to be found from the equation above it, (2.4b). For n_1 large, this presents a computational problem, so one seeks another route offered by iterative inverse filtering or Iterative Constrained Deconvolution (ICD).

To lead into the ICD idea for obtaining the reciprocal data, equation (2.5a) can be rewritten as:

$$\alpha[z-m] - \alpha\Gamma\begin{bmatrix}\beta_1\\ 0\end{bmatrix} = 0 \qquad (2.18)$$

for some choice of non-zero scalar α, where β_2 is constrained to zero because the relationship between z_1 and β_1 is desired.

Define $\beta^{(k)} = [\beta_1^T\ \beta_2^T]^{T\,(k)}$ as the k^{th} iteration of the calculation of β (in what follows), then $C\beta^{(k)}$ is the constrained vector $[\beta_1^T\ 0^T]^{T\,(k)}$. Here, in the matrix analogue of ICD, the (n_1+n_2)-square constraint matrix C consists of an identity matrix of size n_1 in the top left corner with the rest null. At convergence, for large enough k, $C\beta_1^{(k+1)} = C\beta_1^{(k)}$. Adding this to both sides of (18):

$$C\beta^{(k+1)} = \alpha[z-m] - \alpha\Gamma\begin{bmatrix}\beta_1\\ 0\end{bmatrix} + C\beta^{(k)}$$

$$= \alpha[z-m] + [I - \alpha\Gamma]C\beta^{(k)} \qquad (2.19)$$

Convergence of $C\beta_1^{(k)}$ is achieved for a suitable choice of the scalar α, which can be found by trial. Equation (2.19) is couched in matrix form but can be translated into filtering notation, enabling the power of the FFT to be exploited in large data-mapping problems.

2.4.2 Computing Reciprocal Data by Iterative Constrained Deconvolution using FFT.

Equation (2.4a) (or equation (2.10b) or (2.16)) can be viewed as a convolution between the reciprocal data β_1 (or $\beta^{\$}{}_1$) computed directly by equation (2.4b) (or (2.11) or (2.17)) and the covariances g(s) (or $g^{\$}$(s)) linking every target point in D_2 with the reciprocal data in D_1. The estimate of z in D as a whole is thus given by the convolution equation:

$$z(s) = m(s) + g(s)*C\beta(s) \qquad (2.20)$$

where the dependence on (s) emphasizes that the variables are in the space domain and, as introduced in the development of equation (2.19), C is a linear constraint operator, so that $C\beta(s)$ comprises two parts: $\beta_1(s)$ in D_1 and $\beta_2(s)$ in D_2, the latter identically equal to zero. It is straightforward to transform equation (2.20) to Fourier space using the Discrete Fourier Transform (e.g. Press et al., 1992):

$$Z(u) = M(u) + G(u)B(u) \qquad (2.21)$$

where upper-case letters are used to denote the transform of the space-domain variables and the variable u is the wave number (the analogue of frequency in time-dependent problems). Note that equation (2.21) can be used in the Direct Convolution step to obtain z(s), as is shown in Section 4.3 which follows, once β_1 has been obtained by matrix calculation or convolution via FFT.

Although it is tempting to proceed by rearranging equation (2.21) as:

$$B(u) = [Z(u)-M(u)]/G(u) \,, \qquad (2.22)$$

this does not work, because Z(u) in equation (2.22) would be the transform of z(s), which in turn comprises $z_1(s)$ in D_1 and $z_2(s)$ (which is initially identically zero) in D_2. The result would be that $\beta(s)$ would be a "smeared", reverse filtered set of incorrect $\beta_1(s)$ and non-zero $\beta_2(s)$ values as outlined in the discussion leading to equation (2.18). This "obvious" method does not achieve what is desired.

What is required is the set of reciprocal data values in D_1, which after filtering with the covariogram, yield the desired $z_2(s)$ in D_2 and in addition enable the recovery of the original $z_1(s)$ values in D_1.

Dudgeon and Merserau (1984:350) describe a method which is used for image enhancement, which can be modified to derive an iterative algorithm to obtain the correct $\beta(s)$ in the present context. Their algorithm closely follows the pattern of the development of equation (2.19) and can be summarized as follows.

Multiplying equation (2.20) by a constant scalar α, whose value is to be determined by trial, and rearranging:

$\alpha[z(s) - m(s)] - \alpha g(s) * C\beta(s) = 0$ (2.23a)

If convergence of the iterative calculation of $C\beta(s)$ is to be achieved after k iterations of the algorithm below, then for large enough k:

$C\beta^{(k+1)}(s) = C\beta^{(k)}(s)$ (2.23b)

where C is a constraint operator to set the values of $\beta_2^{(k)}(s)$ to zero.

Linearly combining equations (2.23a & b), one obtains an equation analogous to equation (2.19):

$C\beta(s)^{(k+1)} = \alpha\,[z(s) - m(s)] + [\delta(s) - \alpha\, g(s)] * C\beta(s)^{(k)}$ (2.24)

where

α is a scalar which controls stability and convergence of the iterative calculation

$\delta(s)$ is a (Kronecker) delta function, equal to 1 at s=0, zero elsewhere

$g(s)$ is the (symmetrical) covariance function centered on s = 0 and reflected about the Nyquist frequency.

The iterative algorithm based on equation (2.24) proceeds as follows:

Step 0:

- set k = 0
- choose α by trial (positive and usually less than or near to unity)
- set $\beta^{(0)}(s) = \alpha[z(s) - m(s)]$
- set $q(s) = \delta(s) - \alpha g(s)$ and compute its Fourier transform Q(u)

Step 1:

- compute $B^{(k)}(u)$ the Fourier transform of $C\beta^{(k)}(s)$
- compute $B_{\#}^{(k)}(u) = Q(u).B^{(k)}(u)$
- compute $\beta_{\#}^{(k)}(s)$ the inverse Fourier transform of $B_{\#}^{(k)}(u)$
- compute $\beta^{(k+1)}(s) = \beta^{(0)}(s) + \beta_{\#}^{(k)}(s)$
- increment k by 1

Step 2:

If convergence (defined by a suitably small error term in equation 2.23b) of $\beta_1^{(k)}(s)$ has not been achieved, go to Step 1; otherwise stop.

This procedure yields the desired $\beta_1(s)$.

2.4.3 Direct Convolution to obtain Optimal Estimates of the Unobserved Data

To get z(s) in D, compute G(u), the Fourier transform of g(s), and use equation (2.21) in the form

$Z(u) = M(u) + G(u)\, B^{(k)}(u)$ (2.25)

(where $B^{(k)}(u)$ was obtained in the last execution of Step 1 of the algorithm above) and then inverse transform the result. If convergence has been achieved, this procedure yields z_2, the Kriged estimates of z in D_2 and in addition, the correct (collocated) z_1 data in D_1.

2.4.4 Numerical Demonstration of Kriging by FFT

Rainfields measured by radar (or satellite) can be modeled by lognormal random fields with power spectra fitted by power-law functions (Bell, 1987; Schertzer and Lovejoy, 1987; Crane, 1990; Menabde et al., 1997). Adopting this model of the spectrum guarantees that the covariogram is positive definite, avoiding the embarrassment of calculating negative variances, as outlined in Section 2.2.

To demonstrate the technique of Kriging using the FFT, the following numerical experiment involving a simulated rainfield was constructed. The reason for using a simulated image of rain-rates is that individual ground-based radar images are rarely useful over a region 256km square, the range of reasonable accuracy being of the order of 70 to 80 km. The simulated image presented is statistically similar to a radar rainfield but can be quality controlled. The result is that the methodology will transfer to real-world radar-hydrology applications.

Over a square domain D of side 256 pixel units (km), a sample from a stationary lognormal random field was simulated by power-law filtering a white noise field with a spectrum whose negative slope in Gaussian-space was 2.5 (a typical value for instantaneous rainfields in plan as measured by radar; see, for example, Pegram & Clothier, 2001). The spatially dependent Gaussian field was scaled, shifted and exponentiated to produce a lognormal random field, simulating a rainfield of precipitation intensities. This generated field is shown in Figure 2.4a.

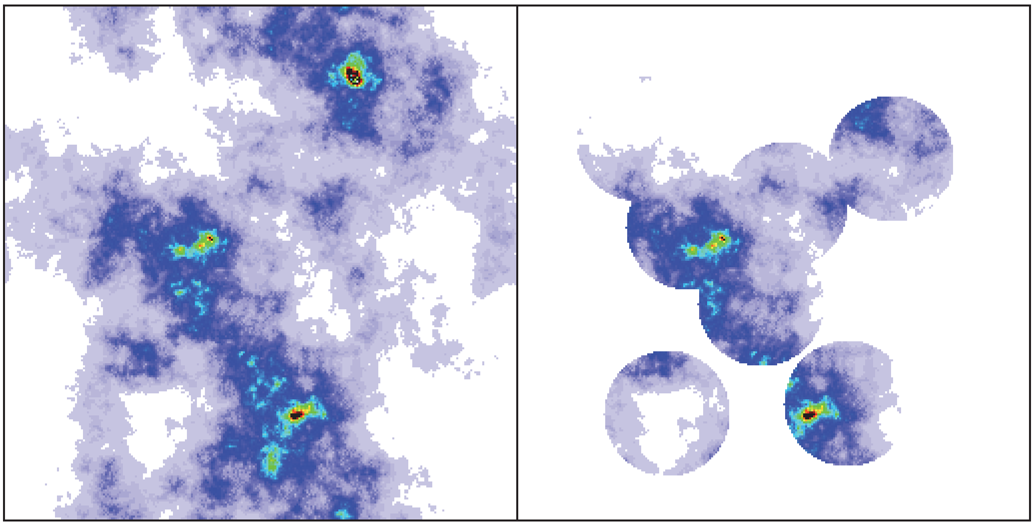


Figure 2.4a (left). Simulated rainfield (spatially correlated lognormal random field) on a 256-sided square region D.

Figure 2.4b (right). The data z_1 in region D_1 selected by sampling from D in Figure 2.4a in order to simulate the union of a set of observation from 7 radars

Next, seven circles of 64 km in diameter were distributed in D to simulate 7 radars whose union is the set D_1. The data in D_1 is the set z_1 of size $n_1 = 20377$, and is shown in Figure 2.4b; note that the size of D_2 is $n_2 = 45159$. Using the Iterative Constrained Deconvolution algorithm in section 4.2, the reciprocal data β_1 (with β_2 set to zero) were obtained after 100 iterations with α set to 2.0 (divergence resulted if α was outside the interval [0, 2.5]) and the Kriged field of Figure 2.5a was calculated using equation (2.25) and inverse transforming the result.

If the Kriged field had been obtained by direct solution using matrix methods, then the number of multiplications would have been in the region of $n_1^3 + n_2n_1^2 \approx 2x10^{13}$. For 100 iterations of the ICD algorithm and the calculation of the Kriged field, the number of multiplications is in the region of $101x2x(n_1+n_2)\log(n_1+n_2) \approx 6x10^7$. The latter calculation took 2 minutes on a 266MHz PC. As a fictional comparison, the matrix-based method (without clever programming) would thus have taken about 200 days, if it had been possible to manipulate the large arrays in the machine's CPU and physical memory without swapping data to disk storage.

After Direct Convolution of β using equation (2.25), the optimal surface was achieved and is displayed in Figure 2.5a.

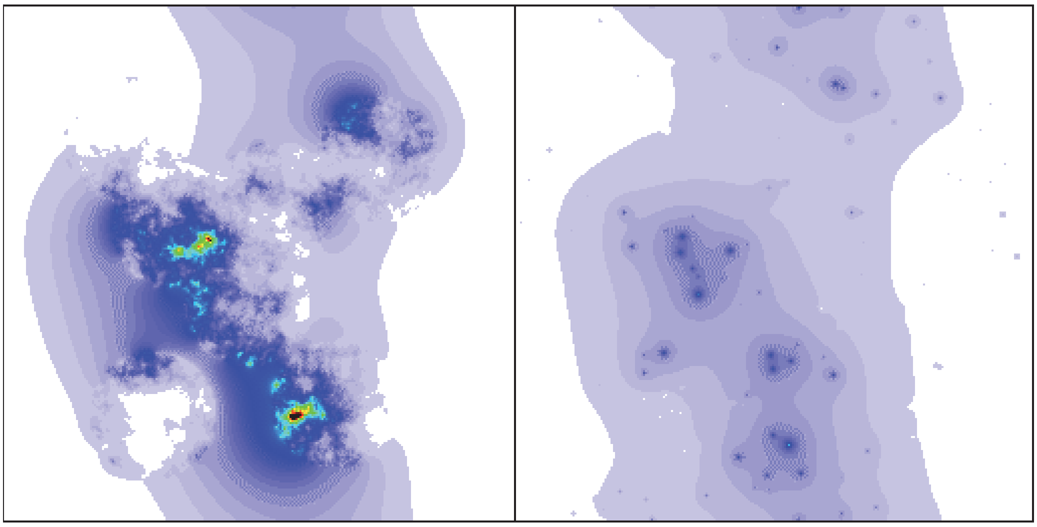


Figure 2.5a (left). The Kriged 'radar' image obtained from the reciprocal data by Direct Convolution using the FFT.
Figure 2.5b (right). The Kriged field of a second data-set consisting of 180 pixels randomly sampled from D in Figure 2.4a.

The z_2 field, shown in Figure 2.5a, is distinguished by its smoothness, the contours in the (integerized) data increasing in steps of 0.5 mm/h, being obvious even in gray scale. What is also apparent in Figure 2.5a is that the level 1.5 mm/h contour at the top of the figure "necks" and then widens again toward the edge. This "necking" is the direct result of the artifact of "wrapping" of the image through opposing edges, a nuisance feature of Fourier filtering. To prevent this, the image should be packed with a border of zeros before transformation, to fade out the extrapolation and prevent the information transfer across opposing edges. (This image was deliberately not zero-packed in order to demonstrate the phenomenon). In addition, note that the extended (Kriged) field z_2 captures some of the gross features of the hidden data and fills in the gaps (particularly between the bottom three circles) quite well. It is noted in passing, that a nice application of this methodology would be the interpolation of a radar field into those areas where information is lost due to ground clutter; this and other practical applications of the method appear in Seed and Pegram (2001).

Figure 2.5b is intended to show how the information in a relatively dense random network of 180 rain gages (1 gage per 364 pixels) can be extended into the sub-region D_2 by Kriging using FFT methods. This would normally have been done by conventional matrix based methods. As expected, the relatively sparse point sampling of the D-field in Figure 2.5b misses the high points. In spite of the sparseness, the (uniformly 'bad') Kriged surface extends the information evenly over the domain D_2 and manages to capture the major features in Figure 2.4a, if at a reduced level of information.

2.4.5 Explained Variance Using The FFT

To obtain $\Delta^{\$}(s)$, the Explained Variance, defined by equation (2.16) using the FFT, two derived fields are required. First, one needs the indicator field $I(s)$ where $I(s) = 1$ where data have been observed in D_1 and $I(s) = 0$ otherwise. Second, the 2-dimensional covariogram g(s) of the data in D_1 is required (obtained using the methods described in the Press et al., 1992) from which the explained variance filter $g^{\$}(s)$ of equation (2.15) is constructed.

$\beta^{\$}(s)$, the reciprocal data field to be used in equation (2.16) is obtained by ICD using the field $g^{\$}(s)$ as filter. $\Delta^{\$}(s)$ results from this computation. The explained variance filter $g^{\$}(s)$ is obtained from the correlogram by multiplying it by the field variance after squaring the elements of the correlogram: $g^{\$}(s) = r(s)^2\sigma_z^2$.

The 1-dimensional correlogram computed from the 2-dimensional spectrum of the data in Figure 2.4b, and the fitted correlogram model, are shown in Figure 2.6a: the column and row sums of the spectrum and their averages were inverse transformed and plotted on the figure; the model fitted to the average is also shown. The model has the equation

$$r(s) = \exp[-(s/L)^a] \qquad (2.26)$$

with L = 12.31 and a = 0.84. This model is shown superimposed on the radial correlogram, the other way of computing the 1-dimensional correlogram, in Figure 2.6b.

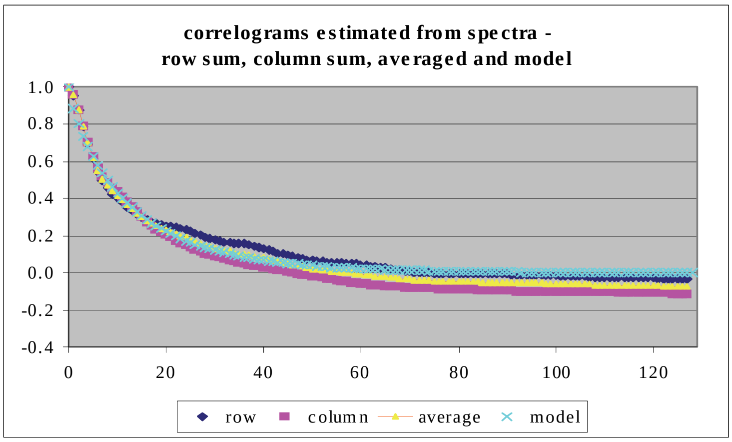


Figure 2.6a. The 1-dimensional correlogram computed from the 2-dimensional spectrum of the data in Figure 2.4b, and the fitted correlogram model; the four curves show the column and row sums of the spectrum and their averages and the model fitted to the average. is also shown. Note that l_{crit} (where $r(l_{crit}) = 0.01$) is 75 km.

In Figure 2.6a the elemental data of the radial correlogram, obtained from the 2-dimensional correlogram by inverse transformation of the 2-dimensional spectrum, are shown as a cloud of points.

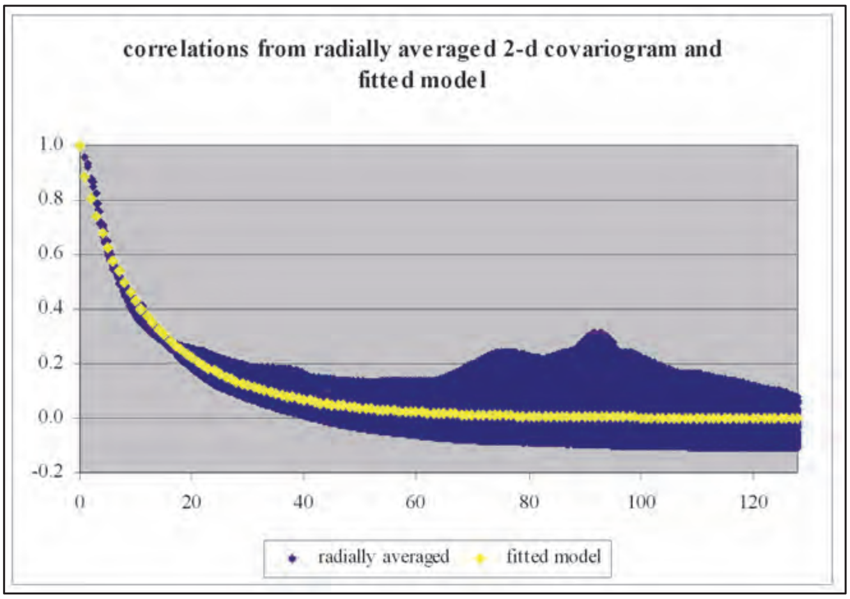


Figure 2.6b. The elemental data of the radial correlogram are shown as a cloud of points; the model fitted to the averaged 1-dimensional correlogram shown in Figure 2.6a is superimposed.

The square of the model of the 1-dimensional correlogram (which is non-negative by design) is used to obtain the 2-dimensional explained variance filter $g^{\$}(s)$ which is computed element by element assuming that the correlogram is a 1-dimensional function of distance, independent of direction (claiming isotropy). The reciprocal data-set $\beta^{\$}(s)$ is obtained by ICD using the technique described in Section 4.2, where now equation (2.24) is rewritten as:

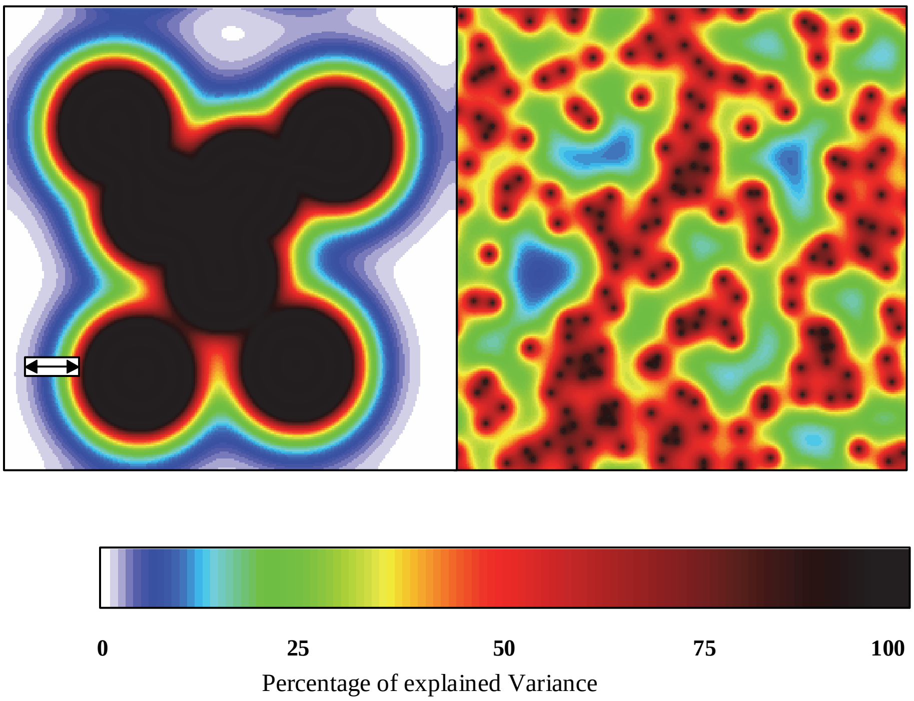
$$C\beta^{\$}(s)^{(k+1)} = \alpha[I(s)] + [\delta(s) - \alpha g^{\$}(s)] * C\beta^{\$}(s)^{(k)}$$

The $\Delta^{\$}(s)$ fields of explained variance for both the "radar" data of Figure 2.4b and the "gauge" data used in deriving Figure 2.5b are shown in Figures 2.7a and b. The scale is logarithmic, with white as zero and black indicating a correlation of 1 in D_1. It is perhaps a surprise to note the rapidity with which the explained variance drops off with distance - by a distance l_{crit} = 37km from the edge of a group of data points it is 0.01, as shown in Figure 2.7a.

Finally, as an interesting application of the explained variance (and one exploited in SIMAR), a merged field is computed by weighting the corresponding Kriged field by its explained correlation (ratio of explained and field variances). The merged field appears in Figure 2.8, as the lower half of the field for easier comparison of the detail of the original field shown in Figure 2.4a. The formula used to compute the combined field is:

$$z(s) = \frac{\Delta^{(1)}(s)z^{(1)}(s)/\sigma_{z_1}^2 + \Delta^{(2)}(s)z^{(2)}(s)/\sigma_{z_2}^2}{\Delta^{(1)}(s)/\sigma_{z_1}^2 + \Delta^{(2)}(s)/\sigma_{z_2}^2}$$

where superscripts (1) and (2) refer to the Kriged radar and gauge fields respectively.



Figures 2.7a and b. The $\Delta^{\$}(s)$ fields of explained variance for both (a) the "radar" data of Figure 2.4b and (b) the "gage" data used in deriving Figure 2.5b. In Figure 2.7a, the boxed arrow shows $l_{crit}/2$, which is 37 km long, which is the distance from the data where the explained variance is negligible (0.01). The scale of the shading appears below the figures.

The boxed arrow in Figure 2.7a shows $l_{crit}/2$ such that $g^{\$}(l_{crit}/2) = 0.01$ and is about 37 km. This can be compared with the distance of l_{crit} = 75 km where g(s) = 0.01 in the fitted covariance model shown in Figures 2.6a and b. The minor differences in the fields shown in Figures 2.8 and 2.5a are due to the additional information imported from Figure 2.5b and evinced in the small bulge in the bottom left-hand corners of the Kriged gauge and merged fields.

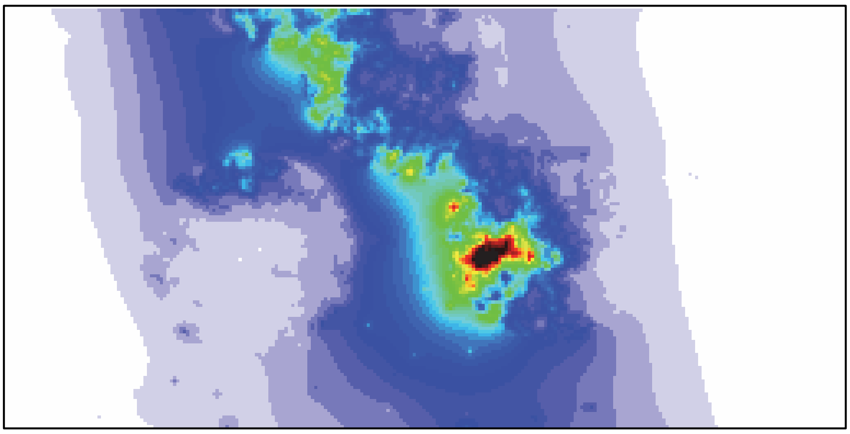


Figure 2.8. The completed merging process showing the lower half of the region

2.5 SUMMARY

The ideas presented in this section are here summarised in algorithmic form. Before commencing the analysis, one might wish to transform the data to make them more Gaussian (symmetrical). However, it was found that covariance functions of the original and transformed data do not materially differ from each other, which means that the linear filtering of the fields will not be much affected by the differences.

Kriging by FFT using Reciprocal Data

- Estimate m_1, the mean of the field of observed data z_1 in D_1, directly.
- Estimate the field variance σ_z^2 and the covariance g(s) from z_1, using direct evaluation for sparse data or the power spectrum for compact fields, in D_1.
- Determine the Reciprocal Data in D_1 by ICD.
- Krige the field from D_1 into D_2 by Direct Convolution.

Obtaining Explained Variance by FFT

- Determine the indicator field I(s), putting 1 where there are data and 0 where data are missing.
- Calculate the Explained Variance filter $g^{\$}(s)$ from g(s).
- Obtain the Reciprocal Data $\beta^{\$}_1(s)$ of I(s) by ICD.
- Calculate the Explained Variance field $\Delta^{\$}(s)$ by Direct Convolution.

The ICD algorithm used to compute either $\beta_1(s)$ or $\beta^{\$}_1(s)$ is identical, except for the input data and filter; $z_1(s)$ and g(s) in the former, I(s) and $g^{\$}(s)$ in the latter. Simple and

Ordinary Kriging also share the same routines; in the first m_1 is assumed, in the latter, m_1 is calculated from the sample mean of $z_1(s)$.

Merging the fields

Evidently when the Δ values become small, care will be needed in the precision of the computation. This was taken care of in K5/1152b.

2.6 INTERIM CONCLUSION

An alternative approach to Kriging has been presented in this section. Kriging can be separated into three steps - Constrained Deconvolution, Direct Convolution and Explained variance estimation, with four consequences:

First, it has been demonstrated that the concept of "reciprocal" data unifies the calculation method for Simple and Ordinary Kriging, as well as the computation of the Explained variance. It also facilitates the iterative constrained deconvolution of large data sets by using the computational efficiency of the Fast Fourier Transform.

Second, Ordinary Kriging is shown to be no more complicated than Simple Kriging for large data sets if reciprocal data, rather than Kriging weights, are calculated. In large data-set applications of Ordinary Kriging, the sample mean may be substituted for the (unknown) mean, with little error.

Third, an approximation to the Explained variance (the difference between the Kriging and Field variances) has been derived, which makes it feasible (as well as desirable) to compute this quantity over the whole Kriging domain in large data mapping problems.

Last, it has been demonstrated that Kriging can be achieved by Fourier transform methods, providing a computationally feasible alternative to matrix methods where large homogeneous data-sets are involved.

To finish this section, there follow in Figures 2.9a, b, c & d images captured in May 2003 from the METSYS web-site which are the best estimates of rainfall from gauges, radar, satellite and the combined field.

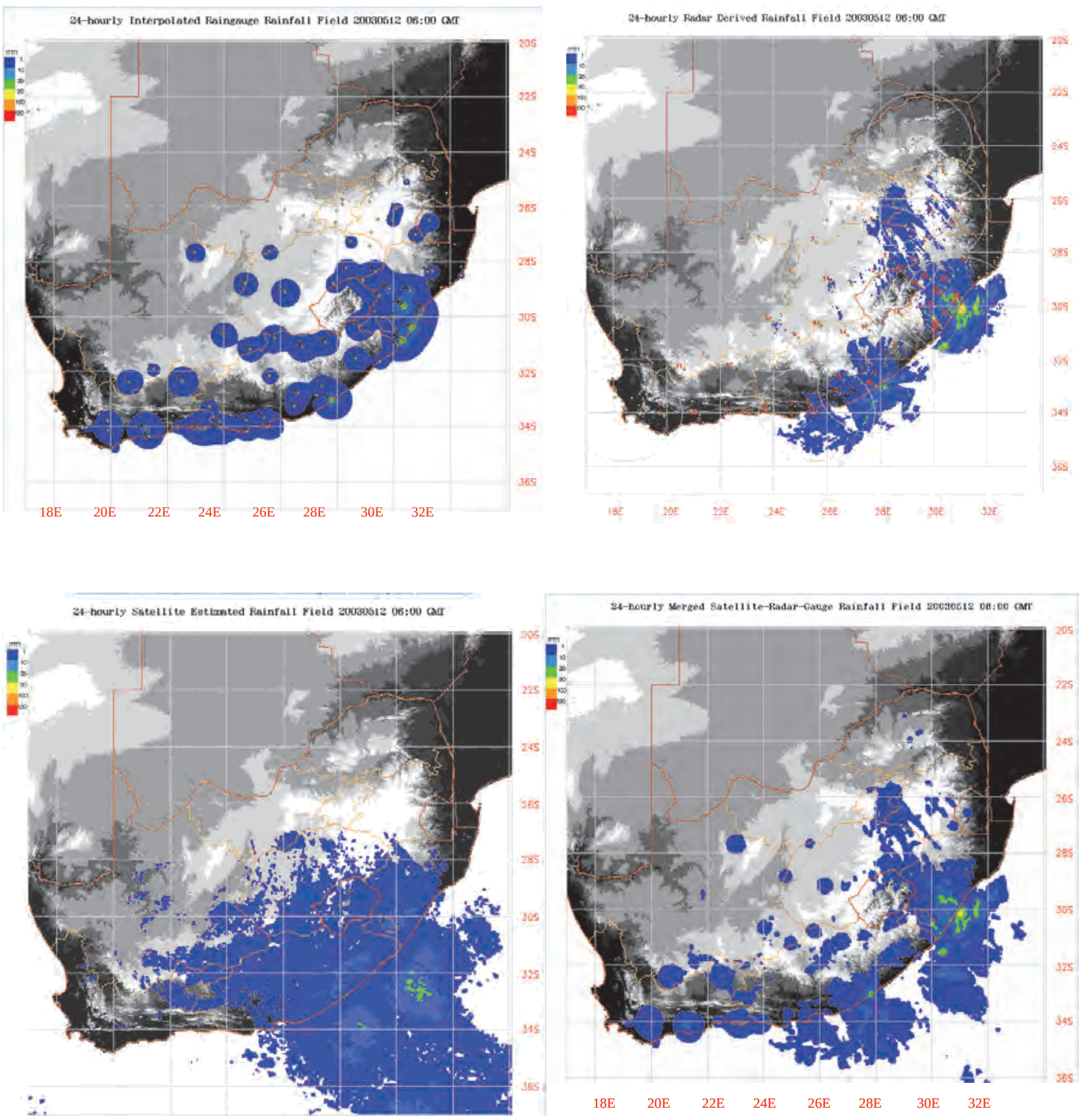


Figure 2.9. Images of rainfall estimation for the 24 hours prior to 8:00 am on 12 May 2003 over Southern Africa at a spatial resolution of 1 minute – the area is 1024 minutes square. In order from the top left and reading to the right, they are Gauge, Radar, Satellite and Merged fields estimated by Kriging using the FFT, the Explained variance and a final combination, as reported in Volume 2.

3. TWO IDEAS WHICH ARE STILL UNDER DEVELOPMENT: CONDITIONAL KRIGING AND IMAGE FORECASTING

The method suggested in Section 2 above using the weighted explained variance does not specifically preserve the gauge readings at points, although it is the method presently used in the SIMAR field construction.

An alternative strategy that seems to have more appeal is suggested. It is to conditionally fit the radar (or satellite) field to the raingauge data, either precisely or with some variance. The argument is in terms of satellite data, but can just as easily apply to radar fields (Pegram, 2001).

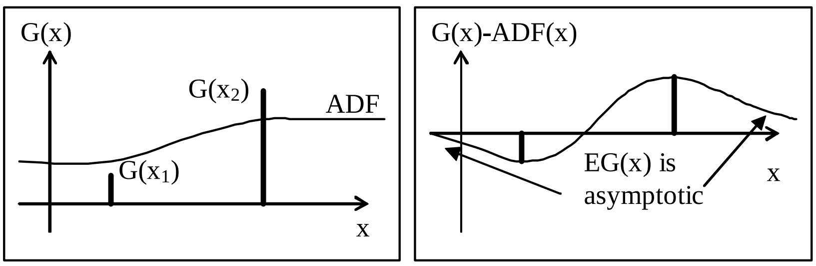
3.1 PRINCIPLE OF CONDITIONAL KRIGING

Assume that the raingauge data G(x,y) are accurately known (or at least they are unbiased estimates of the rainfall with known variance) when they are there (except after a break in reporting like over a weekend - perhaps discard the first reading after the break, or at least try to patch the data?)

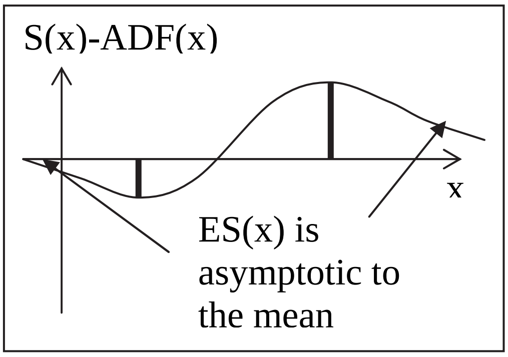
Step 1:
In each of the 8x8 (64 pixel) blocks covering the 1024 square image derived from the satellite image S(x,y) compute the Average Day's Precipitation (ADP) at the centre of each block. The Gaussian linear block smoother introduced in the inaugural meeting of this project will define the Average Day's Field (ADF) over the country. This ADF becomes the reference field for deciding whether it has rained or not at a point: call it ADF(x,y).

Step 2:
Compute the Spectrum/covariance centred on each block after having subtracted the ADF from the satellite data. Then compute the spectrum from the differenced field S(x,y)-ADF(x,y) to yield a "true" correlogram or take an average one based on previous observations. If the radar data are anything to judge by, the spectrum is relatively invariant between weather types and between climates; this observation is based on fitting the String of Beads Model to Bethlehem and Melbourne data.

Step 3:
Krige the shifted gauge data G(x,y)-ADF(x,y) to produce an expected gauge field in each block; call this EG(x,y). The Kriging of the gauge data is done using the covariogram (or semi-variogram) computed from the satellite data - this produces the EG(x,y) field. This might be most conveniently done by Fast Fourier Transform, the methodology developed in this project and outlined in detail in Section 2.



Step 4:
Determine S(x,y) at points corresponding the gauge locations (observations only). Compute the differences S(x,y)-ADF(x,y). Krige these using the satellite covariogram to get the expected satellite field ES(x,y).



Step 5
Combine the various fields as follows to get the best merged rainfall field based on satellite and gauges:

R(x,y) = ADF(x,y)+EG(x,y)-ES(x,y)+S(x,y)

setting all negative values to zero.

This equation, once computed over the field, stitches the satellite field down onto the gauge observations using the satellite field as the estimator of the Wetted Area Ratio and Averaged Daily Rainfall over an area and yet maintaining the correct spatial distribution of rainfall between gauges based on the two dimensional covariogram measured from the satellite data.

3.2 FORECASTING AND MORPHING FIELDS

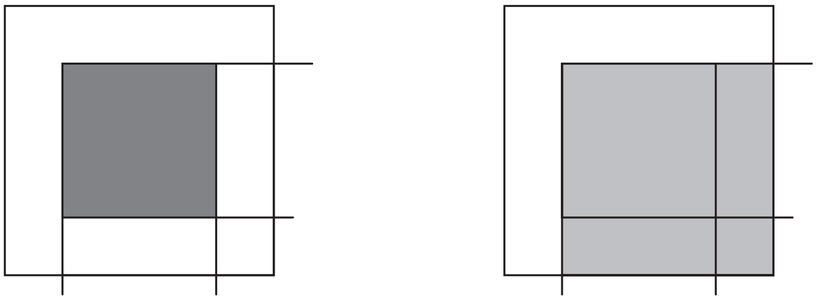
The computation is done block by block as a first step. The next refinement would be to smooth the blockiness by some technique if required. The only problem areas are likely to be where there is a raingauge near the edge of a block because then its effect won't carry over to the adjoining block unless there is some overlap between blocks. This still needs thought.

This problem of 'morphing' images has been considered in a parallel study (K5/1217: Umgeni Nowcasting using Radar) where the difference between (a) average advection and (b) field advection of a radar image makes for a measurable difference in the behaviour of the field – it neatly accounts for some of the circulation effects. The principles carry over from the mesoscale to the sub-continental scale at which modelling is being done in this project.

Thoughts on obtaining the field spectrum for forecasting or smoothing

There will be 8x8 blocks at time ($t-\Delta t$) , each is *128 pixels square and is* to be matched by a corresponding (shifted) area at a subsequent time t.

A possibly neat fast way to do this is to use FFT methods; there are three cases: a corner block, a side block and an interior block. Consider a *top left-hand* corner:



$A_1 = A(t-\Delta t)$

Dark grey square is one of the 128x128 blocks at ($t-\Delta t$)

White border is 64 wide. The part above and to left of the dark square is outside 1024 square. The whole white region is packed with zeros before FFT'ing.

$A_2 = A(t)$

The light grey square is 192x192 pixels and includes all the data from the image at (t). The white area is packed with zeros.

The data in both A_1 and A_2 are standardized relative to the already computed ADF before FFT'ing.

If A_1 is the shifted 128 block packed with zeros to 256 and A_2 is the 256 block centred on A_1, then $A_1(u)A_2(u)^*$ is the cross-spectrum computed using FFT. It is inverse filtered to compute the cross-correlogram. The maximum of the cross-correlogram (or perhaps the average of the last two or three) is found by search and the advection shift Δs is found.

Forecasting

If the maximum cross-correlation is r_{max} then this can be used to calculate an incremental cross-correlogram which can be used for forecasting. Assume that the decay is geometric (average the last 4 covariograms to get a smooth estimate and look perhaps at local Laplacian smoothing as an alternative). To get the appropriate covariogram to infill the intervening times at m sub-intervals the shift will be $\Delta s/m$ and the corresponding covariogram can be calculated in a couple of possible ways. If the maximum correlation between the observed images is r_{max}, multiply the best covariogram by $r_{max}^{1/m}$ and shift it by $\Delta s/m$. This should provide the best forecast of the field into the future if we use the FFT to forecast. Otherwise, fit the String of Beads model to the field and go that route.

Smoothing

For smoothing (infilling between half-hour observations) something similar is needed. Either use the FFT filter (which will degrade the image in between the observed ones) or find the advection vector Δs and then linearly morph the one image into the other along the advection vector in sufficiently short time steps as shown in Figure 3.1.

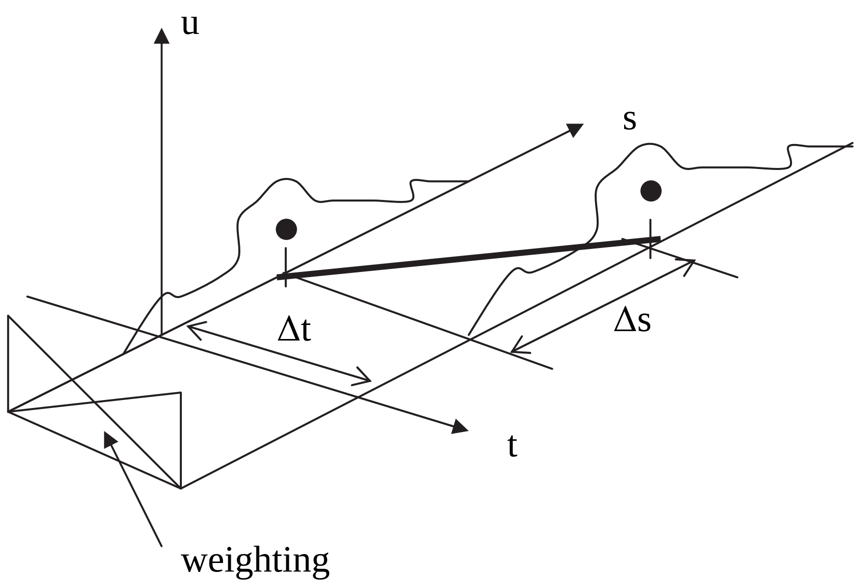


Figure 3.1. The linear morphing scheme for computing accumulation of rainfall between two images, schematically represented in one dimension.

In the one-dimensional sketch shown in Figure 3.1, the two “images” of intensity u have “centroids” which are shifted by Δs in Δt, along the heavy line. The linear weighting functions are used to morph the two images along the dark line at whatever intervals are desired (relative to the smallest of the features which need to be preserved), to achieve a smooth transition. The total depth of rainfall is then equal to the intensity times the subinterval, and the accumulation is then computed by summing these. This should provide a solution to one of the problems identified in the project which was:

Because of the need for infilling between successive satellite images to get rid of “puddles”, to monitor radars and to forecast rainfall fields in intervals shorter than half-hourly, there is a need for field forecasting and/or smoothing.

In Figure 3.2 appears an example of the sort of forecasting achievable with the String of Beads Model and field advection, as computed for K5/1217, without pre-empting the report on that study.

The point is that the principles can be applied to the ½ hour images of rainfall extracted from the METEOSAT images (or the second generation satellite) to enable meaningful daily accumulations to be obtained.This has yet to be done, but the theory and algorithms are in place.

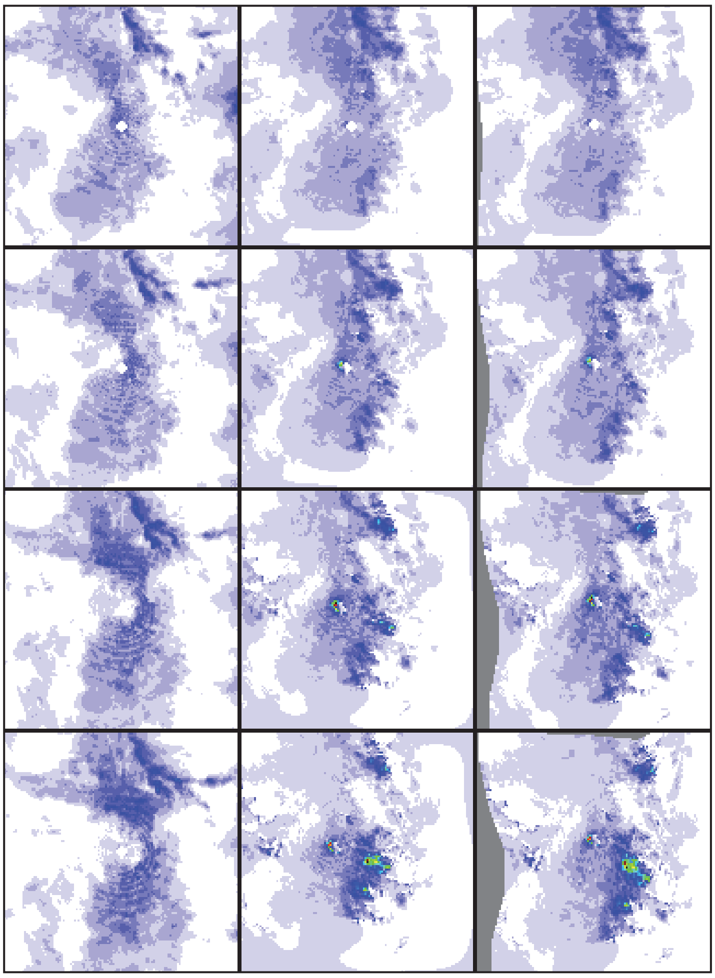


Figure 3.2. Preliminary attempt at forecasting radar rainfields using the String of Beads Model, up to 20 minutes ahead, at 5 minute intervals. The column on the left is the set of observed rainfields, increasing in time from top to bottom. The centre column contains the unadvected fields forecast in time. The column on the right is the set of forecasts (estimates of the real fields on the left) with the field advected in 11x11 panels and then linearly morphed between each panel – the distortion from the original field can be understood by the encroachment of the grey (unknown values) area at the top and to the left.

Scrutiny of the fields shows that the forecast rainrates are growing rather decaying as observed. This is due to the set of parameters chosen to describe the correlation in time. As it happens, in rounding off to 2 significant figures, the stochastic difference equation describing the pixel scale temporal development is marginally non-stationary. This is being rectified and does not detract from using the example to describe field advection.

4. THREE-DIMENSIONAL KRIGING: APPLICATIONS TO CAPPI CLEANSING

While on a 12 week visit in 2001 as a Distinguished Miegunyah Fellow to Melbourne University, Prof Pegram continued his research collaboration with Dr Alan Seed, now at the Bureau of Meteorology. What transpired was a felt need to improve the estimation of 'true' rainfall measured by weather radar in the context of ground clutter, anomalous propagation, beam-blocking, bright band etc. This concern mirrors that expressed frequently by Dr Terblanche and the METSYS group which led to the attempt by Marion Mittermaier (Pegram and Mittermaier, 1998) to tackle the bright band problem by classifying the vertical profile, which work was done as an MScEng with Prof Pegram at the University of Natal.

Dr Seed took hold of the Kriging idea and developed an algorithm which infills the radar data at source in polar coordinates using ordinary Kriging based on nearest neighbours. Prof Pegram has since extended the Kriging by FFT idea to 3 dimensions and the comparison of the two methods was presented at the European Geophysical Society Assembly in Nice on 24th April 2002. A precursor to this presentation was a joint paper at the Kyoto triennial Radar Hydrology conference which both authors attended. An excerpt of the paper presented there (Seed and Pegram, 2001) follows. It was entitled:

4.1 USING KRIGING TO INFILL GAPS IN RADAR DATA DUE TO GROUND CLUTTER IN REAL-TIME

The paper describes the use of an anisotropic Kriging interpolation method to infill gaps in the radar data that arise from ground clutter contamination. The objective is to use all available non-contaminated neighbouring data to estimate the missing value. The structure of rainfall in the vertical and horizontal planes is quite different due to the limited vertical extent of the atmosphere. The Ordinary Kriging equations are used to calculate the weighted sum of from 5 to 40 non-contaminated points that are nearest (in the scaled anisotropic distance sense) to the point to be in filled. Cross-validation comparisons show that the cost of using more points far outweighs the accuracy gained. This paper describes the estimation of the correlation lengths in the horizontal and vertical r_0, z_0, the scaling anisotropy parameter α, and the scaling exponent H. A case study is used to evaluate the method and the results are presented.

4.1.1 Data

Three periods of significant rainfall from the Melbourne radar archive were examined:

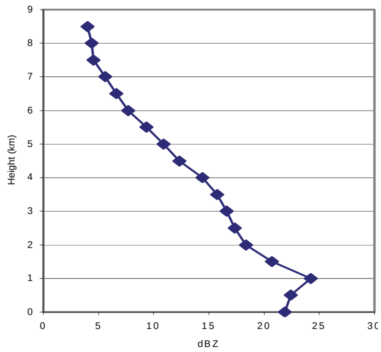
04:00 21 April 2001 to 04:00 22 April 2001
This was at least a 1 in 5 year 24-hour event producing **widespread rainfall** over the area covered by the radar. The mean vertical profile for all range bins within 50 km of the radar is shown in Figure 4.1. The vertical profile shows a strong gradient.

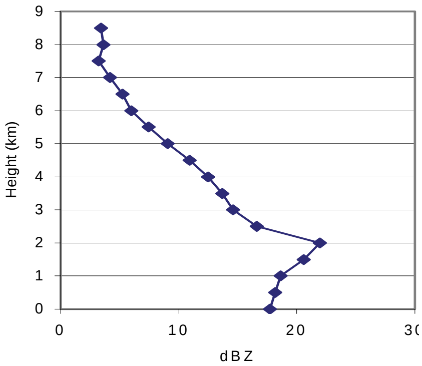
14:00 – 24:00 25 December 1999
This 10-hour period of rainfall is mostly **widespread rainfall** with rainfall intensities generally below 2 mm/h. The mean vertical profile of reflectivity for the 10 hours is shown in Figure 4.2.

03:00 to 09:00 27 December 1999

This period contains some of the most **intense convective** weather in the Melbourne radar archive. The convective nature of the rainfall is evident in highly localised accumulations. The mean vertical profile of reflectivity for the 6 hours is shown in Figure 4.3.

The data used to derive the Figures 4.1 to 4.3 were accumulated from 10-minute CAPPI data with the ground clutter pixels set to zero.





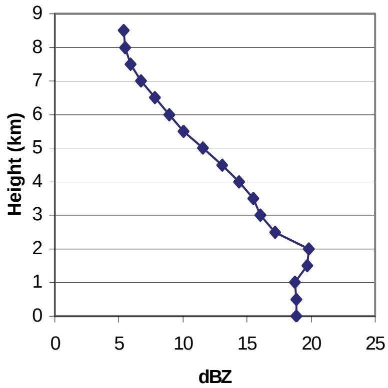


Figure 4.1 (left), **4.2** (centre) **and 4.3** (right). Mean vertical profiles of reflectivity obtained from Melbourne radar archives, courtesy of Dr Alan Seed, for the periods 04:00 on 21 April 2001 to 04:00 on 22 April 2001, 14:00 – 24:00 on 25 December 1999, and 03:00 to 09:00 on 27 December 1999, respectively.

4.1.2 Quality Control

The first step in the analysis was to identify the radar bins in the three-dimensional polar data that were affected by ground clutter. This was done by selecting a 12-hour period with no rain and calculating the probability that the reflectivity at a particular bin exceeded 15 dBZ. The volume scan consists of 15 elevation angles, each with 360 radials of 200 1 km bins. A bin was classified as clutter if the reflectivity exceeded 15 dBZ more than 50% of the time in the period. This mask was applied to each volume scan in the data and the clutter bins were excluded from all subsequent calculations.

4.1.3 Variogram Modelling

The variogram used in Kriging is a measure of spatial dependence and in a stationary random field is the complement of the covariance to the variance of the field; in fields (volumes) with varying mean values (“drift” in kriging parlance) the variogram is a more robust measure that covariance. The horizontal variogram, γ_r, was calculated by identifying all the radar bins in the volume that would be used to construct a CAPPI that was 1500 m above the radar. The distances and the square of the differences in reflectivity between all pairs of bins that exceeded 15 dB were calculated. A variogram was calculated for each volume scan and written to a file.

The variogram for the vertical profile of reflectivity was calculated for each volume scan by first estimating the mean vertical profile of reflectivity for all bins that were within a 50 km range of the radar and had a reflectivity that exceeded 15 dBZ.

The variogram for the vertical profile of reflectivity, γ_h, was calculated using

$$\gamma_h(\Delta h) = \left\langle \left[\left(z(h_1) - \mu(h_1) \right) - \left(z(h_2) - \mu(h_2) \right) \right)^2 \right] \right\rangle \quad (4.1)$$

where

$\mu(h)$ is the conditional mean reflectivity and $z(h)$ is the reflectivity of a polar bin at height h above the radar; the variogram for horizontal reflectivity was computed similarly, relative to the mean at the relevant h. The brackets 〈〉 imply averaging over all pairs $\Delta h = |h_1 - h_2|$ apart.

The 2-dimensional combined variogram model (isotropic horizontally) is that suggested by Smith et al. (2000):

$$\gamma(r,h) = \begin{cases} \sigma^2 \left[\left(\frac{r}{r_0} \right)^2 + \left(\frac{h}{h_0} \right)^{2\alpha} \right]^H, & \left(\frac{r}{r_0} \right)^2 + \left(\frac{h}{h_0} \right)^{2\alpha} < 1 \\ \sigma^2, & \left(\frac{r}{r_0} \right)^2 + \left(\frac{h}{h_0} \right)^{2\alpha} \geq 1 \end{cases} \quad (4.2)$$

where σ^2 is the variance of the field, and α, H, r_0, and h_0 are model parameters.

The parameters were found by using a non-linear routine to minimise the sum of squares of the errors and are shown in Table 4.1. The power law model fits quite well close to the origin where we will be using the variogram. The trick is to know where to stop (maybe 80% of r_0) when using the model.

The r_0 and h_0 parameters are most sensitive to the spatial organization of the field, while the scaling parameter, H, and the (r,h) anisotropy, α, are quite similar for the three events. Since it is difficult to estimate the variogram with any precision, it is probable that the model could be simplified by setting $\alpha = 1$ without much loss in accuracy. The vertical and horizontal components of the three variograms appear in Figures 4, 5 and 6.

Table 4.1 Variogram model parameters

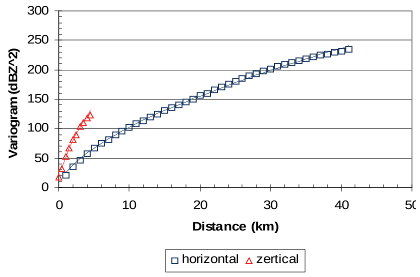
Event	r_0 (km)	h_0 (km)	α	H
21 April 2001	27	8	1.0	0.30
25 December 1999	35	11	1.2	0.25
27 December 1999	12	7	1.3	0.26

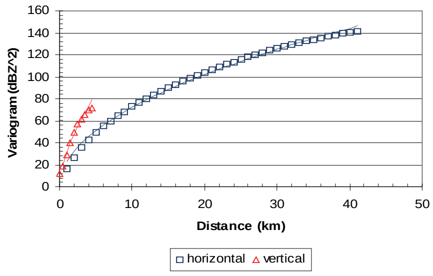
A real-time system would be required to estimate these parameters for each volume scan prior to any data patching. For demonstration purposes it is sufficient to assume that the parameters as given in Table 4.1 remain fixed for each event and simply calculate the variance, σ^2, at each time step. Another possibility is to classify the scene “convective” or “stratiform” based on the CV and wetted area ratio and then switch between the two-parameter sets.

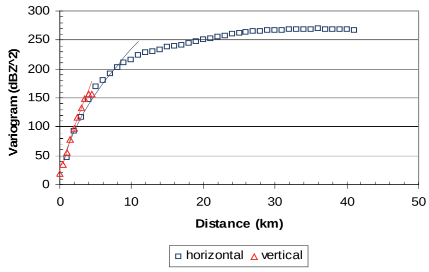
4.1.4 Data Patching

The standard method of infilling clutter pixels is to use the nearest neighbour from the elevation above. The performances of using the nearest neighbour method and 3d Kriging using 5,10,20,40 nearest neighbours were evaluated by hiding raining pixels and calculating the mean standard error (MSE) of estimation. 1000 radar bins with ranges between 20 km and 80 km from the radar and with echoes that exceeded 15 dBZ were drawn at random from the base scan of each radar volume scan and used to calculate the

MSE for that volume. The analysis was repeated for each volume in the three data sets and the median MSE for each event was calculated. For each hidden radar bin, the nearest neighbours in the scaled 3D space surrounding the point were identified. In practice, since the method is to be used to interpolate over clutter, the nearest non-clutter pixels will be used and therefore the minimum distance between the interpolation point and the neighbours in the base scan was set at 3 km.







Figures 4.4, 4.5 and 4.6. Mean horizontal and vertical variograms for the 21 April 2001, 25 December 1999 and 27 December 1999 cases. The solid lines are the model fit based on Equation 2.

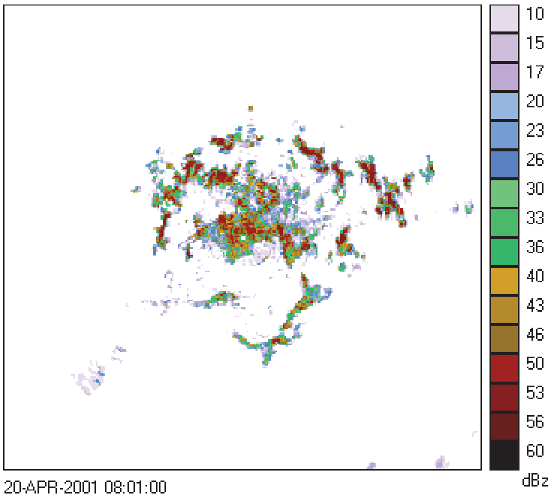
Table 4.2 Median mean standard error (dB) for interpolation using the nearest neighbour (NN) and 3D-Kriging using 5 (K5), 10 (K10), 20 (K20), and 40 (K40) neighbours.

Event	NN	K5	K10	K20	K40
21 April 2001	4.6	3.9	4.0	3.8	3.5
25 December 1999	2.6	2.1	2.2	2.2	2.2
27 December 1999	4.0	3.8	3.9	4.0	4.0

It is apparent from Table 4.2 that neither method is particularly accurate, however the Kriging methods have interpolation errors generally in the 2.1-4.0 dB range depending on the method and meteorology of the day. The large interpolation error for the nearest neighbour method for 21 April is probably due to the low level of the bright band on that day and this would be reduced if the bias due to the vertical profile was accounted for. There is very little difference between the 4 Kriging options and, since the number of computations is proportional to the third power of the number of neighbours, the extra cost in including more interpolation points is considerable.

The method was tested using the 21 April 2001 case. Figure 4.7 shows the ground echoes in the base scan of the Melbourne radar. The city centre is 10-20 km east of the radar in an area of ground echo. Figure 4.8 shows a typical scene with extensive rain over Melbourne and Figure 4.9 shows the same scene with the clutter replaced using a 5-

neighbour 3D Kriging scheme. The interpolations in this scene were performed in less than a second on a standard lap-top PC so it is a perfectly feasible method to use in an operational setting.



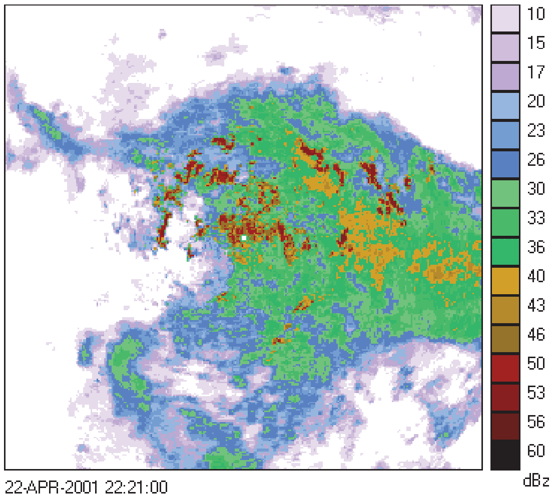
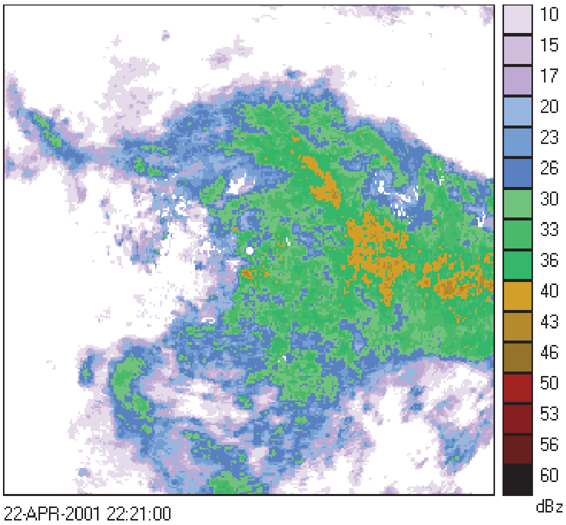


Figure 4.7: Ground echoes in the base scan of the Melbourne radar, showing extensive clutter over the city near the radar (which is in the centre of the image) and to the east.
Figure 4.8: Base scan from the Melbourne radar at 22:21 on 22 April 2001 showing the impact of the ground echoes over Melbourne city.

4.1.5 Interpolation onto a Surface at Ground Level

Since the objective is often quantitative rainfall estimation at ground level, it would be interesting to use an interpolation technique to estimate the radar reflectivity at ground level directly, rather than at some height aloft. By ground level is meant the elevation of the radar above sea level, not a surface that follows the actual topography. The advantages for this method are that the ground clutter is removed, the vertical profile of reflectivity is corrected, and an optimal interpolation scheme is used to estimate the radar reflectivity on the ground. The disadvantages to this scheme are difficulties in estimating the variogram and vertical profile of reflectivity in real-time in a robust manner, and spatial smoothing if too many local radar bins are used in the interpolation.

To test the potential for the idea, the 21 April 2001 event was processed using 3D Kriging to interpolate polar coordinates of radar reflectivity onto a surface at ground level, and then convert these polar data into a Cartesian map. Only polar bins that had non-zero reflectivity aloft were interpolated, the time taken to generate a single map was generally in the order of 1 second, increasing to about 3 seconds in maps with a large raining area. Figure 4.10 shows the interpolated map of radar reflectivity for the same time as Figure 4.9.



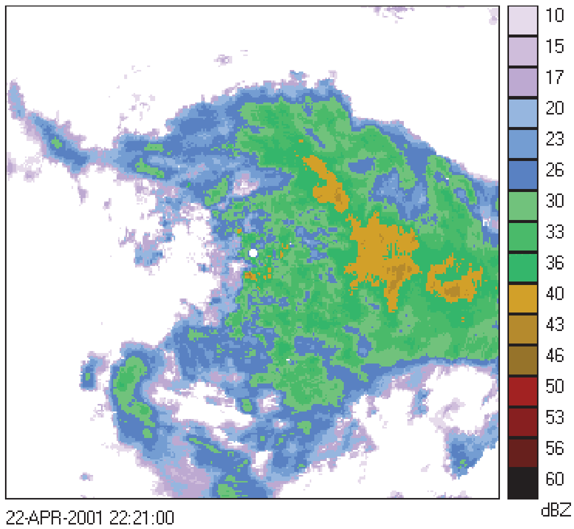


Figure 4.9. Base scan for the Melbourne radar, with the ground echoes replaced using a 5-neighbour 3-D Kriging scheme. **Figure 4.10** Radar reflectivity interpolated onto a plane at ground level using 5-neighbour 3-D Kriging.

Note that in Figure 4.10, bins where the first non-clutter bin exceeded 20 dBZ were patched using Kriging, hence there are occasional holes in the light rain; also evident is a loss of very small-scale variability, particularly at far range where the points used in the interpolation cover a larger area and the interpolation scheme is more like a simple averaging of the nearby data. The region with high reflectivity, say the region where the reflectivity exceeds 40 dBZ, is more contiguous in the interpolated image. The images appear to be very clean, but the only true test will be a reduction in the noise in rain gauge/radar data pairs. This method has been written so that the interpolation can be used to estimate the reflectivity at any (x,y,z) point in the volume, and therefore it can be used to construct other cuts through the volume.

4.2 Three Dimensional Kriging using the FFT

Scrutiny of the text of the paper will reveal that the Kriging described in Section 4.1 was done directly, one point by point, selecting nearest neighbour control points for each missing target data point. As an alternative to direct Kriging, early attempts to infill flagged and missing data using 3-D Kriging and the FFT include the following example from Melbourne radar data. Figure 4.11 shows a stack of 5 levels of 128-square (in half images for economy of presentation space) before infilling. All of the black regions are missing data in the conversion of spherical data to a Cartesian reference frame. The dark rings are caused by the fact that the radar is programmed to perform 15 scans at increasing elevations, so that some of the volume is not sampled. Dr Seed (an experienced practitioner in the weather radar community) expressed some surprise at the high proportion of missing data in the volume scan.

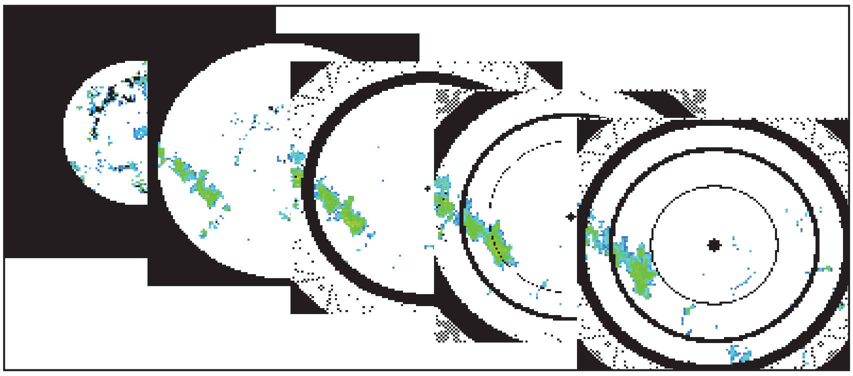


Figure 4.11. 5 layers of radar scan information from ground level to +4 km

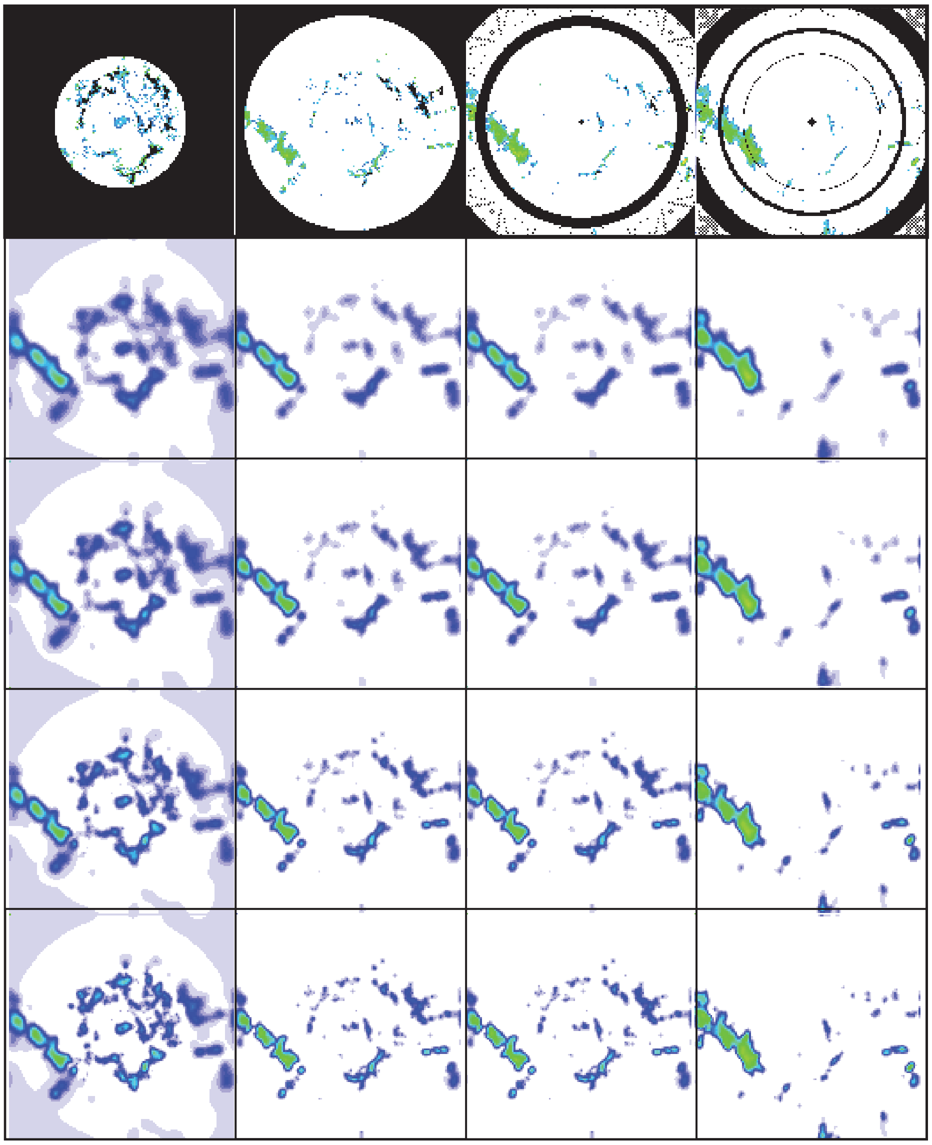


Figure 4.12. Infilling using Kriging by FFT. The top row shows levels 0 to 3 from Figure 4.11; the second to fifth rows are the estimates after 200, 300, 500 & 1000 iterations of the Iterative Constrained Deconvolution algorithm.

Figure 4.12 shows the result of infilling the stack of CAPPIs of Figure 4.11 by Kriging with the FFT as described in Section 2. The pale blue halo around the ground-level images in the leftmost column is the mean (1mm/hr) estimate. The green zone to the left of each image shows development over the full height while the less intense ring of reflectivity around the radar disappears by level 3. Although the fewer number of iterations causes blurring of the image, the structure is relatively well obtained by 500 iterations and even 300 is quite reasonable. It becomes important to find out what the relative benefits of the different computational strategies are. This refinement and evaluation will be deferred to the follow-on project, K5/1425/1.

In the interim, the FFT Kriging ideas have been aired at a prestigious scientific meeting as a presentation (Pegram and Seed, 2002).

5. PRACTICAL IMPLEMENTATION OF THE IDEAS FOR SIMAR

The main aim of the SIMAR project was to provide daily rainfall maps in fine detail on-line on a daily basis. The applications of the Kriging by FFT algorithm developed and outlined in Section 2 to CAPPI cleansing described in Section 4 was an unexpected and potentially extremely valuable contribution to improving the primary rainfall estimates from radar. Exploiting this and the ideas of Section 3 will be done in the follow-on contract, K5/1425/1 and not herein. The blending of raingauge, radar and satellite data to provide the best composite was always at the forefront of the researchers' minds. Consequently, soon after the 2001 steering committee meeting, Prof Pegram and Izak Deysel of METSYS spent some time together (and separately) converting the FORTAN code to C++ to run under LINUX on the METSYS machines.

The original purpose was to use the 2-D FFT Kriging algorithm to spread the information from the radar-rainfields as far as possible over the sub-continent. The results were mildly disappointing for two reasons: the extent of the extrapolation was much less than expected (limited to 30 km beyond the radar circle) and the number of iterations necessary to get good estimates was very high. To emphasize the point about the limited extrapolation, in the centre of Figure 5.1 will be seen a small area of blue (rain above 1mm) outside the radar circle centred at Bloemfontein. In spite of this, the methodology is in place and through Izak Deysel's work with including the satellite estimates, has been used to routinely post the 24 hour accumulated rainfall map on the METSYS website on a daily basis, an example of which was shown in Section 2.6. This can only be improved upon, and will be. In addition, there is to be a great effort to achieve more economical computations of the fields, as well as the "cleansing" of the radar images. This effort will be undertaken in project K5/1425/1.

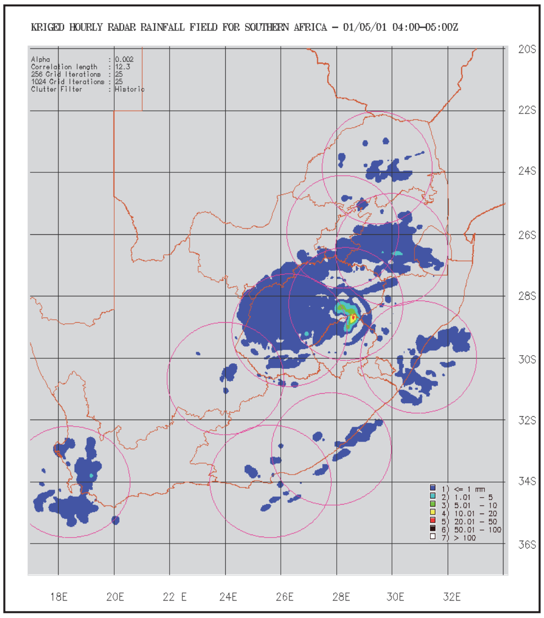


Figure 5.1. 2-D FFT Kriging of radar rainfields, with the variation that the field was first divided into 256x256 pixels of 4 minute squares and iterated 25 times, providing the starting values for the next step with the field described by 1024x1024 pixels of 1 minute square and also iterated 25 times.

6. THE WAY FORWARD

There are four areas of theoretical development that are envisaged for extending the achievements of the SIMAR project which will be undertaken in projects K5/1425/1 and K5/1429:

- Conditional merging of rainfields
- Using forecasting to improve estimation of accumulations
- CAPPI cleansing
- Speeding up the reciprocal data calculation

The conditional merging of rainfields and the applications of forecasting have already been outlined in Section 3 and it remains to see how to make them happen in practical computations. Several bench trials will be needed before the methodology is ready to apply routinely.

It is important to provide as good an estimate of rainfields from radar as is feasible, so it seems reasonable that a secondary (unforeseen) benefit of the development of FFT Kriging is to infill gaps in CAPPIs (due to ground clutter and anomalous propagation) by extrapolating from known data within the volume scan online before the information is stored. This can only be countenanced when it is shown that there is no better current method (in terms of optimal information dispersion) when compared to nearest neighbour kriging etc.

The routine use of 3-D FFT Kriging in this context relies on speeding up the reciprocal data calculation. Of course, by the time PCs have calculation speeds of 5GHz or more, then the ICD algorithm will be convenient. It only at this juncture that it is a little slow (3 minutes for a 128x128x8 CAPPI stack) but that will change.

In the interim, a search is on for a fast and efficient alternative. It is conceivable that algorithms exist for computing the solution of from 50 000 to 100 000 variables directly, especially if one takes advantage of the bandedness of the coefficient matrix. While in Melbourne for 8 weeks at the beginning of 2002, Prof Pegram discovered that there is an algorithm used by Geodesists for solving large network problems based on Least Squares formulations that may provide a key. A great deal of effort is to be made in seeking a solution to this numerical problem to gain the fullest advantage of the Kriging methodology.

REFERENCES

Ahmed, S. and G. de Marsily, Comparison of Geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity, *Water Resources Res*., 23(9), 11717-1739, 1987.

Barnes, S.L., A technique for maximizing details in numerical weather map analysis, *J. Appl. Meteor.*, 3, 396-409, 1964.

Bell, T.L., A space-time stochastic model of rainfall for satellite remote sensing studies, *J. Geophys. Res*., 92(D8), 9631, 1987.

Brandes, E.A., Optimizing rainfall estimates with the aid of radar, *J. Appl. Metorol.*, 14, 1338-1345, 1975.

Crane, R.K., Space-time structure of rain rate fields, *J. Geophys. Res.*, 95(D3), 2011-2020, 1990.

Cressie, N.A.C., *Statistics for Spatial Data*, Wiley, New York, 1991.

Dudgeon, D.E. and R.M. Merserau, *Multidimensional Signal Processing*, Prentice Hall, 1984.

Jenkins, G. M. and D. G. Watts., *Spectral analysis and its Applications*, Holden-Day, San Francisco, 1968.

Krajewski, W.F., Cokriging radar-rainfall and rain gage data, *J. Geophys. Res.*, 92(D8), 9571-9580, 1987.

Lanza, L. G., A Conditional Simulation Model of Intermittent Rain Fields, *Hydrological and Earth System Sciences*, Vol 4 No 1, 173-183, 2000.

Menabde, M., D. Harris, A.W. Seed and G. Austin, Self-similar random fields and rainfall simulation, *J. Geophys. Res.* 102(D12), 13509, 1997.

Mueller, A., Sampling of spatial functions: the aliasing problem and statistical errors, *International Conference HYDROSOFT 86*, 1986.

Pegram, G.G.S. Spatial interpolation and mapping of rainfall: 3. Optimal integration of rain gauge, radar & satellite-derived data in the production of daily rainfall maps. Progress Report to the Water research Commission for the period April 2001 to March 2002.

Pegram, G.G.S. and A.N. Clothier, *High Resolution Space-Time Modelling of Rainfall: the "String of Beads" Model*, Journal of Hydrology, 241(1-2), 26-41, 2001.

Pegram, G.G.S. and M.P. Mittermaier, Estimating Rainfall at Low Altitude, at Distance, from Radar CAPPI Data. 4th International Symposium on Hydrologic Applications of Weather Radar, San Diego, California, April 1998.

Pegram, G.G.S.and A.W. Seed, 3-Dimensional Kriging using FFT to Infill Radar Data. *27th EGS Assembly*, Nice, France. April **,** 2002.

Press, W., S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, *Numerical Recipes in C - The art of scientific computing*, 2nd ed. Cambridge University Press, 1992.

Robin, M.L.J., A.L. Gutjahr, E.A. Sudicky and J.L. Wilson, Cross-correlated random field generation with the direct Fourier transform methods, *Water Resources Res.*, 29(7), 2385-2397, 1993.

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

Seed A.W. and G.G.S. Pegram. Using Kriging to Infill Gaps in Radar Data due to Ground Clutter in Real-Time. *Fifth International Symposium on Hydrologic Applications of Weather Radar - Radar Hydrology*, Kyoto, Japan, November, 2001.

Seo, D.-J., W.F. Krajewski and D.S.Bowles, Stochastic interpolation of rainfall data from rain gages and radar using cokriging, *Water Resources Res.*, 26(3), 469-477, 1990.

Vanmarcke, E., *Random Fields, Analysis and Synthesis,* MIT Press, Cambridge, Massachussetts, 1983.

Zimmerman, D.A., G. de Marsily, C.A. Gotway, M.G. Marietta, C.L. Axnes, R.L. Beauheim, R.L. Bras, J. Carrera, G. Dagan, P.B. Davies, D.P.Gallegos, A. Galli, J. Gómez-Hernándes, P. Grindrod, A.L. Gutjahr, P.K. Kitanidis, A.M. Lavenue, D. McLaughlin, S.P. Neuman, B.S. RamaRao, C. Ravenne and Y. Rubin, A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow, *Water Resources Res.*, 34(6), 1373-1413, 1998.