Techniques for Supplementing Solar Radiation Network Data

A Report of Task 9: Solar Radiation and Pyranometry Studies

Volume 1

September 1992

INTERNATIONAL ENERGY AGENCY
Solar Heating & Cooling Programme

THE INTERNATIONAL ENERGY AGENCY
SOLAR HEATING AND COOLING PROGRAMME

International Energy Agency

The International Energy Agency, headquartered in Paris, was formed in November 1974 as an autonomous body within the framework of the Organination for Economic Cooperation and Development to establish cooperation in the area of energy policy. Twenty-one countries are presently members, with the Commission of the European Communities participating under a special arrangement.

Collaboration in the research, development and demonstration of new energy technologies to help reduce dependence on oil and to increase long-term energy security has been an important part of the Agency's programme. The IEA R&D activities are headed by the Committee on Research and Development (CRD) which is supported by a small Secretariat staff. In addition, four Working Parties (in Conservation, Fossil Fuels, Renewable Energy and Fusion) are charged with monitoring the various collaborative energy Agreements, identifying new areas for cooperation and advising the CRD on policy matters.

Solar Heating and Cooling Programme

One of the first collaborative R&D agreements was the IEA Solar Heating and Cooling Programme which was initiated in 1977 to conduct joint projects in active and passive solar technologies, primarily for building applications. The eighteen members of the Programme are:

Australia
Austria
Belgium
Canada
Denmark
European Community
France (observer)
Germany
Finland
Italy
Japan
The Netherlands
New Zealand
Norway
Spain
Sweden
Switzerland
United Kingdom
United States

A total of eighteen projects or "Tasks" have been undertaken since the beginning of the Programme. The overall programme is managed by an Executive Committee composed of one representative from each of the member countries, while the leadership and mangement of the individual Tasks is the responsibility of Operating Agents. These Tasks and their respective Operating Agents are:

*Task 1: Investigation of the Performance of Solar Heating and Cooling Systems - Denmark
*Task 2: Coordination of Research and Development on Solar Heating and Cooling - Japan
*Task 3: Performance Testing of Solar Collectors - United Kingdom
*Task 4: Development of an Insulation Handbook and Instrument Package - United States
*Task 5: Use of Existing Meteorological Information for Solar Energy Application - Sweden
*Task 6: Solar Heating, Cooling, and Hot Water Systems Using Evacuated Collectors - United States
*Task 7: Central Solar Heating Plants with Seasonal Storage - Sweden
*Task 8: Passive and Hybrid Solar Low Energy Buildings - United States
*Task 9: Solar Radiation and Pyranometry Studies - Federal Republic of Germany
*Task 10: Material Research and Testing - Japan
*Task 11: Passive and Hybrid Solar Commercial Buildings - Switzerland
Task 12: Building Energy Analysis and Design Tools for Solar Applications - United States
Task 13: Advanced Solar Low Energy Buildings - Norway
Task 14: Advanced Active Solar Systems - Canada
Task 15: Advanced Central Solar Heating Plants (In Planning Stage)
Task 16: Photovoltaics in Buildings - Germany
Task 17: Measuring and Modeling Spectral Radiation - Germany
Task 18: Advanced Glazing Materials - United Kingdom

*Completed Task

REPORT No. IEA-SHCP-9D-1

TECHNIQUES FOR SUPPLEMENTING SOLAR RADIATION NETWORK DATA

VOLUME 1: REPORT

Antoine Zelenka
Swiss Meteorological Institute, Krähbühlstrasse 58, CH - 8044 Zürich, Switzerland

Gerhard Czeplak
Deutscher Wetterdienst, Meteorologisches Observatorium Hamburg
Frahmredder 95, D - 22393 Hamburg 65, Germany

Vito D'Agostino
SASIAM, Tecnopolis CSATA Novus Ortus, P.O. Box 775, I - 70010 Valenzano / Bari, Italy

Weine Josefsson
Swedish Meteorological and Hydrological Institute, S - 60176 Norrköping, Sweden

Eugene Maxwell
National Renewable Energy Laboratory, 1617 Cole Boulevard, Gloden, CO 80401-3393, U.S.A.

Richard Perez
Atmospheric Siences Research Center, SUNY at Albany, 100 Fuller Road, Albany, NY 12205, U.S.A.

and (Chapter 11)

Maurizio Noia*, Corrado Ratto, and Roberto Festa
Università di Genova, Dipartimento di Física, Via Dodecaneso 33, I - 16146 Genova, Italy
*Prsent Affiliation: AGIP, S.p.a., Drilling Technologies, R&D Dept., Milano, Italy

September 1992

This report documents work performed within the IEA Solar Heating and Cooling Programme
Task 9: Solar Radiation and Pyranometry Sudies
Subtask D: Techniques for Supplementing Solar Radiation Network Data

Additional copies may be ordered from:

ENET
Swiss Federal Office of Energy
CH - 3003 Bern
Switzerland Cost: 60 CHF (3 Volumes)

Distribution category: Unrestricted

Executive Summary

Solar radiation data are required for resource assessment, simulation model development and validation, system design and evaluation, collector testing, and other activities in the field of solar energy research and application. A complex computer model may be useless if accurate input data from a near-by measuring station are not available. On the other hand, the density of radiometric stations is limited for economical reasons. Thus, there is a need for spatially continuous mapping of solar radiation.

Based on some of the results obtained earlier in Task 9, Subtask D has developed and evaluated techniques for estimating solar radiation received at the earth surface at locations between network sites, using both measured and modelled data. In addition to classical statistical techniques, like interpolation and extrapolation of ground-based data, new methods, such as satellite-based techniques, have been investigated.

The experts of Task 9 are well aware of the fact that many solar energy applications need solar radiation data for the direct and diffuse components on time scales of 1 hour or less. However, providing techniques for mapping such data was out of the scope of Subtask D for two reasons: First, the amount of data to be handled and presented would have exceeded task resources. Second and more important, it is not useful to produce synoptic maps of solar radiation data on an hourly time scale because the high temporal variability of solar radiation, due to moving clouds, induces a corresponding high spatial variability.

Therefore, Subtask D has concentrated on studies of the daily totals of global solar radiation. These restrictions do not curtail the relevance and usefulness of the results obtained by the Subtask. The daily scale is well below the so-called synoptic time scale which represents the average persistence of weathers systems like lows and highs. Further, several well-proven algorithms published in the literature allow for generating short-time variability of radiation data from daily global solar radiation.

The simplest method to estimate daily global solar radiation received at a certain location is to use the data of the next radiometric station. This "extrapolation" has a root-mean-square error (RMSE) which was found to increase linearly with the square root of the distance from the station.

Better results were obtained by weighted interpolation of the data from several radiometric stations surrounding the selected site, whereby the weights are proportional to the inverse squared distance. Even more sophisticated is the so-called kriging method, in which the weights are determined in such a way that the RMSE becomes minimal. Further, interpolation by higher order polynomials of the station coordinates, the so-called smooth surface fitting, was investigated.

Estimates of daily global solar radiation derived from METEOSAT (satellite) data were shown to have a mean bias error (MBE) which increases with altitude above mean sea level. The altitude-induced error was compensated by applying the ratio of clear-day insolation at the given altitude over the clear day insolation at sea level as a correction factor. Satellite-derived hourly global solar radiation data were analyzed to determine whether photovoltaics may effectively contribute to a given utility's power generation capacity.

Both ground-based and satellite-derived data were used in the so-called co-kriging method, in which the weights of both interpolations are determined so that the RMSE becomes minimal. This method yielded the best estimates of daily global solar radiation.

Spatial resolution of the global solar radiation field on a regional scale was obtained by taking one or more additional meteorological quantities into account. For a country with a dense network of sunshine duration stations, the so-called cluster analysis of these data led to the delineation of mesoscale sunshine climates, which were converted to global solar radiation fields.

Cloud observations from weather stations were used to interpolate the cloud field and to derive regional global solar radiation values from that continuous data set. Multi-year cloud observations were used to derive typical diurnal variations of cloud amount for each month and station; these data were used to compute the monthly means of daily direct solar radiation which would be available for sun-tracking solar collectors.

CONTENTS

VOLUME 1: REPORT

1. **Description and purpose of the report** 1

   1.0 **Brief description of Task 9: SOLAR RADIATION AND PYRANOMETRY STUDIES** 1

   1.1 **Objectives of Subtask 9D** 1
      1.1.1 *General* 1
      1.1.2 *Specific* 2

   1.2 **Adopted strategy** 4
      1.2.1 *Background* 4
      1.2.2 *Action plan* 4

   1.3 **Limitations** 4
      1.3.1 *Use of daily totals* 4
      1.3.2 *Transfer of satellite-based estimates* 4

   1.4 **Structure of the report** 5
      1.4.1 *Sequence and content of the chapters* 5
      1.4.2 *Referencing* 5

   1.5 *References* 5

2. **Data description** 7

   2.1 **Data format, collection and dispatch** 7

   2.2 **Networks** 7
      2.2.1 *Germany* 7
      2.2.2 *Swedish mesoscale* 8
      2.2.3 *Swiss ANETZ* 8
      2.2.4 *US Northeast Utilities-State University NY* 8
      2.2.5 *US Pacific Northwest-SERI* 9
      2.2.6 *US West Associates Network-SERI* 9

   2.3 **Satellite-based estimates** 9
      2.3.1 *METEOSAT/DWD/Cologne* 10
      2.3.2 *GOES/NESDIS/Tarpley* 10
      2.3.3 *METEOSAT/CTAMN/"Héliosat"* 11

   2.4 *References* 12

3. **Validation of ground-based methods** 31
   3.1 **Nearest neighbor approximation (Extrapolation)** 31
      3.1.1 *Method and presentation of results* 31
      3.1.2 *Discussion of cross-validation results* 49
      3.1.3 *References* 50

   3.2 **Smooth surface fitting** 53

   3.3 **Weighted averages** 55
      3.3.1 *Method and presentation of results* 55
      3.3.2 *Discussion of cross-validation results* 71
      3.3.3 *References* 72

   3.4 **Kriging** 77
      3.4.1 *Basic considerations* 77
      3.4.2 *The variogram* 78

3.4.3 *The estimation procedure* 80
3.4.4 *Presence of drifts (non-stationary kriging)* 80
3.4.5 *Application of the method and results* 82
3.4.6 *Discussion of cross-validation results* 89
3.4.7 *References* 91

3.5 **Principal components analysis** (Empirical Orthogonal Functions) 93
3.5.1 *Validation of significance tests* 93
3.5.2 *Interpretation of the eigenvectors* 95
3.5.3 *Validation of the EOF method* 95
3.5.4 *Mapping and comparison with other methods* 100
3.5.5 *References* 100

4. **Validation of satellite-based methods** 105

4.1 **Principles of satellite-based methods** 105
4.1.1 *Fundamental considerations* 105
4.1.2 *Fundamental equations* 107
4.1.3 *Practical constraints* 107
4.1.4 *Classification and transportability of methods* 108

4.2 **METEOSAT / DWD** 110
4.2.1 *Application field* 110
4.2.2 *DWD network* 110
4.2.3 *Swiss ANETZ* 111

4.3 **GOES / NESDIS** 113
4.3.1 *Application field* 113
4.3.2 *PNW network* 113
4.3.3 *Northeast network* 114

4.4 **METEOSAT / "Héliosat"** 114
4.4.1 *Application field* 114
4.4.2 *Swiss Anetz* 117

4.5 *References* 119

5. **Validation of compound methods** 121

5.1 **Introduction** 121

5.2 **Use of auxiliary information based on cluster analysis** 122
5.2.1 *Introduction* 122
5.2.2 *Angström regression* 122
5.2.3 *Cluster analysis* 123
5.2.4 *References* 123

5.3 **Cokriging** 131
5.3.1 *Introduction* 131
5.3.2 *Data* 131
5.3.3 *The cokriging method* 134
5.3.4 *Results* 136
5.3.5 *Conclusions* 138
5.3.6 *References* 139

5.4 **Use of interpolated cloudiness from gridded meso-β analysis** 147
5.4.1 *Introduction* 147
5.4.2 *Data sets* 148
5.4.3 *Model* 149
5.4.4 *Correcting computed values using a reference station* 150
5.4.5 *Results of the validation* 150
5.4.6 *Applications to the solar radiation network of Sweden* 151
5.4.7 *References* 152

6. End-use validations 167

6.1 Impact of radiation resources on siting of solar thermal power plants 167
6.1.1 *Background* *167*
6.1.2 *The siting methodology -- An overview* *168*
6.1.3 *Factors affecting site selection* *169*
6.1.4 *Producing a mesoscale solar radiation data base* *170*
6.1.5 *Converting cloud cover to direct normal irradiance* *182*
6.1.6 *Evaluating regional site suitability* *187*
6.1.7 *References* *193*

6.2 Assessing photovoltaic interaction with New York summer-peaking utilities -- Resource assessment logistics 195
6.2.1 *Research objectives* *195*
6.2.2 *Evaluation of GOES-derived irradiance data* *196*
6.2.2.1 Origin of satellite-derived irradiance data *196*
6.2.2.2 Generation of hourly irradiance data *197*
6.2.2.3 Satellite vs. ground interpolation for areal coverage: methodology *197*
6.2.2.4 Evaluation of hourly satellite estimates against controlled ground measurements *198*
6.2.2.5 Results for areal coverage *200*
6.2.2.6 Preparation of a radiation data set for southern New York *202*
6.2.3 *Utility load matching assessment* *202*
6.2.3.1 Experimental data *202*
6.2.3.2 Methodology for experimental evaluation of load matching *203*
6.2.3.3 Preliminary data correction *204*
6.2.3.4 Optimum configuration for the fixed-tilt PV array *204*
6.2.3.5 Effective load carrying capability *205*
6.2.3.6 Auxiliary utility load matching benchmarks *205*
6.2.3.7 Summary of end-use satellite data evaluation *207*
6.2.4 *Conclusions* *207*
6.2.5 *References* *208*
6.2.6 *Illustrations* *210*

7. Summary of validation results 249

8. Conclusions and recommendations 255

9. References and textbooks 259

9.1 Overview of previous work 259
9.1.1 *Spatial interpolation of insolation* *259*
9.1.2 *Satellite-based methods* *259*

9.2 General readings 260

9.3 Related references 260

Appendix: List of Participating Experts 261

VOLUME 2: THEORY

10. Theory of spatial interpolation 1

10.1 Probabilistic and deterministic approaches 2

10.2 Description of the deterministic approaches 4
10.2.1 *Weighted averages* 4
10.2.2 *Akima interpolation* 5

10.3 Description of the probabilistic approaches 5
10.3.1 *Gandin's optimum interpolation* 5
10.3.2 *Geostatistics (kriging)* 7
10.3.2.1 Modelling 8
10.3.2.2 Structural analysis 9
10.3.2.3 Performing structural analysis 14
10.3.2.4 Kriging techniques 17
References 22
10.3.3 *Principal components analysis* 24
10.3.3.1 Introduction 24
10.3.3.2 Motivation and initial data 24
10.3.3.3 The covariance matrix 24
10.3.3.4 Calculating the EOFs 24
10.3.3.5 The multipliers 25
10.3.3.6 Mapping 25
References 26

10.4 Cluster analysis 26
10.4.1 *Introduction* 26
10.4.2 *Procedure* 27
10.4.3 *Mathematical background* 27
10.4.4 *Ward's method* 28
10.4.5 *Defining the optimum number of clusters* 28
References 28

10.5 Technological supports 29

11. Theory of satellite-based methods 31

11.1 Introduction 31

11.2 Physical basis of all methods 31

11.3 General description of the methods 33

11.4 Statistical methods 35
11.4.1 *Hay and Hanson model* 36
11.4.2 *Tarpey model* 36
11.4.3 *Justus, Paris and Tarpley model* 40
11.4.4 *Cano model* 41

11.5 Physical methods 44
11.5.1 *Gautier, Diak and Masse model* 44
11.5.2 *Marullo, Dalu and Viola model* 49
11.5.3 *Möser and Raschke model* 53
11.5.4 *Dedieu, Deschamps and Kerr model* 55

11.6 Comparison between models 57

11.7 *Nomenclature* 60

11.8 *References* 62

VOLUME 3: RESULTS

Table 3.1 Detailed cross-validation results for extrapolated daily global irradiation 1

Table 3.2 Detailed cross-validation results for "gravity"-interpolated daily global irradiation 27

Table 3.3 Detailed cross-validation results for kriged daily global irradiation 51

Table 4.1 Validation of METEOSAT-based daily insolation estimates for the DWD network 59

Table 4.2 Validation of METEOSAT-based daily insolation estimates for the ANETZ network 71

Table 4.3 Validation of GOES-based daily insolation estimates for the PNW network 87

Table 4.4 Validation of GOES-based daily insolation estimates for the Northeast network 93

CHAPTER 1

1. Description and purpose of the report

1.0 Brief Decription of Task 9: *SOLAR RADIATION AND PYRANOMETRY STUDIES*

As activity in the field of solar energy research and application increased, a need was identified for more accurate solar radiation and meteorological data to aid in resource assessment, system design and evaluation, and solar collector testing. Task 9 was initiated to address this need and to expand on the work accomplished in earlier IEA meteorological Tasks (4 and 5). The first phase of the Task ran from October 1982 until June 1987 and consisted of three Subtasks, each coordinated by a lead country:

Subtask A: Small-Scale Time and Space Variability of Solar Radiation (Austria)
Subtask B: Validation of Solar Radiation Models (Canada)
Subtask C: Pyranometry (Canada)

The second phase of the Task ran from July 1987 until June 1991. The work is divided into the following Subtasks:

Subtask D: Techniques for Supplementing Solar Radiation Network Data (Switzerland)
Subtask E: Representative Design Years for Solar Energy Applications (Denmark)
Subtask F: Irradiance Measurement for Solar Collector Testing (Canada)

Based on some of the results obtained in Subtasks A and B, Subtask D was evaluating techniques for estimating solar energy radiation resources at locations between network sites, using both measured and modelled data. In addition to the classical statistical techniques for deriving solar radiation data received at the earth's surface (e.g., interpolation and extrapolation), new methods, such as satellite-based techniques, were investigated.

As an alternative to measured meteorological data, several techniques have been developed recently for simulating representative hourly data over the year using statistical techniques. This approach is particularly suited to simplified design techniques since there are no uncontrolled variables. Subtask E was concerned with developing recommendations on the most appropriate methods for producing representative design years as input to solar energy system and building energy system simulation models.

Building on the work of Subtask C, Subtask F focused on demonstrating the improved quality of solar radiation measurements now available. Participants developed guidelines on solar radiation measurement for solar energy engineers and predicted achievable accuracies for such measurements. Further experimental work on characterization of pyranometers were conducted, and the improvement of terrestrial (longwave) radiation measurement for solar collector testing was investigated.

1.1 Objectives of Subtak 9D

1.1.1 General

Numerous recent solar energy and related applications require access to a geographically continuous solar radiation input. For instance, irradiance on inclined planes can be calculated, provided the horizontal irradiance is known. Efficiency of such calculations has been extensively assessed by Subtask 9B [1.1] and by subsequent, still on-going, mostly 9B-triggered investigations [1.2]. The present project originated as a "natural" extension of 9B after recognition of the facts that, (1), however dense a radiation network may be, there will always be a need for simulation model input at locations not coinciding with network nodes and, (2), the use of an accurate (and often complex) simulation model may be meaningless in the absence of accurate input data. Therefore, there is a general need for spatially continuous mapping of solar irradiance. Each expert of the 9D group has approached this general goal as it best fitted his particular research program.

1.1.2 Specific

V. D'Agostino, CSATA / Tecnopolis Novus Ortus, Italy:

SASIAM (School for Advanced Studies in Industrial and Applied Mathematics) in Tecnopolis is involved in research activities on how data collected at different spatial and temporal scales should be combined to obtain global descriptions of environmental variables. Geostatistics as a probabilistic approach offers a set of methods and techniques able to interpret and describe such variables. Implementation of these techniques calls for parallel processor and expert system technologies and, in this sense, it fits the investigation line of *SASIAM*.

G. Czeplak, DWD, Meteorologisches Observatorium Hamburg, Germany:

To overcome the relative sparcity of the German radiometric network, we designed two procedures which are ultimately intended for country-wide mapping of monthly average daily horizontal global irradiation. The first procedure merges global irradiation and sunshine duration data. Sunshine duration is easier to measure and the station density is markedly higher than for irradiation.

We first determine the coefficients of the Ångström regression, which relates normalized irradiation and sunshine duration, for all sites where both quantities are available. Then, cluster analysis is performed to objectively delineate regions with similar sunshine duration climates. If we postulate that the same Ångström coefficients are valid throughout a whole region, the density of sites with known irradiation becomes large enough to allow gravity interpolation or kriging at a scale adequate for mapping.

In a first step of another entirely different procedure, we determine the altitude dependence of normalized irradition by means of regression analysis. This allows us to reduce all observations to a given reference altitude. In the second step we investigate possible relationships of the reduced values with associated land-use classes. Standard significance tests lead to a subdivision of the data in three categories: rural, urban, and forest regions. The third step consists in arrangement of the averaged irradiation data of each category in matrices with the sites in the rows and months in the columns. From these, we derive covariance matrices whose diagonal elements are proportional to the space variance, as well as the related eigenvectors and eigenvalues. The eigenvectors are called "principal components" or "empirical orthogonal functions", to emphazise their being derived from observations. The eigenvalues represent the corresponding amount of explained variance. In our case, we find that the first four eigenvectors explain 99% of the variance of the radiation field. These functions and, consequently, the irradiation they determine, are readily mapped at the scale of the country-covering digital terrain and land-use model. Furthermore, the principal components can be interpreted in terms of radiation meteorology.

W. Josefsson, Swedish Meteorological and Hydrological Institute (SMHI):

The distance between the stations of the solar radiation network in Sweden varies from 150 to 300 km. The daily global radiation normally varies on a much smaller spatial scale. Therefore, it was helpful to obtain irradiation data from a very dense experimental network, located around Norrköping, which could be combined with the regular network.

The combined data were used to validate a method for computing global irradiation which is based on cloud data extracted from the dense grid (22 x 22 km) of the meso-beta analysis fields of SMHI. The cloud data fit the synoptic three-hourly scheme and is given in octa for three levels (low, middle and high). The simple model generates hourly values which are of interest to many applications.

This approach has a great development potential. It can be improved by using other available information, in addition to cloudiness. Several countries in northern Europe plan a high resolution limited area model (HIRLAM) that will include not only fields of analysis but also fields of forecast. Radiation forecasts are, for example, an important input to models predicting consumption of heat and energy in buildings. There are also plans to incorporate cloud information from satellites in the analysis.

E. Maxwell, Solar Energy Research Institute, U.S.A.:

In large areas of the United States and some other countries, the physical and climatological distances between measurement stations preclude the initial use of mathematical interpolation methods, since these methods are dependent upon the existence of spatial correlations between the measurements made at adjacent or nearby locations. Without these correlations, the mathematical interpolation methods produce data that contain artifacts of the methodologies, rather than data that are related to the climate and topography of the region. For those countries faced with this situation, satellite methods and/or simulation models must first be employed to estimate solar irradiance at a sufficient number of locations to represent the general topographical and climatological changes within the country. Use of these methods can then be followed by mathematical interpolation if a greater spatial resolution is desired.

R. Perez, ASRC, State University of New York at Albany, U.S.A.:

Geographically continuous solar radiation data are essential to properly assess/monitor the impact of dispersed photovoltaic power production on the impact of various daylighting strategies on the load of electric utilities.

Here, we take initial steps in providing the needed spatially continuous irradiance data that may be subsequently processed via appropriate models [1.2] into the needed solar or daylight system input data.

We address this initial objective by evaluating currently available methodologies relying on ground-based networks and geostationary satellites. Beyond the physical evaluation of these methods, our concern extends to the assessment of their end-use accuracy, that is their ability to address the objectives of data users. In particular, we inquire about their potential to provide utilities with a proper assessment of the impact of photovoltaic penetration on their service area.

A. Zelenka, Swiss Meteorological Institute:

Within the Swiss automatic meteorological network ANETZ, about four dozen sites can be assimilated to nodes of a network measuring global horizontal irradiation. In comparison to some other radiation monitoring networks, the station density may appear to be rather high. However, the Alps and major lakes generate a variety of micro-climates, which in turn, imply an extremely variable insolation field, even for horizontal distances as small as one kilometer. Thus, mapping useful to solar system design ought to be preceeded by identification of the most appropriate technique for supplementing the network data.

Purely network-based interpolation techniques, i.e. without external constraints such as orography etc., are doomed to failure because they filter out all variations with scales smaller than the separation of the pair of mutually closets stations. Nevertheless, determining the magnitude of the failure is of practical interest.

The spatial resolution of insolation estimates based on the imagery of the geostationary satellite METEOSAT corresponds to the pixel size, which is, for Switzerland, 2km in longitude by 4km in latitude. These estimates are able to render the spatial variability of the insolation field much better than the network. Yet, they have several shortcomings and substantial effort was devoted to identifying and tentatively correcting these.

Eventually, the best supplementing technique was identified as cokriging with the (accurate but under sampled) network data together with the (less accurate but better sampled) METEOSAT derived, empirically corrected estimates.

1.2 Adopted strategy

1.2.1 Background

Clearly, the main focus is on the experimental evaluation and comparison of different methodologies for estimating global horizontal irradiation when neither this parameter nor any meteorological quantity from which it could be derived is measured. In some instances, our inquiry also pertains to the direct beam as well as global irradiation.

The methodologies include two basic categories as well as their hybrids. The basic categories are (1) the so called "ground-based" methods, involving different interpolation techniques within a monitoring network, and (2) the "satellite-based" methodologies. The hybrids integrate information from both categories.

Some of the algorithms also use cloud observations or sunshine duration. However, when this happens, it is to increase the number of sites with "known" insolation and, thus, to render interpolation more accurate rather than to avoid it. Interpolation of cloud observations with subsequent processing via an appropriate model [1.3] is another concievable scheme [1.4] where interpolation methodology plays a major role.

1.2.2 Action plan

The cooperative work plan included the following steps: (1) Preparing select data sets from national or regional radiation networks; (2) identifying, reviewing and selecting ground-based interpolation methodologies; (3) reviewing currently operational satellite-based methods; (4) validating and comparing all methodologies; and (5) formulating recommendations. As to this last step, it was not expected that one unique technique would emerge as a *panacea.* Rather, recommendations must be based on cost-benefit analyses of each procedure.

1.3 Limitations

1.3.1 Use of daily totals

To reach the goals assigned in 1.1 we would need to estimate at least hourly horizontal irradiance. However, to cope with limited time and funding, evaluation of the methodologies was (mostly) restricted to daily totals. This is not a serious drawback. From the point of view of insolation meteorology, the daily scale is well below the synoptic time scale (persistence of weather patterns) and, *a fortiori,* well below the climatological scale (long term means). Also, satellite-based estimation methods are, today, primarily focused on daily totals. Even with this restriction, the available resources were not sufficient to evaluate all techniques for all networks.

1.3.2 Transfer of satellite-based estimates

Almost any given interpolation technique may be tested on any given network. Only insufficient spatial density of measuring sites may prevent one from doing so. On the contrary, estimates based on a given satellite method can be validated only with measurements from sites located within the useful field of view of the satellite. The models which convert satellite data into irradiances at the Earth's surface depend upon the spectral properties of the satellite's sensor, as well as upon its resolution. Thus, a method which was conceived for the GOES spacecraft to yield estimates for the American continent, cannot be transfered one-to-one to the METEOSAT platform to yield estimates for Europe and Africa. Although we report about such an experiment, we could not investigate methodology transfers among satellites within the frame of this project. Moreover, we must confine our study to readily available, i.e. operational, methods.

1.4 Structure of the report

1.4.1 Sequence and content of the chapters

After a brief overwiew of the data gathered for validation purposes (Chapter 2), we introduce each interpolation method in a rather heuristic way, starting with first principles (Chapter 3). At each step, performance of the method is documented by cross-validation as far as available.

Cross-validation is the current technique for assessment of spatial interpolation accuracy. One at a time, each station is artificially removed from the network. Interpolation using the remaining stations yields a predicted value for the missing station. This predicted value is then compared to the actual, measured value. The Mean Bias Error (MBE) and Root Mean Square Error (RMSE) for all stations provide a measure of the accuracy of each method. The satellite-based methods are introduced and evaluated in the same manner (Chapter 4) except that usual validation replaces cross-validation.

After an excursion towards methods which merge network and satellite measurement as well as methods which use supplemental information sources (Chapter 5), we present several examples of practical applications (end-use validations) in Chapter 6. Finally, Chapters 7 and 8 summarize and discuss the results.

Theory pertaining to spatial interpolation and to satellite-based estimation is delivered in two chapters (10 and 11) that are deliberately placed in a second volume. Questions which may arise during studies of the other chapters will, hopefully, all be answered there.

1.4.2 Referencing

References specific to a chapter, or even to a voluminous section, are displayed immediately at the end of the pertaining text. This makes referencing somewhat more cumbersome but speeds up reference title consultation. References to investigations or to textbooks of general interest are grouped in a final Chapter 9 of Volume 1, where a special section is also devoted to previous work in the field, which was often a valuable guide to our investigations.

1.5 References

[1.1] J.E. Hay: 1986, "Calculation of solar irradiances for inclined surfaces: verification of models which use hourly and daily data". Final Rept. for the IEA Solar Heating and Cooling Programme Task 9, Phase I, Subtask B. Canadian Climate Centre, Atmospheric Environment Service, Downsview, Ontario M3H 5T4.

[1.2] R. Perez, P. Ineichen, R. Seals, J. Michalsky, and R. Stewart: 1990, "Modeling daylight availability and irradiance components from direct and global irradiance". *Solar Energy* **44**, 271-289.

[1.3] J.A. Davies, D.C. McKay, G. Luciani, and M. Abdel-Wahab: 1988. "Validation of modesl for estimating solar radiation on horizontal surfaces", Final Rept. for IEA Solar Heating and Cooling Programme Task 9, Phase I, Subtask B. Canadian Climate Centre, Atmospheric Environment Service, Downsview, Ontario M3H 5T4.

[1.4] A. Zelenka and D. Lazic: 1987, "Supplementing network global irradiance data". Advances in Solar Energy Technology, Proc. 1987 Biennial Congres ISES, Hamburg. Vol. 4, pp. 3861-3865, Pergamon 1988.

CHAPTER 2

2. Data description

2.1 Data format, collection and dispatch

In a first step, the experts had to prepare a data set with which performance of select interpolation methods could be assessed. Available network data and corresponding satellite estimations were prepared in a uniform, self-explaining format. One file per network was supplied, plus, where available, a corresponding file for satellite estimates. Table 2.1 displays the format and collation sequence.

Sunshine duration was included to aid selection of specific conditions for the German and Swiss networks. It was decided that top of the atmosphere values for insolation and sunshine duration had to be calculated by the same algorithm [2.13] in all cases. This work was done in Zurich, where all the files were collected, supplemented and merged to the common validation data set which was, then, re-distributed among the participants. The schedule of these activities is reproduced in Table 2.2.

It was decided that one year of daily sums for each network would yield a tractable amount of bytes. At the beginning, an effort was undertaken to homogenize the period to the year 1988, but this could be realized only in half of the cases. Also, it was not possible to gather a full year of data everywhere. Details about the full content of the final validation data set are to be found in Tables 2.3. and 2.4.

2.2 Networks

The networks are succintly described in the following subsections. Table 2.3 displays their main characteristics. There are many ways to define the mean horizontal distance within a network. The one selected for Table 2.3 is the mean of the distances to the closest neighbour. When locations of the network stations are not given here, Chapter 3 should be consulted.

It is implied that all data have been quality-controlled according to the routine rules of the suppliers' institutions. In the following, this point is merely evoked, if at all, with a reference, while concentrating primarily on noteworthy characteristics of the networks.

2.2.1 *Germany*

The Meteorologisches Observatorium Hamburg (MOH) of Deutscher Wetterdienst (DWD)has operated a solar radiation monitoring network in the former Federal Republic of Germany since 1949. The number of stations was increased in 1958 and 1972 and since 1980, 29 stations (Fig. 2.1 and Table 2.5) record both global radiation (G) and sunshine duration (S). Additionally, 15 stations are also measuring diffuse radiation (D) [2.1]. The maximum distance between the stations is less than 200 km. Most sites are equipped with Kipp & Zonen CM 11 ventilated and wintertime-heated pyranometers, but a few stations still use CM 6 instruments. All stations are manned so that controlling of the pyranometers and checking of the data acquisition and processing systems can be done daily.

The sites have been carefully selected. Elevation of the entire horizon does not exceed 5° and no obstacles are tolerated within the sun's diurnal arc. Inspection of the stations by members of the radiation center of MOH and replacement of the pyranometers by freshly calibrated ones are done at least every second year.

The stations supply MOH with hourly sums of G, D, and S every month. MOH checks, assembles and converts the data to an archive format. A data book is issued yearly with a condensed summary of all data (daily sums and monthly mean hourly sums) [2.2].

2.2.2 *Swedish mesoscale*

The network is actually twofold. One is the experimental dense network around Norrköping (where the SMHI main office is located) while the other is part of the national solar radiation network. All stations are automatic. The locations are plotted in Fig. 2.2 and they are listed in Table 2.6. All pyranometers are ventilated. Kipp & Zonen CM 5 are used in the experimental network and CM 11 in the other network. Further details on data and the network are given in [2.3] and [2.4].

2.2.3 *Swiss ANETZ*

The pyranometers equipping the stations of the Swiss automatic meteorological network ANETZ are Kipp CM5, ventilated instruments. Mode of operations and calibrations are briefly discussed in [2.5]. One important feature is that the instruments are unattended for extended periods of time with all hazards that this type of operation encompasses. However, occasional comparisons between measurements with attended and unattended instruments, done with pyranometers at several stations, indicate that the network data are generally rather reliable [2.6, 2.7].

Care has been taken to select, among the 60 available for the chosen period, only those sites with acceptable views of the horizon. Masking of the sun's diurnal arc by obstacles was not tolerated if it resulted in a loss of more than 2% of the average possible daily total insolation. Other probable source of inaccuracies have not been assessed in detail. Among these, possible effects of the thick poles on which the mountain top instruments are fixedmay be most significant.

The 41 stations fulfilling the selection criterion and, consequently, retained for this study (see Fig. 2.3 and Table 2.7), are not as evenly distributed as desired for straightforward mapping purposes. A dearth of valley sites is obvious, especially in the southeastern Alps. This has little consequence as long as only the estimation methodologies are investigated. For the mapping procedure itself, the selction criteria may be relaxed within reasonable limits to supplement the basic set with enough relevant stations.

2.2.4 *US Northeast Utilities-State University NY*

A seven-station ground network was assembled from existing data [2.8]. This covers a triangular area extending from Albany and New York City, NY, to southeastern Massachusetts. It was possible to assemble such a network because three distinct measurement campaigns happened to be coincident in time: (1) data from the Northeast Utilities network were made available for this project (Waterford, Middleton and Milford, Connecticut); (2) data from the Woods Hole Oceanographic Institution in Woods Hole, MA; and (3) data from Albany, Queens and Farmingdale, NY, collected by the authors under funding from the New York State Energy Research and Development Authority (NYSERDA).

A total of 134 daily total global irradiance measurements from October 1986 to July 1987, for each of the seven sites, were assembled after screening for quality control and time coincidence between stations. Calibration coherence within the network was attempted by performing clear day analysis and adjusting data in reference to the New York instruments.

Fig. 2.4 and Table 2.8 describe the network sites.

2.2.5 *US Pacific Northwest-SERI*

The University of Oregon Solar Monitoring Laboratory has been operating a solar radiation network in the states of Oregon, Washington, and Idaho since 1977. Support for the network operation has come from the Eugene (Oregon) Water and Electric Board, the Bonneville Power Administration, the U.S. Department of Energy, and certain state agencies. The University of Oregon provided SERI with daily total global horizontal insolation data for 1984 for 9 sites in Oregon. Table 2.9 and Fig. 2.5 give the names and locations of the stations. Data acquisition and quality control are described in [2.9].

The mountainous terrain in Oregon and the limited number of stations restrict the application of mathematical interpolation methods. Therefore, satellite estimates of daily total insolation were obtained from NOAA-NESDIS for the 1° x 1° latitude-longitude grid covering Oregon (see Fig. 2.5 as well as Section 2.3.2). The Solar Energy Research Institute assembled and converted these data sets to the units and format required by Subtask 9D. A preliminary comparison of the surface and satellite data showed reasonable agreement. Exact agreement would not be expected since the satellite estimates are obtained from an analisys of pixels over a 50 km x 50 km region around the 1° x 1° grid positions. In mountain and coastal locations such as those in Oregon, the solar irradiance can change a great deal within 50 km. It should be noted that the NOAA-NESDIS satellite estimates of solar irradiance were designed primarily to be a service to agriculture.

2.2.6 *US West Associates Network-SERI*

In 1975, Southern California Edison (SCE) initiated solar radiation monitoring in Southern California. Subsequently, SCE organized the solar monitoring program of the Western Energy Supply and Transmission (WEST) Associates. Eleven of the members of that organization participated in the solar monitoring program, at stations in six southwestern states. Only in southern California was the station density adequate to support mathematical interpolation between stations. SERI has all of the solar radiation data collected by WEST Associates from 1976 through 1980. The solar monitoring program was discontinued at most locations after 1980. Data from 1979 were selected from 12 California stations, and stations at Las Vegas, Nevada, and Page, Arizona. The locations of the stations are given in Table 2.10 and are shown in Fig. 2.5. These stations are separated by distances as small as 7.5 km (Alhambra and Los Angeles) and as large as 333 km (Las Vegas and Page). This provided a wide range of conditions for evaluating mathematical interpolation methods.

2.3 Satellite-based estimates

In Section 2.2 familiarity with pyranometry was implicitly assumed. The same familiarity cannot be assumed with satellite-based measurements. Chapters 4 and 10 describe, at different depth levels, the principles which guide conversion of the satellite data into insolation estimates. Here, we only briefly describe what must be known about those estimates that are part of the validation data sets.

The images are first collected and processed at the respective satellite operations centers, from which they are disseminated in two different forms (digital and analog), again via the satellite. Reception of the digital signal requires an expensive and manpower extensive station, while the analog, less accurate, signal can be received with cheaper, standard electronics. The latter signal follows the "weather facsimile", WEFAX format.

Each image consists of "pixels" (picture elements) that are defined by the angular resolution (aperture) of the scanning radiometer on board the satellite. The geographical size (e.g. in km) of a pixel depends upon its location at the Earth's surface with respect to the sub-satellite point. Insolation estimates are to be understood as average values for the pixel's area.

2.3.1 *METEOSAT/DWD/Cologne*

Operations control is at ESOC in Darmstadt, Germany. Reception (and storage) of the digital and analog signals requires so-called Primary Data User Stations (PDUS) and Secondary Data User Stations (SDUS) respectively.

The digital images are received by the PDUS at the German Weather Service (DWD) in Offenbach and, then, treated essentially following the model of Möser and Raschke [2.13, see also Chapter 10], originally developed at the University of Cologne, but subsequently modified in several steps in collaboration with DWD to be rendered operational [2.11].

An hourly sequence of 5 images, centered about 11 UTC is required for the calculation of a daily insolation sum. When added to other contingencies inherent to the satellite methods (transmission, sensor response, system failures etc.), this condition, in our case of 1988, restricted the number of processed daily sums to 148, and the useful interval from March until October.

The values included to the data sets from Germany and Switzerland are those pertaining to the pixels within which the network stations are located. Pixel size is about 8km (in W-E) by 8km in (S-N). The value assigned to a pixel is not exactly the outcome of the model for that very pixel. Rather, weighted spatial averaging is performed over the 12 neighboring pixels (see Fig. 4.3 in Chapter 4).

The data has been supplied by G. Czeplak, DWD, Met. Obs. Hamburg, resp. by W. Benesch, DWD, Offenbach.

2.3.2 *GOES/NESDIS/Tarpley*

Operations control is in NOAA's CDA at Wallops Island, VA. The retransmitted digital signal is called "stretched VISSR" and it is available to any user with an appropriate ground terminal (equivalent to the METEOSAT PDUS). The retransmitted analog signal is in WEFAX format.

The digital images are accessed by the Satellite Applications Laboratory of NESDIS in Washington, D.C. from a 24-hour rotating data base maintained on the NOAA Central Computing Facilty and, then, are processed following the model of Tarpley and Justus [2.12, see also Chapter 11].

Standard operations are based on two GOES spacecrafts. For the western United States (104-125 deg. W longitude), six GOES-West images between 16:45 and 23:45 UTC are used in the product. For the eastern United States, seven GOES-East images between 13:00 and 23:00 UTC are accessed. The hourly estimates are interpolated to the remaining daylight hours and the resulting hourly values summed to give the daily insolation. From the failure of the GOES-East platform in August 1984 and until its replacement in March 1987, the slightly eastwards displaced GOES-W provided the coverage of the whole American continent. Hurricane monitoring made seasonal shifts of position necessary, thus rendering geographical positioning of the pixels more tedious. A similar situation arose in January 1989 as the GOES-W decayed and the new GOES-E had to take over the entire monitoring task.

The NESDIS insolation estimates are for an array of 1° x 1° latitute-longitude locations through the continental US. The so-called target areas over which the estimate are determined consist of 5x5 8-km resolution pixels. Bilinear interpolation is recommended for interpolating values between the grid locations. Grid locations above water surfaces have an assigned value of zero. Figs. 2.4 and 2.5 show the satellite grid points in the Northeast and the Pacific Northwest networks respectively. Routine estimations in the present form began in1982. The whole estimation program (AgRISTARS) was initiated in 1980. Thus, no estimations are available for the Southwest network in 1979.

In the Northwest, coverage is sufficiently uniform to allow application of the recommended bilinear interpolation for inference of the values at the surface network sites; these inferred values have been included in the network's data set. The situation is less straightforward for the Northeast. No attempt has been made to include inferred values to the network's data set. But, in both cases, the satellite data files, i.e. the NESDIS estimates at the grid points, have been merged into the validation data set to allow various retrieval schemes to be tested. Table 2.4 gives the main characteristics of these files.

The data for the Northwest and Northeast were supplied to SERI and to ASRC respectively by D. Tarpley of NESDIS.

2.3.3 *METEOSAT/CTAMN/Héliosat*

The "Héliosat" system relies on the METEOSAT WEFAX image dissemination, that is, on a Secondary Data User Station (SDUS). It was developed by the Centre de Télédétection et d'Analyse des Milieux Naturels (CTAMN), École des Mines de Paris in Sophia-Antipolis after recognition that methods based on digital (primary) data are mostly operated within national meteorological organizations with large computing means and that they, therefore, "are not suitable for use and operation by people involved in the solar energy business" [2.14]. Requirements that the system had to meet were low cost, simplicity, and ease in use and maintainance. Since 1988, the commercially available version [2.15] of the Héliosat system has been comprised of:

-an HF receiver with antenna to collect the analog signal (WEFAX format),
-a PC with AT compatible bus,
-a satellite signal digitizing unit (to re-digitize the analog signal for further numerical processing),
-a PC graphic board to display color maps on a RGB monitor,
-software to convert satellite data into maps of global irradiation (based on the Cano *et al.* method [2.16, see also Chapter 11]),
-software providing several features in image processing.

Such a station was installed at the Swiss Meteorological Institute within the frame of subtask 9D. The software was purchased from CTAMN which also acted as consultant for the hardware evaluation and ordering. A close collaboration was initiated because the system was to be tuned to hitherto untested operation conditions, i.e., full spatial resolution and a narrow geographical area. Both conditions placed stringent demands on the system for successful operation amidst the Swiss topography and they called for software adaptations that could only be done by CTAMN. This, as well as further tuning due to deficiencies of the data acquisition line and hardware failures (the digitizing unit had to be returned to the manufacturer) took the whole year 1990, so that routine operational tests began only in February 1991.

Daily irradiation maps of 512 times 512 pixels are stored in files of approx. 260kBytes.

2.4 References

[2.1] F. Kasten, K. Dehne, H.D. Behr, and U. Bergholter: 1984. "Die räumliche und zeitliche Verteilung der diffusen und direkten Sonnenstrahlung in der Bundesrepublik Deutschland". *Rept. BMFT-FB-T84-125*, Deutscher Wetterdienst, Meteorologisches Observatorium Hamburg.

[2.2] Deutscher Wetterdienst: 1988. "Ergebnisse von Strahlungsmessungen in der Bundesrepublik Deutschland sowie von speziellen Messreihen am Meteorologischen Observatorium Hamburg". DWD, Meteorologisches Observatorium Hamburg.

[2.3] W. Josefsson: 1984. "The solar radiation measuring program in Sweden". Proc. IEA Solar R&D Task IX Symposium *Recent Advances in Pyranometry*, Norrköping, Jan. 1984, D. Wardle & D.C. McKay Eds., pp. 41-48.

[2.4] W. Josefsson: 1989. "Computed global radiation using interpolated, gridded cloudiness from the meso-beta analysis compared to measured global radiation". Rept. *SMHI Meteorologi* **No. 101**, Norrköping.

[2.5] A. Zelenka: 1984. "The Swiss automatic network". Proc. IEA Solar R&D Task IX Symposium *Recent Advances in Pyranometry*, Norrköping, Jan. 1984, D. Wardle & D.C. McKay Eds., p. 13.

[2.6] A. Zelenka: 1984. "Accuracy checks on pyranometers in the Swiss ANETZ network". Proc. IEA Solar R&D Task IX Symposium *Recent Advances in Pyranometry*, Norrköping, Jan. 1984, D. Wardle & D.C. McKay Eds., pp. 218-219.

[2.7] A. Heimo, 1985. "Quality control of the Swiss network radiation data", *Arbeitsbericht der Schweiz. Meteorol. Anstalt* **No 134**, Zürich.

[2.8] R. Perez, R. Seals, and R. Stewart: 1990. "Assessing photovoltaic interaction with New York summer-peaking utilities: An evaluation of existing solar resource assessment capabilities from measuring networks and geostationary satellites". Rept. to the New York Power Authority, Atmospheric Sciences Research Center, State University of NY at Albany, Albany NY 12222. See also Section 6.2.

[2.9] F. Vignola: 1989. "Pacific Northwest solar radiation data". Solar Monitoring Lab., Phys. Dept., Univ. of Oregon, Eugene, OR.

[2.10] W. Möser and E. Raschke: 1984. "Incident solar radiation over Europe estimated from METEOSAT data", *Jour. Climat & Applied Meteorol.* **23**, 166.

[2.11] F.-J. Dieckmann, S. Happ, M. Rieland, W. Benesch, G. Czeplak, and F. Kasten: 1988. An operational estimate of global solar irradiance at ground level from METEOSAT data: results from 1985 to 1987. *Meteorol. Rdsch.* **41**, 65-79.

[2.12] C.G. Justus, M.V. Paris, and J.D. Tarpley: 1986, "Satellite-measured insolation in the United-States, Mexico, and South America". *Remote Sensing of Environment* **20**, 57-83.

[2.13] J.J. Michalsky: 1988, "The *Astronomical Almanac*'s algorithm for approximate solar position (1950-2050)". *Solar Energy* **40**, 227-235. Errata *Solar Energy* **41**, 113, 1988 and *Solar Energy* **43**, 323, 1989.

[2.14] L. Diabaté, G. Moussu, and L. Wald: 1989, "Description of an operational tool for determining global solar radiation at ground using geostationary satellite images". *Solar Energy* **42**, 201-207.

[2.15] L. Diabaté, G. Moussu, and L. Wald: 1988, "An operational tool for the fine-scale mapping of the incident solar radiation using satellite images: the Heliosat station", *Solar'88*, Proc. ASES Annual Meeting, June 1988, Cambridge MA, M.J. Coleman Ed., pp.11-17.

[2.16] D. Cano, J.M. Monget, M. Albuisson, H. Guillard, N. Regas, and L. Wald: 1986, "A method for the determination of the global solar radiation from meteorological satellite data". *Solar Energy* **37**, 31-39.

Table 2.1

Format description of validation data set.

Tape : 9 Tracks, ASCII coded, 6250 bpi.

Blocking:
50 characters per record.
60 records per block = 3000 characters per block.

Record image:

character from-to	1-5	6-7	8-10	11-14	15-19	20-23	24-27	28-31	32-35	36-38	39-41	42-50
Content	WMO-ID	YEAR	DAY	LAT	LONG	ALT	GH	G0	GSAT	SD	S0	-----
number of characters	5	2	3	4	5	4	4	4	4	3	3	9

Reading FORTRAN format (1I5,1I2,1I3,1I4,1I5,4I4,2I3,[9X])

Sequence of records: Station i...365 resp. 366 days; Station i+1...365 resp. 366 days; etc...

Meaning of content:

WMO-ID...WMO station identifier 'IIiii', where II stands for country or region.*
FRGermany II=10, Switzerland II=06, Sweden II=02, USA II=72 to 74.
YEAR in 20th century; e.g. 88.
DAY ...Daynumber in YEAR (1 to 366).
LAT ...Station's latitude in hundredths of degrees. No sign (always positve).
LONG. ... " longitude " " " " No sign (positive only for USA).
ALT ... " altitude a.s.l. in meter. (If WMO-ID IIiii=00000 -see(*)- then set ALT to 9999).
GH ...Daysum global horizontal irradiance (Jcm^{-2} resp. hunderdth of MJm^{-2}) measured.
Missing value = 9999
G0 ...Daysum global horizontal irradiance (Jcm^{-2} resp. hunderdth of MJm^{-2}) extraterrestrial.
GSAT ...Daysum global horizontal irradiance (Jcm^{-2} resp. hunderdth of MJm^{-2}) satellite estimate.
SD ...Sunshine duration in tenth of hours measured. *Missing value = 999*
S0 ...Sunshine duration in tenth of hours possible (astronomical).

*) Identifier is not WMO standard for:
a) USA radiation networks under consideration as they are not integrated in national meteo network. Then set: IIi = -72, and the last two digits ii: 00-49 for US/SW, 50-79 for US/NW, and 80-99 for US/NE.
b) Sites with satellite estimates alone. Then set IIiii=00000 (and ALT to default value *9999*).

Table 2.2

Summary of main actions during elaboration of the validation data.

Item	Action	Date
Test file DWD March 88	Magnetic tape (MT) from Hamburg to Zurich	DEC 88
Test collection & dispatching	Merge DWD test file with ANETZ March 88 file, homogenise, send MT to Bari and Hamburg	JAN 89
Test readability of MTs	Tapes OK in Bari and Hamburg	JAN 89
SMHI net 88	MT from Norrköping to Zurich	MAR 89
DWD net 88, without METEOSAT	MT from Hamburg to Zurich	MAR 89
USSWN net 79 PNW net 84 + GOES	MT from Golden CO to Zurich	MAY 89
PROVISIONAL SET	Merge with ANETZ net 88, homogenise, MTs to Albany, Bari, Golden, Hamburg and Norrköping	JUN 89
DWD net 88 corrected	MT from Hamburg to Zurich	JUN 89
USNEU net 86/87 + GOES	MT from Albany NY to Zurich	JAN 90
METEOSAT for DWD & ANETZ nets 88	MT from Hamburg to Zurich	MAR 90
SMHI net 88 corrected		APR 90
DEFINIVE SET	Homogenize, merge or replace. Send MTs to Albany, Bari, Golden, Hamburg and Norrköping.	**APR 90**

Table 2.3

Network characteristics

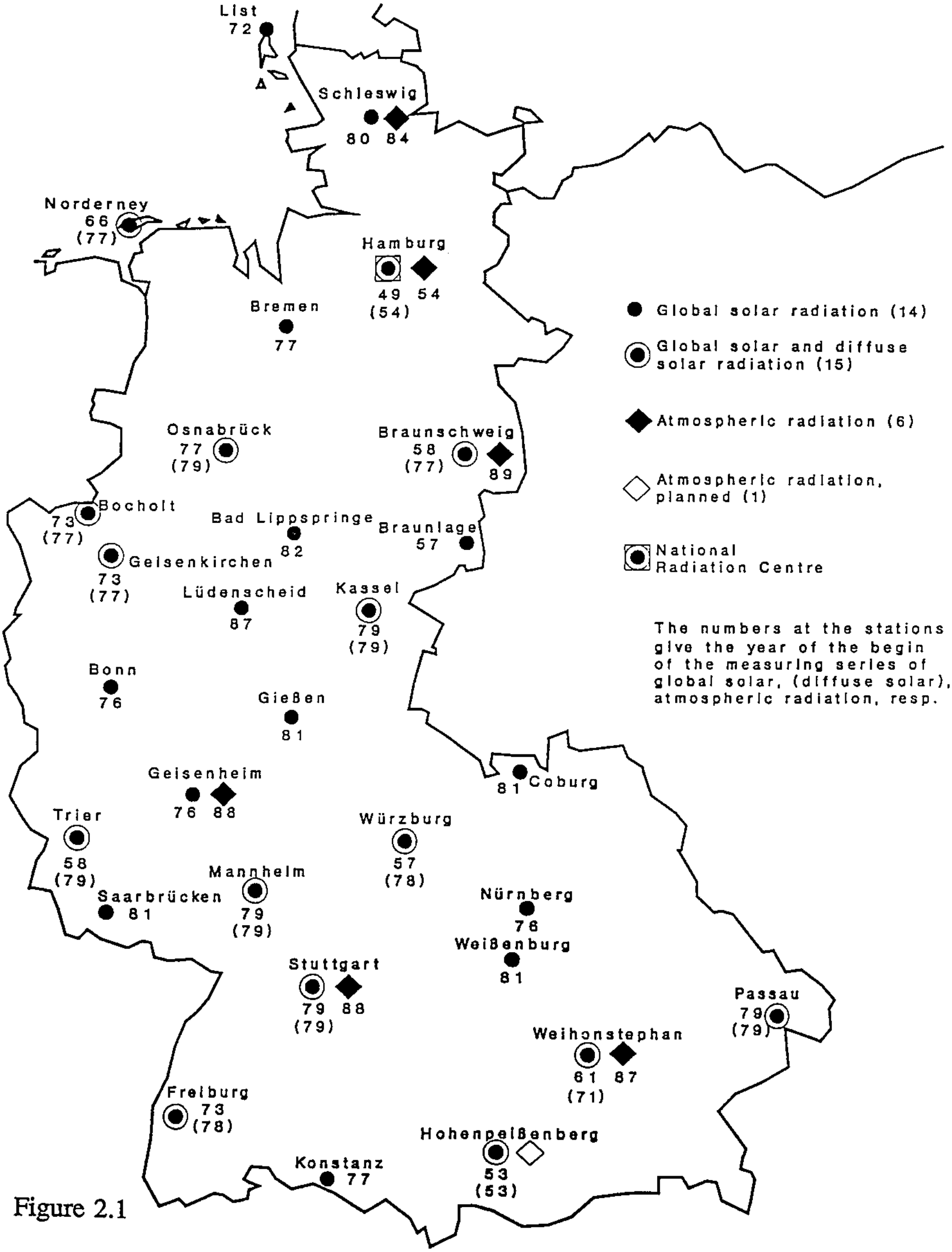
Network	Shortname	Period	No. of stations	No. of observed daily totals	Horizontal station separation [km] mean/min./max.	Altitudes [m] min. / max.
Pacific Northwest / SERI	PNW	JAN 84 - DEC 84	9	2993	120 / 91 / 147	35 / 1325
Southwest/SERI	USSWN	JAN 79 - DEC 79	14	5005	86 / 7 / 173	40 / 1000
USSWN without "Las Vegas" & "Page"	USWN	JAN 79 - DEC 79	12	4335	53 / 7 / 173	40 / 900
Northeast/SUNY	USNEU	OCT 86 - JUL 87	7	938	71 / 31 / 165	15 / 87
German Weather Service "DWD"	DWD	JAN 88 - DEC 88	29	10539	80 / 53 / 132	26 / 990
Swedish Meteo & Hydrol. Inst.	SMHI	APR 88 - SEP 88	15	2573	34 / 16 / 89	4 / 107
Swiss Meteorol. Institute	ANETZ	JAN 88 - DEC 88	41	12687	21 / 3 / 46	192 / 3315

Table 2.4

Characteristics of GOES/NESDIS 1°x 1° grids.
(Locations above water surface are not considered as grid points).

Network subtended	Shortname	Period	No. of grid points	No. of estimated daily totals
Pacific Northwest / SERI	PNW	JAN 84 - DEC 84	38	12768
Northeast/SUNY	USNEU	OCT 86 - JUL 87 (134 days)	12	1608

Radiation Network of Deutscher Wetterdienst (as of 31 Dec 1989)



The numbers at the stations give the year of the begin of the measuring series of global solar, (diffuse solar), atmospheric radiation, resp.

Figure 2.1

Table 2.5

Radiation Network of Deutscher Wetterdienst
(as of 31. Dec. 1989)

Station	geographical co-ordinates			hourly data available since		
	ϕ	λ	H	G	D	A
List	55°01'N	08°25'E	33 m	8/72		
Schleswig	54°32'N	09°33'E	59 m	10/80		12/84
Norderney	53°43'N	07°09'E	29 m	11/66	11/77	
Hamburg	53°39'N	10°07'E	49 m	7/49	1/64	1/64
Bremen	53°03'N	08°48'E	24 m	1/77		
Braunschweig	52°18'N	10°27'E	83 m	7/62	10/77	1/89
Osnabrück	52°15'N	08°03'E	104 m	3/77	4/79	
Bocholt	51°50'N	06°32'E	24 m	2/73	12/77	
Bad Lippspringe	51°47'N	08°50'E	162 m	1/82		
Braunlage	51°43'N	10°37'E	615 m	7/57		
Gelsenkirchen	51°30'N	07°05'E	63 m	4/74	11/77	
Kassel	51°18'N	09°27'E	237 m	2/79	2/79	
Lüdenscheid	51°13'N	07°38'E	465 m	1/87		
Bonn	50°42'N	07°09'E	65 m	5/76		
Gießen	50°35'N	08°42'E	201 m	1/81		
Bad Nauheim *	50°22'N	08°45'E	177 m	11/71		
Coburg	50°16'N	10°57'E	357 m	11/81		
Geisenheim	49°59'N	07°57'E	131 m	5/76		4/88
Würzburg	49°46'N	09°58'E	275 m	5/57	2/78	
Trier	49°45'N	06°40'E	278 m	1/64	1/79	
Mannheim	49°31'N	08°33'E	106 m	1/79	2/79	
Nürnberg	49°30'N	11°05'E	312 m	6/76		
Saarbrücken	49°13'N	07°07'E	325 m	8/81		
Weißenburg	49°01'N	10°58'E	428 m	9/81		
Stuttgart	48°50'N	09°12'E	318 m	10/79	10/79	5/88
Passau	48°35'N	13°28'E	412 m	3/79	3/79	
Weihenstephan	48°24'N	11°43'E	469 m	1/61	12/71	10/87
Freiburg	48°00'N	07°51'E	308 m	1/73	2/78	
Hohenpeißenberg	47°48'N	11°01'E	990 m	1/53	4/53	
Konstanz	47°41'N	09°11'E	450 m	10/77		

* closed end of 4/77

H: height of the instrument above msl
G: global radiation
D: diffuse solar radiation
A: atmospheric radiation

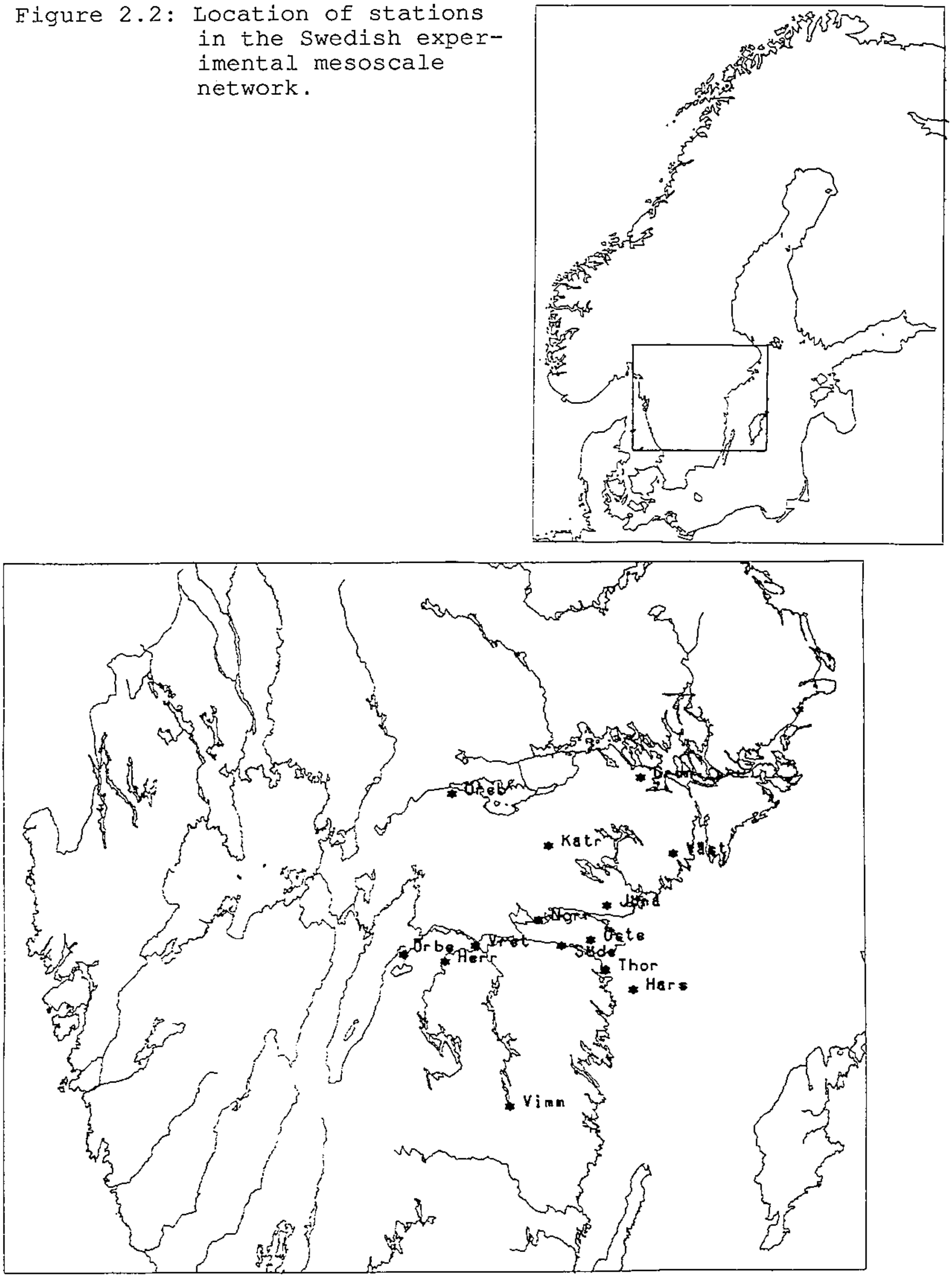


Figure 2.2: Location of stations in the Swedish experimental mesoscale network.

Table 2.6

List of stations in the Swedish mesoscale network

Station's name	Short name	Geographical coordinates E long./N lat. [°][']/[°][']	Elevation [m]
Norrköping	NOR	16 09/58 35	43
Grunbro	GRU	17 07/59 19	30
Örebro	ORE	15 17/59 16	23
Katrinenholm	KAT	16 14/58 59	62
Västerljung	VAE	17 26/58 56	34
Jönåker	JOE	16 46/58 45	30
Örberga	ORB	14 50/58 26	105
Herrberga	HER	15 13/58 24	85
Vreta Kloster	VRK	15 31/58 29	52
Söderköping	SOD	16 21/58 28	27
Östermem	OST	16 37/58 30	30
Thorönsborg	THO	16 45/58 21	4
Harstena	HAR	17 00/58 15	17
Vimmerby	VIM	15 51/57 40	107
Stockholm	STO	18 04/59 21	30

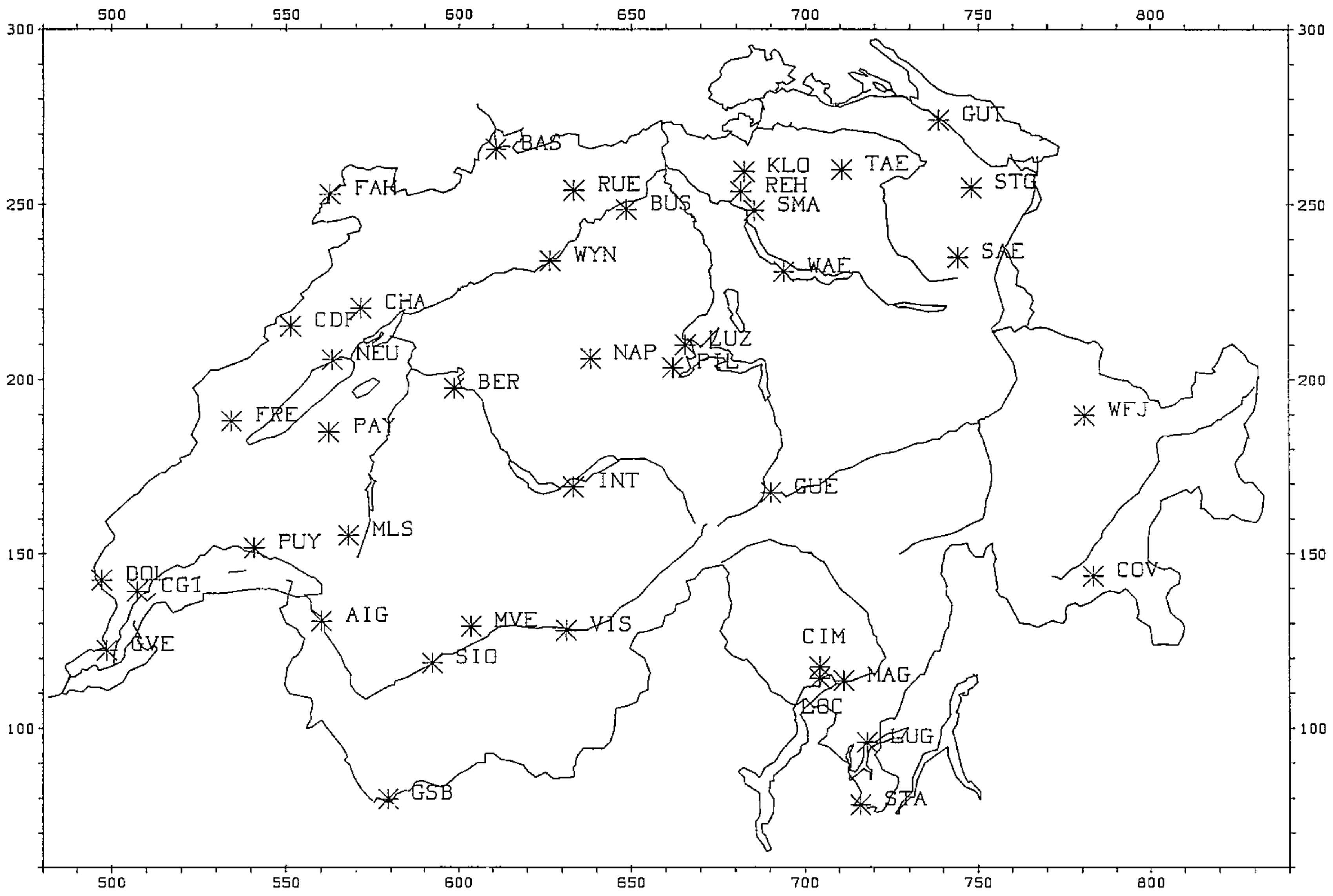


Figure 2.3: Location of stations in the Swiss ANETZ network. Axes are in km with arbitrary origin rather than longitude, resp. latitude.

Table 2.7

List of stations in the Swiss network ANETZ

Station's name	Short name	Geographical coordinates E long./N lat. [°][']/[°][']	Elevation [m]
La Dole	DOL	06 06/46 26	1670
Payerne	PAY	06 57/46 49	490
Wynau	WYN	07 47/47 15	422
Saentis	SAE	09 21/47 15	2490
Aigle	AIG	06 55/46 20	381
Moleson	MLS	07 01/46 33	1972
Fahy	FAH	06 57/47 26	596
Montana	MVE	07 29/46 19	1508
Chasseral	CHA	07 04/47 08	1599
Pilatus	PIL	08 15/46 59	2106
Lugano	LUG	08 58/46 00	273
Napf	NAP	07 56/47 00	1407
Sion	SIO	07 20/46 13	482
Locarno-Magadino	MAG	08 53/46 10	197
Neuchatel	NEU	06 57/47 00	485
Stabio	STA	08 56/45 51	353
Interlaken	INT	07 52/46 40	580
St.Gallen	STG	09 24/47 26	779
Geneve-Cointrin	GVE	06 08/46 15	420
Zuerich-Kloten	KLO	08 32/47 29	436
Guetsch	GUE	08 37/46 39	2287
Pully	PUY	06 40/46 31	461
Grand St.Bernhard	GSB	07 10/45 52	2472
Visp	VIS	07 51/46 18	640
La Chaux-de-Fonds	CDF	06 48/47 05	1018
Ruenenberg	RUE	07 53/47 26	610
Buchs-Suhr	BUS	08 05/47 23	387
Luzern	LUZ	08 18/47 02	456
Zuerich SMA	SMA	08 34/47 23	556
Weissfluhjoch	WFJ	09 49/46 50	2690
Corvatsch	COV	09 49/46 25	3315
Basel-Binningen	BAS	07 35/47 33	316
Changins	CGI	06 14/46 24	430
La Fretaz	FRE	06 35/46 50	1202
Bern-Liebefeld	BER	07 25/46 56	565
Guettingen	GUT	09 17/47 36	440
Waedenswil	WAE	08 41/47 13	463
Taenikon	TAE	08 54/47 29	536
Reckenholz	REH	08 31/47 26	443
Locarno-Monti	LOC	08 47/46 10	366
Cimetta	CIM	08 48/46 12	1672

Figure 2.4: Location of stations in the US Northeast Utilities & SUNY network, together with locations of target areas for NESDIS insolation estimates based on the GOES-Satellite.

Table 2.8

List of stations in the U.S. Northeast Utilities - SUNY network

Station's name	Short name	Geographical coordinates W long./N lat. [°] / [°]	Elevation [m]
Waterford	WAT	72.17/41.30	15
Middleton	MID	72.57/41.53	35
Milford	DEV	73.10/41.22	15
Woods Hole	WOH	70.66/41.52	15
Albany	ALB	73.80/42.70	87
Farmingdale	FAR	73.43/40.72	25
Queens	QUE	73.79/40.76	25

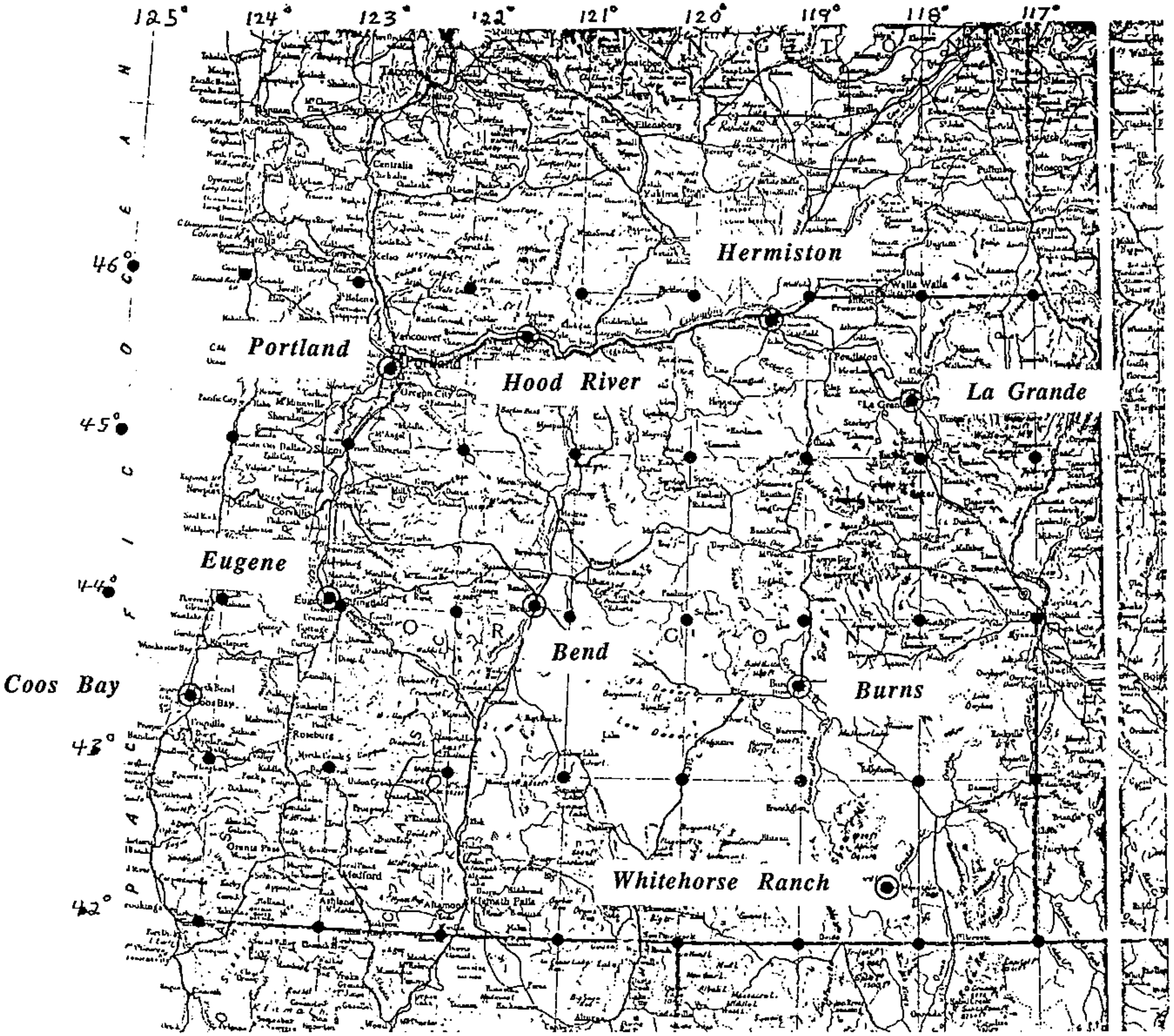


Figure 2.5: Location of stations in the US PNW network (encircled dots), together with locations of target areas for NESDIS insolation estimates based on the GOES-Satellite (dots).

Table 2.9

List of stations in the U.S. Pacific Nothwest network

Station's name	Short name	Geographical coordinates W long./N lat. [°] / [°]	Elevation [m]
Portland	POR	122.64/45.45	45
Hood River	HOO	121.52/45.70	140
Hermiston	HER	119.28/45.83	180
La Grande	LAG	118.08/45.33	885
Coos Bay	COB	124.25/43.40	35
Eugene	EUG	123.07/44.05	150
Bends	BEN	121.30/44.06	1115
Burns	BUR	119.02/43.52	1265
Whiterhorse	WHR	118.23/42.33	1325

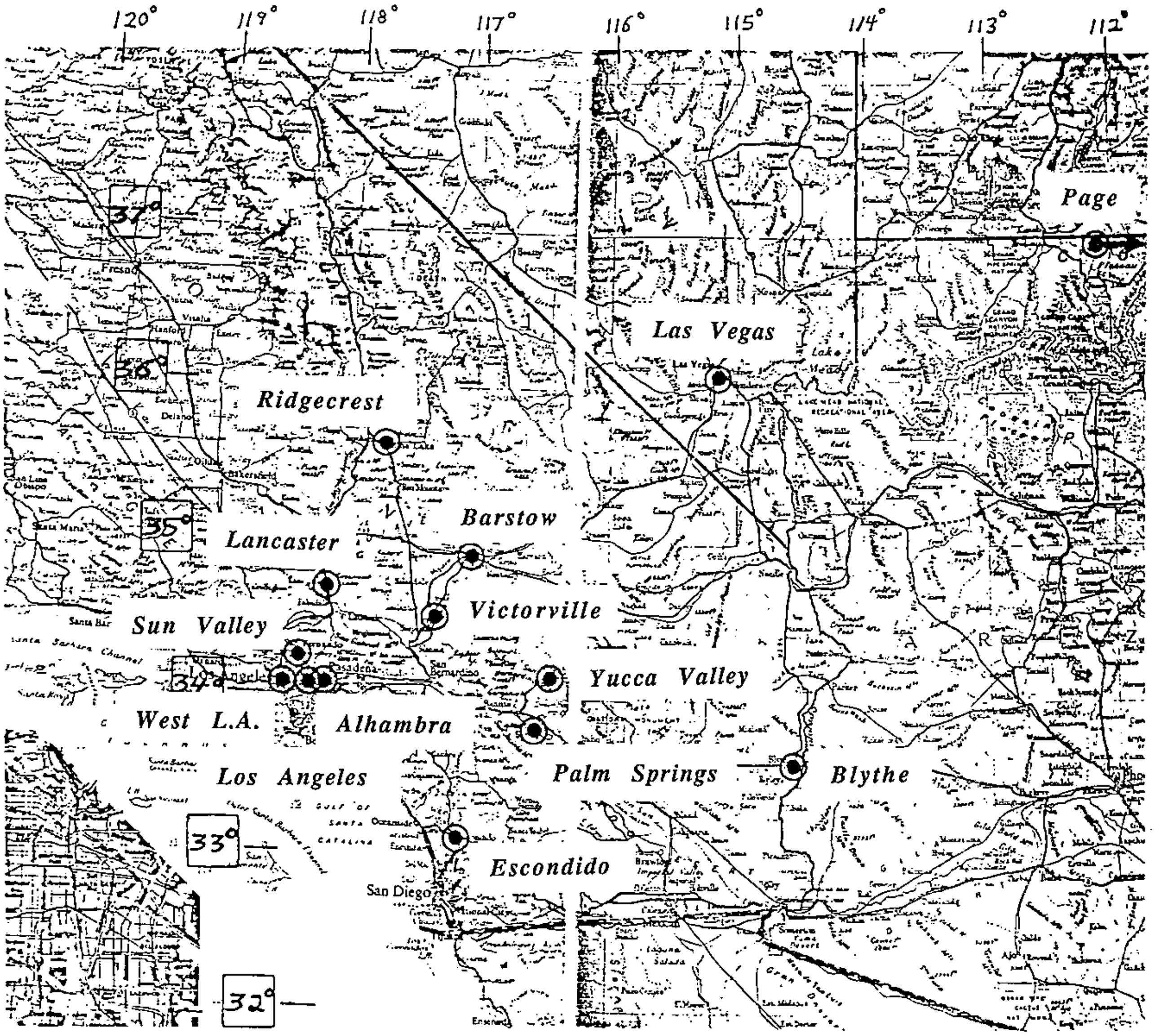


Figure 2.6: Location of stations in the US West Associates + NOAA network.

Table 2.10

List of stations in the U.S. Southwest - SERI network

Station's name	Short name	Geographical coordinates W long./N lat. [°] / [°]	Elevation [m]
Ridgecrest	RID	117.67/35.62	800
Barstow	BAR	117.00/34.88	600
Lancaster	LAN	118.15/34.70	800
Victorville	VIC	117.28/34.55	800
Sun Valley	SUV	118.38/34.25	200
Yucca Valley	YUV	116.42/34.12	900
Alhambra	ALH	118.15/34.08	100
West Los Angeles	WLA	118.45/34.07	40
Los Angeles	ANG	118.23/34.07	80
Palm Springs	PAS	116.46/33.78	200
Escondido	ESC	117.10/33.13	200
Las Vegas	VEG	115.10/36.05	659
Page	PAG	111.45/36.92	1000
Blythe	BLY	114.60/33.60	80

CHAPTER 3

3. Validation of ground-based methods

Although our data are devised so that we can work anytime with the clearness index K_T = GH/GH0, all the validations reported below were deliberarely computed with GH only. Doing so can introduce systematic errors when hourly intervals are considered. But we deal here with daily sums and the maximum distance between station pairs within a network corresponds at most to about 1° on the meridian. The extraterrestrial sum GH0 does not change much for such a latitude difference and, therefore, validations with K_T yield only marginally different results. And we realize that users will not care about computing GH0 when it is not imperative.

Distance between sites plays a key role in this chapter. The horizontal distance d can either be read from a geographical map, or be calculated from the geographical coordinates of both locations. In the latter case, let ϕ_1,λ_1 and ϕ_2,λ_2 be the latitude and longitude of site 1 and site 2 respectively. Then

$$\cos\alpha = \sin\phi_1\sin\phi_2 + \cos\phi_1\cos\phi_2\cos(\lambda_1-\lambda_2) \tag{3.1.1}$$

and

$$d = R\alpha \tag{3.1.2}$$

where R stands for an appropriate value of the Earth's radius for the considered region. In all the results reported below d has been calculated with R = 6366km.

Cross-validations are always expressed in terms of Mean Bias Error MBE = $(1/n)\Sigma(P_i-M_i)$ and in terms of Root Mean Square Error RMSE = $\sqrt{[(1/n)\Sigma(P_i-M_i)^2]}$ with summation over the pertaining number of stations and days of observation. Here, P stands for the predicted, and M for the measured value. When the MBE and RMSE are expressed in percent of the observed mean $(1/n)\Sigma M_i$, then they are denoted with MBE% and RMSE%.

3.1 Nearest neighbor approximation (Extrapolation)

Study prepared by:	*Antoine Zelenka and Draginja Lazic* *Swiss Meteorological Institute*
Under sponsorship from:	*Swiss Federal Office of Energy*

3.1.1 *Method and presentation of results*

Solar energy system design requires prediction of irradiances that a network station would measure if it were located at the system's site. The only way to avoid such an evaluation is to use data from some adjacent site. But then, as Hay [3.1.1] states, "the user must question the applicability of the data to his specific location or, in the more general sense he must assess the degradation in the quality of the information as a result of the *extrapolation* of data from the site of origin to the site of use."

The conception of closest network station can be extended to encompass *a priori* climatological knowledge. The horizontal distance d between a valley site and a mountain summit station can indeed be small, but the insolation climate of the closest valley station is certainly the better approximation.

Zelenka and Lazic [3.1.2] have shown that this can be accomodated for with the definition of a new "distance" D as

$$D^2 = d^2 + [v(h_2 - h_1)]^2 \qquad (3.1.3)$$

with h_1 and h_2 being the respective altitudes of the sites. The magnitude of the vertical scale factor v depends upon the variability of the insolation field. Typically, v is found to lie around 100 (see Section 3.3), indicating that on the average, a 100 meter change in height results in a change in insolation climate about the same as a 10km movement horizontally.

Clearly, this concept of "climatic distance" can be further refined with introduction of East-West and North-South scale factors in the horizontal distance. We shall not dwell on this too long because, as we shall see in Section 3.4, the whole kriging technique relies on this very recognition of the field's structure.

The cross-validation results for extrapolation presented in the tables of Volume 3 and the diagrams below *are for horizontal distances only,* more precisely for D with v=1, which is very close to the horizontal distance d. This is for two reasons: (1) use of D with v>>1 improves validation results only for the Swiss ANETZ where altitude differences are very pronounced and, (2), we deliberately aimed at the crudest possible estimation, corresponding to straightforward use of extrapolation. Doing so has had the advantage of defining an accuracy threshold that any other technique will have to surpass in order to be considered.

The structure of the validation results displayed for each network is as follows.

In this Volume 1 (Report): Two map sketches (with approximate scale) show the geographic distribution of the stations and display their yearly validation statistics (MBE% and RMSE% in Figs. 3.1.1 to 3.1.14);

Volume 3 (Results) contains tables (in this case Table 3.I) consisting of

1) Front pages: Identifying the network.

Displaying min. & max. of relative (% of measured mean) MBE, RMSE and $U95=\sqrt{[(MBE)^2+(2RMSE)^2]}$. U95 is a measure of the uncertainty of the estimate.

The absolute values are in $MJm^{-2}d^{-1}$ and they correspond to the min. & max. noted above.

Min. & max. selection are only for months with at least 15 observed days.

2) Entries for each station and month with:

MM (ND)
MBE / MBE%
RMSE / RMSE%

where MM = the measured mean in $MJm^{-2}d^{-1}$; ND = the number of observed days;
MBE and RMSE = the absolute MBE and RMSE in $MJm^{-2}d^{-1}$; and
MBE% and RMSE% = the relative MBE and RMSE in % of MM.

3) "Bottom lines" provide monthly (for all stations) and station-wise (yearly) entries as in 2).

The WEST Associates-NOAA network is validated twice. Once completely (Figs. 3.1.3 and 3.1.4) and once without the station Las Vegas and Page (Figs. 3.1.5 and 3.1.6). The large distance between the latter and the large distances to the rest of the network preclude the use of methods other than extrapolation. Thus, the restricted validation here is mainly for later comparisons.

VALIDATION OF EXTRAPOLATION

YEAR MBE [% MEAS.MEAN OF SITE]

NETWORK

NO. OF DAYS	2993
MEAS. MEAN [MJm^{-2}]	14.36
MBE % OF MEAS.MEAN	-1.6
RMSE % OF MEAS.MEAN	27.0

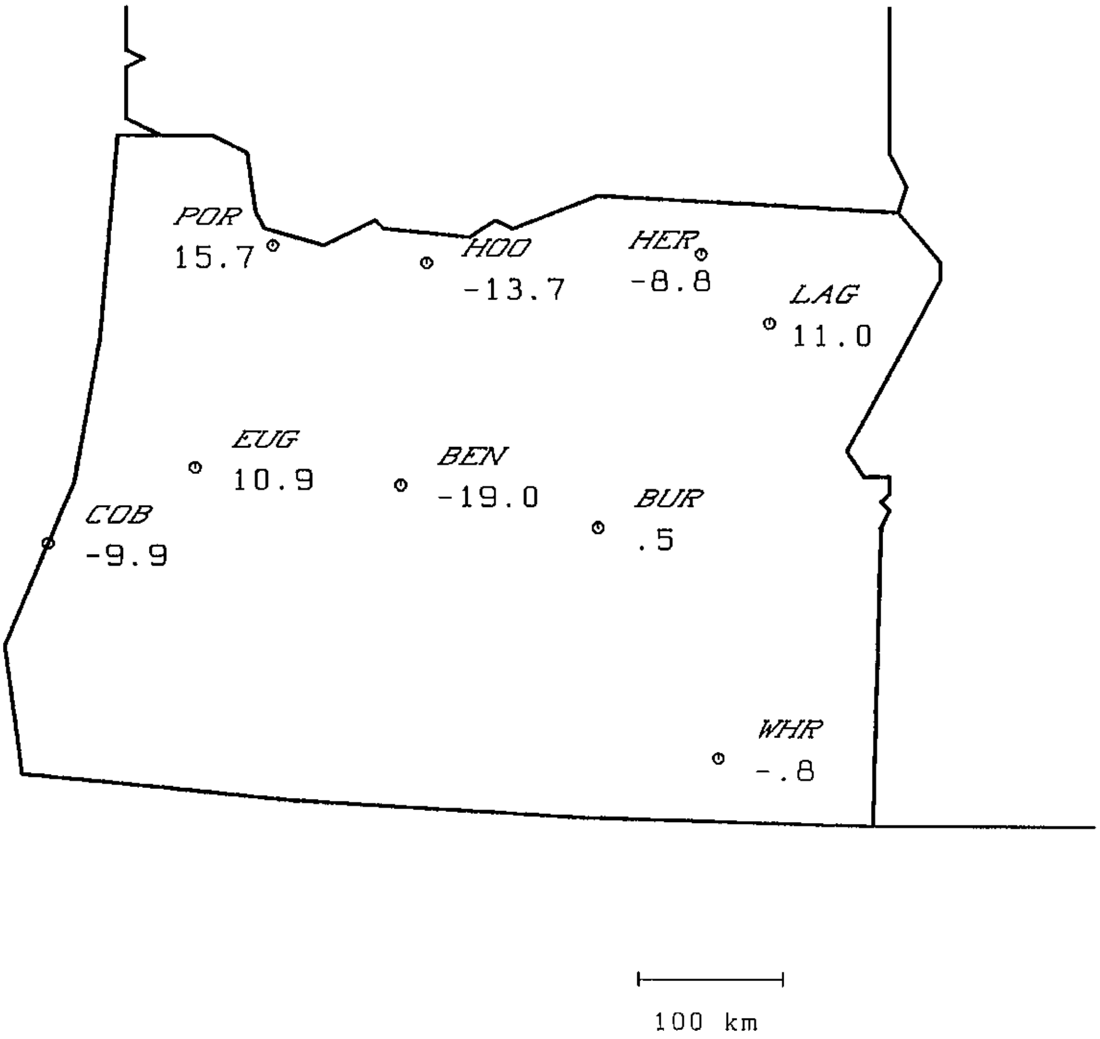


Fig. 3.1.1 : MBE% Pacific Northwest.

VALIDATION OF EXTRAPOLATION

YEAR RMSE [% MEAS.MEAN OF SITE]

NETWORK

NO. OF DAYS	2993
MEAS. MEAN [MJm^{-2}]	14.36
MBE % OF MEAS.MEAN	-1.6
RMSE % OF MEAS.MEAN	27.0

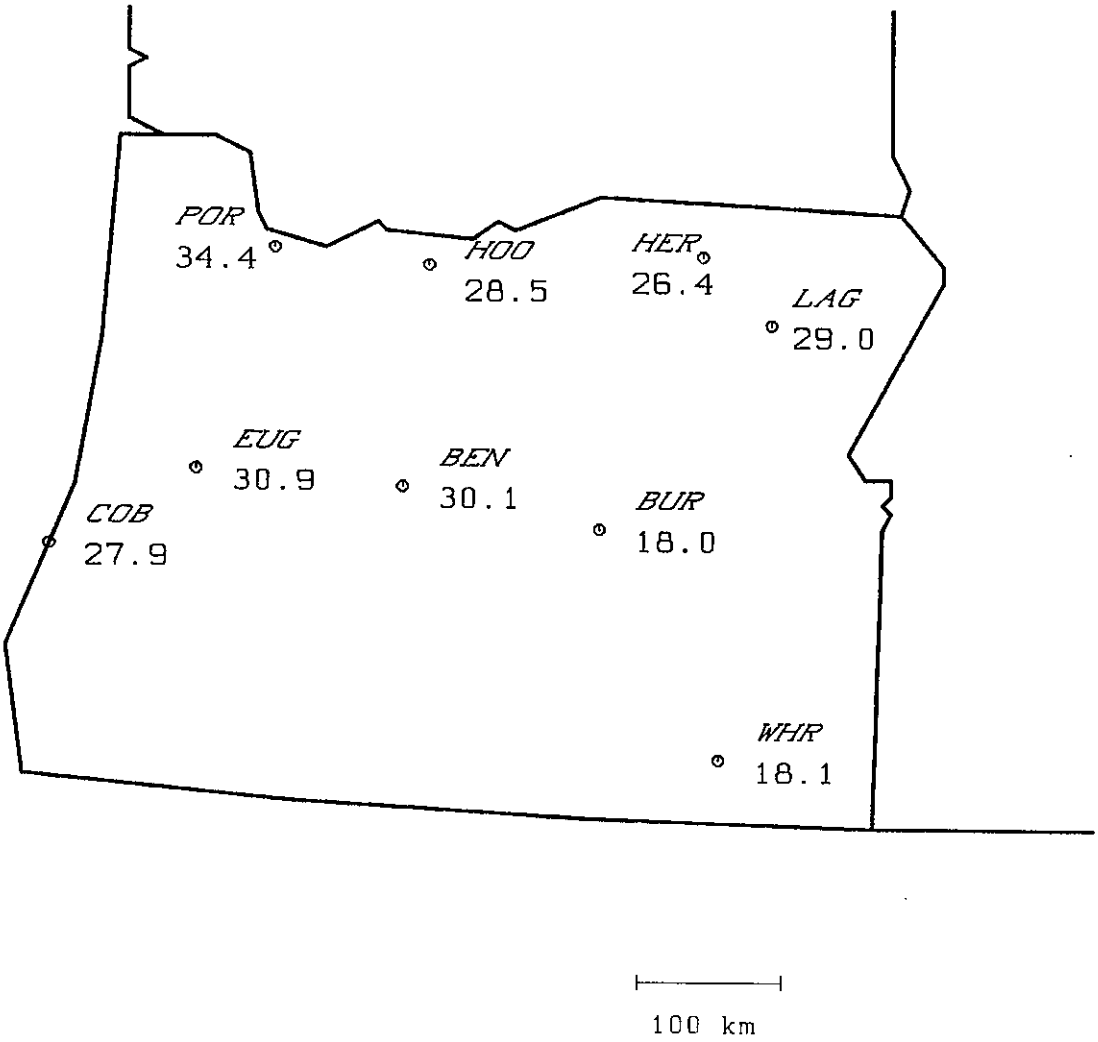
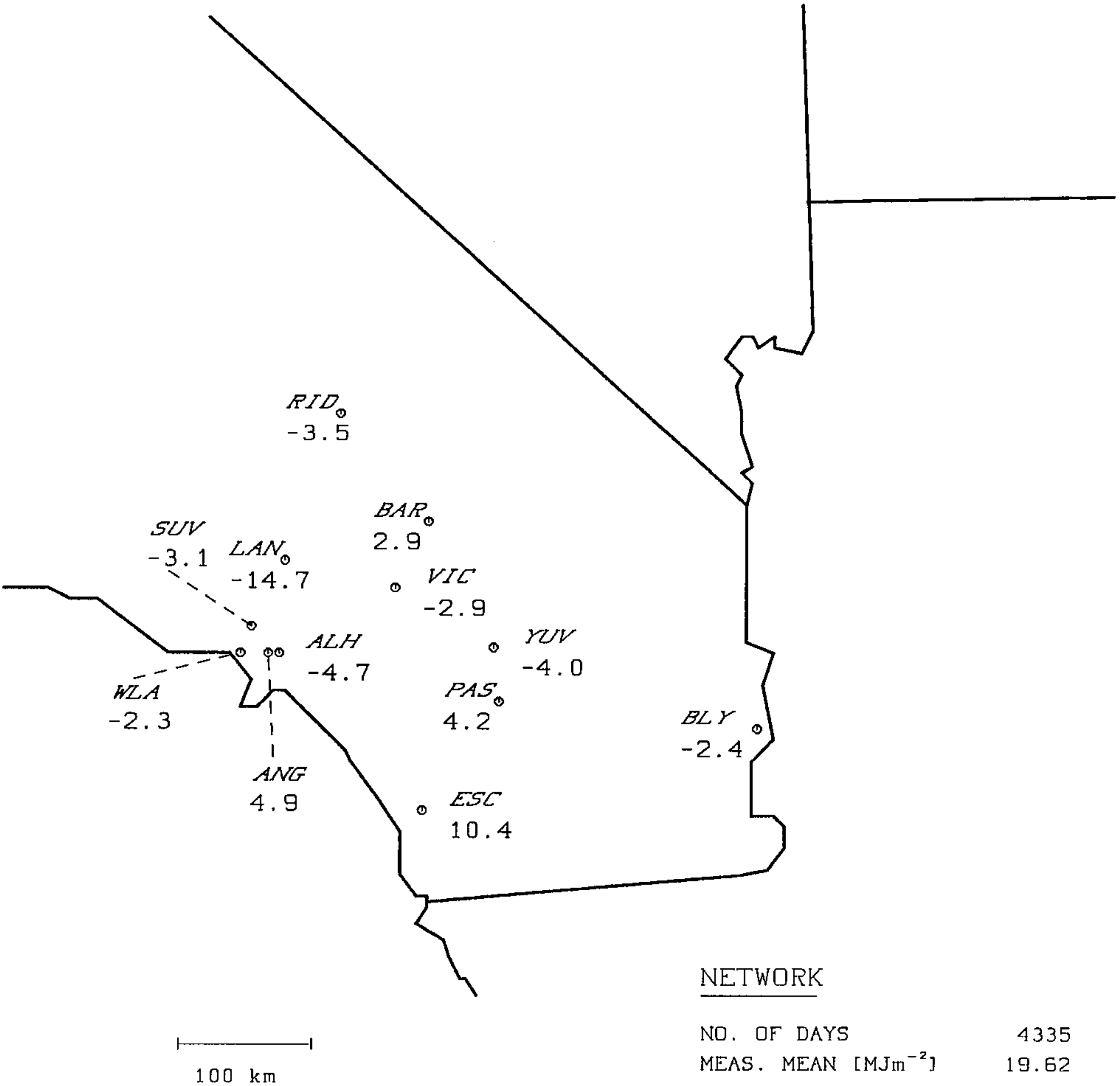


Fig. 3.1.2 : RMSE% Pacific Northwest.

VALIDATION OF EXTRAPOLATION

YEAR MBE [% MEAS.MEAN OF SITE]



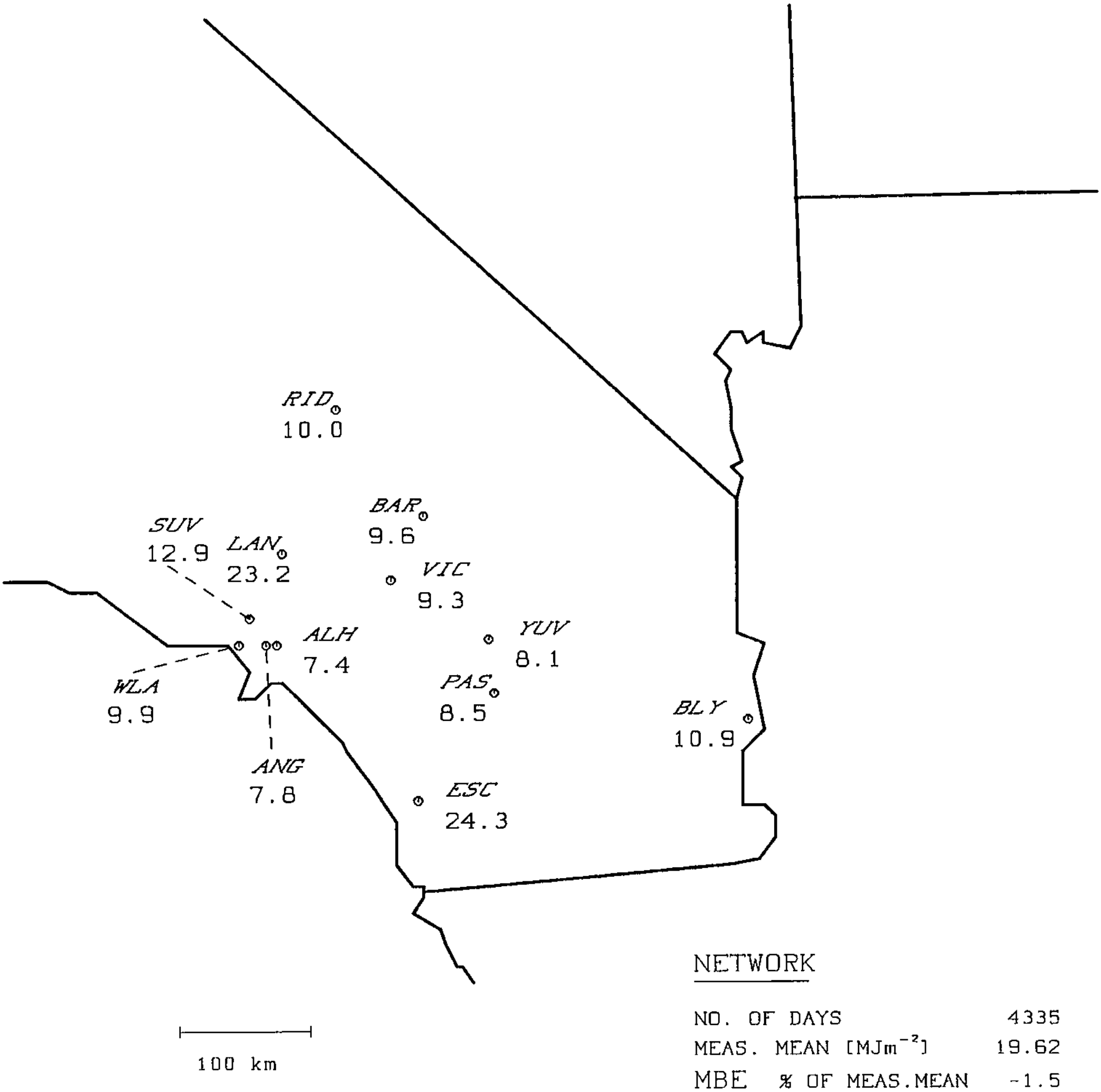
NETWORK

NO. OF DAYS	4335
MEAS. MEAN [MJm^{-2}]	19.62
MBE % OF MEAS.MEAN	-1.5
RMSE % OF MEAS.MEAN	13.0

Fig. 3.1.3 : MBE% West Associates.

VALIDATION OF EXTRAPOLATION

YEAR RMSE [% MEAS.MEAN OF SITE]



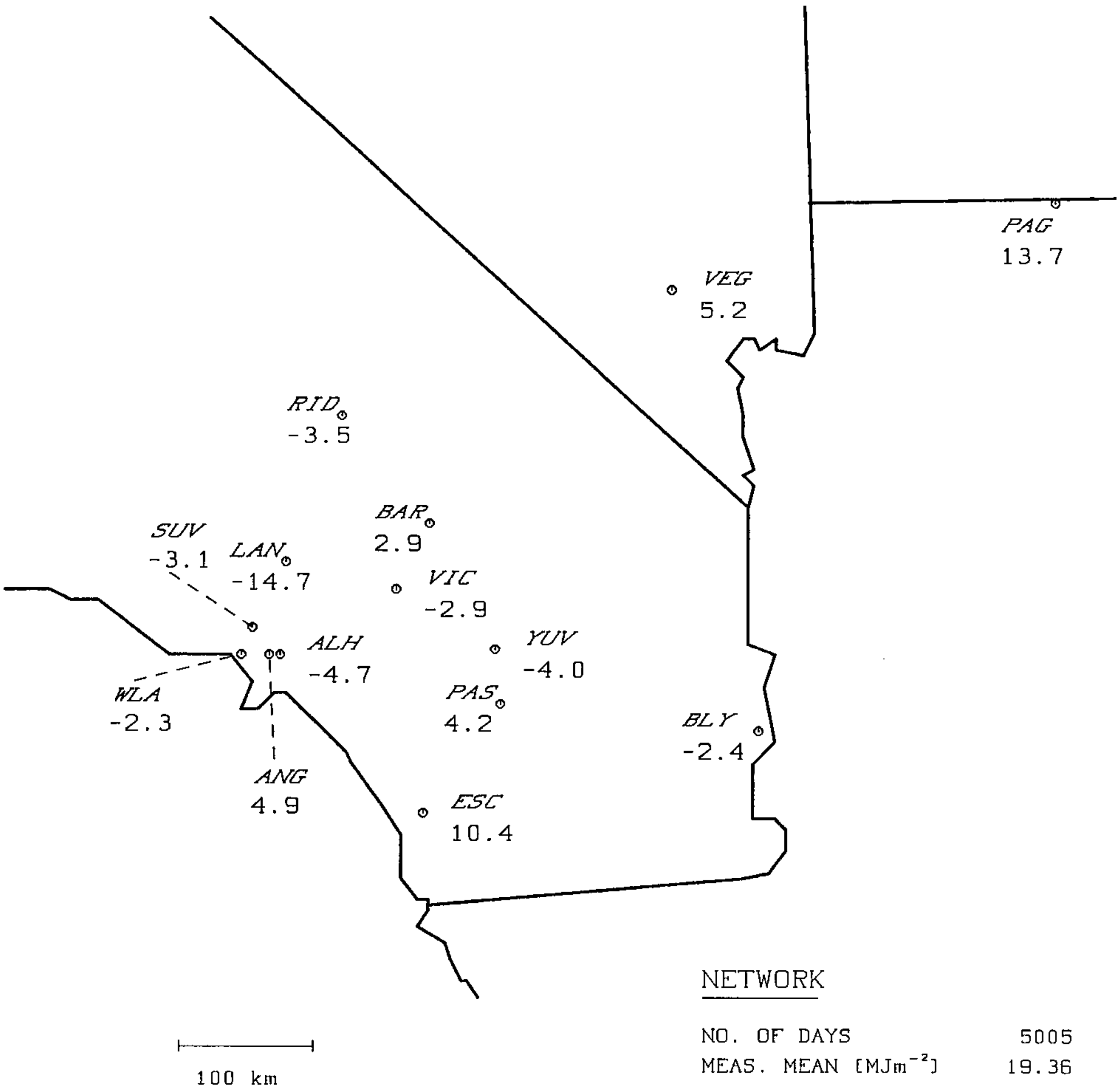
NETWORK

NO. OF DAYS	4335
MEAS. MEAN [MJm^{-2}]	19.62
MBE % OF MEAS.MEAN	-1.5
RMSE % OF MEAS.MEAN	13.0

Fig. 3.1.4 : RMSE% West Associates.

VALIDATION OF EXTRAPOLATION

YEAR MBE [% MEAS.MEAN OF SITE]



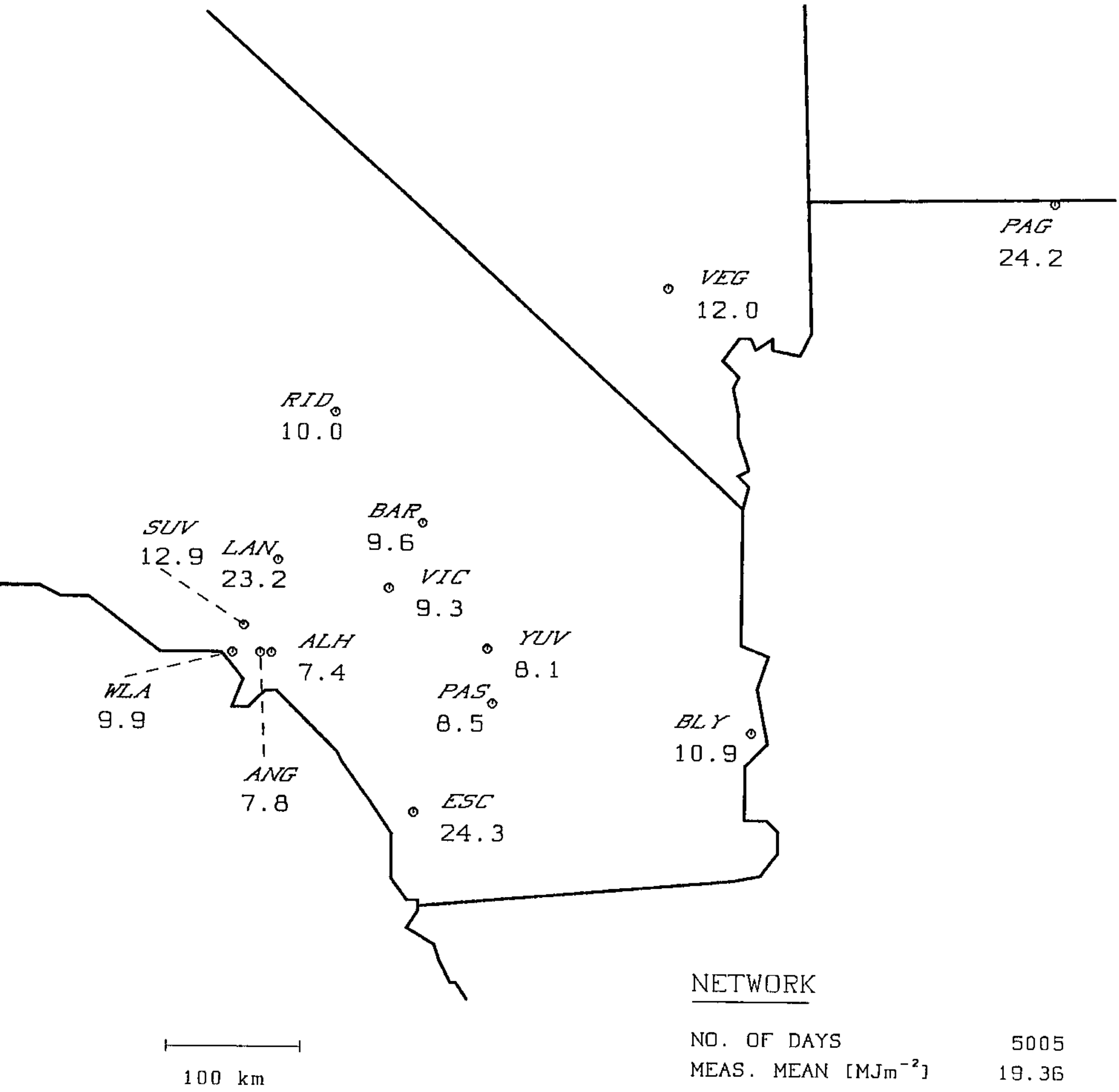
NETWORK

NO. OF DAYS	5005
MEAS. MEAN [MJm^{-2}]	19.36
MBE % OF MEAS.MEAN	-.2
RMSE % OF MEAS.MEAN	13.7

Fig. 3.1.5 : MBE% West Associates & NOAA.

VALIDATION OF EXTRAPOLATION

YEAR RMSE [% MEAS.MEAN OF SITE]



NETWORK

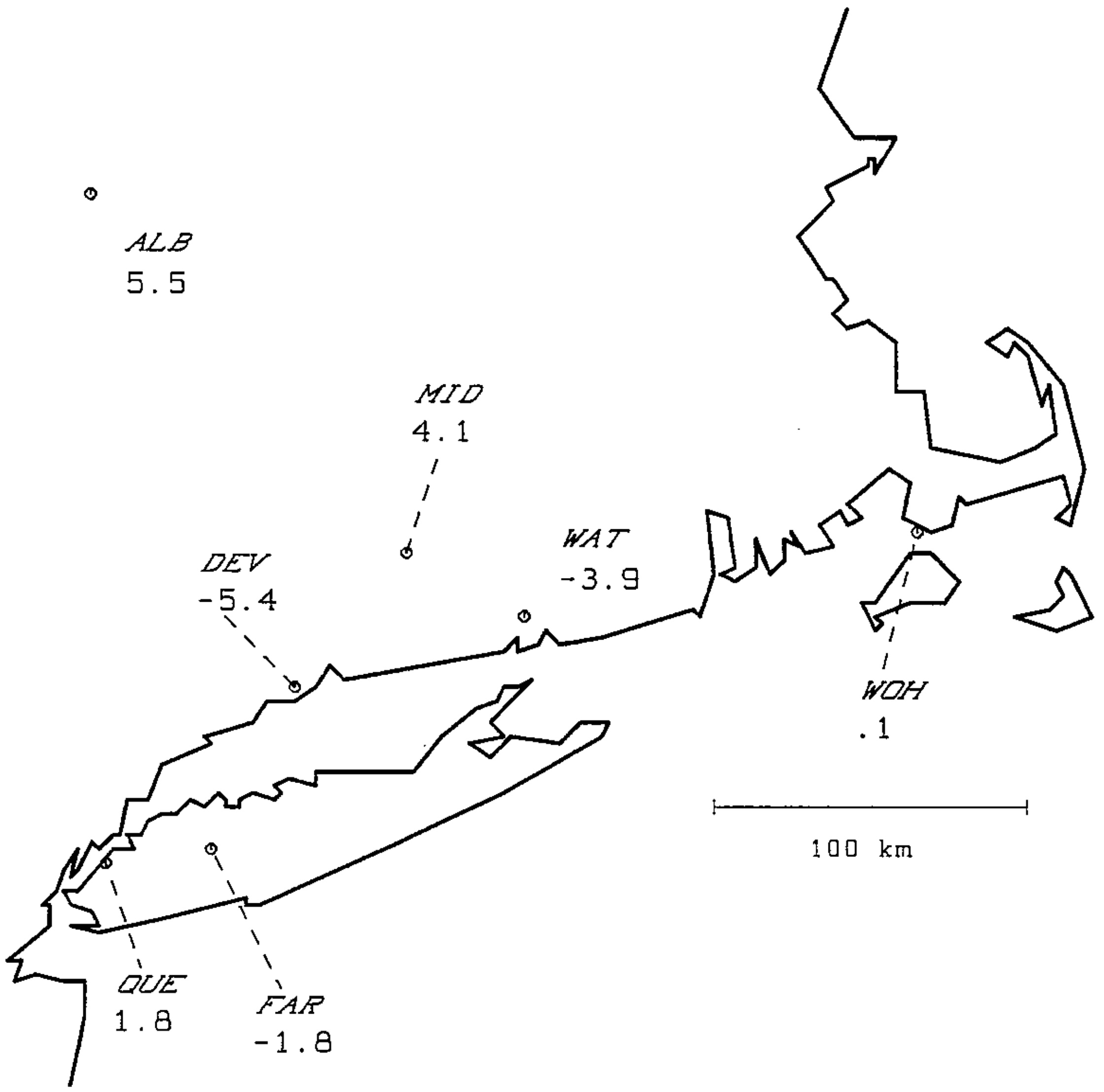
NO. OF DAYS	5005
MEAS. MEAN [MJm^{-2}]	19.36
MBE % OF MEAS.MEAN	-.2
RMSE % OF MEAS.MEAN	13.7

Fig. 3.1.6 : RMSE% West Associates & NOAA.

Fig. 3.1.7 : MBE% Northeast Utilities & SUNY.

VALIDATION OF EXTRAPOLATION

YEAR MBE [% MEAS.MEAN OF SITE]



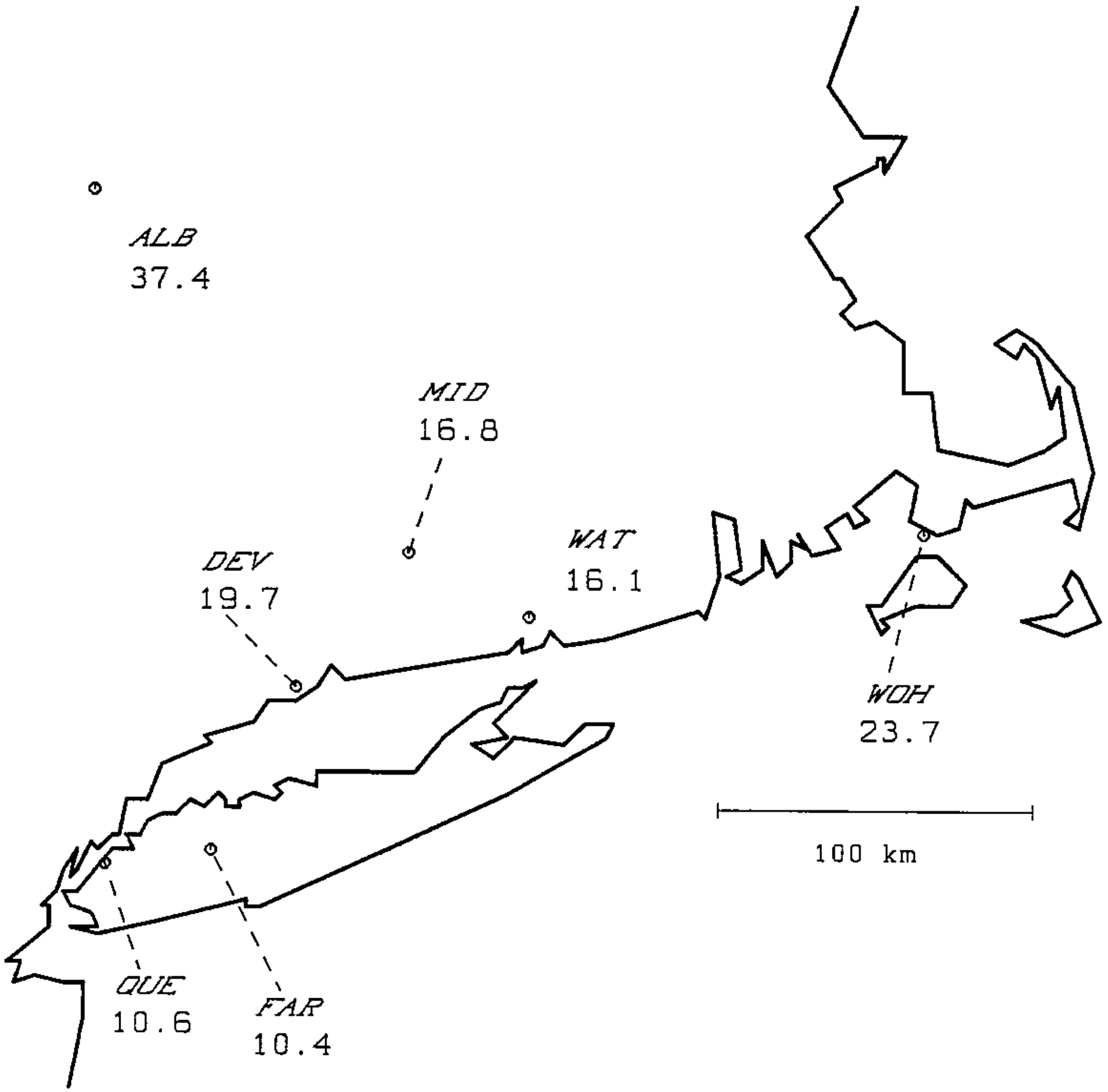
NETWORK

NO. OF DAYS	938
MEAS. MEAN [MJm^{-2}]	12.59
MBE % OF MEAS.MEAN	-.1
RMSE % OF MEAS.MEAN	20.6

Fig. 3.1.8 : RMSE% Northeast Utilities & SUNY.

VALIDATION OF EXTRAPOLATION

YEAR RMSE [% MEAS.MEAN OF SITE]



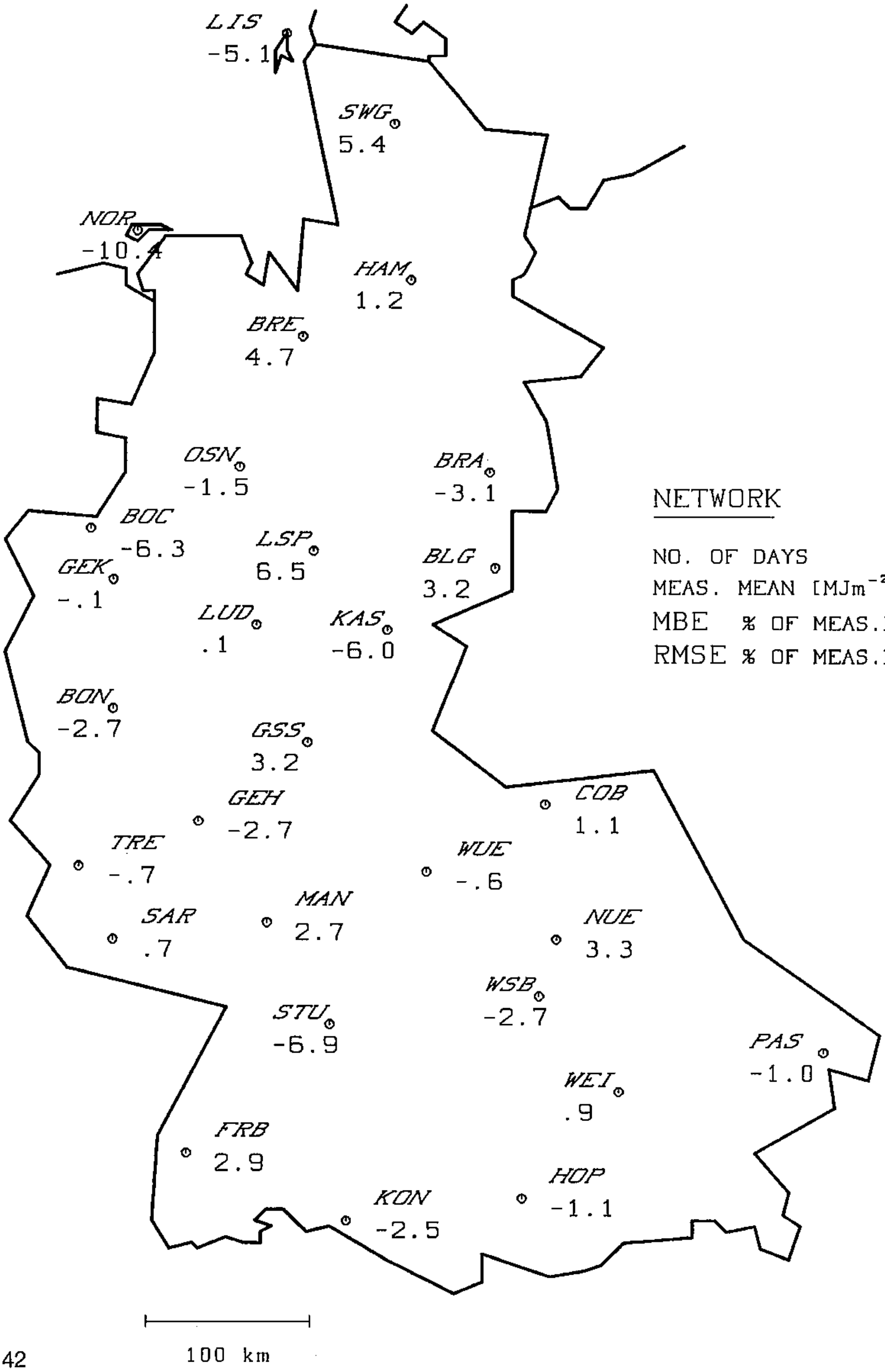
NETWORK

NO. OF DAYS	938
MEAS. MEAN [MJm^{-2}]	12.59
MBE % OF MEAS.MEAN	-.1
RMSE % OF MEAS.MEAN	20.6

Fig. 3.1.9 : MBE% Germany DWD.

VALIDATION OF EXTRAPOLATION

YEAR MBE [% MEAS.MEAN OF SITE]



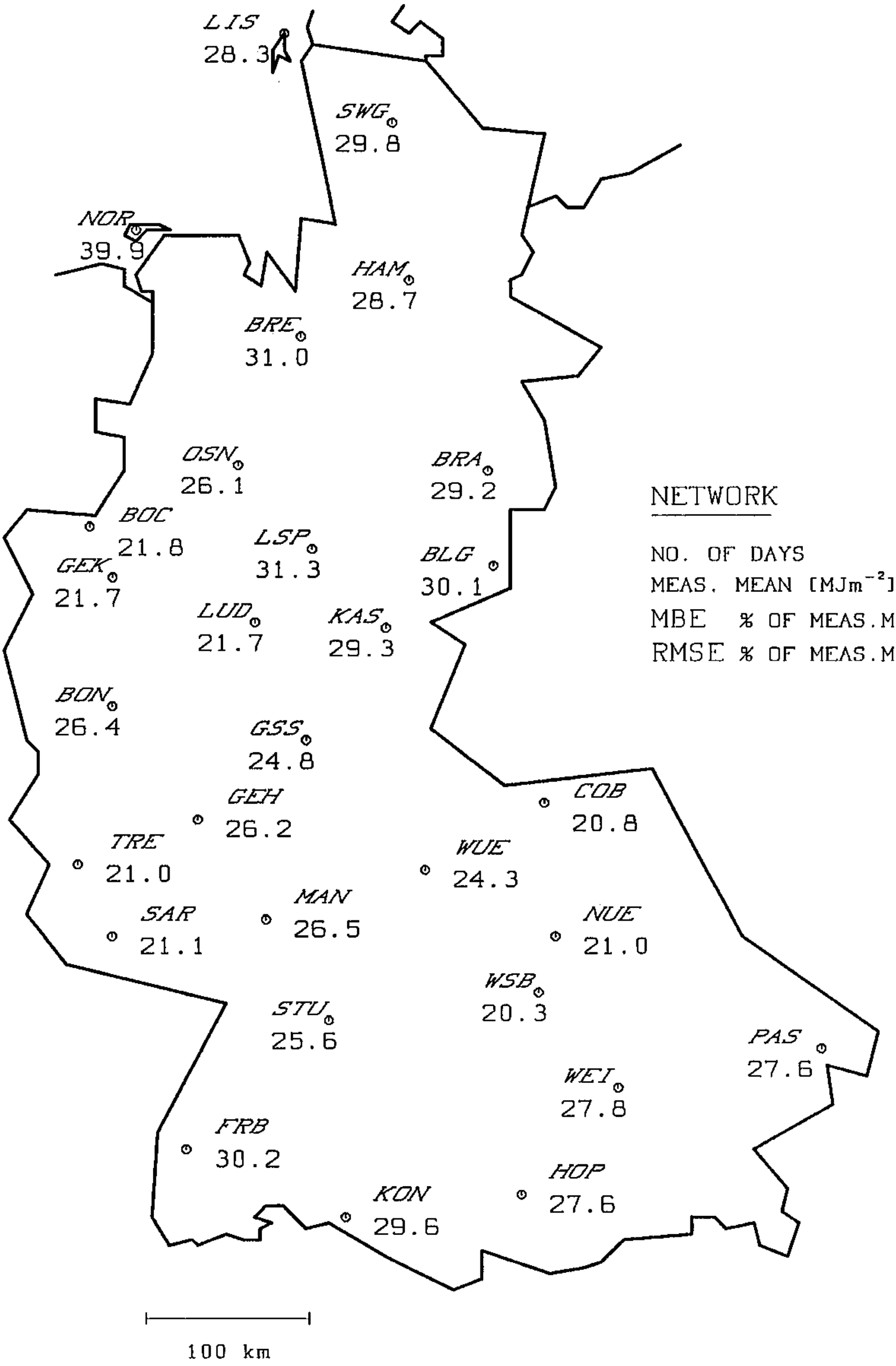
NETWORK

NO. OF DAYS	10539
MEAS. MEAN [MJm^{-2}]	9.80
MBE % OF MEAS.MEAN	-.7
RMSE % OF MEAS.MEAN	26.9

Fig. 3.1.10: RMSE% Germany DWD.

VALIDATION OF EXTRAPOLATION

YEAR RMSE [% MEAS.MEAN OF SITE]



NETWORK

NO. OF DAYS	10539
MEAS. MEAN [MJm^{-2}]	9.80
MBE % OF MEAS.MEAN	-.7
RMSE % OF MEAS.MEAN	26.9

VALIDATION OF EXTRAPOLATION

YEAR MBE [% MEAS.MEAN OF SITE]

NETWORK

NO. OF DAYS	2573
MEAS. MEAN [MJm^{-2}]	15.81
MBE % OF MEAS.MEAN	-.4
RMSE % OF MEAS.MEAN	13.3

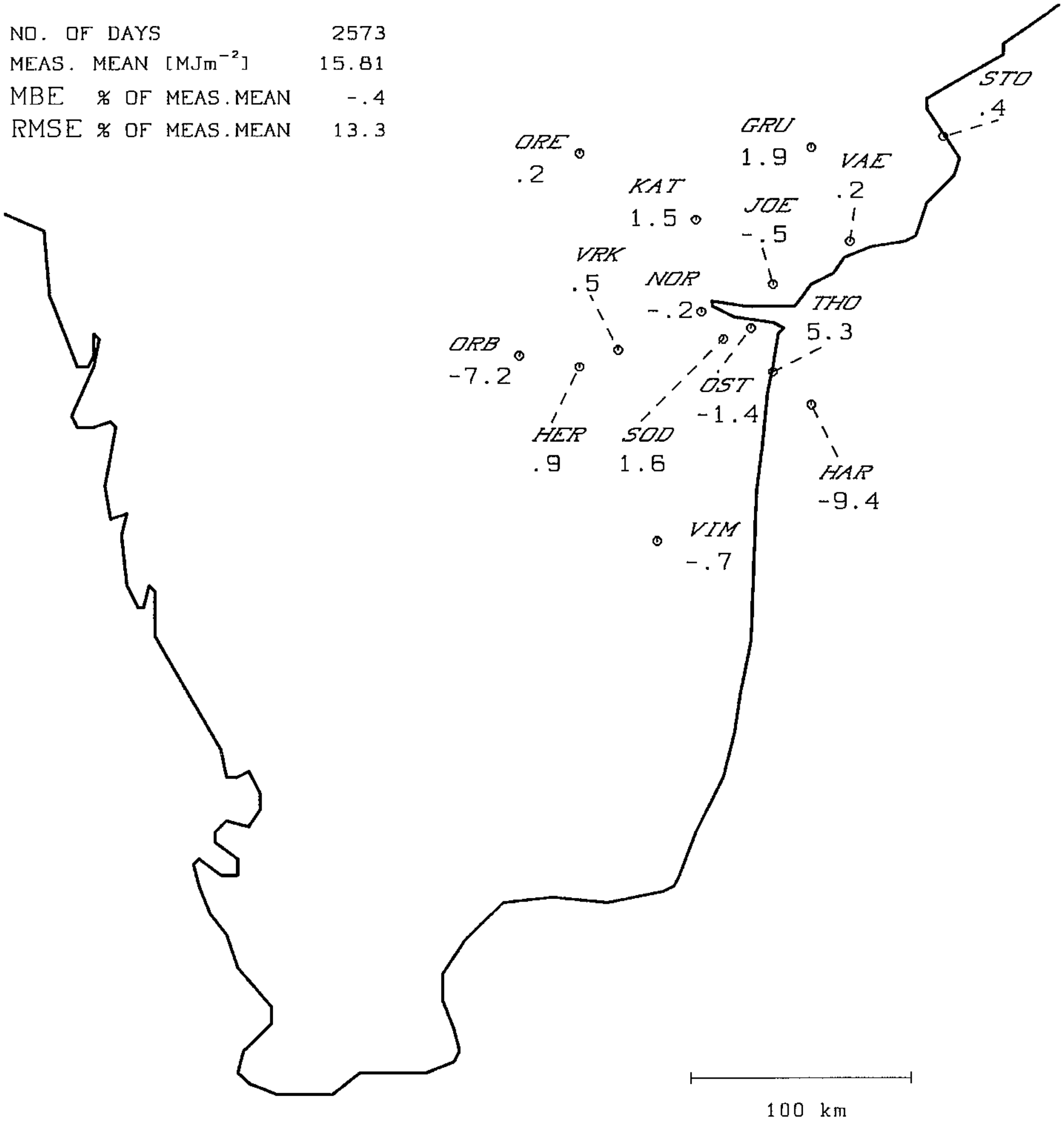


Fig. 3.1.11: MBE% Sweden mesoscale.

VALIDATION OF EXTRAPOLATION

YEAR RMSE [% MEAS.MEAN OF SITE]

NETWORK

NO. OF DAYS	2573
MEAS. MEAN [MJm^{-2}]	15.81
MBE % OF MEAS.MEAN	-.4
RMSE % OF MEAS.MEAN	13.3

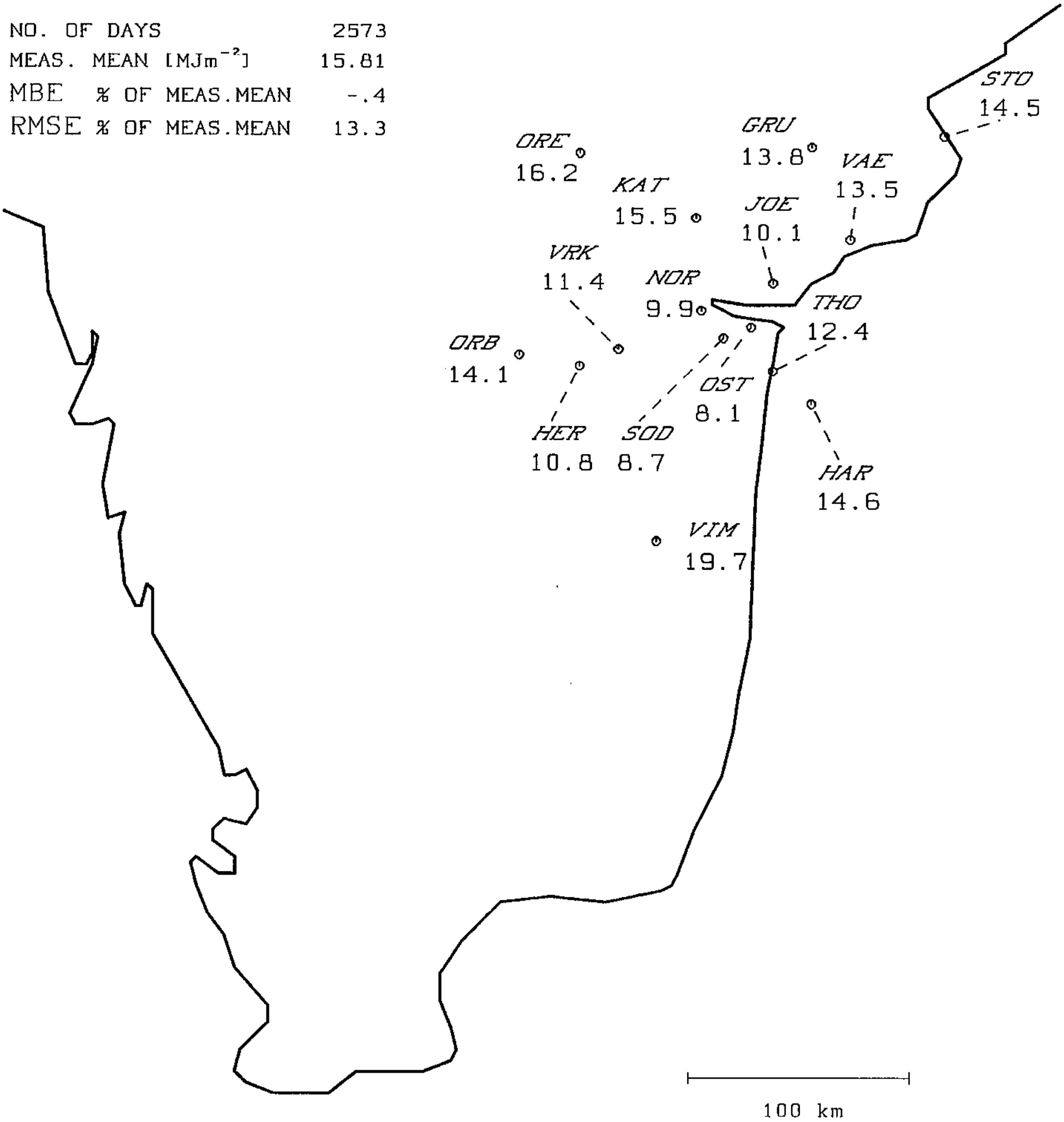


Fig. 3.1.12: RMSE% Sweden mesoscale.

Fig. 3.1.13: MBE% Switzerland ANETZ.

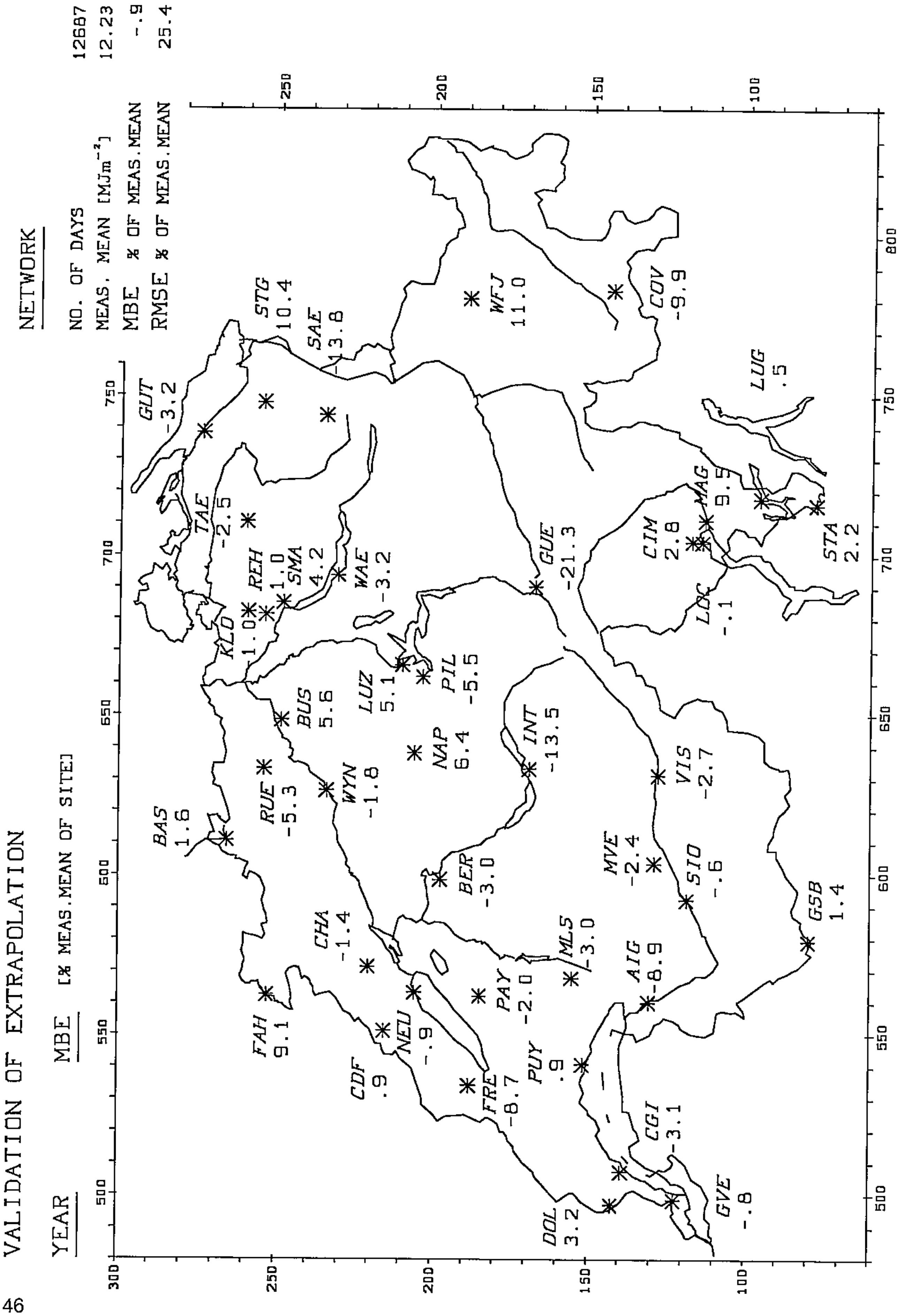
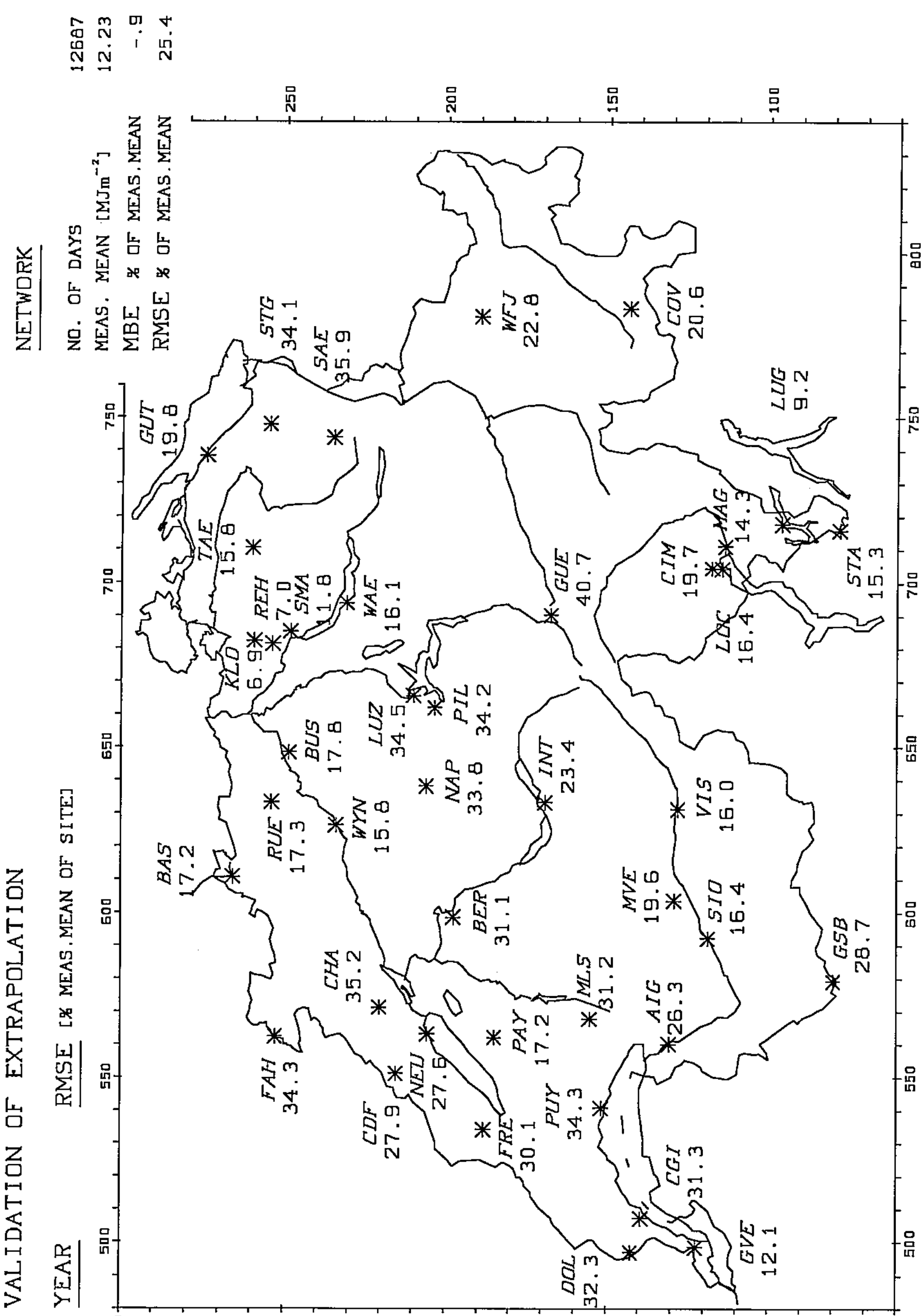


Fig. 3.1.14: RMSE% Switzerland ANETZ.



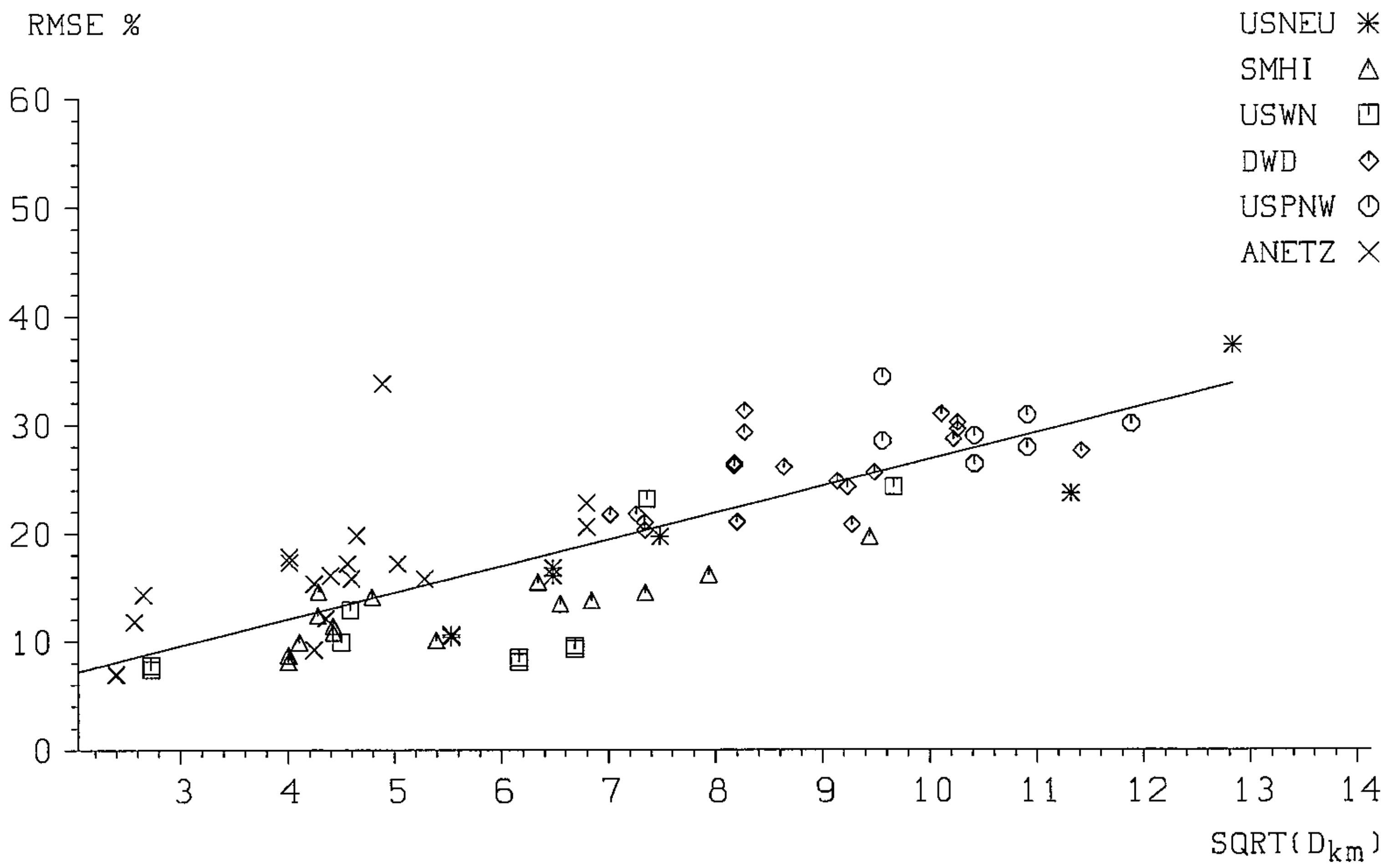


Fig. 3.1.15: Correlation between RMSE (in % of pertaining, observed mean) of one year of extrapolated daily global irradiation and the square root of extrapolation distance D[km].

3.1.2 *Discussion of cross-validation results*

The overall MBE (all stations and all days) is always close to zero. This is simple to understand as stations mostly aggregate in pairs (although it cannot always be so, it happens rather often, that when site 2 is closest to site 1, then site 1 is also closest to site 2.)

Seasonal trends are obvious for those regions where seasons have a signature in the insolation climate. RMSE% then vary between 50% in winter to 20% in summer. The large winter value are due to the lower means to which the percent pertain, to reduced measurement precision, and, in certain cases (e.g., in the Swiss network), to the reduced number of sites contributing to the statistics. However, the primary cause is the more erratic behavior of the insolation field.

As can be seen on the front page of the Tables, individual RMSEs can become prohibitive in the sense that any educated guess would be as good as the prediction of the algorithm (if this crude procedure deserves such a name). This last point emphasizes the potential danger of this very popular (because of its simplicity) technique.

Most interesting is the behavior of the RMSE% with respect to extrapolation distance as already stated in the introduction.

Perez *et al.* [3.1.3] (Chapter 6, Section 6.2) found for the NEU-SUNY network that yearly (and seasonal) RMSE% increase linearly with log(d[km]). However, when all networks are considered, linear regression is slightly better with $\sqrt{d}$[km]. Figure 3.1.15 shows this relation which reads

$$\mathrm{RMSE\%} = 2.2 + 2.5\sqrt{d_{km}}\ [\%],\ \ 4\mathrm{km} \leq d_{km} \leq 165\mathrm{km} \qquad (3.1.4)$$

with a correlation coefficient of 0.813, well above the critical value at the 1% confidence level.

Such a clear relation, however, emerges only when pairs from more or less climatically homogeneous regions are accepted and pairs from clearly heterogeneous regions are rejected. Table 3.1.1 contains the list of closest neighbors (as defined by horizontal distance) which were rejected after study of available insolation climate descriptions [3.1.4 - 3.1.6]. Rejected pairs come mostly from coast-to-inland and valley-to-summit locations, as well as sites separated, among others, by hill/mountain ranges.

The requirement for homogeneity, as well as relation (3.1.4), are strongly reminiscent of Hay and Suckling's (1978) relation as reported by Wilson [3.1.7], and of Atwater and Ball's [3.1.8] relation between distance and coefficient of variability for northern America, respectively reminiscent of the relations reported in WMO Technical Note No. 172 [3.1.9] for other temperate latitudes.

Finally, we assess the gain in prediction accuracy when distances other than the mere horizontal one are used to define the "nearest" neighbor. Table 3.1.1 shows that the Swiss ANETZ is the climatically most heterogeneous network. If we use distances D as defined by (3.1.3) with a vertical scale factor v=100 from November to January and v=50 for the other months, we bring stations with about the same altitude always nearer one to another than to any station at a markedly different level. Overall validation then results in MBE% = 0% and RMSE% = 24% which is not really different from the previous case. However, the largest individual RMSE% is reduced from 115% to 72%, while the largest individual MBE% drops from 92% to 55%. Obviously, even these figures are too large and more powerful prediction techniques are needed.

3.1.3 *References*

[3.1.1] J.E. Hay: 1984. An assessment of the mesoscale variability of solar radiation at the Earth's surface. *Solar Energy* **32**, 425-434.

[3.1.2] A. Zelenka and D. Lazic: 1987. Supplementing network global irradiance data. *Advances in Solar Energy Technology*, Proc. 1987 Biennial Congres ISES, Hamburg. **Vol. 4**, pp. 3861-3865, Pergamon 1988.

[3.1.3] R. Perez, R. Seals, and R. Stewart: 1990. Assessing photovoltaic interaction with New York summer-peaking utilities: An evaluation of existing solar resource assessment capabilities from measuring networks and geostationary satellites. Rept. to the New York Power Authority, Atmospheric Sciences Research Center, State University of NY at Albany, Albany NY 12222.

[3.1.4] S. Kunz, R. Volz und F. Wittmer: 1984. Sonnenenergie Nutzungszonen Schweiz, SONUS. Bericht Geogr. Inst. Univ. Bern, CH-3012 Bern, Switzerland.

[3.1.5] R. Trapp and F. Kasten: 1988. Kleinskalige Variabilität der Sonnenstrahlung. Berichte des DWD, Deutscher Wetterdienst, D-6050 Offenbach a.M., FR Germany.

[3.1.6] F. Vignola and D.K. McDaniels: 1989. Pacific Northwest Solar Radiation Data. Solar Monitoring Lab., Physics Dept., Iniversity of Oregon, Eugene OR 97403.

[3.1.7] R.G. Wilson, 1978. "Radiation Network Assessment and Design". *Proceedings 1st Canadian Solar Radiation Workshop*, J.E. Hay and T.K. Won Eds., Ministry of Supply and Services Canada 1980, pp.108-110.

[3.1.8] M.A. Atwater and J.T. Ball, 1978. "Intraregional Variations of Solar Radiation in the Eastern United States". *Jour. Applied Meteorol.* **17**, 1116-1125.

[3.1.9] WMO - No. 557, 1981: "Meteorological Aspects of the Utilization of Solar Radiation as an Energy Source". *WMO Technical Note* **No. 172**, CH-1211 Geneva 20.

Table 3.1.1

List of rejected station pairs
in the study of RMSE% as a function of extrapolation distance
(closest neighbors in horizontal distance but
not belonging to the same insolation climate)

Site	Neighbour	reciprocity	reason
ANETZ			
La Dôle	Changins	yes	Alts. 1670/430m
Säntis	St. Gallen	yes	Alts. 2490/779m
Moléson	Aigle	yes	Alts. 1972/381m
Fahy	Chasseral		Alts. 596/1599m
Montana	Sion	yes	Alts. 1508/482m
Chasseral	Neuchâtel		Alts. 1599/485m
Pilatus	Luzern	yes	Alts. 2106/456m
Neuchâtel	La Chaux-de-fonds	yes	Alts. 485/1018m
Interlaken	Napf		Alts. 580/1407m
Gütsch	Pilatus		differ. regions
Pully	Moléson		Alts. 461/1972m
Gd. St. Bernard	Sion		Alts 2472/482m
Visp	Montana		Alts. 640/1508m
La Frêtaz	Payerne		Alts. 1202/490m
Bern	Chasseral		Alts. 565/1599m
Locarno	Cimetta	yes	Alts. 366/1672m
DWD			
Braunschweig	Braunlage	yes	Alts. 83/615m
Hohenpeissenberg	Weihenstephan	yes	Alts. 990/469m
List	Schleswig	yes	Coast
Nordeney	Bremen		Coast
USWN			
Ridgecrest	Barstow		differ. regions
Blythe	Palm springs		differ. regions
PNW			
Burns	Whitehorse Ranch	yes	differ. regions

3.2 Smooth surface fitting

Study prepared by:	*Antoine Zelenka* *Swiss Meteorological Institute*
Under sponsorship from:	*Swiss Federal Office of Energy*

The fitting of polynomial surfaces

$$GH(x,y) = \sum_{j=0}^{n-2} \sum_{k=0}^{n-2-j} a_{jk} x^j y^k \qquad (3.2.1)$$

to n points becomes an ungrateful task as soon as n≥4 (recall that, e.g., the Swiss ANETZ consists of 41 stations in this study!). It is also a frustrating endeavour because an undesirable feature of these surfaces is to develop wavy edges in order to fit points in the centre. This "fluttering" is well known in y=f(x) polynomial fitting, but it is more severe in two dimensions where there is more boundary to be affected.

The tendency, therefore, is to turn to the so called *spline* methods. Many one-dimensional interpolation techniques are based on fitting some curve in each of the intervals between the data points and choosing the parameters of the curve to give continuity to a certain number of derivatives at each data point.

Defining spatial splines is, comparatively, a more difficult subject. Whereas the data points in one dimension divide the real line into intervals, it is not obvious how to use data points to divide a plane into regions.

Figure 3.2.1 illustrates one of the two possible basic contructs. To each point is associated a Dirichlet cell (also called Thiessen polygon or tile), which is that part of the plane that is nearer to that data point than to any other.

Note at this stage that, for extrapolation, a Dirichlet cell strictly defines the region around a site where values extrapolated from the site are valid. (Seen in a perspective plot, we have a discontinuous fit in which the plane over each tile is assigned the value at its defining data point.)

The second basic construct is the Delaunay triangulation on which the Akima quintic splines interpolation is based. This interpolation technique on an irregular grid is the only one available in mathematical software packages such as the IMSL etc.. A theoretical description of it is to be found in the theory chapter (Section 10.2.2).

The practical application of this method can be covered briefly in our context. Within a sample of 55 ANETZ stations for 30 days in April 1982, it predicted two negative daily totals (Zelenka [3.2.1]). Significantly enough, both sites where this happened were located on the convex border of the data points. Obviously, even by restricting the polynomial surfaces of segments of the plane to lower orders (whereby the resulting surface for the whole network is continuous and smooth), does not prevent the "fluttering" of the surface's edges.

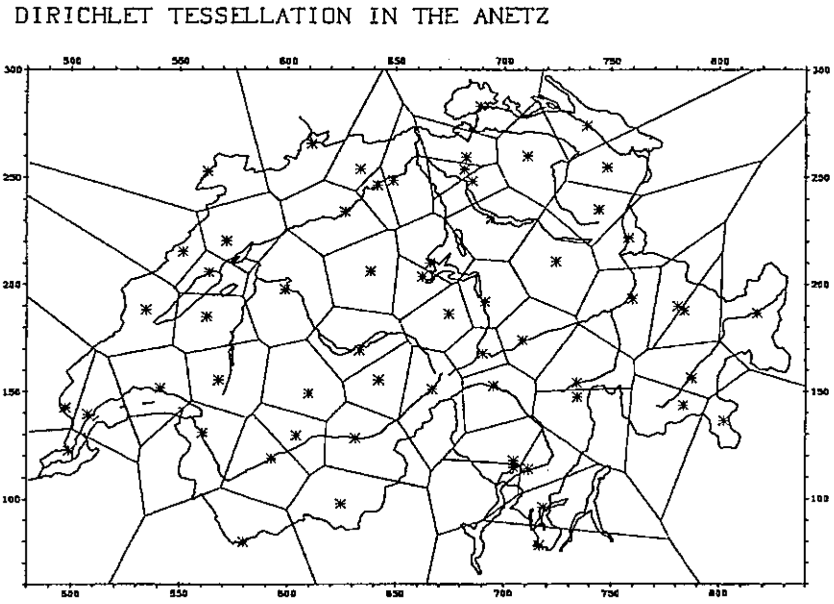


Fig. 3.2.1. One of the basic constructs to divide the plane into regions in the case of the Swiss ANETZ (the * mark the stations whose number, in this example, is larger than the 41 selected in this study). Note that each polygon delineates the validity domain of the extrapolated values discussed in Section 3.1!

Suggestions have been made to artificially extend the grid beyond the border defined by the network sites, thus placing the zone where the fluttering occurs outside the region of interest. However, introduction of fictitious observations at judiciously distributed points necessitates some extrapolation, e.g., by the method of weighted means to be discussed next in Section 3.3. But then, why not operate the whole interpolation with that technique? Finally, as the location of the supplementary points may have to change from case to case, the technique of smooth surface fitting was not selected for cross-validation in our study.

A last point is worthy of mention in the context of this section. Bi-cubic splines are a widespread tool for perspective (3-D) plots and for contouring. Yet, they necessitate data on a *regular* x,y grid. In this report, we are precisely concerned with this crucial problem of *gridding*.

References

[3.2.1] A. Zelenka, 1985: Satellite versus ground observation base model for global irradiation. *INTERSOL 85*, Proc. 1985 Biennial Congress ISES, K.G.T. Hollands & E. Bilgen Eds., Vol. 4, pp. 2513-2517, Pergamon 1986.

3.3 Weighted averages

Study prepared by: *Antoine Zelenka and Draginja Lazic*
Swiss Meteorological Institute

Under sponsorship from: *Swiss Federal Office of Energy*

3.3.1 *Method and presentation of results*

This technique is also known under the name of moving averages. As all those to be discussed in the forthcoming sections, it is based on the concept of *linear estimators*. Insolation $GH(\underline{x})$ at a location defined by its vector position $\underline{x}$ is estimated as a linear combination of the measured values $GH(\underline{x}_i)$ at the network locations $\underline{x}_i$:

$$GH(\underline{x}) = \Sigma\, w_i(\underline{x},\underline{x}_i)GH(\underline{x}_i) \qquad (3.3.1)$$

or, short,

$$GH = \Sigma\, w_i GH_i \qquad (3.3.2)$$

where the removal of bias is guaranteed by the constraint

$$\Sigma\, w_i = 1 \qquad (3.3.3)$$

We first note that extrapolation, as discussed in Section 3.1, is a special case of (3.3.2), namely $w_1 = 1$ for the closest neighbor and $w_i = 0$ for all other stations.

Intuitively one will argue that the further away a station lies from the site, the smaller should be its contribution to the predicted value. Thus, weights which are inversely proportional to a power of distance, viz.,

$$w_i \sim (1/d_i)^p \qquad (3.3.4)$$

are certainly a possible choice. Indeed, numerous studies which are reviewed by Hay [3.3.1] & [3.3.2] indicate that the standard deviation of daily differences of insolation varies as a power of station separation distance.

From a purely formal argument concerning the convergence of the sum (3.3.2), $p=2$ should in fact be a lower bound. The proof is to be found in Ripley [3.3.3]. Another possibility to guarantee convergence is to introduce a "search radius" R beyond which w_i is set to zero.

Both conditions are fulfilled in **Shepard's "gravity interpolation"** [3.3.4] where

$$w_i = [(1-\delta_i)^2/\delta_i^2]/\Sigma w_k \text{ , with } \delta_i = d_i/R \quad \text{for } d_i < R$$

and (3.3.5)

$$w_i = 0 \text{ otherwise.}$$

Here, Σw_k is the sum over all weights. It enforces the unbiased condition $\Sigma w_i=1$. The special form $(1-\delta_i)^2/\delta_i^2$ rather than $(1/d_i)^2$ is to enforce smoothness of the resulting, continuous interpolation surface.

The horizontal distance d_i between site and station i is determined by Eqns. (3.1.1)-(3.1.2) of Section 3.1. Alternatively, if *a priori* knowledge about the mean variabilty of the insolation field is given, the generalized distance D_i (3.1.3) can replace d_i.

The cross-validation results displayed in the following Figs. 3.3.1 to 3.3.12 are for gravity interpolation with generalized distances. The presentation format is exactly the same as in Section 3.1. The only difference is in the detailed Table 3.2 of Volume 3, where an additional list of the monthly search radii R ("S.R.") and vertical scale factors v ("V.S.F.") is displayed on the front page identifying each network.

Selection of these two parameters has been on a monthly basis by trial and error. In fact, Ripley [3.3.3] gives recipes about how to proceed analytically, but the domain of R's and V.S.F.'s which optimize the cross-validation is rather broad and this reduces the number of necessary iterations to less than 6 per network. Also, the selected pair is not the one which makes the overall validation statistics (all stations and all days) best in the last decimal. Rather, care has been taken to find that combination which keeps the largest individual monthly RMSE% as small as possible.

VALIDATION OF GRAVITY INTERPOLATION

YEAR MBE [% MEAS.MEAN OF SITE]

NETWORK

NO. OF DAYS	2993
MEAS. MEAN [MJm^{-2}]	14.36
MBE % OF MEAS.MEAN	-1.1
RMSE % OF MEAS.MEAN	22.7

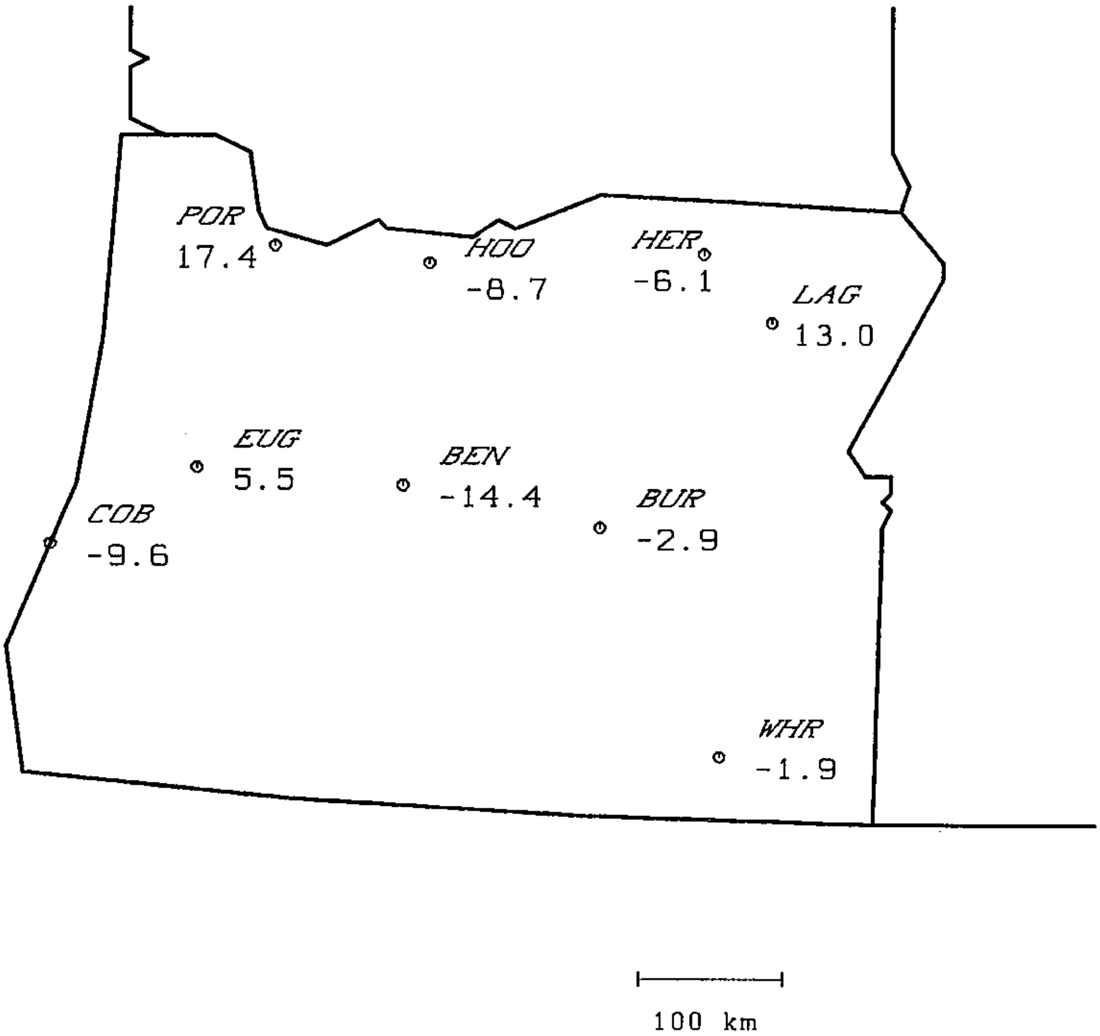


Fig. 3.3.1 : MBE% Pacific Northwest.

VALIDATION OF GRAVITY INTERPOLATION

YEAR RMSE [% MEAS.MEAN OF SITE]

NETWORK

NO. OF DAYS	2993
MEAS. MEAN [MJm^{-2}]	14.36
MBE % OF MEAS.MEAN	-1.1
RMSE % OF MEAS.MEAN	22.7

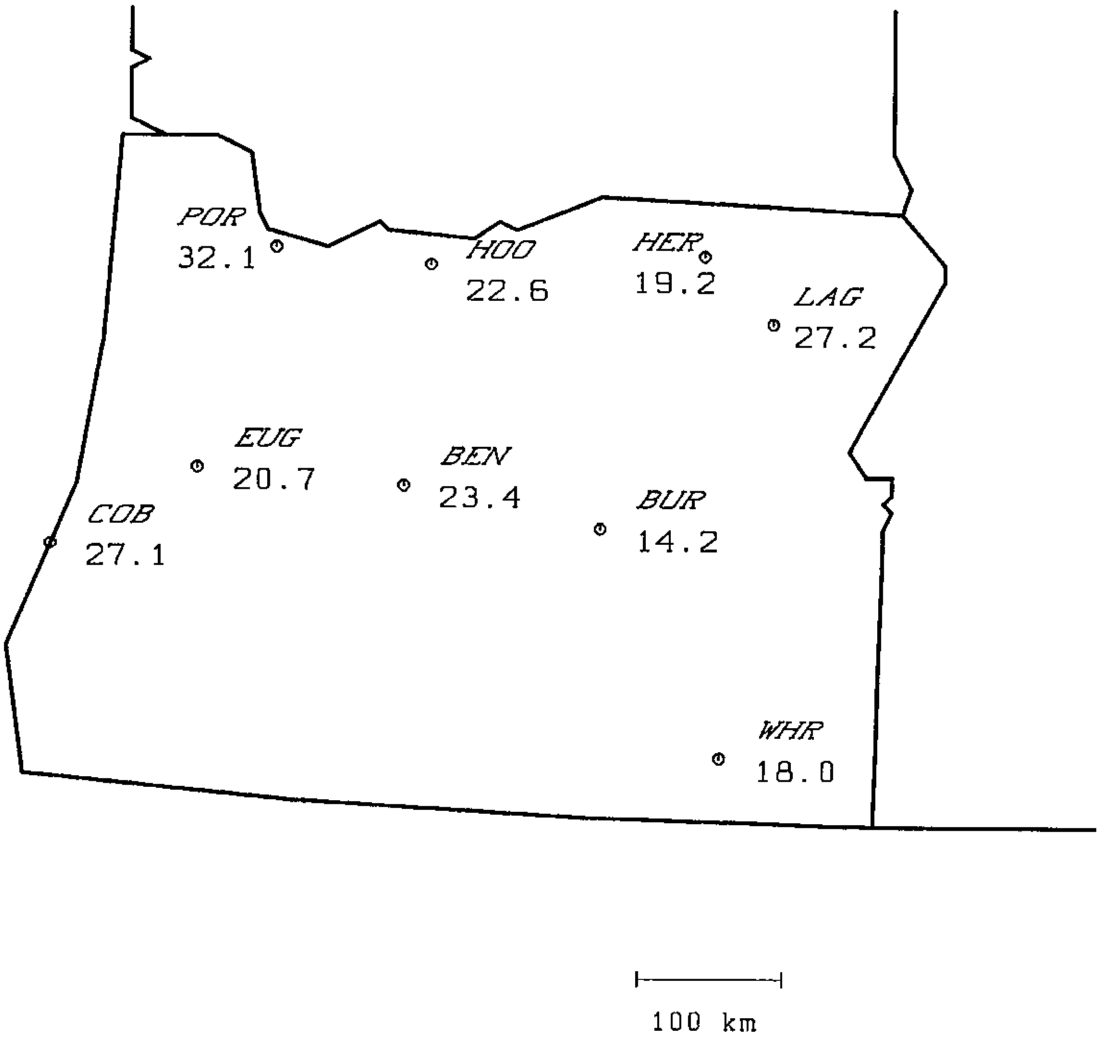
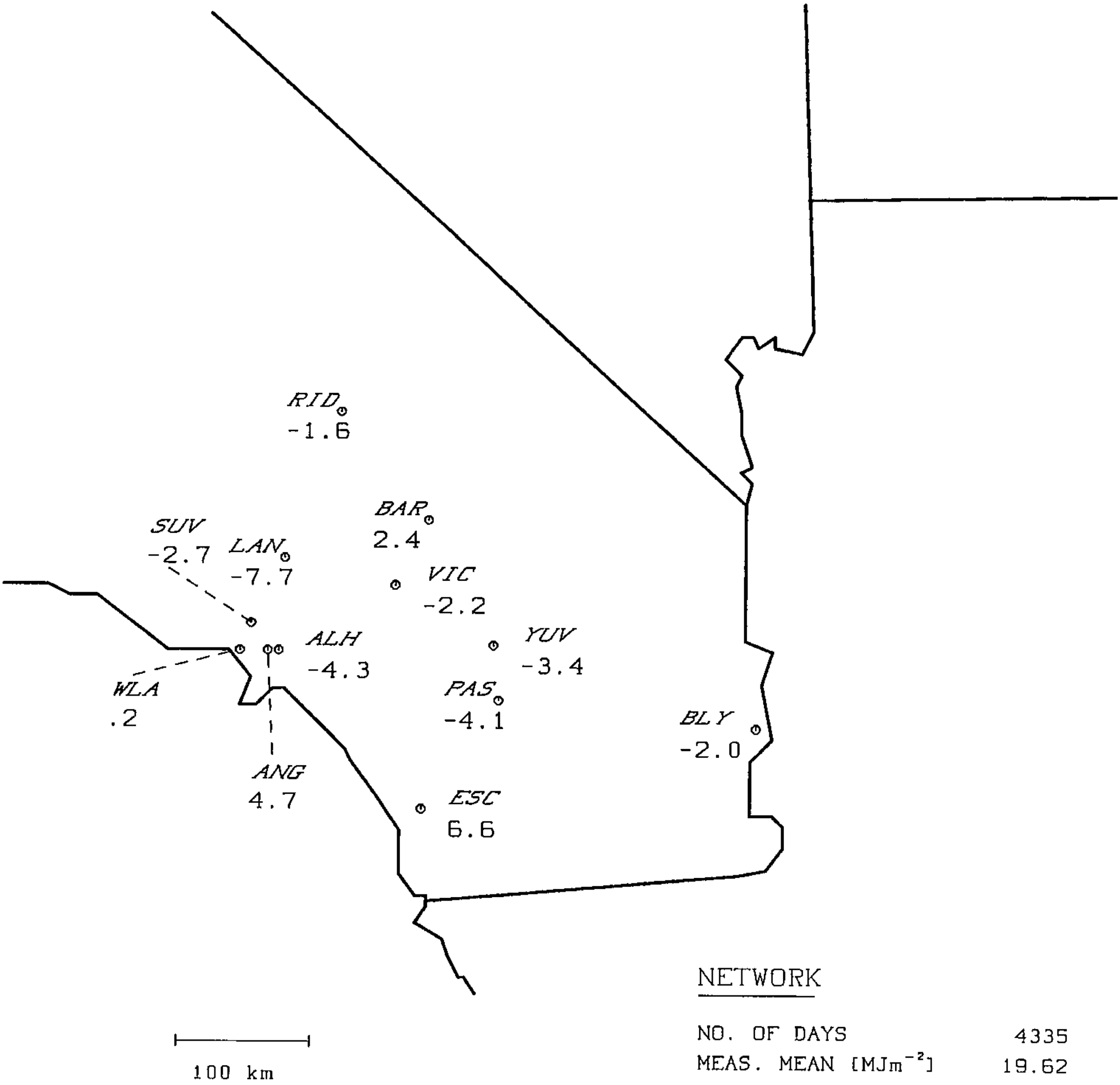


Fig. 3.3.2 : RMSE% Pacific Northwest.

VALIDATION OF GRAVITY INTERPOLATION

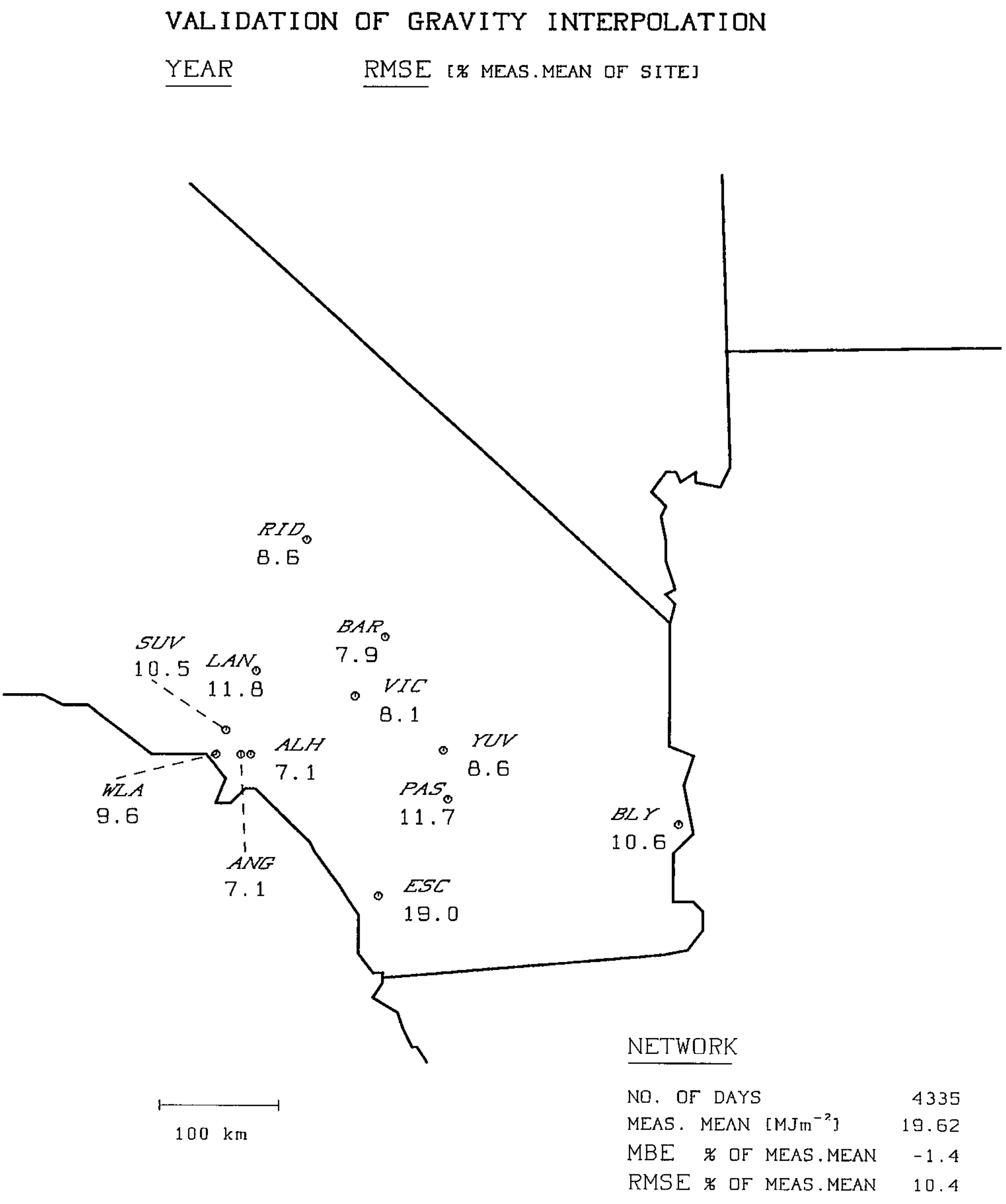
YEAR MBE [% MEAS.MEAN OF SITE]



NETWORK

NO. OF DAYS	4335
MEAS. MEAN [MJm^{-2}]	19.62
MBE % OF MEAS.MEAN	-1.4
RMSE % OF MEAS.MEAN	10.4

Fig. 3.3.3 : MBE% West Associates.



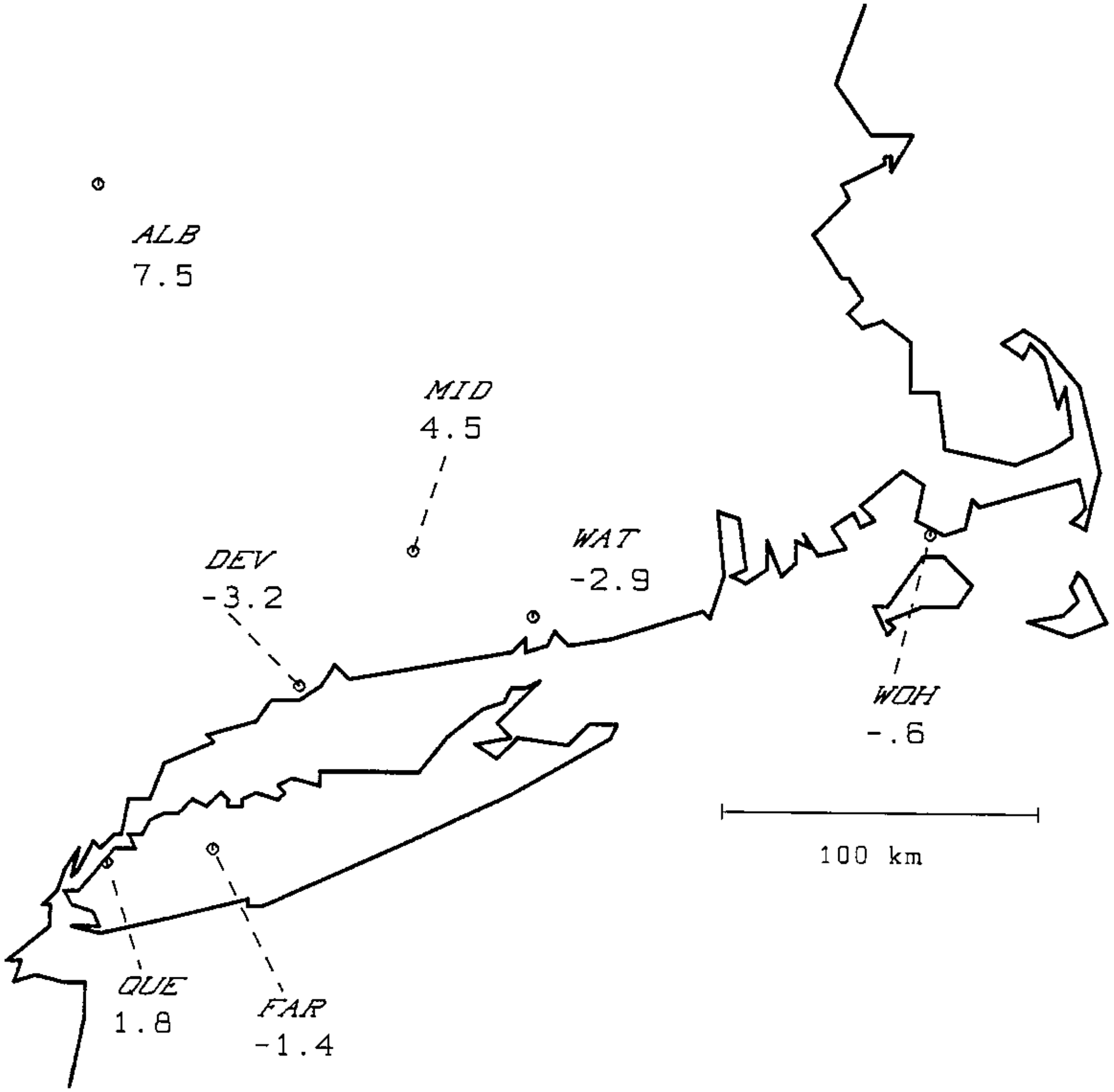
NO. OF DAYS	4335
MEAS. MEAN [MJm^{-2}]	19.62
MBE % OF MEAS.MEAN	-1.4
RMSE % OF MEAS.MEAN	10.4

Fig. 3.3.4 : RMSE% West Associates.

Fig. 3.3.5 : MBE% Northeast Utilities & SUNY.

VALIDATION OF GRAVITY INTERPOLATION

YEAR MBE [% MEAS.MEAN OF SITE]



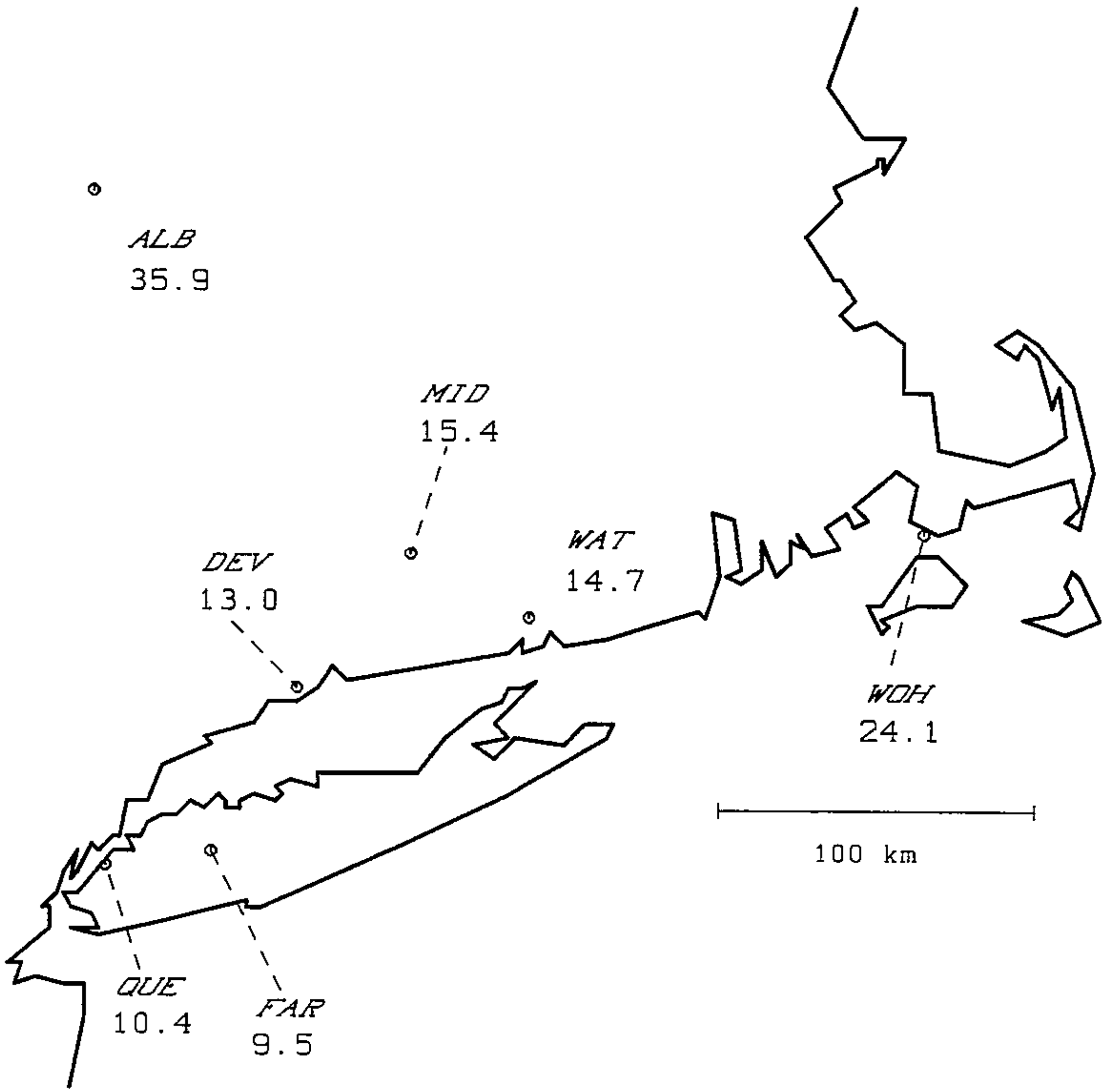
NETWORK

NO. OF DAYS	938
MEAS. MEAN [MJm^{-2}]	12.59
MBE % OF MEAS.MEAN	.7
RMSE % OF MEAS.MEAN	19.1

Fig. 3.3.6 : RMSE% Northeast Utilities & SUNY.

VALIDATION OF GRAVITY INTERPOLATION

YEAR RMSE [% MEAS.MEAN OF SITE]



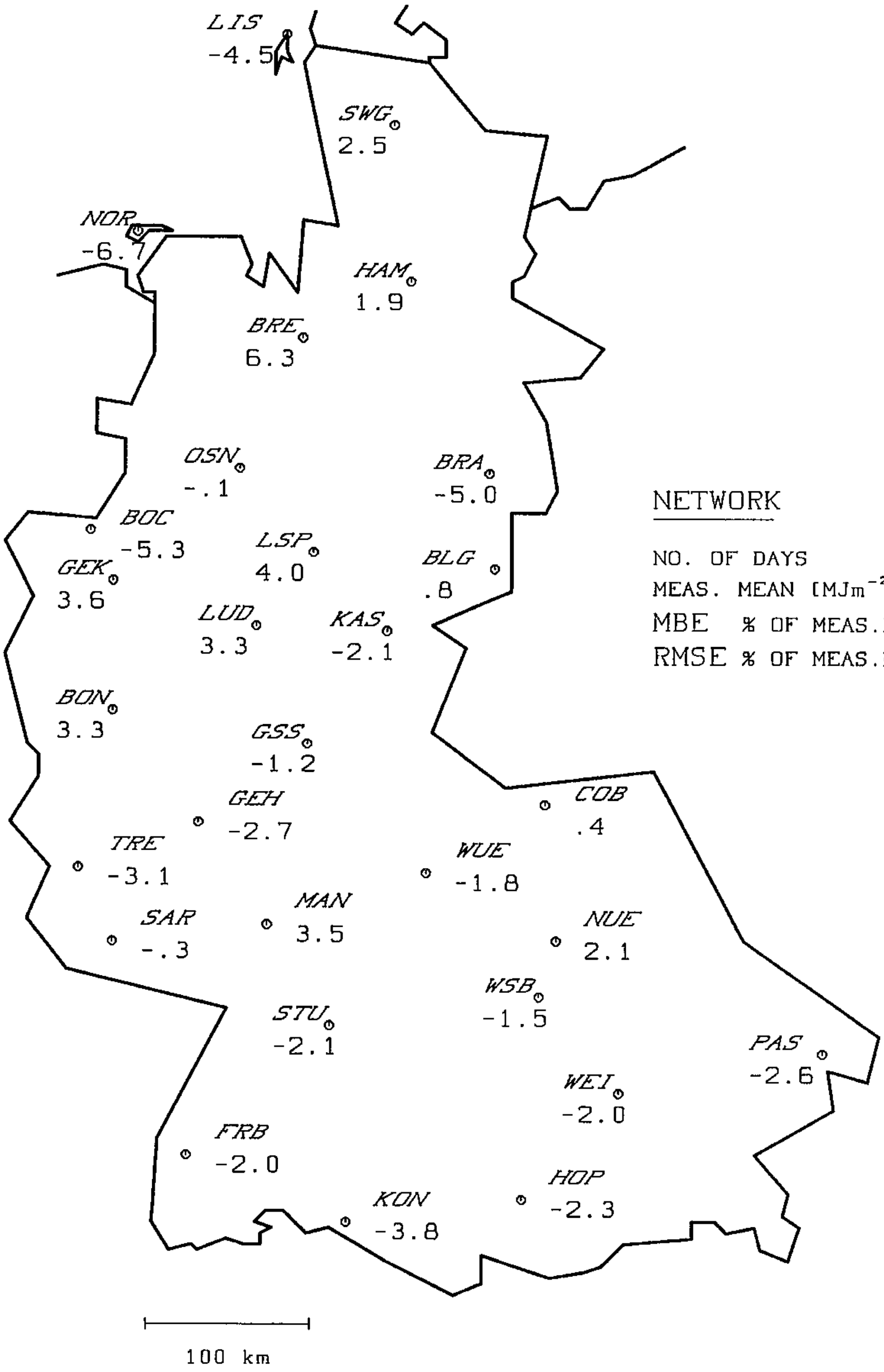
NETWORK

NO. OF DAYS	938
MEAS. MEAN [MJm^{-2}]	12.59
MBE % OF MEAS.MEAN	.7
RMSE % OF MEAS.MEAN	19.1

Fig. 3.3.7 : MBE% Germany DWD.

VALIDATION OF GRAVITY INTERPOLATION

YEAR MBE [% MEAS.MEAN OF SITE]



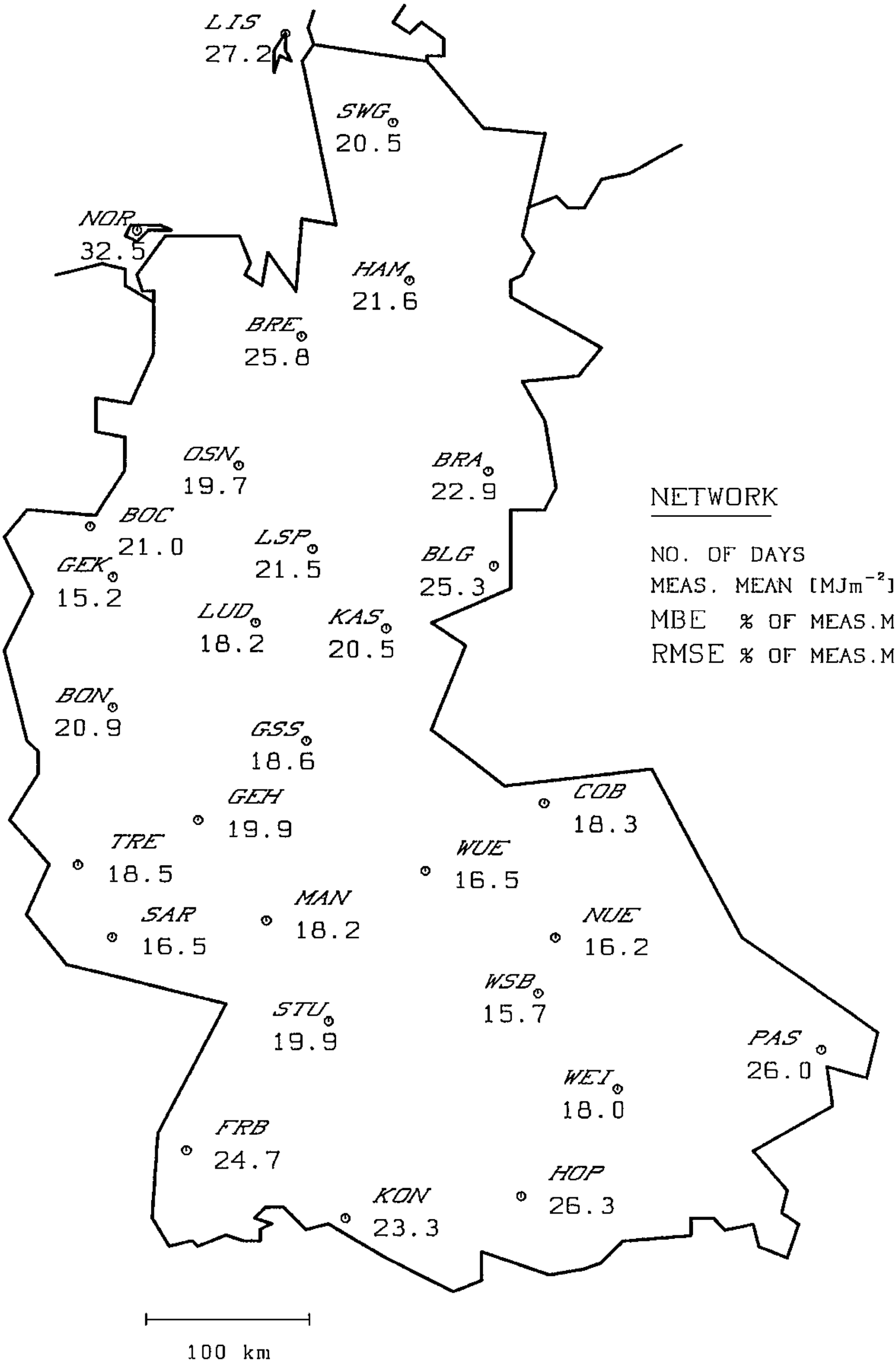
NETWORK

NO. OF DAYS	10539
MEAS. MEAN [MJm^{-2}]	9.80
MBE % OF MEAS.MEAN	-.8
RMSE % OF MEAS.MEAN	21.5

Fig. 3.3.8 : RMSE% Germany DWD.

VALIDATION OF GRAVITY INTERPOLATION

YEAR RMSE [% MEAS.MEAN OF SITE]



NETWORK

NO. OF DAYS	10539
MEAS. MEAN [MJm^{-2}]	9.80
MBE % OF MEAS.MEAN	-.8
RMSE % OF MEAS.MEAN	21.5

VALIDATION OF GRAVITY INTERPOLATION

YEAR MBE [% MEAS.MEAN OF SITE]

NETWORK

NO. OF DAYS	2573
MEAS. MEAN [MJm^{-2}]	15.81
MBE % OF MEAS.MEAN	-.1
RMSE % OF MEAS.MEAN	11.8

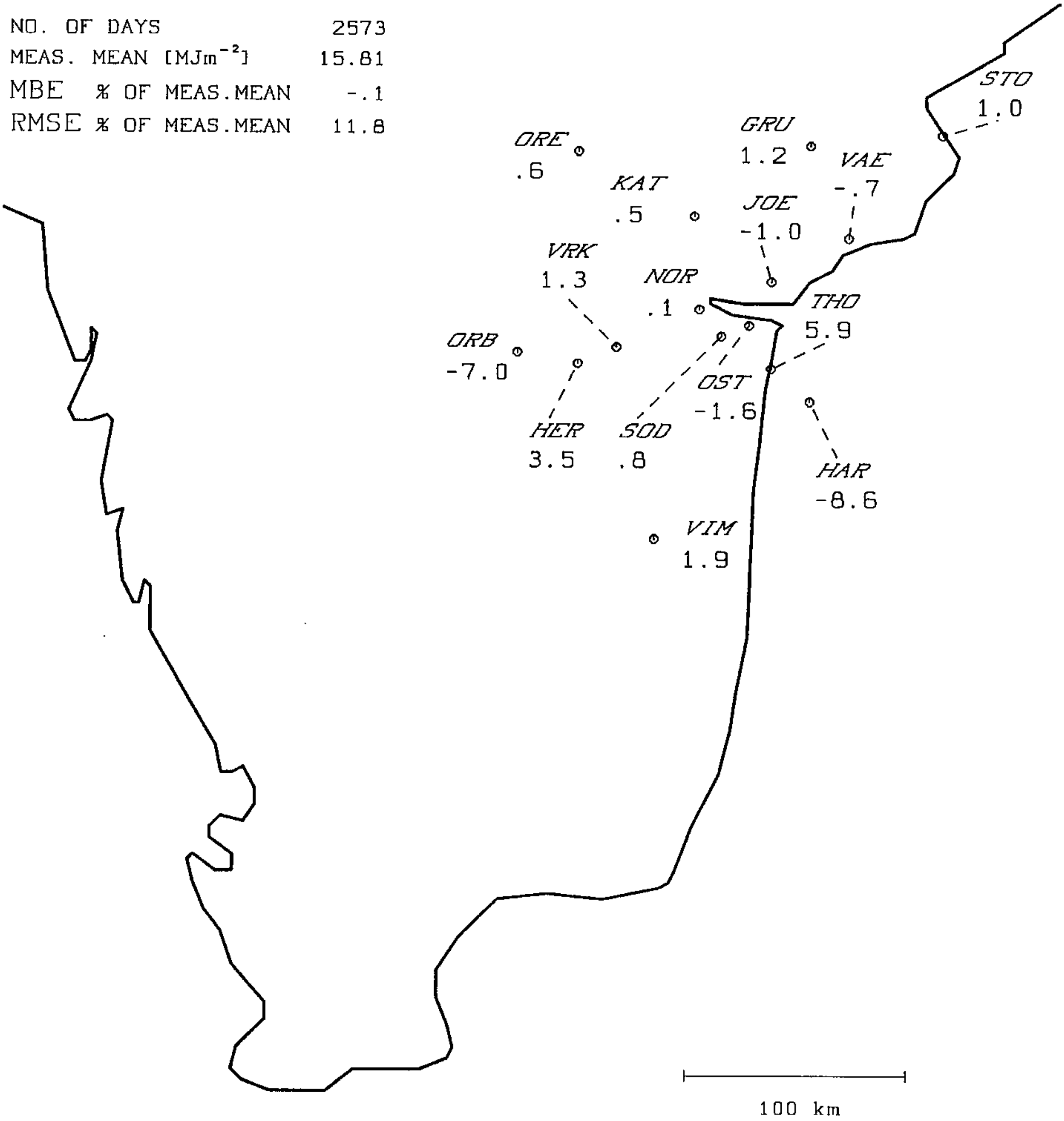


Fig. 3.3.9 : MBE% Sweden mesoscale.

VALIDATION OF GRAVITY INTERPOLATION

YEAR RMSE [% MEAS.MEAN OF SITE]

NETWORK

NO. OF DAYS	2573
MEAS. MEAN [MJm^{-2}]	15.81
MBE % OF MEAS.MEAN	-.1
RMSE % OF MEAS.MEAN	11.8

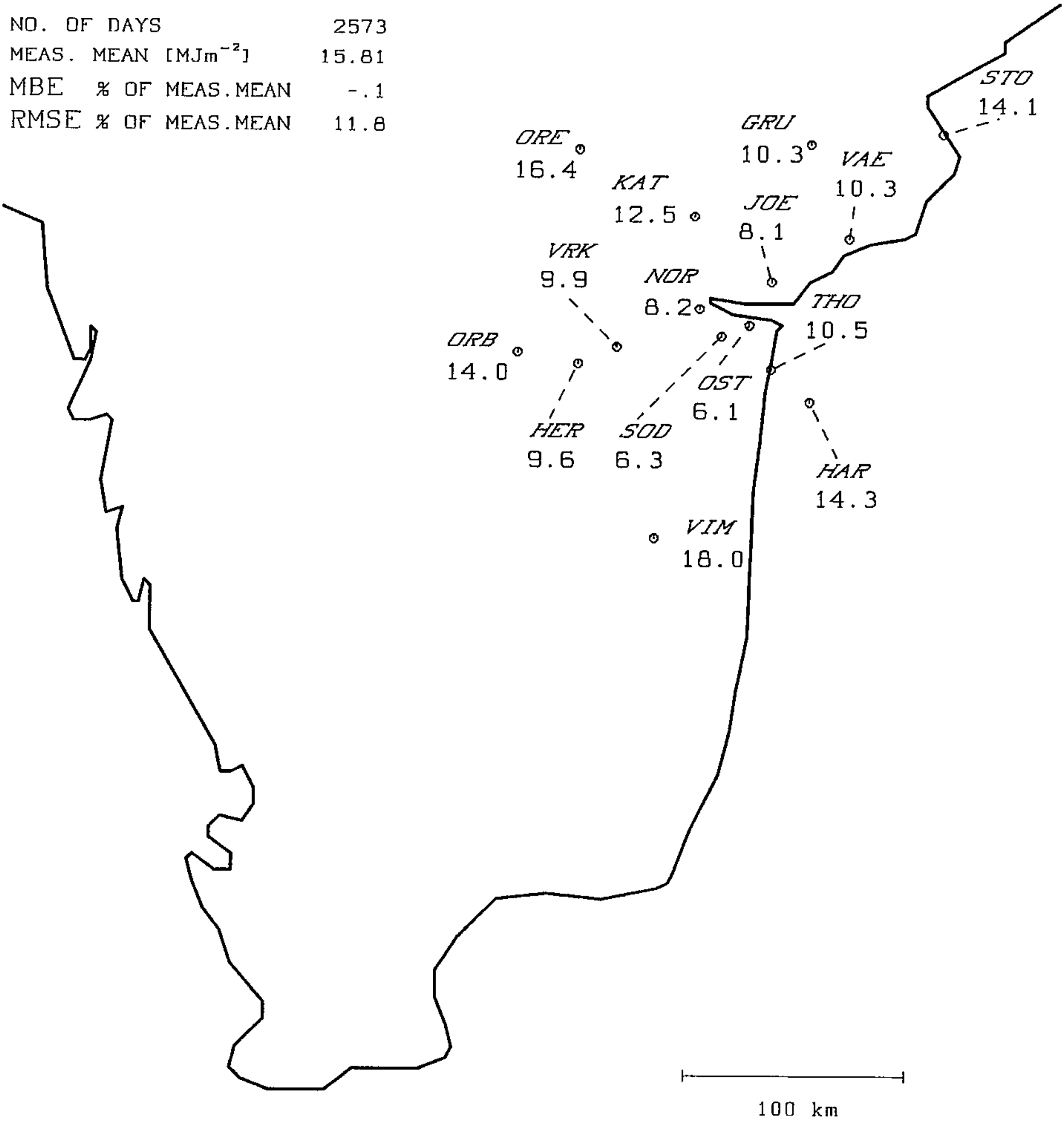
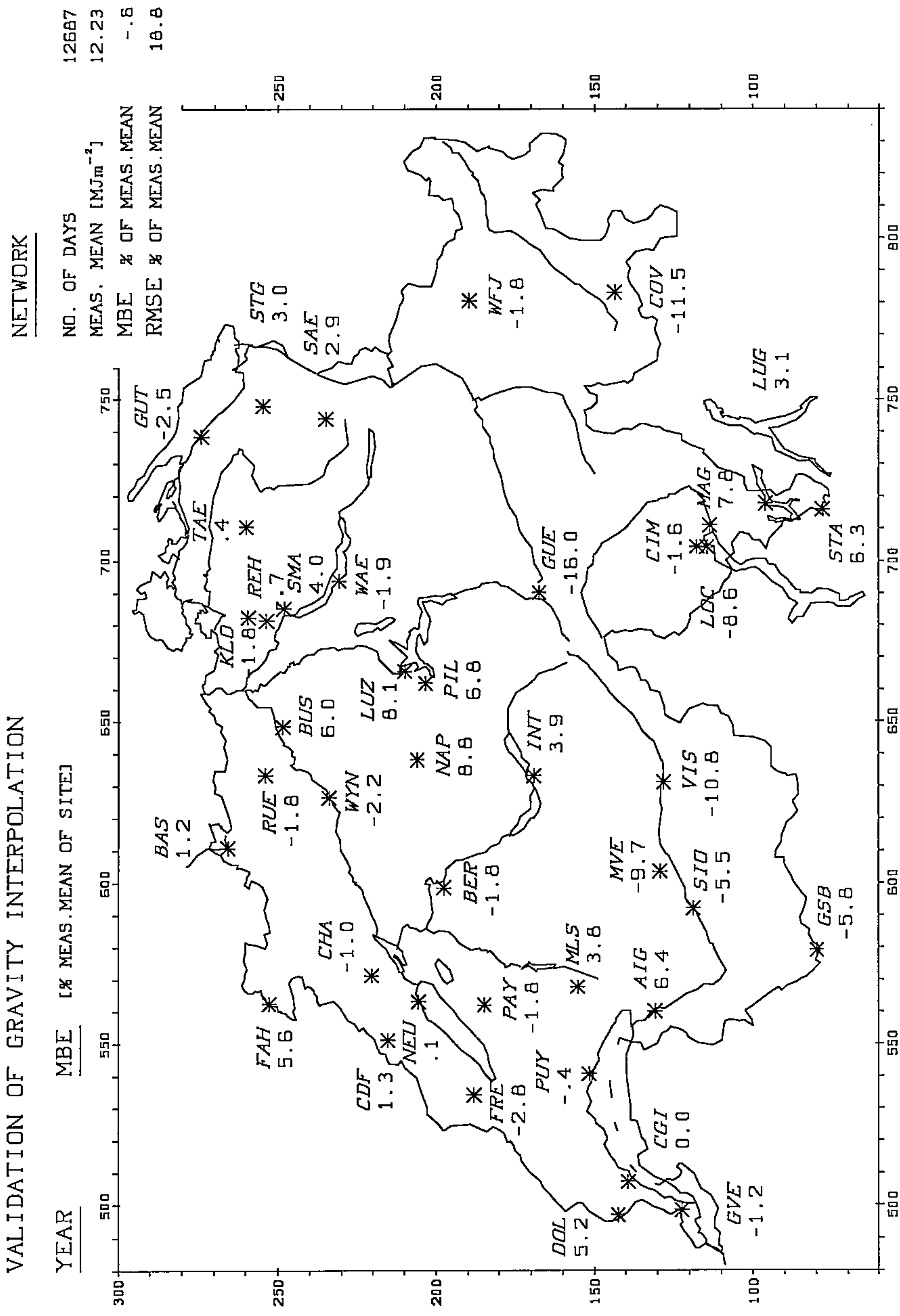


Fig. 3.3.10: RMSE% Sweden mesoscale.

Fig. 3.3.11: MBE% Switzerland ANETZ.



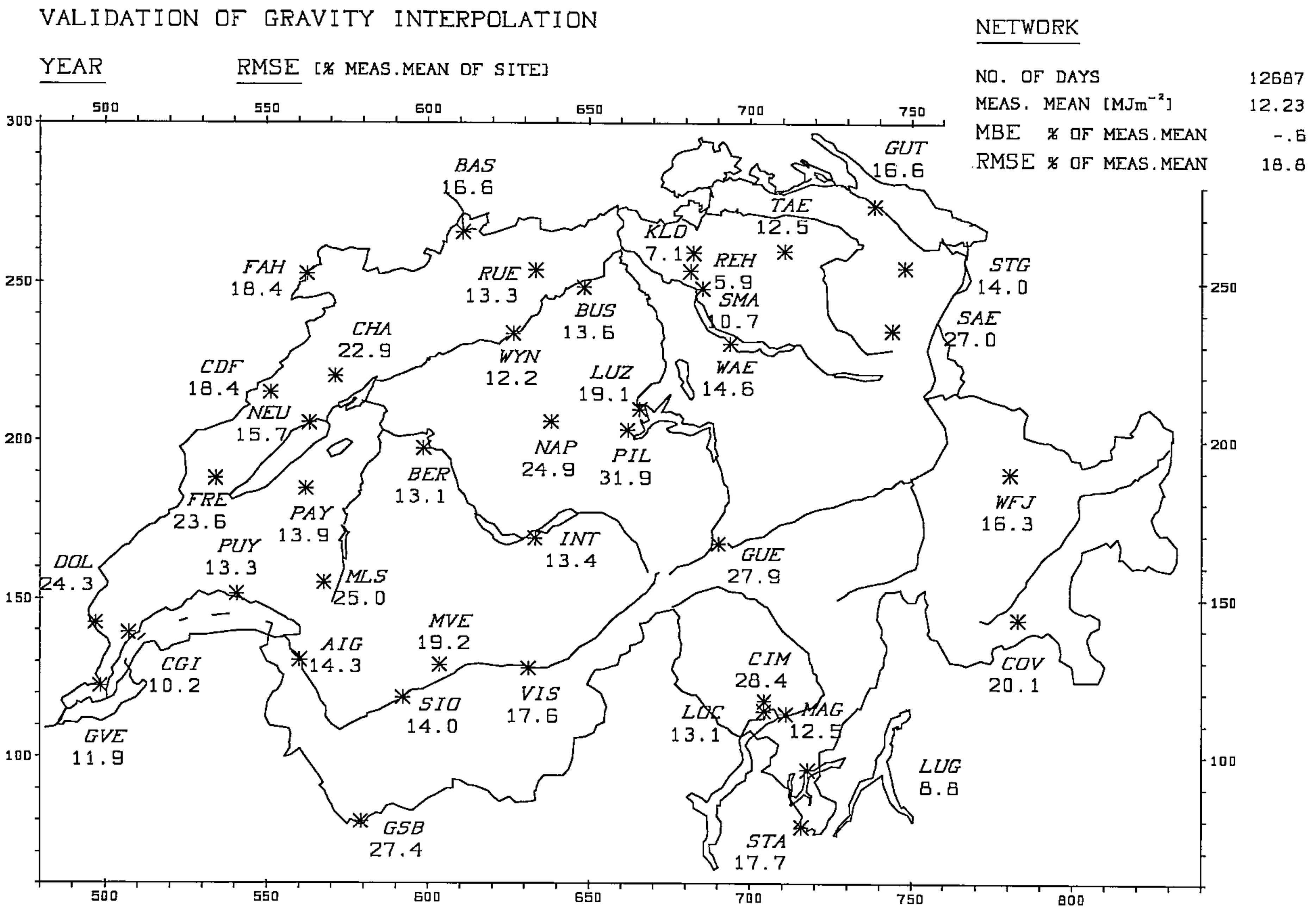


Fig. 3.3.12: RMSE% Switzerland ANETZ.

3.3.2 *Discussion of cross-validation results*

For all networks, gravity interpolation yields better cross validation results than does mere extrapolation. Seasonal trends are still present, but the whole network RMSE% now vary between 40% in winter and 15% in summer. Maximum individual monthly RMSE% can still be large in winter, but the November 126% for Bremen in the DWD network are an (explainable) exception.

As there were some indications that variability as function of station separation could follow a linear law in the DWD network [3.3.5], weights proportional to $(1-d_i)/d_i$ have also been tested for it. However, the square law eventually appeared to yield the better results.

Tuning of the search radius ("S.R.") R and of the vertical scale factor ("V.S.F") v on the monthly time scale significantly improved the results only in the case of the Swiss ANETZ. There, due to the very pronounced orography, the radiation climate varies across a hectometer altitude difference as much as across a 10 km (approximate) horizontal distance ($v \approx 100$).

The stations Las Vegas and Page were removed from the U.S. Southwest network to allow real selection of the search radius. When they are parts of the network, their separation as well as their distance from the other stations implies such a large R that discrimination among the other stations is no longer effective.

Yet in all cases, the search radius was always kept large enough to guarantee contribution of at least two stations to the estimates (two terms in the sum (3.3.2)). This condition was imposed because it was found that if "pollution" of the gravity method by extrapolation is indulged, then the cross-validation results lose their generality as they, then, depend upon the occurence of few specific site pairs within the network.

The selected search radii range from 150km for the SMHI network to 600km for the PNW network. Only in this last case is R found equal to the size of the whole network, leading to the situation that all stations always contribute to each estimate. This is probably a consequence of the low density of the network. Such an interpretation is sustained by the behavior of the RMSE% as a function of the sum of weights, a relation first suggested by Perez *et al.* [3.3.6] (see also Section 6.2) for the USNEU network.

Figure 3.3.13 reproduces the finding of Perez *et al.*. The yearly RMSE% for each station is plotted versus $\log(\Sigma w_i')$, the w_i' being the weights (before normalization, i.e., (3.3.5) not divided by Σw_k) entering the estimate. Clearly, for a given R, these w_i' grow larger as distances d_i from the contributing stations get smaller.

Whereas a similar relation emerges for all the other networks, or at least, for their climatically homogeneous parts (Figs. 3.3.14 and 3.3.15), it fails for the PNW.

RMSE% as function of $\log(\Sigma w_i')$ cannot be so straightforwardly interpreted as relation (3.1.4), i.e., as RMSE% $= 2.2 + 2.5\sqrt{d_{km}}$ which pertains to extrapolation. Such a relation emerges if an "average distance"

$$D = \Sigma w_i D_i \qquad (3.3.6)$$

is introduced where the D_i are generalized distances (3.1.3) for the cases where V.S.F. > 1. Then, linear regression analysis of RMSE% versus $\sqrt{D_{km}}$, results in

$$\text{RMSE\%} = 2.6 + 1.7\sqrt{D_{km}} \ [\%], \qquad 9 \text{ km} \le D_{km} \le 196 \text{ km} \qquad (3.3.7)$$

Because of possible monthly variations of either search radius and vertical scale factor, or of the number of contributing sites, the average distance is taken as the mean of the possible 12 monthly average distances.

Figure 3.3.16 displays the relation (3.3.7). A substantial part of the scatter can be traced to differences among the networks, but this point will not be elaborated further here. Rather, attention is drawn to the close agreement of this relation with Hay's analysis [3.3.7], which was based on optimum interpolation in a 12 sites mesoscale network located in the Vancouver region.

Because the moving average estimations encompass contributions by sites located beyond the next neighbor, the domain of average distances is necessarily wider than that of extrapolation distances. Yet, at the interval's lower limit (D=9 km) the RMSE% of gravity interpolation amounts to 7.7%, while for extrapolation it is 9.7%. Also, in the overlapping zone of both domains, i.e., for 9 km ≤ D ≤ 165 km, the interpolation's RMSE% is systematically below the extrapolation's RMSE%. At the interval's upper limit the former amounts to 24% while the latter reaches 34%. This corresponds to the lower boundary of the interval stated in WMO Technical Note No. 172 [3.3.8], where interpolation between sites in temperate latitudes is reported to reduce the estimation error (as compared to extrapolation) by 1.5 to 2 times in the case of daily irradiation.

3.3.3 *References*

[3.3.1] J.E. Hay: 1984. An assessment of the mesoscale variability of solar radiation at the Earth's surface. *Solar Energy* **32**, 425-434.

[3.3.2] J.E. Hay and K.J. Hanson: 1985. Evaluating the solar resource: a review of problems resulting from temporal, spatial and angular variations. *Solar Energy* **34**, 151-161.

[3.3.3] B.D. Ripley: 1981. Spatial Statistics. Wiley & Sons.

[3.3.4] D.S. Shepard: 1984. Computer mapping: the SYMAP interpolation algorithm. In G.L. Gaile and C.J. Willmott Eds.: 1984, Spatial Statistics and Models. D. Reidel.

[3.3.5] R. Trapp and F. Kasten: 1988. Kleinskalige Variabilität der Sonnenstrahlung. Berichte des DWD, Deutscher Wetterdienst, D-6050 Offenbach a.M., FR Germany.

[3.3.6] R. Perez, R. Seals, and R. Stewart: 1990. Assessing photovoltaic interaction with New York summer-peaking utilities: An evaluation of existing solar resource assessment capabilities from measuring networks and geostationary satellites. Rept. to the New York Power Authority, Atmospheric Sciences Research Center, State University of NY at Albany, Albany NY 12222.

[3.3.7] J.E. Hay: 1891. The mesoscale distribution of solar radiation at the Earth's surface and the ability of satellites to resolve it. In R.J. Bahm Ed.: Satellites and forecasting of solar radiation. Proc. 1st Workshop on Terrestrial Solar Resource Forecasting and on the Use of Satellites for Terrestrial Solar Resource Assessment, Newark, DE. ASES 111pp.

[3.3.8] WMO - No. 557, 1981: "Meteorological Aspects of the Utilization of Solar Radiation as an Energy Source". WMO Technical Note No. 172, CH-1211 Geneva 20.

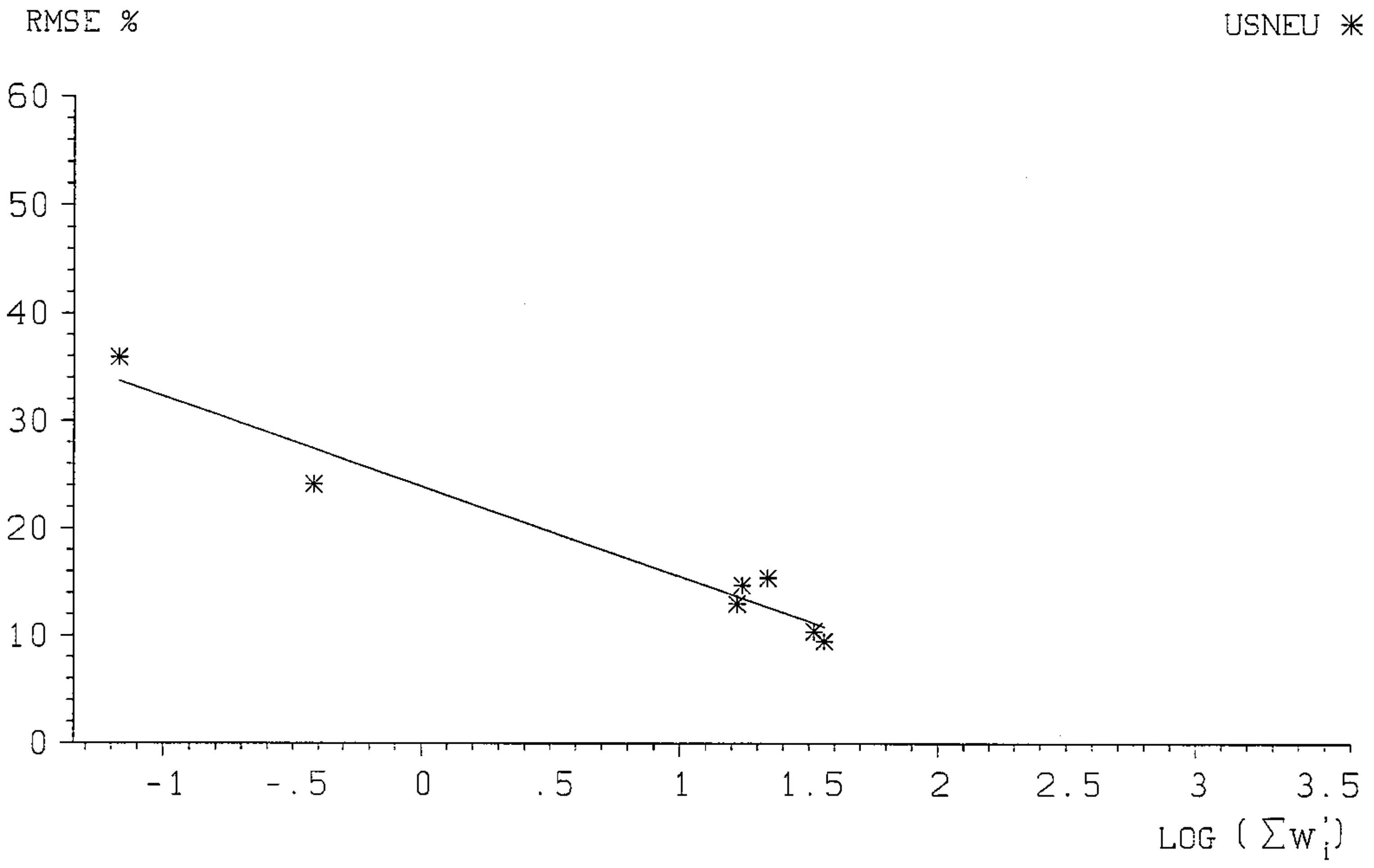


Fig. 3.3.13: Variation of interpolation RMSE (in % of pertaining, observed mean) as a function of the sum of interpolation weights for the Northeast Utilities & SUNY network.

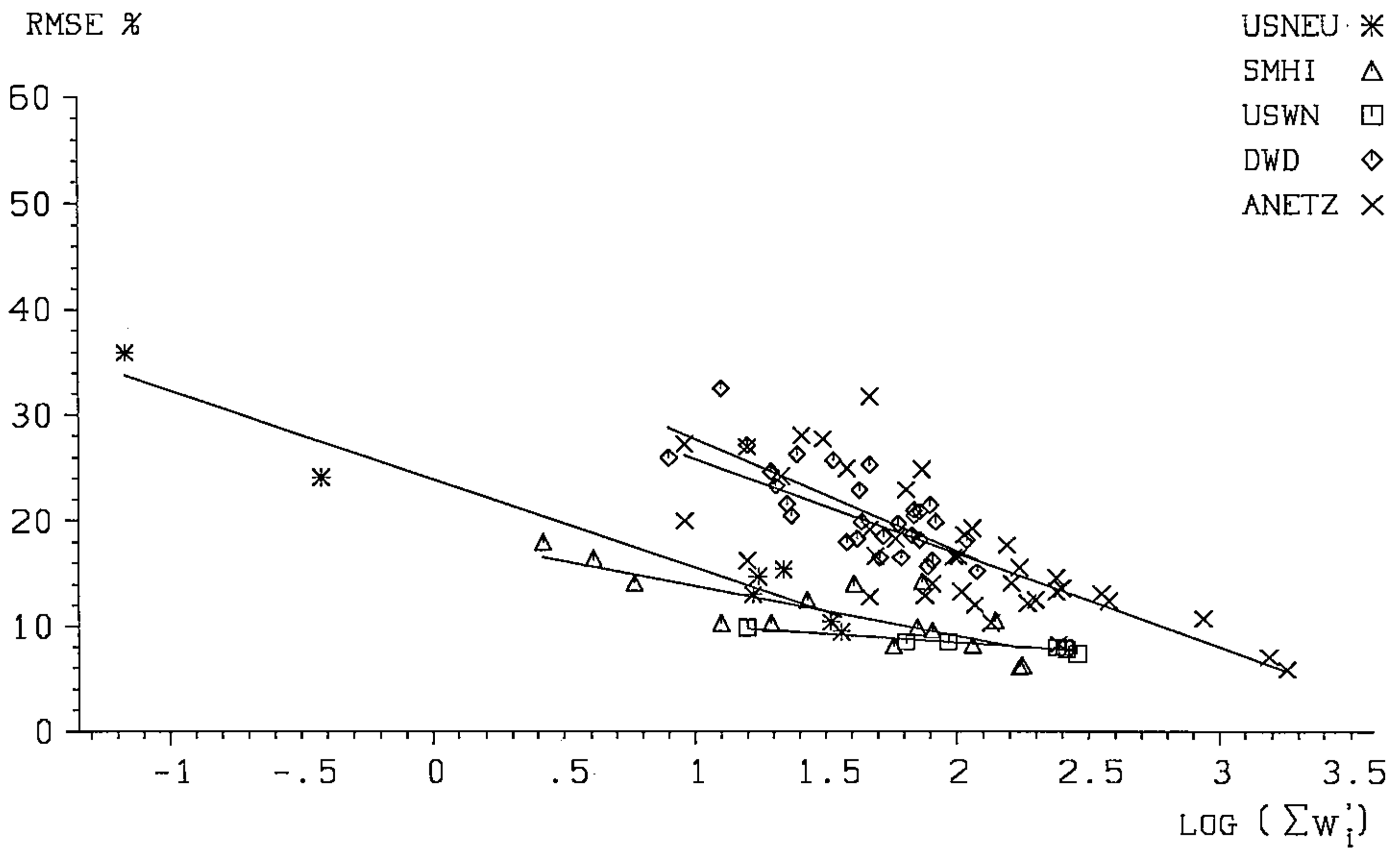


Fig. 3.3.14: Variation of interpolation RMSE (in % of pertaining, observed mean) as a function of the sum of interpolation weights for five different networks. Linear regression is for each network separately.

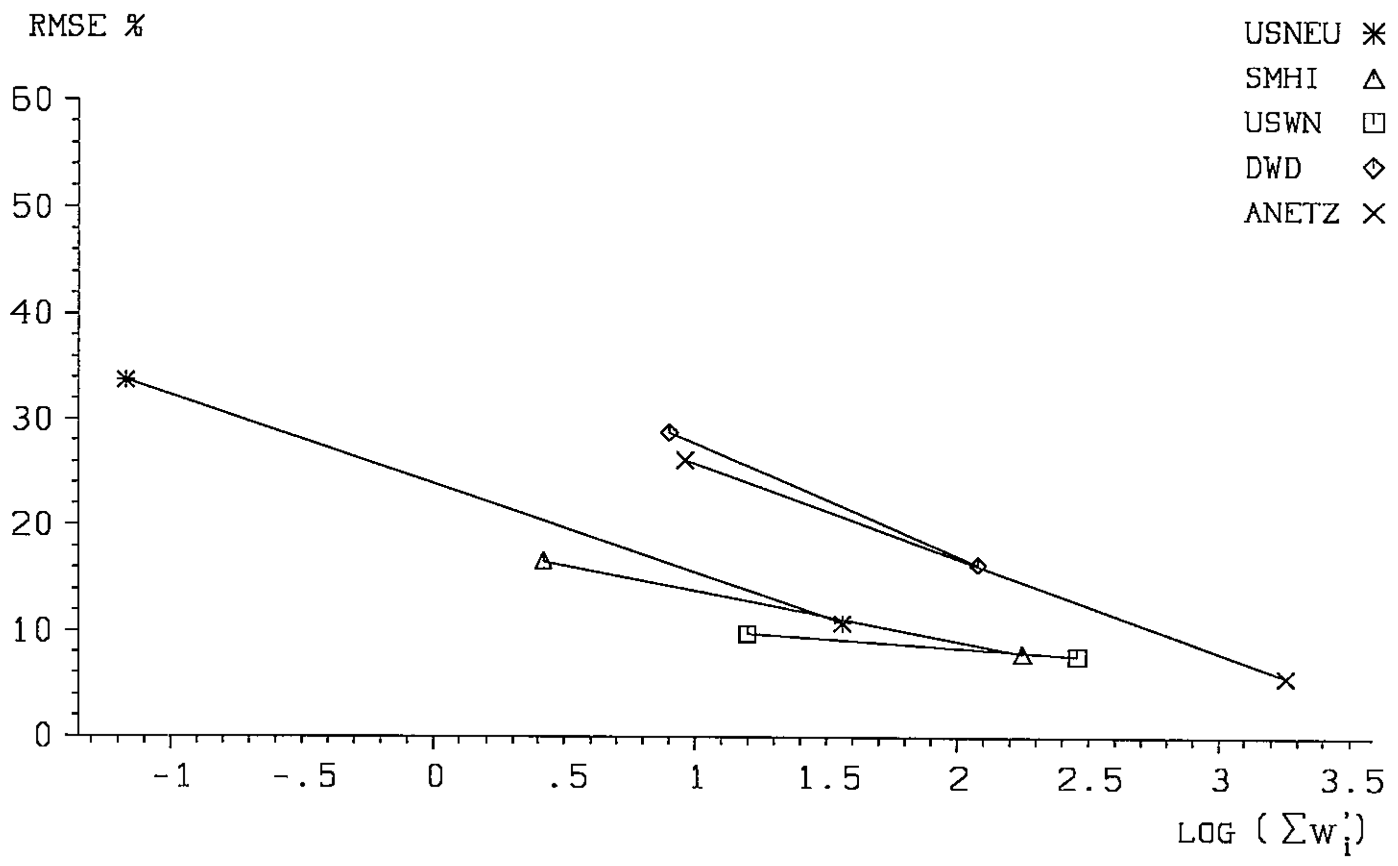


Fig. 3.3.15: Same as Figure 3.3.14, without individual points but with network identifiers at the ends of each regression line.

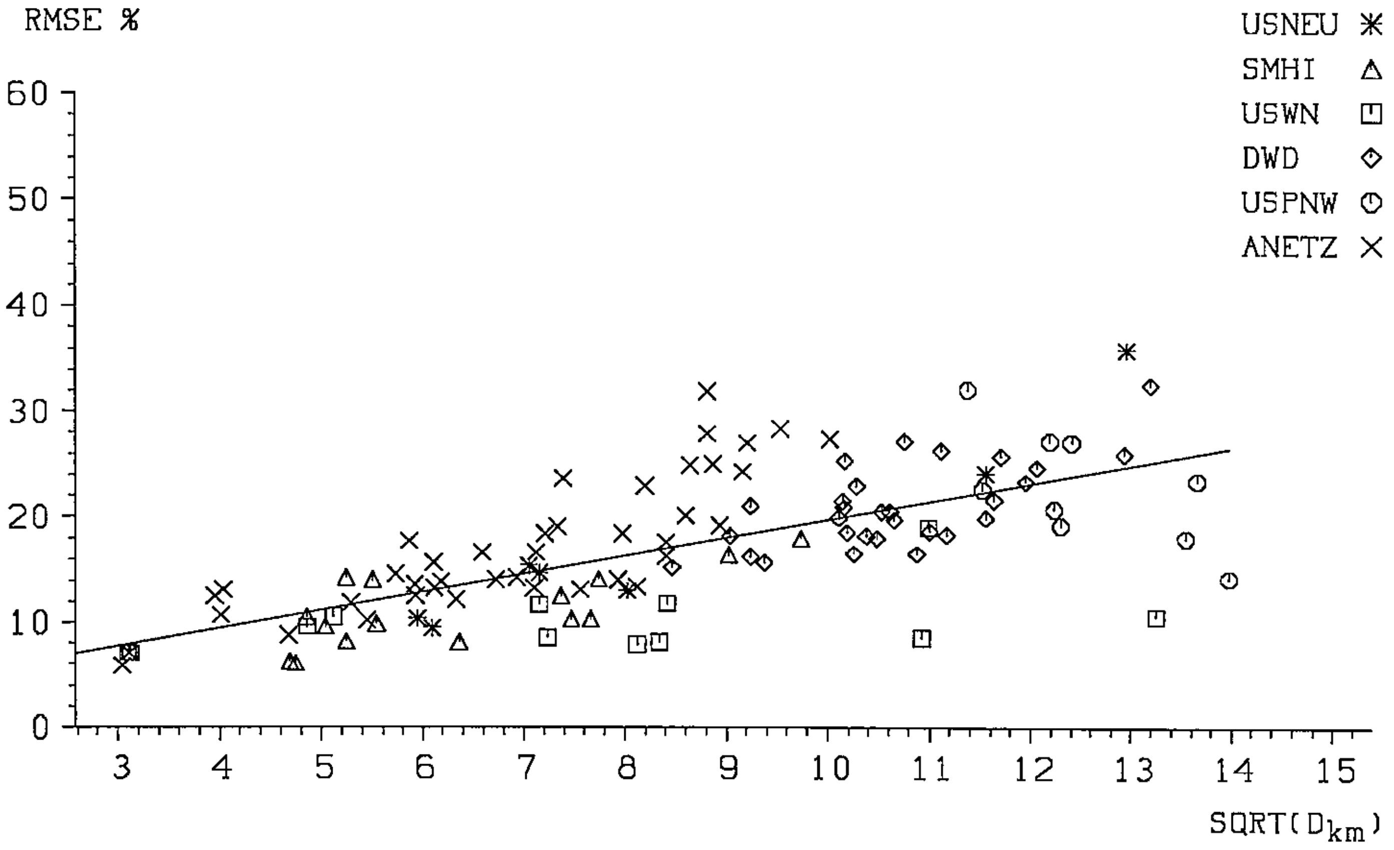


Fig. 3.3.16: Correlation between RMSE (in % of pertaining, observed mean) of one year of interpolated daily global irradiation and the square root of average interpolation distance D[km].

3.4 Kriging

Study prepared by: *Antoine Zelenka and Richard Perez*
Swiss Meteorological Institute and Atmospheric Sciences Research Center respectively

Under sponsorship from: *Swiss Federal Office of Energy and New York Power Authority, National Renewable Energy Laboratory (SERI) & The Mobil Foundation respectively*

3.4.1 *Basic considerations*

As in the case of gravity interpolation, the insolation GH for a given location is approximated by the linear estimator (3.3.1), respectively (3.3.2), the outcome of which we now designate GH*. Thus

$$GH^* = \Sigma \lambda_i GH_i \qquad (3.4.1)$$

where the GH_i are, as before, the values measured at the network stations, while now, the related weights are written λ_i to be consistent with the current literature on kriging.

Now, rather than postulating some *a priori* fixed relation between the weights λ_i and the distances to the stations, as was done for moving averages, one considers the actual structure of the insolation field for determining the relationship so that GH* reproduces GH optimally.

The optimality conditions read

$$E[GH^* - GH] = 0 \qquad (3.4.2)$$

and

$$Var[GH^* - GH] = Minimum \qquad (3.4.3)$$

where the operator "E" stands for the mathematical expectation and "Var" stands for the variance of the expressions in square brackets. The first condition requires the estimate to be unbiased, while the second requires it to have minimum variance.

The actual structure of the insolation field is primarily characterized by the n samples GH_i but this is barely enough for a straightforward determination of the λ_i from Eqns. (3.4.2) and (3.4.3). Additional hypotheses are required, which are discussed in detail in Chapter 10 (Theory Volume) and of which only the most widely used is sketched below.

Basically, each sample point (observed value at a station) GH_i is considered as the outcome of a random variable, which means that GH is considered as a random function. Specific hypotheses as to the two first moments of this random function allow statistical inference or, in other words, lead to enough conditions to allow a unique determination of the λ_i.

The usual hypothesis in random function theory is that of stationarity, that is, of invariance of the distribution law under translations. *Stationarity in the weak sense* holds when only the first two moments of the law are invariant under translation. In kriging, the most useful working hypothesis is the *intrinsic hypothesis* with a constant mean. There, it is assumed that *only the increments* of the function are stationary in the weak sense. For any $\underline{h}$, the increment $GH(\underline{x}+\underline{h})-GH(\underline{x})$ has an expectation and a variance which are independent of the location $\underline{x}$:

$$E[GH(\underline{x}+\underline{h}) - GH(\underline{x})] = 0 \qquad (3.4.4)$$

$$Var[GH(\underline{x}+\underline{h}) - GH(\underline{x})] = 2\gamma(\underline{h}) \qquad (3.4.5)$$

The advantage of this hypothesis is that, although it is accepted as constant, the mean does not have to be explicitly known. The function $\gamma(\underline{h})$ is called the *(semi-)variogram* and it is the basic tool for structural interpretation as well as for the estimation process itself. Indeed, when an acceptable function $\gamma(h)$ can be fitted to the data, then conditions (3.4.4) and (3.4.5) yield enough constraints to guarantee a unique solution to Eqns. (3.4.2) and (3.4.3), and a unique estimator GH* for any considered site.

3.4.2 *The variogram*

Under the intrinsic hypothesis,

$$\gamma(h) = (1/2)E[\{GH(x+h)-GH(x)\}^2] \qquad (3.4.6)$$

because of assumption (3.4.4) (recall that $Var[Z(x)] = E[Z(x)^2] - E[Z(x)]^2$). In practice, the *experimental variogram* is computed from the available data GH_i as

$$\tilde{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [GH(x_i + h) - GH(x_i)]^2 \qquad (3.4.7)$$

with N(h) being the number of data pairs belonging into the lag (station's interdistance) h within a prescribed tolerance Δh. This experimental variogram, thus, appears as a series of triplets each comprising:

- the number of contributing pairs within h±Δh,
- the average distance between the pairs,
- the average value of the squared differences,

which can be represented as demonstrated in Fig. 3.4.1 for a specific day from the Swiss ANETZ record.

In order to satisfy condition (3.4.5), a continuous, analytical expression for $\gamma(h)$ must be fitted to the experimental points. This procedure is also depicted in Fig. 3.4.1. The number of possible expressions is limited by several conditions of mathematical consistency without which one would possibly end with negative estimation variances. Allowed expressions are standard and ready for use in current kriging software packages.

The example of Fig. 3.4.1 has been chosen so that several properties of typical variograms are clearly displayed. In general $\gamma(h)$ has the following behavior:

1) It starts at 0 when GH(x+h) -> GH(x) for h -> 0, or it is offset from zero by an amount termed the "nugget effect" in kriging terminology (which comes from prospection) and which may originate, e.g., from measurement inaccuracies, insufficient information about the behavior of the field at small scale etc.;

2) It increases with increasing h;

3) It either stabilizes at large h or it doesn't. When it stabilizes, then the upper bound (the "sill") roughly corresponds to the variance of the whole population of sample points. When it remains unbounded, then growth must be less than parabolic. When growth is faster than or equal to parabolic growth, then a drift is present in the field and the hypothesis of weak stationarity is not fulfilled. This latter case will be discussed further, right after the following, brief sketch of the estimation procedure.

4) It's shape may depend upon the direction in which h is taken.

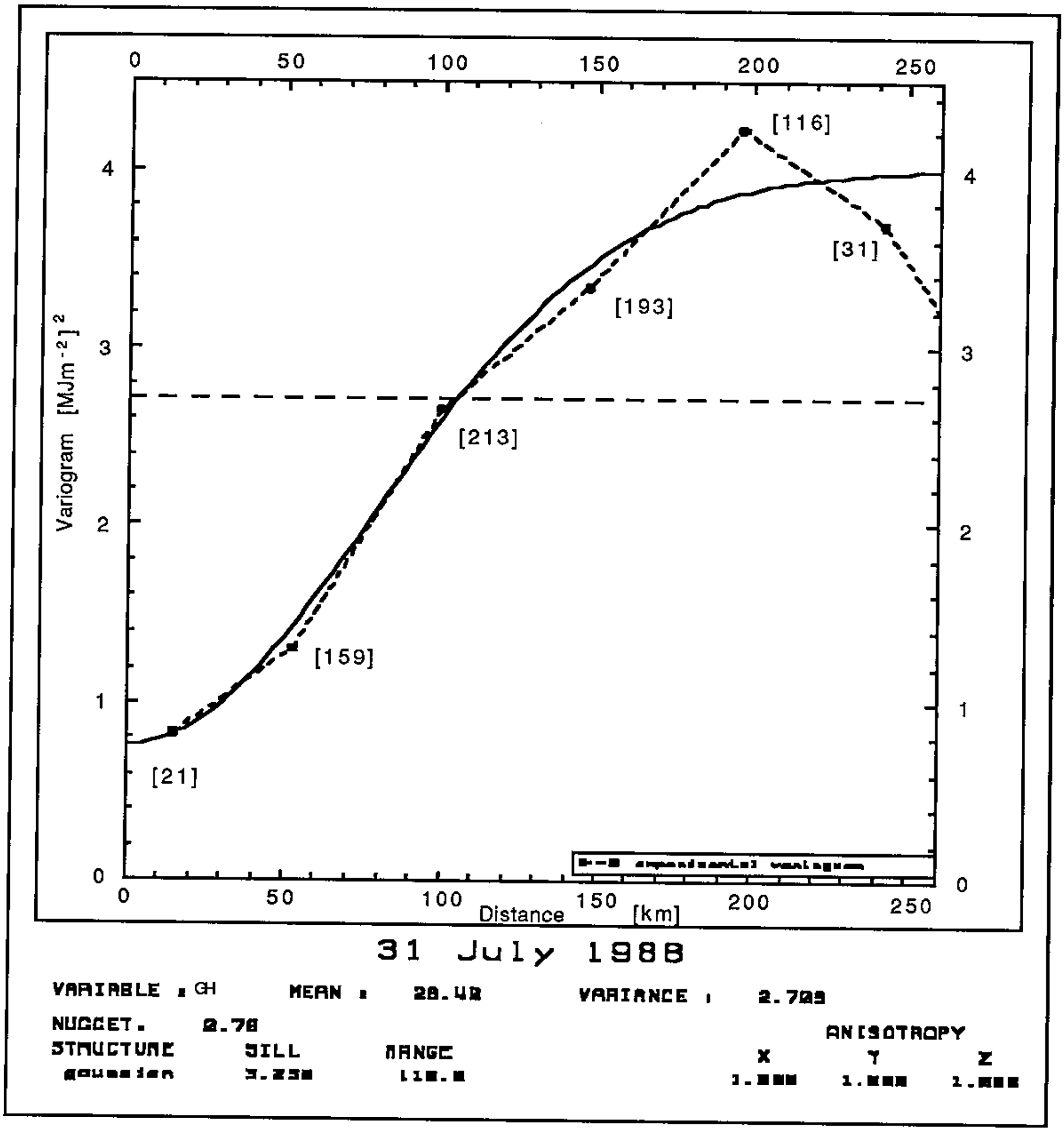


Fig. 3.4.1: Experimental variogram for 31 July 1988 (dotted line). Swiss network with 39 active stations.
The lag comprises 50 km with a tolerance of ± 25 km.
The [..] display the number of pairs contributing to the experimental mean value represented by the dots.
The fitted γ(h) is a *gaussian model* with sill (upper limit) 3.25 [MJm^{-2}]^2, nugget 0.76 [MJm^{-2}]^2 and range 110 km.

3.4.3 *The estimation procedure*

Given the analytical form $\gamma(h)$, the estimation procedure is rather straightforward (at least in the intrinsic case discussed here) because the relevant value γ_{ij} can be calculated for any distance between location i and location j.

Let GH_o be the insolation at the site of interpolation and let GH_o^* be the best linear unbiased estimator of GH_o. Then, GH_o^* as expressed by (3.4.1) ought to fulfil condition (3.4.2), that is

$$E[GH_o^* - GH_o] = \Sigma\lambda_i E[GH_i] - E[GH_o] = 0$$

or, as long as hypothesis (3.4.4) and $E[GH] \neq 0$ hold,

$$\Sigma\lambda_i = 1 \qquad (3.4.8)$$

The estimate also ought to fulfil condition (3.4.3), which states that

$$E[(GH_o^* - GH_o)^2] = \sigma^2$$

must be minimum. Development of this expression while recalling (3.4.6) leads to

$$\sigma^2 = - \sum_i \sum_j \lambda_i \lambda_j \gamma_{ij} + 2\sum_i \lambda_i \gamma_{io} - \gamma_{oo} \qquad (3.4.9)$$

To minimize this variance under the constraint (3.4.8), we introduce a Lagrange multiplier μ_o as

$$\phi = \sigma^2 - 2\mu_o(\Sigma\lambda_i - 1) \qquad (3.4.10)$$

and set the partial derivatives $\delta\phi/\delta\lambda_i$ and $\delta\phi/\delta\mu_o$ equal to zero. This leads to the following system of linear equations for the λ_i and for μ_o (the "kriging system")

$$\begin{aligned} &\sum_{j=1}^{N} \lambda_i \gamma_{ij} + \mu_o = \gamma_{io} \qquad i=1,2,\ldots,N \\ &\sum_{j=1}^{N} \lambda_j = 1 \end{aligned} \qquad (3.4.11)$$

which has a unique solution. Thus, the estimate $GH_o^* = \Sigma\lambda_i GH_i$ assumes a unique value and the variance of this estimate (the "kriging variance") is

$$\sigma^2 = - \gamma_{oo} + \mu_o + \Sigma\lambda_i \gamma_{io} \qquad (3.4.12)$$

3.4.4 *Presence of drifts (non-stationary kriging)*

When the insolation field contains a systematic trend, it cannot be assumed that the mean is constant (hypothesis 3.4.4). If the trend happens to be a slowly varying function, then $m(\underline{x}) = E[GH(\underline{x})]$ may well be approximated by a polynomial of the form (3.2.1). Such an approach, known as **"universal kriging"** accepts that the insolation field may be split into 2 components: one structured and the other random. For example, a linear drift would be

$$m(x,y) = a_o + a_1 x + a_1 y$$

and a quadratic drift would be

$$m(x,y) = a_0 + a_1x + a_2y + a_3x^2 + a_4y^2 + a_5xy$$

The coefficients a_0 to a_k are of course unknown but their determination can be avoided if additional constraints are introduced into the kriging system (see Section 9.3.4.2 in the Theory Volume). Then, the λ_i are again solutions of a linear system which now contains as many Lagrange multipliers as there are cofficients in the polynomial.

Universal kriging is similar to **trend surface analysis** in as much as the field is split into 2 subfields with structures of different scales. But, as in trend surface analysis, the mathematical operations used for the splitting seldom have anything to do with physical reality. This is best illustrated in the practical application of universal kriging which must necessarily be of the iterative type:

1) The "raw" data lead to an experimental variogram that indicates the presence of a drift;
2) A model is set up for the drift $m(\underline{x})$;
3) $m(\underline{x})$ is substracted from the data and the residuals are tested for weak stationarity;
4) $m(\underline{x})$ is adapted until the variogram of the residual converges to what may be looked at as the "true" variogram underlying the field.

This approach may have a chance of success if the range of fluctuation of the residuals is smaller than the domain over which $m(\underline{x})$ is modelled. But if the scales are comparable, then the approach becomes problematic.

Complex drifts are more efficiently treated by the **intrinsic random functions of order k** (short, IRF-k), the theory of which may be found in [3.4.1] and which are implemented in the geostatistical software package *BLUEPACK-3D* [3.4.2] that has been used in parts of the present work. Based on the hypothesis of stationarity of increments of order k, the IRF approach does not require the field to be split into drift and (hopefully stationary) residuals. However, as it relies on local structure recognition in a moving neighborhood, its application is restricted to networks with sufficient station density. The structural and estimation tool (besides the structure recognition) is now termed generalized covariance K(h) and it takes the form

$$K(h) = C + A_1|h| + A_2|h|^2\ln(|h|) + A_3|h|^3 \qquad (3.4.13)$$

Here C accounts for the nugget effect, A_1 is the coefficient of the linear term, A_2 the coefficient of the spline term and A_3 that of the cubic term. Depending upon k only certain combinations of these terms are authorized. As the IRF-k procedure in some way "filters out" the drifts, the remaining structure is assumed to be isotropic. Finally, $k = 0$ corresponds to the intrinsic hypothesis discussed at some length in the previous subsections.

3.4.5 *Application of the method and results*

The foregoing, heuristic presentation of the models underlying kriging is rather long when compared to the presentation of, e.g., moving averages. Correspondingly, kriging is also complex to implement at the programming level. Yet, the tool is so widely used (from geology, to geophysics, to environmental sciences etc.) that numerous software packages exist at different levels of sophistication and cost. Several of the widespread commercial packages are briefly reviewed at the end of Chapter 10 in the Theory Volume, but not just with respect to kriging.

For familiarization with the kriging technique and to ease experience exchanges, the experts group selected the PC-based GEO-EAS (Geostatistical Environmental Assessment Software, Eglund and Sparks [3.4.3]) which is distributed free of charge by the US Environmental Protection Agency. GEO-EAS supports only ordinary kriging, that is only the model based on the intrinsic hypothesis discussed above. Drifts must be treated with the universal kriging approach and this is the reason why the latter has been sketched at some length in the last paragraph.

Because structural analysis cannot be automated, and because of limited manpower, a systematic cross-validation of all available data sets was not possible. Also, because each network must be treated day by day, we cannot present a complete validation for all networks as in the previous cases. In fact, there are only two validations.

Table 3.4.1

Comparison of cross-validation results for three interpolation methods in the US Northeast network.

NORTHEAST US NETWORK			
41 day sample from 4/87 to 6/87			
Network supplementing technique			
operational error (MJ/sq.m/day)			
MOVING AVERAGE			
mean	abe	mean	true
mbe		rmse	rmse
0.07	**0.77**	**2.67**	**2.94**
EXTRAPOLATION			
mean	abe	mean	true
mbe		rmse	rmse
-0.08	**1.05**	**2.90**	**3.28**
KRIGING			
mean	abe	mean	true
mbe		rmse	rmse
0.08	**0.46**	**2.54**	**2.91**

True rmse is the RMSE used hitherto; Mean rmse is the mean of the daily RMSEs;
Mean mbe is the mean of the daily MBEs (equivalent to the MBE used hitherto);
abe is the mean of the daily |MBE|'s.
The sampling period should read 5/87 to 7/87.

The first one, performed with GEO-EAS, is for the US Northeast network and it encompasses a sample of 41 days taken between May and July 1987. This period covers several insolation situations pertaining to the network and the measured mean is 19.3 MJm^{-2}/day. Validations of extrapolation and of moving averages are performed for the same sample to allow for comparisons. The overall results are assembled in Table 3.4.1

Day by day validation results are compared in Figs. 3.4.2 to 3.4.5. Figure 3.4.2 displays absolute bias errors of kriging as a function of absolute bias errors of moving averages. Figure 3.4.3 displays the RMSEs in the same way. The latter comparison is repeated in Fig. 3.4.4, and Fig. 3.4.5, but now for high and low to intermediate insolation conditions separately.

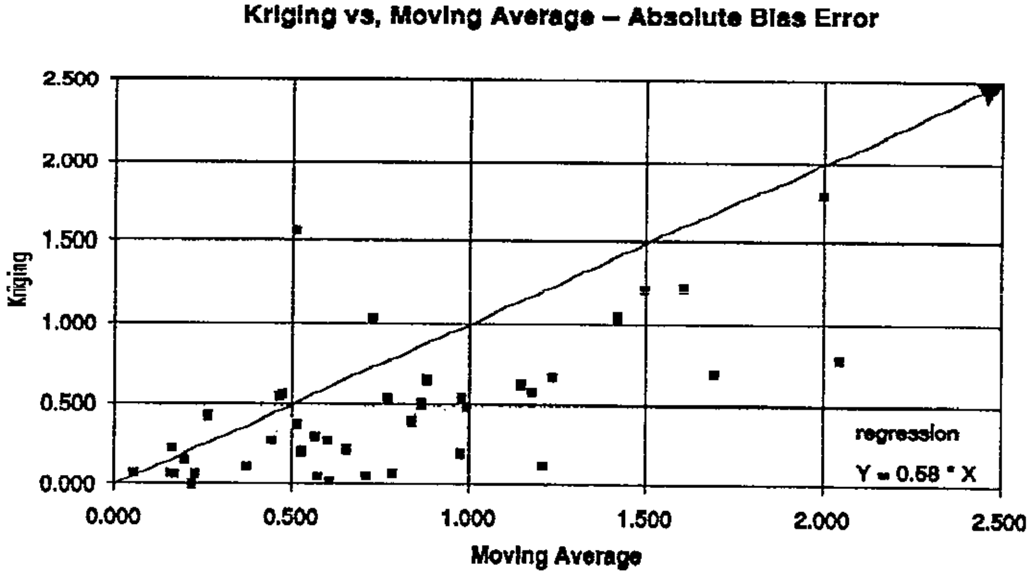


Fig. 3.4.2: Day by day comparison of absolute bias errors: Kriging as a function of moving averages. US Northeast network May - July 1987.

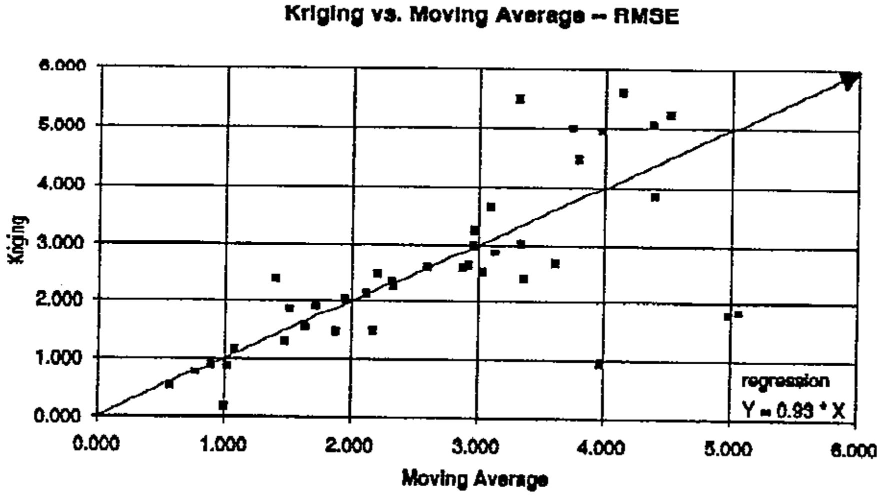


Fig. 3.4.3: Day by day comparison of cross-validation RMSE: kriging as a function of moving averages. US Northeast May-July 1987.

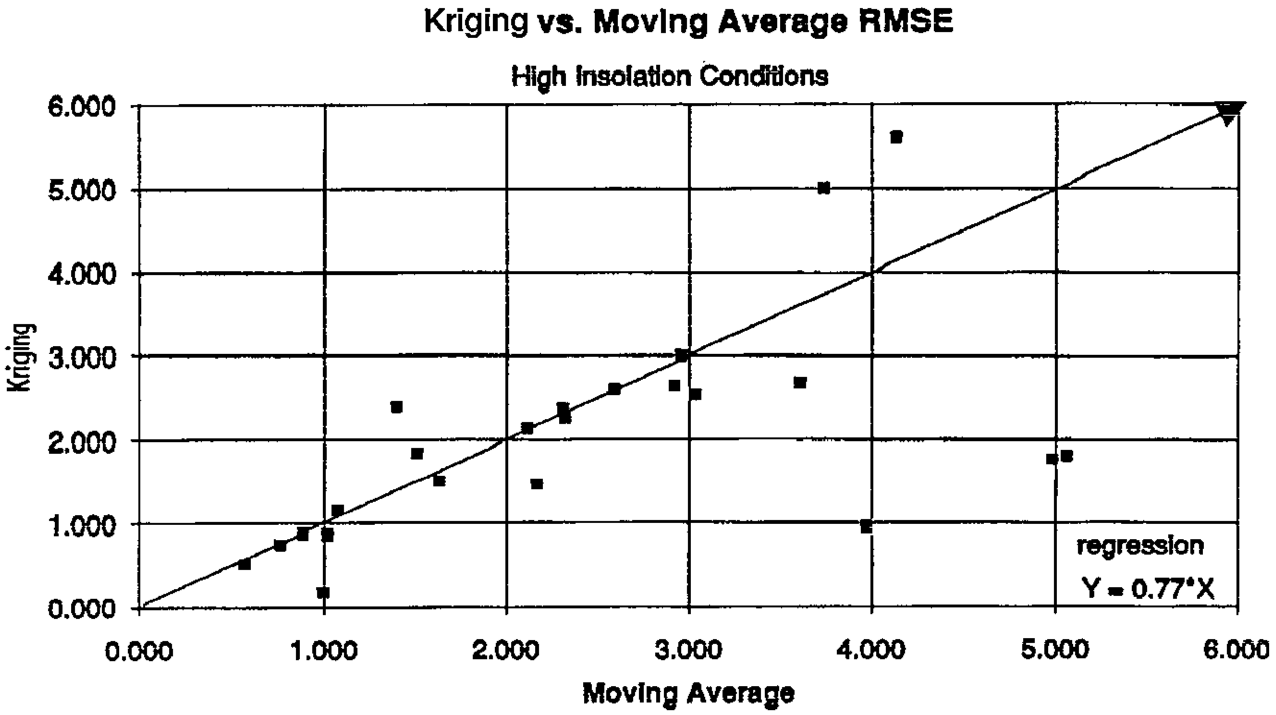


Fig. 3.4.4: Day by day comparison of cross-validation RMSE: Kriging as a function of moving averages for high insolation conditions. US Northeast network, May-July 1987.

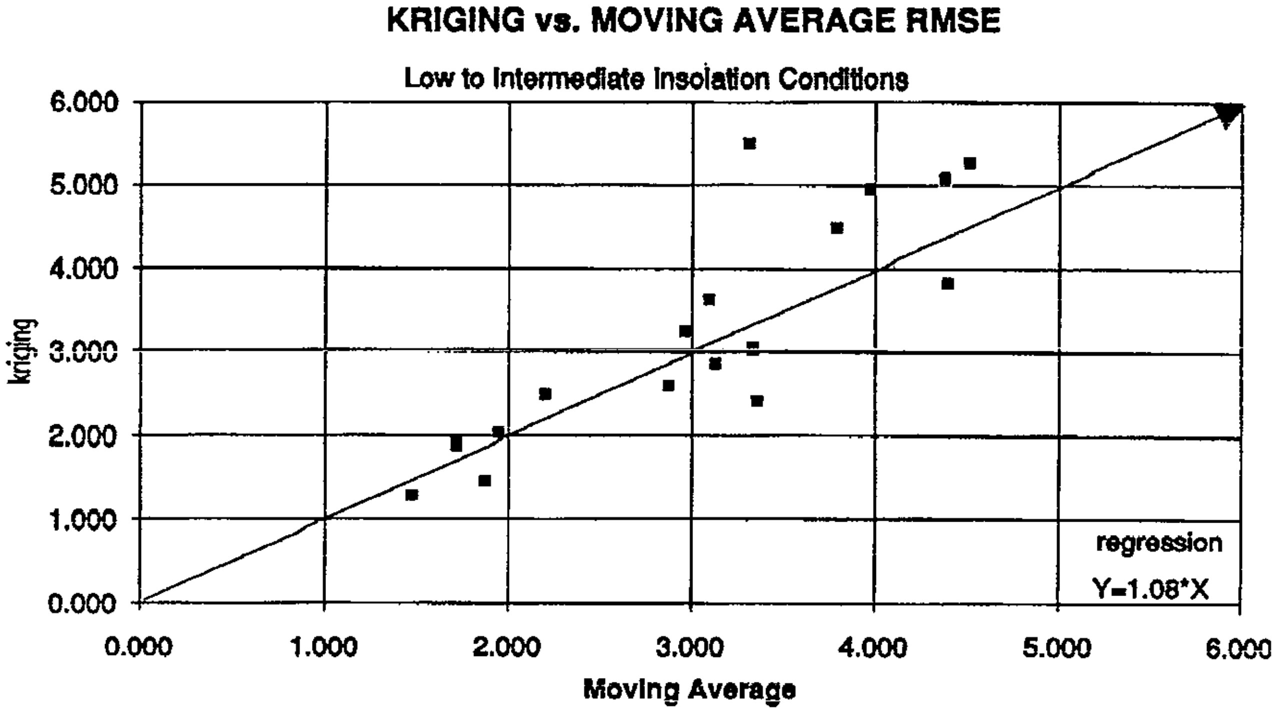


Fig. 3.4.5: Day by day comparison of cross-validation RMSE: Kriging as a function of moving averages for low to intermediate insolation conditions. U.S. Northeast network, May-July 1987.

The second validation is for the Swiss network ANETZ and it encompasses the whole data record, i.e., the year 1988. Here, kriging is performed with the commercial *BLUEPACK-3D* software of Géovariances [3.4.2] accessible to us on a DEC/VAX mainframe at the computing centre of the Swiss Federal Institute of Technology in Zurich.

Depending upon the month, several of the 41 stations retained for subtask 9D must be discarded because of temporary, severe horizon obstruction (see Section 2.2.3). The effectively available number of sites for each month is displayed in Table 3.4.2. The lowest number of 25 active sites in December is still large enough to allow application of the automatic structure recognition feature ("clique" "reco") in the segment "b3drm" of *BLUEPACK-3D*. Therefore, IRF-k kriging can be systematically applied through the whole year. Division into two subsets (sites below either 1000 m or 2000 m elevation and sites above these limits) is performed whenever the situation requires it (fog, strong turbidity gradients, clouded summits, etc.). Structural analysis and kriging is then performaed on each subset separately, with the restriction that below a minimum number of sites in a given subset, the degree k of the possible IRFs must also be restricted . The number of days in each month where this division had to (and could) be applied is also listed in Table 3.4.2.

Table 3.4.2

Monthly number of active sites (1) and of days with altitude subdivision (2) in the ANETZ network.

	J	F	M	A	M	J	J	A	S	O	N	D
(1)	26	34	38	38	39	38	39	38	39	36	26	25
(2)	10	12	8	15	19	18	15	15	24	26	25	25

Cross-validation results are presented in the previous format. Figures 3.4.6 and 3.4.7 show the maps of the annual MBE and RMSE respectively, while the detailed validation will be found in Table 3.3 of the Results Volume. Additionally, Table 3.4.3 gives a summary of the frequency of occurence of the different components of the generalized covariance K(h) (Eqn. 3.4.13) on which IRF-k kriging is based. The table also displays the degree k of the increments associated to each of the encountered K(h). Recall that an increase of k means an increase of complexity of the underlying drift, and that, depending upon k not all components are authorized. For example, in the order 0, K(h) may only be composed of a nugget and/or a linear term.

Fig. 3.4.6: MBE% Switzerland ANETZ.

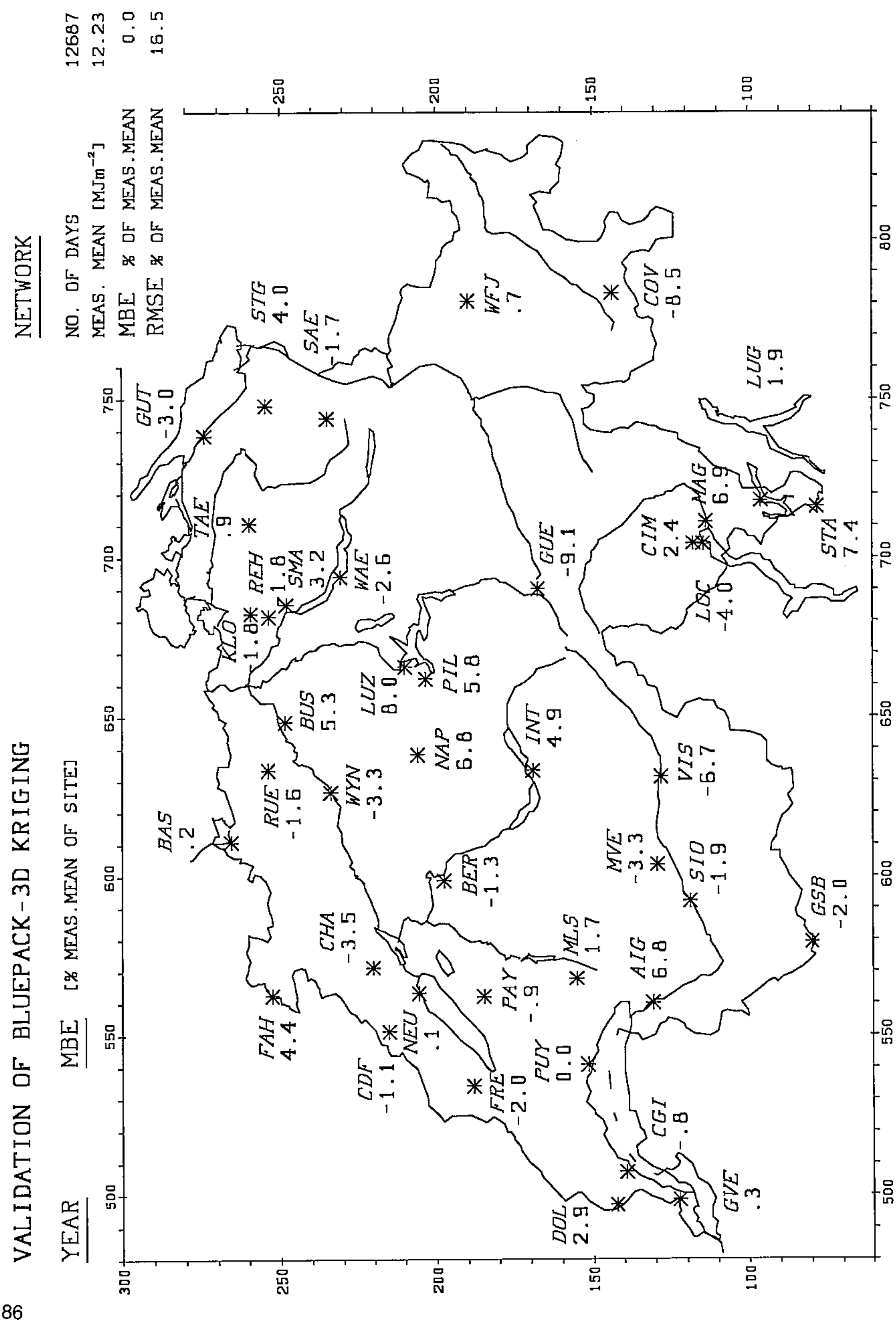


Fig. 3.4.7: RMSE% Switzerland ANETZ.

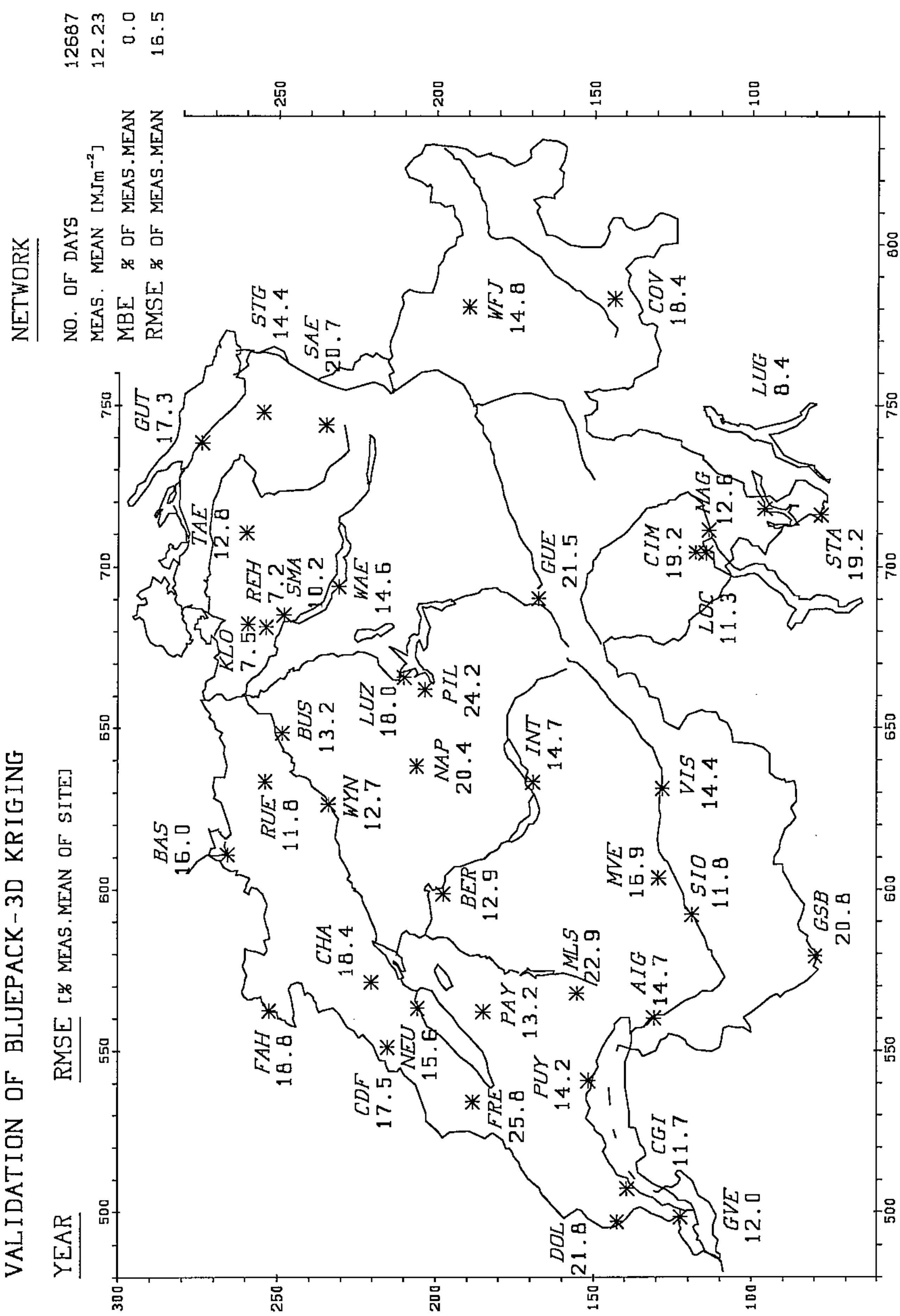


Table 3.4.3

Frequency of occurence of the 4 possible terms (n: nugget, l: linear,s: spline,c: cubic) in the generalized covariance K(h) within 366 days (212 days with 2 covariances), and associated frequency of occurence of the degree k of the increments, for the Swiss ANETZ network

K(h) term				#cases	associated k		
n	l	s	c		0	1	2
x				184	82	97	5
	x			211	106	97	8
		x		32		27	5
			x	19		16	3
x	x			43	25	14	4
x		x		16		15	1
x			x	39		35	4
	x	x		04		04	
	x		x	13		12	1
		x	x	01		01	
x	x	x		-			
x	x		x	01		01	
x		x	x	01		01	
	x	x	x	04		04	
x	x	x	x	-			
Modelled variogram				10	10		

3.4.6 *Discussion of cross-validation results*

In the results presented above, kriging yields slightly (USNEU) to markedly (ANETZ) better cross-validation results than moving averages. The improvement is less pronounced for RMSEs than for MBEs. Obviously, MBEs profit from kriging's ability to reduce bias. For example, in the ANETZ, MBEs were as high as 4% of the network's monthly mean with moving averages, while they are consistently less than 0.5% now. Figure 3.4.2 demonstrates that in the USNEU, the absolute value of the network's daily MBE is almost halved in comparison to moving averages.

In the ANETZ, the maximum reached by one station's monthly RMSE drops from 3.3 MJm^{-2}/day for moving averages down to 2.3 MJm^{-2}/day and the overall RMSE is reduced by 12%. In the USNEU, where the comparison is done day by day, the scatter of the network's RMSEs (Fig. 3.4.3) show little advantage for either interpolation technique. There is a slight advantage for kriging for days with high insolation conditions (Figs. 3.4.4 and 3.4.5), probably because the field structure is better defined under these conditions.

However, comparison of maps 3.3.11 and 3.4.6 for the station's yearly MBEs, as well as maps 3.3.12 and 3.4.7 for the RMSEs, reveals that (in the ANETZ) degradation of performance can occur for individual stations (even if results for the network are improved), especially for high altitude sites.

When the networks only (without the individual stations) are considered, the results could suggest that performance improvement of kriging over the moving averages is not more than marginal. Furthermore, since the required manpower has significantly increased, from a cost/benefit analysis there has been little, if any, improvement. This is perhaps true as long as efficiency is understood only in the narrow sense of "prediction accuracy". However, if efficiency is considered in the wider sense of "prediction plus assignement of confidence limits on the predicted value", then the increased effort required by kriging may well be justified. Indeed, one should recall that the outcome of a kriging procedure is not only the interpolated value GH* but, also, the variance σ^2 of this interpolation (Equ. 3.4.12). The significance of this variance is best illustrated by Fig. 3.4.8 where a kriged map of daily totals in the ANETZ is depicted together with the variance of the estimation σ^2. Obviously, kriging yields physical information that was not present in the methods discussed hitherto and that has several most welcome applications. For example, persistence of a local maximum of σ^2 across several different configurations of the field will indicate the *preferential location for a supplementary station in a network.*

Also, when comparing performances, one should be aware that kriging is applied here under somewhat adverse conditions, while moving averages benefit from a rather high degree of sophistication.

For instance, the U.S. Northeast network (defined for this study) encompasses only 7 sites. This is probably the lowest limit for reliable structural analysis because some important points in the experimental variogram may not represent the structure of the field. Some kriging software packages even issue warning messages when entry is attempted with so few samples.

Furthermore, to speed validation progress, application of *BLUEPACK-3D* to the ANETZ has deliberately been limited to automatic operations whenever possible. Table 3.4.3 shows that variograms were interactively modelled (e.g., Fig. 3.4.1) in only 10 cases, whereas interactive modelling should also have been used for the 213 cases where the intrinsic hypothesis is valid, that is, where the order of increments is zero. Namely because in this situation, interactively modelled variograms present much more flexibility (more parameters) than $K(h)=C+A_1|h|$ which has only two.

On the other hand, for moving averages, no time was spared when evaluating the most adequate inverse distance weighting law, and when adjusting vertical scale factors and search radii. As already noted in the discussion in Section 3.3, this is not really a burden provided the data are available. This was the case here given the very nature of the study, but otherwise, enough time ought be alloted for gathering and conditioning the network data (see Chapter 2!) necessary for "tuning" the method's parameters.

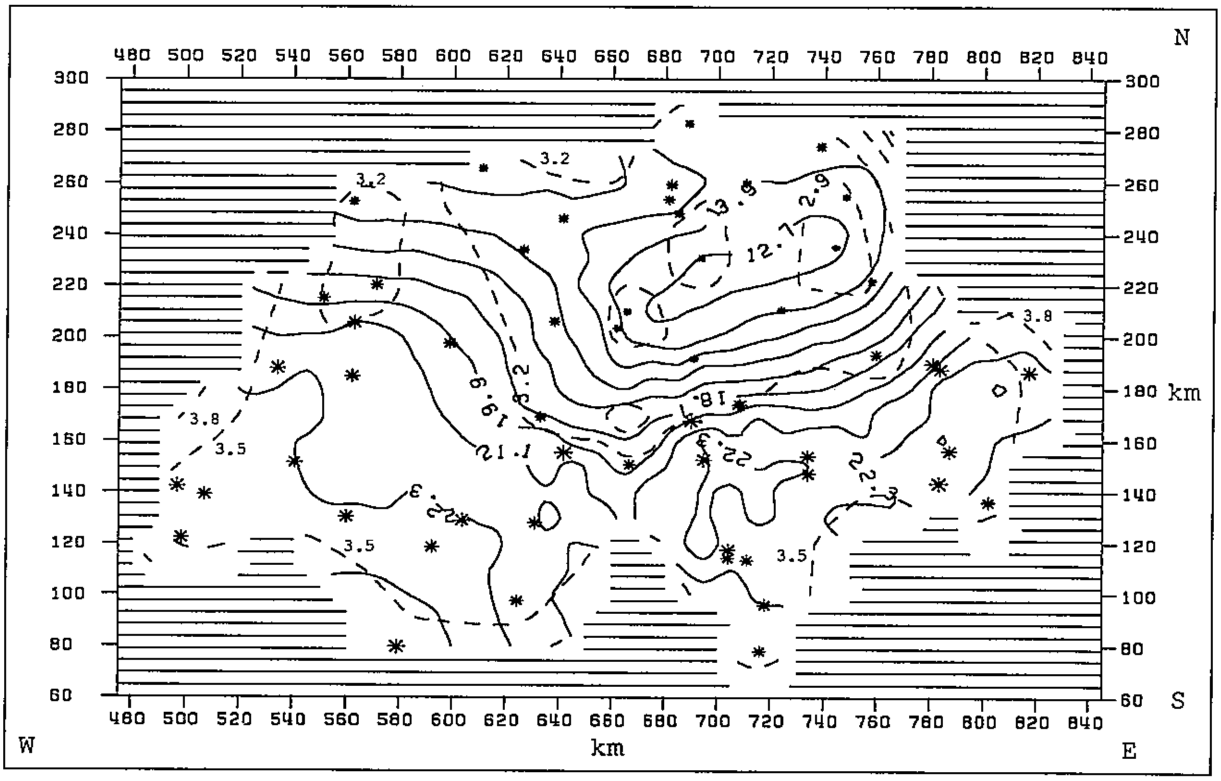


Fig. 3.4.8: Isolines of daily insolation [MJ/m^2] in Switzerland on 10 April 1982. Dashed isolines are for the variance of estimation. The latter increases towards the border of the network (stations = *).

While the altitude effect inherent to the ANETZ are accounted for in the moving averages as well as the method allows it, they are only approximated in kriging as performed here. Division of the sample in sites below and above an elevation level (typically 1000 m) could work well if there were always enough sites present in each category. In fact, it does work well in Summer (see Table 3.4.2) when convective cloudiness attenuates insolation as a function of altitude, but it doesn't in Winter, in the presence of fog, because then there are only a few elevated stations with an unobstructed horizon (less than 10 sites in a sample allow detection of only linear drifts, while k=0 is assumed for less than 6 sites).

At this stage it is worth noting that moving averages applied without altitude modulation (i.e., with vertical scale factor equal 1) yield cross-validation's "bottom line" MBE = -1.2% and RMSE = 20.7%, which are markedly worse than with kriging. However, this comparison is again biased by the presence of an (admittedly rough) altitude component in the kriging application.

As its name suggests, *BLUEPACK-3D* is expected to optimize interpolation of F(x,y,z) functions. However, the z-coordinate seems to be constrained to drillhole geometry found in mineral exploration. We have not investigated if it is really so, because the technique of cokriging with insolation estimates from satellite images (Chapter 4), to be presented in Chapter 5, is preferred to any interpolation of GH(x,y,h) only.

Some of the experience gathered while performimg structural analysis in the ANETZ are worth reporting before concluding this discussion. The synoptic weather pattern was found to properly propagate into the structure of the insolation field. If for example, as a result of a frontal passage, the structure of a given day leads to a nugget and a cubic term in K(h) under k=1, then the next day will present this structure again rather than a completely different one. Also, certain structures are found to be characteristic of specific conditions (e.g., pure nuggets with k=0 correspond to clear days), a finding that confirms previous, more limited investigations by Jaquet [3.4.4] and by Zelenka *et al.* [3.4.5].

3.4.7 *References*

[3.4.1] P. Delfiner and coworkers, 1987. "Applied Geostatistics for the Petroleum Industry". Course Manual 1987 Summer School, Centre de Géostatistique, École des Mines de Paris, Fontainebleau, p.213ff.

[3.4.2] BLUEPACK-3D User's Manual, 1989. Géovariances, 38 Av. F. Roosevelt, F-77210 Avon-Fontainebleau.

[3.4.3] E. Eglund and A. Sparks, 1988. GEO-EAS (Geostatistical Environmental Assessment Software) User's Guide. Environmental Monitoring Systems Laboratory. Office of Research and Development, U.S. Environmental Protection Agency, Las Vegas, Nevada 89139.

[3.4.4] O. Jaquet, 1991. "Étude géostatistique de l'insolation énergétique mesurée par le réseau météorologique Suisse". *Sciences de la Terre,* Série Inf. **30**, pp. 57-70.

[3.4.5] A. Zelenka, D. Lazic, and O. Jaquet, 1989. "Solar resource assessment with geostatistical tools". *Clean and Safe Energy Forever,* Proc. 1989 Biennial Congress ISES, Kobe, **Vol.3** pp. 2077-2081. Pergamon 1990.

3.5 Principal Components Analysis (Empirical Orthogonal Functions)

Study prepared by: *Gerhard Czeplak and Jörg Schneider*
DWD, Meteologisches Observatorium Hamburg

Under sponsorship from: *Bundesministerium für Forschung und Technologie (BMFT)*

This technique has been selected for calculation of monthly average global insolation maps for western Germany. The final product will refer to the period 1976-1985, but here, results are presented for 1988, for the data contributed to subtask 9D (Section 2.2.1). A survey of the composite procedure has been given in Section 1.1.2 and theoretical aspects are developed in appropriate context partly in Chapter 5 but mainly in Chapter 10 (Theory Volume). Although presented here for a temporal series of 12 monthly-mean days, the EOFs could be set up for any sequence of individual days, at least from a technical point of view. The pertinence of such an approach, as well as the availability of the required data are not subject to discussion here.

The data to which the EOF-technique is applied is based on measured sunshine duration data and global radiation at 188 stations of the German Weather Service (DWD) for the year 1988. For stations where only sunshine duration was measured, the sunshine duration data were converted into global radiation by using "Ångström's" relation (see Section 5.2.7).

As the influence of topography on global radiation was to be considered, the total sum of 188 stations was partitioned into classes according to the land-use around the stations. Three land-use classes were made up of stations in rural, urban and forest regions. To prove that there are significant differences between these three land-use classes, they have to pass one or if necessary two significance tests.

3.5.1 Validation of Significance Tests

Methods found suitable for determining whether the differences between these classes are fortuitous or due to the land-use of the areas around the stations are the F-test and T-test. They yield a necessary and sufficient condition for establishing the significance [3.5.1].

First of all, the dependence of the measured data on the station altitude has to be removed by reducing all data to a fixed altitude, chosen so as to minimize the reduction for the maximum number of stations [3.5.2]. Then the monthly means of each class were converted into annual means. The tests were applied alternately on two classes at confidence levels of 95, 99 or 99,9 percent.

Results for forest and urban regions:

$$\text{F-test: } PF=1.24 < F(95\%, f1=61, f2=36) \approx 1.87 \ ,$$

where $PF = (s_1/s_2)^2$, s_1 and s_2 being the standard deviations of data set 1 (forest) and data set 2 (urban) respectively, with $s_1 > s_2$ mandatory, where further $f_1 = n_1-1$, $f_2 = n_2-1$, with n_1, n_2 being the number of stations in data set 1 and 2, respectively. The test quantity PF being lower than the corresponding value in the F(95)-table yields no significant differences. This means that both data sets can be merged into one and thus, a subsequent T-test is allowed. The sufficient condition of the T-test result was

$$\mathrm{Tau} = 0.66 < T\ (95\%, f=97) \approx 1.98$$

where $f = f_1 + f_2$,

$$\mathrm{Tau} = \left| \frac{m_{x1} - m_{x2}}{s_d} \right| \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

and

$$s_d = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

with m_{x1} and m_{x2} being the means of the values in data sets 1 and 2, respectively.

The test quantity Tau being less than the corresponding value in the T-table proves that there are *no* significant differences between these classes.

Results for forests and open land:

$$\text{F-test: } PF=1.02 < F(95\%,\ f_1=61,\ f_2=88) \approx 3.36$$

This result proves that there is no significant difference between the standard deviations s_1 and s_2. A subsequent T-test is therefore allowed.

$$\text{T-test: } TAU = 3.41 > T(99{,}9\%\ f=149) \approx 3.36.$$

In this case, the difference between the means m_{x1} and m_{x2} are highly significant. Therefore, the two classes should not be combined into a single class.

Results for urban areas and open land:

$$\text{F-test: } PF = 1.22 < F(95\%,\ f_1=36,\ f_2=88) \approx 1.82$$
$$\text{T-test: } Tau=2.27 > T(95\%,\ f=124) \approx 1.98$$

Thus, these classes probably differ and they can not be merged into one.

Final result:

It is reasonable to partition the initial data according to the land-use classes of the areas around the stations.

3.5.2 Interpretation of the Eigenvectors

The mapping of global radiation was realized for each month by a principal components analysis according to Granger's empirical orthogonal functions (EOF) method (see Chapter 10.3.3 in the Theory Volume).

The eigenvectors and multipliers of each of the land-use classes were calculated separately. Then the land-use at each grid-point of the digital terrain map was evaluated by determining the dominant land-use at the surrounding 24 grid points which corresponds to an area of about 12km x 12km. The global radiation at each grid point was calculated from the eigenvectors and multipliers corresponding to the land-use assigned to the grid point. It appeared that the first 4 eigenvectors explained 99% of the variance of the radiation field.

As shown in Fig. 3.5.1 to 3.5.3, the seasonal variations of the first 2 eigenvectors of each land-use class are very similar.

Eigenvector 1 represents the seasonal variation of solar insolation and the lengths of the days. Using the monthly means of just one year incorporates the special weather of that year and makes the curve less smooth than with long term averages.

Eigenvector 2 is related to the cloud climatology and the patterns of synoptic scale systems during the rainy periods and thus represents the variation of the weather conditions for the year 1988.

Eigenvector 3 represents the convection processes of the atmosphere. It can be clearly seen that the pattern of eigenvector 3 for urban stations is opposite to the pattern for stations in open land. This is due to the atmospheric convection over ground of different absorption conditions. The pattern of eigenvector 3 for forest stations differs primarily from the pattern of stations in rural areas in the months of March and August. That may easily be accounted for by the low albedo of forests in wet seasons.

The interpretation of eigenvector 4 is uncertain. It might be related to the patterns of fog and stratus. But as the loading of eigenvector 4 on the total variance is relatively unimportant (see Section 10.3.3.4), the lack of an interpretation is not significant.

3.5.3 Validation of the EOF Method

To judge the quality of the EOF method, a cross-validation was performed for the 29 stations of the DWD radiometric network. Global radiation calculated from the measured sunshine duration at the majority of the 188 sites by means of Ångström's formula were used as input for the EOFs, but were not considered as measurements for the cross-validation.

The cross validation method alternately leaves out one of the N stations of a single land-use class and calculates the value of the global radiation at that station from the N-1 remaining stations of that respective land-use class, according to the EOF method. Then for each month of the year, the root mean square error (RMSE) of the 29 global radiation stations was calculated. For interpolations required by the cross-validation procedure, as well as in the calculation of the global radiation according to the EOF method, the gravity interpolation by Zelenka and Lazic [3.5.3] was used.

Figure 3.5.4 displays the seasonal variation of the RMSE as the percentage of the measured global radiation. It can be seen that in months with short days and low solar elevations, the RMSE is largest but becomes better in summer. Figure 3.5.5 shows the monthly means of daily sums of the global radiation in Wh/m^2 with respect to the root-mean-square-error (RMSE) for the same 29 stations. The RMSE is largest for the largest values of global radiation.

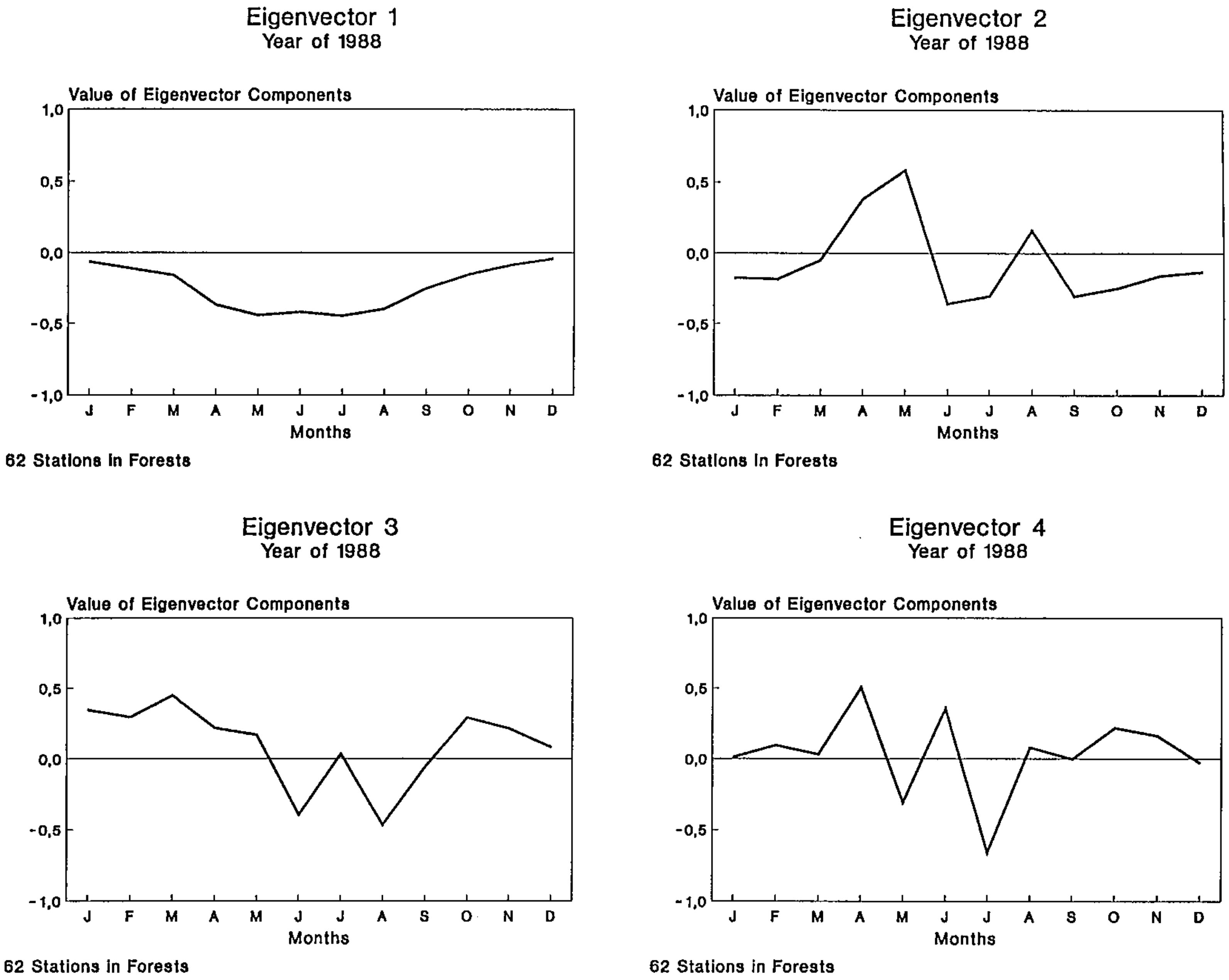

Eigenvector 1
Year of 1988
Value of Eigenvector Components
1,0
0,5
0,0
-0,5
-1,0
J F M A M J J A S O N D
Months
62 Stations in Forests
Eigenvector 2
Year of 1988
Value of Eigenvector Components
1,0
0,5
0,0
-0,5
-1,0
J F M A M J J A S O N D
Months
62 Stations in Forests
Eigenvector 3
Year of 1988
Value of Eigenvector Components
1,0
0,5
0,0
-0,5
-1,0
J F M A M J J A S O N D
Months
62 Stations in Forests
Eigenvector 4
Year of 1988
Value of Eigenvector Components
1,0
0,5
0,0
-0,5
-1,0
J F M A M J J A S O N D
Months
62 Stations in Forests

Fig. 3.5.1

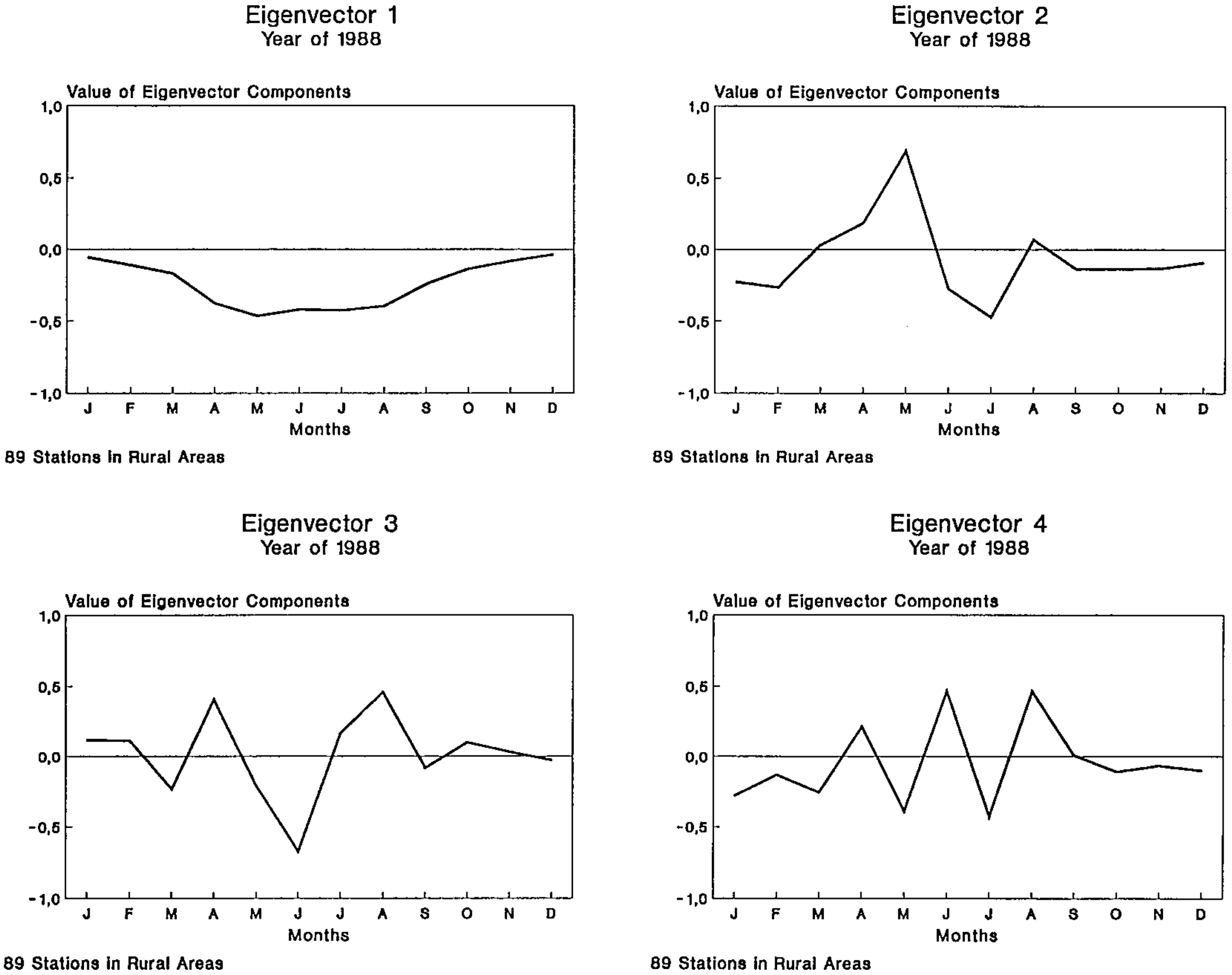
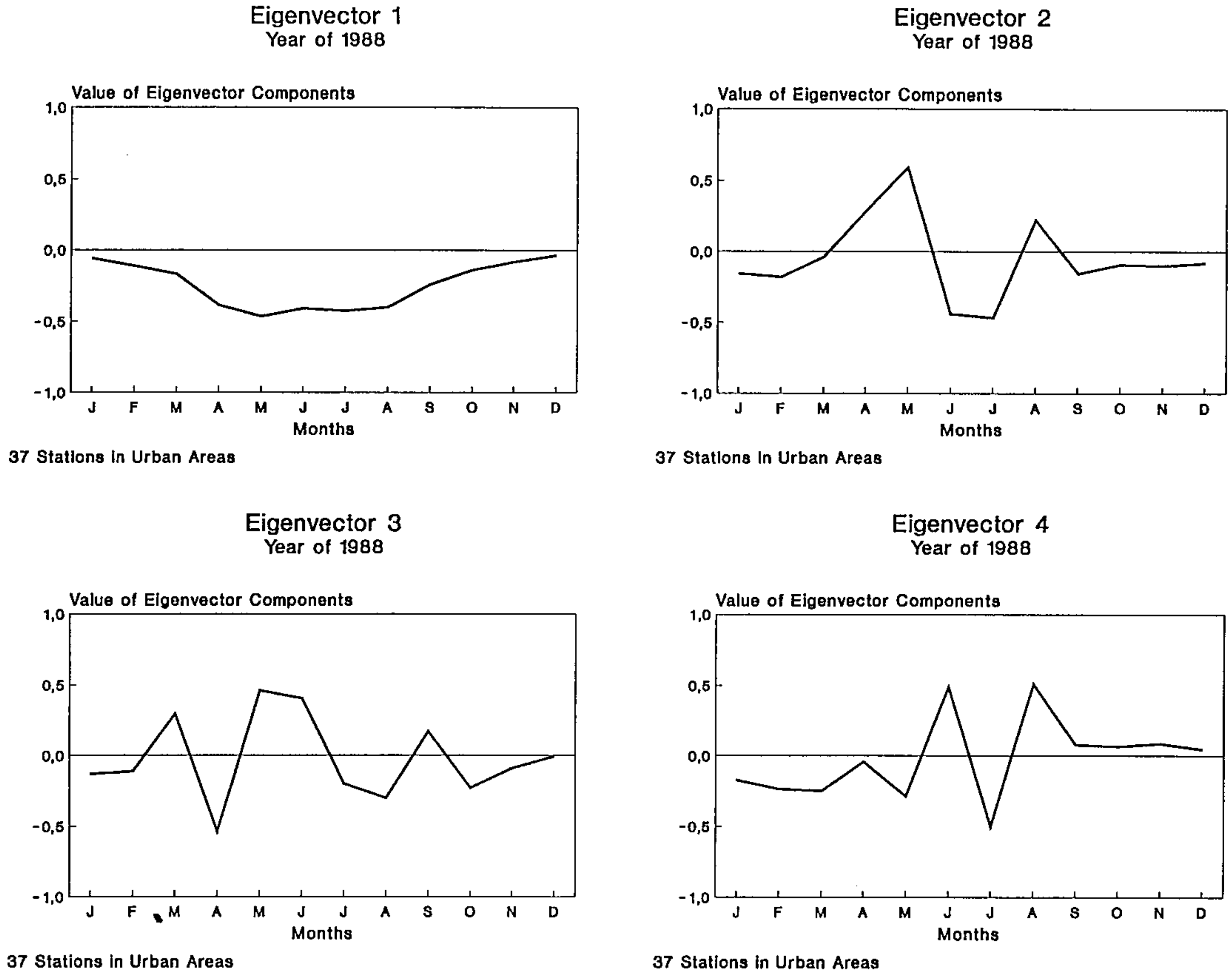

Eigenvector 1
Year of 1988
Value of Eigenvector Components
1,0
0,5
0,0
-0,5
-1,0
J F M A M J J A S O N D
Months
89 Stations in Rural Areas
Eigenvector 2
Year of 1988
Value of Eigenvector Components
1,0
0,5
0,0
-0,5
-1,0
J F M A M J J A S O N D
Months
89 Stations in Rural Areas
Eigenvector 3
Year of 1988
Value of Eigenvector Components
1,0
0,5
0,0
-0,5
-1,0
J F M A M J J A S O N D
Months
89 Stations in Rural Areas
Eigenvector 4
Year of 1988
Value of Eigenvector Components
1,0
0,5
0,0
-0,5
-1,0
J F M A M J J A S O N D
Months
89 Stations in Rural Areas

Fig. 3.5.2



Eigenvector 1
Year of 1988

Eigenvector 2
Year of 1988

Eigenvector 3
Year of 1988

Eigenvector 4
Year of 1988

Fig. 3.5.3

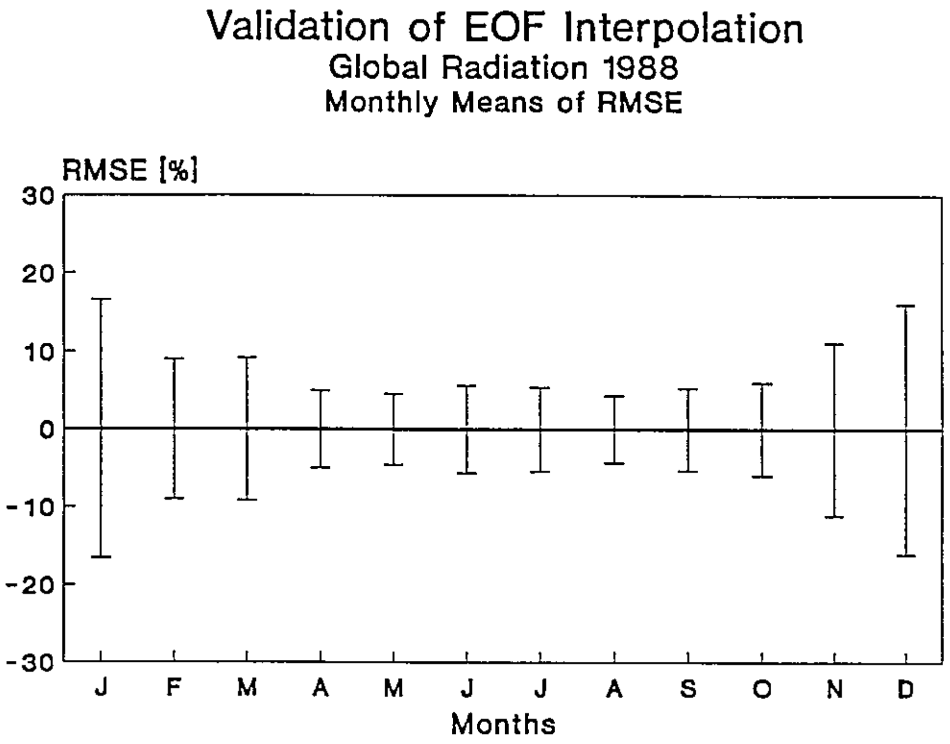


Fig. 3.5.4

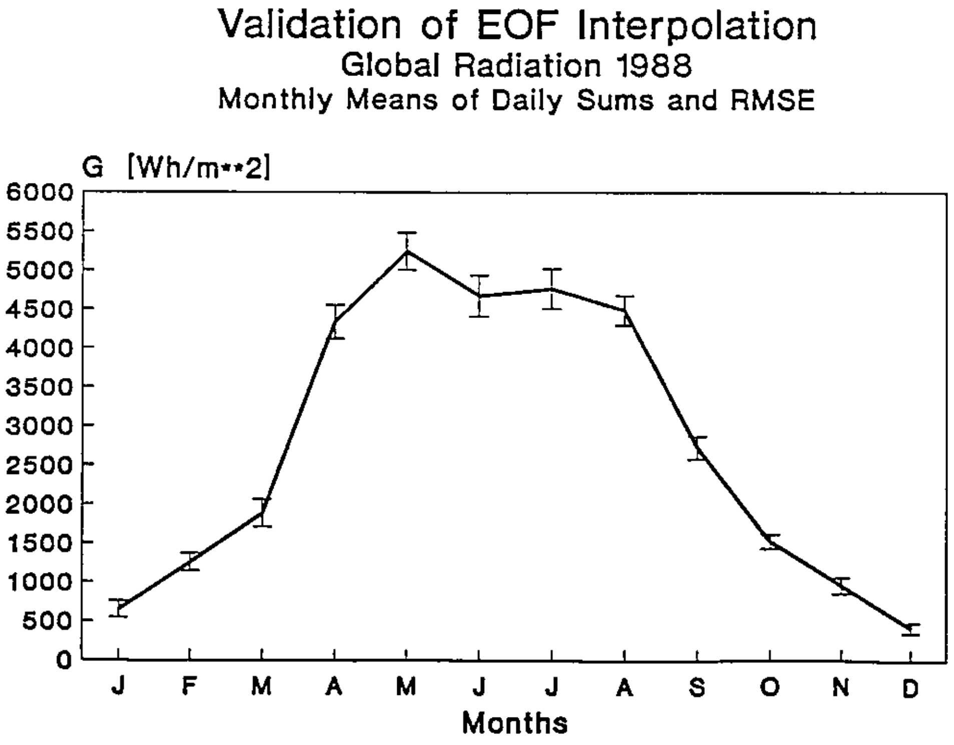


Fig. 3.5.5

3.5.4 Mapping and comparison with other methods

Maps of global radiation for western Germany as calculated with Granger's EOF method (see Section 10.3.3.6), in which the the multipliers are determined by gravity interpolation (search radius 400 km, vertical scale factor 100), is first discussed for the two typical months June and December.

In the month of June (Fig. 3.5.6), a high rate of global insolation can be found on the islands in the North Sea and Baltic Sea and on the leeside of the mountainous regions in the south of Germany (east of the Black Forest and the Suelic Alb, and north of the Alps). The lowest rate of global insolation can be found in the uplands between 50 and 52 degrees latitude north, over the Black Forest and also within and north of the Alps. This is due to cloud accumulation within and in front of the mountains. *It is obvious that the detailed structure is based on the land use-classes of the topographical data set.*

The meridional distribution dominates the month of December, that is to say, the global radiation increases continously from north to south according to the increasing solar elevation. Orographical effects are insignificant (Fig. 3.5.7). At this stage, the comparison with June maps obtained with satellite-based methods is very favorable.

Figure 3.5.8 shows the coarse outcome of gravity interpolation when the sites which contribute to the moving average are restricted to the 29 stations of the radiometric network. Despite use of a vertical scale factor of 100, no topographical details are resolved due to the small number of adequate stations.

Figure 3.5.9 documents the outcome of the same method when all of the 188 available sites (i.e., the radiometric network and the sunshine duration network for which global radiation was deduced from the Ångström relation) inside the search radius of 400 km are allowed to contribute. Several features appear which are typical for the convective cloudiness typical of June in mountainous regions.

Figure 3.5.10 displays the result of kriging with GEO-EAS [3.5.4]. The underlying variogram is shown in Fig. 3.5.11. A slight elliptical anisotropy between E-W and N-S was detected during the structural analysis but, as no plausible physical cause could be found for it, isotropy was eventually assumed. The wealth of details contained in the map is rewarding as, opposite to gravity interpolation (topography) and to EOFs (land-use/topography), no external constraint is implied here.

Finally, Fig. 3.5.12 reproduces the METEOSAT4-based map as obtained by the Möser and Raschke method, which will be discussed in Chapters 4 and 11. Some ideas about how to combine this information with the former one are given in Chapter 5.

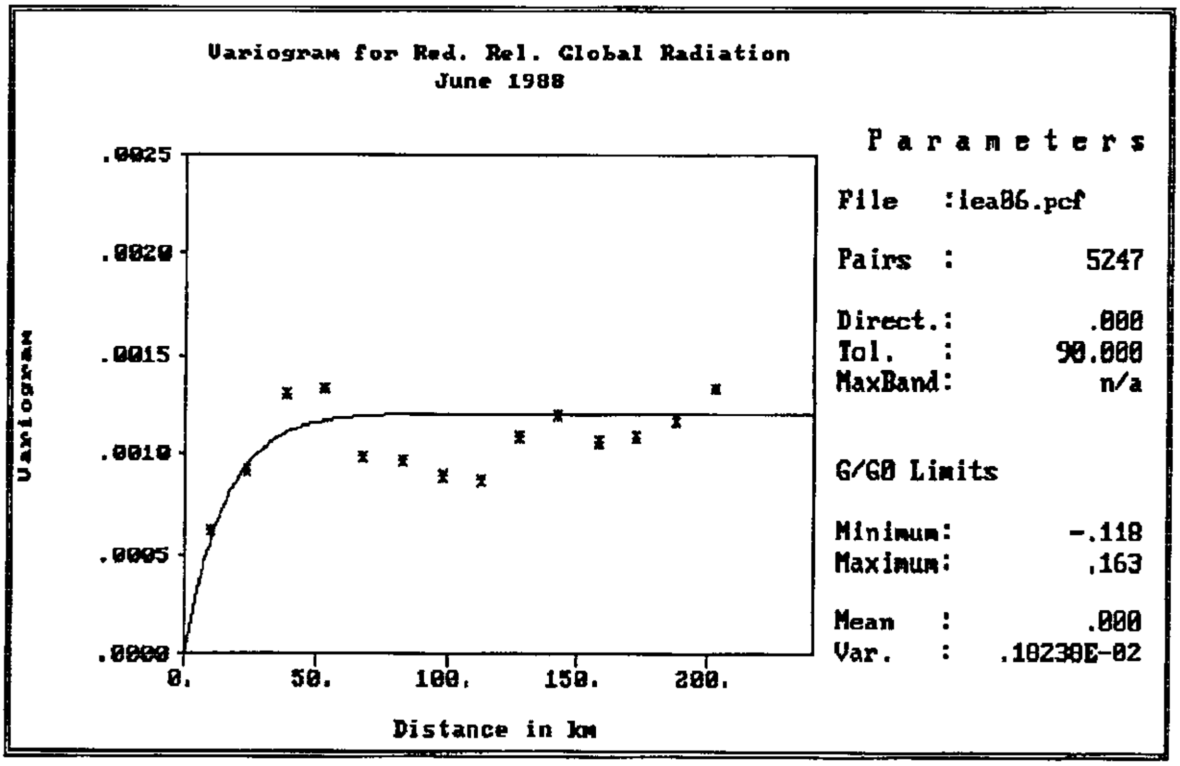
3.5.5 *References*

[3.5.1] Kaiser, R.E. und Möhlbauer, J.A., 1983. "Elementare Tests zur Beurteilung von Messdaten". 2. Aufl., Mannheim 1983, pp. 24-30, 56 ff., 60-64.

[3.5.2] Gerth, W.-P., 1986. "Klimatische Wechselwirkungen in der Raumplanung bei Nutzungsänderungen". Ber. d. DWD **Nr. 171**, Offenbach a. M. 1986, p. 18.

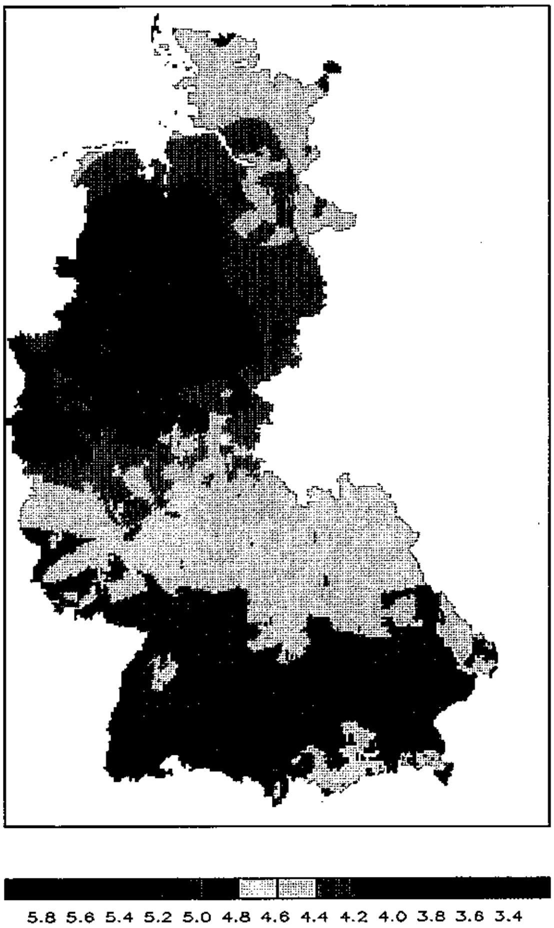
[3.5.3] A. Zelenka and D. Lazic: 1987. "Supplementing network global irradiance data". *Advances in Solar Energy Technology*, Proc. 1987 Biennial Congres ISES, Hamburg. **Vol. 4**, pp. 3861-3865, Pergamon 1988.

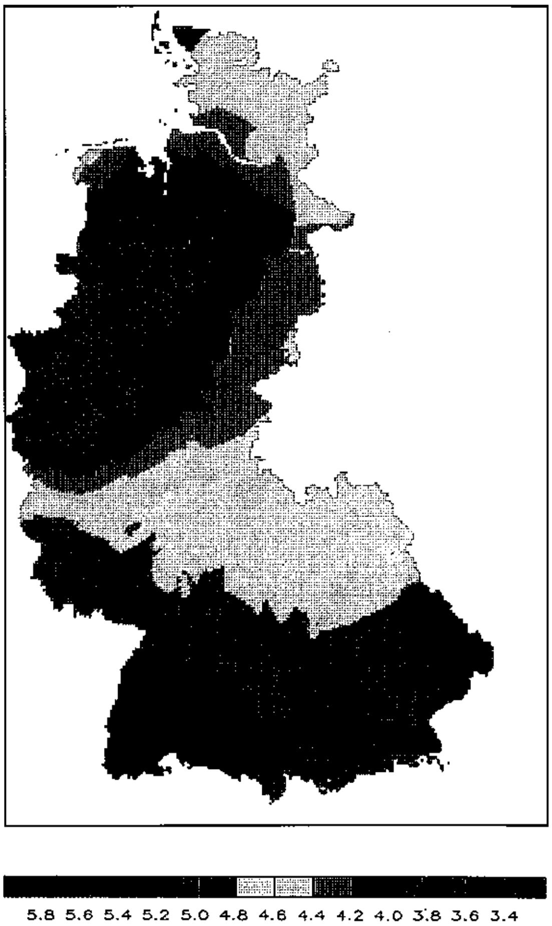
[3.5.4] E. Eglund and A. Sparks, 1988. GEO-EAS (Geostatistical Environmental Assessment Software) User's Guide. Environmental Monitoring Systems Laboratory. Office of Research and Development, U.S. Environmental Protection Agency, Las Vegas, Nevada 89139.



```
Parameters

File    :iea06.pcf

Pairs   :          5247

Direct.:           .000
Tol.    :        90.000
MaxBand:            n/a

G/G0 Limits

Minimum:          -.118
Maximum:           .163

Mean    :          .000
Var.    :   .10230E-02
```

Fig. 3.5.8: Experimental variogram (*) and spherical model (full line) with June 1988 average daily relative global insolation from 188 sites in western Germany. Parameters of the model are: sill = 0.0012 and range 45 km; nugget =0. The exponential model supports the kriged map in Fig. 3.5.7.





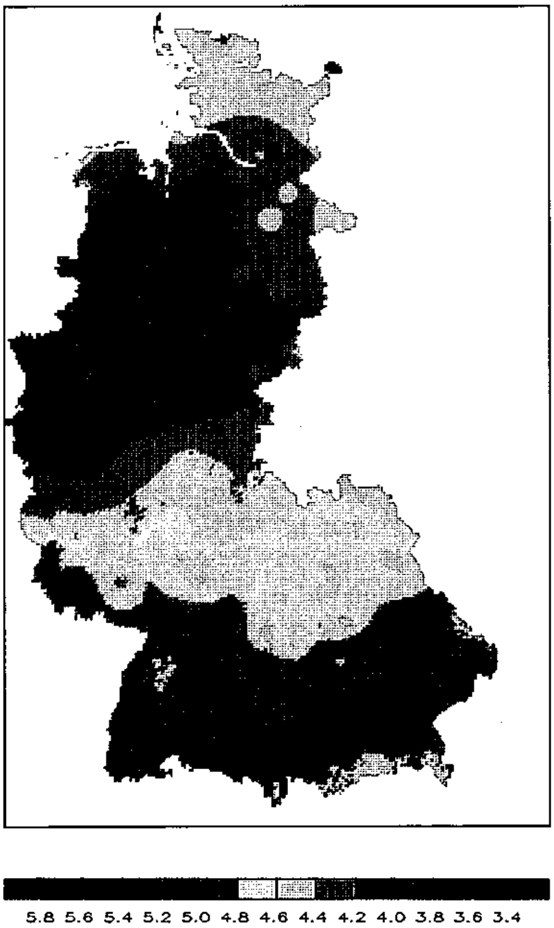


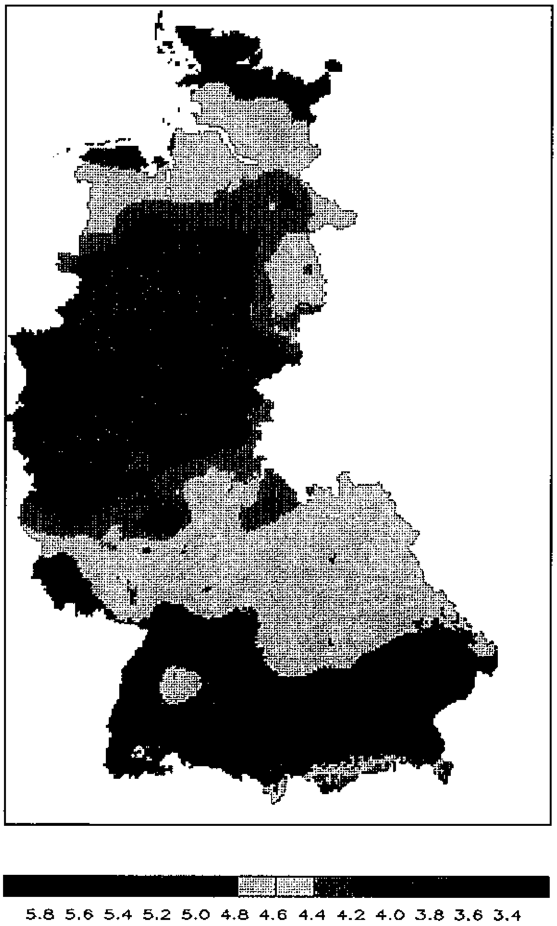
Fig. 3.5.6

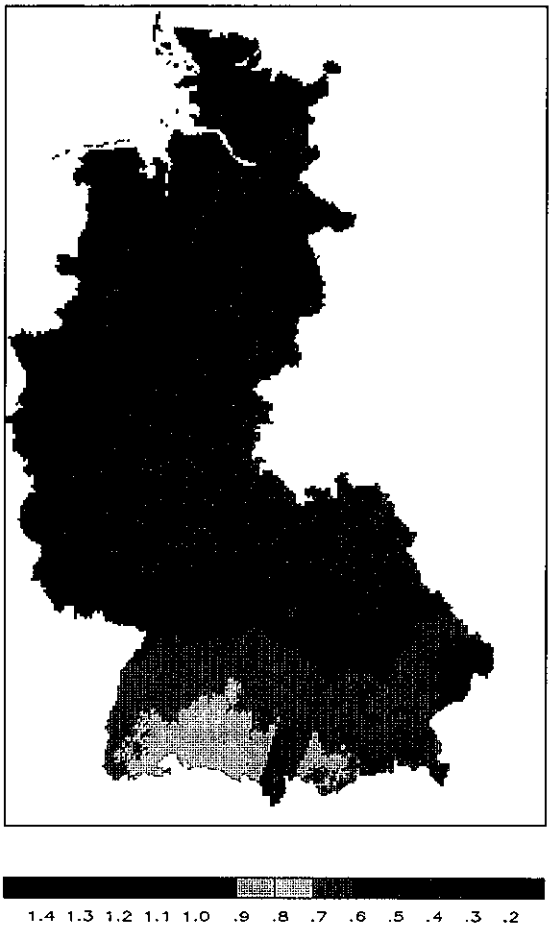
Maps of monthly averaged daily insolation totals for June 1988 in $kWhm^{-2}$:

A (upper left) : EOF method;

B (upper right) : Gravity interpolation with 29 stations;

C (lower left) : Gravity interpolation with 188 stations.





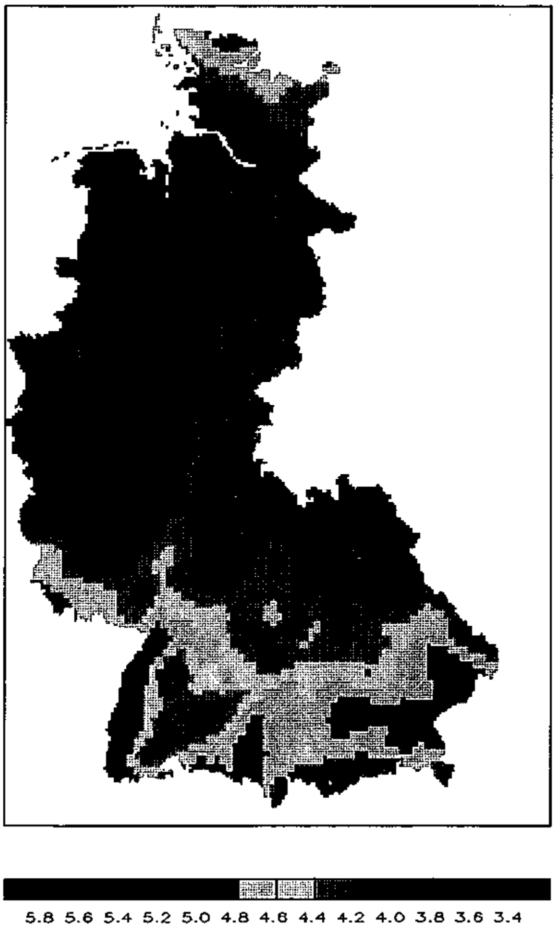


Fig. 3.5.7

Maps of monthly averaged daily insolation totals in $kWhm^{-2}$:

A (upper left) : Kriging for June 1988; see also Fig. 3.5.8 for the variogram;

B (upper right) : Kriging for Dec. 1988;

C (lower left) : METEOSAT-based estimates with the DWD/Cologne method for June 1988.

CHAPTER 4

4. Validation of satellite-based methods

Study prepared by: *Antoine Zelenka*
Swiss Meteorological Institute

Under sponsorship from: *Swiss Federal Office of Energy*

Recognition that radiance measurements by spacecraft-borne radiometers can serve to estimate global irradiance incident on the Earth's surface goes back to the mid-sixties. Developments since then are described in several of the references listed in Chapter 11 (Theory Volume).

Most important is the assessment of the ability of the results to reproduce mesoscale variability of global irradiance as it is known from specially dedicated, ground-based measurements [4.1.1, 4.1.2].

Among the two types of Earth-observing spacecraft, i.e., polar orbiters and geostationary, meteorological platforms, only the latter provide adequate time coverage (at 30 minute intervals) to yield the type of estimates necessary to most solar energy applications. Sun-synchronous, polar orbiters operate much closer to the ground, thus allowing a substantially better spatial resolution of the scanned regions. However, a given platform can deliver only one daylight scan of a given location. Therefore, even a system of several platforms will hardly provide an adequate coverage of daily variations [4.1.3]. Consequently, operational methods, i.e., routinely performed estimates, all rely on geostationary satellites and in the following the generic term "satellite" will always imply the latter type of spacecraft.

The scope of this chapter is definitely the presentation of the validation results. Yet, their discussion is not possible without some references to basic knowledge of the involved methodologies. Readers who are not familiar with the subject will find this knowledge at the necessary depth in Section 4.1, which may also serve as an introduction to Chapter 11.

4.1 Principles of satellite-based methods

4.1.1 Fundamental considerations

Satellite radiometers record that fraction of radiation incident from the Sun which is scattered back to space by the Earth and its atmosphere. This scattered fraction is called the *planetary albedo* and it should not be confused with the more widely known ground albedo. The planetary albedo increases with increasing atmospheric turbidity and cloudiness (enhanced backscattering), while, at the same time, the flux reaching the ground decreases. This fundamental complementarity between the satellite observation and the global irradiance to be estimated from it, is sketched and also semi-quantitatively depicted in Fig. 4.1. There, global irradiance is expressed with the clearness index K_T (i.e., as a fraction of the irradiance available at the top of the atmosphere), while turbidity is parametrized with the Linke factor and cloudiness is expressed in octa.

Readers who are exclusively interested in the performance of satellite-based, operational estimates regardless of the estimation method, can now skip the next sub-sections and proceed directly to section 4.2.

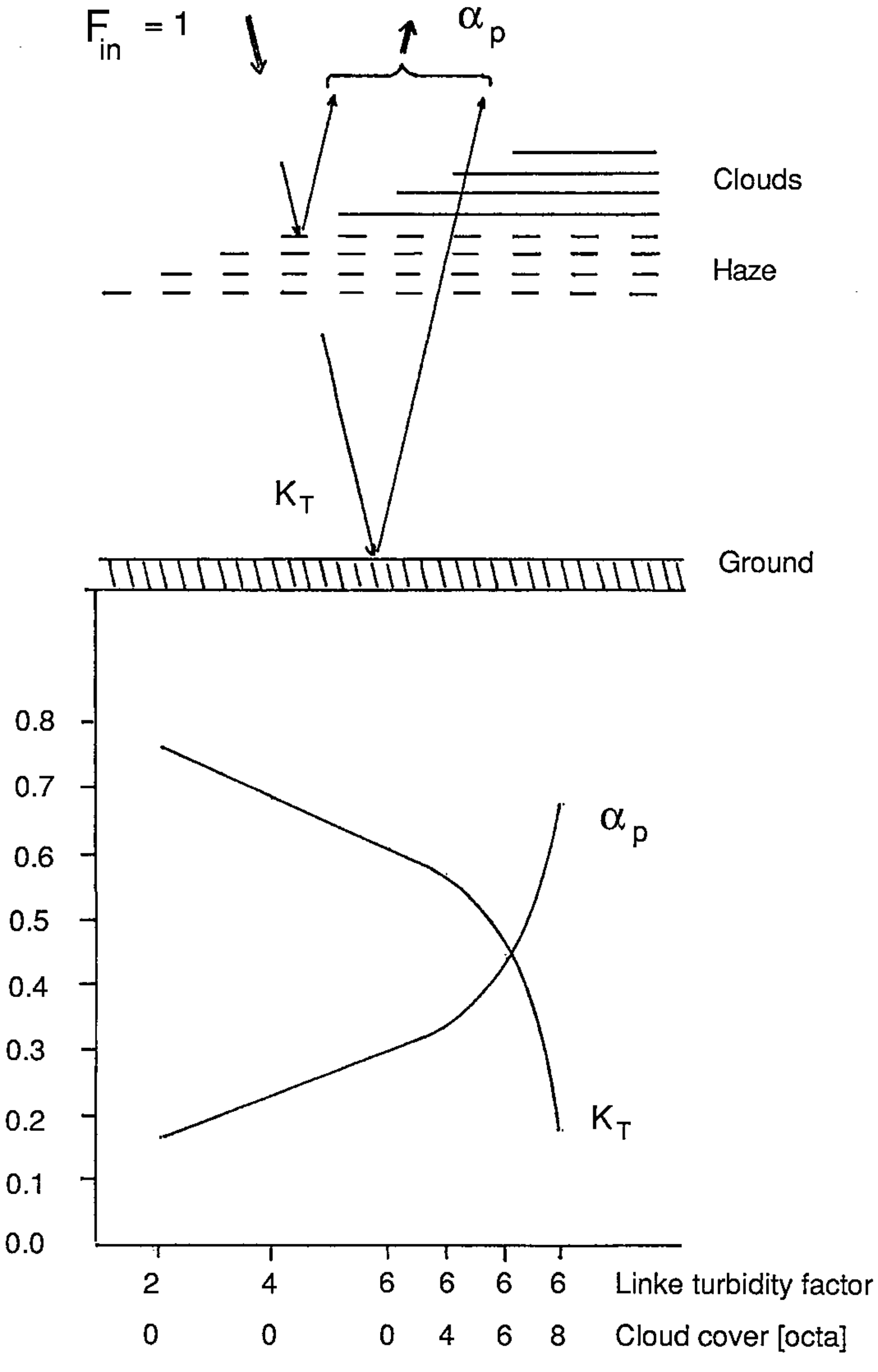


Figure 4.1 Upper part: sketch of the relation between planetary albedo α_p (satellite observation) and global irradiance (resp. clearness index K_T) at ground level.
Lower part: An example of the quantitative relation $K_T(\alpha_p)$. The scale of the horizontal axis is arbitrary, with increasing turbidity (Linke factor) and cloud cover (in octa) to the right. The solar flux incident on the top of the atmosphere is set equal to one, so that the vertical axis can express the involved fluxes in fractions of the incident flux.

4.1.2 Fundamental equations

The relation between planetary albedo and clearness index is readily expressed in terms of shortwave radiant flux conservation, because, to a very good approximation, the related infrared fluxes can be balanced separately. The conservation equation states that the difference between the net flux at the top of the atmosphere and the net flux at ground level equals the flux absorbed between these two boundaries, that is

$$F^o_{in} - F^o_{out} = F^g_{in} - F^g_{out} + F_a \tag{4.1}$$

with

F^o_{in} = Solar flux incident on the top of the atmosphere,
F^o_{out} = Outgoing flux at the top of the atmosphere,
F^g_{in} = Global horizontal irradiance at ground level,
F^g_{out} = Outgoing flux at ground,
F_a = Sum of fluxes absorbed by the atmosphere.

Division of Equ. (4.1) by F^o_{in}, re-arrangement of terms and introduction of the standard definitions

$F^o_{out} / F^o_{in} = \alpha_p$ = planetary albedo,
$F^g_{out} / F^g_{in} = \alpha_g$ = ground albedo,
$F^g_{in} / F^o_{in} = K_T$ = clearness index,
$F_a / F^o_{in} = \alpha_a$ = absorbed fraction,

lead to the relation

$$K_T = (1 - \alpha_a - \alpha_p)/(1 - \alpha_g) \tag{4.2}$$

Theoretically, introduction of climatological (seasonal) values for the absorbed fraction α_a (typically 0.19) and for the ground albedo (typically 0.17), would lead to a unique $K_T(\alpha_p)$. Practically, however, many peculiarities of the satellite imaging system propagate into the relation and the major ones deserve at least qualitative consideration, as they will be needed for the discussion of results in section 4.2.

4.1.3 Practical constraints

The various radiometers on board the satellites have different apertures (angular resolution) and they operate in several spectral ranges with different spectral response functions. The spectral ranges are termed "channels". The infrared (IR) channel has been occasionally used, together with the visible (VIS) channel, mainly to differentiate cloud types. But the yield has been too modest, at least in the operational mode, to justify the effort to process both images. Therefore, all of the routine methods rely exclusively on the VIS channel. Depending upon the satellite, the latter ranges from 550nm to 750nm (GOES, GMS), and from 400nm to 1100nm (METEOSAT).

Measurements (α_p) in this spectral window, with a definitely non-uniform response function should, according to Equ. (4.2), be related to measurements (K_T) by pyranometers with a black (uniformly responding) sensor in the the range from 300-3300nm. From this, we see that the purely theoretical form (4.2) must be generalized into

$$K_T = a - b\,\alpha_p \tag{4.3}$$

where the, *empirical*, coefficients a and b are, among others, functions of the spectral specifications of the VIS channel.

The radiometers scan the targets with a finite aperture in order to yield a "picture", or "image" of them. Each digital picture element ("pixel") consists of a count (typically 0-63), which is a function of the target's area *radiance* (fig. 4.2). However, Equ. (4.3) contains irradiances (radiant flux densities) - not radiances - and the only operational link between both radiance and irradiance is the assumption of isotropy for all involved scattering processes. Only in this case is the radiance proportional to the irradiance (irradiance = π x radiance) and the VIS counts simply related to α_p. Obviously, the isotropy assumption must also hold for the outgoing irradiance at ground level. Departures from these assumptions will lead to changes in the coefficients of (4.3), and if the latter are kept constant at average values, to erroneous predictions.

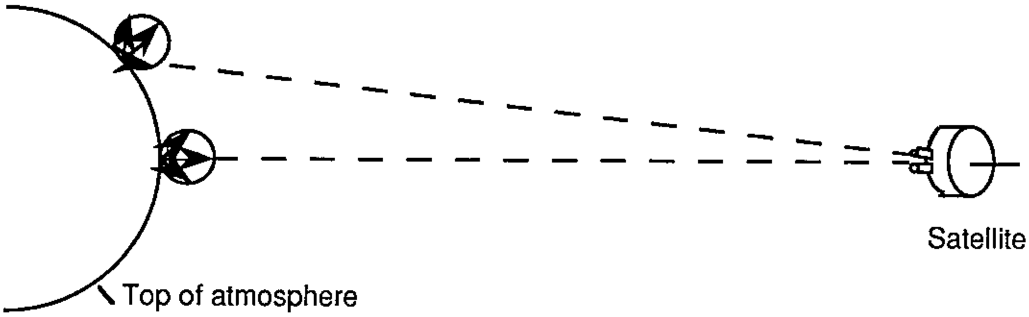


Figure 4.2 Sketch of the difference between the outgoing irradiance at the top of the atmosphere and the radiance recorded by the satellite. The outgoing flux is a weighted sum of all radiances as (represented by the circles) while individual radiances are represented by the arrows

Isotropy is usually satisfied if the scattering angles (Sun-target-satellite) are not too large. This constraint, in turn, narrows the satellite field of view for which estimates are possible.

Pixel sizes at the sub-satellite point, in the VIS channel, range from 1 km x 1 km (latitude-longitude) for GOES to 2.5 km x 2.5 km for METEOSAT and they increase with increasing angular distance from this point. Irradiance estimates are always meant to be average values over the whole pixel. In fact, the averaging is usually extended over several neighboring pixels. First at the stage of the "raw" data, to avoid erroneous values due to measurement or transmission errors. Then during the conversion procedures, in order to ensure that instantaneously recorded values are properly transformed into hourly integrals. This is implicitly done by arguing that several adjacent pixels display those conditions which are experienced by one sole pixel during the hour around the instant at which the image is recorded.

4.1.4 Classification and transportability of methods

A. *Statistical approaches*

The "statistical", empirical approach exploits the fundamental relation in its free adaptable form (4.3). Simultaneous satellite brightness and surface irradiance observations are required over an extended period of time to allow determination of the coefficients a and b by regression analysis. The relative simplicity of operation (if one neglects the calibration phase with ground observations, which are not available everywhere) is offset by limited regional validity [e.g., 4.1.4].

B. *Physical approaches*

More or less sophisticated radiative transfer calculations, with various model atmospheres and ground albedos, are performed to predict the relation (4.2) between K_T and α_p as a function of parameters which are unique to the satellite image. Among these parameters, cloud cover related indices obviously play the dominant role.

The most elaborate radiation transfer calculations are performed spectrally, thus taking into account the radiometer's spectral sensitivity, provided the latter is known. Several methods skip this crucial step, at the cost of requiring ground-truth measurements, by which the factors first left open can be determined.

C. *Relation between method and satellite*

From all the foregoing, it should be apparent that each method must be calibrated for a specific satellite's radiometer, as well as its scanning scheme, before it can yield practical results. Furthermore, this calibration is a time consuming undertaking. So it shouldn't be surprising that these methods are usually tied to the satellite for which they have been developed and that their performance can only be assessed with ground data collected within the satellites useful field of view.

Simultaneous performance assessments of different methods for a given satellite with the same ground data have been performed by Raphael and Hay [4.1.5], Grüter *et al.* [4.1.6] and, more recently, by Whitlock *et al.* [4.1.7].

Testing the transportability of select methods from one satellite to another is a current challenge [4.1.8], but clearly outside of the scope of the present study, which only deals with results obtained by routine operations. However, we can also report on a preliminary investigation, which originated as a spinoff from this work.

Noia [4.1.9] and, subsequently, Czeplak, Noia and Ratto [4.1.10] have attempted a one-to-one transfer of Tarpley's method to METEOSAT data for western Germany. The empirical coefficients duly calibrated for a given GOES radiometer and for the United States were applied entirely unchanged to counts of the METEOSAT2 radiometer for European scenes. Validation with 26 stations of the DWD network for 19 days (in November 1986) yielded a RMSE of 22% and a MBE of -5% of the measured average daily total. These values are slightly larger than those to be discussed in the next sections, but they are definitely in the same range. Station-wise 19-day averages are within 10% of the averages obtained from the METEOSAT/DWD method. This indicates that satellite-based methods for estimating global insolation at the Earth's surface are rather "forgiving", but it also implies a "natural" lower limit for the precision of these methods. Clearly, further investigations on a broader scope are needed to confirm the preliminary findings reported here.

Also, Pesin [4.1.11] recently used METEOSAT2 data to compare daily estimates using Cano's statistical method (see Section 11.4.4 in Volume 3: Theory) with those obtained with the Gautier physical method (see Section 11.5.1). Validation was for 31 stations of the French radiometric network. Unfortunately, the comparison remains somewhat uncertain because the performance of Cano's model was assessed only for July 1986 while Gautier's model was tested for January and July 1986. Even more recently, the physical method of Dedieu *et al.* (see Section 11.5.4), originally derived for the METEOSAT imagery, has been very successfully tuned to the GOES imagery [4.1.12].

We are fortunate to have one operational example of both the physical and statistical methods. This confers a certain generality to our performance assessments even without method transfer studies.

4.2 METEOSAT / DWD

4.2.1 *Application field*

The EUMETSAT spacecrafts of the METEOSAT generation describe a geostationary orbit which keeps them at the vertical of the origin of the geographical coordinates (zero degrees latitude and longitude). The field of view useful for surface irradiation estimations encompasses West and East Europe (except Nothern Scandinavia), the Middle East, the whole African continent, and the most Eastern part of South America [4.2.1].

The area treated by DWD (see Section 2.3.1) until 1987 comprised all of Germany (the formerly FRG and GDR). In the South it was limited by the northern border of the Alps and in the West it covered parts of the Benelux and of Eastern France. From April 1987, to accomodate IEA Subtask 9D, the area was extended about 120 km to the South and to the West. Extension to the South made ground-truthing with the Swiss ANETZ possible. The current area consists of 112 (E-W) x 144 (N-S) pixels with a mean pixel size of 7.5 km x 7.5 km. The area boundaries, in geographical coordinates are:

	2.15° E	15.91° E	
55.26° N	x	x	55.40° N
45..58° N	x	x	45.68° N
	3.99° E	14.51° E	

The number of available days for ground-truthing were restricted to 148 (see Section 2.3.1) for the period from March until October. Monthly validation tables, such as those presented in Chapter 3, are of limited value here. In this case, it is preferable to present validations for all days together, but with details which may be of interest for the discussion. Thus, Tables 4.1 and 4.2 in Volume 3 (Results) display, for each station of the considered network, not only the MBE and RMSE values, but also standard deviations for measurements and estimates, minima and maxima of the time series, and coefficients for the correlation equation

pyranometer measurements = $a_0 + a_1$ x satellite-based estimates.

4.2.2 DWD network

Table 4.1 reveals a behavior of the estimates which is perfectly consistent with the results of the longer validation period (1985-1987) discussed in [4.2.2]. There, the modelled daily sums for 16 stations and 22 months lead to a MBE of 5.3%, and a RMSE of 17.9%, while we obtain -4.1%, and 16.8% with 29 stations and about 6 months. Ref. [4.2.2] reports a seasonal dependence of the validation results with degradation in Winter. The slightly lower RMSE reported here may be explained by the fact that our sample is concentrated on the Summer months.

While the mean biases are comparable in absolute value, the change in sign is conspicuous. It can be speculated that sensitivity changes in the visible channel between METEOSAT2 (until 1987) and METEOSAT3 (1988) may be responsible for this behavior.

4.2.3 *Swiss ANETZ*

Results of the overall validation (Table 4.2) are comparable in magnitude with those of the DWD network: a slightly better MBE of -0.8% is accompanied by a degradation of the RMSE to 20.5%. This result was to be anticipated as the terrain within large regions of the network does not fulfill several conditions implicitely required by the model.

In section 4.1.3, isotropic backscattering, of the atmosphere and the ground within the area subtending the pixel, was required for optimum performance of the satellite models. At a scale of several km^2, a flat rural region is certainly a better isotropic diffusor than a rugged mountainous area. Furthermore, as already stated in section 2.3.1, the estimated insolation is not for one isolated pixel, but, rather, is a weighted mean of several pixels surrounding the site. The weighting scheme of the Cologne/DWD model is depicted in Fig. 4.3. Considering their pixel size, it is easy to understand why the estimates deteriorate at stations located between mountains and lakes.

In this respect, subdividing all sites into 3 categories, namely "normal", "elevated", and "lake shore and/or valley" leads to a very instructive performance assessment. Sixteen stations belong to the "normal" category (PAY, WIN, FAH, MAG, STA, STG, KLO, RUE, BUS, SMA, LUZ, BAS, BER, WAE, TAE, REK). In this subset the MBE amounts to 0.7% and the RMSE to 16.2%, similar to the results for the DWD network. Fourteen stations with altitude above 1000m belong to the category "elevated" (DOL, SAE, MLS, MVE, CHA, PIL, NAP, GUE, GSB, CDF, WFJ, COV, FRE, CIM), where MBE = 2.5%, while the RMSE increases to 25.3%. Finally, the 11 stations located near "lake shores and/or valleys" (LUG, LOC, NEU, INT, GVE, PUY, LUZ, CGI, GUT, AIG, SIO, VIS; only major lakes, where the water surface fills more than a relevant half-pixel are considered) have a MBE of -7.0% and a RMSE of 18.4%. The large underestimation is the result of the "pollution" of the estimate by pixels lying above water and snowy surfaces).

The results for "elevated" stations require further comments. The original estimates by DWD lead to a whole network MBE of -5.3%. This bias is introduced by the estimates for stations with altitudes larger than 1000m, for which individual MBEs amount to some -15% to -20%. The problems were traced by Möser [4.2.3] as early as 1982 to, first, an inadequacy of his radiation transfer calculations and, second, to the intermingling of snow with cloudiness in the satellite images. Stuhlmann [4.2.4] suggested correcting the first major cause (namely the lack of air mass depletion with altitude in the model calculations) by use of an empirical altitude gradient for clear days. Indeed, the structure of the estimation scheme for daily totals shows that the estimates can *a posteriori* be multiplied by the ratio of clear day insolation at a given altitude to clear day insolation at sea level. Zelenka and Lazic [4.2.5] have investigated the dependence on altitude of some 930 clear sky daily sums recorded by the ANETZ between 1982 and 1990 and they deduced that

$$K_T(h) = K_T(0)[1+(p_k h)/100] \tag{4.4}$$

with h in [hm] and

$$p_k = 1.06 + 0.36 \sin[(\pi/6)(m+2.843)] \tag{4.5}$$

where

$$m = 12N/365 \tag{4.6}$$

N being the day number of the year. *The valid interval of this relation is $4\ hm \leq h \leq 35\ hm$.*

According to (4.5), clear day irradiation increases between 0.7% (Summer) and 1.4% (Winter) per hectometer elevation above sea level. The results of Table 4.2 and those reported in the

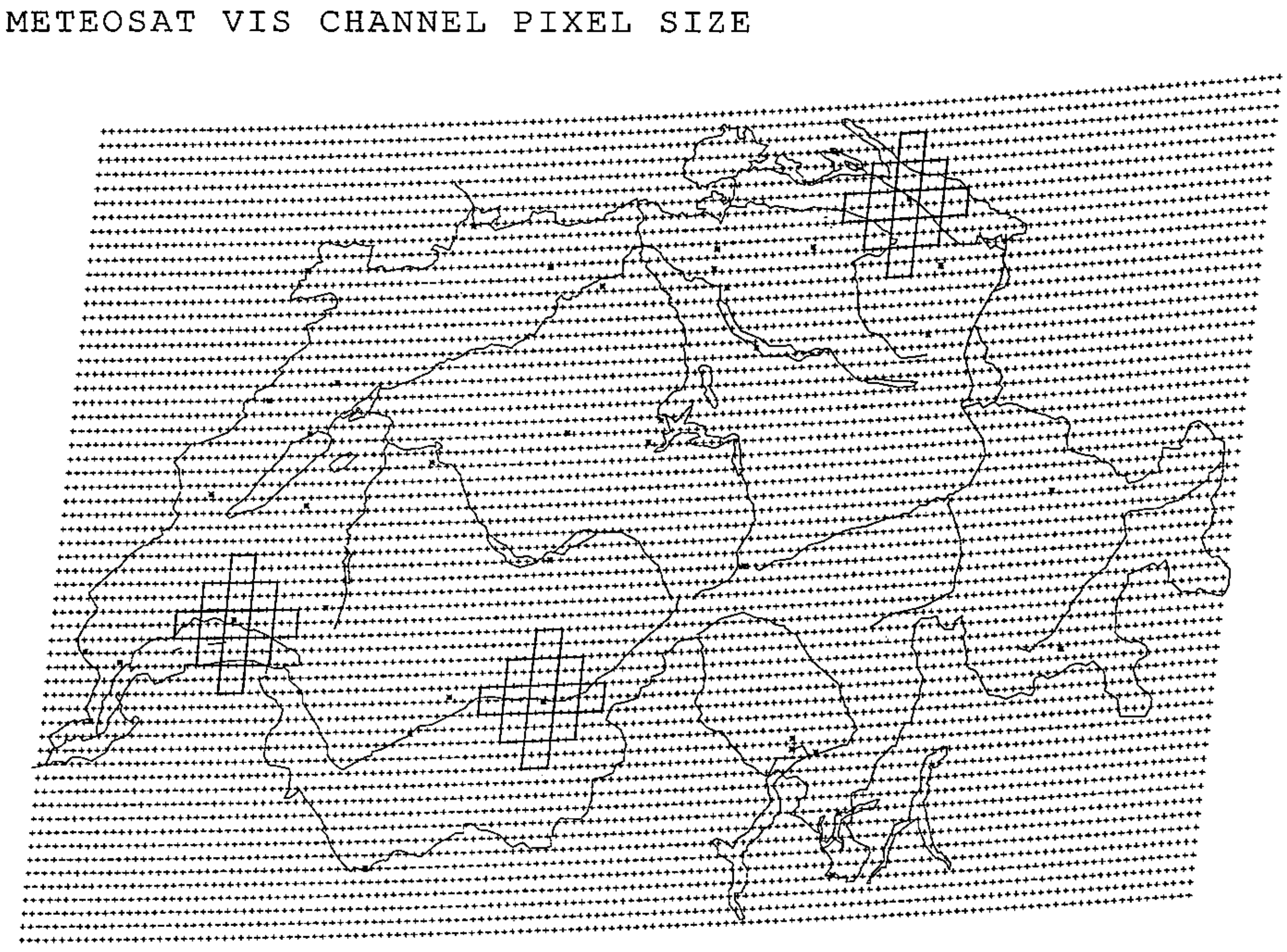


Figure 4.3: Weighting scheme in the METEOSAT/Cologne/DWD model. Elementary pixels (delimited by crosses) correspond to the highest VIS resolution. Insolation is calculated for each area of 7.5 x 7.5km, corresponding to 6 elementary pixels, but estimated insolation for a station (marked with a star) is a weighted average over 13 blocks, as depicted in the 3 examples above. The weights are: 2 for the (central) block containing the site, 1 for blocks having a contiguous side with the central block, 1/√2 for blocks having a contiguous edge with it, and 1/2 for the outermost blocks.

foregoing paragraph, are obtained when the original estimates are multiplied by the factor $(1+p_k h/100)/(1+p_k 4/100)$ whenever h>4 hectometers.

Note that the Cologne method (from which the DWD method was derived) has recently been updated and that it now accounts for the elevation of the insolated surface (Stuhlmann *et al.* [4.2.1]).

4.3 GOES / NESDIS

4.3.1 *Application field*

The NOAA/AgRISTARS insolation product (see Section 2.3.2) covers the 48 contiguous United States (67-125°W, 25-49°N), and the agriculturally significant parts of Mexico (90-110°W, 16-30°N) and South America (40-67°W, 15-41°S). We recall that estimates are made from 40 km x 40 km target areas, with targets centered on a 1° x 1° latitude-longitude grid spacing.

4.3.2 *PNW network*

For 1984 an average of 336 daily sums is available at each grid point (Table 2.4). Estimates for a given site follow from bilinear interpolation between the 4 adjacent grid points. If one of these points is reported missing, estimation for the site is not performed. The number of available values for validation then ranges between 336 (EUG, POR) and 245 (COB). While the first number would justify monthly validation tables, the second does not. Therefore, the format of Tables 4.1 and 4.2 from section 4.2 is maintained for presentation of validations in Table 4.3 of Volume 3 (Results).

The estimated daily totals for the 9 stations of the network reproduce the observed values within a MBE of -7.3% and a RMSE of 19.9% of the observed mean of 13.94 $MJm^{-2}day^{-1}$, a performance very similar to that of the DWD method applied to Switzerland.

The somewhat large underestimation is, again, mainly due to sites above 1000m, namely to Bend (MBE=-15.8%), Burns (-12.5%) and Whitehorse Ranch (-10.4%). The NESDIS estimation method has not been calibrated with measurements from elevated sites, so some adjustment for such sites is required. But use of clear day gradients, which is obvious for the DWD procedure, is not necessarily the correct step here. Rather, use of some mean gradients would seem more appropriate. Yet, if we consider Eq. (4.4) as a first approximation, we definitely improve prediction quality. For, e.g., Burns, underestimation drops to -5.8% and this propagates into the RMSE reducing it from 19.9% to 17.4%. The whole network's MBE is reduced to -4.6% while RMSE decreases to 18.9%. However, it must be noted that underestimation is not only induced by altitude, but also by snow cover, which can not be distinguished from cloud cover (see, e.g., [4.3.4]).

The best individual validation result is for Hood River (MBE=-1.1%, RMSE=14.5% related to an observed average of 15.22 $MJm^{-2}day^{-1}$), while the poorest is for Coos Bay (MBE=-10.3%, RMSE=27.0% related to an observed average of 14.06 $MJm^{-2}day^{-1}$). Actually, the best performance would have been expected for Eugene where the pyranometer site is in the immediate vicinity of the target area. However, there, MBE=-7.0% and RMSE=21.5% (related to an observed average of 12.80$MJm^{-2}day^{-1}$), so that the better performance in Hood River must be credited to better climatic conditions ([4.3.4]; notice the higher measured average) during the shorter sampling period. The smoothing influence of the interpolation must be considered in our evaluation, but this will hardly be the dominant factor because other sites behave similar to Eugene.

The station at Coos Bay lies slightly outside the grid mesh and is close to the ocean. Application of the bilinear scheme then results in extrapolation in the very mathematical sense (not to be confused with the "extrapolation" of Chapter 3), so that assessment of the influence of other interpolation schemes is mandatory.

Validation of moving averages (Section 3.3) with weights proportional to inverse distance squared, to inverse distance, and to inverse square root of distance, with search radii around 200 km, do not improve prediction quality. On the contrary, they even worsen the validation results for Coos Bay while the whole network results change only marginally. In all cases, estimates for sites in the West degrade, while they improve for sites in the East, and the effect is more pronounced the lower the power of the inverse distance. Obviously, coastal sites do not profit from the admixture of inland estimates. Conversely, elevated sites seem to benefit from estimates for surrounding, elevated sites (see map 2.5).

Comparison of this validation with previous ones for the NESDIS method are given in the next section because results for the NEU/SUNY network are similar.

4.3.3 *Northeast network*

A complete validation *including hourly sums and direct beam component* is given in Chapter 6.2, Section 6.2.2.5. We repeat here only the essentials concerning daily totals and devote more space to comparisons with previous validations of the NESDIS method.

Given the distribution of the target areas with respect to the network stations, it is obvious that the bilinear scheme cannot be applied. Differences between moving averages with different inverse distance power laws for the weights are small, so that we retain the "gravity" law (Section 3.3) and the search radius of 250 km which are applied in Section 6.2.2.5. The complete validation Table 4.4 (Volume 3: Results) is, therefore, only a more detailed version of Table 6.2.4 in 6.2.2.5. The small differences between the MBE% and RMSE% displayed in both places result from the use of "detrended" GOES-based estimates in 6.2, while the original estimates have been used here.

The whole network MBE is 2.2% and the RMSE is 20.8% of the measured average of 12.59 $MJm^{-2}day^{-1}$. These values are almost identical to the PNW network's values and comparison with previously published, select validations for equivalent record lenghts allow useful conclusions.

The authors of the GOES/NESDIS method report RMSE% of 10% for not completely independent data [4.3.1] and of 13% for independent data [4.3.2]. Dugas and Heuer [4.1.2] obtain 16% for 20 sites in Texas with individual site values ranging from 11% to 21%, while Frulla et al. [4.3.3] obtain 19.5% for 13 sites in Argentina with individual values between 16% and 25%. Klink and Dollhopf [4.3.4] report 22.5% for 8 sites in Ohio and they assess seasonal and climatic factors in this value. In fact, RMSE% is barely 12% for the snow-free period (1 May - 31 October) but it can exceed 35% at sites with persistent snow cover during the snow cover season.

Thus, the results presented here for two networks located in the northern United States agree well with previous assessments. The advantage here is that the validation is directly comparable with the outcome of interpolation based on the network stations only.

4.4 METEOSAT/"Héliosat"

4.4.1 Application field

Dissemination of Weather Facsimile (WEFAX) formatted METEOSAT images for Secondary Data User Stations (SDUS) follow a fixed schedule. Formats are reproduced from the METEOSAT System Guide in Fig. 4.4. Additional, so-called CD formats are not explicitly displayed in Fig. 4.4.

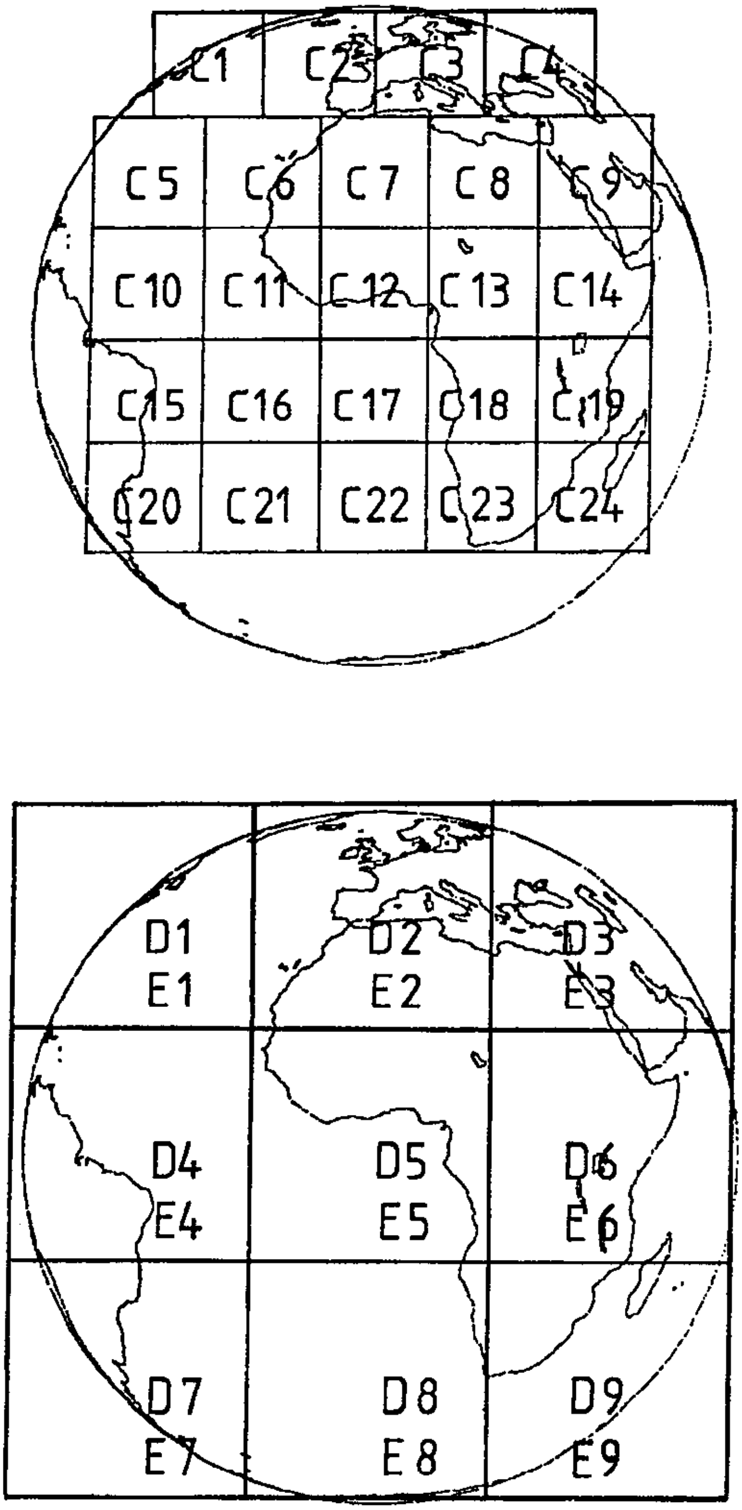


Fig. 4.4: WEFAX visible formats C (above) and infrared (D), resp. water vapor (E) formats below.
From: METEOSAT System Guide, Part I
ESOC, Darmstadt, 1978.

Table 4.1

METEOSAT DISSEMINATION SCHEDULE S9008M01 - CHANNELS A1 (1691MHz) AND A2 (1694.5MHz)

VALID FROM 1 AUGUST 1990

GMT HH	00		03		06		09		12		15		18		21		HH GMT
MM	CH A1	CH A2	CH A1	CH A2	CH A1	CH A2	CH A1	CH A2	CH A1	CH A2	CH A1	CH A2	CH A1	CH A2	CH A1	CH A2	MM
2	D1 48	AI 48	D1 6	AI 6	C02 12	AIVH 12	C02 18	AIVH 18	C02 24	AIVH 24	C02 30	AIVH 30	D1 36	AI 36	D1 42	AI 42	2
6	D3 48	" "	D3 6	" "	C03 12	" "	C03 18	" "	C03 24	" "	C03 30	" "	D3 36	" "	D3 42	" "	6
10	D4 48		D4 6		D1 12	" "	D1 18	" "	D1 24	" "	D1 30	" "	D4 36		D4 42		10
14	D5 48	DTOT 48	D5 6	DTOT 6	D3 12	DTOT 12	D3 18	DTOT 18	D3 24	DTOT 24	D3 30	DTOT 30	D5 36	DTOT 36	D5 42	DTOT 42	14
18	D6 48	ETOT 48	D6 6	ETOT 6	D4 12	ETOT 12	D4 18	CTOT 18	D4 24	CTOT 24	D4 30	CTOT 30	D6 36	ETOT 36	D6 42	ETOT 42	18
22	D7 48		D7 6		D5 12	WEFA 2	D5 18	ATEST 0	D5 24		D5 30		D7 36		D7 42	ATEST 0	22
26	D8 48		D8 6		D6 12	" "	D6 18	" "	D6 24		D6 30		D8 36		D8 42	" "	26
30	D2 1	BIW 1	D2 7	BIW 7	D2 13	BIV 13	D2 19	BIV 19	D2 25	BIV 25	D2 31	BIV 31	D2 37	BIV 37	D2 43	BIW 43	30
34	D9 1	AI 1	D9 7	AI 7	C02 13	AIVH 13	C02 19	AIVH 19	C02 25	AIVH 25	C02 31	AIVH 31	D9 37	AI 37	D9 43	AI 43	34
38	D1 1	" "	D1 7	" "	C03 13	" "	C03 19	" "	C03 25	" "	C03 31	" "	D1 37	" "	D1 43	" "	38
42	D3 1	LY 1	D3 7	LY 7	C3D 13	" "	C8D 19	" "	C3D 25	" "	C8D 31	" "	D3 37	LY 37	D3 43	LY 43	42
46		LR 1		LR 7	C2D 13	LY 13	C9D 19	LY 19	C2D 25	LY 25	C9D 31	LY 31		LR 37		LR 43	46
50					D3 13	LR 13	C2D 19	LR 19	C1D 25	LR 25	D3 31	LR 31		WEFA 5			50
54					D1 13		D1 19		D1 25		D1 31			" "			54
58	D2 2	BIW 2	D2 8	BIW 8	D2 14	BIVW 14	D2 20	BIVW 20	D2 26	BIVW 26	D2 32	BIVW 32	D2 38	BIVW 38	D2 44	BIW 44	58

GMT HH	01		04		07		10		13		16		19		22		HH GMT
MM	CH A1	CH A2	CH A1	CH A2	CH A1	CH A2	CH A1	CH A2	CH A1	CH A2	CH A1	CH A2	CH A1	CH A2	CH A1	CH A2	MM
2	D1 2	AI 2	D1 8	AI 8	C02 14	AIVH 14	C02 20	AIVH 20	C02 26	AIVH 26	C02 32	AIVH 32	D1 38	AI 38	D1 44	AI 44	2
6	D3 2	" "	D3 8	" "	C03 14	" "	C03 20	" "	C03 26	" "	C03 32	" "	D3 38	" "	D3 44	" "	6
10			E1 8		D7 14	" "	D7 20	" "	D7 26	" "	D7 32	" "	E1 38				10
14		LXI 1	E2 8	LXI 7	D8 14	LXI 13	D8 20	LXI 19	D8 26	LZ 25	D8 32	LZ 31	E2 38	LZ 37		LXI 43	14
18		AW 2	E3 8	AW 8	D9 14	AW 14	D9 20	AW 20	D9 26	AW 26	D9 32	AW 32	E3 38	AW 38		AW 44	18
22		" "	E4 8	" "	D3 14	" "	D3 20	" "	D3 26	" "	D3 32	" "	E4 38	" "		" "	22
26			E5 8										E5 38				26
30	D2 3	BIW 3	D2 9	BIW 9	D2 15	BIV 15	D2 21	BIV 21	D2 27	BIV 27	D2 33	BIV 33	D2 39	BIV 39	D2 45	BIW 45	30
34	D1 3	AI 3	D1 9	AI 9	C02 15	AIVH 15	C02 21	AIVH 21	C02 27	AIVH 27	C02 33	AIVH 33	D1 39	AI 39	D1 45	AI 45	34
38	D3 3	" "	D3 9	" "	C03 15	" "	C03 21	" "	C03 27	" "	C03 33	" "	D3 39	" "	D3 45	" "	38
42			E6 9		D1 15	" "	D1 21	" "	D1 27	" "	D1 33	" "	E6 39				42
46			E7 9		D3 15		D3 21		D3 27		D3 33		E7 39				46
50		WEFA 1	E8 9				C1D 21		C1D 27	LXIV 25	C1D 33	LXIV 31	E8 39	LXIV 37			50
54		" "	E9 9		C2D 15		C2D 21		C2D 27	" "	C2D 33	" "	E9 39	" "			54
58	D2 4	BIW 4	D2 10	BIW 10	D2 16	BIVW 16	D2 22	BIVW 22	D2 28	BIVW 28	D2 34	BIVW 34	D2 40	BIW 40	D2 46	BIW 46	58

GMT HH	02		05		08		11		14		17		20		23		HH GMT
MM	CH A1	CH A2	CH A1	CH A2	CH A1	CH A2	CH A1	CH A2	CH A1	CH A2	CH A1	CH A2	CH A1	CH A2	CH A1	CH A2	MM
2	D1 4	AI 4	D1 10	AI 10	C02 16	AIVH 16	C02 22	AIVH 22	C02 28	AIVH 28	C02 34	AIVH 34	D1 40	AI 40	D1 46	AI 46	2
6	D3 4	" "	D3 10	" "	C03 16	" "	C03 22	" "	C03 28	" "	C03 34	" "	D3 40	" "	D3 46	" "	6
10		AW 4		AW 10	C3D 16	" "	C3D 22	" "	C3D 28	" "	C1D 34	" "		AW 40		AW 46	10
14		" "		" "			C4D 22		C4D 28		C4D 34			" "		" "	14
18	TEST 0	ADMIN 0	ADMIN 0	TEST 0	TEST 0	ADMIN 0	ADMIN 0	TEST 0	TEST 0	ADMIN 0	ADMIN 0	TEST 0	TEST 0	ADMIN 0	ADMIN 0	TEST 0	18
22																	22
26																	26
30	D2 5	BIW 5	D2 11	BIW 11	D2 17	BIV 17	D2 23	BIV 23	D2 29	BIV 29	D2 35	BIV 35	D2 41	BIW 41	D2 47	BIW 47	30
34	D1 5	AI 5	D1 11	AI 11	C02 17	AIVH 17	C02 23	AV 23	C02 29	AIVH 29	D1 35	AI 35	D1 41	AI 41	D1 47	AI 47	34
38	D3 5	" "	D3 11	" "	C03 17	" "	C03 23	" "	C03 29	" "	D3 35	" "	D3 41	" "	D3 47	" "	38
42			E1 11		C5D 17	" "	E1 23	" "	C5D 29	" "	E1 35				E1 47		42
46			E2 11		C6D 17	WEFA 3	E2 23	" "	C6D 29	WEFA 4	E2 35			WEFA 6	E2 47		46
50			E3 11		C7D 17	" "	E3 23	" "	C7D 29	" "	E3 35			" "	E3 47		50
54	CTH 0				CTH 0			" "	CTH 0				CTH 0				54
58	D2 6	BIW 6	D2 12	BIVW 12	D2 18	BIVW 18	D2 24	BIVW 24	D2 30	BIVW 30	D2 36	BIVW 36	D2 42	BIW 42	D2 48	BIW 48	58

SOME VISIBLE IMAGE FORMATS ARE SUBJECT TO SUPPRESSION DUE TO SEASONAL UNDERILLUMINATION

They have the extent and the spatial resolution of the D infrared images but contain visible counts. The dissemination schedule valid at present is reproduced in Table 4.1. Available formats (channel A1) can be selected for treatment by entering the appropriate emission time in the Heliosat acquisition table. After reception of the image and completion of the conversion process the insolation map can be accessed for extraction of estimates for those pixels in which the ground stations are located.

4.4.2 Swiss ANETZ

Five (winter) to seven (summer) C03 (C3 in Fig. 4.4) even numbered slots were scheduled for daily treatment between February 1991 and January 1992. Image losses were important at times and the monthly averages needed for validation could not be calculated with all days of the months. The minimum numbers of valid days occurred in February (15 days) and September (17 days), while a maximum was reached in August with 29 days. Most data losses are inherent to the procedure itself. An image must be rejected when the landmarks necessary for automatic superposition with the reference albedo (composite minimum reflectance) image are not clear enough. Further causes for losses were hardware (from antenna to the PC's hard disk) problems, shut-downs for software adaptations, and emission interruptions (METEOSAT reconfiguration, decontamination with mission transfer between METEOSAT4 & 5, and ground segment problems).

Figure 4.5 shows the comparisons between the Heliosat estimates and the observed averages (restrained to the valid days). The RMSE between estimates and observations amounts to 16% for sites below 800 m (294 observations with a mean of 11.81 $MJm^{-2}day^{-1}$) and to 18% for all sites (418 observations with a mean of 12.89 $MJm^{-2}day^{-1}$) after application of an altitude correction. The Cano *et al.* method (see Section 11.4.4 in Volume 3: Theory) used in Heliosat accepts the altitude correction (4.5). In both cases a systematic overestimation is expressed by an MBE of 10%.

These RMSEs for monthly averages are only slightly below those reported in the preceding sections for individual days, so that they may be qualified as large. This is the price to be paid for "personalized" operations with a "desktop" system (whereby, in our case, part of the instabilty is due to the antenna segment on which we have no influence, and this cannot be imputed to the system). At its place of origin and in the configuration in which it was developed, its performances were reported as more stable ([2.14], [2.15], [2.16]). This supports the statement by Schmetz [4.4.1] that precision assessment by authors of models are prone to be optimistic because most techniques have already been tuned on the basis of preliminary investigations before starting the comparisons. Note also that RMS differences for monthly averages gathered by Schmetz have a median value around 5%.

A discussion of possible error sources is given by Zelenka [4.4.2]. Also, an analysis of series of individual days in different seasons reveals that overestimation is due to overcast situations. Snow cover can be fairly well accounted for if the reference (albedo) images are adequately updated. However, each adaptation requires about a day of work which must be added to the hour invested each day in monitoring and servicing the system.

The resulting insolation maps display the correct regional variations at an unequaled spatial resolution level. After correction with differences to the ground measurements (e.g., by cokriging - see Section 5.3) they are best suited for spatial interpolation with high resolution.

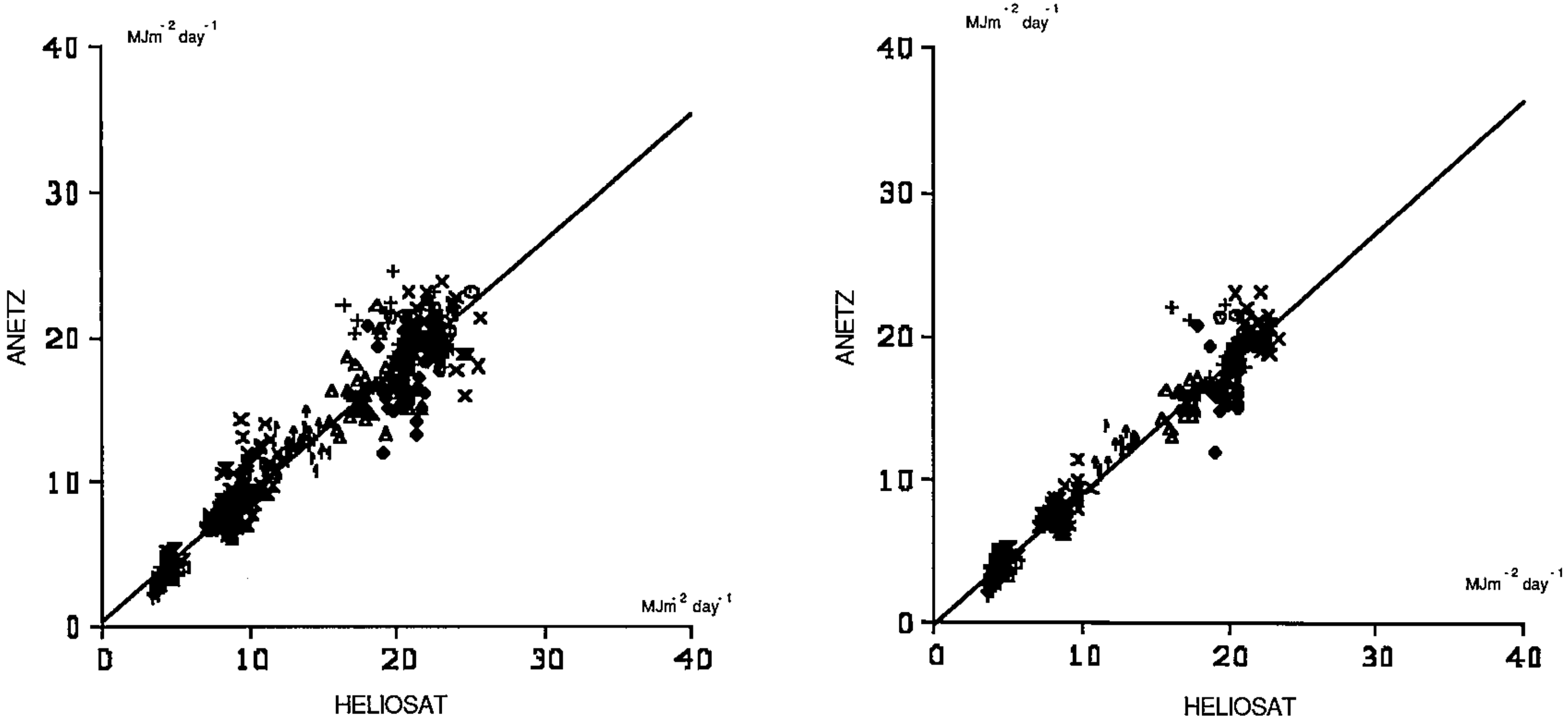


Fig. 4.5: Heliosat estimated and ANETZ observed monthly mean days between February 1991 and January 1992. Left for all 41 stations, right only for stations below 800 m. Full lines are the regression lines.

4.5 References

[4.1.1] N. Benchimol and J.E. Hay, 1986. "An Assessment of the Ability of a Geo-stationary Satellite-Based Model to Characterize the Mesoscale Variability of Solar Irradiance Over the Lower Fraser Valley". *Atmosph.-Ocean* **24**, 128-144.

[4.1.2] W.A. Dugas and M.L. Heuer, 1985. "Relationship between Measured and Satellite-Estimated Solar Irradiance in Texas". *J. Climate & Appl. Meteorol.* **24**, 751-757.

[4.1.3] W.L. Darnell, W.F. Staylor, S.K. Gupta, and F.M. Denn, 1988. Estimation of surface insolation using Sun-synchronous satellite data. *J. Climate* **1**, 820-835.

[4.1.4] M. Nunez, T.L. Hart, and J.D. Kalma, 1984. "Estimating solar radiation in a tropical environment using satellite data". *Jour. of Climatology* **4**, 573-585.

[4.1.5] C. Raphael and J.E. Hay, 1984. "An assessment of models which use satellite data to estimate solar irradiance at the Earth's surface". *J. Climate & Applied Meteor.* **23**, 832-844.

[4.1.6] W. Grüter, H. Guillard, W. Möser, J.M. Monget, W. Palz, E. Raschke, R.E. Reinhardt, P. Schwarzmann, and L. Wald, 1986. "Solar Radiation Data from Satellite Images". *Solar Energy R&D in the European Community,* Series F Solar Radiation Data, **Vol. 4**, pp. 74-85, Reidel.

[4.1.7] C.H. Whitlock, W.F. Staylor, W.L. Darnell, M.D. Chou, G. Dedieu, P.Y. Deschamps, J. Ellis, C. Gautier, R. Frouin, R.T. Pinker, I. Laslo, W.B. Rossow, and D. Tarpley, 1990. "Comparison of surface radiation budget satellite algorithms for downwelled shortwave irradiance with Wisconsin FIRE/SRB surface-truth data". *Proc. 7th AMS Conf. on Atmospheric Radiation, San Francisco, CA, July 23-27, 1990,* 237-242. Published by the American Meteorological Society, Boston, MA.

[4.1.8] C. Delorme, 1987. "Le gisement solaire", G.D.T.A. , Toulouse, F-31055, France, p.141.

[4.1.9] M. Noia: 1990. "Valutatione della radiatione solare al suolo a partire da immagini da satelliti geostazionari". Ph.D. Thesis Dipartimento di Fisica, Università di Genova.

[4.1.10] G. Czeplak, M. Noia, and C. Ratto, 1991. "An assessment of a statistical method to estimate solar irradiance at the Earth's surface from geostationary satellite data". *Renewable Energy* **1**, pp. 737-743.

[4.1.11] E. Pesin, 1988. "Comparison of different models to estimate surface insolation", *Proc. 7th METEOSAT Scientific User's Meeting,* Madrid. EUMETSAT, D-6100 Darmstadt.

[4.1.12] G. Dedieu, 1992. "Surface radiation budget in the shortwaves using the METEOSAT ISCCP B2 data set", *Proc. 9th Meteosat Scientific User's Meeting,* Locarno, 69-76. EUMETSAT, D-6100 Darmstadt. Also private communication and [4.1.7].

[4.2.1] R. Stuhlmann, M. Rieland, and E. Raschke: 1990, "An improvement of the IGMK model to derive total and diffuse solar radiation at the surface from satellite data". *Jour. Applied Meteorol.* **29**, 586-603.

[4.2.2] F.-J. Dieckmann, S. Happ, M. Rieland, W. Benesch, G. Czeplak, and F. Kasten: 1988. An operational estimate of global solar irradiance at ground level from METEOSAT data: results from 1985 to 1987. *Meteorol. Rdsch.* **41**, 65-79.

[4.2.3] W. Möser, 1982: Personnal communication.

[4.2.4] R. Stuhlmann, 1987: Personnal communication.

[4.2.5] A. Zelenka and D. Lazic, 1990: Not yet published.

[4.3.1] C.G. Justus and J.D. Tarpley: 1984. Atlas of satellite-measured insolation in the United-States, Mexico, and South America. Rept. for the USDOE SOLERAS Program.

[4.3.2] C.G. Justus, M.V. Paris and J.D. Tarpley: 1986. "Satellite-measured insolation in the United-States, Mexico, and South America", *Remote Sensing of Environment* **20**, 57.

[4.3.3] L.A. Frulla, D.A. Gagliardini, H. Grossi Gallegos, R Lopardo, and J.D. Tarpley: 1988. Incident solar radiation on Argentina from the geostationary satellite GOES: comparison with ground-measurements. *Solar Energy* **41**, 61.

[4.3.4] J.C. Klink and K.J. Dollhopf: 1986. "An evaluation of satellite-based insolation estimates for Ohio", *J. Climate & Applied Meteor.* **25**, 1741.

[4.4.1] J. Schmetz: 1989. "Towards a surface radiation climatology: retrieval of downward irradiances from satellites", *Atm. Res.* **23**, 287.

[4.4.2] A. Zelenka: 1992. "Estimating monthly average daily insolation with the SDUS-based "Héliosat" system: Experiences and results from one year of operations", *Proc. 9th Meteosat Scientific User's Meeting*, Locarno, 109-114. EUMETSAT, D-6100 Darmstadt.

CHAPTER 5

5. Validation of compound methods

5.1 Introduction

Until now, either the network data (Chapter 3) or the satellite data (Chapter 4) have been considered in the process of predicting global insolation for sites where adequate measurements are missing, but not both. In Section 3.5, however, monthly sums derived (by means of Ångstroem's relation) at sunshine duration measuring sites were added to those measured by the DWD radiometric network. Spatial density was then sufficient to permit principal components analysis. This chapter will now describe several methods where available *additional information is input in the estimation* process to increase its precision. Further such methods will be discussed in *Chapter 6* in the context of immediate application of their outcome.

Section 5.2 will elaborate on the use of sunshine duration measuring sites and this will introduce *cluster analysis*. Then, Section 5.3 will deal with merging of satellite-based estimates with network data, such that spatial resolution of the former and calibration accuracy of the latter are optimally exploited. This will introduce *cokriging*. Finally, Section 5.4 will introduce a bulk radiation model based on meteorological input parameters (like the models investigated by Subtask 9B - see reference [5.4.2]), which is unique in that the parameters originate from the *analysis* step required for running numerical weather forecasting models. This analysis consists of *gridding irregularly distributed meteorological observations* by means of *multivariate optimum interpolation* and by means of satellite-based measurements. Consequently, hourly global insolation can be retrieved by the radiation model at each grid point and eventually be forecast there.

It is important to realize that cluster analysis need not be confined to sunshine duration, and that the kriging covariable may be any quantity related to the network, not just satellite-based estimates. In other words, the techniques described are not limited to just the variables with which they are introduced here.

5.2 Use of Auxiliary Information Based On Cluster Analysis

Study prepared by: *Gerhard Czeplak*
DWD, Meteorologisches Observatorium Hamburg

Under sponsorship from: *Bundesministerium für Forschung und Technologie (BMFT)*

5.2.1 Introduction

Because sunshine duration can be measured more easily than irradiation, the station density is markedly higher than for irradiation. Therefore, sunshine duration can be used as auxilliary information to compensate for the relative sparsity of the German radiometric network. Two procedures are performed. The first one merges global radiation and sunshine duration measured at the same station by using Ångstroem's formula. The second one delineates regions with similar sunshine duration by cluster analysis (see chapter 10.4 in Volume 3: Theory). If we postulate that the same Ångstroem coefficients are valid throughout a whole region, we can calculate the global radiation at each sunshine site.

5.2.2 Ångstroem Regression

Ångstroem's formula is the simplest and best known relation between global radiation and sunshine duration. Its linear approximation reads

$$G/G_0 = a + b(S/S_0) \ ,$$

where G = global radiation to be calculated, G_0 = extraterrestrial insolation on a horizontal surface, S = measured sunshine duration, S_0 = astronomical sunshine duration, and a and b are coefficients to be determined empirically.

For the year 1988, a linear regression is performed between the daily sums of global radiation and the sunshine duration measured at 29 stations of the German radiometric network. Tables 5.1 and 5.2 display the calculated coefficients a and b for each month of 1988, as well as the respective correlation coefficients R (Table 5.3) which are criterions for the quality of the linearity between G and S. The mean value of R is about 0.95, except for the winter months. In any case, the lowest value is 0.88. Due to these relative high correlation coefficients, the assumption of a linear relation between G and S, even in the winter months, seams quite acceptable.

Besides the statistical errors, further errors occur in calculating global radiation from sunshine duration [5.2.1] because:

1. The threshold sensitivity of the Campbell-Stokes sunshine recorder depends on the turbidity and the water vapor concentration of the atmosphere.

2. The interpretation of the recorder chart might be erroneous.

3. Extremely high values of the daily sums of G under cloudy sky conditions are due to reflections by clouds.

4. The low-energy insolation in the morning and in the evening are weighted in the same way as the high-energy insolation around noon.

In spite of these constraints, the application of Ångstroem's regression formula yields good results, if it is used for the calculation of monthly means, but it fails for daily sums. In our case (year of 1988) the MBEs are lower than 1 %.

The calculated coefficients a and b for the 29 stations of the German radiometric network vary slightly from month to month and their seasonal variations are to some extent opposite (Tables 5.1 and 5.2). Coefficient a can be interpreted as the transmissivity of the atmosphere under overcast sky conditions and the sum of a and b as the transmissivity of a clear sky. In general, coefficient a shows a slight increase during the summer months which is likely the result of the high solar elevation (short optical path through the clouds). Since the values of a and b are determined from the daily sums of S/S_0, the sum of a + b indicates the mean transmissivity of the atmosphere. Fluctuations of a + b within each of the months are due to the variability of the optical characteristics of the atmosphere.

5.2.3 Cluster Analysis

As the first step for this type of analysis, the daily sunshine durations (the variables) for about 200 stations (the objects to be clustered) of the German Radiation Network for the year 1988 were arranged in data matrices, one for each month. Because it makes no sense to perform a cluster analysis with correlated variables, correlation matrices were calculated (see Section 10.4.3.2). The results show that, with a few exceptions, the variables are not correlated with each other. The results for the months of June and December are in Table 5.4 and 5.5. Therefore, the square of the Euclidian distance (sunshine space) can be chosen as a criterion for the similarity of stations concerning irradiation (see Section 10.4.3.3). Due to the squaring, large "distances" are weighted stronger than short distances. To merge similar objects or clusters, Ward's algorithm was selected. This is one of the hierarchic agglomerative methods that is looked upon as a standard that yields good partitions (see Section 10.4.4).

After these preliminary examinations, a decision on the number of clusters or regions has to be made. These regions should contain stations with similar sunshine duration (see Section 10.4.5). Plotting the sum of the square of the Euclidian distances or the sum of the squares of errors of every merging step versus the number of clusters yields a so called structogram. A sudden rise of the error sum of squares at a certain merger indicates the forming of a heterogeneous cluster from two homogeneous clusters.

As an example, Fig. 5.2.1 displays the structogram for the month of June 1988. In this case, the process should be stopped after the forming of either 7 or 13 clusters. Each cluster or region must contain at least one station of the global radiation network.

Finally, each of the clusters or regions with their sunshine duration stations can be determined by means of a so-called dendrogram. It describes the hierarchical structure of this method. In Fig. 5.2.2 the dendrogram for the month of June 1988 is shown. The percentage of the total sum of squares (SQT) that is represented by the error sum of squares (SQI) is plotted versus the last 32 mergers. The branch number indicates the cluster or station name. Each cluster includes one or more well known stations. If we insert the result of the structogram from above (dashed straight line parallel to x-axis crossing 13 cluster branches), we obtain the clusters sought after. Fig. 5.2.3 displays the clusters or regions of this analysis (interpolation method: nearest neighbor).

If a single cluster contains more than one station of the global radiation network, Ångstroem's coefficients of this cluster are calculated as the average of the coefficients of the global radiation stations in that cluster.

Finally, the monthly means of daily sums of the global radiation of all of the sunshine stations are calculated by using the averages of the coeffcients of the respective cluster. This prepares them for further statistical examination to allow gravity interpolation or EOF interpolation or kriging at a scale adequate for mapping.

5.2.4 Reference

[5.2.1] Golchert, H.-J., Mittlere monatliche Globalstrahlungsverteilungen in der Bundesrepublik Deutschland. *Meteor. Rdsch.* **34** (1980) 143-151.

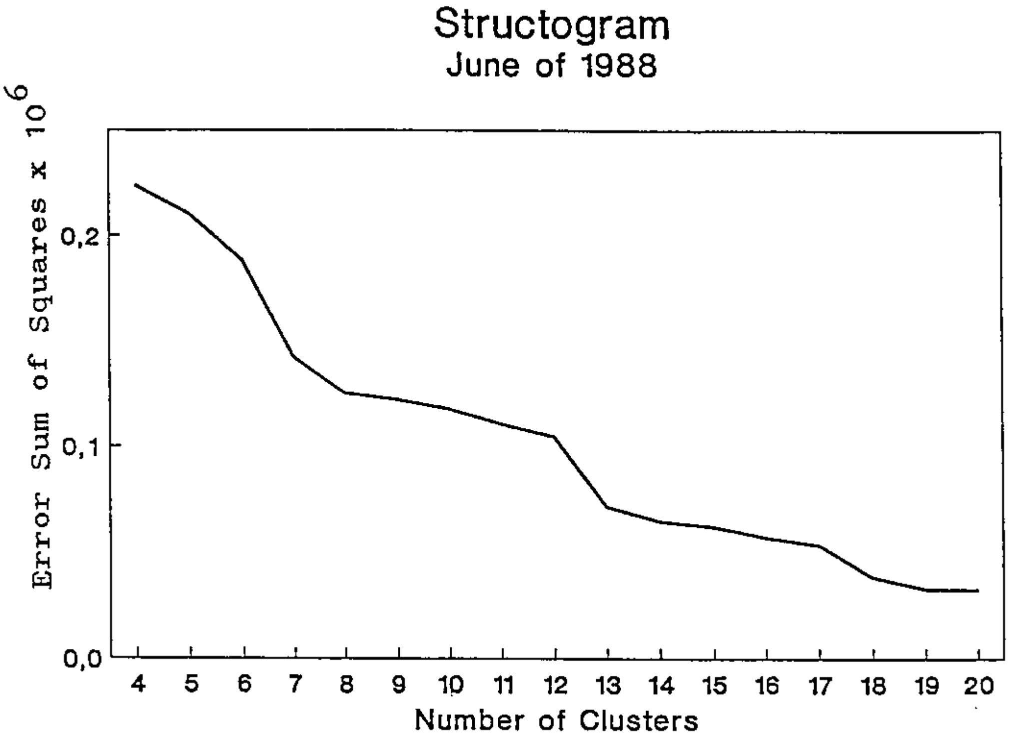


Fig. 5.2.1

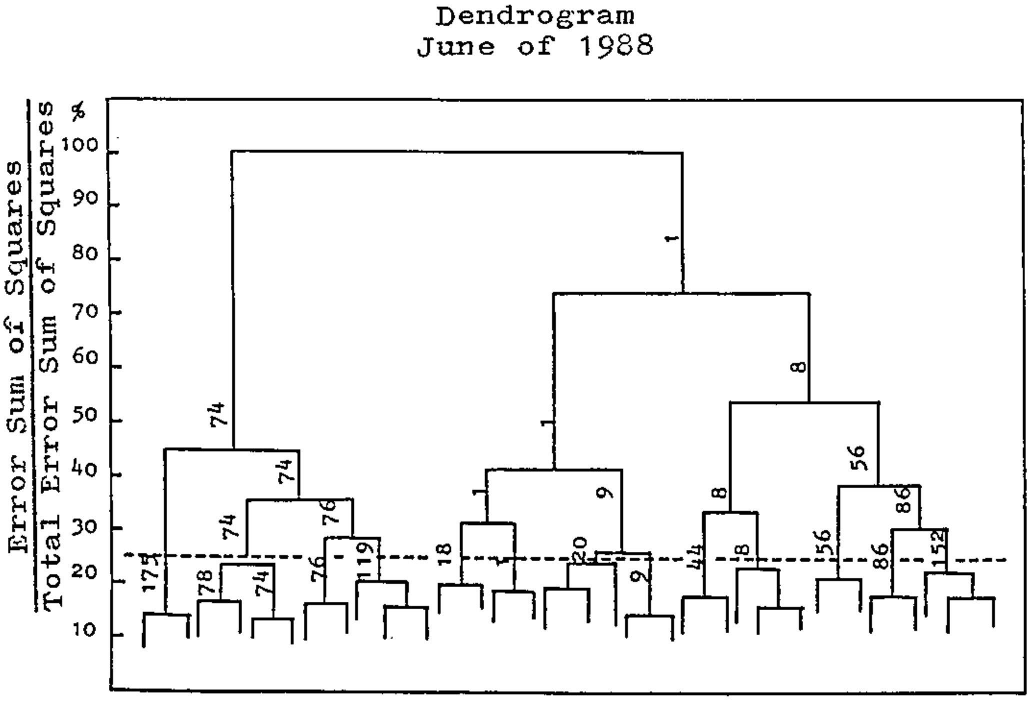


Fig. 5.2.2

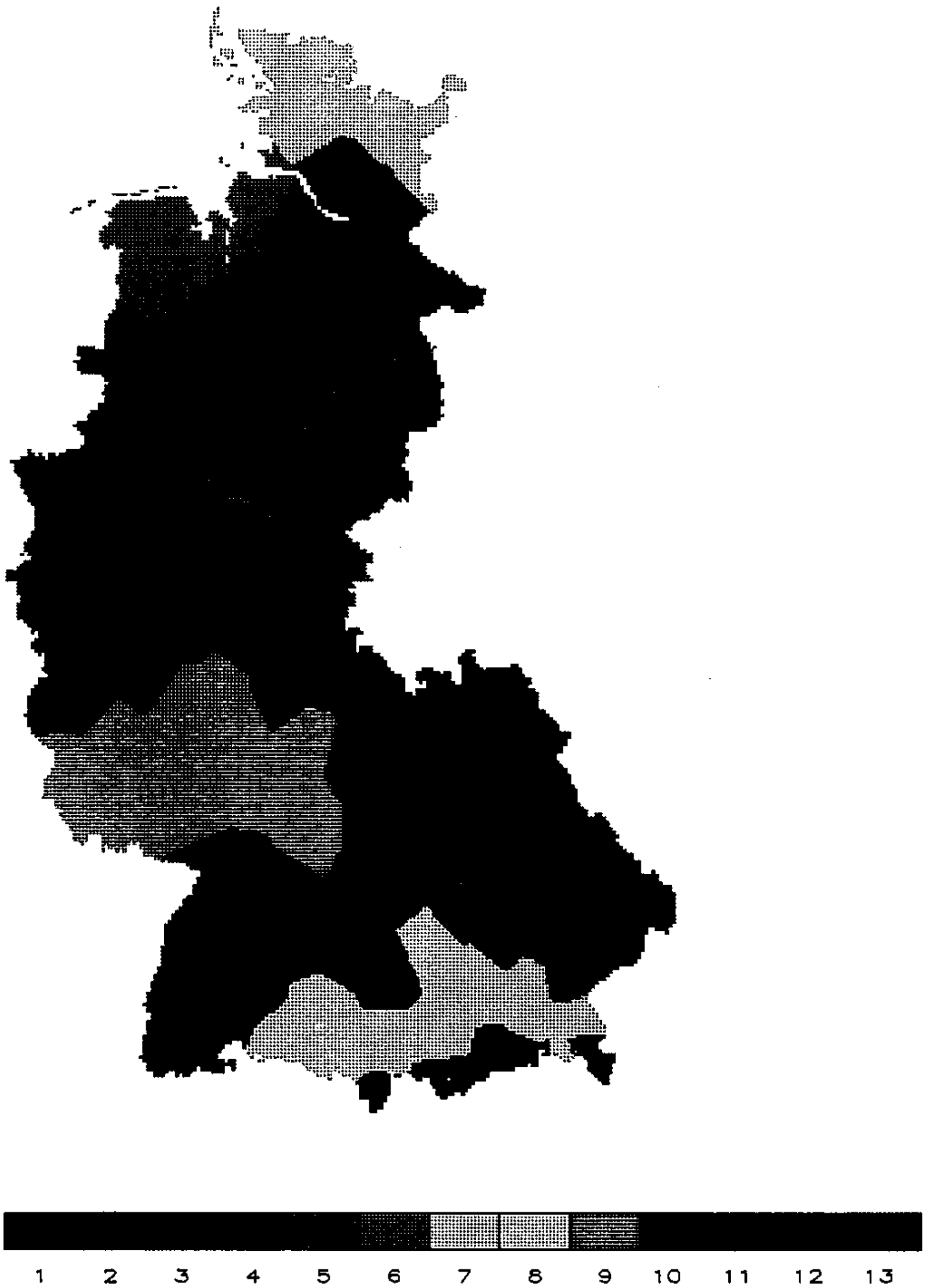


Fig. 5.2.3 Clusteranalysis
Sunshine Duration
June 1988
13 Clusters

Table 5.1

Monthly Ångstroem Regression Coefficient a of the Regression G/G0 = a + b·S/S0
German Radiometric Network 1988

Station	J	F	M	A	M	J	J	A	S	O	N	D
Bocholt	0.156	0.179	0.196	0.240	0.224	0.224	0.179	0.257	0.189	0.191	0.195	0.149
Bonn	0.209	0.189	0.185	0.207	0.223	0.222	0.182	0.227	0.203	0.179	0.181	0.138
Braunschweig	0.158	0.166	0.177	0.215	0.254	0.185	0.226	0.256	0.199	0.177	0.179	0.148
Braunlage	0.121	0.176	0.194	0.272	0.211	0.193	0.177	0.194	0.182	0.137	0.186	0.134
Bremen	0.151	0.194	0.188	0.197	0.247	0.186	0.197	0.239	0.191	0.150	0.207	0.163
Coburg	0.152	0.190	0.206	0.233	0.231	0.205	0.187	0.210	0.186	0.181	0.169	0.127
Kassel	0.159	0.174	0.208	0.169	0.237	0.212	0.216	0.240	0.196	0.169	0.198	0.144
Freiburg	0.172	0.161	0.156	0.208	0.231	0.198	0.158	0.171	0.189	0.160	0.193	0.146
Gelsenkirchen	0.154	0.167	0.158	0.197	0.207	0.199	0.180	0.243	0.173	0.192	0.182	0.134
Geisenheim	0.141	0.174	0.176	0.202	0.259	0.230	0.176	0.215	0.194	0.165	0.175	0.139
Gießen	0.137	0.156	0.181	0.161	0.215	0.234	0.182	0.197	0.218	0.171	0.177	0.134
Hamburg-Sasel	0.148	0.160	0.192	0.211	0.256	0.220	0.216	0.244	0.183	0.185	0.164	0.151
Hohenpeißenberg	0.235	0.226	0.201	0.221	0.223	0.196	0.179	0.155	0.184	0.204	0.207	0.189
Konstanz	0.196	0.208	0.183	0.221	0.257	0.202	0.173	0.175	0.204	0.173	0.188	0.158
Lüdenscheid	0.128	0.144	0.158	0.203	0.197	0.203	0.152	0.237	0.155	0.174	0.150	0.097
Lippspringe, Bad	0.122	0.152	0.165	0.163	0.211	0.202	0.174	0.237	0.190	0.181	0.161	0.099
List	0.151	0.182	0.188	0.237	0.253	0.256	0.178	0.196	0.167	0.162	0.176	0.158
Mannheim	0.164	0.163	0.158	0.191	0.193	0.200	0.155	0.168	0.198	0.160	0.166	0.128
Nürnberg	0.182	0.177	0.186	0.188	0.229	0.222	0.192	0.160	0.181	0.185	0.189	0.142
Norderney	0.153	0.177	0.192	0.260	0.250	0.233	0.209	0.213	0.185	0.162	0.200	0.182
Osnabrück	0.143	0.171	0.163	0.179	0.221	0.199	0.199	0.254	0.192	0.180	0.172	0.137
Passau	0.195	0.196	0.178	0.210	0.267	0.202	0.169	0.162	0.176	0.150	0.163	0.138
Schleswig	0.139	0.173	0.176	0.194	0.215	0.215	0.204	0.213	0.153	0.164	0.166	0.146
Saarbrücken	0.140	0.159	0.162	0.219	0.222	0.245	0.169	0.238	0.204	0.179	0.000	0.000
Stuttgart	0.180	0.201	0.202	0.200	0.228	0.196	0.168	0.171	0.171	0.164	0.198	0.159
Trier	0.168	0.159	0.157	0.216	0.226	0.229	0.179	0.231	0.195	0.176	0.205	0.126
Weißenburg	0.202	0.192	0.200	0.210	0.250	0.211	0.195	0.161	0.179	0.161	0.000	0.000
Weihenstephan	0.208	0.219	0.224	0.228	0.248	0.228	0.217	0.189	0.214	0.177	0.190	0.175
Würzburg	0.182	0.203	0.223	0.207	0.263	0.245	0.253	0.239	0.231	0.179	0.206	0.141

Table 5.2

Monthly Ångstroem Regression Coefficient b of the Regression $G/G0 = a + b \cdot S/S0$
German Radiometric Network 1988

Station	J	F	M	A	M	J	J	A	S	O	N	D
Bocholt	0.569	0.548	0.578	0.526	0.543	0.557	0.628	0.457	0.563	0.534	0.549	0.674
Bonn	0.541	0.518	0.659	0.556	0.544	0.582	0.629	0.551	0.602	0.577	0.519	0.756
Braunschweig	0.615	0.621	0.702	0.554	0.497	0.647	0.529	0.465	0.556	0.572	0.558	0.645
Braunlage	0.692	0.629	0.718	0.517	0.554	0.648	0.616	0.557	0.577	0.680	0.574	0.571
Bremen	0.751	0.510	0.551	0.509	0.415	0.534	0.513	0.462	0.515	0.548	0.649	0.522
Coburg	0.797	0.651	0.695	0.565	0.520	0.577	0.568	0.536	0.582	0.589	0.573	0.688
Kassel	0.547	0.560	0.615	0.639	0.497	0.556	0.546	0.481	0.572	0.551	0.518	0.586
Freiburg	0.553	0.591	0.653	0.563	0.522	0.596	0.594	0.562	0.549	0.608	0.519	0.690
Gelsenkirchen	0.543	0.541	0.606	0.536	0.528	0.592	0.628	0.484	0.558	0.494	0.491	0.700
Geisenheim	0.657	0.557	0.598	0.585	0.473	0.550	0.559	0.508	0.554	0.588	0.549	0.639
Gießen	0.639	0.582	0.582	0.611	0.532	0.544	0.566	0.513	0.508	0.549	0.560	0.585
Hamburg-Sasel	0.610	0.625	0.651	0.584	0.511	0.591	0.633	0.522	0.604	0.564	0.550	0.509
Hohenpeißenberg	0.466	0.529	0.725	0.565	0.544	0.621	0.595	0.599	0.590	0.510	0.523	0.538
Konstanz	0.585	0.533	0.712	0.581	0.531	0.644	0.616	0.580	0.545	0.607	0.548	0.576
Lüdenscheid	0.676	0.605	0.696	0.578	0.572	0.595	0.601	0.463	0.621	0.572	0.539	0.886
Lippspringe, Bad	0.625	0.564	0.724	0.659	0.545	0.610	0.623	0.506	0.560	0.572	0.568	0.612
List	0.598	0.544	0.594	0.538	0.507	0.526	0.571	0.566	0.563	0.566	0.529	0.542
Mannheim	0.510	0.551	0.568	0.566	0.558	0.518	0.561	0.531	0.513	0.571	0.532	0.556
Nürnberg	0.557	0.579	0.696	0.571	0.508	0.514	0.543	0.578	0.561	0.573	0.593	0.647
Norderney	0.667	0.616	0.618	0.521	0.519	0.603	0.593	0.574	0.586	0.589	0.538	0.720
Osnabrück	0.677	0.543	0.640	0.626	0.516	0.591	0.561	0.473	0.558	0.576	0.574	0.675
Passau	0.625	0.571	0.721	0.593	0.509	0.627	0.604	0.631	0.645	0.660	0.591	0.802
Schleswig	0.714	0.550	0.599	0.571	0.541	0.569	0.525	0.493	0.573	0.560	0.529	0.600
Saarbrücken	0.611	0.602	0.636	0.546	0.525	0.492	0.561	0.474	0.529	0.534	0.000	0.000
Stuttgart	0.545	0.510	0.601	0.594	0.549	0.569	0.586	0.582	0.606	0.627	0.536	0.606
Trier	0.649	0.593	0.732	0.570	0.529	0.559	0.585	0.503	0.587	0.629	0.533	0.769
Weißenburg	0.554	0.573	0.805	0.590	0.516	0.584	0.561	0.584	0.586	0.617	0.000	0.000
Weihenstephan	0.459	0.493	0.599	0.533	0.513	0.556	0.547	0.566	0.518	0.536	0.537	0.546
Würzburg	0.588	0.580	0.622	0.603	0.533	0.592	0.518	0.455	0.545	0.644	0.593	0.764

Table 5.3

Monthly Correlation Coefficient R of the Regression G/G0 = a + b·S/S0
German Radiometric Network 1988

Station	J	F	M	A	M	J	J	A	S	O	N	D
Bocholt	0.884	0.908	0.924	0.971	0.958	0.944	0.961	0.928	0.953	0.948	0.926	0.859
Braunschweig	0.933	0.922	0.915	0.952	0.922	0.958	0.964	0.940	0.959	0.883	0.952	0.921
Braunlage	0.929	0.928	0.883	0.963	0.954	0.960	0.954	0.955	0.969	0.946	0.912	0.932
Bremen	0.856	0.838	0.895	0.953	0.952	0.924	0.934	0.928	0.957	0.947	0.488	0.905
Coburg	0.813	0.917	0.887	0.934	0.960	0.959	0.951	0.956	0.961	0.933	0.923	0.820
Kassel	0.882	0.947	0.903	0.966	0.959	0.953	0.957	0.964	0.969	0.919	0.922	0.906
Freiburg	0.882	0.982	0.908	0.963	0.960	0.973	0.972	0.978	0.964	0.963	0.934	0.868
Gelsenkirchen	0.916	0.936	0.966	0.970	0.956	0.955	0.965	0.958	0.932	0.923	0.953	0.717
Geisenheim	0.935	0.938	0.936	0.978	0.864	0.952	0.966	0.980	0.957	0.930	0.939	0.882
Gießen	0.895	0.960	0.916	0.979	0.950	0.962	0.959	0.963	0.942	0.941	0.935	0.879
Hamburg-Sasel	0.897	0.912	0.932	0.967	0.963	0.935	0.934	0.941	0.952	0.946	0.938	0.935
Hohenpeißenberg	0.896	0.936	0.934	0.955	0.958	0.936	0.968	0.967	0.967	0.922	0.931	0.878
Konstanz	0.896	0.939	0.910	0.969	0.971	0.943	0.963	0.951	0.964	0.959	0.855	0.894
Lüdenscheid	0.957	0.963	0.880	0.962	0.966	0.939	0.947	0.820	0.965	0.927	0.965	0.878
Lippspringe, Bad	0.967	0.938	0.850	0.963	0.965	0.952	0.928	0.953	0.946	0.914	0.937	0.912
List	0.852	0.912	0.954	0.968	0.976	0.956	0.962	0.929	0.967	0.968	0.952	0.911
Mannheim	0.890	0.944	0.880	0.978	0.954	0.953	0.971	0.977	0.968	0.963	0.949	0.864
Nürnberg	0.921	0.954	0.884	0.951	0.960	0.944	0.964	0.974	0.940	0.955	0.919	0.941
Norderney	0.843	0.918	0.943	0.966	0.962	0.943	0.936	0.964	0.981	0.969	0.911	0.833
Osnabrück	0.924	0.887	0.924	0.950	0.962	0.958	0.943	0.945	0.959	0.908	0.909	0.899
Passau	0.929	0.956	0.928	0.950	0.971	0.948	0.967	0.984	0.966	0.966	0.944	0.849
Schleswig	0.883	0.938	0.921	0.952	0.958	0.948	0.926	0.917	0.943	0.956	0.939	0.926
Saarbrücken	0.904	0.970	0.902	0.941	0.957	0.945	0.971	0.960	0.963	0.945	0.000	0.000
Stuttgart	0.907	0.945	0.887	0.966	0.973	0.968	0.964	0.956	0.966	0.966	0.943	0.856
Trier	0.800	0.970	0.942	0.962	0.952	0.960	0.983	0.982	0.955	0.911	0.923	0.812
Weißenburg	0.914	0.944	0.869	0.954	0.963	0.948	0.973	0.963	0.981	0.958	0.000	0.000
Weihenstephan	0.929	0.905	0.850	0.968	0.962	0.928	0.973	0.969	0.970	0.934	0.950	0.826
Würzburg	0.917	0.933	0.871	0.953	0.940	0.935	0.933	0.898	0.964	0.954	0.910	0.904

Table 5.4

Correlation between the Daily Sums of Sunshine Duration

June 1988

German Radiometric Network

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	1.00	-.09	-.31	.28	.04	-.27	-.05	-.18	-.21	-.07	-.14	-.10	.25	.18	.03	.14	.10	.16	-.06	-.11	-.09	-.01	.06	-.23	-.22	-.14	-.11	.06	.14	.02
2	-.09	1.00	-.04	.24	.53	.54	-.11	-.20	-.32	.16	-.02	.60	.17	-.58	-.21	-.29	.42	-.46	-.44	.03	.46	.16	-.29	-.13	.08	.20	-.34	.09	-.34	-.03
3	-.31	-.04	1.00	-.70	-.48	-.23	.16	.69	.70	.23	.55	-.05	-.74	-.04	-.06	.21	-.24	.15	.55	.57	.11	.06	.28	.80	.54	.28	.65	.14	.42	.38
4	.28	.24	-.70	1.00	.60	.34	-.25	-.50	-.66	-.07	-.32	.26	.52	-.17	-.26	-.43	.20	-.41	-.59	-.49	.07	.03	-.22	-.62	-.35	-.20	-.65	-.21	-.47	-.17
5	.04	.53	-.48	.60	1.00	.70	-.24	-.56	-.65	-.12	-.13	.61	.50	-.62	-.24	-.44	.48	-.69	-.67	-.35	.36	.08	-.33	-.52	-.29	-.14	-.69	-.15	-.65	-.28
6	-.27	.54	-.23	.34	.70	1.00	-.19	-.36	-.44	-.17	.01	.67	.26	-.57	-.08	-.37	.41	-.65	-.56	-.16	.41	.05	-.34	-.25	-.07	-.02	-.49	-.02	-.54	-.30
7	-.05	-.11	.16	-.25	-.24	-.19	1.00	.44	.31	.21	-.03	-.43	.05	.28	.05	.33	-.06	.22	.17	-.18	-.39	-.28	-.36	-.14	-.03	.28	.19	.36	.07	.33
8	-.18	-.20	.69	-.50	-.56	-.36	.44	1.00	.69	.21	.42	-.35	-.57	.33	-.05	.26	-.34	.33	.56	.40	-.08	-.04	.09	.54	.51	.42	.67	.34	.46	.54
9	-.21	-.32	.70	-.66	-.65	-.44	.31	.69	1.00	.24	.52	-.46	-.58	.31	.25	.61	-.28	.44	.75	.41	-.11	-.16	.41	.67	.45	.30	.76	.23	.60	.42
10	-.07	.16	.23	-.07	-.12	-.17	.21	.21	.24	1.00	.08	-.03	-.11	-.06	-.16	.11	-.03	-.01	.22	.13	.19	.13	.12	.24	.37	.24	.21	.04	.21	.24
11	-.14	-.02	.55	-.32	-.13	.01	-.03	.42	.52	.08	1.00	.04	-.37	-.09	.11	.36	.18	-.03	.44	.35	.20	-.21	.37	.55	.15	.29	.32	.15	.35	.29
12	-.10	.60	-.05	.26	.61	.67	-.43	-.35	-.46	-.03	.04	1.00	.08	-.63	-.24	-.46	.43	-.62	-.56	.22	.61	.33	-.12	.04	.08	.07	-.34	-.04	-.39	-.19
13	.25	.17	-.74	.52	.50	.26	.05	-.57	-.58	-.11	-.37	.08	1.00	-.00	.11	-.05	.40	-.14	-.42	-.59	-.07	-.13	-.40	-.79	-.52	-.06	-.61	.09	-.35	-.19
14	.18	-.58	-.04	-.17	-.62	-.57	.28	.33	.31	-.06	-.09	-.63	-.00	1.00	.33	.37	-.32	.69	.37	-.02	-.51	-.23	.07	.09	-.01	.13	.39	.23	.45	.21
15	.03	-.21	-.06	-.26	-.24	-.08	.05	-.05	.25	-.16	.11	-.24	.11	.33	1.00	.65	.14	.38	.33	.09	-.06	-.27	.20	.05	-.14	-.02	.22	.24	.43	-.15
16	.14	-.29	.21	-.43	-.44	-.37	.33	.26	.61	.11	.36	-.46	-.05	.37	.65	1.00	.12	.47	.62	.17	-.15	-.32	.31	.23	.02	.21	.47	.43	.59	.22
17	.10	.42	-.24	.20	.48	.41	-.06	-.34	-.28	-.03	.18	.43	.40	-.32	.14	.12	1.00	-.30	-.29	-.01	.22	-.16	-.07	-.19	-.21	.12	-.26	.18	-.06	-.04
18	.16	-.46	.15	-.41	-.69	-.65	.22	.33	.44	-.01	-.03	-.62	-.14	.69	.38	.47	-.30	1.00	.54	.22	-.32	-.17	.28	.25	.11	.06	.50	.20	.55	.19
19	-.06	-.44	.55	-.59	-.67	-.56	.17	.56	.75	.22	.44	-.56	-.42	.37	.33	.62	-.29	.54	1.00	.41	-.12	-.06	.56	.56	.28	.15	.75	.18	.69	.30
20	-.11	.03	.57	-.49	-.35	-.16	-.18	.40	.41	.13	.35	.22	-.59	-.02	.09	.17	-.01	.22	.41	1.00	.43	.40	.52	.74	.46	.08	.64	.22	.50	.17
21	-.09	.46	.11	.07	.36	.41	-.39	-.08	-.11	.19	.20	.61	-.07	-.51	-.06	-.15	.22	-.32	-.12	.43	1.00	.40	.14	.21	.31	.05	-.03	.10	-.05	-.12
22	-.01	.16	.06	.03	.08	.05	-.28	-.04	-.16	.13	-.21	.33	-.13	-.23	-.27	-.32	-.16	-.17	-.06	.40	.40	1.00	.20	.12	.30	-.22	.10	-.08	-.07	-.13
23	.06	-.29	.28	-.22	-.33	-.34	-.36	.09	.41	.12	.37	-.12	-.40	.07	.20	.31	-.07	.28	.56	.52	.14	.20	1.00	.57	.20	-.17	.52	-.20	.57	.04
24	-.23	-.13	.80	-.62	-.52	-.25	-.14	.54	.67	.24	.55	.04	-.79	.09	.05	.23	-.19	.25	.56	.74	.21	.12	.57	1.00	.63	.20	.70	.01	.55	.25
25	-.22	.08	.54	-.35	-.29	-.07	-.03	.51	.45	.37	.15	.08	-.52	-.01	-.14	.02	-.21	.11	.28	.46	.31	.30	.20	.63	1.00	.26	.52	.17	.33	.20
26	-.14	.20	.28	-.20	-.14	-.02	.28	.42	.30	.24	.29	.07	-.06	.13	-.02	.21	.12	.06	.15	.08	.05	-.22	-.17	.20	.26	1.00	.23	.45	.17	.42
27	-.11	-.34	.65	-.65	-.69	-.49	.19	.67	.76	.21	.32	-.34	-.61	.39	.22	.47	-.26	.50	.75	.64	-.03	.10	.52	.70	.52	.23	1.00	.25	.70	.35
28	.06	.09	.14	-.21	-.15	-.02	.36	.34	.23	.04	.15	-.04	.09	.23	.24	.43	.18	.20	.18	.22	.10	-.08	-.20	.01	.17	.45	.25	1.00	.22	.28
29	.14	-.34	.42	-.47	-.65	-.54	.07	.46	.60	.21	.35	-.39	-.35	.45	.43	.59	-.06	.55	.69	.50	-.05	-.07	.57	.55	.33	.17	.70	.22	1.00	.28
30	.02	-.03	.38	-.17	-.28	-.30	.33	.54	.42	.24	.29	-.19	-.19	.21	-.15	.22	-.04	.19	.30	.17	-.12	-.13	.04	.25	.20	.42	.35	.28	.28	1.00

Table 5.5

Correlation between the Daily Sums of Sunshine Duration

December 1988

German Radiometric Network

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	1.00	.25	-.14	-.07	-.16	.09	.13	-.08	.05	-.05	-.02	.53	-.04	.22	-.07	-.03	.29	-.04	.37	.01	.54	.05	.00	-.03	-.02	.08	.09	-.09	.07	.03	.34
2	.25	1.00	-.00	-.02	-.10	-.10	-.04	-.00	.01	-.02	.03	-.05	-.06	.22	.08	-.05	.06	-.02	.15	.03	.10	-.05	.07	-.07	.15	.14	-.07	-.07	.06	-.02	.06
3	-.14	-.00	1.00	-.05	.02	.13	-.11	.24	-.15	-.01	.09	-.14	.33	-.08	-.27	.10	.08	.09	-.02	-.23	-.10	.16	.24	-.16	.10	.11	.17	.37	-.07	.14	.15
4	-.07	-.02	-.05	1.00	.23	.00	.24	-.08	.13	-.02	-.03	.10	-.06	.09	.17	-.16	.15	-.04	.05	.25	-.06	-.07	-.07	.23	-.07	-.06	-.02	-.11	-.10	-.09	.07
5	-.16	-.10	.02	.23	1.00	.15	.36	-.19	.21	.22	-.08	.09	-.11	-.08	-.01	-.22	.20	.06	.02	.16	-.22	-.25	-.12	.37	-.22	-.13	-.03	-.23	-.22	-.28	-.15
6	.09	-.10	.13	.00	.15	1.00	.08	.15	-.22	.08	.00	.14	.15	-.15	-.12	.19	-.01	.18	-.02	-.09	.12	.12	.09	-.08	.08	-.10	-.03	.14	-.27	-.18	-.12
7	.13	-.04	-.11	.24	.36	.08	1.00	-.35	.54	.02	-.13	.44	-.11	.19	.30	-.34	.40	-.13	.44	.63	.09	-.24	-.14	.52	-.28	-.31	-.02	-.28	-.19	-.32	.01
8	-.08	-.00	.24	-.08	-.19	.15	-.35	1.00	-.30	-.05	.04	-.28	.39	.08	-.21	.30	-.08	.13	-.17	-.28	-.04	.20	.28	-.21	.43	.08	.07	.40	-.03	.17	.04
9	.05	.01	-.15	.13	.21	-.22	.54	-.30	1.00	.08	-.12	.28	-.24	.36	.11	-.65	.38	-.09	.49	.60	.11	-.30	-.30	.43	-.30	-.24	-.01	-.44	-.13	-.36	.01
10	-.05	-.02	-.01	-.02	.22	.08	.02	-.05	.08	1.00	-.02	-.01	.07	-.03	-.10	-.11	.15	-.04	.19	-.07	-.02	-.02	-.05	.26	-.05	.01	.20	-.05	-.08	-.05	.01
11	-.02	.03	.09	-.03	-.08	.00	-.13	.04	-.12	-.02	1.00	-.09	.30	.09	-.09	.19	-.06	.18	-.06	-.11	.22	.37	.43	-.05	.18	.43	.26	.25	.21	.17	.22
12	.53	-.05	-.14	.10	.09	.14	.44	-.28	.28	-.01	-.09	1.00	-.11	.21	.16	-.20	.21	-.13	.48	.29	.42	-.18	-.18	.02	-.25	-.22	-.08	-.30	-.24	-.36	.08
13	-.04	-.06	.33	-.06	-.11	.15	-.11	.39	-.24	.07	.30	-.11	1.00	-.02	-.31	.23	.13	.30	.01	-.25	.08	.22	.25	-.07	.52	.17	.42	.52	-.05	.17	.16
14	.22	.22	-.08	.09	-.08	-.15	.19	.08	.36	-.03	.09	.21	-.02	1.00	.24	-.18	.16	-.04	.42	.35	.18	-.02	-.02	.13	.12	-.04	-.04	-.10	-.14	-.12	.22
15	-.07	.08	-.27	.17	-.01	-.12	.30	-.21	.11	-.10	-.09	.16	-.31	.24	1.00	-.01	-.21	-.22	.01	.45	-.01	-.32	-.20	.09	-.19	-.26	-.29	-.38	-.05	-.16	-.24
16	-.03	-.05	.10	-.16	-.22	.19	-.34	.30	-.65	-.11	.19	-.20	.23	-.18	-.01	1.00	-.32	-.02	-.37	-.37	-.05	.33	.28	-.37	.32	.28	.09	.43	.26	.40	.07
17	.29	.06	.08	.15	.20	-.01	.40	-.08	.38	.15	-.06	.21	.13	.16	-.21	-.32	1.00	.05	.54	.12	.34	.05	-.07	.47	-.01	.03	.32	-.06	-.25	-.13	.38
18	-.04	-.02	.09	-.04	.06	.18	-.13	.13	-.09	-.04	.18	-.13	.30	-.04	-.22	-.02	.05	1.00	-.07	-.19	.04	.25	.21	-.00	.26	.24	.18	.31	.06	.12	.14
19	.37	.15	-.02	.05	.02	-.02	.44	-.17	.49	.19	-.06	.48	.01	.42	.01	-.37	.54	-.07	1.00	.27	.49	-.06	-.12	.35	-.14	-.06	.10	-.18	-.23	-.19	.33
20	.01	.03	-.23	.25	.16	-.09	.63	-.28	.60	-.07	-.11	.29	-.25	.35	.45	-.37	.12	-.19	.27	1.00	-.04	-.34	-.26	.35	-.29	-.33	-.22	-.40	-.14	-.33	-.11
21	.54	.10	-.10	-.06	-.22	.12	.09	-.04	.11	-.02	.22	.42	.08	.18	-.01	-.05	.34	.04	.49	-.04	1.00	.21	.01	.14	-.04	.11	.23	-.02	-.03	.01	.44
22	.05	-.05	.16	-.07	-.25	.12	-.24	.20	-.30	-.02	.37	-.18	.22	-.02	-.32	.33	.05	.25	-.06	-.34	.21	1.00	.22	-.01	.41	.46	.33	.61	.20	.45	.54
23	.00	.07	.24	-.07	-.12	.09	-.14	.28	-.30	-.05	.43	-.18	.25	-.02	-.20	.28	-.07	.21	-.12	-.26	.01	.22	1.00	-.16	.23	.20	.06	.41	.09	.25	.11
24	-.03	-.07	-.16	.23	.37	-.08	.52	-.21	.43	.26	-.05	.02	-.07	.13	.09	-.37	.47	-.00	.35	.35	.14	-.01	-.16	1.00	-.13	.04	.31	-.20	-.12	-.13	.26
25	-.02	.15	.10	-.07	-.22	.08	-.28	.43	-.30	-.05	.18	-.25	.52	.12	-.19	.32	-.01	.26	-.14	-.29	-.04	.41	.23	-.13	1.00	.23	.25	.51	.08	.21	.09
26	.08	.14	.11	-.06	-.13	-.10	-.31	.08	-.24	.01	.43	-.22	.17	-.04	-.26	.28	.03	.24	-.06	-.33	.11	.46	.20	.04	.23	1.00	.45	.40	.40	.54	.56
27	.09	-.07	.17	-.02	-.03	-.03	-.02	.07	-.01	.20	.26	-.08	.42	-.04	-.29	.09	.32	.18	.10	-.22	.23	.33	.06	.31	.25	.45	1.00	.37	.12	.28	.52
28	-.09	-.07	.37	-.11	-.23	.14	-.28	.40	-.44	-.05	.25	-.30	.52	-.10	-.38	.43	-.06	.31	-.18	-.40	-.02	.61	.41	-.20	.51	.40	.37	1.00	.13	.53	.41
29	.07	.06	-.07	-.10	-.22	-.27	-.19	-.03	-.13	-.08	.21	-.24	-.05	-.14	-.05	.26	-.25	.06	-.23	-.14	-.03	.20	.09	-.12	.08	.40	.12	.13	1.00	.49	.10
30	.03	-.02	.14	-.09	-.28	-.18	-.32	.17	-.36	-.05	.17	-.36	.17	-.12	-.16	.40	-.13	.12	-.19	-.33	.01	.45	.25	-.13	.21	.54	.28	.53	.49	1.00	.46
31	.34	.06	.15	.02	-.15	-.12	.01	.04	.01	.01	.22	.08	.16	.22	-.24	.07	.38	.14	.33	-.11	.44	.54	.11	.26	.09	.56	.52	.41	.10	.46	1.00

5.3 Cokriging

Study prepared by:	*Vito D'Agostino and Antoine Zelenka* *SASIAM, Tecnopolis CSATA Novus Ortus and* *Swiss Meteorological Institute repectively*
Under sponsorship from:	*Italian Space Agency and Swiss Federal Office of Energy respectively*

5.3.1 Introduction

From Chapters 3 and 4 we have determined that there seems to be a limit to the precision of predicted yearly sequences of daily sums of insolation. A minimum RMSE% of 17% is obtained regardless of the evaluation method used, i.e. satellite-based or network sustained interpolation.

Interpolation benefits from the relative accuracy of pyranometric measurements, but it lacks the spatial resolution of satellite-based estimates. Satellite-based estimates, however, can produce subtstantial biases under unfavorable conditions (e.g., presence of snow, discontinuous sampling in time, and inhomogeneous albedo or rugged terrain (large relief) within pixels). Yet, despite its sensitivity to systematic errors, the information gained from satellite images can by no means be neglected, because even an interpolation method as sophisticated as kriging is unable to resolve spatial features with dimensions less than the smallest distance between two stations in the network.

One possible way to combine the advantages of both techniques is to apply a uniform correction to the satellite estimates corresponding to the correlation between G(measured) and G(sat). This may help remove biases, but the scatter of the resulting estimates will essentially remain that of the correlation, i.e., about 17% in the favorable cases.

As suggested by the pioneering work of Creutin (1987) [5.3.1] and of Galle (1987) [5.3.2], cokriging is another way to interpolate within the network, while taking advantage of the full structural information of the satellite estimates, and to simultaneously *reduce the prediction's RMSE*. In the simplest case, cokriging (coK in the following) reduces to kriging the residuals between measurements and satellite estimates. Application of the location-dependent correction forces the estimates to fit the measured values at the network sites while preserving any spatial structure that may be present, as illustrated in Fig. 5.3.1.

In this study, coK is performed with the PC-based software *GEOPACK* of Yates and Yates (1989) [5.3.3]. The limited capacities of even a powerful PC do not yield conclusions about the ultimate precision achievable by the method, but they certainly do provide conclusions as to its adequacy. Originally, this study was being considered for follow-on work to Subtask 9D (D'Agostino and Zelenka (1990) [5.3.4]). Therefore, the data are not from the 9D-set. Nevertheless, they are not really different as they still originate from the Swiss ANETZ network and from METEOSAT according to the Cologne method, i.e., from sources which were duly presented in the foregoing chapters. Only the period, which is April 1982, and some pre-treatment of the data differ. As will be seen, results for mere kriging are very similar between April 1988 and April 1982. Furthermore, April validations happen to be representative of validations for whole year sequences (see, e.g., The ANETZ Tables in Volume 3), a fact which confers the required generality to the results to be reported.

After a short presentation of the peculiarities of the data, the essentials of the cokriging method are exposed and results are presented and discussed, with the conclusion that coK helps reduce the prediction RMSE% substantially below 17%.

5.3.2 Data

In April 1982, 55 ANETZ stations were in operation. From these, 46 were retained for this investigation. The 5 sites added to the 41 sites of the 9D set have slightly more severe local

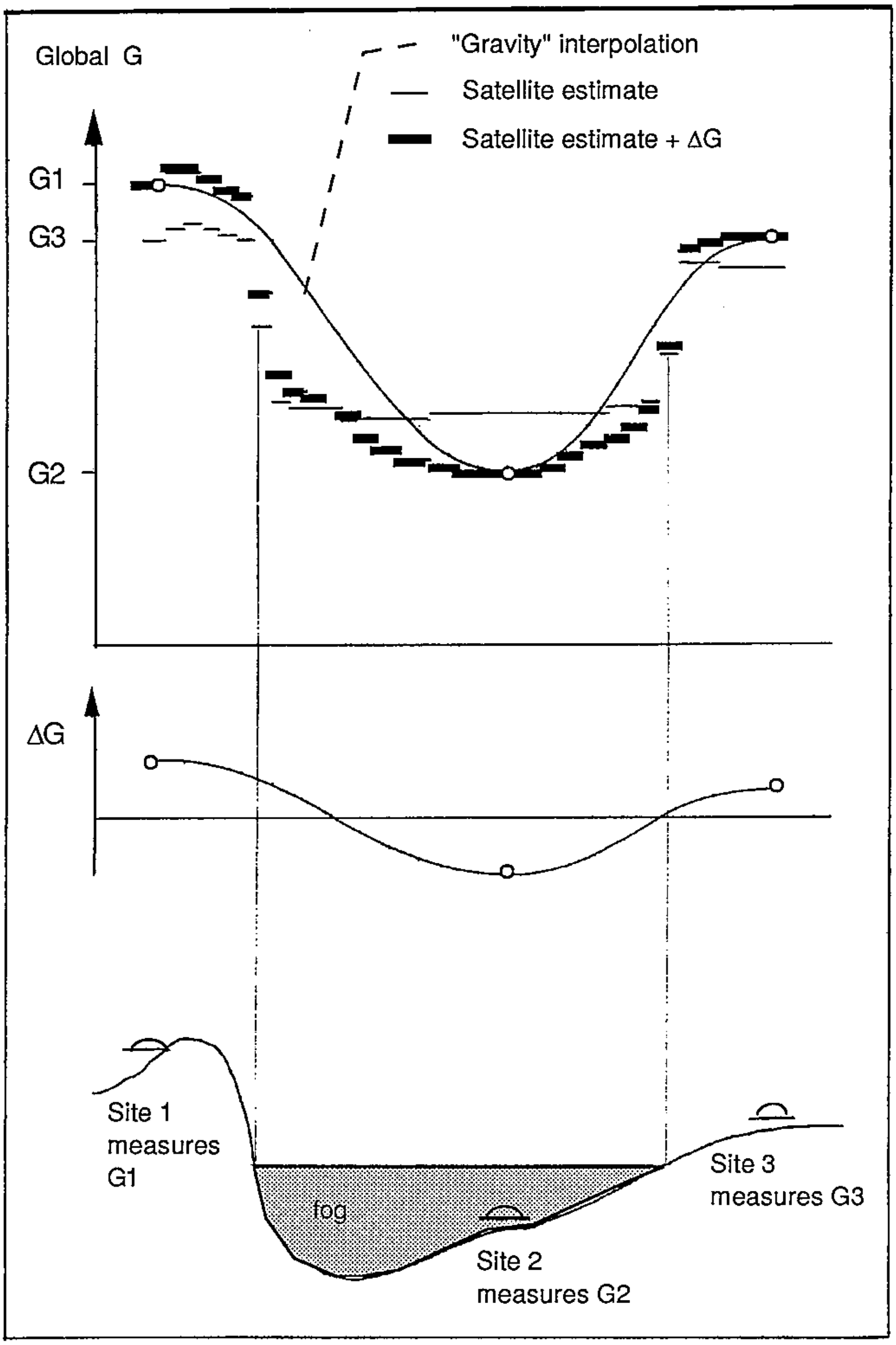


Fig. 5.3.1: Mere interpolation between sites 1, 2 and 3 misses the sharp transitions at the fog boundaries. Satellite estimates render them correctly in location but not necessarily in absolute value. Application of an interpolated, i.e., location-dependent correction ΔG = G(meas)-G(sat) leads to predictions which reproduce the measurements without real loss of spatial resolution. The discrete, discontinuous shape of the satellite estimates is intended to mimic the pixel size of the satellite image.

horizon restrictions. A 5% loss of the daily total due to horizon restrictions is now tolerated, because this investigation is closer to enventual applications, whereas only a 2% loss was tolerated in the 9D set (see Section 1.1.2). Within this subset of 46 sites, horizontal distances between neighboring stations range from 3.4 km to 44 km, while altitudes range from 192 m to 3315 m.

The satellite irradiation totals have been calculated from three METEOSAT 2 images (visible and infrared) collected each day by Möser (1983) [5.3.5] with the physical model of Möser and Raschke (1984) [5.3.6]. Although the full spatial resolution of the visible (reference) channel is 2.4 km x 4.2 km (E-W x S-N), some averaging is compulsory to mitigate the influence of erroneous pixel values and positioning inaccuracies. Therefore, Möser's estimates are for blocks of 6 x 8 pixels, corresponding to areas of 19.2 km x 25.2 km (see Fig. 4.3 in Chapter 4). Estimates for any given location, especially for each network station but also for each full resolution pixel, are derived from these blocks by inverse distance averaging with a search radius corresponding to the dimension of 2.4 blocks (Möser, 1983 [5.3.5]). Finally, the values obtained are corrected for the site elevation effect as described in Section 4.2.3, the altitude being the average of all the pixels in a block.

In this way we obtained 6628 values for the structural analysis of the satellite information at full resolution. This is, by far, more than *GEOPACK* can manage. Therefore, we returned to the scale of Möser's original blocks by averaging over every 8 x 6 (altitude corrected) pixels, while keeping in mind that this procedure probably leads to a loss of precision when the variance of the blocks is not constant. The locations of the averaged estimates depicted in Fig.5.3.2 are referenced to the centers of the blocks. The same averaging is performed around each pixel containing each of the network's stations. Thus, for each of the 46 observing sites depicted in Fig.5.3.2, both surface

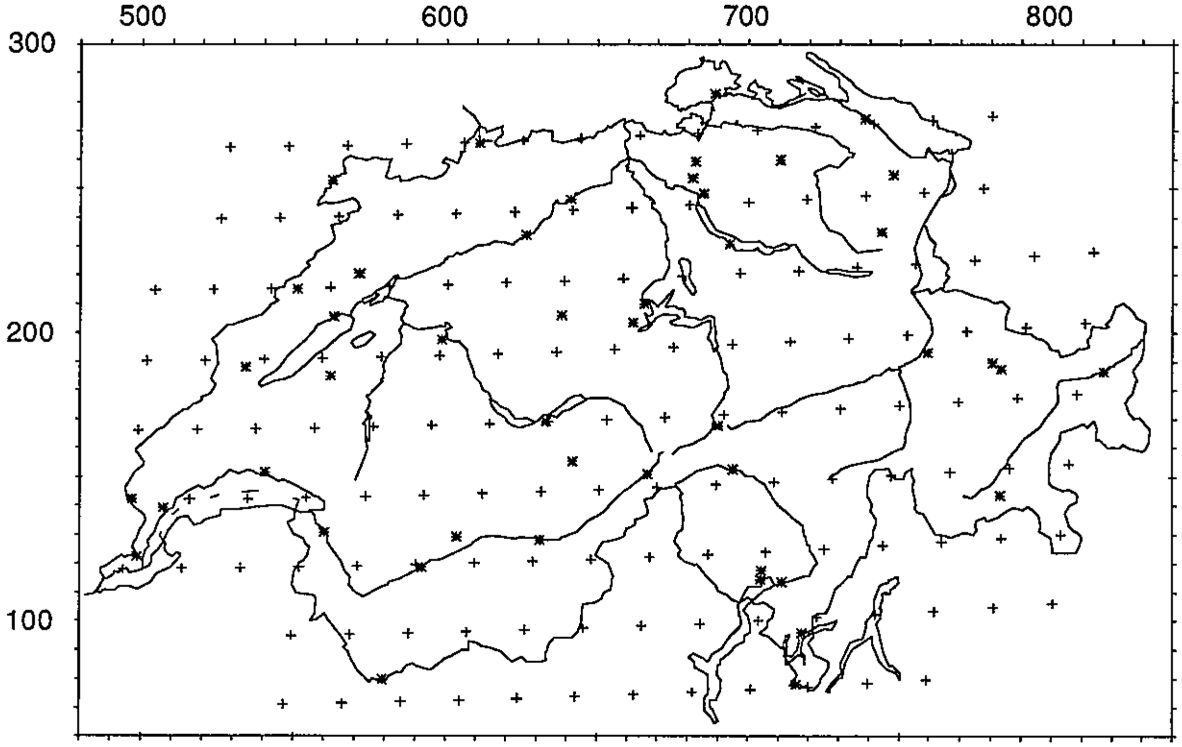


Fig. 5.3.2: Location of the 46 network stations (stars) and of the averaged satellite estimates (crosses).

measurements and satellite estimates are available for the determination of the experimental cross-variograms.

5.3.3 The cokriging method

Cokriging is the extension of kriging to the situation where variables are spatially interdependent and the one of particular interest is undersampled. We recall that the kriging philosophy is to locally weight the data such that it provides estimates without bias and with minimum variance. Theoretically, the cokriging variance decreases with respect to the kriging variance, improving the quality of the estimate (Journel and Huijbregts (1978) [5.3.7]). This is accomplished by transferring the spatial information of the frequently sampled auxiliary variable to the primary variable through the cross-variogram. This requires a significant number of locations where both functions are sampled. Matheron (1971) [5.3.8] discusses the theory of coregionalization. We develop here only some aspects of the equations and we explain the adopted simplifications.

Let $Z_1(x)$ be the function to be estimated, where x represents the spatial coordinates in the plane. We assume that $Z_1(x)$ is known at $(x_{11},...,x_{1n})$ irregularly distributed points. Let $Z_2(x)$ be the complementary function, known at m regular blocks, each of them identified by the coordinates $(x_{21},...,x_{2n})$ of a regular grid. Finally, assume that the first set is contained in the second one and, as a *major simplification,* ignore that Z_1 and Z_2 have different supports.

Cokriging provides an estimator Z_1^* of Z_1 from a linear combination of the available sample values, making use of an estimated spatial correlation between the property of interest Z_1 and the complemetary one, Z_2 :

$$Z_1^* = \sum_{i=1}^{n} \beta_i Z_{1i} + \sum_{j=1}^{m} \alpha_j Z_{2j} \qquad (5.3.1)$$

where $Z_{1i} = Z_1(x_{1i})$ and $Z_{2j} = Z_2(x_{2j})$.

The weighting factors β_i and α_j are determined by requiring that Z_1^* meets the standard bias and optimality conditions and by assuming that Z_1 and Z_2 fulfill the intrinsic hypothesis. This latter hypothesis assumes first and second order stationarity for the increments of the functions and allows statistical inference of their two first moments (see also Section 3.4 in this Volume and Chapter 10 in Volume 2).

Then, the first moments of Z_1 and Z_2 are constants, at least locally, and the second moments depend only on the separation h between the sample points. We then call the (auto-) variogram the set of expectations E of the squared increments of Z_1 and Z_2, designate them with γ_1 and γ_2 respectively , and by analogy with the single variable case, we call cross-variogram $\gamma_{12}(h)$ the expectation of the incremental products between Z_1 and Z_2. Thus, for any pair of variables Z_1 and Z_2, γ at lag h is defined as:

$$\gamma_{12}(h) = (1/2)\ E[(Z_1(x)-Z_1(x+h))(Z_2(x)-Z_2(x+h))] \qquad (5.3.2)$$

If $Z_1(x) = Z_2(x)$ the above equation denotes the auto-variogram. Note that when the stationary hypothesis holds, then the cross-variogram may be written in terms of the cross-covariance function $C_{12}(h)$:

$$\gamma_{12}(h) = C_{12}(0) - (1/2)[C_{12}(h) + C_{21}(-h)] \qquad (5.3.3)$$

and this may be simplified if $C_{12}(h)$ is assumed to be symmetrical:

$$\gamma_{12}(h) = C_{12}(0) - C_{12}(h) \qquad (5.3.4)$$

Under these hypotheses, at each point x_o, the standard conditions imposed on $Z_1{}^*$ lead to the cokriging system (coK-system). To save space, the detailed structure of this system is given in Chapter 10 in Volume 3 (Theory). By solving this coK-system at every location x_o we obtain the estimate $Z_1{}^*(x_o)$ and the cokriging variance ($\sigma^2{}_{coK}$):

$$\sigma^2{}_{coK} = \mu_1 + \sum_{i=1}^{n} \beta_i \gamma_1(x_{1i} - x_o) + \sum_{j=1}^{m} \alpha_j \gamma_{12}(x_{2j} - x_o) \qquad (5.3.5)$$

where μ_1 is one of the Lagrange factors originating from the minimum variance condition.

If Z_2 is correlated to Z_1 then σ_{coK} is generally *less then the auto-kriging variance* (Eq. 3.4.12) because the additional term involving γ_{12} allows more flexibility in the minimizing process.

Fixing of the coefficients in the rows and columns of the coK-system (essentially $\gamma_1(\Delta x)$, $\gamma_2(\Delta x)$ and $\gamma_{12}(\Delta x)$) requires valid models of the variograms which fit the data. For an interval around a given distance h, and, if appropriate, for each direction class, an experimental cross-variogram $\gamma_{12}{}^*$ is estimated from the sample data as the ordered set of values $\{h,\gamma^*(h)\}$ (which can be presented as a graph of γ^* versus h) using the equation:

$$\gamma_{12}{}^*(h) = \{ \sum_{k=1}^{n(h)} [Z_1(x_k+h)-Z_1(x_k)][Z_2(x_k+h)-Z_2(x_k)]\}/[2n(h)] \qquad (5.3.6)$$

with n(h) being the number of pairs having a separation within the given interval. Again if $Z_1 = Z_2$ for all x, then $\gamma_{12}{}^*$ becomes the experimental auto-variogram $\gamma_1{}^*$ (or $\gamma_2{}^*$).

The cross-variogram $\gamma_{12}{}^*(h)$ follows from those locations where values of both Z_1 and Z_2 are sampled. A valid model (e.g. exponential, gaussian, spherical, linear etc.) must be fitted to it either visually, by means of a least squares fit (as in GEOPACK) or by some other valid procedure, and this model can be validated by the jacknife technique.

For the auto-variograms γ_1 and γ_2 the allowed theoretical models ensure certain conditions of mathematical consistency, without which one might come up with negative variances. Similary, for a cross-variogram, the Cauchy-Schwarz relation

$$| \gamma_{12}(h) | \leq \sqrt{[\gamma_1(h)\gamma_2(h)]} \qquad \text{for all } h > 0 \qquad (5.3.7)$$

must be fulfilled to ensure a positive cokriging variance in all circumstances. *GEOPACK* tests the Cauchy-Schwarz inequality using a discrete method. It calculates the inequality at about 50 points between zero lag and the range, and if the inequality fails, a warning message is issued. The number of points from which an experimental variogram can be reliably computed limits the investigative power and, sometimes, the credibility of geostatistical studies. To avoid discarding applications on the sole criterion of the sample size, several empirical rules exist for the reliability assessment of variograms. In particular, the weights assigned to less representative points $\gamma_i{}^*$, while fitting a model $\gamma(h)$, are strongly dependent upon the knowledge one has about the physical properties of the phenomenon. A further empirical procedure determines if a modelled γ(h) yields a better cross-validation than a pure nugget (γ(h)=constant) model. As this has been demonstrated by Jaquet (1991) [5.3.9] for the Swiss network, we needn't worry about the modest size of our samples. Our only limitation is that we must assume isotropy in order to involve all points at one time in the structural analysis. Valuable information contained in the (satellite) data is perhaps lost but, as already stated in the introduction, this is not critical to this application.

5.3.4 Results

To estimate daily sums of solar radiation through cokriging, we modelled day by day (from 1 to 30 April) the experimental variogram for the ground data (the primary variable), the Möser-Raschke estimates corrected for site altitude (the auxiliary variable), and the experimental cross-variogram between them. For the variograms involving the ground data we adopted a lag of

TABLE I: Crossed variograms between ground based measurements and METEOSAT-based estimates corrected by altitude. Days are referred to April1982 (G=Gaussian, S=Spherical, E=Exponential, L=Linear model).

Days	Model	Nugget	Sill-Nugget	Range
		$[MJm^{-2}day^{-1}]^2$		[km]
1	G	.0	-2	80.5
2	G	1.0	375.3	918.2
5	S	.5	14.2	154.4
6	E	.1	20.5	184.2
7	G	2.0	14.1	169.7
8	G	.9	27.2	98.8
9	E	2.3	29.2	75.6
10	G	.8	15.3	101.3
11	G	.1	27.3	98.9
12	G	.1	37.2	81.5
13	G	1.1	28.3	767.5
14	L	.1	11.8	142.4
15	G	.2	23.9	96.3
16	G	.1	46.6	90.6
17	E	.1	4.9	23.0
18	E	.0	3.3	29.5
19	S	.0	10.0	155.9
20	G	.1	27.8	286.8
21	G	.2	12.2	85.7
22	G	.5	17.3	108.6
23	E	.0	3.7	31.6
24	E	.1	10.9	164.3
25	G	.0	8.9	80.8
26	G	1.3	13.4	82.5
27	G	.1	24.7	116.2
28	G	.3	17.9	81.6
29	G	.0	18.3	94.3
30	G	.0	18.7	88.8

75 km, whereas the lag for the more dense satellite estimates was 65 km. The different types of experimental variograms encountered for the ground data have been discussed by Jaquet (1991) [5.3.9] and by Zelenka *et al.* (1989) [5.3.10]. On April 3 the METEOSAT images were missing, and April 4 did not show the expected correlation between the variables (failure of the satellite model). Thus, the cross-variogram modelling was performed for only 28 days. The results of the structural analyses are presented in Table I. The frequent occurrence of the Gaussian model confirms the continuous nature of the solar irradiation field. However, if used without a nugget effect, this Gaussian model leads to numerical instabilities in the coK-system. In these cases, the nugget value is chosen different from zero only to avoid the instabilities (e.g. Bardossy 1988 [5.3.11]).

The trend at infinity of the experimental variogram has always been treated with the linear variogram modelling. The following alternative steps were tested: (i) drift recognition, (ii) drift subtraction, and (iii) residual analysis; this approach did not yield significant differences from the linear variogram modelling.

Each model for each day has been evaluated by cross-validation. Those models which satisfied unbiased and consistent estimation were adopted, and the RMSE of the estimates with respect to the measurements were calculated. As an example, Figs. 5.3.3, 5.3.4 and 5.3.5 present the experimental and modelled auto- and cross-variograms of April18. In this case, we can observe the shape characteristics of the ground-based experimental variogram in the experimental cross-variogram. The satellite data diminish the sill (from 4.3 to 3.3 $[MJm^{-2}day^{-1}]^2$) and increase the range (from 26.9 to 29.5 km) of the cross-variogram model, thereby reducing the RMSE from 9.75% to 6.88% for a daily total measured mean of 22.7 $MJm^{-2}day^{-1}$.

Table II gives the monthly mean RMSE, as expressed in percent of the measured mean, together with the RMSE and Mean Bias Error (MBE) of the satellite and kriging estimates.

TABLE II: RMSE and MBE in % of the measured mean for the METEOSAT (MR), kriging (K) and cokriging (co-K) estimates. Subset of 28 days in April1982.

	MR	K	co-K
No. of days	1287	1287	1287
Meas.mean $[MJm^{-2}d^{-1}]$	18.8	18.8	18.8
MBE %	-1.6	0.1	-0.2
RMSE %	16.9	16.0	13.7

In Figs. 5.3.6 and 5.3.7, the RMSE% of kriging (K) and cokriging (co-K) cross-validations are compared to the RMSE% of the Möser-Raschke (MR) method. On 8 days the co-K was worse than the MR method (April 5, 6, 7, 16, 21, 23, 26 and 28). Finally, Fig. 5.3.8 presents the comparison between co-K and K: the former performs worse than the latter only on April 1 and 30. As to April 1, this result can be understood as a failure of the MR model: the negative sill of the cross-variogram (Table I) displays a negative correlation between both sets of data. Then, obviously, the RMSE of co-K behaves better than the RMSE of the MR estimates, but both are necessarily worse than the RMSE of auto-Kriging. Consequently, days where the satellite model fails (April 1 and 4) must be considered as days with missing auxiliary information (April 3) and estimates must originate from Kriging with the network data alone. Validation results of these nearly routine operations are reported in Table III.

TABLE III: RMSE and MBE in % of the measured mean for the METEOSAT (MR), kriging (K) and cokriging (co-K) esimates. April 1982 without April 3.

	MR	K	co-K*
No. of days	1333	1333	1333
Meas.mean [$MJm^{-2}d^{-1}$]	18.94	18.94	18.94
MBE %	-1.8	0.1	-0.3
RMSE %	16.7	15.8	13.4

* Contains the K estimates for April 1 and 4.

Figures 5.3.9 and 5.3.10 show contouring maps for April 1 obtained from the co-K and K procedures. Despite the partial failure of the MR estimates, the better resolution of the satellite data results in enhanced meso-structures in the co-K map, especially in the Alpine regions where network stations are sparse. But from a qualitative point of view, both maps are very similar: the ill-conditioned covariable does not jeopardize the estimation procedure.

5.3.5 Conclusions

Multivariate geostatistics (or coregionalization) simultaneously use two ore more correlated random variables so that the spatial information contained in each of them is used in the estimation (cokriging) of one of them.

We have assessed the ability of cokriging to supplement network data based on a sample of insolation totals for 29 days from 46 sites in Switzerland. Use of the network measurements as the principal variable and estimates derived from satellite measurements as the auxiliary variable permits more accurate prediction (13.7% RMSE) than the network (Kriging, 16.0% RMSE) or satellite model (Möser and Raschke, 16.9% RMSE) by themselves.

Improvement of prediction precision in terms of RMSE would not be achieved by application of a uniform correction factor (whose main function is to remove a possible bias). The regionalized correction inherent to cokriging degenerates to the uniform factor correction only when both variables are intrinsically correlated, that is, when their auto- and cross-variograms are proportional to the same model $\gamma(h)$. Several of the cases mentioned in the foregoing discussion demonstrate that this is not systematically fulfilled in the present context. The prediction RMSE is improved by at least 14%, although only the essential part of the available satellite information could be used because our software runs on a PC and because the model failed on 3 days. Despite the obvious computational shortcoming of the present study, we conclude that cokriging is a promising tool for global insolation estimation since it reproduces this insolation without bias and with close to mesoscale resolution.

Finally, although already implicitly explained in Fig. 5.3.2, it should be stressed that the satellite information must not be applied as a "continuous" image, i.e., in the form of estimates for contiguous pixels. For example, the GOES-based estimates for a 1°x1° grid available at NOAA/NESDIS can be just as useful as "continuous" estimates when the insolation field does not behave too erratically.

Acknowledgemnts

The PC-based software package *GEOPACK* was kindly supplied by its authors S.R. and V.M. Yates. We thank Marcello Raspone for his assistance in using *GEOPACK*. All reported opinions, conclusions and recommendations are ours and not those of the Italian Space Agency.

5.3.6 *References*

[5.3.1] Creutin J.D., 1987, "Validation et étalonnage d'images de télédétection à l'aide de mesures ponctuelles et selon une approche géostatistique: application à la mesure des précipitations et de l'insolation". Thèse Sciences- Physiques, Université J. Fourier et Institut National Polytechnique de Grenoble, 245 pp.

[5.3.2] Galle S., 1987, "Mesures journalières de l'insolation en France à l'aide d'un système de télédétection et d'un réseau conventionnel: étude statistique de la cohérence et de la combinaison de ces mesures". Thèse Sciences-Physiques, Université J. Fourier, Grenoble, 257 pp.

[5.3.3] Yates S.R. and Yates M.V., 1989, "Geostatistics for Waste Management: A User's Manual for the *GEOPACK* (Version 1.0)". Internal Report. Dept. of Soil & Environmental Sciences. University of California. Riverside CA.

[5.3.4] D'Agostino and A. Zelenka A., 1992, "Estimating Solar Global Irradiance by Integration of Satellite and Network Data: The Cokriging Approach". *2nd CODATA Conference on Geomathematics and Geostatistics*, The University of Leeds, September 1990. *Sci. de la Terre, Sér. Inf.*, **31**, 461.

[5.3.5] Möser W., 1983, "Globalstrahlung aus Satellitenmessungen". *Mitt. Inst. Geophys. Meteorol. Universität zu Köln*, Heft **37**, and private communication.

[5.3.6] Möser W. and Raschke E., 1984, "Incident Solar Radiation over Europe Estimated from METEOSAT Data". *Jour. Climate & Appl. Meteorol.* **23**, 166-170.

[5.3.7] Journel A. and Huijbregts C., 1978, "Mining Geostatistics". Academic Press, London.

[5.3.8] Matheron G., 1971, "The Theory of Regionalized Variables and its Applications". *Les Cahiers du Centre de Morphologie Mathematique de Fontainebleau* **5**, Ecoles des Mines de Paris, 211 pp.

[5.3.9] Jaquet, O., 1991, "Étude Géostatistique de l'Insolation Énérgétique Mesurée par le Réseau Météorologique Suisse". *Sci. de la Terre, Série Inf.* **30**, 57-70.

[5.3.10] Zelenka A., Lazic D. and Jaquet O., 1989, "Solar Resource Assessment with Geostatistical Tools". *Clean and Safe Energy Forever*, Proc. 1989 Biennal Congres ISES, Kobe, Vol. **3**, pp 2077-2081, Pergamon 1990.

[5.3.11] Bardossy A., 1988,'Notes on the Robustness of the Kriging System". *Mathematical Geology* **20**, 189-203.

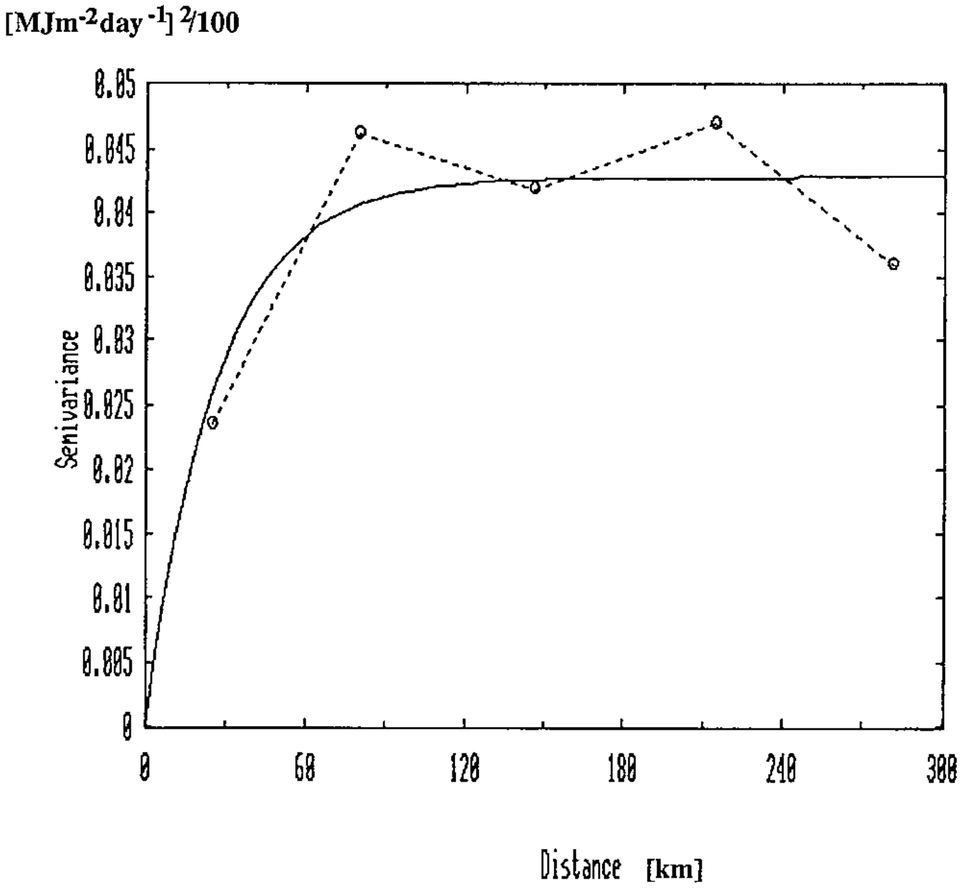


Fig. 5.3.3: Exponential model (full line) fitted to the experimental variogram (dashed line) derived from ***ground-data*** of April 18. The model parameters are: Sill = 4.3 [MJm^{-2}day^{-1}]^2; Range = 26.9 km.

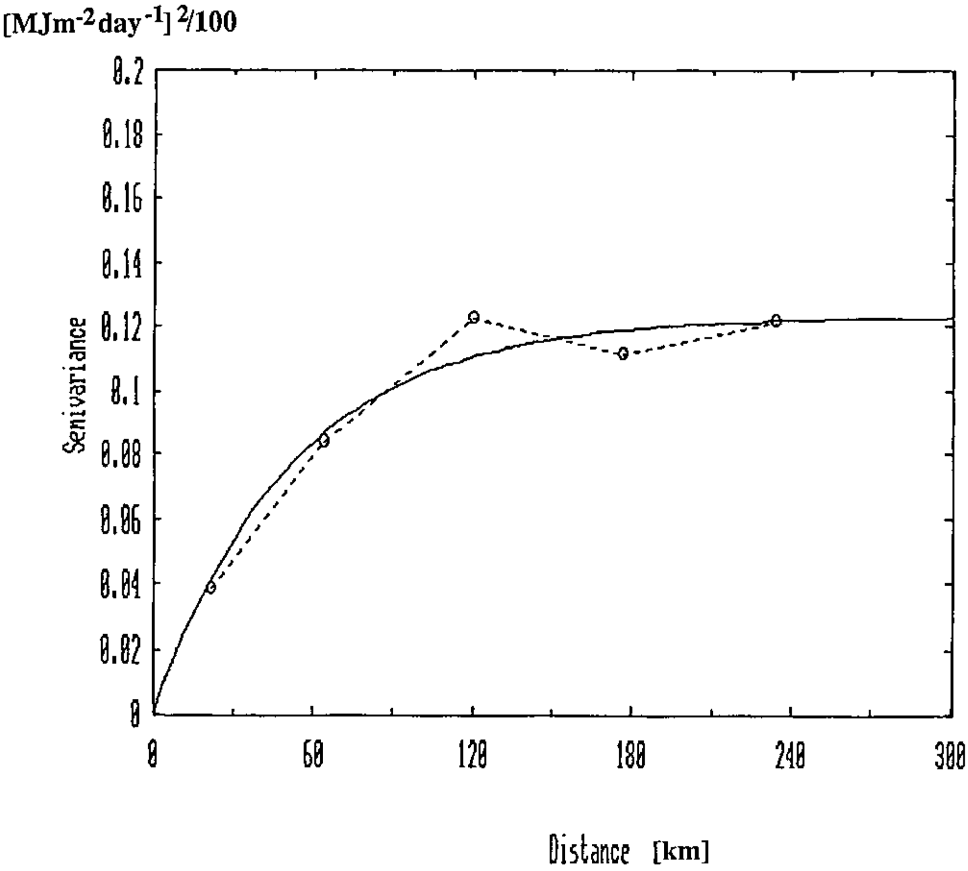


Fig. 5.3.4: Exponential model (full line) fitted to the experimental variogram (dashed line) derived from ***METEOSAT 2-data*** of April 18. The model parameters are:
Sill = 12.3 [MJm^{-2}day^{-1}]^2; Range = 52.7 km.

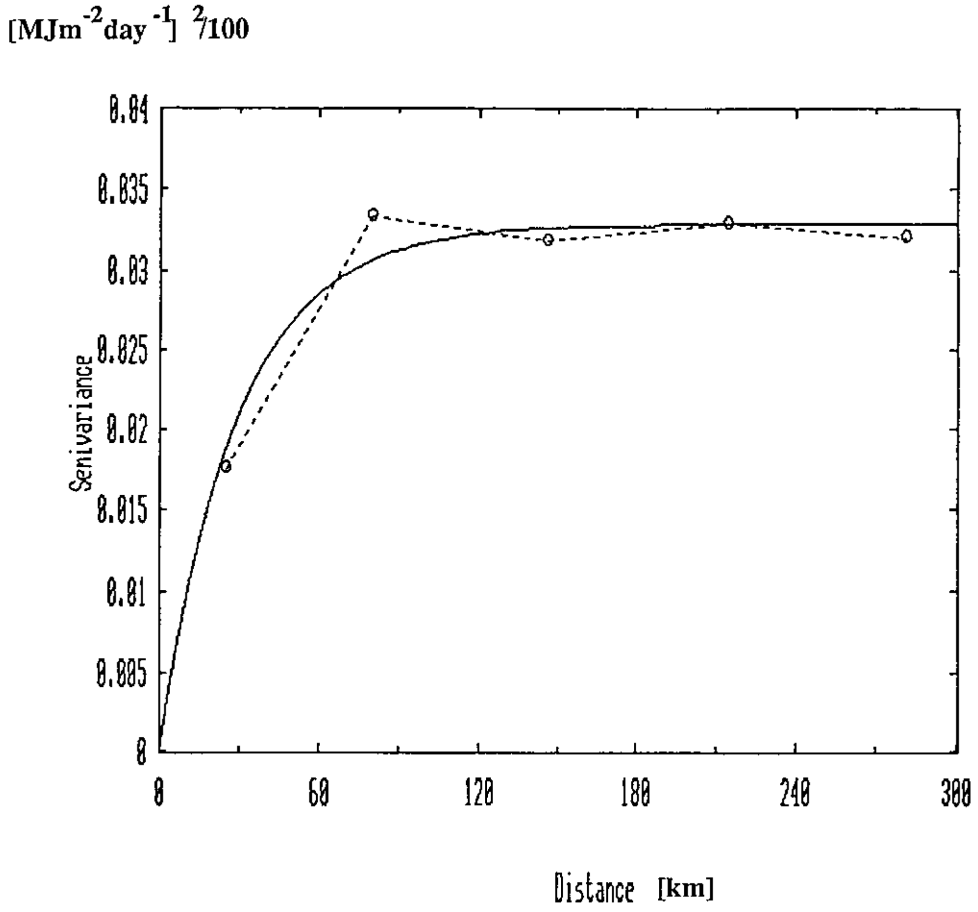


Fig. 5.3.5: Exponential model (full line) fitted to the experimental *cross-variogram* (dashed line) derived from ***ground- and METEOSAT 2-data*** of April 18. The model parameters are:
Sill = 3.3 $[MJm^{-2}day^{-1}]^2$; Range = 29.5 km.

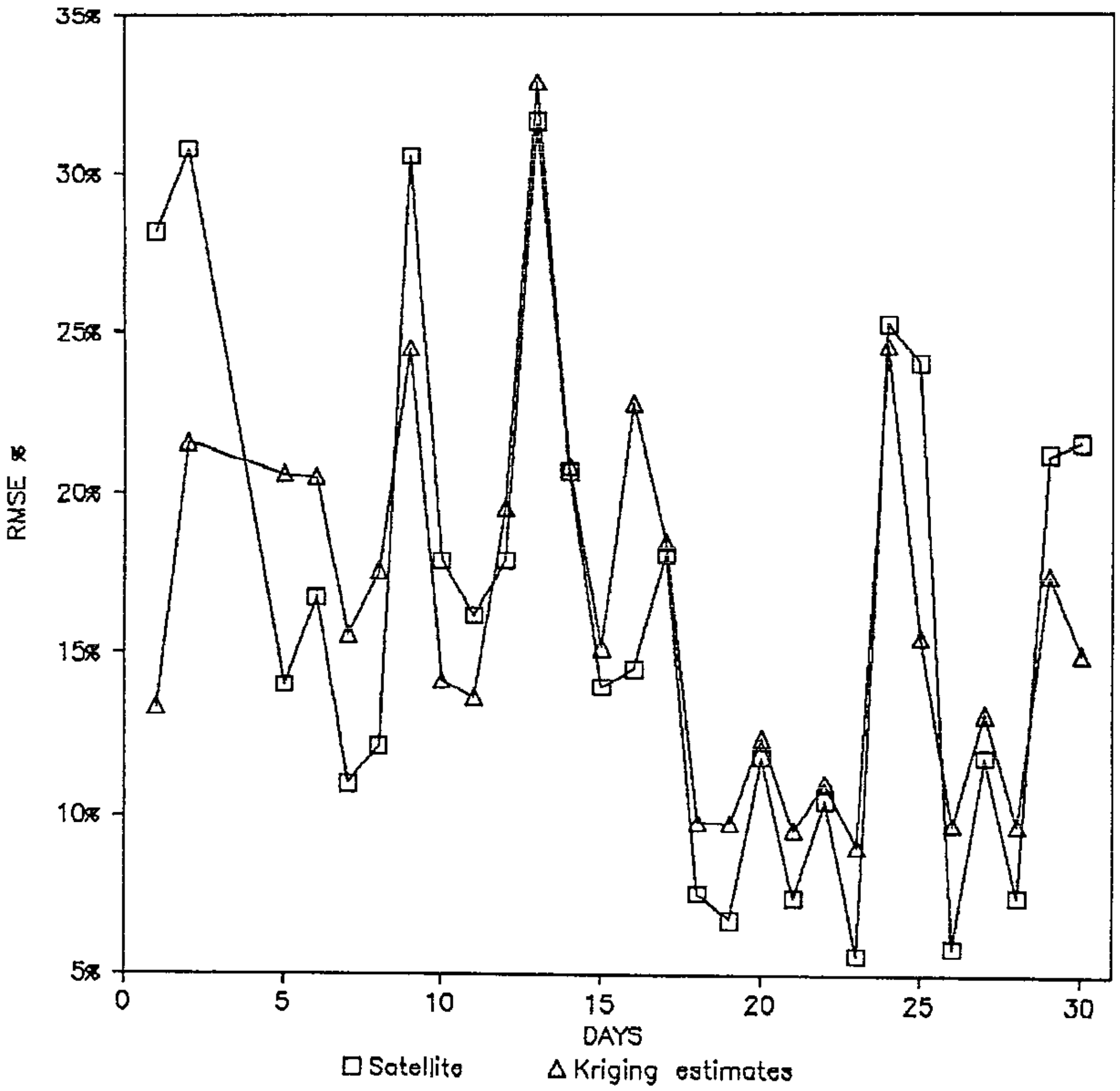


Fig. 5.3.6: RMSE% of Möser-Raschke and of kriging estimates for each day of April 1982.

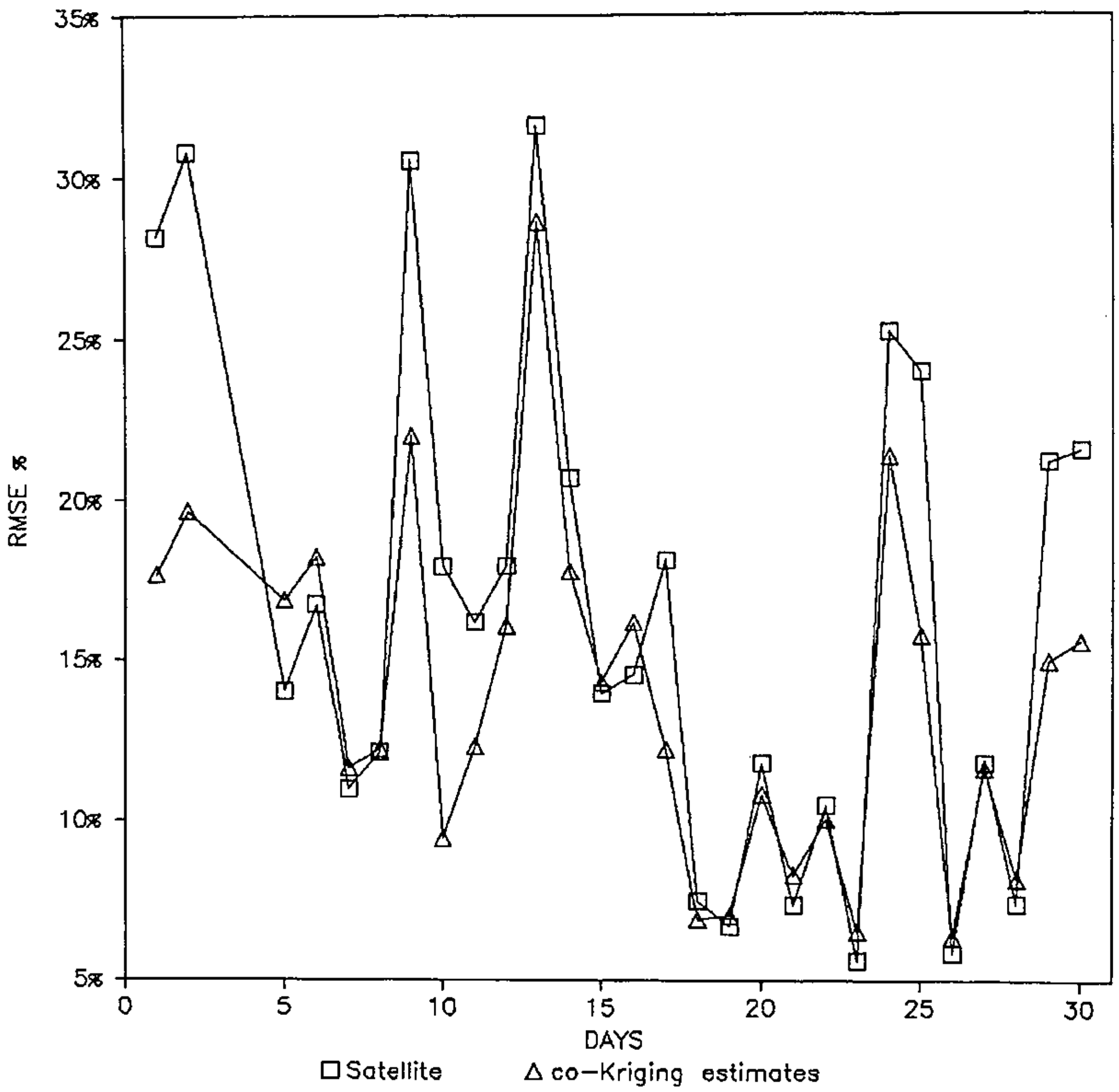


Fig. 5.3.7: RMSE% of Möser-Raschke and of cokriging estimates for each day of April 1982.

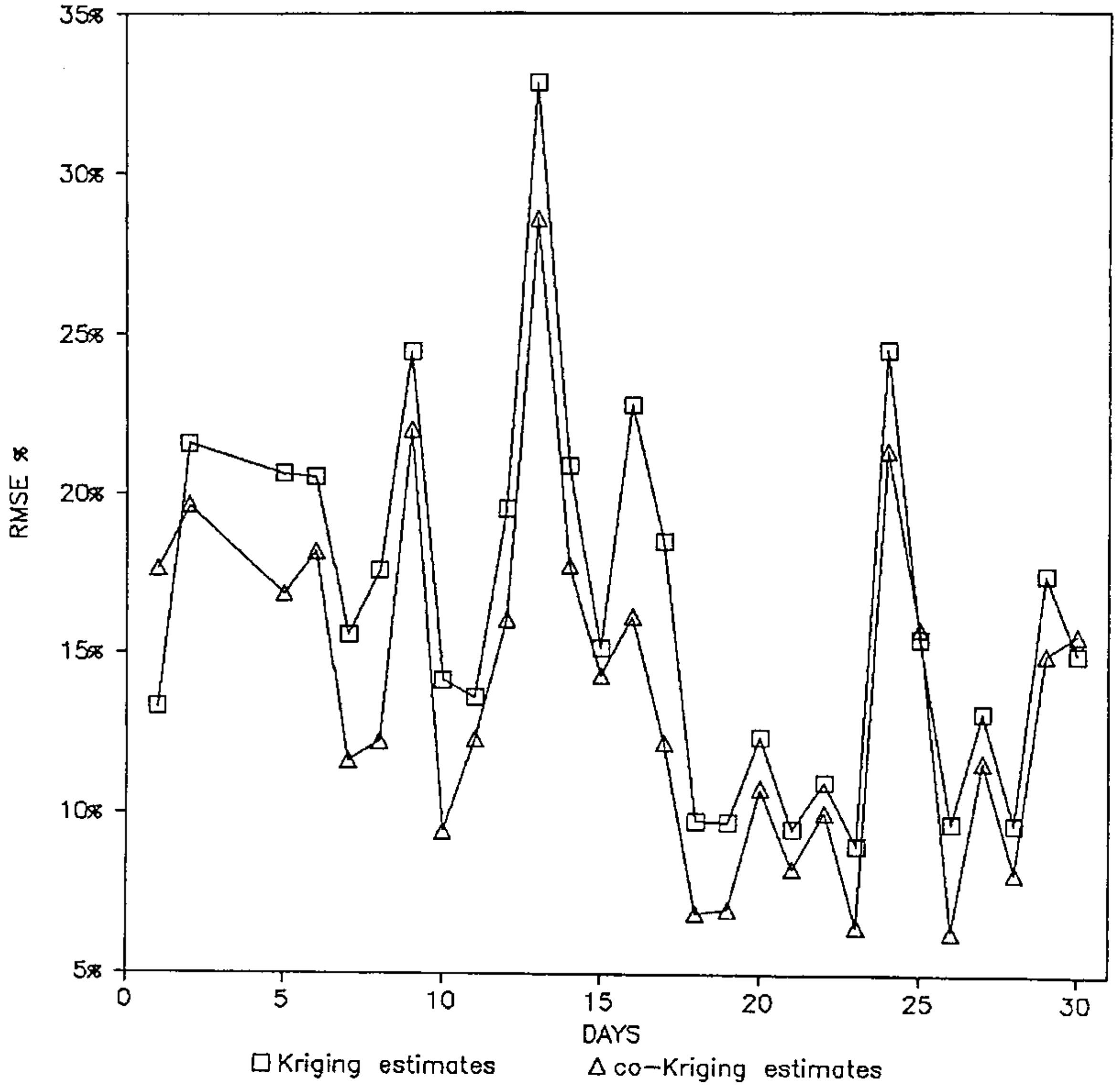
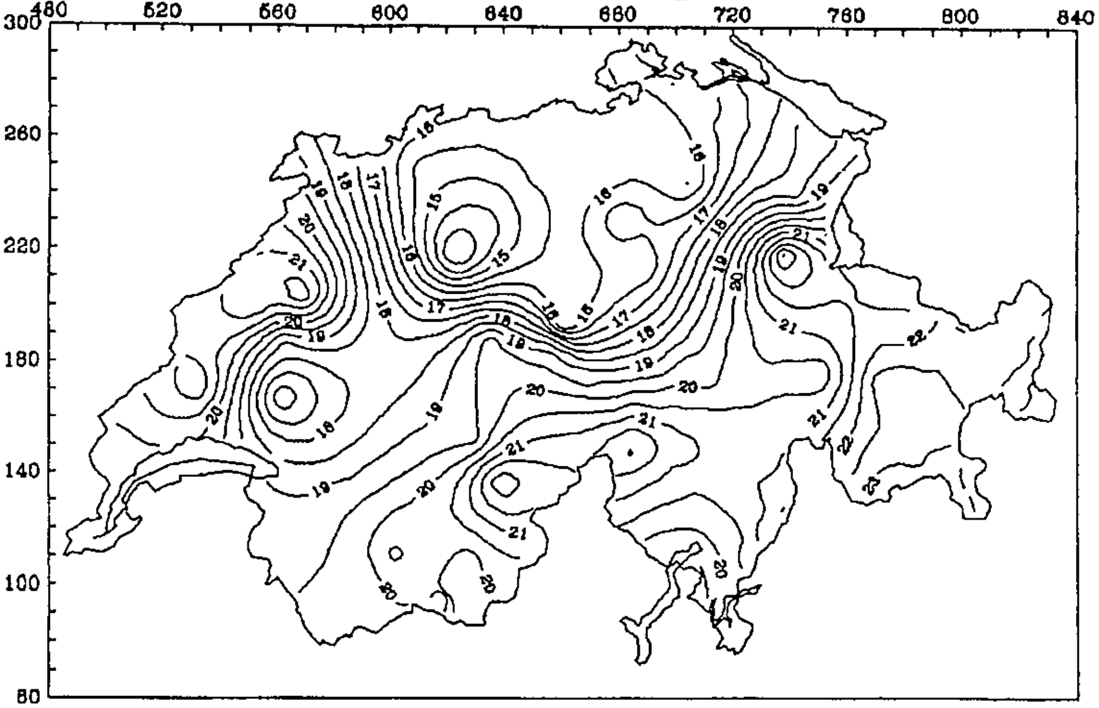


Fig. 5.3.8: RMSE% of kriging and of cokriging estimates for each day of April 1982.

Kriging estimates April 1, 1982



co–Kriging estimates April 1, 1982

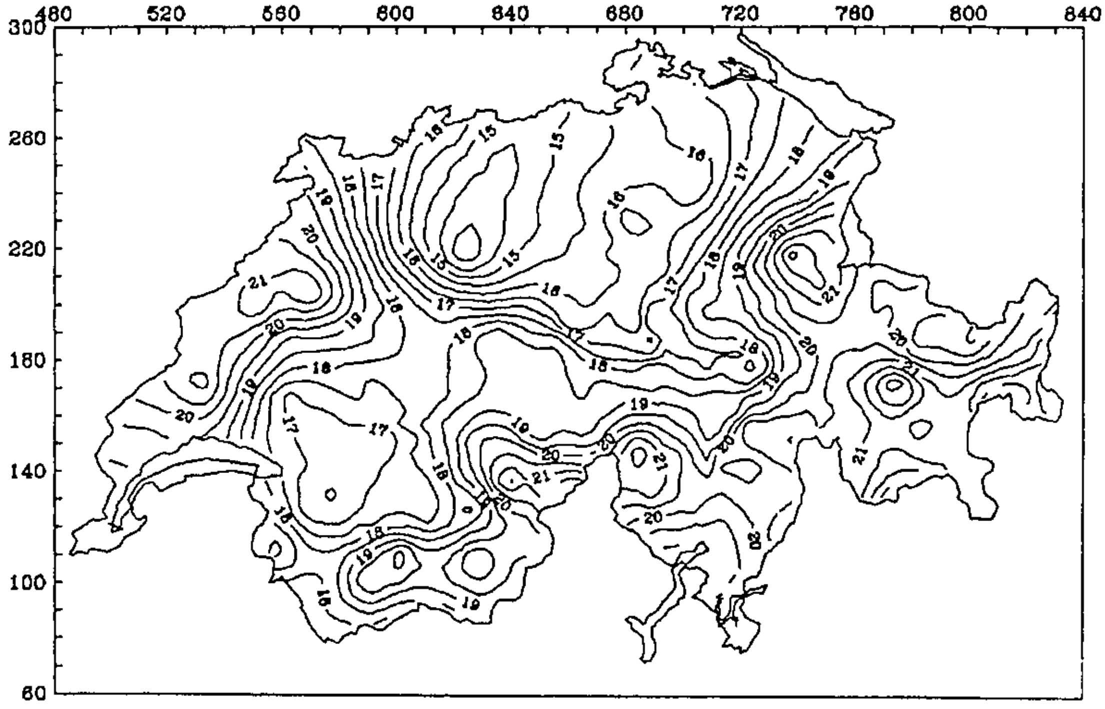


Fig. 5.3.9: Contour levels of April 1, 1992 for kriging (above) and for cokriging (below). Each level corresponds to 0.5 MJm^{-2}day^{-1}. CoK-contours are closer to reality in the Eastern Alps and in the NE of the Lake of Geneva.

5.4 Use of interpolated cloudiness from gridded meso-β analysis

Study prepared by: *Weine Josefsson*
Swedish Meteorological and Hydrological Institute

Under sponsorship from: *Swedish Council for Building Research*

5.4.1 Introduction

A simple and rough method of computing global radiation using dense, gridded cloudiness data from the Swedish Meteorological and Hydrological Institute (SMHI) meso-β analysis have been compared to other methods of obtaining global radiation at sites with no measurements. One method is the Nearest Neighbor, which extrapolates the data from the nearest station. The radiation climate of the two stations is not considered. Another method uses weighted averages of values from surrounding stations. The weights are proportional to the inverse square of the distance from the point of interest. The distance referred to in this section is the horizontal distance, usually given in km.

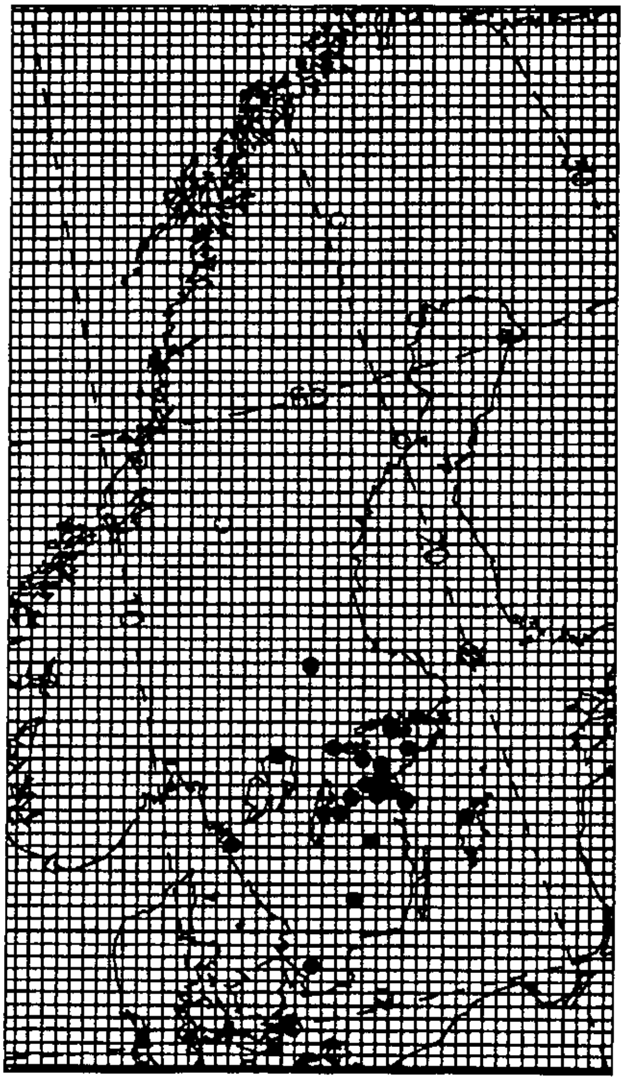


Figure 5.4.1: The grid used in the meso-β analysis (22 x 22 km). The network of global radiation measurements used in the validation are plotted as dots.

There are plans to implement cloud information retreived from satellites into the meso-β analysis. Therefore, the gridded cloud information can be from any source including observations from synoptic stations and analyzed humidity fields. The future potential of this analysis for computing fields of global radiation, even at an hourly time scale, is very good.

5.4.2 Data sets

Two data sets have been compiled to compare the different methods. The first set contains global radiation from 21 stations in southern Sweden. The stations, listed in Table 5.4.1, belong to two networks. One of them is part of an experiment and, therefore, is rather dense.

Table 5.4.1

Stations in the validation. The N:o is used in some of the plots. The latitude is given in degrees and minutes. The altitude is in m. A star in column two indicates that data from the station also was included in the data set from 1988.

N:o	WMO-Nr	Station name	Latitude	Longitude	Altitude
1	2071 *	Norrköping	58°35'	16°09'	43
2	2703 *	Grundbro	59 19	17 07	30
3	2705 *	Örebro	59 16	15 17	23
4	2708 *	Katrineholm	58 59	16 14	62
5	2710 *	Västerljung	58 56	17 26	34
6	2717 *	Jönåker	58 45	16 46	30
7	2721 *	Örberga	58 26	14 50	105
8	2723 *	Herrberga	58 24	15 13	85
9	2724 *	Vreta Kloster	58 29	15 31	52
10	2725 *	Söderköping	58 28	16 21	27
11	2726 *	Östermem	58 30	16 37	30
12	2728 *	Thorönsborg	58 21	16 45	4
13	2734 *	Harstena	58 15	17 00	17
14	2740 *	Vimmerby	57 40	15 51	107
15	2627	Lund	55 43	13 13	73
16	2641	Växjö	56 56	14 44	182
17	2513	Göteborg	57 42	12 00	5
18	2091	Visby	57 40	18 21	51
19	2483 *	Stockholm	59 21	18 04	30
20	2415	Karlstad	59 22	13 28	46
21	2749	Borlänge	60 29	15 26	140

The other one is the southern part of the national solar radiation network in Sweden. The sampled period is April 28 to October 31 1989. There are gaps in the measurements and the number of available daily sums for each station varies from 140 to 180. This data set should not be confused with a similar data set for the period from April to September 1988, which was distributed among the participants of IEA Task IX.

The reason for compiling a new set was the availability of the meso-β information. From this analysis, cloud cover data for three levels and for every third hour have been extracted. Data are available in a very dense grid (22 km x 22 km) covering Scandinavia. Each station of the surface radiation network has been identified with the nearest gridpoint and data have been collected for the same period of 1989.

The validation is performed on daily values although the original data are for shorter time periods. The measured global radiation is based on six-minute averages and hourly grid estimates were obtained by interpolating between three-hourly computations. There were no missing data from the meso-β analysis. However, there are several gaps in the measured data. Very short ones have been interpolated to get complete daily values. Also note that the validation period only includes the summer half of the year.

5.4.3 Model

The cloud information from the meso-β analysis have been used in a very simple model, here referred to as the meso-β model. Note that the gridpoint closest to the radiation station has been used in the validation. The two 'sites' should be within 15 km from each other. Details of the model may be found in [5.4.1]. However, the principal approaches and some important considerations will be presented.

Using time, latitude and longitude, it is possible to compute the global extraterrestrial radiation, G_{ex}. To compute the 'clear' day global irradiance, a simple relation between the clear day transmittance, t_0, and the solar elevation, h_{sun}, was used. It includes correction terms for snow covered ground and a rough seasonal variation. This type of empirical approach could easily be adapted to any climate using local relationships. An important improvement should be to include daily information of the aerosol optical depth (turbidity) input to the model.

$$t_0 = 0.50 + 0.30 \sin(h_{sun})^{0.75} + seas + snow$$

The meso-β cloud amount for each level (low=cl, middle=cm and high=ch) is used to compute the total cloud transmittance, t_c, for every third hour, including the effect of multiple reflections, g_{mult}. The screening effect between different cloud layers is taken into account and also the overestimation of cloudiness by the observer. The calculation is simplified by using a constant transmittance for each cloud level. A simple way to improve the model in this respect would be to include information on the occurrence of precipitation.

$$\begin{aligned} t_c = {} & ((1-ch) + ch*th)* \\ & ((1-cm) + cm*tm)* \\ & ((1-cl) + cl*tl)*gmult \end{aligned}$$

$th = 0.90$ high level cloud transmittance

$tm = 0.37$ middle level cloud transmittance

$tl = 0.28$ low level cloud transmittance

The transmittances of the other hours are found by linear interpolation. The reflectance of the ground is very important, especially if snow cover is present. The months involved in this test do not include cases with snow cover.

Applying the transmittance of the clouds to the clear day global radiation will give *hourly* values of the global radiation at each grid point:

$$G_h = G_{ex} * t_0 * t_c$$

These hourly values were summed, giving daily values to be used in the validation.

5.4.4 Correcting computed values using a reference station

The next step investigated the use of a reference station or a 'ground truth' station to improve the computed global radiation using the meso-β model. The station in Norrköping was chosen as reference. It is situated centrally in the dense part of the network and has a good performance record. The following ratios between daily values were determined for the reference station

$$R_{Ref} = G_m / G_c$$

$$X_{Ref} = G_c / G_{ex}$$

where G_m is the measured daily global radiation, G_c is the computed daily global radiation and G_{ex} is the daily global extraterrestrial radiation. The other stations in the network were corrected according to this ratio such that all stations having a lower or equal daily total transmittance than Norrköping, $X_{Stn} < X_{Ref}$, were corrected by the same ratio, R_{Ref}. At stations having a higher total transmittance than the reference, $X > X_{Ref}$, the global radiation were corrected by a smaller factor. This factor was determined by a linear interpolation between the ratio found for Norrköping and the ratio for a total daily transmittance equal to 0.85, which roughly corresponds to a perfectly clear day. To avoid gross errors, the corrected total global transmittance was limited to values between 0.05 and 0.77. One must note that all these adjustments are applied to daily totals.

5.4.5 Results of the validation

The validation is based on daily totals only. The measures of the performance are the mean bias error, MBE, and the root mean squared error, RMSE, or more precisely the mean bias deviation between calculated and measured global radiation and the corresponding root mean squared deviation.

The methods that have been used to calculate daily global radiation in this validation are:

A. nearest neighbor / extrapolation,

B. interpolation by weights proportional to the inverse squared distance
  - using all available stations
  - using the three closest stations,

C. using meso-β
  - model
  - model + one reference station

Full data sets of the validation using each method are given in Tables 5.4.2 to 5.4.6. The RMSE relative to the average global radiation for each method are presented in the following figures as a function of the distance. Using the nearest neighbor method the distance is simply the distance to the nearest neighbor. For weighted interpolation, the distance is the weighted value of the distances to the stations used in the interpolation. Using the meso-β model the distances are relative to the Norrköping station, which also has been used as a reference station.

The nearest neighbor method shows a significant increase of the RMSE with distance. From about 10% at 25 km it reaches 25% at 200 km. Weighted interpolation using three stations seems to give slightly lower RMSE. From about 8% at 25 km it will be about 22% at 200 km. Using all stations will not significantly lower the RMSE at smaller distances, but will only change the computed distances, as defined above. At longer distances the use of all stations seems to increase the RMSE.

At very short distances, in this case less than 25 km, extrapolation and interpolation methods may reach RMSE-levels of 5-10%. This is close to the level of accuracy of the measurements in the dense network. Probably, this is also close to the level of variability on the spatial scale of 0-10 km for daily values. The typical scale of convective cloud formations is in this range and the clouds are the main cause of radiation climate variability at this scale in this network.

One must note that the network is not in a homogenous area. The coastal zone and the large lakes of Sweden as well as even the low mountains will produce large gradients in the radiation field. Therefore, using the station at minimum distance as the nearest neighbor for extrapolation and interpolation will not always be the best choice. Knowledge of the radiation climate will be an important input to select among the stations and it will, most probably, improve the final result.

The meso-β model produced daily global radiation values with an RMSE in the range of 11-20%. The median value was close to 14%. Once again it should be noted that the validation only includes the summer half of the year. Including the winter will increase the relative RMSE. It is interesting to compare this result with the results of different models in a validation study performed by IEA Task 9 [5.4.2]. The RMSE of daily values of global radiation using all the available stations of different climates is in the range 10-20% using rather sophisticated models based on hourly synoptic observations. Simpler models more comparable to the meso-β model in complexity produced RMSE in the range 10-30%. This would probably be the range of the present meso-β model applying it to a full year of data in the Swedish climate. Using a slightly more complex model that uses snow cover and precipitation information will improve the performance.

Using a reference station to correct the meso-β computed values decreased the scatter in the RMSE. If the RMSE of each station was plotted versus the distance from the reference station it was obvious that the overall performance was improved only for stations within about 75 km from the reference station.

5.4.6 Applications on the solar radiation network of Sweden

The national solar radiation network in Sweden consists of twelwe stations. The inter-distances between the stations varies from 160 to 300 km. The stations next to each other may be situated in different climates, e.g. coastal or inland. In such a sparse network, large areas will be far away from the existing stations and also outside the net. Therefore, interpolation methods must be applied under these circumstances.

In the area close to an existing station (<25 km) extrapolation is probably the simplest and best method to estimate the global radiation. This method may be used up to 50 km depending on the local climate.

Use of the meso-β model is definitely the best method for areas where the radiation stations are more than 100 km away and also in situations where the climate is very different from the neighboring stations. Also in the range 50-100 km the meso-β model will produce values as good as those obtained with the other methods. Although there is an indication that interpolation methods may perform better in the range from 50-100 km, it certainly is more practical to use one method, i.e. meso-β, for all distances larger than 50 km.

5.4.7 References

[5.4.1] Josefsson W.: 1989. "Computed Global Radiation Using Interpolated, Gridded Cloudiness from the Meso-β Analysis Compared to Measured Global Radiation", SMHI, Meteorologi, No. **101**.

[5.4.2] Davies J.A. et al: 1988. "Validation of Models for Estimating Solar Radiation on Horizontal Surfaces",IEA Task IX Report IEA-SHCP-9B-1, AES, Canada.

STN	GLOBAL (Whm^{-2})	MBE (Whm^{-2})	MBE (%)	RMSE (Whm^{-2})	RMSE (%)	N:O	DIS (km)
1	4225.1	-56.8	-1.3	598.1	14.2	179	0
2	4362.0	-37.1	-0.9	471.5	10.8	149	99
3	4407.3	-196.7	-4.5	670.2	15.2	162	91
4	4171.3	-4.9	-0.1	521.9	12.5	168	45
5	4241.1	45.1	1.1	554.3	13.1	171	84
6	4155.0	5.3	0.1	543.2	13.1	163	40
7	4339.6	-268.2	-6.2	753.8	17.4	170	78
8	4253.2	-196.9	-4.6	669.8	15.7	170	58
9	4159.5	-6.7	-0.2	581.2	14.0	170	38
10	4162.6	60.0	1.4	613.9	14.7	170	17
11	4206.7	37.6	0.9	543.7	12.9	170	29
12	4379.9	44.5	1.0	520.7	11.9	163	43
13	4699.5	-101.7	-2.2	570.6	12.1	160	62
14	4211.9	110.3	2.6	560.7	13.3	157	103
15	4442.9	-3.4	-0.1	601.0	13.5	177	364
16	4116.1	201.5	4.9	613.8	14.9	158	202
17	4205.6	113.2	2.7	571.4	13.6	161	262
18	4532.8	41.3	0.9	598.2	13.2	175	164
19	4163.7	68.6	1.6	500.8	12.0	179	139
20	4266.4	-267.8	-6.3	675.4	15.8	179	177
21	4196.8	-437.7	-10.4	832.5	19.8	179	215

Table 5.4.2 Daily global radiation **computed** from the **meso-β** cloud information compared with the measured radiation at 21 stations. The stations (STN) are presented in table 5.4.1. GLO is the average global radiation of the validation period including N:o daily values. MBE is the mean bias error and RMSE is the root mean square error. The distance is given from the station at Norrköping.

STN	GLOBAL (Whm^{-2})	MBE (Whm^{-2})	MBE (%)	RMSE (Whm^{-2})	RMSE (%)	N:o	DIS (km)
1	4225.1	0.0	0.0	0.0	0.0	179	0
2	4362.0	78.6	1.8	624.7	14.3	149	99
3	4407.3	-127.2	-2.9	606.9	13.8	162	91
4	4171.3	51.0	1.2	406.1	9.7	168	45
5	4241.1	113.9	2.7	598.2	14.1	171	84
6	4155.0	75.6	1.8	507.1	12.2	163	40
7	4339.6	-208.6	-4.8	633.1	14.6	170	78
8	4253.2	-136.3	-3.2	551.4	13.0	170	58
9	4159.5	59.5	1.4	472.7	11.4	170	38
10	4162.6	128.5	3.1	429.2	10.3	170	17
11	4206.7	112.2	2.7	500.7	11.9	170	29
12	4379.9	130.4	3.0	527.4	12.0	163	43
13	4699.5	-16.1	-0.3	551.9	11.7	160	62
14	4211.9	164.2	3.9	613.3	14.6	157	103
15	4442.9	34.6	0.8	697.2	15.7	177	364
16	4116.1	237.1	5.8	655.0	15.9	158	202
17	4205.6	178.1	4.2	699.3	16.6	161	262
18	4532.8	84.1	1.9	690.9	15.2	175	164
19	4163.7	133.7	3.2	649.9	15.6	179	139
20	4266.4	-221.8	-5.2	700.7	16.4	179	177
21	4196.8	-399.3	-9.5	796.3	19.0	179	215

Table 5.4.3 Daily global radiation computed from the meso-β cloud information and corrected by using the station at Norrköping (N:o 1)as reference station compared with the measured radiation at 21 stations. The stations (STN) are presented in table 5.4.1. GLO is the average global radiation of the validation period including N:o daily values. MBE is the mean bias error and RMSE is the root mean square error. The distance is given from the station at Norrköping.

STN	GLOBAL (Whm-2)	MBE (Whm-2)	MBE	RMSE (Whm-2)	RMSE	N:O	DIST
1	4247.9	-90.9	-2.1	416.2	9.8	172	17.4
2	4354.3	46.3	1.1	661.9	15.2	151	46.3
3	4393.3	-53.6	-1.2	694.4	15.8	161	62.6
4	4113.4	35.1	0.9	579.4	14.1	162	40.1
5	4167.9	-18.4	-0.4	581.2	13.9	165	43.4
6	4149.5	-10.6	-0.3	399.7	9.6	165	29.1
7	4332.0	-84.6	-2.0	400.5	9.2	172	22.6
8	4247.5	-93.5	-2.2	372.9	8.8	172	19.7
9	4153.9	93.5	2.3	372.9	9.0	172	19.7
10	4157.0	43.8	1.1	320.8	7.7	172	16.0
11	4200.9	-43.8	-1.0	320.8	7.6	172	16.0
12	4371.8	-74.8	-1.7	481.4	11.0	165	18.4
13	4682.5	-248.0	-5.3	548.2	11.7	161	18.4
14	4205.7	111.7	2.7	818.9	19.5	159	89.5
15	4513.9	-391.9	-8.7	1061.1	23.5	159	164.3
16	4203.5	94.3	2.2	800.5	19.0	141	105.5
17	4227.2	-157.5	-3.7	1015.7	24.0	143	184.7
18	4713.9	-10.0	-0.2	902.3	19.1	159	102.6
19	4322.4	31.9	0.7	609.4	14.1	151	54.0
20	4439.5	-42.3	-1.0	810.3	18.3	164	103.6
21	4384.8	12.4	0.3	997.0	22.7	164	135.4

Table 5.4.4 Daily global radiation **extrapolated** using **next neighbour** station as **reference** compared with the measured radiation. The stations (STN) are presented in table 5.4.1. GLO is the average global radiation of the validation period including N:o daily values. MBE is the mean bias error and RMSE is the root mean square error. The distance is the distance to the next neighbour.

STN	GLOBAL (Whm-2)	MBE (Whm-2)	MBE (%)	RMSE (Whm-2)	RMSE (%)	N:o	DIST (km)
1	4247.9	-49.1	-1.2	329.1	7.7	172	38.1
2	4354.3	28.8	0.7	517.5	11.9	151	84.0
3	4393.3	-1.2	0.0	703.7	16.0	161	107.5
4	4113.4	49.6	1.2	410.4	10.0	162	67.4
5	4167.9	-2.4	-0.1	546.1	13.1	165	74.1
6	4149.5	19.1	0.5	320.9	7.7	165	51.0
7	4332.0	-106.7	-2.5	448.3	10.3	172	55.1
8	4247.5	-18.0	-0.4	319.2	7.5	172	40.9
9	4153.9	93.8	2.3	332.1	8.0	172	45.4
10	4157.0	81.7	2.0	308.0	7.4	172	30.2
11	4200.9	32.4	0.8	283.2	6.7	172	30.2
12	4371.8	22.9	0.5	354.9	8.1	165	32.9
13	4682.5	-281.6	-6.0	576.6	12.3	161	43.4
14	4205.7	107.6	2.6	575.9	13.7	159	117.6
15	4513.9	-243.4	-5.4	1069.4	23.7	159	314.4
16	4203.5	172.5	4.1	757.7	18.0	141	190.1
17	4227.2	25.3	0.6	973.0	23.0	143	250.2
18	4713.9	-286.1	-6.1	1058.2	22.4	159	165.6
19	4322.4	73.1	1.7	714.6	16.5	151	106.0
20	4439.5	-36.1	-0.8	897.9	20.2	164	176.0
21	4384.8	18.2	0.4	1028.8	23.5	164	208.6

Table 5.4.5 Daily global radiation interpolated using all the other stations in the network compared with the measured radiation. The weights in the interpolation are proportional to the inverse of the squared distance. The stations (STN) are presented in table 5.4.1. GLO is the average global radiation of the validation period including N:o daily values. MBE is the mean bias error and RMSE is the root mean squared error. The distance is the weighted distance of the stations used in the interpolation.

STN	GLOBAL (Whm-2)	MBE (Whm-2)	MBE (%)	RMSE (Whm-2)	RMSE (%)	N:o	DIST (km)
1	4247.9	-81.0	-1.9	361.4	8.5	172	22.8
2	4354.3	1.7	0.0	451.6	10.4	151	52.6
3	4393.3	-41.7	-0.9	718.6	16.4	161	75.9
4	4113.4	42.5	1.0	437.1	10.6	162	45.9
5	4167.9	-29.5	-0.7	558.7	13.4	165	49.0
6	4149.5	-26.1	-0.6	381.6	9.2	165	34.6
7	4332.0	-105.8	-2.4	412.6	9.5	172	29.9
8	4247.5	-15.9	-0.4	281.2	6.6	172	23.2
9	4153.9	107.2	2.6	325.2	7.8	172	26.2
10	4157.0	64.4	1.5	290.6	7.0	172	18.5
11	4200.9	12.5	0.3	291.0	6.9	172	18.8
12	4371.8	45.1	1.0	347.8	8.0	165	20.0
13	4682.5	-277.3	-5.9	575.3	12.3	161	24.7
14	4205.7	93.8	2.2	622.9	14.8	159	91.6
15	4513.9	-331.4	-7.3	1017.9	22.5	159	203.4
16	4203.5	178.1	4.2	613.5	14.6	141	132.2
17	4227.2	-5.2	-0.1	939.2	22.2	143	190.7
18	4713.9	-172.2	-3.7	963.5	20.4	159	117.0
19	4322.4	46.7	1.1	589.9	13.6	151	62.4
20	4439.5	-19.5	-0.4	824.6	18.6	164	123.6
21	4384.8	6.1	0.1	936.0	21.3	164	158.7

Table 5.4.6 Daily global radiation **interpolated** using the **three** closest **neighbours** compared with the measured radiation. The weights in the interpolation are proportional to the inverse of the squared distance. The stations (STN) are presented in table 5.4.1. GLO is the average global radiation of the validation period including N:o daily values. MBE is the mean bias error and RMSE is the root mean square error. The distance is the weighted distance of the three stations used in the interpolation.

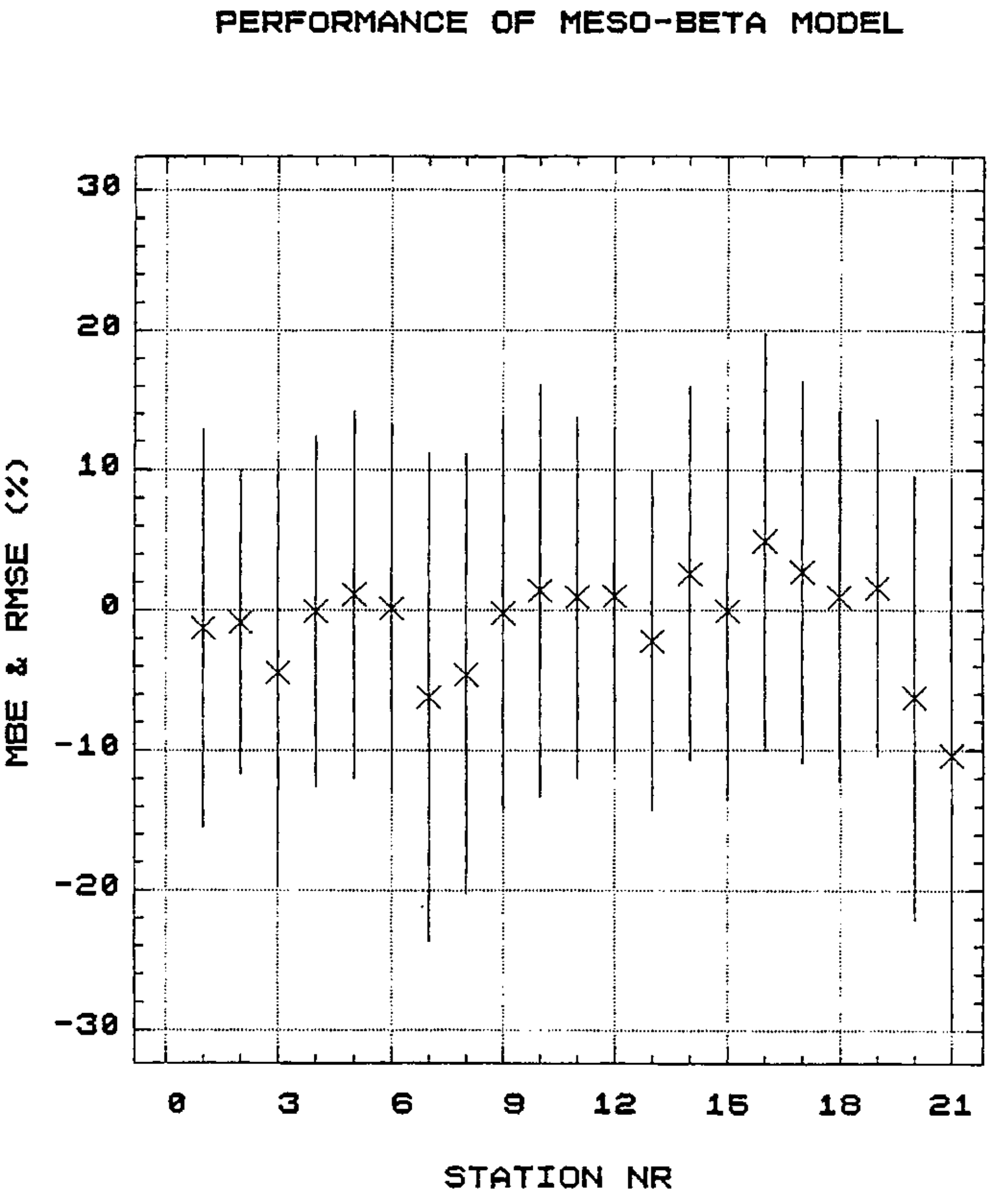


Fig. 5.4.2: Performance of the meso-β model for the 21 stations and for the daily values of the test data period 28 April to 31 October 1989. Stations are given in Table 5.4.1. The mean bias (MBE, x) and the root-mean-squared-error (RMSE, bars) are related to the average global radiation.

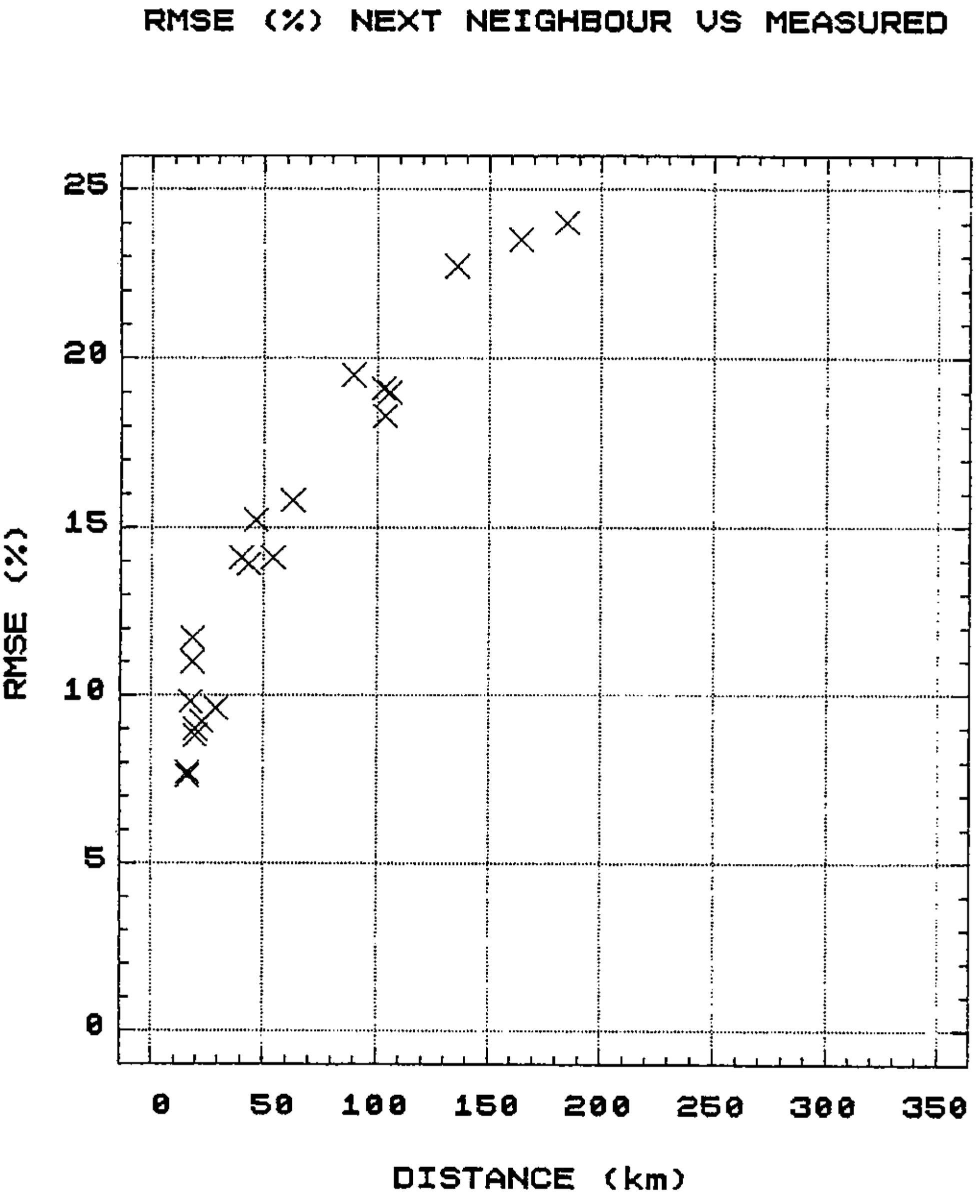


Fig. 5.4.3: The RMSE (%) of daily values of global radiation with the method of extrapolation, i.e. using the next neighbour, plotted versus the distance (km) to the next neighbour.

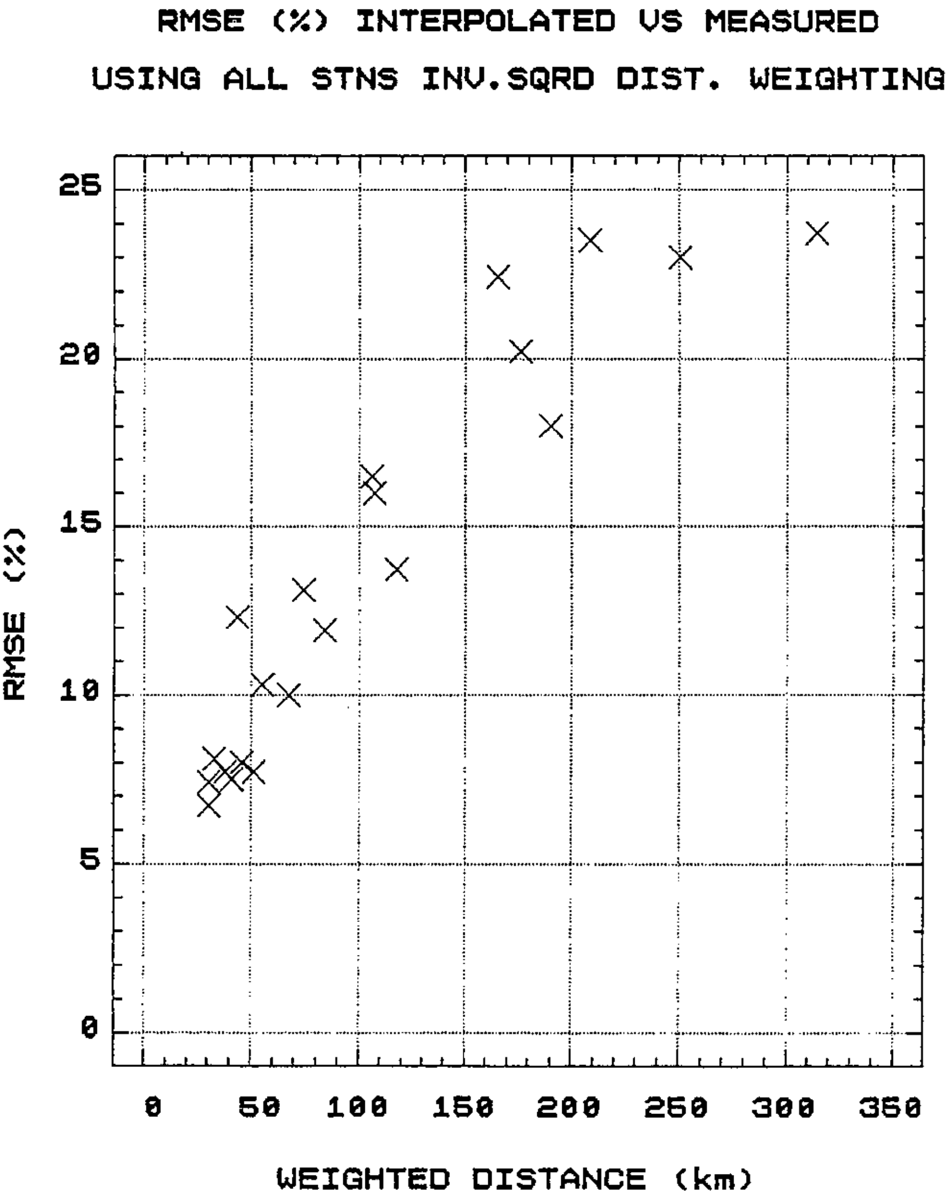


Fig. 5.4.4: The RMSE (%) of daily values of global radiation of the interpolation method using **all** available sations plotted versus the weighted distance (km). The weights applied are proportional to the inverse of the squared distance.

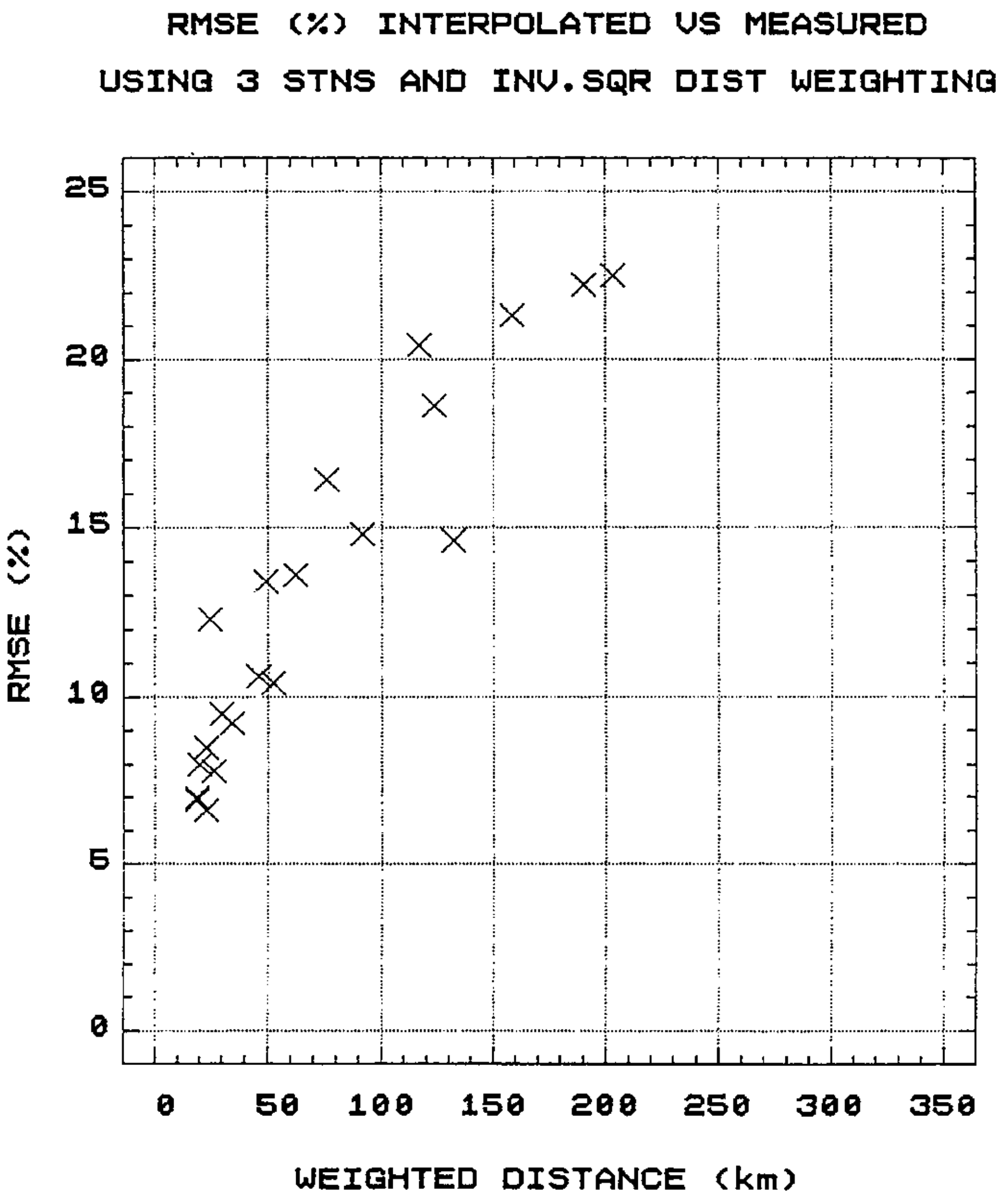


Fig. 5.4.5: The RMSE (%) of daily values of global radiation using the interpolation method and **three** nearby sations plotted versus the weighted distance (km). The weights applied are proportional to the inverse squared distance.

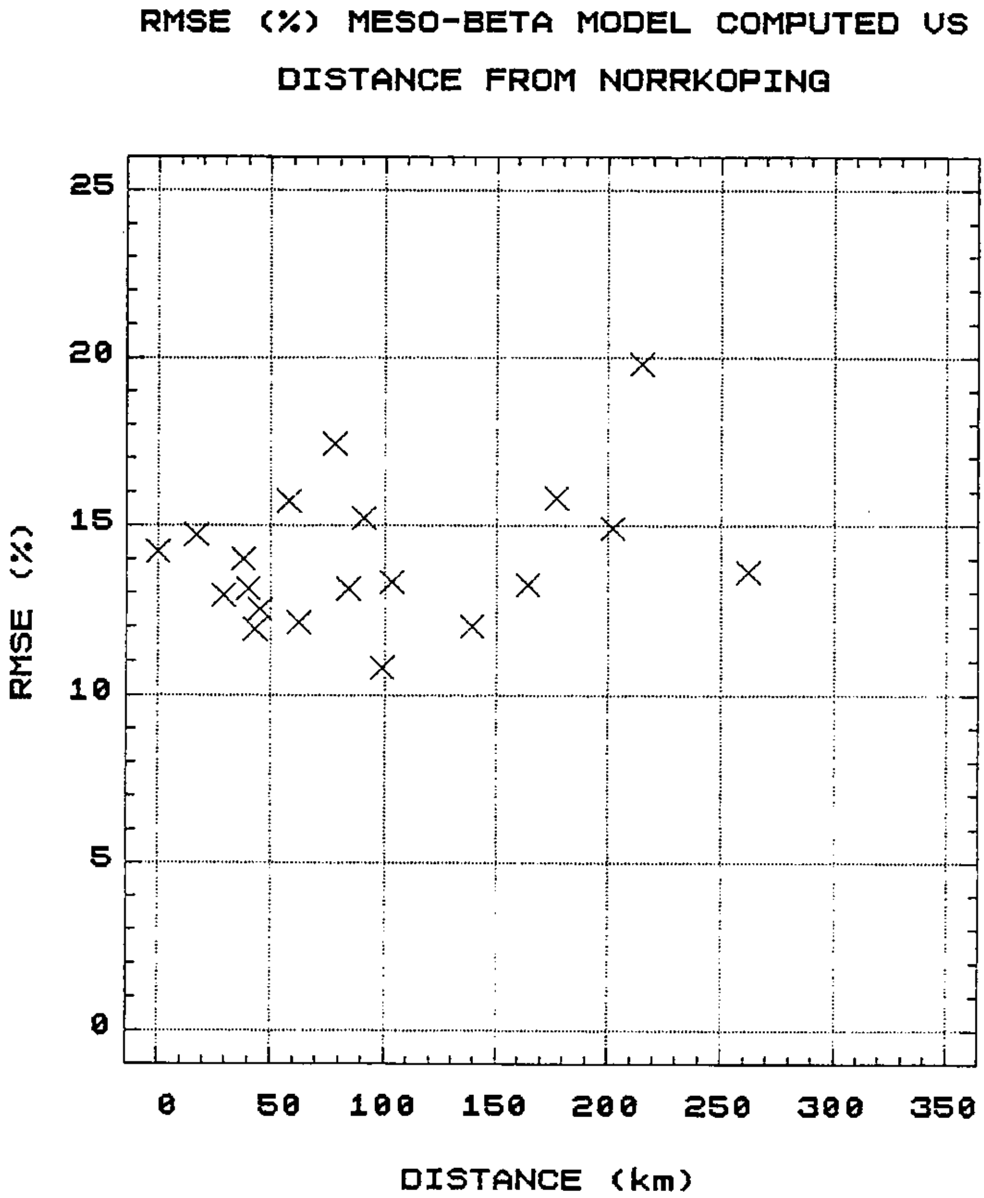


Fig. 5.4.6: The RMSE (%) of daily values of global radiation using the Meso-β model wersus the distance from Norrköping as reference for the next figures.

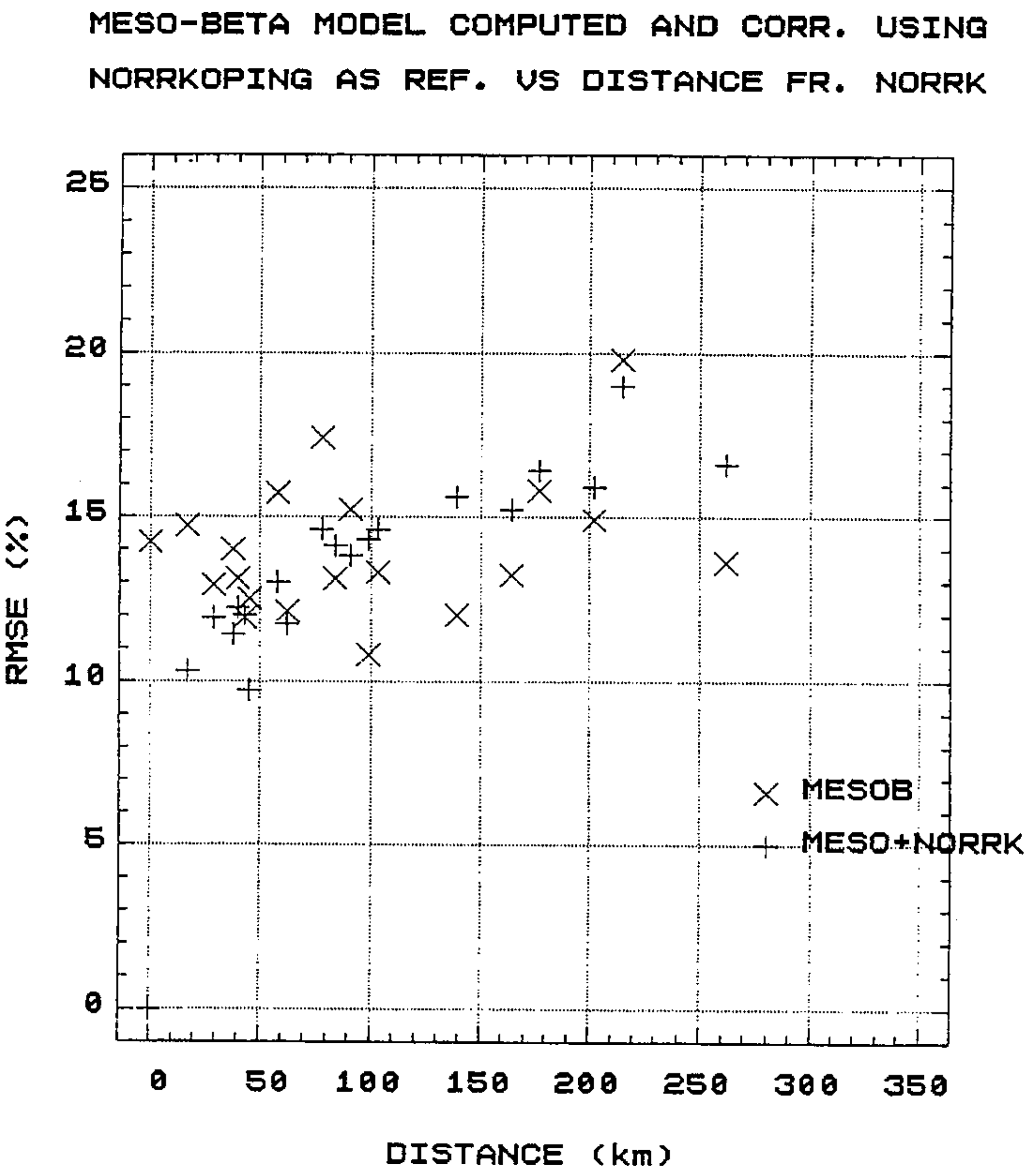


Fig. 5.4.7: The RMSE (%) of daily values of global radiation using the meso-β model only (plotted with x) and using the model with a reference station as described in section 5.4.4 (plotted with +). Both are plotted versus the distance from the reference station Norrköping.

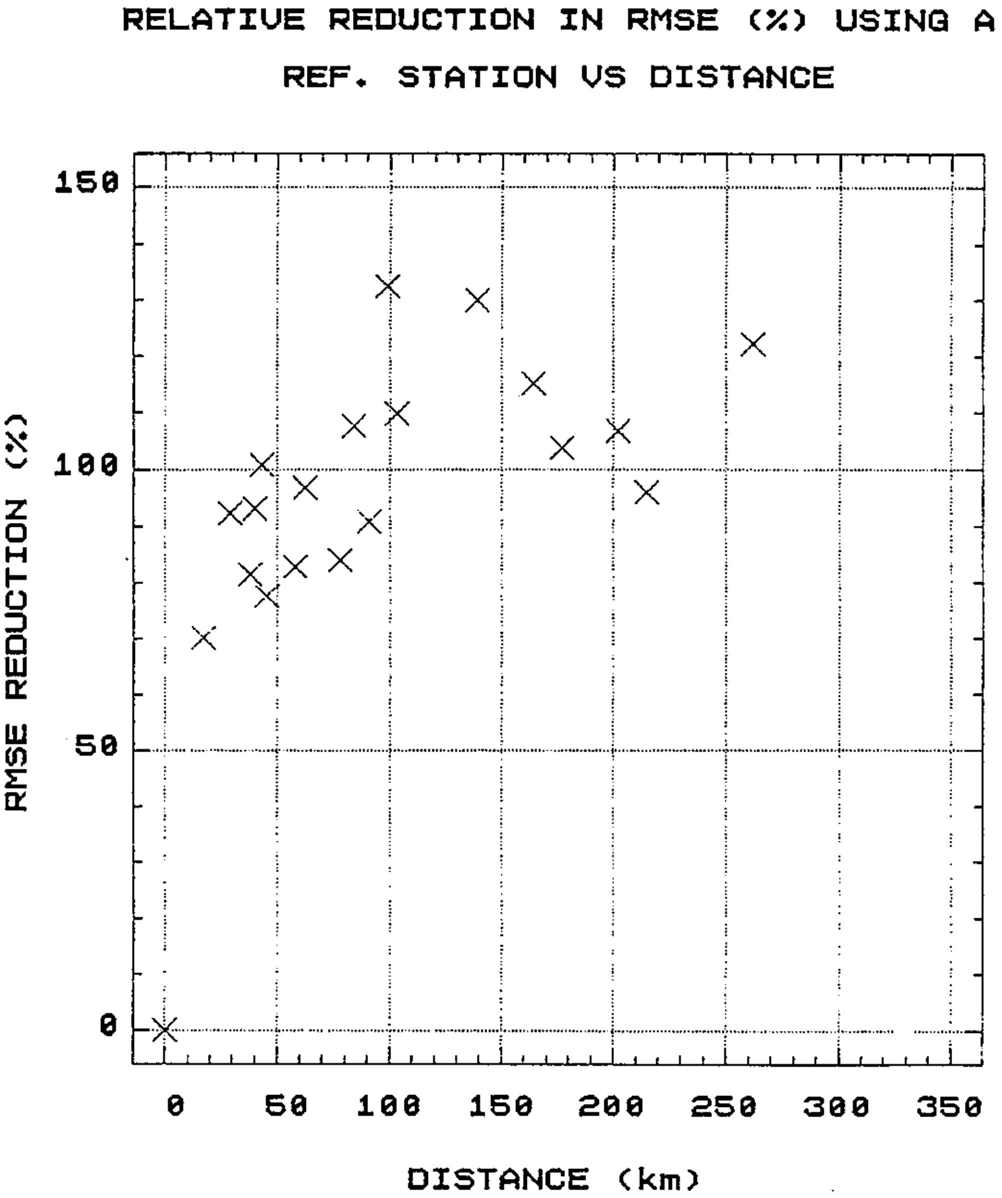


Fig. 5.4.8: The effect of using a reference station together with the meso-β model as described in section 5.4.4 is expressed as a relative change of the RMSE as compared with using only the model (= 100 %).

SUMMARY OF THE DIFFERENT METHODS

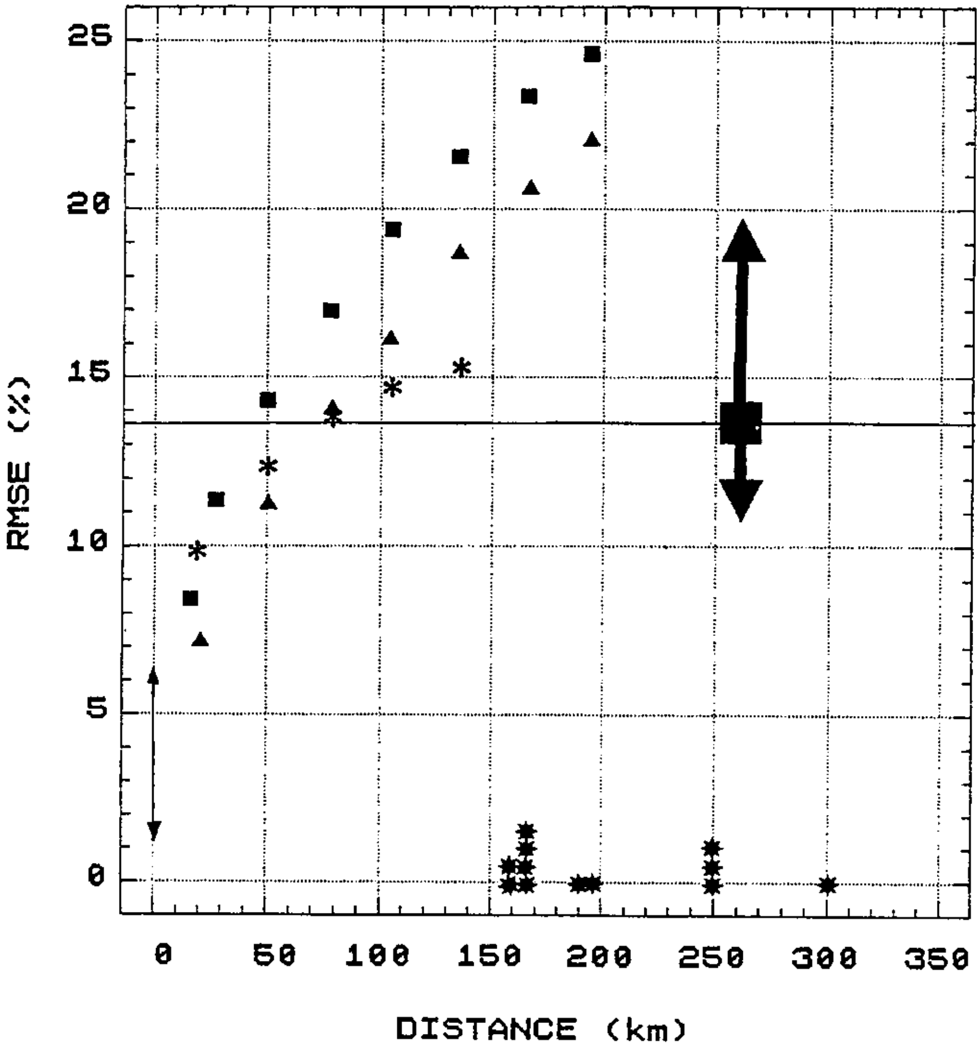


Fig. 5.4.9: Summary of the RMSE (%) of the different methods to retrieve daily global radiation and their smoothed dependence on distance (km). Symbols are: squares for extrapolation, triangles for interpolation, and stars for the meso-β model with reference station. The large black square indicates the median of the meso-β model and the attached arrows its range. The thin vertical bar with small arrows denotes an estimation of the precision of the measured values in the validation data set. Stars on the x-axis represent the distances to the next neighbour in the Swedish radiometric network.

CHAPTER 6

6. End-use validations

6.1 Impact of Solar Radiation Resources on the Siting of Solar Power Plants

Study prepared by:	*Eugene L. Maxwell and Martin D. Rymes*
Under sponsorship from:	*United States Department of Energy* *Solar Thermal and Resource Assessment Programs*

6.1.1 Background

The selection of a site for any power plant is a multiphase, multifaceted problem that generally reduces to a cost/benefit analysis. The fundamental benefit is providing energy when and where it is needed. The major factors that affect the cost of that energy generally are dominated by fuel and construction costs. Other factors — such as the need for a water supply, community acceptance, and government regulations that deal with safety and environmental issues — can exclude many locations from serious consideration.

For solar power plants, the fuel (solar radiation) is free for the taking. However, the collectors needed to acquire the fuel will probably entail the single largest cost of the entire power plant. The environmental and natural resource variables that will affect the construction costs of a solar power plant include

- Land availability
- Solar radiation (which affects the collector field size)
- Topography
- Soils and surface geology
- Seismic history.

Of these natural resource variables, only land availability and solar radiation have been included in this study. These are the variables important for the first phase of power plant siting, the objective of which is to determine the relative suitability of one region versus another. These results significantly reduce the area to be subjected to more costly environmental, economic, and engineering analyses, which require topographic, geologic, soils, and seismic data. Land availability, which is determined from the nature of land use/land cover and land management/land ownership, eliminates large regions such as national parks, cities, and military reservations from further consideration. Solar radiation data narrow the search even more by identifying those regions having the best energy resource for the power plant design under consideration.

Land availability in the Southwest was determined from satellite images and existing maps. Available solar radiation data were found to be severely lacking, which led to an emphasis on improving the assessment and characterization of solar radiation resources. This emphasis was directed at achieving mesoscale (10 km to 200 km) resolution with a combination of surface measurements and derived cloud-cover data as input to a solar irradiance model.

Most of the work under Subtask 9D has concentrated on interpolation of daily totals of global horizontal solar radiation. This effort departs from that and considers long term (10 years) monthly mean direct normal solar irradiance. This is the most important technical variable for siting a solar power plant using concentrating collectors.

Although the demonstration of a methodology for siting solar power plants incorporated a number of other variables, these will only be mentioned briefly since they are not relevant to Subtask 9D. The use of all readily available information affecting the solar energy resource is pertinent and will be discussed in some detail.

6.1.2 The Siting Methodology — An Overview

The methodology that has been developed for regional evaluations of site suitability employs a computerized Geographic Information System (GIS) to overlay and combine mappable variables, using an algorithm that calculates site suitability on a cost/benefit basis. The overlaying of data planes is illustrated in Figure 6-1. The figure provides a pictorial representation of multiple land uses overlain with six categories of land management. In a computerized implementation of this process, each of the rectangular cells shown in this figure would be assigned a land management and a land use/land cover designation. One can also envision overlays of numerous other geographically related information, such as solar radiation resources, topography, water resources, population densitities, etc. Once all of the variables of interest have been geographically overlain in the computer, they may be combined or used in any multivariate type of analysis that the computer is capable of performing. This is most commonly referred to as compositing, i.e., the forming of a composite map.

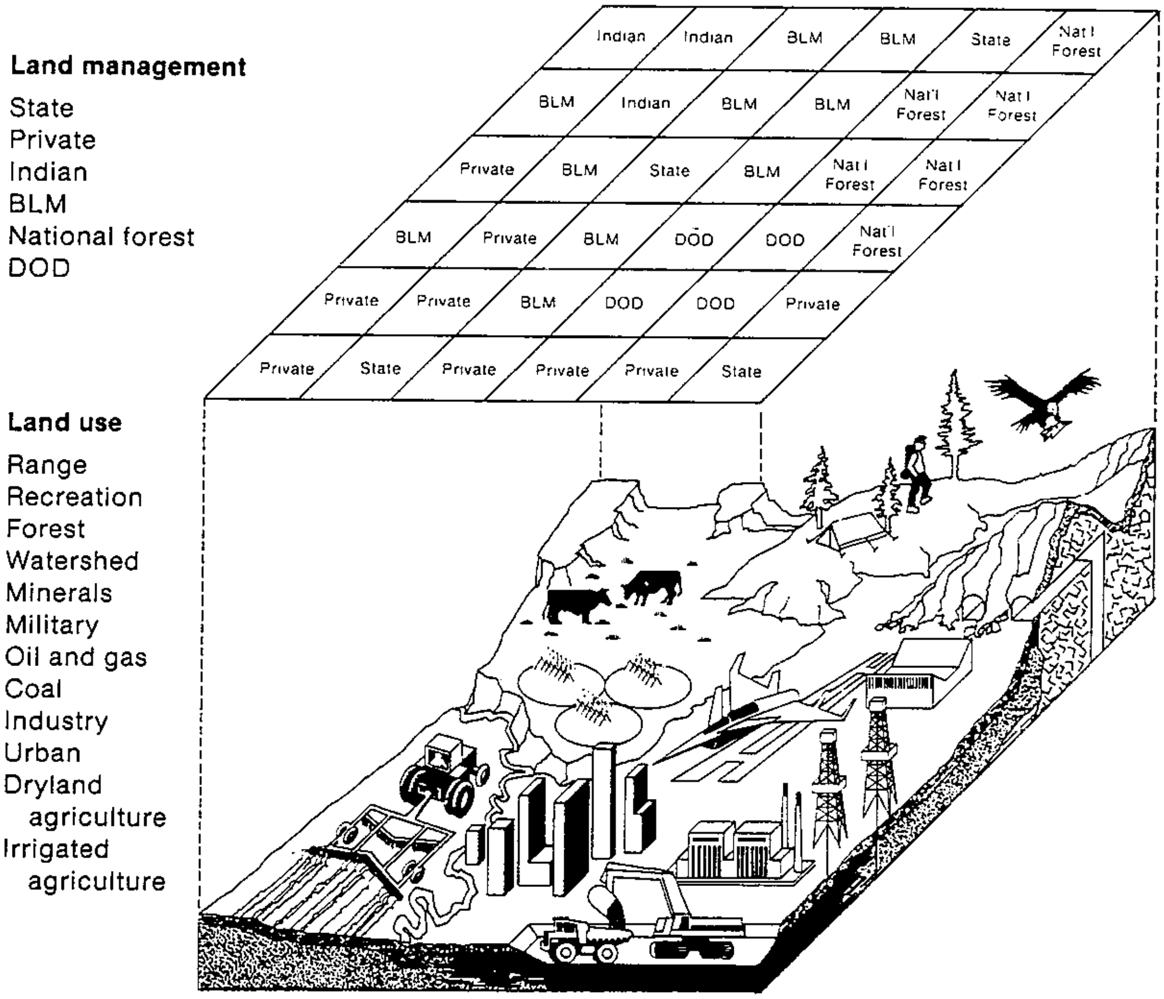


Figure 6-1. Overlay of Land Use/Land Cover with Land-Management Classes

McHarg (1969) was one of the first to employ the overlay and compositing process. He prepared transparent gray-scale maps of the variables needed and physically overlayed them to obtain a composite gray-scale map indicating the relative value of suitability of each location within the mapped region. This was a labor-intensive, inflexible process involving cartographic machine transformations of map scales and projections and the manual creation of gray-scale representations of the original maps. Both the manual and computer versions of this process have been used for many applications, such as environmental impact analysis (Turner 1976), selection of routes for highways, power lines, etc., and site selection for microalgae production systems (Maxwell, Folger, and Hogg 1985).

6.1.3 Factors Affecting Site Selection

In general, the factors important to the siting of a solar power plant are no different from those that must be considered in siting hydroelectric, coal, gas, or nuclear plants. These factors include the following:

- The regional supply and demand for electric power
- The energy resources to be used in generating electricity
- The transmission of electricity from the plant to the loads
- Water resources for generating steam and disposing of waste heat
- Land resources for constructing plants and substations
- Topographic and geologic factors affecting construction costs
- Environmental, legal, and political constraints.

The importance of these factors and the methods of dealing with them vary for different types of plants and from one region to another. For example, the builder of a hydroelectric plant is not concerned with disposing of waste heat; rather, the water resources become the energy resources. And, of course, the environmental impacts associated with the various types of plants are greatly different.

Diurnal and seasonal variations in the demand for electricity are always important concerns to electric utilities. In terms of solar power plants, this concern is complemented by the diurnal and seasonal variations in the primary energy resource. For some energy demands (e.g., heating of homes at high latitudes), the availability of the solar resource is out of phase with the demand. For air conditioning, however, the peak demand is generally correlated with the peak supply of the solar resource. This correlation of supply and demand could lead to favorable economics for solar power in the U.S. Southwest or Southeast (where air-conditioning demands are high) during this century.

Estimating the Effect of Air Conditioning. A large proportion of the summer demand for electricity in the southern states can be attributed to air conditioning. This demand is related to the cooling-degree-days for the month and region of interest and the temperature-humidity index.

Cooling-degree-days are directly related to the energy required to maintain a constant indoor temperature, and the temperature-humidity index provides a measure of human discomfort, which is roughly related to the demand for air conditioning.

Land Resources. Land resources are needed for any type of power plant. In addition to the land required for the switching gear, control systems, transformers, etc., a solar power plant must have a great deal of land for the collectors that capture the solar radiation.

The potential availability and use of land for solar power plants encompasses a variety of physical, economic, legal, and political factors. Potentially available land must be physically suitable, politically available, and economically affordable. Political availability is considered an important factor in much of the Western U.S. because of the large amount of publicly owned land.

The land-use maps used for this project were prepared from satellite images (Maxwell, Folger, and Hogg 1985). Land-management information was available from the Bureau of Land Management (BLM) of the U.S. Department of Interior.

Electric Power Transmission. The cost of transmitting electric power is low, compared with the cost of generation, but it is not insignificant. Therefore, the distance over which the plant output might be dissipated was made of part of the regional suitability algorithm. A hypothetical but empirically reasonable equation for estimating the demand for electric power is

$$\text{kWh/person/day} = 6 + (c\ \text{CDD THI}) \qquad (6-1)$$

where

CDD is the cooling-degree-days
THI is the temperature-humidity index
c is a scaling constant (9×10^{-5}).

Population, cooling-degree-days, and temperature-humidity index data were used with Eq. 6-1 to calculate the radial distance within which the output of 30-MW and 100-MW power plants would be dissipated. The average distance over which the plant output must be transmitted was one of the factors used in assessing relative site suitability.

Solar Radiation Energy Resources. Economic analyses of power plants inherently reflect the competition between the different means of generating electricity. Usually, the regional cost of the plant's primary energy resource (coal, gas, nuclear power, etc.) is the most important factor in determining which one will be chosen for a particular plant. Actually, the analysis involves the combined cost of the fuel and the plant that must use it. Although solar energy itself is free, the cost of the collection system required to capture the resource will be a major part of the overall plant cost. For this reason, the development of a cost-effective method for obtaining an accurate assessment of available solar energy resources at a mesoscale resolution was given top priority.

6.1.4 Producing A Meoscale Solar Radiation Data Base

Topography and climate are the dominant factors producing variations in solar radiation resources. In the southwestern United States (hereafter referred to as the Southwest) topographic controls occur along the Pacific Coast and in and around the numerous mountain ranges. Anyone who has visited the West Coast during the summertime is familiar with the nighttime buildup of clouds along the coastline. These clouds may extend inland over distances of a few miles to tens of miles, depending on the local terrain. Normally, the clouds are dissipated by the radiation from the sun by mid-morning, except for low-lying regions immediately adjacent to the coast.

The mountainous regions in the Southwest also exhibit extreme variability in the availability of solar radiation resources. During the summertime, cumulus clouds start to build-up over the mountains by mid-morning, and the afternoon is characterized by frequent rain showers and thunderstorms. This phenomenon itself is quite variable, depending upon the flow of moisture over the region. A 3-to-1 variability in the occurrence of afternoon clouds could be found over distances less than 100 km at a number of locations.

Superimposed on the topographic controls of cloud cover and, hence, solar radiation, one finds climatological factors not associated with topography. Within the Southwest, the most important climatological control is associated with the movement of moisture into the region during the monsoon season, which usually begins in July. This monsoon moisture reaches as far north as Colorado and Utah, but the dominant effect is felt in Arizona and certain parts of New Mexico. Although topography has some effect on the flow of moisture, the winds aloft, particularly the jet stream, provide most of the control.

Even a cursory examination of the weather records for the Southwest leads one to recognize that a mesoscale resolution for the solar radiation resource is a minimum requirement for siting major solar energy facilities. In many localized regions, even microscale resolution (<10 km) will be needed to optimize site selection.

Solar Radiation Measurements in the Southwest. During the period from 1976 through 1980, a mesoscale solar radiation network was established in part of the Southwest by a consortium of public utilities (WEST Associates). The locations of the stations in this network are shown in Figure 6-2. Stations operated by the National Weather Service (NWS) are also shown. It is obvious from this map that mesoscale resolution was achieved for part of southern California and Arizona. However, because WEST Associates stations were located at electric power facilities, they tend to be clustered around the population centers and do not provide uniform or adequate coverage. Since cloud cover is the dominant factor limiting the availability of solar radiation at the earth's surface, the production of a mesoscale cloud-cover map became the first step in the development of mesoscale solar radiation data.

Surface Observations of Cloud Cover. Surface observations of cloud cover have been made in the United States on a regular basis since the establishment of the Weather Bureau in the late 1800s. The majority of the stations established since 1870 are indicated in Figure 6-2, from which it is apparent that surface observations of cloud cover will not achieve mesoscale resolution in the Southwest. In addition to the large separation between observing stations, the radius around the station within which the observer can determine cloud cover is limited and variable. Under ideal conditions, a ground observer might be able to accurately assess percent cloud cover within a radius of approximately 30 km. However, this would depend on the types of clouds present and the extent to which they obscure higher level clouds and clouds at more distant locations. Obviously, fog, smoke, terrain, precipitation, and other factors can limit the observers' ability to accurately assess cloud cover. We are not questioning the observers ability, since our analyses of cloud-cover data lead us to the conclusion that the observers are very conscientious and consistent in performing their task. Neither do we want to cast doubt on the general quality of the cloud-cover data collected by the National Weather Service (NWS). The only qualification that we are making is simply that these data represent cloud cover for a circular region around the observer, which may have a radius as small as 5 to 10 km and rarely larger than 30 km.

The limited range of ground observations of clouds is particularly significant at locations exhibiting high cloud-cover gradients, e.g., mountain and coastal sites. The use of mathematical interpolation techniques to estimate cloud cover between stations becomes useless when the distance between stations is much larger than the topographic features affecting the development of clouds.

Some key advantages of surface observations of cloud cover include the ability to identify cloud type, the ability to make observations at any hour of the day, and the long historical record of cloud observations extending back to the 1870s. These benefits of surface observations cannot be lightly set aside in favor of satellite observations of cloud cover.

Satellite Observations of Cloud Cover. There are three satellite systems that could be used to obtain cloud cover data for the Southwest. There is a network of geosynchronous satellites orbiting the earth above the equator at an altitude of 22,000 miles. These include the GOES East and GOES West satellites, which obtain imagery for North America. GOES satellites are capable of

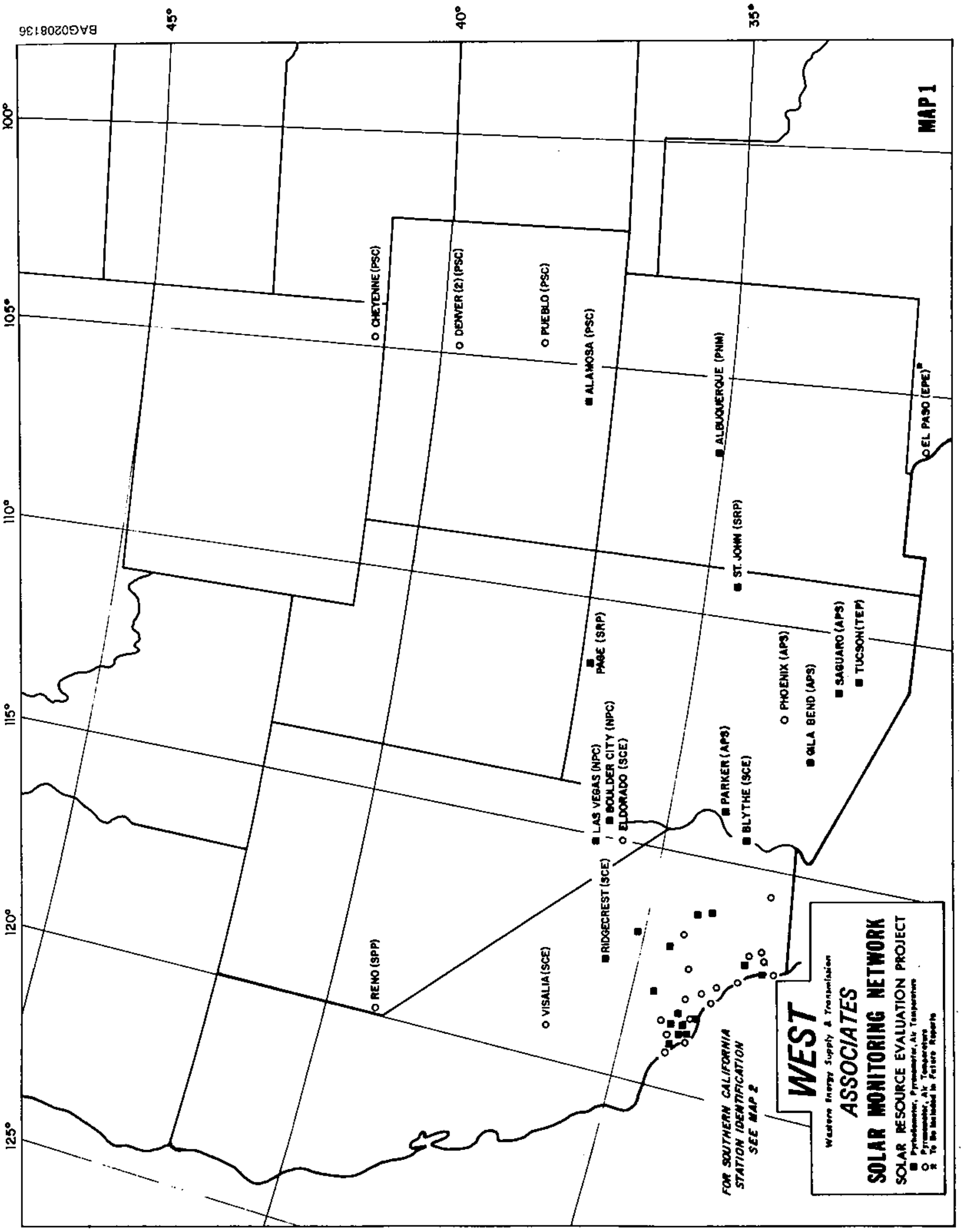


Figure 6-2. Solar Radiation and NWS Meteorological Observation Sites in the Southwest

obtaining images every half-hour. NOAA also operates polar orbiting satellites of the NIMBUS type, which obtain images of the surface of the earth from a much lower elevation, twice a day. Both GOES and NIMBUS satellites have resolutions on the order of 1 to 8 km, depending on the sensor used and the angle at which the earth's surface is viewed. The LANDSAT satellites are also polar orbiters, but they scan a smaller portion of the earth's surface on each overpass at a much higher resolution (79 m). Each LANDSAT satellite obtains images of the earth's surface every 16 to 18 days. The first choices among these satellites for cloud-cover observations are the GOES satellites that have adequate resolution and that are capable of obtaining an image every 30 minutes.

However, for site selection one should use 5 to 10 years of data to obtain climatic averages. Because of limited budgets for this work, we could not process that much data.

Fortunately, the LANDSAT images offered a unique and very inexpensive source of cloud-cover data covering an 11-year period of time. As each LANDSAT image (185 km by 185 km) is obtained, the percentage of the image covered by clouds is determined both by computer analysis and human photointerpretation. The cloud-cover estimate is recorded along with other data describing the position, quality, and other characteristics of the image. This descriptive information is kept on computer tape and is sent on request to potential purchasers of the LANDSAT imagery. The cloud-cover data were obtained by purchasing (for only $300) a computer listing of all LANDSAT images for the southwestern United States from July 1972 through 1983). During part of this period, two satellites on staggered orbits imaged each location every 9 days. This produced a maximum of about 30 images for each month of the year. Figure 6-3 shows the location of the center of each of the LANDSAT scenes over the Southwest. Because of the overlap from scene to scene, the nominal separation of the images in the United States is approximately 125 km.

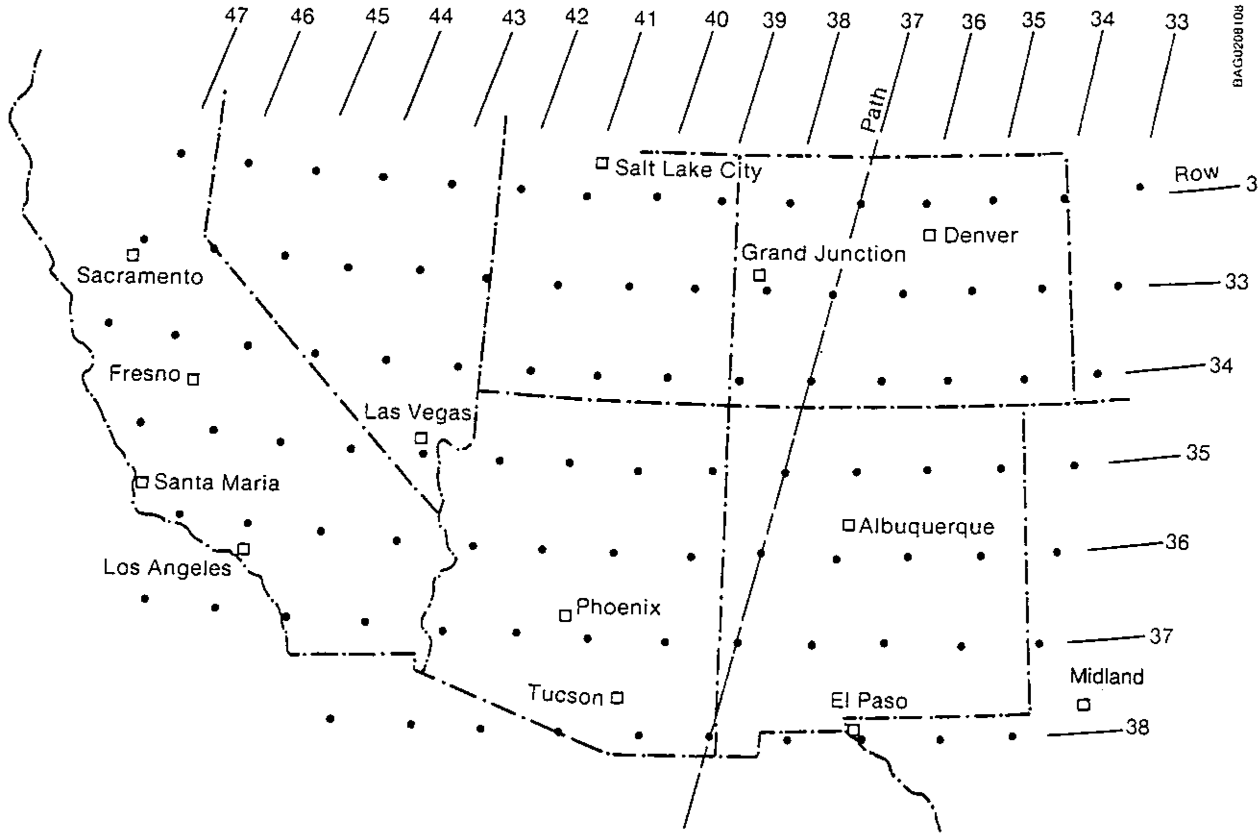


Figure 6-3. **Nominal Centers of LANDSAT Images over the Southwest and Surface Observation Stations Used in this Study**

Although a resolution of 125 km is not ideal, these cloud-cover data have several advantages over surface observations. First, the cloud observations cover the entire national uniformly. Second, whereas the surface observer is recording cloud cover within an average radius of perhaps 20 km, the interpretation of the LANDSAT image is uniform over the entire 185 by 185 km. Furthermore, the uniform spacing of the images, including some overlap from image to image, is ideal for the preparation of a computer-generated cloud cover map. The surface fitting algorithms function best when presented with a uniform raster or grid of data. The nonuniform spacing of surface observations and the large areas between the small regions actually observed result in cloud-cover values which are artifacts of the surface fitting algorithm. Therefore, the spatial quality of the LANDSAT-derived cloud-cover data was expected to be superior to that available from ground observations. The final results obtained under this project support that expectation. Although the results also indicate the need for higher resolution data in mountain and coastal areas, there is much of the United States for which the resolution obtained from the LANDSAT data is quite adequate.

A significant problem accompanying the use of LANDSAT cloud-cover data results from the observation time. Weather satellites and surface observations are capable of obtaining cloud-cover information for every hour of the day. LANDSAT, however, obtains all of its images for the continental United States at approximately 9:30 AM. Furthermore, in order to limit the recording of cloud-cover information to one digit, nine-tenths and ten-tenths cloud cover were both recorded as nine-tenths. This is of little concern to the users of LANDSAT images, because they are concerned mostly with their ability to see the surface of the earth and nine-tenths is just as bad as ten-tenths. It was necessary to use surface observations to correct for these shortcomings of LANDSAT cloud-cover data.

Evaluating and Correcting LANDSAT Cloud-Cover Observations. Surface observations of cloud cover were obtained from the National Climatic Data Center (NCDC) for the 13 NWS stations shown on Figure 6-3. Hourly observations of total and opaque cloud cover were available for the period from 1948 through 1983, with the exception of the period from 1964 through 1980, when only 3-hour data were recorded. It should be noted that thin cirrus or stratus clouds are difficult to detect on satellite images, depending on the albedo of the earth's surface under the clouds. Therefore, satellite-derived cloud-cover data are expected to correlate better with opaque cloud cover than with total cloud cover. The following analyses were undertaken to verify the usefulness of the LANDSAT data and to extrapolate the 9:30 AM observations to each hour of the day.

We first set out to determine if cloud-cover data for 1973 through 1983 were representative of long-term mean cloud cover for the Southwest. Although it appears that cloud cover has been increasing at some locations over the last 6 to 8 years, the average change over all of the 13 stations used for this study is less than two hundredths. Therefore, the cloud cover-data from LANDSAT should be very representative of long-term means.

As previously noted, in order to restrict the indication of cloud cover to one digit, both nine-tenths and ten-tenths were designated as nine-tenths. In order to correct the nine-tenths cloud-cover recordings, the frequency distribution of surface observations of cloud cover between 9:00 and 10:00 in the morning were determined for each month of the year for each location. Figure 6-4 through 6-7 show histograms of hourly opaque cloud cover for Albuquerque, N. Mex., Fresno, Calif., Los Angeles, Calif., and Salt Lake City, Utah. These are typical of the results for all months and all of the stations in the Southwest. Although this analysis was done only to determine the relative frequency of occurrence of nine-tenths and ten-tenths cloud cover, the results were far more interesting than expected. It is particularly significant to note the preponderance of clear and overcast cloud conditions at virtually all locations for all months of the year. Exceptions occur at locations such as Albuquerque, where cumulus clouds occur frequently in the summer, resulting in frequent observations of one- to five-tenths cloud cover.

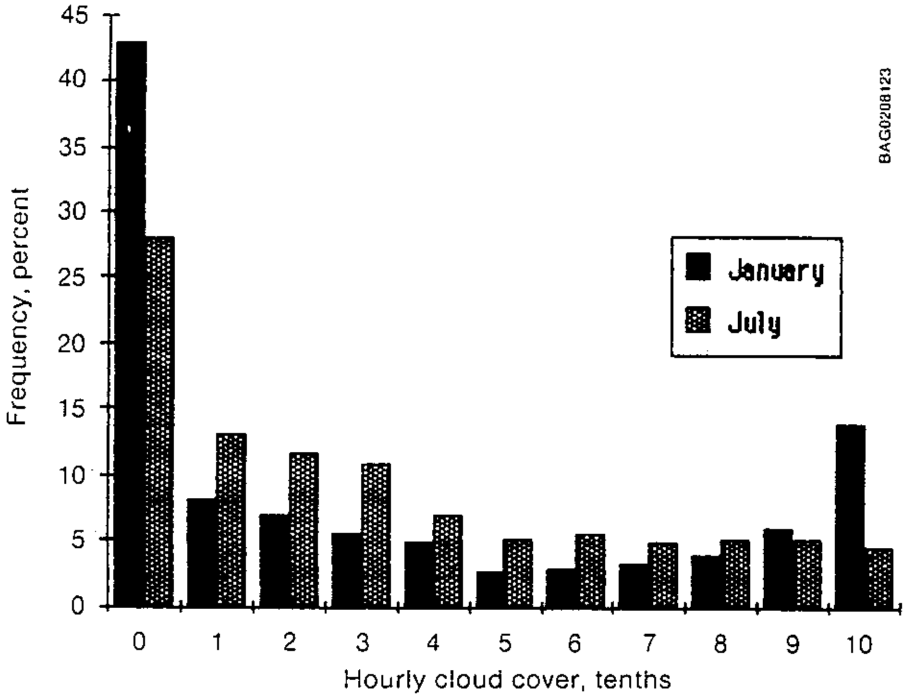


Figure 6-4. Histogram of Hourly Opaque Cloud Cover at Albuquerque

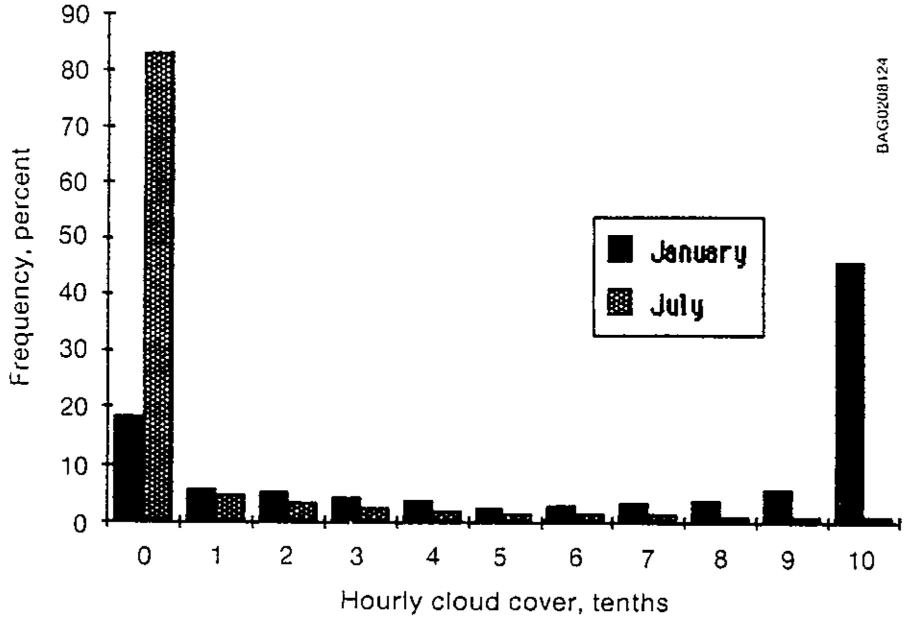


Figure 6-5. Histogram of Hourly Opaque Cloud Cover at Fresno

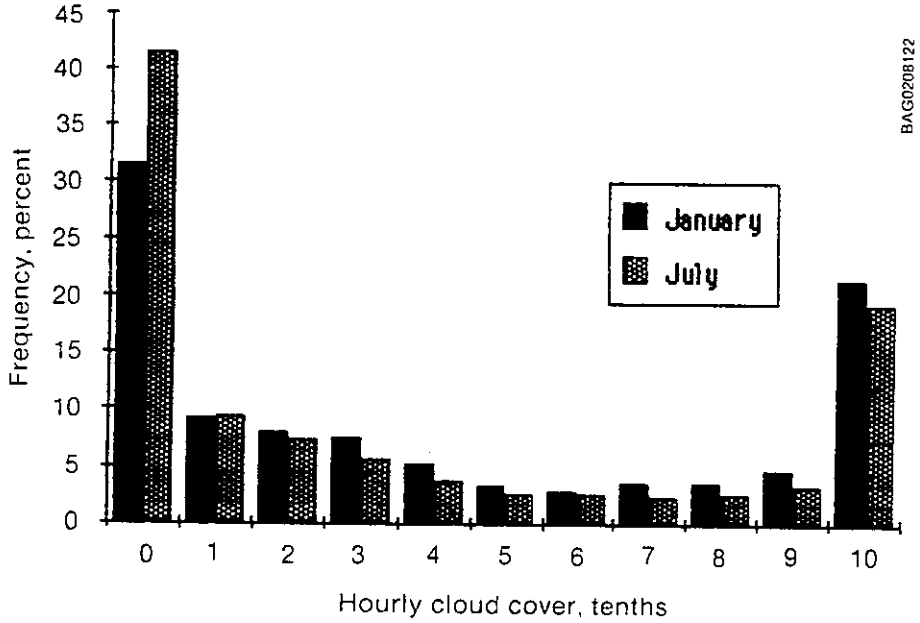


Figure 6-6. Histogram of Hourly Opaque Cloud Cover at Los Angeles

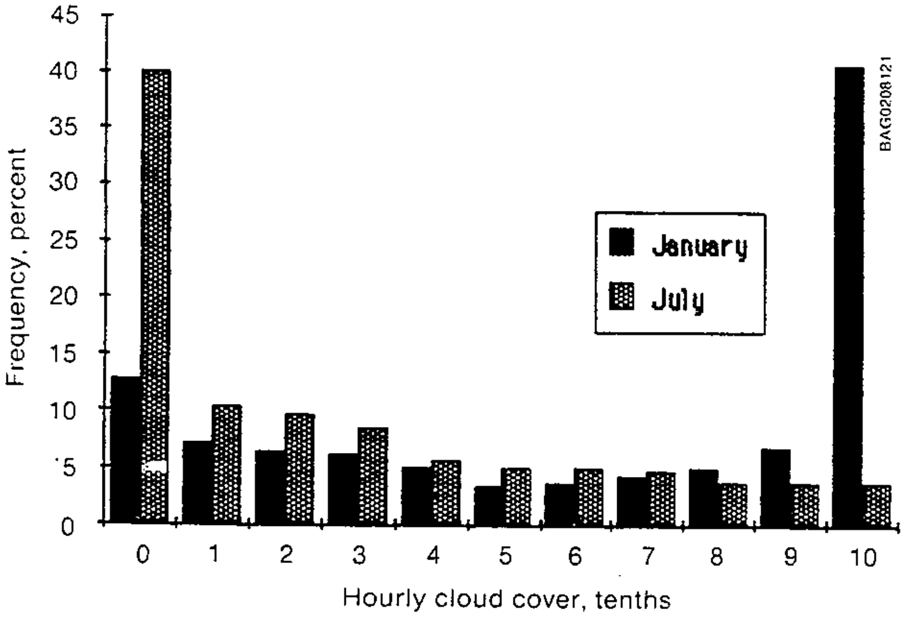


Figure 6-7. Histogram of Hourly Opaque Cloud Cover at Salt Lake City

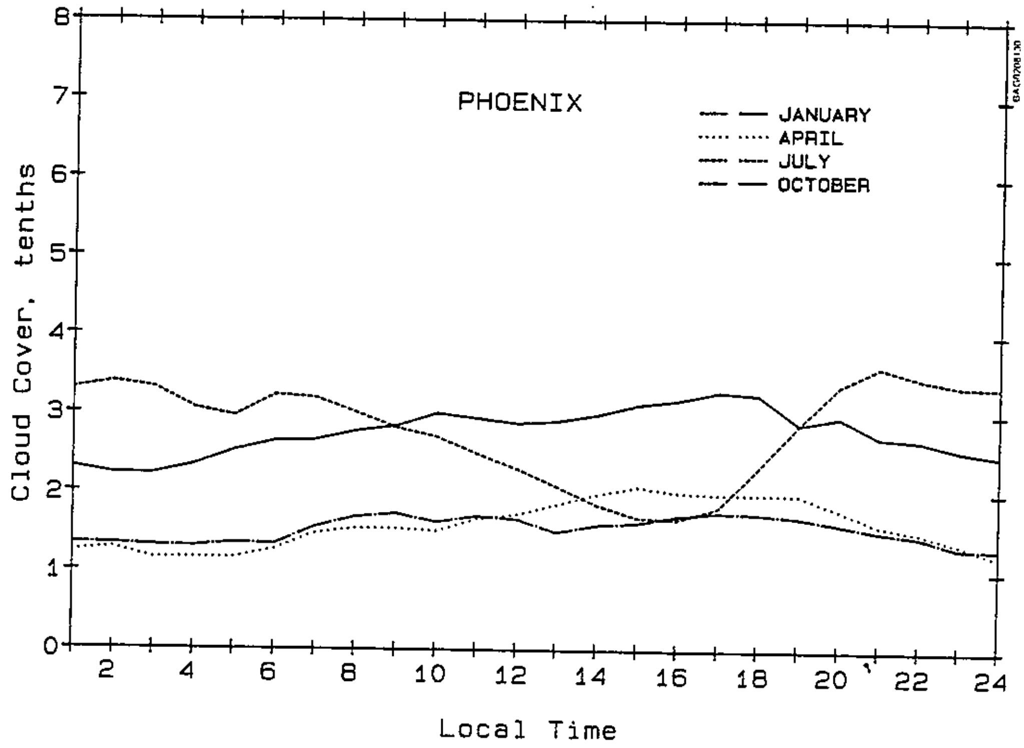


Figure 6-10. Diurnal Variation of Opaque Cloud Cover at Phoenix

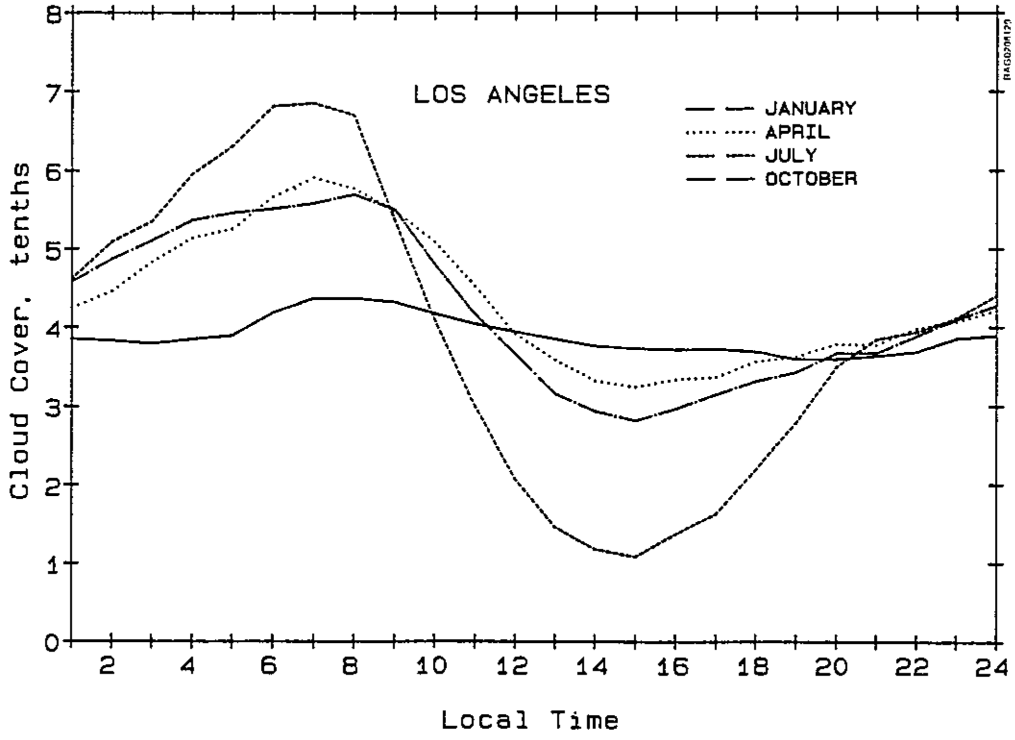


Figure 6-11. Diurnal Variation of Opaque Cloud Cover at Los Angeles

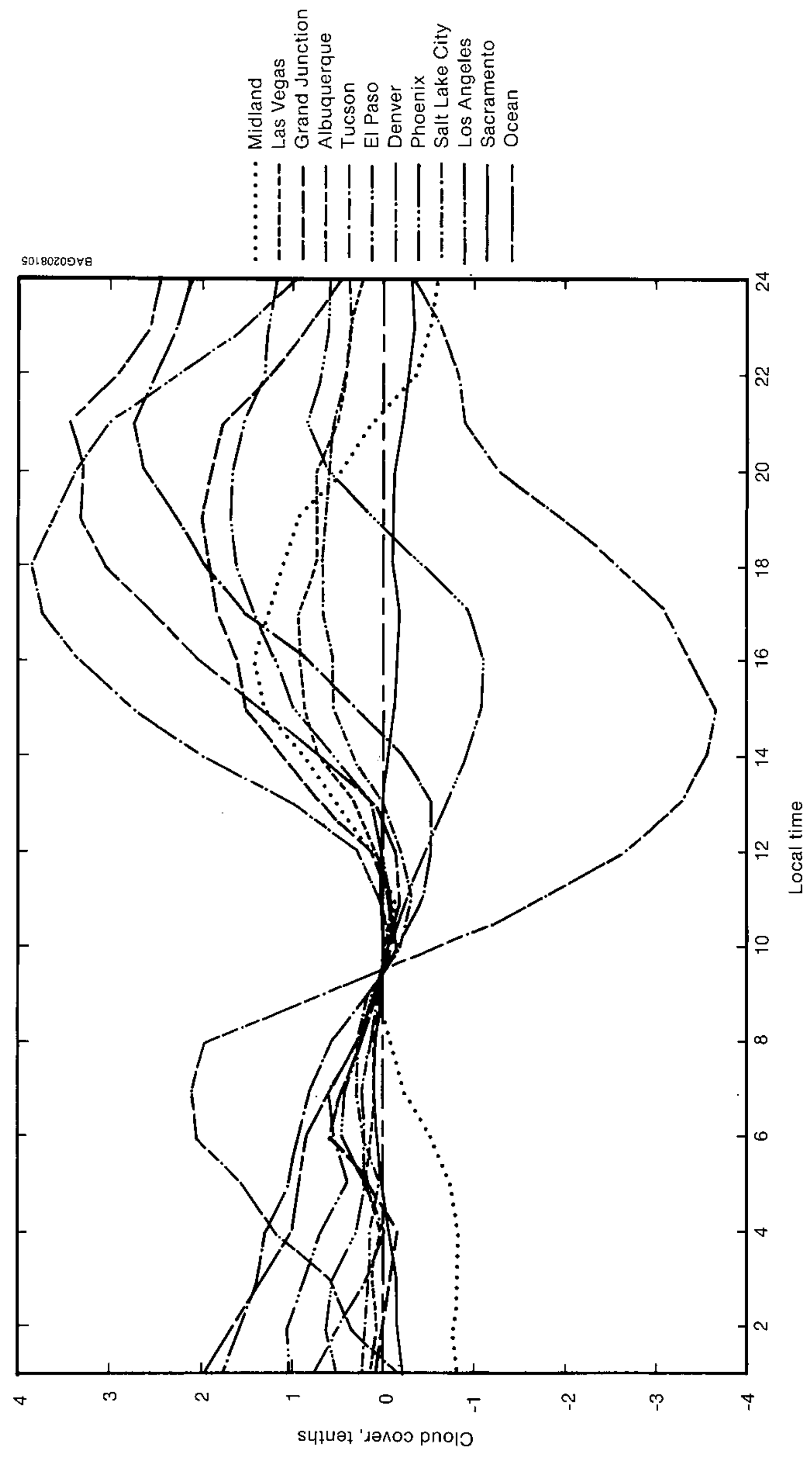


Figure 6-12. **Diurnal Variation of Opaque Cloud Cover for Eleven Southwest Sites--Normalized to 9:30 AM**

At first one might think that the observers were biased and recorded 0 cloud cover and overcast conditions more frequently than they actually occurred. Upon examining the results for many stations at all times of the year, e.g., Albuquerque and Fresno in July, one must conclude that this was not the case.

These results are particularly significant relative to the availability of solar energy resources. For example, at Fresno and Salt Lake City it appears that the sky will be overcast for approximately 50% of the time during the winter. Since these results are for opaque cloud cover, one would have to be concerned about the persistence of these overcast conditions when designing a solar energy system. It is obvious that one should be concerned about more than average cloud cover when designing a solar energy system for any location.

Extrapolating LANDSAT Observations to Other Hours of the Day. The surface observations for the 13 National Weather Service stations from the Southwest were used to obtain diurnal cloud-cover patterns such as those shown in Figure 6-8 through 6-11. These four stations exhibit the major patterns found in the Southwest. Data for Denver exhibit the effects of frequent thunderstorms in the summertime; hence, the buildup of cloud cover in the afternoon. At Fresno the diurnal variation of cloud cover is essentially the same all year. Phoenix shows the effect of the monsoons in July, where clouds build up during the evening hours and then are dissipated during the day, reaching a minimum in July around 4:00 PM. Los Angeles illustrates the typical pattern found along the West Coast during the summertime, when radiative cooling produces clouds during the nighttime and then burn off during the day. It is interesting to note that the diurnal variation of cloud cover during the winter months is essentially the same at all locations, indicating very little variation during daylight hours. It is very likely, in fact, that the decrease in cloud cover during nighttime hours may be more a matter of the difficulty experienced in seeing clouds under nighttime conditions. Seasonal variations in cloud cover are also apparent from these figures.

The diurnal cloud-cover patterns for all of the 13 stations in the Southwest were normalized by calculating the average cloud cover at 9:00 AM and 10:00 AM and then subtracting this mid-morning cloud-cover value from the values for all hours of the day. Figure 6-12 shows the resulting normalized diurnal variation of cloud cover for 11 of the 13 stations for the month of July. A LANDSAT observation at 9:30 AM at similar locations can now be added to these normalized data to produce cloud cover for each hour of the day. This procedure could not be used to obtain cloud-cover values for a specific day, but it works well to obtain monthly means for a 10-year period. This procedure could also be used to extrapolate cloud cover from the twice daily overpasses of NIMBUS satellites.

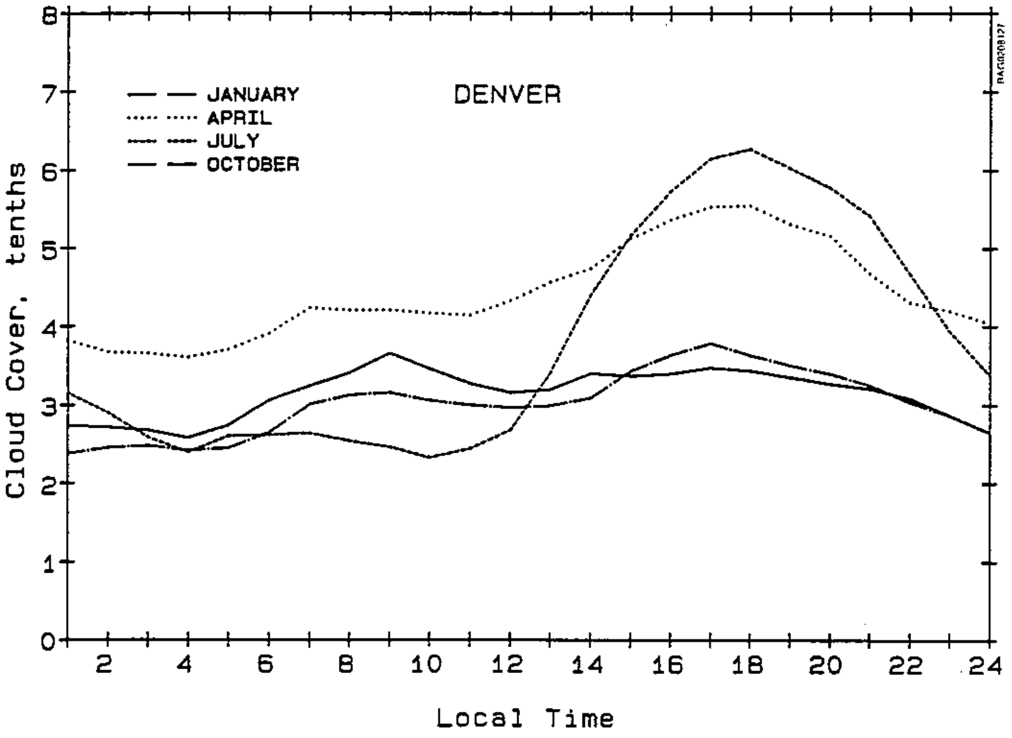


Figure 6-8. Diurnal Variation of Opaque Cloud Cover at Denver

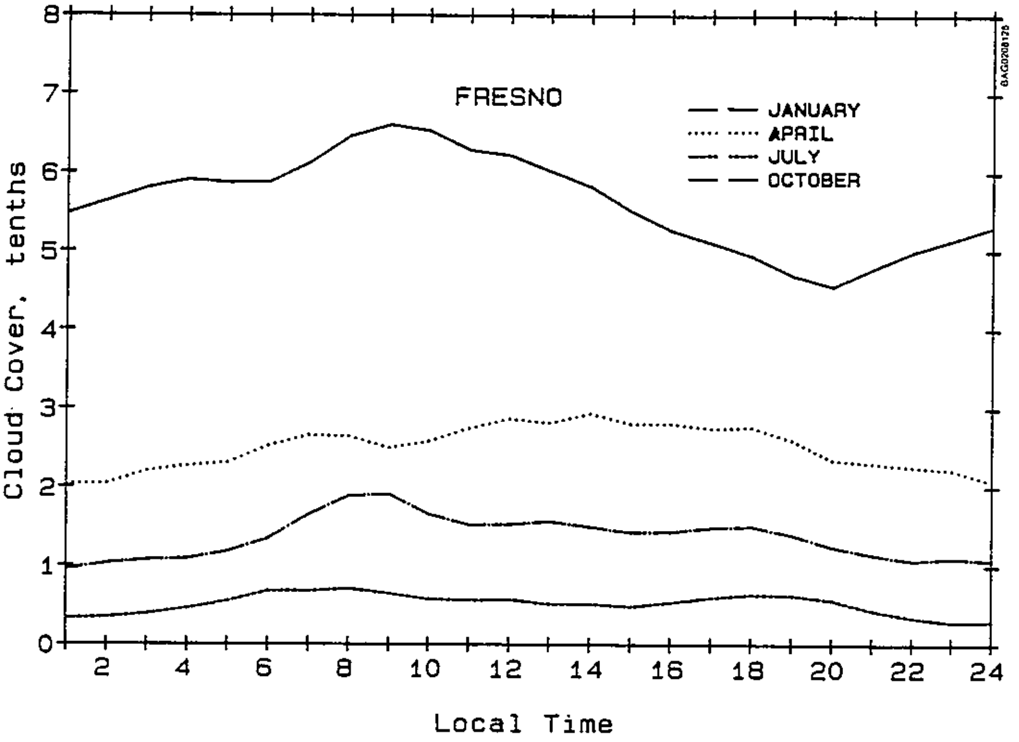


Figure 6-9. Diurnal Variation of Opaque Cloud Cover at Fresno

Next, from an examination of thunderstorm data (Changery 1981) and the topography of the region, the cloud climate zones shown in Figure 6-13 were formed. The zones were formed around the locations of NWS stations (for which diurnal patterns had been obtained) and the LANDSAT scene centers. Diurnal cloud-cover patterns for each zone were established. For example, Zone 13 encompasses Denver and the Rocky Mountain region, within which the number of thunderstorms and their time of initiation are similar to Denver. Researchers assumed, therefore, that the diurnal variations of cloud cover in Denver could be applied to all locations within that zone. Zone 17 encompasses the region immediately east of Denver and a large region to the south, all of which have about the same number of thunderstorms during the summer and within which the time of initiation of thunderstorms is about 2 hours later than that in Denver. For Zone 17, the diurnal pattern of cloud cover for Denver (see Figure 6-8) was shifted by two hours (to peak later) and was decreased in amplitude by a factor of 0.77. The Denver diurnal pattern was also used for Zone 18, where it was shifted by 4 hours, and its amplitude was decreased by a factor of 0.71. The changes in amplitude were based on a relationship between thunderstorm frequency and opaque cloud cover. Similar analysis were used to establish the diurnal patterns of cloud-cover occurrence for all of the climate zones in Figure 6-13. Table 6-2 gives the source of the diurnal cloud-cover patterns for each of the cloud climate zones and indicates the time and amplitude shifts that were employed. The diurnal variation of cloud cover for Zone 20, the ocean, was assumed to be flat. No data were readily available, but this seemed like a reasonable assumption.

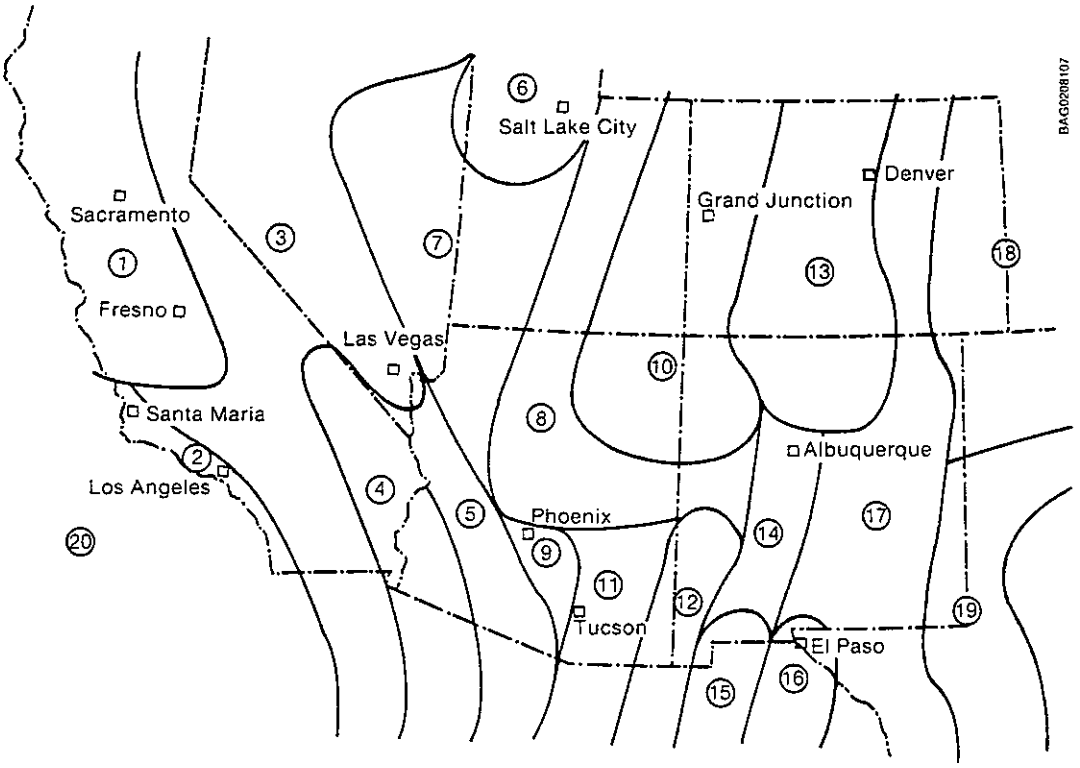


Figure 6-13. Cloud Climate Zones for the Southwest—July

Table 6-2. Summer Cloud Climate Zones in the Southwest (see Figure 6-13) and Diurnal Pattern Modifiers

No.	Climate Zone	Reference Station	Amplitude Multiplier	Time Shift
1	San Joaquin Valley	Sacramento	1.00	0
2	So. Calif. Coast	Los Angeles	1.00	0
3	Nevada/Calif. Desert	Las Vegas	1.00	0
4	Yuma, Ariz.	Las Vegas	0.89	+2
5	Cila Bend, Ariz.	Phoenix	0.72	+1
6	Salt Lake Desert	Salt Lake City	1.00	0
7	Great Basin	Las Vegas	1.27	-3
8	Utah/Arizona Mtns.	Crand Junction	1.22	-3
9	Phoenix, Ariz.	Phoenix	1.00	0
10	Four Corners Region	Grand Junction	1.00	0
11	Tucson, Ariz.	Tucson	1.00	0
12	Tombstone, Ariz.	Tucson	1.34	-1
13	Central Rocky Mtns.	Denver	1.00	0
14	Rio Grande Valley	Albuquerque	1.00	0
15	Ciudad Juarez	El Paso	1.17	+1
16	El Paso, Tex.	El Paso	1.00	0
17	Colo./New Mex. Plains	Denver	0.77	+2
18	Colo./Kansas Border	Denver	0.71	+4
19	West Texas	Midland	1.00	0
20	Ocean	None	--	--

A map of mean opaque cloud cover (sunrise to sunset) in July for the Southwest, derived from LANDSAT scene listings using the procedures described above, is shown in Figure 6-14. Similar maps for 9:00 AM and 6:00 PM were compared with surface observations of opaque cloud cover at the NWS stations. Although some variations were seen, no large anomalies were noted. Researchers decided, therefore, that no modifications of the data would be made. The LANDSAT grid of cloud-cover data were smoothed using a two-dimensional Gaussian filter, and the results were used to prepare the map. The close agreement of the data with opaque cloud cover rather than total cloud cover is not surprising, since thin cirrus and stratus clouds are hard to detect on the LANDSAT images.

6.1.5 Converting Cloud Cover to Direct Normal Irradiance

The July hourly opaque cloud-cover data for the Southwest were converted to hourly direct normal irradiance values using a modified Solar Irradiance Model (SIM) (SERI 1980). The original model divided the day into 10 segments of varying length, depending on percent sunshine and/or cloud-cover data. The modified SIM model divided each hour into 60 1-minute segments. The opaque cloud-cover value for the hour was used to determine the number of minutes the sun was obscured by clouds, and the hourly direct normal irradiance was calculated according to the equation

$$I_N = c \sum_{i=1}^{n} I_0 \exp(-\tau \cos\theta) \qquad (6\text{-}2)$$

where

$n = 60(1 - Cq)$ and Cq = opaque cloud cover

c = clearness number

I_0 = direct normal extraterrestrial solar irradiance in W/m^2

τ = the optical depth of a typical atmosphere

θ = solar zenith angle

I_N = hourly direct normal irradiance in W/m^2.

For example, if the opaque cloud cover were 0.7, 18 of the 1-minute segments would be unobscured ($n = 18$) and 18 computations would be distributed evenly over the hour and summed to determine the direct normal insolation for the hour. The hourly values were then summed up over the day to obtain the daily average direct normal irradiance for July for each LANDSAT scene center shown in Figure 6-3.

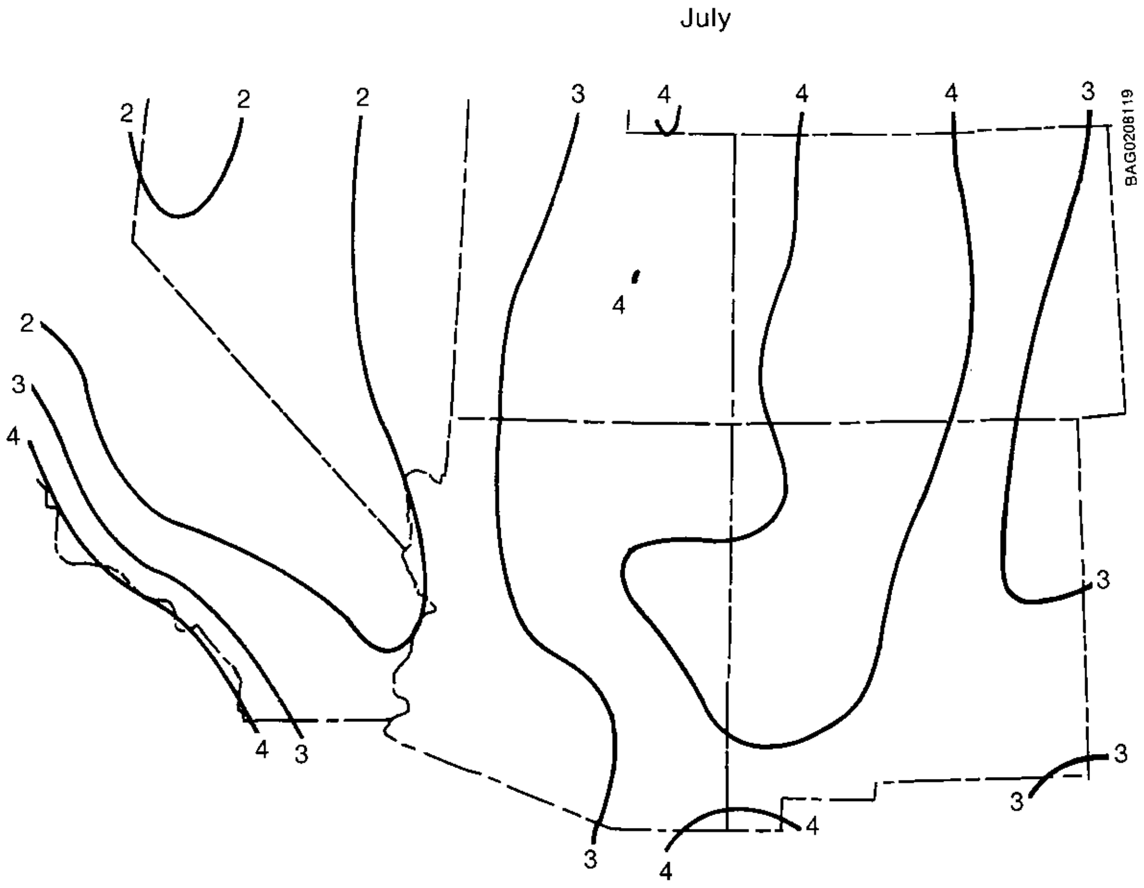


Figure 6-14. Mean Opaque Cloud Cover (tenths) in July for the Southwest (From 10 years of LANDSAT scene analyses)

Clearness numbers (an empirical number used to account for regional variations in atmospheric turbidity) were assigned to each LANDSAT scene location using data from Threlkeld and Jordan (1958) and SERI (1980). Equation 6-2 was then used to compute the mean daily total direct normal irradiance for the month of July. These initial computations were then compared with measured data obtained at NOAA and WEST Associates stations during the 1977 to 1980 time period. The initial values appeared to be about 10% to 15% high, so the clearness numbers were reduced by that amount. In effect, an across-the-board adjustment of clearness numbers was needed to bring the satellite/model values into agreement with surface measurements. No attempt was made to obtain exact agreement with measurements at specific locations. Rather, regional trends were noted and clearness numbers for the entire Southwest were adjusted accordingly.

For a number of reasons, it was necessary to smooth the calculated insolation data during the contouring and map-making process. As noted earlier, the LANDSAT cloud cover data provided a maximum of 30 and an average of 23 samples over the 10-year period used for this project. This is marginal for determining a mean cloud-cover value. Furthermore, the compositing process required the assignment of a solar radiation value to each 10-mi.^{2} cell used in digitizing the other variables. The smoothing procedure established values of insolation for each cell. A Gaussian filter function with a radius of 150 km around each cell was used. The resultant map of direct normal insolation is shown in Figure 6-15.

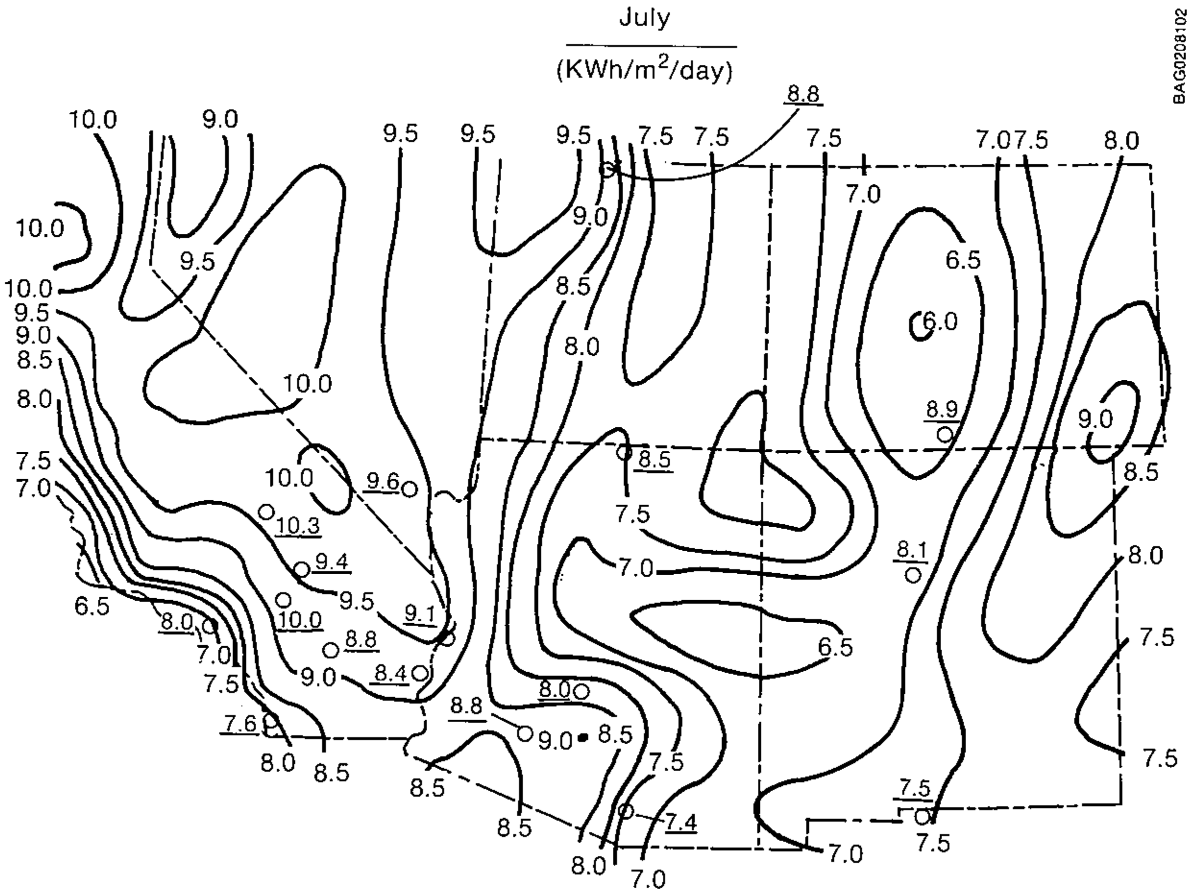


Figure 6-15. **Mean Daily Total Direct Normal Insolation in July for the Southwest [Modeled using LANDSAT-derived cloud cover; regionally adjusted to surface measurements (underlined)]**

Two types of comparisons were made with measured data. First, for those locations where a LANDSAT scene center was very close to a measurement station (within 30 km), a comparison was made between the nonfiltered computed values and the measurements. Table 6-3 gives the results of these comparisons. Because of the uncertainty associated with the LANDSAT cloud-cover data and the small number of samples available for the measured data (1 to 4 years of data), only two significant figures are given for both measured and calculated values.

Table 6-3. Comparison of Measured and Calculated Direct Normal Insolation for July (Wh/m^2/day)

Location	Measured	Calculated	Error (%)
Alamosa, Colo.	8900	6400	-28
El Paso, Tex.	7500	7500	0
Las Vegas, Nev.	9600	9900	3
Phoenix, Ariz.	8000	9200	15
Gila Bend, Ariz.	8800	8600	- 2
Ridgecrest, Calif.	10300	9700	- 6
Victorville, Calif.	10000	9500	- 5

With the exception of Alamosa, Colo., and Phoenix, Ariz., these results are very good, well within the uncertainty of the measurements, considering the few years for which measured data were available. An analysis of the results for the two locations exhibiting high errors showed that these differences can be explained. For instance, Alamosa, Colo., is located in the middle of the San Luis Valley. The Sangre de Cristo Mountains are less than 50 km to the east and the San Juan Mountains are less than 50 km to the west. Therefore, since the center of the LANDSAT scene at Path 36, Row 34 is almost congruent with Alamosa, the scene boundaries include the mountain ranges on both sides. As anyone who has visited Colorado in the summer knows, clouds occur over the mountains with a much higher frequency than over the plains or wide valleys like the San Luis Valley. This is apparent from Figure 6-16, which shows a portion of a GOES image. The approximate location of LANDSAT Scene 36-34 is shown along with the location of Alamosa. It is not surprising that the average cloud cover for Scene 36-34 is higher than the cloud cover observed at Alamosa. This in turn results in the estimate of insolation being low compared to that measured at Alamosa.

The situation at Phoenix is similar. The elevation of the terrain northeast of Phoenix increases rapidly from about 300 m at Phoenix to 2000 m on the Colorado Plateau. This results in a high gradient in the occurrence of clouds. Given this high gradient, the cloud cover recorded for LANDSAT Scene 39-37 could have been higher or lower than that observed at Phoenix, depending on the exact nature of the variations occurring within the scene. Based on the comparisons of the calculated insolation for Scene 39-37 and the measured insolation at Phoenix, one must assume that the average cloud cover over the scene is less than that observed at Phoenix.

A comparison between measured and smoothed calculated data can be made by examining Figure 6-15. Measured values are given for each of the locations shown. Data for Grand Junction, Colo., Ely, Nev., and Fresno, Calif., are not included because the poor data collection record at these stations resulted in a small sample size and data of questionable quality. The sites where poor agreement exists between measured data and the smoothed mesoscale map can all be related to the need for better spatial resolution in mountainous or coastal locations. For example, Albuquerque

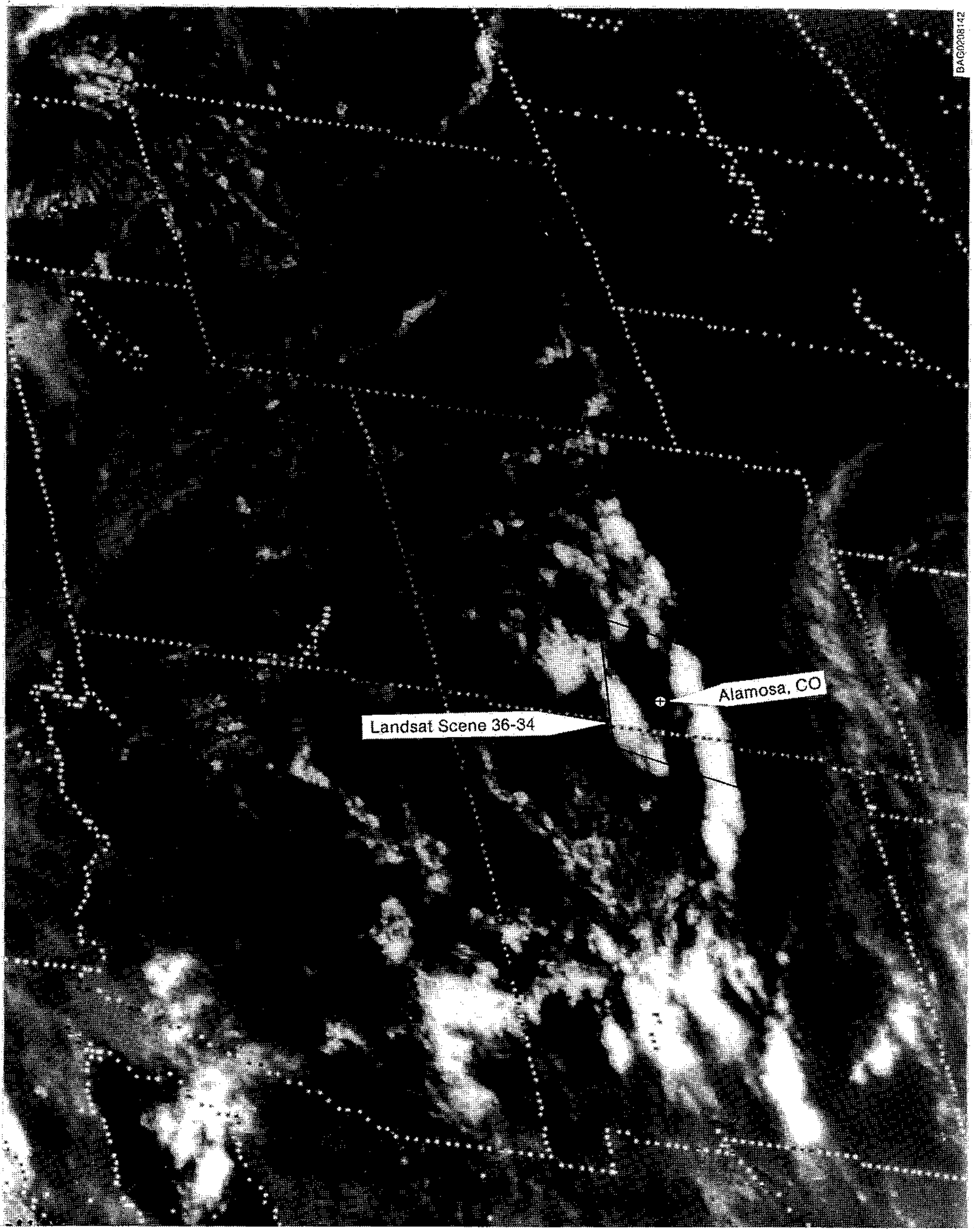


Figure 6-16. GOES Image Showing Typical Cloud Cover Surrounding the San Luis Valley in Southern Colorado

lies between LANDSAT Scenes 36-35 and 36-36. Scene 36-35 encompasses the southern extensions of both the Sangre de Cristo and San Juan Mountains, and Scene 36-36 encompasses mountains and high plateaus to the southwest of Albuquerque, which experience a high frequency of thunderstorms. Albuquerque itself lies along the Rio Grande River and experiences less cloud cover than most of the surrounding higher elevation plateaus and mountains.

Page, Ariz., lies between Scenes 40-34 and 40-35 and is situated at the Clen Canyon Dam. The situation is similar to Albuquerque's, since the Wasatch Range lies north of Page and the Kaibab and other high-elevation plateaus lie to the south. An examination of weather satellite images during the summer shows a lower frequency of cloud cover at Page than 0ver the surrounding terrain.

Obviously, the photointerpretation of satellite images (GOES, NIMBUS, LANDSAT) on a point-by-point grid could achieve much greater resolution and much better agreement between measured and modeled data. The combination of satellite-derived cloud-cover data, model calculations, and macroscale measured data (to make regional adjustments of absolute values) could be used to achieve more accurate and much higher resolution maps of solar radiation resources. A comparison of these results with the *Solar Atlas* (SERI 1981) show that these scene interpretations of LANDSAT images have made significant improvements, which are adequate in areas of low cloud-cover gradients.

Comparisons with SOLMET Data. The map shown in Figure 6-17 was taken from the *U.S. Solar Atlas* (SERI 1981). The maps in the atlas were developed from the SOLMET national insolation data base (SOLMET, Vol. 1). These data were rehabilitated from historical global horizontal measurements made from 1953 to 1975. The direct normal data in these data bases were all derived from regression models based on less than 2 years of data at five sites (SOLMET, Vol. 2). These data bases have been the primary source of solar radiation resource information for the nation since they were completed in 1978.

A comparison of the maps in Figures 6-15 and 6-17 show some regions of good agreement, such as northern Nevada, and some regions of serious discrepancies. For example, the atlas map shows a value near Sacramento, Calif., of about 8.9 kWh/m^2/day, whereas the mesoscale map shows a value of about 11.0 kWh/m^2/day. This indicates that the atlas values in this region are low by as much as 20%. The two maps also show significant differences in Colorado. The mesoscale map developed during this project shows a high level of insolation in southeastern Colorado, whereas the lowest values in Colorado occur in the central Rocky Mountains. These regional patterns are in good agreement with weather observations in Colorado. The atlas map, on the other hand, indicates that the lowest insolation in Colorado occurs along the eastern border, whereas the highest value is found in northwest Colorado. This is totally inconsistent with the weather patterns observed in Colorado.

One must conclude, therefore, that the mesoscale direct normal insolation map shown in Figure 6-15 represents a significant improvement in the accuracy of the assessment of solar radiation resources. The improved spatial resolution is also of considerable significance, although further improvements are needed for coastal and mountainous regions.

6.1.6 Evaluating Regional Site Suitability

The methodology used to determine regional site suitability was based on a cost-benefit analysis. Much of the cost data was estimated, however, so this should be considered a demonstration of methods, not an actual site selection process.

The Compositing Process. An overview of the compositing process is shown in the diagram of Figure 6-18. The mesoscale map of solar radiation resources, created from a combination of surface measurements and modeled data based on cloud-cover values derived from satellite images, is the most important single source of information.

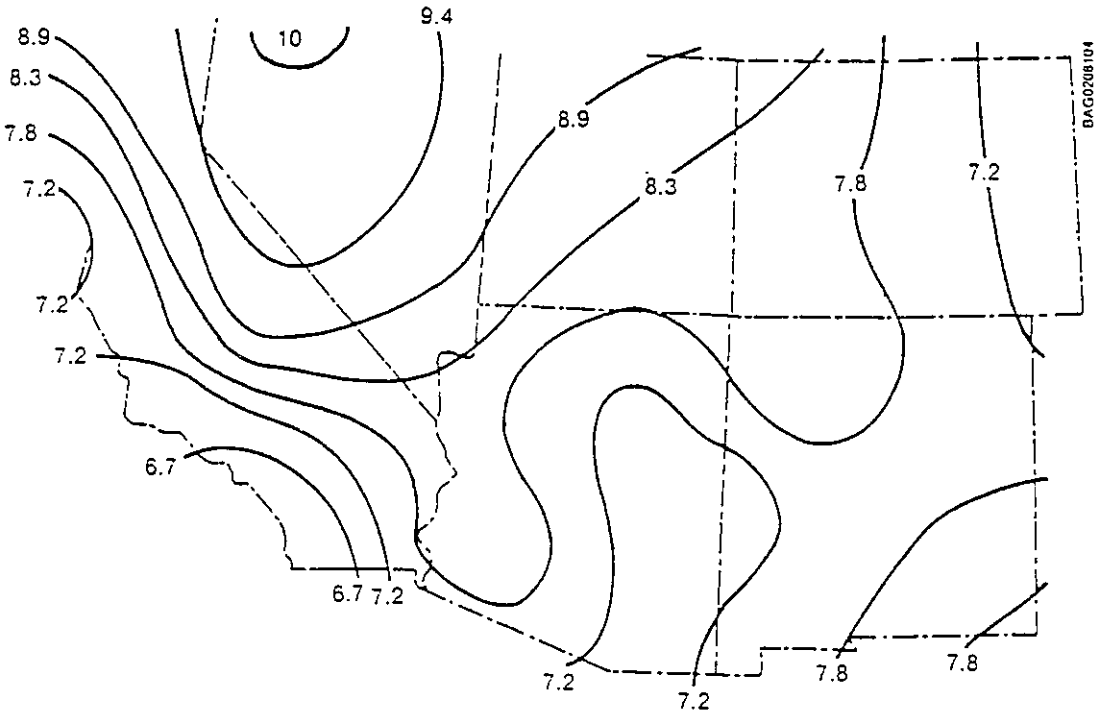


Figure 6-17. Mean Daily Total Direct Normal Insolation in July for the Southwest (From the *Solar Atlas*, 1981)

The distance constraints were calculated from estimates of electrical energy demand using Eq. 6 - 1 and a radial integration of the product of energy demand with population. This procedure determined the average distance over which the entire plant output could be dissipated.

Land suitability is a composite of land use and land management. The land values in Table 6-4 were assigned to each of the land-use types and these were multiplied by the land-management factors in Table 6-5, which were assigned to each of the land-management types. These products were used to estimate land suitability.

Table 6-4. Amortized Cost of Land over a 20-Year Period (Rough estimates)

Land Use	$/Ha
Urban	$300,000
Mining	50,000
Irrigated Ag.	25,000
Dry Land Ag.	10,000
Poor Rangeland	3,000
Mixed Rangeland	6,000
Good Rangeland	10,000
Forest Lands	6,000
Lakes and Reservoirs	∞
Wetlands	50,000
Barren Lands	3,000
Snow and Alpine	50,000

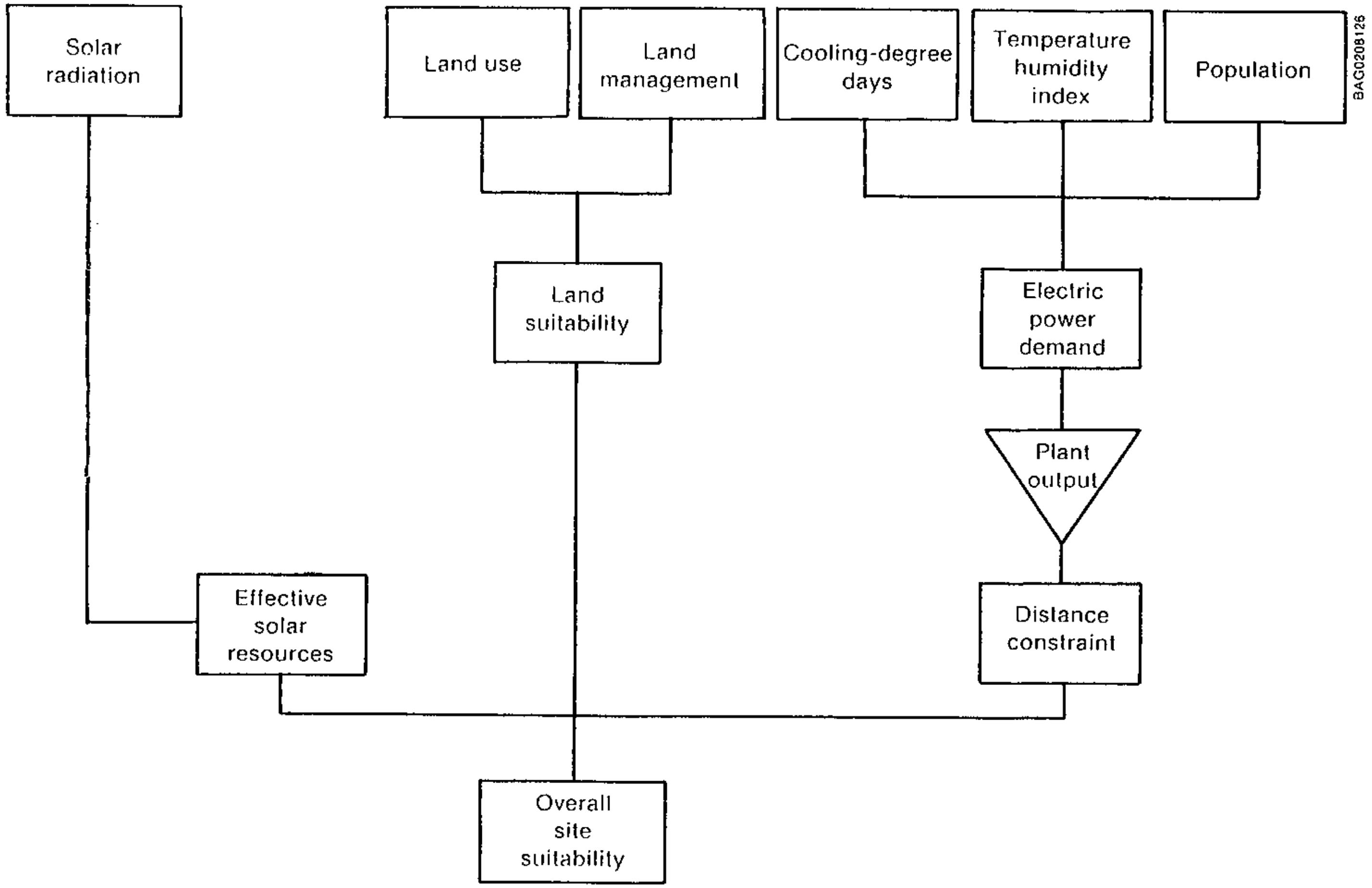


Figure 6-18. Overwiev of the Map Compositing Proces Used to Assess Regional Suitability of Sites for a Solar Thermal Power Plant

Table 6-5. Land-Management Multiplication Factors

Land Management	Multiplier
Urban	10
National Forest	2
Indian Reservation	2
Nat. Parks and Mon.	∞
Military Reser.	4
National Wildlife Ref.	∞
Private, State, Public	1

Defining the Siting Algorithm.

As noted earlier, the siting algorithm is based on a cost-benefit approach. In order to equate costs and benefits, it was necessary to estimate the value (V) of the energy produced per hectare of land over a 20-year plant lifetime. This was calculated according to the expression,

$$V = I \cdot (10{,}000\ \mathrm{m^2/Ha})\ e\ (20\ \mathrm{yrs} \cdot 365.\ 25\ \mathrm{days})\ R \tag{6-3}$$

$$= I \cdot 2.92\ 10^5\ (\$/\mathrm{ha})\ ,$$

where

I is the direct normal insolation in kWh/m^2/day

e is the efficiency factor

ha is area in hectares (10,000 m^2)

R is the electrical energy rate in \$/kWh, set at \$0.10 for this example.

The overall efficiency (e) of a solar power plant was defined as

$$e = E/I\ , \tag{6-4}$$

where E is the electrical energy output in kWh/m^2 (based on the collector field area) and I is the incident solar insolation in kWh/m^2. A typical arrangement of collectors for a solar power plant covers about 20% of the collector field. Using a generous 20% efficiency for the solar collectors, we arrived at a 4% efficiency factor, as defined above.

The first cost to be calculated is the cost of the land on which the power plant will be built. This cost is obtained from the composite of land use and land management as described above, i.e.,

$$L = \text{land value x land mgt. factor } (\$/\mathrm{ha}), \tag{6-5}$$

where rough estimates of land value, amortized over 20 years, are those found in Table 6-4 and the land management multiplication factors are those found in Table 6-5. This algorithm is used to estimate the actual cost of the land required for a power plant, but it is also used to eliminate certain land categories from further consideration. For example, lakes and reservoirs are eliminated by assigning an infinite cost to this land use. Similarly, National Parks and Monuments and National Wildlife Refuges are eliminated from consideration for power plant sites. This is not unrealistic, because it would probably be very costly to pass a bill through Congress to open up these lands for such use, a cost many times greater than the actual value of the land. Similarly, inflated costs have been assigned to wetlands and snow and alpine to account for the excessive costs that would be involved in constructing a power plant at these locations.

The other cost factor considered is related to the cost of transporting the energy from the power plant site to the loads. We have previously described the method employed to determine the average distance over which the plant output would be transmitted. This is defined as a distance constraint. The average transmission costs are estimated to be \$0.03/kWh/1000 km. In order to express these costs in \$/ha, such that they can be combined with the value of the electricity produced and the cost of the land, it was necessary to know how much electricity would be produced over a 20-year period for each hectare of land employed for the power plant.

Using all but the R term in Eq. 6-3 we arrive at the expression

$$\text{kWh/ha} = I \cdot 2.92\ 10^6$$

for calculating 20-year plant production per hectare. The expression for calculating transmission costs then becomes

$$T = 0.03 \cdot I \cdot 2.92\ 10^6 \cdot d \tag{6-6}$$

$$= I \cdot 8.77\ 10^4 \cdot d\ (\$/\text{ha})\ ,$$

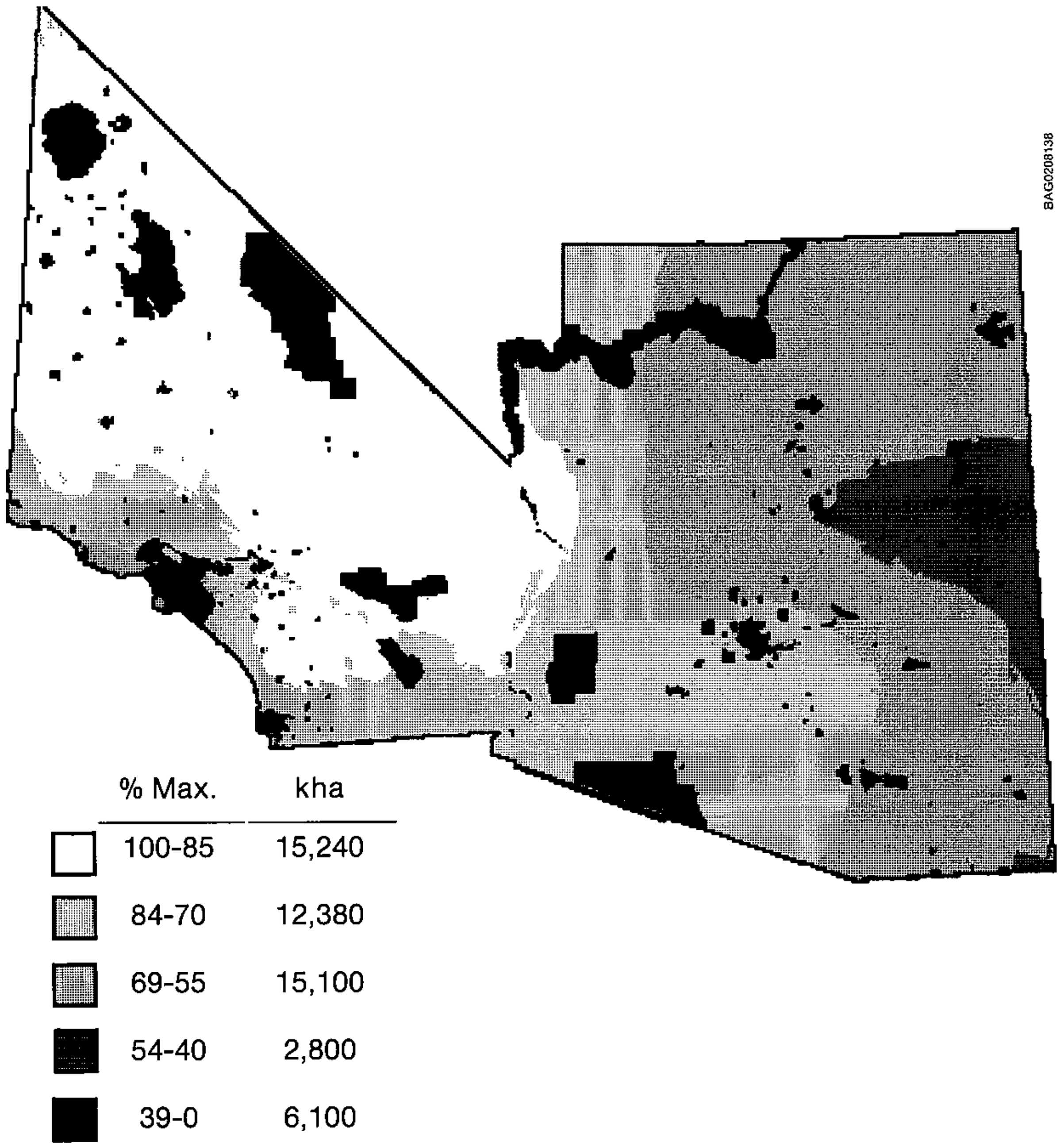
where d is the average distance (in Mm) over which the plant output will be transmitted.

The final algorithm for computing the suitability (S) of one location relative to another is given by

$$S = V - L - T\ . \tag{6-7}$$

This algorithm was used to prepare the overall suitability map. Figure 6-18 indicates that a map of effective solar resources was combined with maps of distance constraints and land suitability to create a map of overall site suitability.

The different gray levels on Figure 6-19 represent changes in suitability of 15%, with the exception of black, which designates areas that have a suitability rating less than 40% of the maximum. The isolated, smaller black areas are the national parks and monuments, lakes and reservoirs, and urban areas that were essentially eliminated from consideration on the basis of land-use and land-management criteria. The total area of each of the five levels of suitability was computed and is indicated in hectares. It must be emphasized that these results are based on solar radiation for July and the consideration that the primary objective for the power plant was to offset peak summertime loads associated in part with air conditioning.



	% Max.	kha
☐	100-85	15,240
░	84-70	12,380
▒	69-55	15,100
▓	54-40	2,800
■	39-0	6,100

Figure 6-19. **Overall Site Suitability for a Solar Power Plant. (July performance only)**

6.1.7 References

Changery, Michael J., 1981, National Thunderstorm Frequencies for the Contiguous United States, NUREG/CR 2252-RB, R6 for the U.S. Nuclear Regulatory Commission, Asheville, NC: National Climatic Data Center.

Maxwell, E. L., A. G. Folger, and S. E. Hogg, 1985, Resource Evaluation and Site Selection for Microalgae Production Systems, *SERI/TR-215-2484*, Golden, CO: Solar Energy Research Institute.

Maxwell, E. L., and M. D. Rymes, 1987, An Initial Development of a Methodology for Siting Solar Power Plants, *SERI/TR-215-3117*, Golden, CO: Solar Energy Research Institute.

McHarg, I. M., 1969, Design with Nature, Garden City, NY: Natural History Press.

Olea, R. A., 1974, "Optimal Contour Mapping Using Universal Kriging", *Journal of Geophysical Research*, Vol. **79**, No. 5, pp. 695-702.

SERI, 1980, SIM (Solar Irradiance Model) User's Manual, *SERI/TR-06022-1*, Golden, CO: Solar Energy Research Institute.

SERI, 1981, Solar Radiation Energy Resource Atlas of the United States, *SERI/SP-642-1037*, Golden, CO: Solar Energy Research Institute.

SOLMET, Vol. **1**, 1977, User's Manual -- Hourly Solar Radiation -- Surface Meteorological Observations, TD-9724, Asheville, NC: National Climate Center.

SOLMET, Vol. **2**, 1979, Final Report -- Hourly Solar Radiation -- Surface Meteorological Observations, TD-9724, Asheville, NC: National Climatic Data Center.

Threlkeld, J. L., and R. C. Jordan, 1958, "Direct Radiation Available on Clear Days", *ASHRAE Transactions*, Vol. **64**, p. 45.

Turner, A. K., 1976, "Computer Aided Environmental Impact Analysis, Part I", MER19 #2, Golden, CO: Colorado School of Mines.

6.2 Assessing Photovoltaic Interaction with New York Summer-Peaking Utilities -- Resource Assessment Logistics

Study Prepared by: *Richard Perez, Robert Seals and Ronald Stewart*
Atmospheric Sciences Research Center

Under Sponsorship from: *New York Power Authority*
Solar Energy Research Institute
The Mobil Foundation

6.2.1 Research Objectives

The objective of this study was to evaluate the interaction between photovoltaic (PV) power generation and the needs of utilities located in southeastern New York. Special attention was focused on evaluating the ability of existing solar resource assessment capabilities to address this question satisfactorily. These capabilities include insolation measurements performed from ground stations and remote sensing from geostationary satellite.

The central issue is the experimental determination of the load-matching potential of photovoltaics for four downstate New York summer peaking utilities: Long Island Lighting Company (LILCO) , Consolidated Edison (ConEd), the south eastern sector of the New York Power Authority (NYPA-SENY) and the Village of Freeport Municipal Utility.

The study spans two summer periods (1987 and 1988). A preliminary study had shown that, because of the unique sun-synchronous and feed-back relationships between commercial loads and solar power availability, photovoltaics had a substantial capacity value. The present results confirm and reinforce those of our preliminary findings. The consistency of results obtained throughout two complete summers is particularly encouraging.

Three benchmarks were used to quantify PV's load-matching potential: the effective load carrying capability (ELCC) of PVs, the relative value of energy produced and the minimum buffer storage needed to guaranty that peak loads are reduced by a given portion of installed PV capacity. For LILCO, effective load carrying capabilities of the order of 50% of the rated PV capacity were found for PVs making up to 10% of the utility's generation mix, that is 400 MW; at the 10 MW level the ELCC reaches 55%. PV energy was found to be time coincident with highest loads for each utility studied. Finally, a buffer storage of the order of one to one and a half hours was found to be sufficient to guarantee a firm reduction (up to 10%) of peak loads; by contrast, achieving equivalent peak shaving without PVs would require three to four times more storage.

The second issue is the evaluation of the accuracy of currently available satellite-derived insolation in southern New York State. Actual insolation data are needed for simulating PV power plants and investigating their interaction with utilities. Geostationary satellites constitute at present the only routinely available source of current insolation data, continuous over large geographical extents.

We first proceeded with a physical evaluation of satellite-derived insolation. The main findings are the following:

- At the hourly data level, the agreement between satellite estimates of global irradiance and ground-measured values was found to be consistent with other studies, with negligible overall bias (long term error) and a dispersion (short term error) of the order of ±20%. For the direct component, a much larger dispersion was found. We suggest that more appropriate methods be sought to directly convert the satellite signal into direct beam estimates.

- In climatological terms however, satellite predictions were found to be quite representative, both for global and direct irradiance; the ability of the satellite information to recreate typical hourly insolation conditions at the locations tested in this report was found to be remarkable; indeed, the satellite could well constitute a better source of radiative data than many of the current Typical Meteorological Year (TMY) sets.

- In terms of extended regional coverage, the satellite was found to be a better source of irradiance data than extrapolation from a nearby ground station if the distance between the considered location and the station exceeds 65 km (40 miles). In the case of a ground measurement network, the satellite's precision would better that of the interpolated groun measurements for points with a mean interpolation distance exceeding 95 km (60 miles) when using a moving average approach. The study also points to the fact that a satellite-ground network combination could yield both measurement precision and thorough aerial coverage.

Second, we assessed the accuracy of the satellite-derived insolation in terms of end-use viability, that is, in terms of providing answers to specific PV utility interaction questions. The end-use criteria are the three photovoltaic load-matching benchmarks with respect to the four utilities. Better agreement was found at this level than would have been expected from the above physical assessment of short term errors. The satellite data appear to meet the goal of determining whether PVs may effectively contribute to a given utility's generation capacity. It is likely that other utilities could use a similar approach to perform a preliminary assessment of PV load matching potential for their service area before addressing more specific issues through a ground measurement network.

6.2.2 Evaluation of GOES-derived irradiance data

6.2.2.1 Origin of Satellite-derived Irradiance Data

Several methods have been developed to make use of the earth-monitoring sensors aboard geostationary satellites to infer solar radiation at the earth's surface (e.g., [2,4,5,6]). These methods vary greatly in terms of complexity, from physical radiative transfer-based model to simpler statistical approaches. The simpler models rely only on the on-board visible radiation sensor, others also make use of infrared sensors. No comprehensive model inter-comparisons have been performed to date. However partial evaluations have shown that complex models did not perform substantially better than the simpler ones [7] as long as the simple models are first calibrated for a given geographical area.

Besides implementing sometimes complex transfer models, a major difficulty with current satellite models is that one must access the primary satellite data (digitized images), and perform tedious navigation calculations (as satellites are not absolutely fixed, the on-board coordinate system and the earth's surface coordinate do not always match and images have to be adjusted). Adjusted digitized satellite frames are available in the United States, but at a prohibitive price within the context of this study. Note that, as a low cost alternative, a European team [6] has developed a desktop package to readily produce satellite-derived irradiance from the analog images (WEFAX) sent via commercial satellite by the National Weather Service [8] (most local weather forecasting firms already receive this signal). This methodology was developed for the European geostationary satellite Meteosat. It could be extended for operation in our area.

For the present study, we were fortunate that GOES measurement of earth-emitted visible radiation (earth's brightness) are routinely processed by the National Oceanic and Atmospheric Administration (NOAA) into navigationally conformed global irradiance estimates at the earth's surface -- global irradiance is defined as the solar power density impinging a horizontal surface).

The satellite transfer methodology used by NOAA is one of the simpler statistical ones [2]: this correlates the observed brightness with inferred global irradiance; for a given solar geometry, lowest brightness levels correspond to cloudless situations hence highest irradiances; as brightness increases, due to cloud cover, irradiance is correspondingly reduced (snow cover on the ground is a potential source of error but this can be minimized [9]. The methodology was not developed for over-water applications, hence only over-land estimates are made available. Also, the satellite data is processed in such a way as to provide a geographical resolution of only 1° longitude by 1° latitude, that is a 70X95 km grid (45X60 miles). Each satellite grid point is an integrated value of 25 image pixels surrounding each grid point, amounting to a ground area of about 2600 sq. km (1000 sq. miles).

Figure 6.2.1 includes an artist rendering of the GOES satellite [8], showing the earth sensor (Visible-infrared spin-scan radiometer Atmospheric Sounder, V.A.S.) used to generate the ground insolation estimates.

6.2.2.2 Generation of Hourly Irradiance Data

The satellite-derived global irradiance estimates are provided by NOAA on an instantaneous basis (satellite snapshots) five to seven times daily.

A *key component of the work* performed for this project was *to generate hourly time series from these readings, and further, to split global irradiance into diffuse and direct components* suitable to perform photovoltaic simulations. Diffuse and direct (beam) irradiances are the solar power densities originating respectively from the sky and the sun.

Hourly global irradiance time series were obtained through physically sound time-interpolation/extrapolation. Physically sound interpolation signifies that global irradiance itself was not interpolated, since it is dependent on varying solar geometry. Instead, we interpolated a solar position-independent quantity: a normalized clearness index Kt'. This index was recently proposed by the authors and their colleagues [10], as a more accurate means of parameterizing global insolation conditions at a particular point in time.

Figure 6.6.2 shows an example of the time interpolation/extrapolation performed on a specific day with variable insolation conditions.

Once hourly global values were generated, they were processed into beam and diffuse components using a model developed by the authors and co-workers [11] as an upgrade to SERI's DISC model [12]. The generated global, direct and diffuse components were subjected to a further correction depending on solar geometry, insolation levels and the presence of snow on the ground. This statistical correction was derived from the ground-truth validation of the uncorrected satellite values in two locations.

6.2.2.3 Satellite vs. Interpolation for Aereal Coverage: Methodology

Although ground measurements are in principle more precise than any satellite estimate at the point where they are performed, the satellite's main advantage is one of extensive geographical coverage. Beyond a given distance from a ground station, the satellite ought to become a better predictor. We investigated this particular aspect using a ground-truth network covering southern New York and New England; this network had been assembled for the IEA task mentioned above [1].

This aspect of the study investigated daily total global irradiance rather than hourly values, in accordance with the current IEA scope, and because of time resolution limitations within the current network. Satellite daily global estimates were compared with ground values obtained through both interpolation between and extrapolation from neighboring ground stations.

6.2.2.4 *Evaluation of Hourly Satellite Estimates Against Controlled Ground Measurements*

Ground-based hourly global and direct irradiance data measured in Queens, NY, and Albany, NY, in 1987 [13] were used as the primary ground-truthing data for this analysis. The sites, operated by the authors, were carefully monitored and the data subjected to stringent quality control. Moreover, each site is located in the immediate vicinity of satellite grid points.

Plots of satellite-estimated hourly global and beam irradiance against ground values are provided in Fig. 6.2.3 for each test site. As would be expected, a much better level of prediction is achieved for the global component than for the direct one (since the latter goes through an additional simulation step). The accuracy of the satellite estimates may be summarized in terms of root mean square and mean bias errors (resp. RMSE and MBE); the two terms describe respectively the short term accuracy (dispersion) of the estimates and their long term accuracy (bias). If x_i and x_i' are respectively the measured and estimated values for each of a total of n observations, MBE and RMSE are defined as:

$$\mathrm{MBE} = \Sigma(x_i - x_i')/n \qquad \mathrm{RMSE} = \sqrt{[\Sigma(x_i - x_i')^2/n]}$$

It is sometimes useful to refer to these two quantities in relative (%) terms rather than in absolute terms by normalizing them with respect to the variable's mean value. The MBEs and RMSEs corresponding to the current evaluation are presented in Table 6.2.1. Mean values for each quantity at each site are given in Table 6.2.2.

Table 6.2.1

RMSE and MBE
for Uncorrected and Corrected Satellite-Derived
Global and Direct Irradiance

	QUEENS		ALBANY	
	RMSE	MBE	RMSE	MBE
Uncorrected				
Global Irradiance	90 W/m^2	-6 W/m^2	95 W/m^2	1 W/m^2
Direct Irradiance	190 W/m^2	50 W/m^2	196 W/m^2	73 W/m^2
Corrected				
Global Irradiance	87 W/m^2	-4 W/m^2	88 W/m^2	-1 W/m^2
Direct Irradiance	167 W/m^2	-4 W/m^2	166 W/m^2	9 W/m^2

Model Adjustment: Based on the validation of the uncorrected global and direct irradiance at each site, a correction depending on solar elevation, clearness index (i.e., uncorrected global), and prevailing ground reflectance (i.e., snow/no-snow) was derived. A correction function for global irradiance (a discrete matrix) was optimized through least square fitting to data from the two reference ground sites in Albany and Queens. In addition, a simple bias detrending, also common to both sites, was applied to the direct beam component derived from the corrected global values.

Corrected satellite estimates of hourly global and direct irradiance are plotted against ground-truth values in Fig. 6.2.4. Some improvement may be noted, particularly at low global irradiance values. For the direct component, the dispersion for intermediate cases, while still large, has been noticeably reduced. Bottom line RMSEs and MBEs are presented in Table 6.2.1 for these corrected estimates.

Discussion: The validation results are quantitatively consistent with independent tests (e.g., [1,2,14]). The relatively large dispersion for the beam component would lead one to caution the use of this component for evaluating concentrating solar collector real-time performance. Concerning inclined sun/south facing flat plate collectors, the use of an anisotropic solar diffuse radiation model tends to compensate for errors in the direct beam as shown by [15] (this would not be true for non-sun facing surfaces). Hence, the level of accuracy achieved at this point was deemed satisfactory enough to pursue the study to its load matching phase. Note that the satellite vs. ground-truth comparison continues through section II. The evaluation of the satellite's precision in estimating photovoltaic load matching benchmarks constitute a very important complement to the present physical comparison.

It should be noted that the level of precision achieved here is representative of the satellite transfer methodology available for this study rather than the ultimate capability of remote irradiance sensing from geostationary satellites. In particular, a better estimation of beam irradiance should be possible by modeling this component directly from the satellite image rather than by extrapolating from global irradiance [14,16]. Also, using the satellite as a complement to a ground truth network should be an optimum way of tapping effectively into the remote sensor's main asset -its access to radiation fields' geographical structure - while insuring quantitatively accurate predictions; preliminary results on this matter are quite encouraging [1,7].

An interesting finding is made on the climatological scale: The results presented above pertain to the satellite's ability to predict current hourly insolation levels at a specific point in time -- a crucial consideration when performing load matching assessments. If one looks instead at the satellite's ability to reproduce typical rather than current local hourly insolation conditions the satellite's precision becomes quite remarkable, both for the global and direct components; typical conditions are needed for solar system design and optimization. Frequency distribution histograms and their associated statistical characteristics (mean, skewness and kurtosis) are an appropriate measure of climatological representativeness. Such histograms are presented in fig. 6.2.5 and 6.2.6 (resp. Albany and Queens), comparing the direct and global components measured at the ground and estimated from the satellite. Statistical characteristics for these histograms are provided in Table 6.2.2. These results show that the GOES estimates provide, in the considered area of the northeastern U.S., a sound measure of local hourly insolation conditions.

Note that the typical hourly radiation data currently available for engineering purposes are the Typical Meteorological Year (TMY) tapes prepared by the USDOE in the mid 70s [17]; in most cases these data were not measured but derived from other meteorological measurements; their local representativeness in terms of radiation has been found in many instances to be well below the level of accuracy achieved here with the satellite's data [18].

Table 6.2.2

Statistical Summaries for Histograms
Presented in Figs. 6.2.5 and 6.2.6

	GLOBAL		DIRECT	
ALBANY	GROUND	SATELLITE	GROUND	SATELLITE
Mean (W/m^2)	365	364	336	335
Std.Dev.(W/m^2)	234	256	299	327
* Skewness	0.764	0.688	0.272	0.294
* Kurtosis	2.408	2.296	1.591	1.509
QUEENS				
Mean (W/m^2)	408	403	396	391
Std.Dev.(W/m^2)	262	269	325	340
* Skewness	0.540	0.446	-0.026	0.112
* Kurtosis	1.985	1.983	1.404	1.446

* Skewness is a distribution's third central moment.It is a measure oft he distribution's disymmetry (Skew = 0 for a symmetrical distribution).

* Kurtosis is a distribution's fourth central moment.It is a measure of the distribution's peakedness (Kurt = 3 for a normal distribution).

6.2.2.5 *Results for Areal Coverage*

A seven-station ground network was assembled from existing data. This covers a triangular area extending from Albany and New York City, NY, to southeastern Massachusets. A map of this network is shown in Fig. 6.2.7. It was possible to assemble such a network because three distinct measurement campaigns happened to be coincident in time: (1) Data from the North East Utilities network were made available to us [19] for the IEA effort (Waterford, Middleton and Milford in Connecticut were used here) . (2) Data from the Woods Hole Oceanographic Institution in Woods Hole, MA, are regularly published by the Institute [20]. (3) Data from Albany, Queens and Farmingdale, NY, were recorded by the authors for an ongoing daylight project sponsored by NYSERDA [13].

A total of 134 daily total global irradiance measurements from 10/86 to 7/87 for each of the seven sites were assembled after screening for quality control and time coincidence between stations. Calibration coherence within the network was insured by performing clear day analysis and adjusting data in reference to the New York instruments.

Interpolation and Extrapolation: The methodology selected by the IEA, to evaluate both interpolation and extrapolation accuracy, is to utilize, in turn, each station in the network as a reference point for its closest neighbor (extrapolation) or a combination of the other stations (interpolation): The interpolated or extrapolated estimate at that point is compared to the measured value.

The interpolation method used here is the so-called moving average or gravity interpolation method [21], where the interpolated estimate at point x , G(x), obtained from other stations (G_i) is given by

$$G(x) = \Sigma\, w_i G_i \,/\, \Sigma\, w_i ,$$

where $w_i = (1-D_i)^2/D_i^2$ with $D_i = d_i/R$, R being the search radius -- set at 250 km for this study -- and d_i being the distance from the ith site to point x.

Interpolation weights assigned to each neighboring station from each neighboring station are reported in Table 6.2.3 along with extrapolation distances.

Other interpolation techniques, such as Kriging, have been retained by the IEA for evaluation. However they are less systematic than the above method and could not be incorporated within the present study's time frame. Also, the network we were able to assemble here is not ideally suited to take full advantage of these more sophisticated techniques.

The accuracy achieved through interpolation was compared to that achieved through nearest neighbor extrapolation and to the accuracy achieved by the satellite estimates. The satellite estimates at each station were themselves obtained by interpolation (moving average) of the $1^{\circ}X1^{\circ}$ grid point values. Results for extrapolation from a station's closest neighbor are presented in Fig. 6.2.8. The relative RMS error has been plotted as a function of the distance between the considered station and its closest neighbor. As expected, one observes a monotonic increase in error as distance increases. Errors range from less than 10% at 30 km distance to over 35% at 180 km extrapolation. Note that local climatic effects are apparent through a small deviation from the general trend: The error degradation is below the trend for Woods Hole relative to Albany, the former site being climatically more similar (coastal) to its closest neighbor than the latter. Also note a larger seasonal effect in Albany, further north than the other stations, likely caused by snow cover and winter weather patterns differences with the sites to the south.

Interpolation results are presented in Fig. 6.2.9. Relative RMS errors have been plotted as a function of log(Σwi), a measure proportional to the inverse of the mean interpolation distance for each reference point. Note that errorsare noticeably reduced only for those sites near the center of the network (interpolation and extrapolation tend to converge for peripheral points such as Albany, Woods Hole or Queens).

Table 6.2.3

Interpolation Weights
and Extrapolation Distances
for the North East U.S. Network

	Interpolation Weight From							Nearest	Distance
Site	Wat	Mid	Mil	W H	Alb	Far	Que	Neighbor	(km)
Waterford		23.7	4.9	0.9	0.0	1.1	0.5	Middleton	43
Middleton	23.7		12.4	0.3	0.3	1.4	0.8	Waterford	43
Milford	4.9	12.4		0.0	0.2	9.4	5.1	Middleton	55
Woods Hole	0.9	0.3	0.0		0.0	0.0	0.0	Waterford	127
Albany	0.0	0.3	0.2	0.0		0.0	0.0	Middleton	166
Farmingdale	1.1	1.4	9.4	0.0	0.0		51.0	Queens	31
Queens	0.5	0.8	5.1	0.0	0.0	51.0		Farming	31

Estimation from Satellite: The satellite's accuracy is evaluated for each of the seven sites and compared to interpolation and extrapolation performance.

The MBE and RMS errors for the satellite estimates are reported in Table 6.2.4. Because of the regular satellite grid, differences between satellite estimates for each site are not very large. A RMS error of the order of 20% is found for the prediction of daily total global irradiance.

Note that the satellite relative error in estimating daily total is comparable to that observed for the hourly values (see Table 6.2.1-2). This is to be expected since there are no physical reasons why satellite errors should average out throughout a day: the cause of the error - measurement inaccuracy - should persist as long as a given type of insolation prevails. On the other hand, there are physical reasons (i.e., moving cloud systems) for interpolation and extrapolation accuracy to degrade as one goes toward shorter time intervals. Hence, the present comparison should be viewed as conservative with respect to the satellite relative capabilities at the hourly level.

Table 6.2.4

Root Mean Square and Mean Bias Error
of Daily Total Global Satellite Estimates
For seven sites in southern New York and New England

Site	Relative MBE	Relative RMSE
Waterford	-0.1%	22.3%
Middleton	2.6%	21.2%
Milford	-1.4%	20.6%
Woods Hole	3.0%	22.4%
Albany	3.4%	18.7%
Farmingdale	3.0%	20.6%
Queens	4.9%	20.3%

In summary, based on the limited results of this study, the satellite is found to be more accurate than ground station extrapolation for distances greater than about 65 km (40 miles) from the station. In the case of interpolation, the break-even interpolation distance would be of the order of 95 km (60 miles); the break-even distance is larger than for extrapolation since the considered points has access, through the interpolating process, to some information about the irradiance field's structure. As explained above, these distances are likely to be shorter for hourly estimates.

Finally, note that the level of accuracy observed here is representative only of the type of data available at this point in time. Satellite-aided ground station interpolation should perform considerably better. Evidence from the current incomplete data set shows for instance that, for Albany, the satellite error may be noticeably reduced if a ground truth station is present in New York City, 230 km (140 miles) away.

6.2.2.6 Preparation of a Radiation Data Set for southern New York

A data base containing hourly global and diffuse irradiance for the year 1987 was prepared. Four locations in southern New York were selected. These are (1) The eastern tip of Long Island at 72° longitude, 41.1° latitude, (2) central Long Island at 73° longitude, 40.8° latitude, (3) the Western end of Long Island (Brooklyn) at 74° longitude, 40.6° latitude, and (4) the lower Hudson Valley (near Newburgh) at 74° longitude and 41.6° latitude.

The data base was derived solely from the satellite source for the following reasons: (1) completeness of the satellite data base with respect to the ground base, (2) overall consistency, and (3) acceptable level of prediction from the satellite. Indeed, the validated hourly ground data for the area are available for only a fraction of the year 1987 and all sites are clustered near the western end of long Island, hence extrapolation would have had to be used for the other points, with the loss of accuracy trade-off mentioned above. The presentchoice is further justified by the end-use satellite validation results presented in the following section.

The choice of four ground locations is consistent with the space resolution currently made available for satellite irradiance estimates (1° X 1°). A more refined processing of satellite images could eventually yield higher resolution. Values for each of the four locations were obtained by spatially interpolating the Kt' values at available neighboring satellite grid points. In addition, the interpolated estimates were linearly adjusted to account for the yearly geographic trend observed for the coastal region (this adjustment did not exceed 3%). Note that, as for the temporal interpolations described in section 6.2.2.2, these spatial interpolations were performed using the solar geometry-independent clearness index Kt' rather than global irradiance.

6.2.3 Utility load matching assessment

6.2.3.1 Experimental Data

The photovoltaic load matching potential was determined experimentally from existing hourly load and PV generation data. The data cover two summer seasons (June to September 1987 and 1988).

Hourly utility load data were provided by the New York Power Pool for four southern New York utilities: Consolidated Edison (ConEd), Long Island Lighting Company (LILCO), the Village of Freeport (V. Freeport) and the southeastern New York sector of the New York Power Authority (NYPA-Seny).

Each utility is summer peaking, with peak requirements driven mostly by air conditioning. For information, the mean and peak daily load profiles of each utility are shown in Fig. 6.2.10. Note the marked daily peak for ConEdison, linked to commercial activity, and the increased evening shoulders for LILCO and V. Freeport, where residential loads, late in the day, are relatively more important. As previously noted, the NYPA daily peak is also very pronounced but two secondary peaks are superimposed in the early morning and late afternoon (corresponding to rush hours) because of this utility's power delivery to the New York City's transit system.

Hourly photovoltaic generation data were obtained by means of simulation from hourly irradiance, wind and temperature data. Sandia National Laboratories' PVFORM [22] was used for the simulations. This simulation program was validated for the New York City area in a previous research phase [23, 24]. Two comparative sets of input irradiance data were used: one was derived from the GOES satellite observations as explained above; the other is ground data set from Queens and/or NYPA's experimental Lincoln Center photovoltaic station, where measurements were performed through both summers with a data recovery rate exceeding 90%. In order to perform a sound comparison between the ground and satellite input, the satellite data were estimated for the Lincoln Center's area. The temperature and wind speed data were also recorded at the Lincoln Center's Station (missing points were obtained from nearby LaGuardia airport).

Two geometric configurations were considered for the photovoltaic array. The first is a two-axis tracking configuration; this constitutes the upper limit of collectable energy at all time; any other configuration would deliver less per unit of collecting area (although not per unit of occupied ground area). Note that a horizontal one axis tracking system will approach the performance of the two axis tracker in summer. The second configuration is a fixed flat plate structure; both slope and azimuth are optimized for peaking purposes.

6.2.3.2 Methodology for Experimental Evaluation of Load Matching

Effective Load Carrying Capability: The principal criterion used to assess load matching is the determination of the PV system's Effective Load Carrying Capability (ELCC) as specified by Garver [3]. The Garver method is readily applicable when time series of the utility load requirements and the considered unit's capacity are available (e.g., see [25]).

This method was derived primarily to handle large "island utilities" that can not rely on outside sources to meet their requirements. The utility's generation mix is characterized by a single coefficient, the Garver characteristic value [3], defined as the increase in annual peak load which would cause a loss-of-load risk e (e = 2.718) times greater than before. LILCO is such an island utility and utilizes the Garver methodology for generation planning purposes. This utility provided us with its Garver characteristic value.

ConEdison does not use the Garver method for generation planning. However, to put LILCO's results in perspective, we also investigated the photovoltaic ELCC with respect to the ConEd grid. ConEd's load profile is driven more extensively by commercial demand than LILCO's; hence it offers an interesting contrast. A Garver characteristic value was estimated for ConEd by prorating LILCO's Garver characteristic to ConEd's capacity; this was justified because the mean forced outage rate of their generation units was found to be comparable and Garver characteristic values are proportional to mean forced outage rate of installed capacity [3].

The ELCC criterion was not applied to NYPA-Seny and V. Freeport, because of their smaller loads and different generation mixes.

Additional Load Matching Criteria: Two additional load matching criteria were used to complement and contrast the ELCC estimates. These criteria are (i) the resource's relative energy value and (ii) the minimum energy storage requirement to ensure that the PV+storage system will effectively (i.e., in the worst case) reduce the utility's peak load requirement by a given fraction of its capacity. Both parameters may be readily derived from the hourly load and generation time series as explained in [23].

The relative energy value summarizes the distribution of the energy produced with respect to the utility's energy production rate. In the present case, because hourly production costs are not readily available for each utility and in an attempt to present normalized results, the value of energy to a utility is assumed to be in direct relation to the utility's load duration curve: this may be either proportional to the energy axis or to the time axis as illustrated in Fig. 6.2.11. In a recent study by Hoff et al. [25] the time axis was used; both options are compared here.

The minimum storage requirement parameter is interesting on two accounts. First, it provides a worst-case measure by which to assess PVs reliability for supplying power when needed. Such a worst case measure is particularly important for an non-controllable resource such as PVs. Second, because PV is a modular resource likely to be developed in a dispersed rather than a centralized fashion, the use of PV systems as a demand-side (user-side-of-the-meter) application is a possibility that offers both practical and economical advantages [24,26]. As such, photovoltaics would complement other DSM options such as cold water or ice storage. Storage reductions resulting from PV operation could make these economically viable but space consuming storage options more attractive.

6.2.3.3 Preliminary data Correction

During the process of comparing ELCC values obtained from the Lincoln Center station data and the satellite data, it became evident that a larger differencewas found than would have been expected from the results presented in section 6.2.2. We decided to perform additional hourly global irradiance comparisons between the satellite and the Lincoln Center data (that had not been used in section 6.2.2 validations): the results are presented in Fig. 6.2.12 for the two summer seasons under study. They indicate a substantial overprediction from the satellite, that had not been observed with the other data sets. As some of the Lincoln Center's data overlapped with that from our station in nearby Queens, we compared data from the two sites and found a quantitatively similar discrepancy, as can be seen in Fig. 6.2.13: in fact a linear regression performed between the Lincoln Center's data and the satellite yields a 1.047 slope while a 1.044 slope is obtained between Lincoln Center and Queens.

It was therefore concluded that the Lincoln Center data were slightly underpredicting the regional solar yield. Several reasons probably combined to produce this effect: (1) the higher atmospheric turbidity within Manhattan due to the pollution canopy, (2) the higher degree of instrument soiling, also due to greater pollution, (3) the high horizon line and loss of diffuse radiation due to the presence of sky-scrapers, particularly in the instrument's southern field of view, and (4), the very low albedo of Manhattan and the resulting loss of atmospheric backscattering, particularly during overcast conditions.

Because the Lincoln Center's data was the only source of ground data covering both summers, the decision was taken to account for the measurement anomaly by correcting the data using the regression function derived with the controlled data from the Queens station.

6.2.3.4 Optimum Configuration for the fixed-tilt PV array

The azimuth and slope of the fixed-tilt PV array were optimized to provide the best possible load matching with respect to LILCO's load. Both ELCC and relative energy value criteria were considered for this purpose. The variations of the fixed array's ELCC and relative energy value with azimuth and slope are respectively illustrated in Fig. 6.2.14 and 6.2.15. These were obtained by computing ELCC and relative energy value for a set of possible slope/azimuth combinations.

Optimum ELCC was achieved at about 35° slope and 65° azimuth-west while optimum energy value was achieved at only 20° slope and 25° azimuth-west. This difference is understandable since optimum peaking is best achieved for a late afternoon optimization, but at the expense of morning and midday energy production. The relative energy value penalty at optimum ELCC condition is 3% while the ELCC penalty at optimum energy value is 8.5%. Note that for south facing 15° tilt collector, the optimum configuration for overall summer energy yield, the penalty is 16% in terms of ELCC and 5% in terms of energy value.

The optimum fixed-tilt configuration selected for the remainder of this analysis features 40°-West for the azimuth and 30° for the slope. With this configuration, the penalties with respect to optimum ELCC and energy value configurations are only 2% and 1.2%, respectively.

6.2.3.5 Effective Load Carrying Capability

The difference observed between the ELCC derived from the satellite and the reference data from Queens (Jun-Jul, 1987) has been plotted in Fig. 6.2.16 as a function of the installed PV capacity for both ConEdison and LILCO.

Part of the satellite's more optimistic view may be legitimately attributed to the fact that the satellite provides a regional (25 image pixels = 2600 sq. km) rather than a local value --the presence of localized clouds at a critical time will be more detrimental to the ELCC if one considers one single PV system rather than an ensemble of systems distributed over a large area. Nevertheless, this difference was used as a conservative correction function for the satellite-derived ELCC values for 1987 and 1988.

These results are presented beginning with the two-axis tracking array in Fig. 6.2.17 where the ELCC variations as a function of installed PV capacity have been plotted for LILCO. A similar plot is provided for ConEd in Fig. 6.2.18. A plot of relative ELCC vs. installed PV capacity is provided in Fig. 6.2.19 for both utilities. Figures 6.2.20 to 6.2.22 include similar results but for the fixed-tilt photovoltaic array.

Discussion: The results are qualitatively consistent with our previous preliminary assessment based on a limited data set. Quantitative agreement with the previous analysis is also found, although, because of the small bias found with the Lincoln Center data, the ELCC estimate has been slightly increased.

It is encouraging to note that 1987 results differ only marginally from 1988's. This consistency indicates that the relationship between PV production and the demand for electrical power on the ConEd and LILCO grids is strong and predictable.

As would be expected from the load profiles, the ELCC with respect to LILCO's load is found to be smaller than in the ConEd case, although the numbers remain encouragingly high.

A gradual ELCC decrease with installed PV capacity is noted. However, note that PV penetration would have to be at least one to two orders of magnitude above current world PV production levels (currently 40 MW per year) to substantially affect effective load carrying capability. With two-axis tracking at the 10 MW penetration level, the ELCC is near maximum at 65% for ConEd and 57% for LILCO.

If the system does not track the sun, the ELCC reduction is not found to be substantial. With an optimized fixed tilt system, the ConEd and LILCO values become respectively 56% and 50% at 10 MW penetration. ELCC decrease with penetration is relatively less marked than for the two-axis tracking configuration.

6.2.3.6 Auxiliary Utility Load-Matching Benchmarks

Relative Energy Value

Histograms showing the distribution of the energy produced by the PV system as a function of a utility's load requirements (load duration curve's time axis) are shown for each considered utility in Figs. 6.2.23-26 for the 2-axis tracking configuration and in Figs. 6.2.27-30 for the optimized fixed-tilt configuration. Each histogram includes results for 1987, 1988, the satellite and the ground-based PV simulations. Summary relative energy values (%), obtained by assigning a linear scale (0-100%) to each load duration bin, are also provided in each diagram.

Results are summarized in Figure 6.2.31 and 6.2.32 (resp. tracking and fixed array) for the four utilities.

The marked asymmetry of each distribution indicates an unquestionable correlation between solar availability and times of greatest electrical demand for each utility.

The agreement between satellite and ground-derived results is quite satisfactory; note a small tendency for the ground-based values to exceed the satellite relative energy values. This minor disagreement (conservative for the satellite) may partly be caused by the fact that the satellite base is complete for each summer, but about 5% of the data were unfortunately missing from the ground station.

The relative energy values are found to be higher for the fixed-tilt than for the two-axis tracking system. This is a direct result of the optimization process where most of the fixed-tilt energy production is centered around peak load time. However, it must be remembered that these energy values are only relative (i.e., per unit of energy produced) and that one must multiply by the total energy from the array to obtain the actual worth of the energy produced by the system. In terms of total energy produced, the fixed array falls short of the 2-axis tracker by almost 30% for the June-September period.

Energy distribution histograms, sorted with respect to the load duration curve's load axis (Bins #2), are shown in Figs. 6.2.33 and 6.2.34 for the tracking and the fixed array; results have been summarized for all the considered utilities. These figures may be directly compared to the time axis distributions in Fig. 6.2.31 and 6.2.32. Again, an asymmetry toward the highest load bins is noted, but this is substantially less marked than when the time axis is considered (except for NYPA-SENY). This is a consequence of the fact that the considered load duration curves are sharply peaked at the high end, in other words, that there are relatively few very high load events.

Minimum storage requirements

Minimum storage requirements were calculated for two photovoltaic penetration levels for each utility: 5% and 10%. For the sake of clarity, it was assumed that the PV-plus-storage systems were on the user-side-of-the-meter and that their assigned requirement were to reduce the highest system-wide load the utility would have had to meet in their absence by a firm portion of the total installed PV capacity -- 80% was selected here. The minimum total customer-side storage needed to meet this requirement was calculated in terms of total installed capacity-hours (e.g., for 20 MW of installed PV capacity on the LILCO grid, one hour of storage would be 20 MWh). Remark that the results can easily be interpreted in terms of utility-operated PV and/or storage systems.

Calculations were performed both for the optimized fixed-tilt and the 2-axis tracking array, comparing both the satellite and the ground-simulated systems. Finally, the calculated minimum storage values were compared to the storage that would have been needed to accomplish the same peak load reduction goal without the PV system. All storage systems were assumed to be recharged at night and to be full at the beginning of each day.

Results for the two-axis tracking and the fixed-tilt system at 10% PV penetration are presented in Figs. 6.2.35 and 6.2.36, respectively. Results for the 5% penetration level are presented in Figs. 6.2.37 and 6.2.38.

Overall, the results confirm that a small amount of storage (one to two hours) is sufficient to make photovoltaic production reliable at all times to accomplish their peak shaving requirement in relation to the utilities under study. As can be seen, levels of storage necessary for peak load reduction through supply or demand side management could be substantially reduced with the help of PVs. It is interesting to note that the fixed array's performance approaches that of the tracker in this case.

The consistency of the results between 1987, 1988 and our limited previous one-month study [23,24] is quite remarkable given the "zero tolerance" nature of this storage benchmark (i.e., only one event with high cloudiness during high demand time would considerably modify the results). Again, these results tend to confirm that the solar resource-load relationship is a solid, predictable one.

Although the satellite-derived storage estimates are a little more optimistic than the ground's, the agreement between the two sets is quite satisfactory, particularly when compared to the storage needed without a PV system.

6.2.3.7 Summary of End-Use Satellite Data Evaluation

Reviewing results from the three benchmarks, the end-use agreement between the satellite and the ground-based simulation is better than the simple physical comparison in this study's first section would have suggested. Although the satellite has a tendency to provide a little more optimistic rating for PVs, the overall agreement is satisfactory.

The largest disagreement between ground and satellite was found for the evaluation of LILCO's relative ELCC where the satellite derived relative ELCC was found to 6-10% more optimistic.

This tendency for the satellite results to be a little more optimistic than ground-derived values may be partly attributed to the fact each satellite data point represents an ground area of 2600 sq. km (1000 sq. miles). Such an extended area is less sensitive to localized cloud development (e.g., thunderstorms), hence will not experience as sharp short term variations as a point source would. In some respects, the extended area may be more useful to a utility than the point source information. Also contributing to this "smoothing" effect, is the fact that the satellite hourly data were obtained by time interpolation from only five to seven points per day.

In summary, the satellite data appear to meet the goal of adequately addressing the general question of whether PVs may effectively contribute to a given utility's power supply options. Therefore, it is believed that other utilities could use this information to perform a preliminary assessment of PVs load matching potential for their service area.

6.2.4 Conclusions

This study addressed three general issues. A set of conclusions may be drawnfor each.

First, the accuracy of satellite-derived insolation at the earth surface was validated on a physical basis. At the hourly level, agreement between currently available satellite-estimated global irradiance and ground-measured global irradiance at a given location was found to be consistent with other investigations, with little overall bias and a dispersion of the order of 20%. For the direct component, dispersion was found to be much more significant; better methods should be sought to directly convert the satellite signal into direct beam estimates. In climatological terms however, satellite predictions were found to be quite representative both for global and direct irradiance; the ability of the satellite information to recreate typical insolation conditions at the locations tested is remarkable; indeed the satellite could well constitute a better source of radiative data than many of the current Typical Meteorological Year (TMY) sets which contain synthesized solar data [27]. Concerning geographically extended solar resource assessment, the satellite was found to be a better source of data than a ground station if the distance from the ground station exceeds about 65 km (40 miles). For a network, the satellite's precision would exceed that of the interpolated ground measurements if the considered point features a mean interpolation distance exceeding 95 km (60miles) - note that this distance may likely be increased if more sophisticated interpolation techniques are used. The present study also pointed to the fact that a satellite-ground network combination can result in increased precision compared to either data set alone.

Second, the accuracy of the satellite-derived insolation data was judged in terms of end-use viability. A better level of agreement was found than would have been expected from the physical assessment. The satellite data appear to meet the goal of adequately measuring PV's capacity value to a given utility. It is likely that other utilities could use a similar approach to perform a preliminary assessment of PV load matching potential for their service area before addressing more specific issues through a ground measurement network.

Finally, the last set of conclusions pertains to the evaluation of PV's load- matching potential for four downstate New York utilities. Results of our preliminary study [23, 24] were confirmed and reinforced by the present investigation. The consistency of results obtained throughout two complete summers is particularly encouraging. Three benchmarks were used to quantify PV's potential. Effective load carrying capabilities of the order of 60% of PV rated capacity were found for ConEd for up to 10% PV penetration, while the figure is 50% in the case of LILCO. PV energy was found to be time coincident with highest load instances for each utility studied. Finally, buffer storage systems of the order of one to two hours were found to be sufficient for PV to meet a firm portion of utility load; by contrast, achieving equivalent peak shaving without PVs would require two to four times more storage.

6.2.5 *References*

1. International Energy Agency Solar Heating and Cooling Program, Task IX-D: Techniques for Supplementing Network Data for Solar Energy Applications. (1988-1992). IEA, Paris, France.

2. Justus, C. G., M. V. Paris and J. D. Tarpley, (1986): Satellite-Measured Insolation Data in the United States, Mexico and South America. *Remote Sensing of Env.*, **20**, 57-83.

3. Garver, L. L., (1966): Effective Load Carrying Capability of Generating Units. *IEEE Transactions, Power Apparatus and Systems*, **Vol. Pas-85**, no. 8.

4. Diak, G. R., and C. Gautier, (1983): Improvements to a Simple Physical Model for Estimating Insolation from GOES Data. *Jour. of Climate and Appl. Meteorol.*, **22**, pp. 505-508.

5. Möser, W., and E. Raschke, (1983): Mapping of Global Radiation and Cloudiness from METEOSAT image Data - Theory and Ground Truth Comparisons. *Meteorol. Rdsch.* **36**, pp. 33-41.

6. Diabaté, L., G. Moussu and L. Wald, (1989): Description of an Operational Tool for Determining Global Solar Radiation at Ground Using Geostationary Satellite Images. *Solar Energy* **42**, pp. 201-207

7. Zelenka, A., (1990): Personal Communication. IEA-SHCP Subtask 9D leader. Swiss Meteorological Institute, Zurich, Switzerland.

8. NOAA Satellite Program Briefing, (1984). U.S. Department of Commerce National Oceanic and Atmospheric Administration, National Environmental Satellite, Data, and Information Service, Washington, DC.

9. Diekmann, F.-J., S. Happ, M. Rieland, W. Benesch, G. Czeplak and F. Kasten, (1988): An Operational Estimate of Global Irradiance at Ground Level from METEOSAT data: results from 1985 to 1987. *Meteorol. Rdsch.* **41**, 65-79.

10. Perez, R., P. Ineichen, R. Seals and A. Zelenka, (1990): Making Full Use of the Clearness Index for Parameterizing Hourly Insolation Conditions. *Solar Energy* **45**, 111.

11. Perez, R., R. Seals, A. Zelenka and P. Ineichen, (1990): Climatic Evaluation of Models that Predict Hourly Direct Irradiance from Hourly Global Irradiance: Prospects for Performance Improvements. *Solar Energy* **44**, pp. 99-108

12. Maxwell, E. L., (1987): A Quasi-Physical Model for Converting Hourly Global Horizontal to Direct Normal Insolation. Solar Energy Research Institute Report # SERI/TR-215-3087. SERI, Golden, CO.

13. Daylight Availability Resource Assessment Program, (1985-1989), Final Report to the New York State Energy Research and Development Authority, NYSERDA, Albany, NY

14. Zelenka, A., and R. Perez, (1986): Predicting Direct Solar Irradiance from Satellite from Estimated Global Irradiance and Total Cloud Cover. Application of Satellite Technology Annual Progress Report, WMO, Geneva, Switzerland.

15. Reindl, D. T., W. A. Beckman and J. A. Duffie, (1989): Diffuse Fraction Corrections. Proc. of Int. Sol. Energy Soc. Biennial Congress, Kobe, Japan.

16. Star, D. (1987): Personal Communication, NASA, Goddard Center, Washington,DC.

17. Typical Meteorological Year (TMY) Data Tapes, National Climatic Center, Asheville, NC.

18. Maxwell, E. L. (1989): Personal Communication. Solar Energy Research Institute, Golden, CO.

19. North East Utilities Solar Monitoring Network Data (1980-1987): North East Utilities, Hartford, CT.

20. Woods Hole Oceanographic Institution, Solar Radiation Station Data Summaries, (1986-1987). Woods Hole, MA.

21. D' Agostino, V., (1989): Theory of Spatial Interpolation. Draft Final Report IEA Solar Heating and Cooling Program, Subtask 9D. This Report (Chapter 10, Volume 2).

22. Menicucci, D. F., J. P. Fernandez, (1988): User's Manual for PVFORM: A Photovoltaic System Simulation Program for Stand-Alone and Grid-Interactive Applications. Report # SAND85-0376-UC-276, Sandia National Laboratories, Albuquerque, NM

23. Perez, R., W. Berkheiser III and R. Stewart, (1989): Analysis of Lincoln Center Experimental Data for Investigation of Photovoltaic Peak Load Matching Potential. Report to the New York Power Authority. ASRC Pub # 1281, Atmospheric Sciences Research Center, Albany, NY.

24. Perez, R., W. Berkheiser III, R. Stewart, M. Kapner and G. Stillman (1989): Photovoltaic Load Matching Potential for Metropolitan Utilities and Large Commercial Users in the Northeastern United States. 9th European PV Conference, Freiburg, Germany.

25. Hoff, T. (1987): Calculating Photovoltaic Value: A Utility Perspective. Proc. 19th IEEE PV Specialist Conference.

26. Perez, R. and R. Stewart, (1988): An Opportunity for Photovoltaic Development in the North East -- Non Remote Non Grid-Interactive Systems. ASRC pub. #1322, Atmospheric Sciences Research Center, Albany, NY.

27. Randall, C. M., and M. E. Whitston, Jr., (1977): Hourly Insolation and Meteorological Data Bases Including Improved Direct Insolation Estimates. Aerospace Rept. # ATR-78(7592)-1, The Aerospace Corp., El Segundo, CA.

Acknowledgements

This work was funded by the New York Power Authority (NYPA). It was performed in association with an International Energy Agency (IEA) Task [1]; as such, it received partial support from the Solar Energy Research Institute (SERI). A grant from the Mobil Foundation was helpful in enhancing the depth of this effort.

We are thankful to J. Dan Tarpley of NOAA NESDIS for providing us with satellite irradiance estimates and to Robert Goodrich of Northeast Utilities for providing us with the Northeast Utilities network data. Insolation data published by the Woods Hole Oceanographic Institution and data previously recorded by the authors on behalf of NYSERDA and NYPA were also used extensively in this work. Utility load data were made available by the New York Power Pool.

Frequent discussions with Mark Kapner, NYPA project manager, and Antoine Zelenka, IEA-SHCP subtask 9D leader, were particularly helpful at many stages of this work.

6.2.6 Illustrations

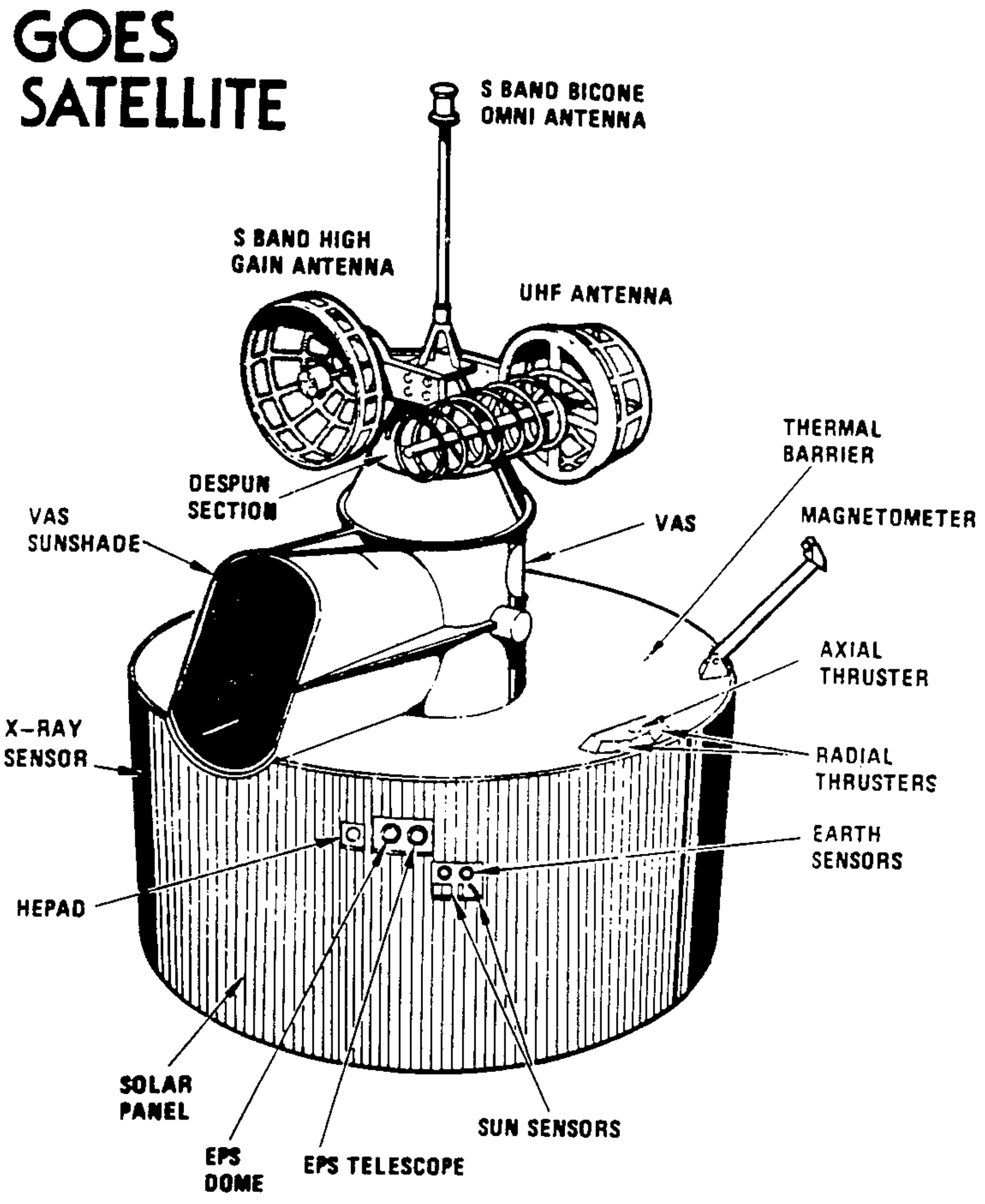


Figure 6.2.1: An aerial view of the GOES Satellite. The scanning visible sensor recording earth's surface brightness is protected from interference by the sun (V.A.S. sunshade). Reproduced from [8].

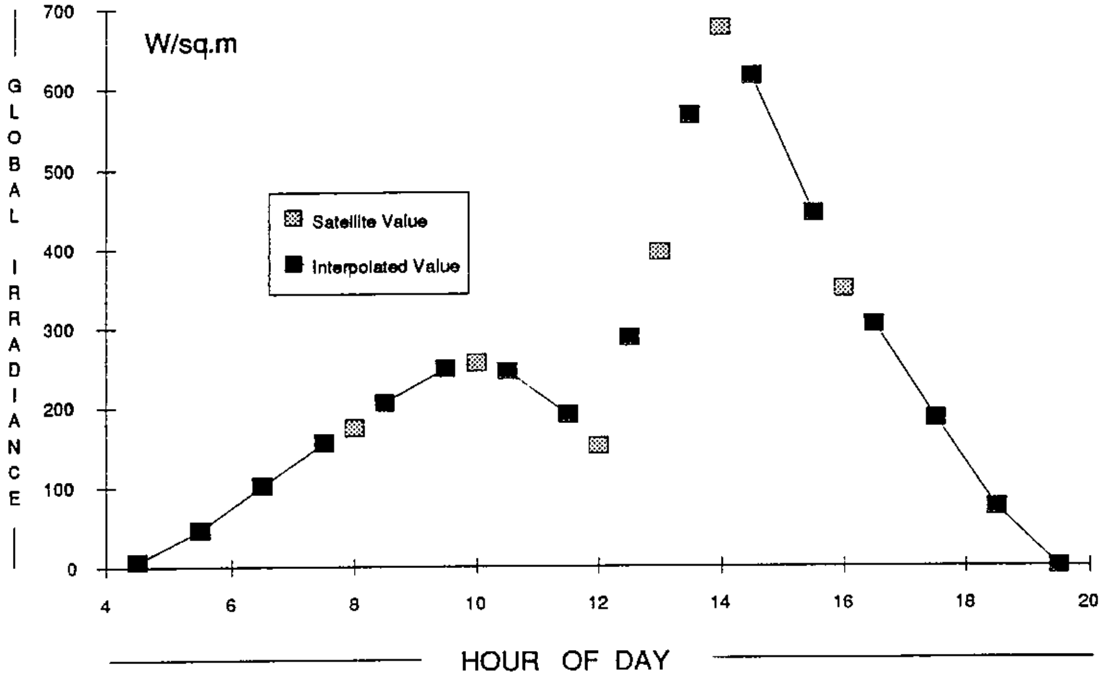


Figure 6.2.2: An example of hourly irradiance data obtained through time interpolation and extrapolation from six satellite points (6/30/87, New York City area)

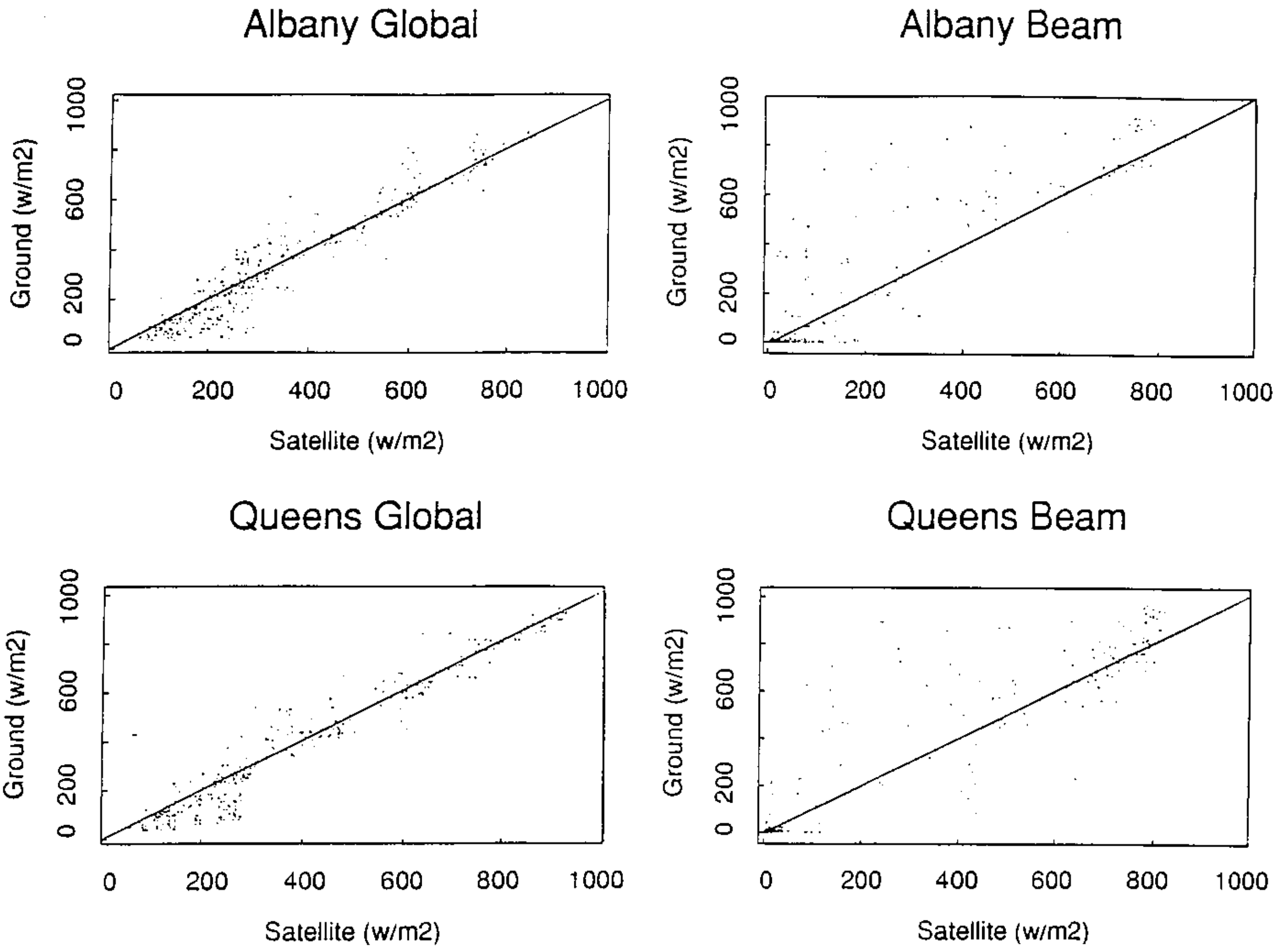


Figure 6.2.3: Comparison between hourly global and direct irradiance measured at the ground and uncorrected values derived from the satellite at two reference stations (Albany, NY and Queens, NY) for a period of one year.

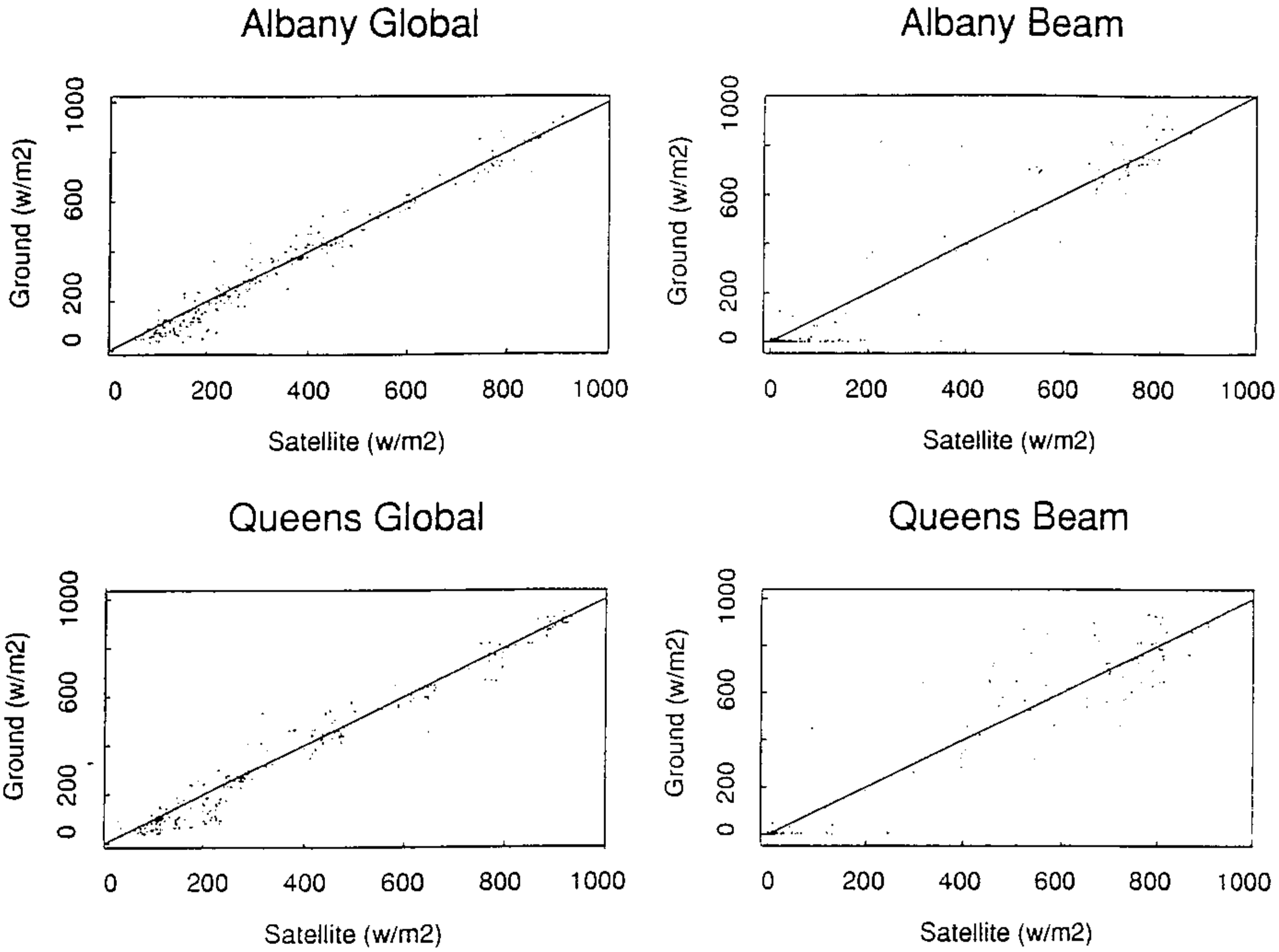


Figure 6.2.4: Comparison between hourly global and direct irradiance measured at the ground and corrected values derived from the satellite at two reference stations (Albany, NY and Queens, NY) for a period of one year.

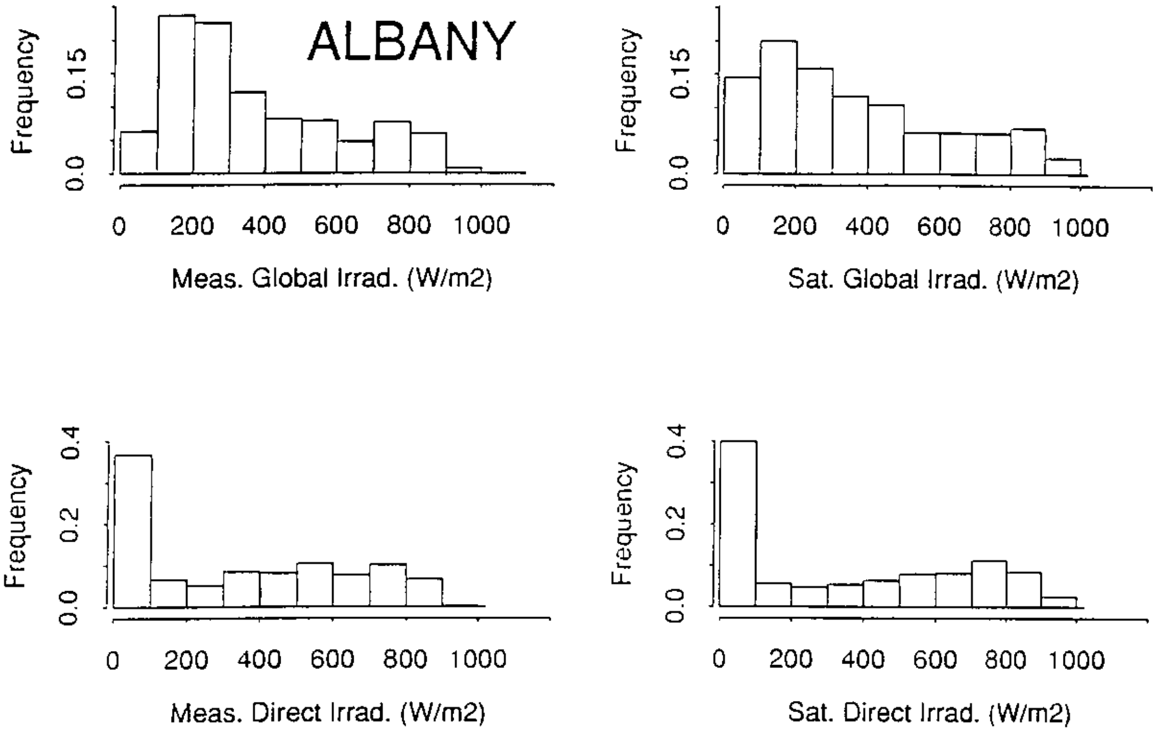


Figure 6.2.5: Comparison between the frequency distribution of hourly global and direct irradiance measured at the ground and estimated from satellite data in Albany. Data cover a period of about one year.

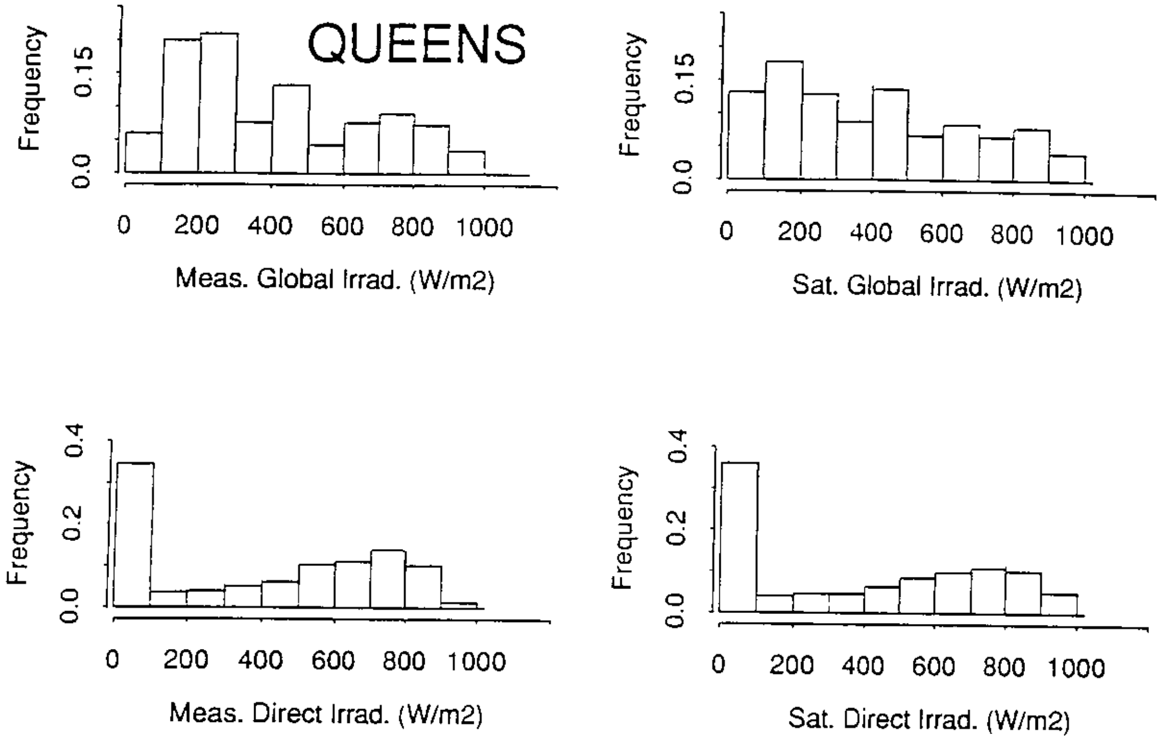


Figure 6.2.6: Comparison between the frequency distribution of hourly global and direct irradiance measured at the ground and estimated from satellite data in Queens. Data cover a period of about one year.

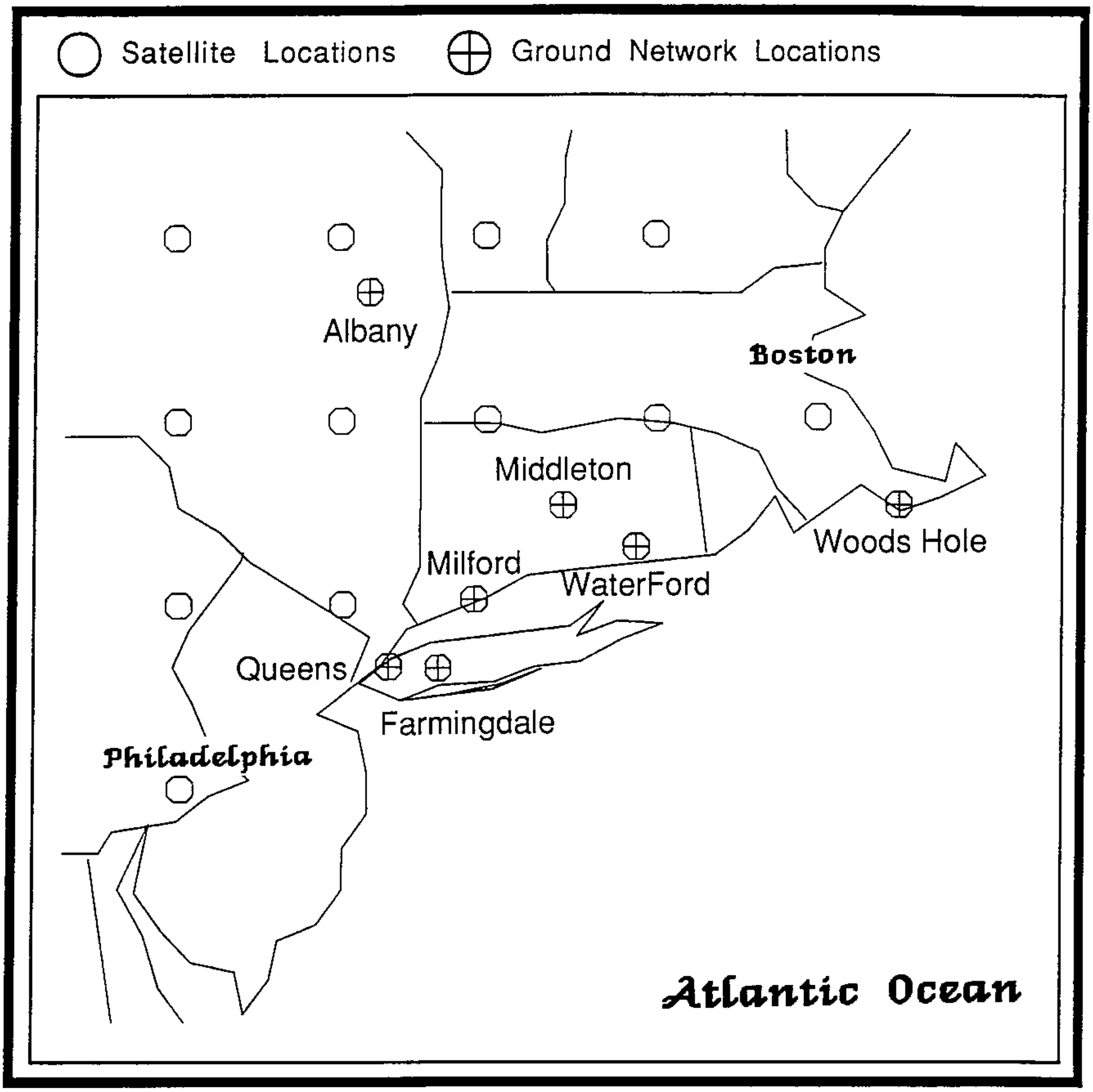


Figure 6.2.7: Map of the southern New England and northern Mid-Atlantic region showing the location of the ground-truth stations and satellite estimated points used in this study.

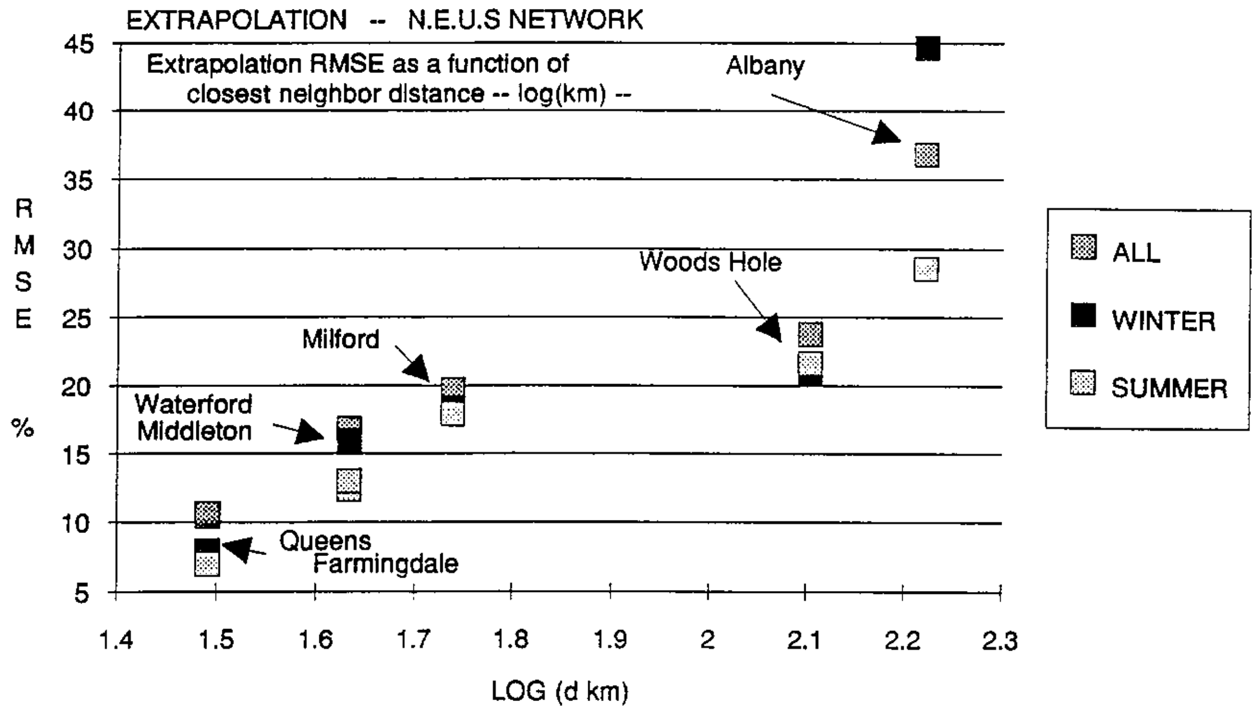


Figure 6.2.8: Variations of nearest neighbor extrapolation relative RMS error as a function of the log of the extrapolation distance for the 7-station northeastern U.S. ground network.

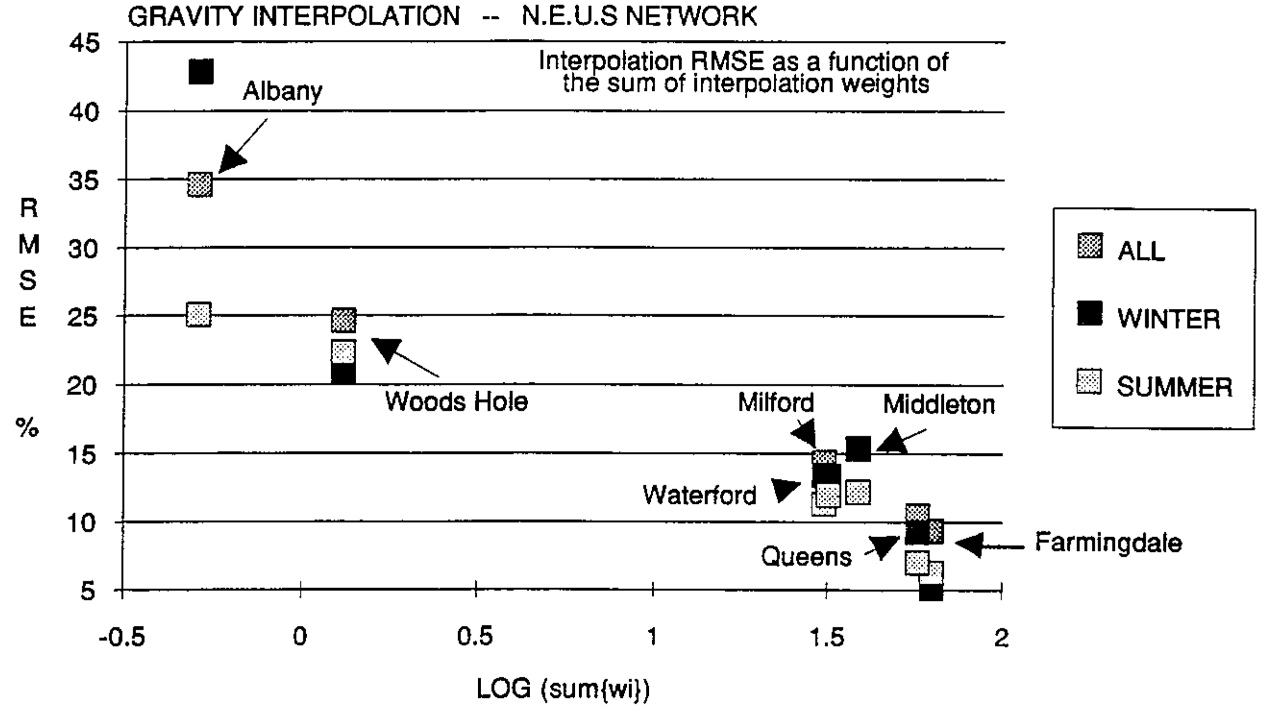


Figure 6.2.9: Variations of interpolation relative RMS error as a function of the sum of interpolation weights within the 7-station northeastern U.S. ground network. Interpolation methodology consists of gravity interpolation (moving averages)

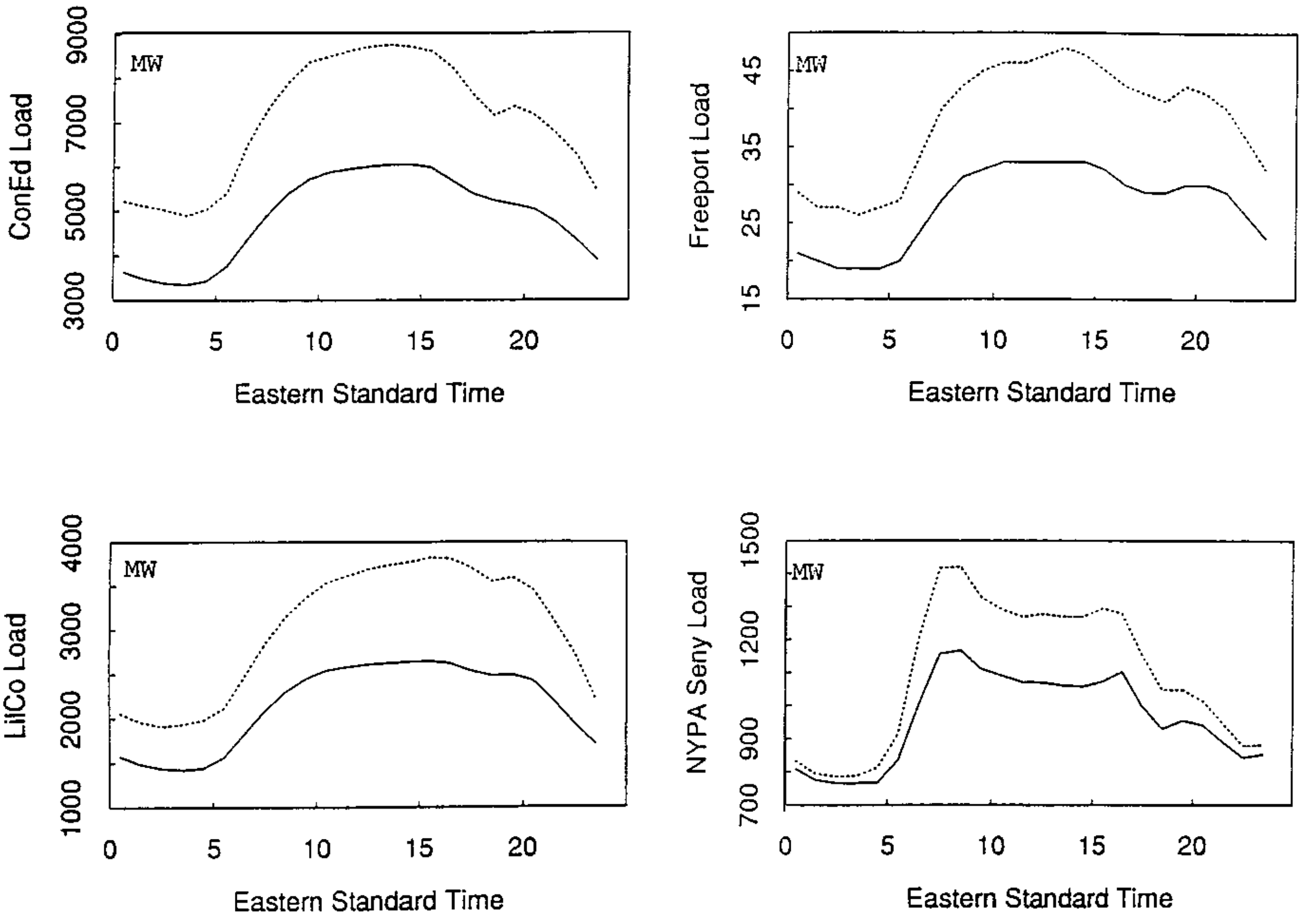


Figure 6.2.10: Daily mean (solid line) and peak (dotted line) load profiles for the four considered utilities

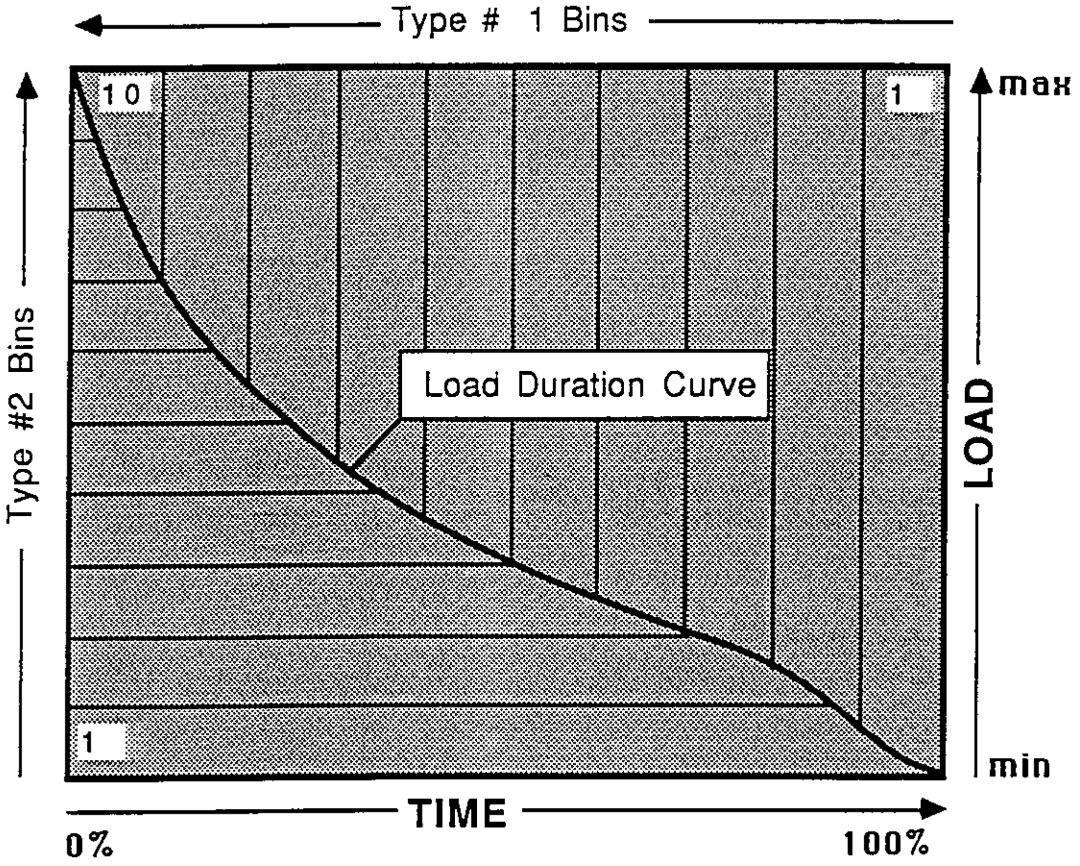


Figure 6.2.11: Example of a load duration curve reporting the fraction of time (from 0 to 1) the load is above a given value (from minimum to maximum load). Bins can be defined either proportionally to the time axis (bin # 1) or to the load axis (bin # 2) to estimate a relative value (0 to 100%) for coincident PV energy.

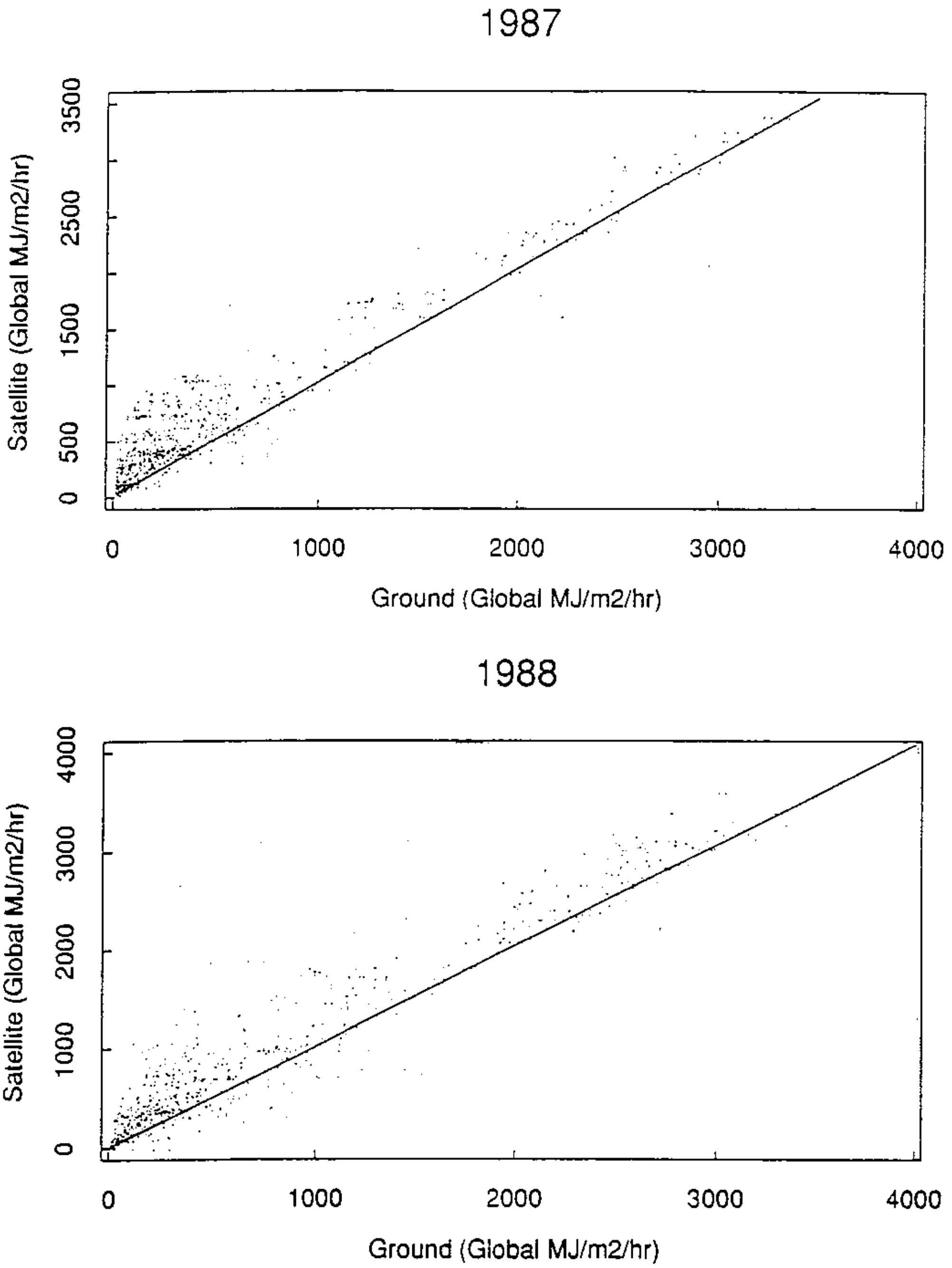


Figure 6.2.12: Comparison between satellite-estimated global irradiance and values recorded at the Lincoln Center station in Manhattan in 1987 and 1988

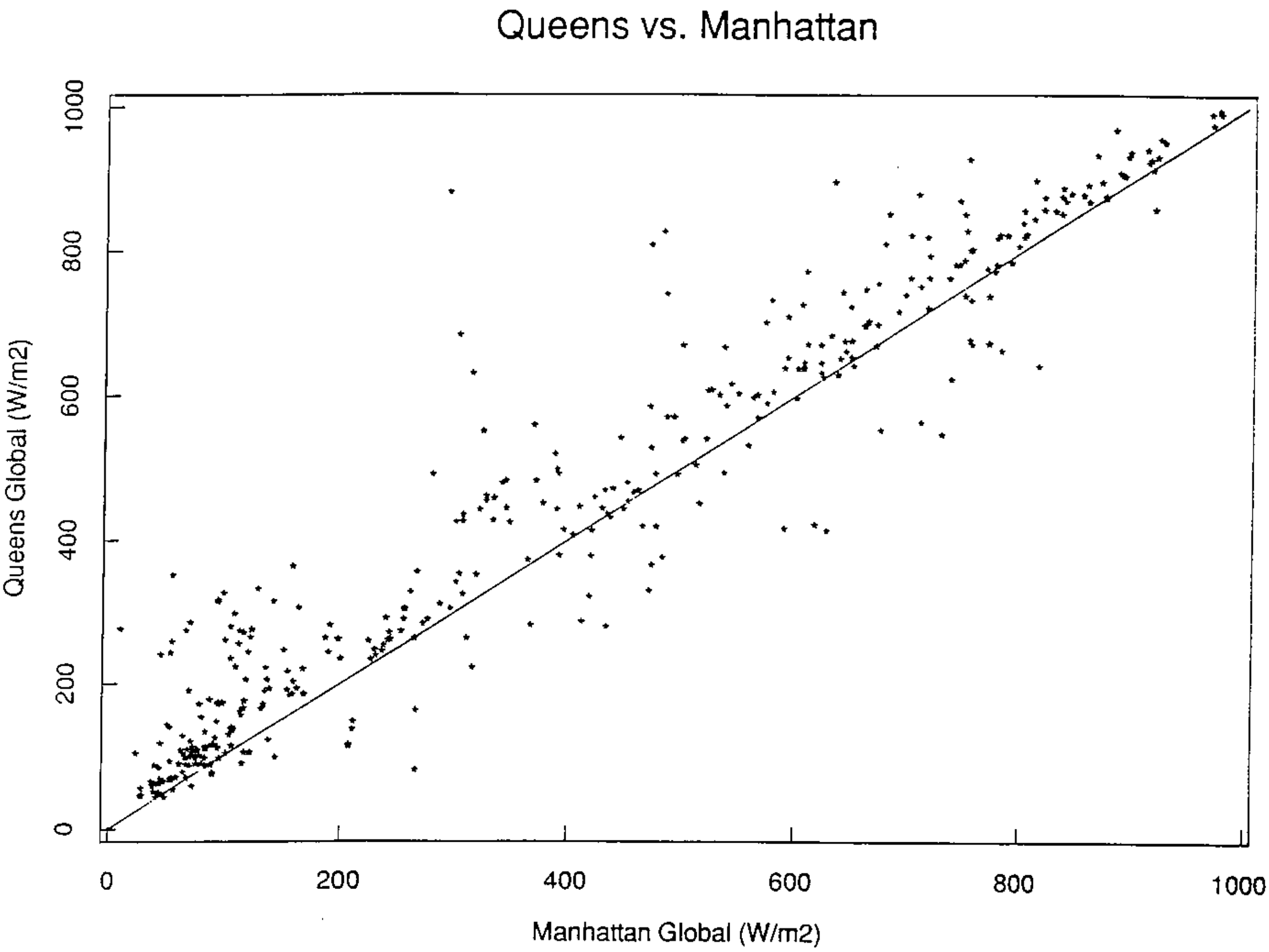


Figure 6.2.13: Comparison between global irradiance measured at the Queens station and values recorded at the Lincoln Center stations in Manhattan in 1987

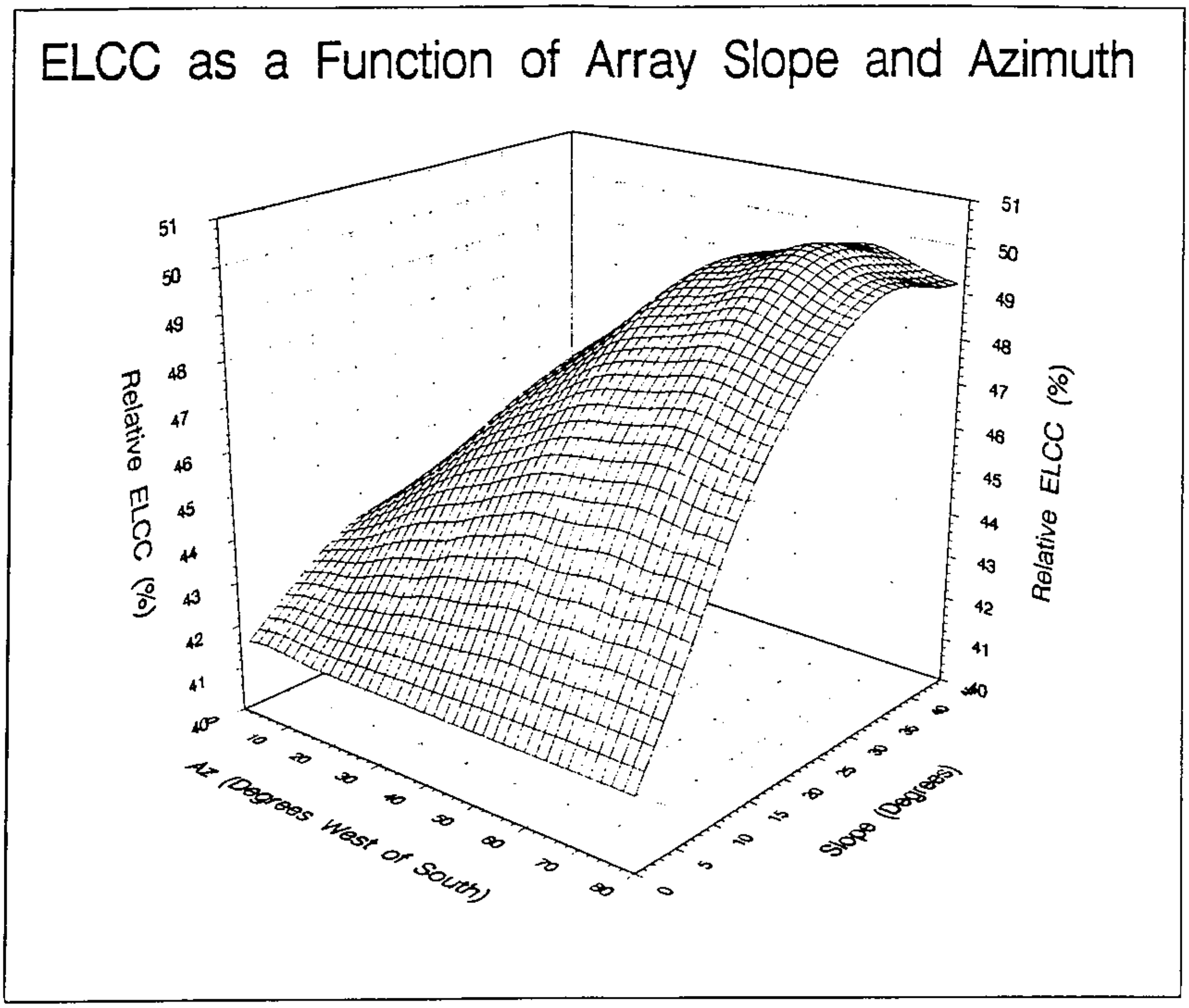


Figure 6.2.14: Variations of PV's relative ELCC with respect to LILCO as a function of array slope and azimuth. Relative ELCC is defined here as the percentage ratio of the ELCC to the installed PV capacity.

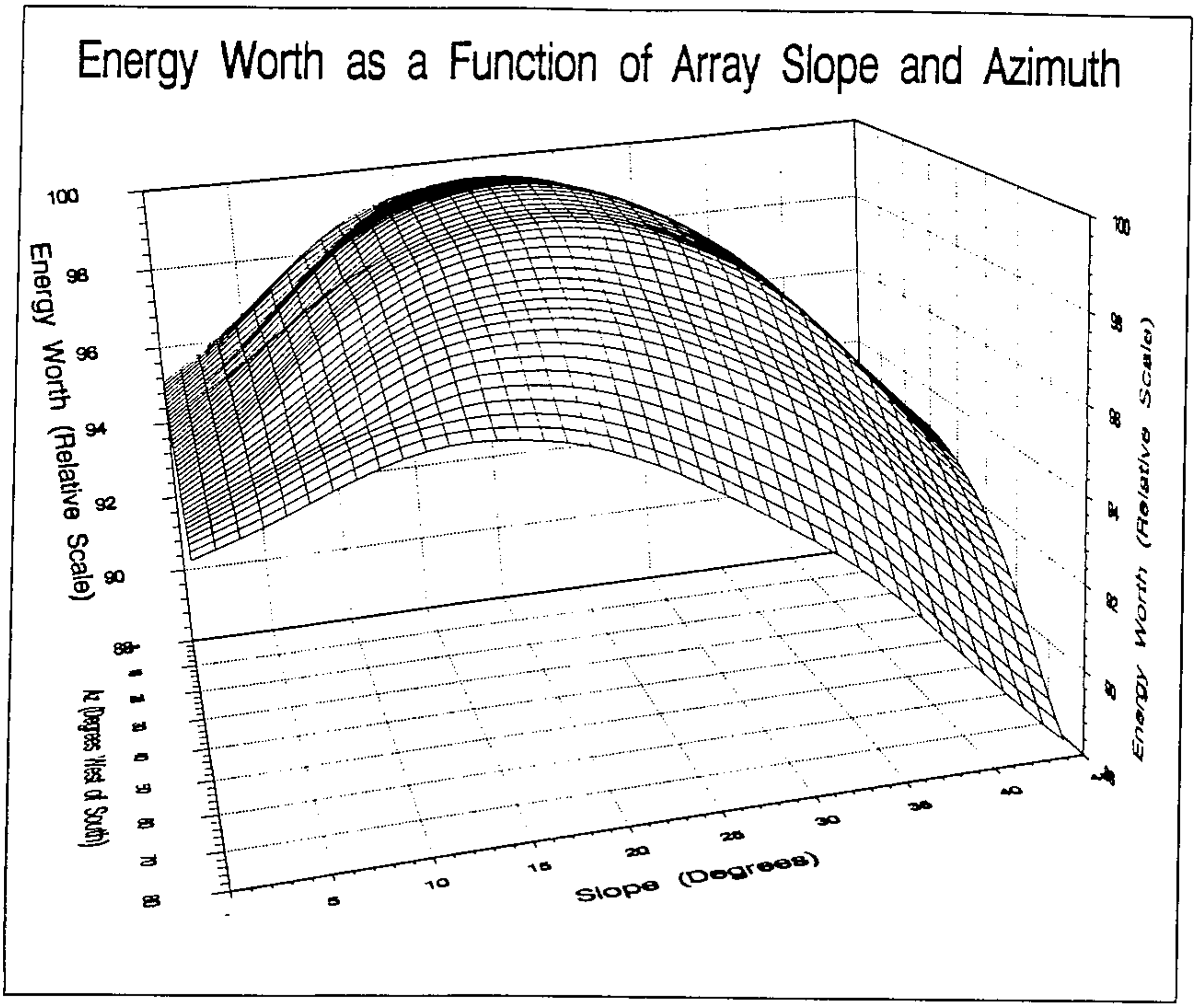


Figure 6.2.15: Energy worth of PV with respect to LILCO's load as a function of array slope and azimuth. Energy worth is reported here relatively to the highest value found (100%)

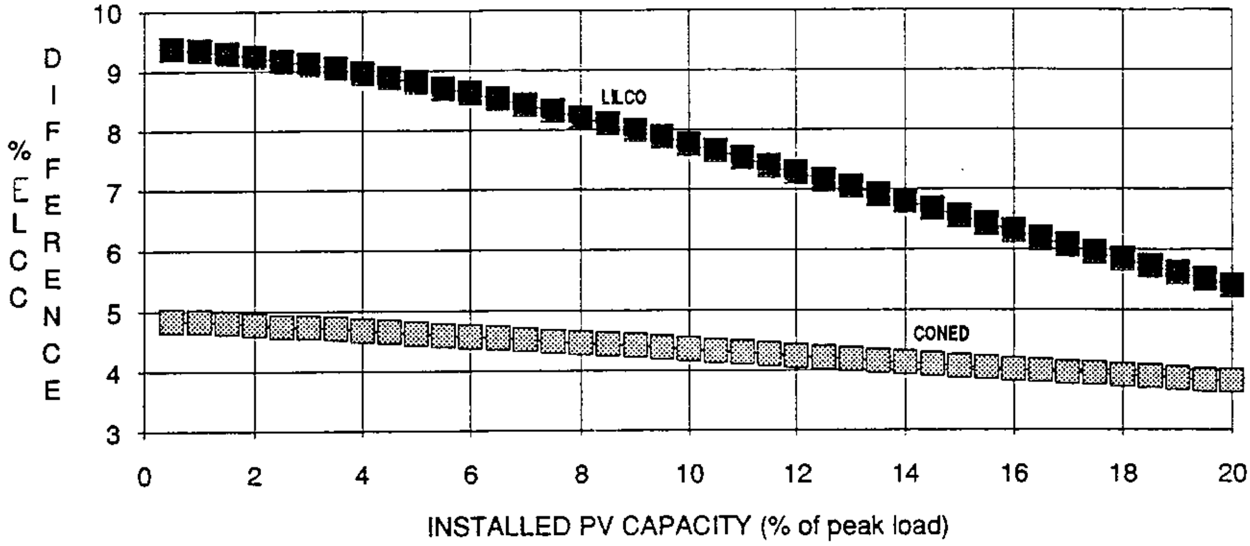


Figure 6.2.16: Observed difference between relative ELCC determined from the satellite and the ground data for LILCO and ConEd as a function of installed PV capacity

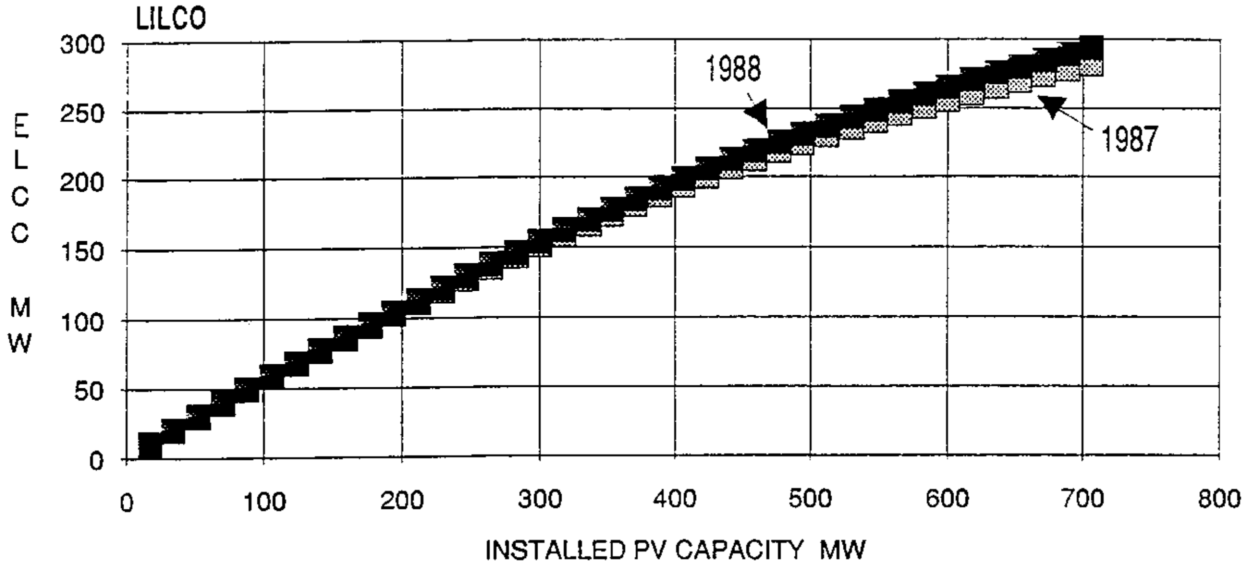


Figure 6.2.17: Effective PV Load Carrying Capability (ELCC) as a function of installed PV capacity for LILCO -- tracking PV array

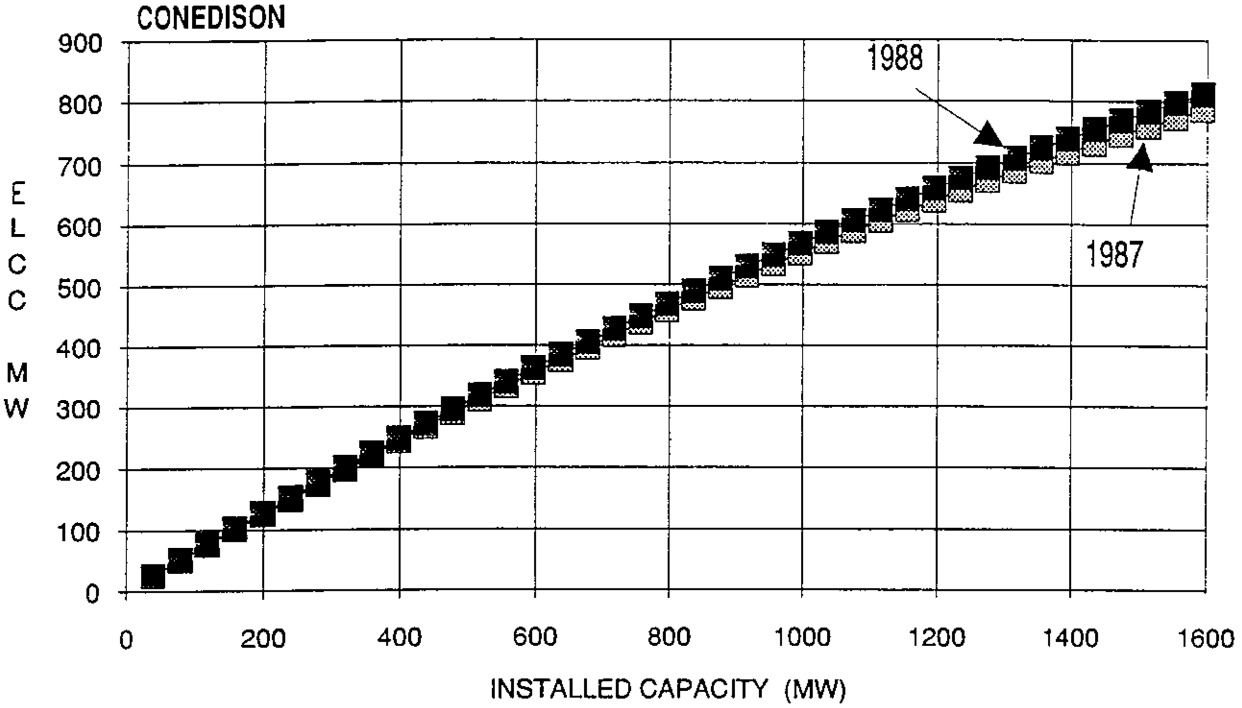


Figure 6.2.18: Effective Load Carrying Capability (ELCC) as a function of installed PV capacity for ConEd -- tracking PV array

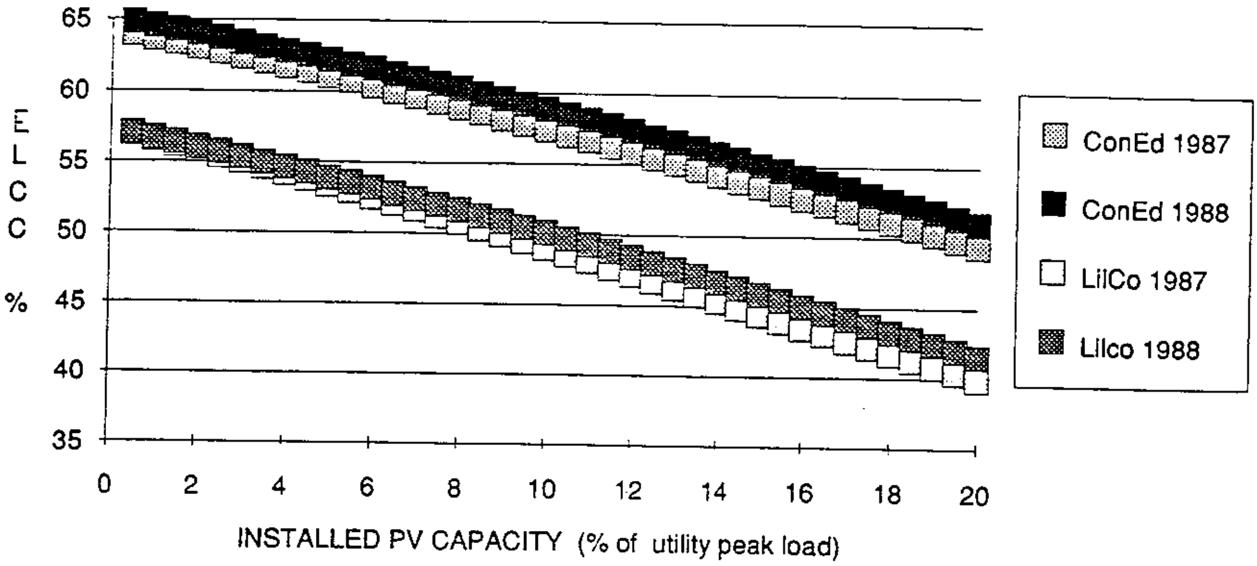


Figure 6.2.19: Relative ELCC as a function of installed PV capacity for LilCo and ConEd in 1987 and 1988 -- tracking PV array

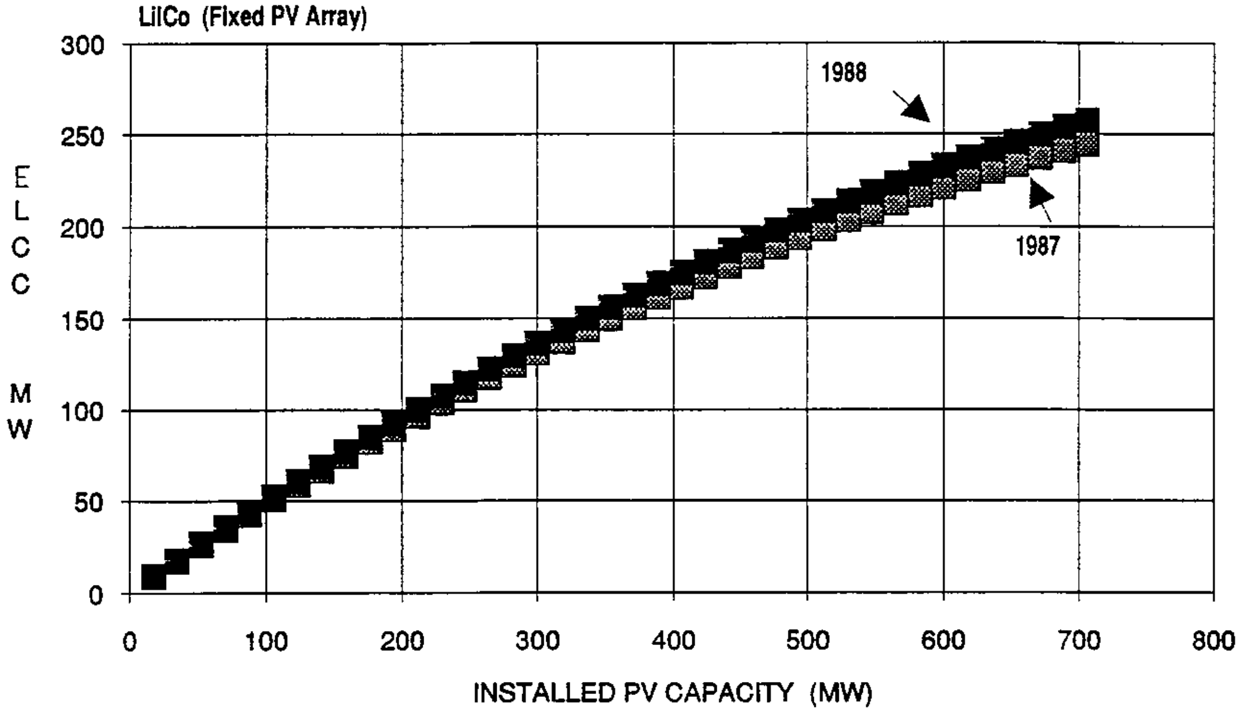


Figure 6.2.20: Effective PV Load Carrying Capability (ELCC) as a function of installed PV capacity for LILCO -- Fixed optimized PV array

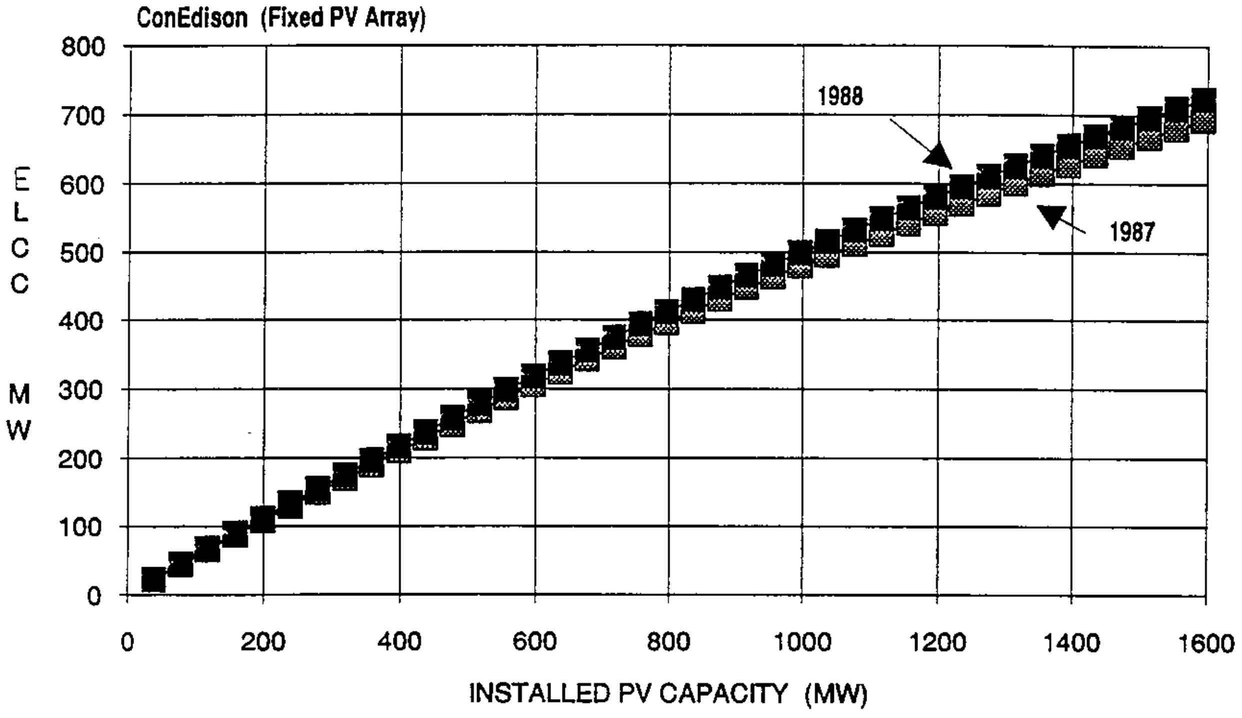


Figure 6.2.21: Effective Load Carrying Capability (ELCC) as a function of installed PV capacity for ConEd -- Fixed optimized PV array

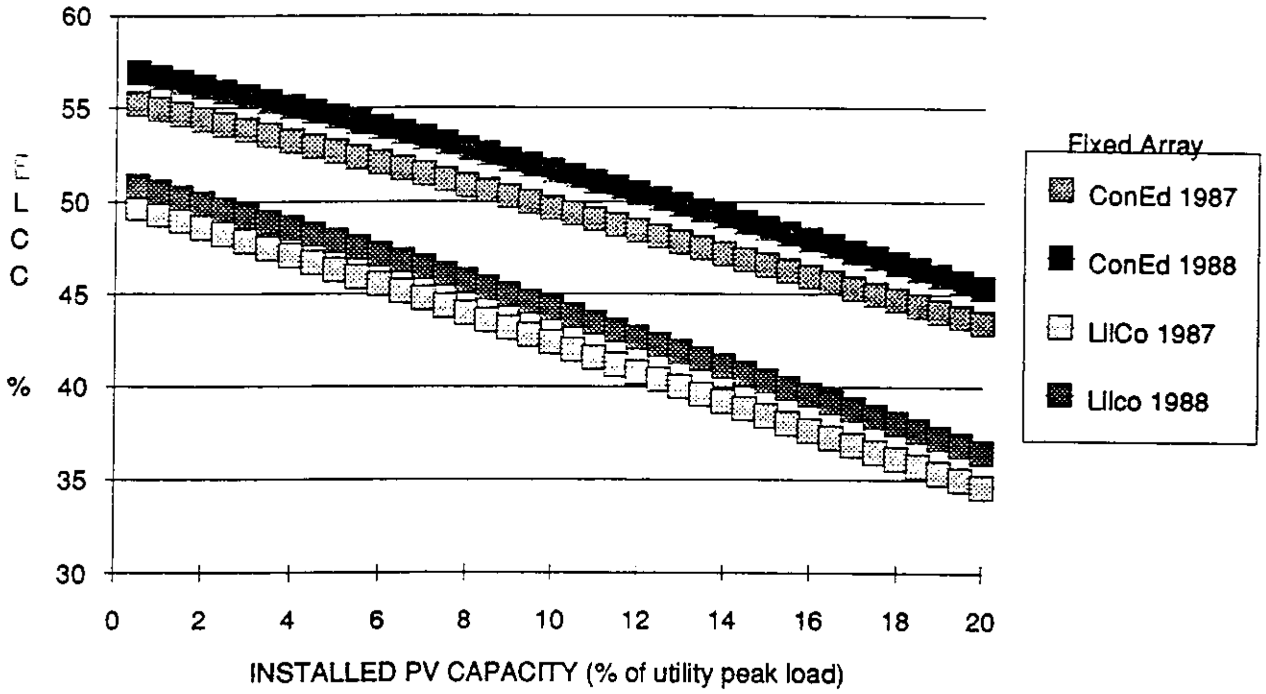


Figure 6.2.22: Relative ELCC as a function of installed PV capacity for LilCo and ConEd in 1987 and 1988 -- Fixed optimized PV array

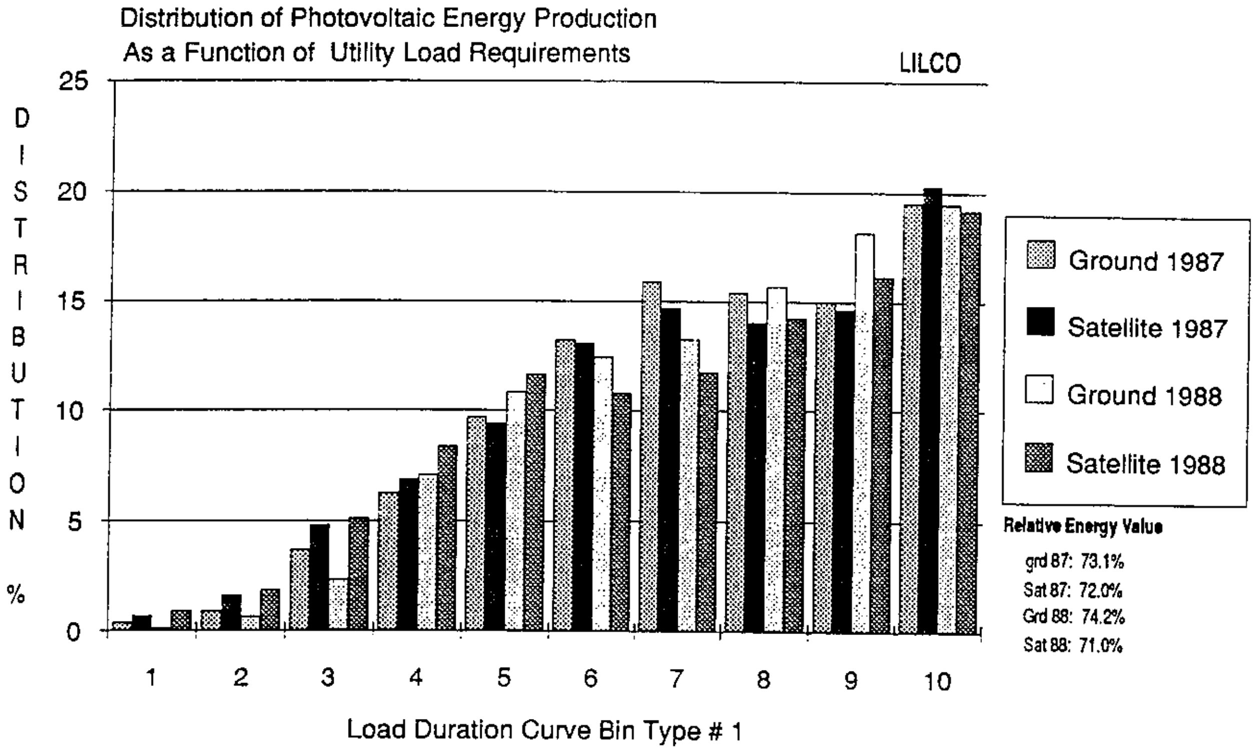


Figure 6.2.23: Distribution of the energy produced by a tracking PV array as a function of LILCO's load requirements (time axis bins on Fig. 11) in 1987 and 1988 as calculated from ground and satellite data

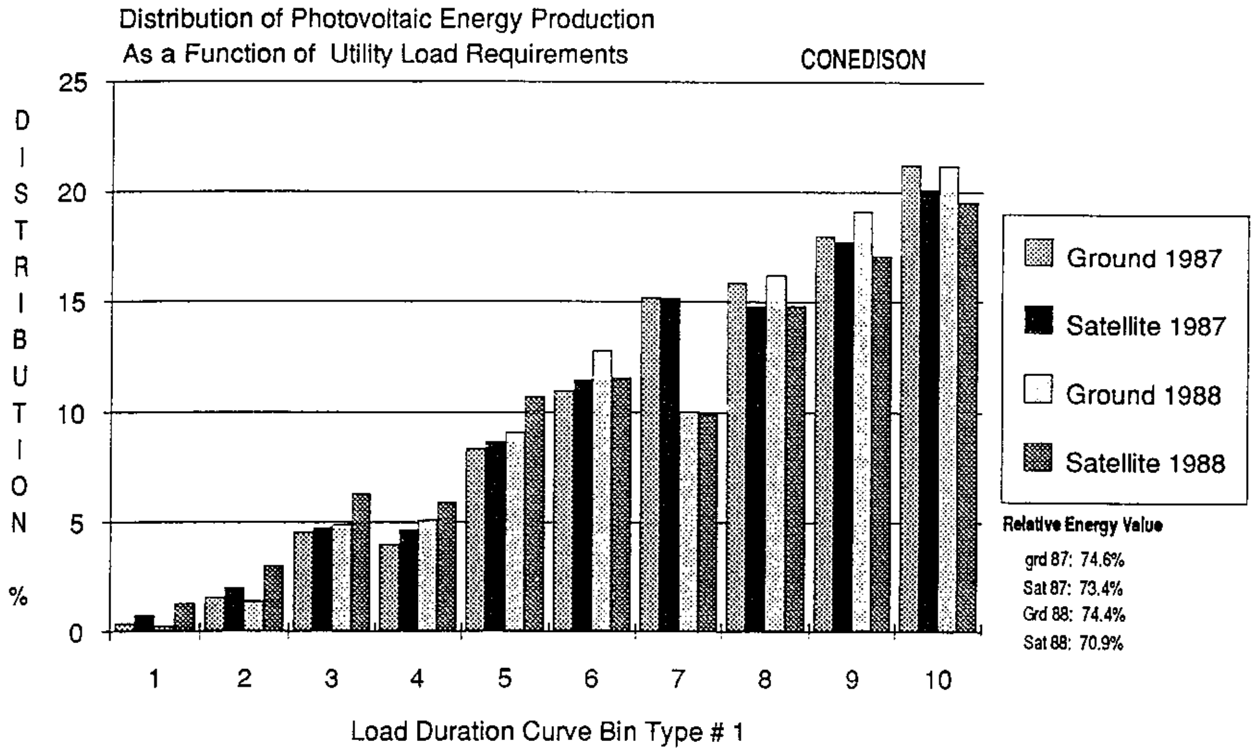


Figure 6.2.24: Distribution of the energy produced by a tracking PV array as a function of ConEd's load requirements (time axis bins on Fig. 11) in 1987 and 1988 as calculated from ground and satellite data.

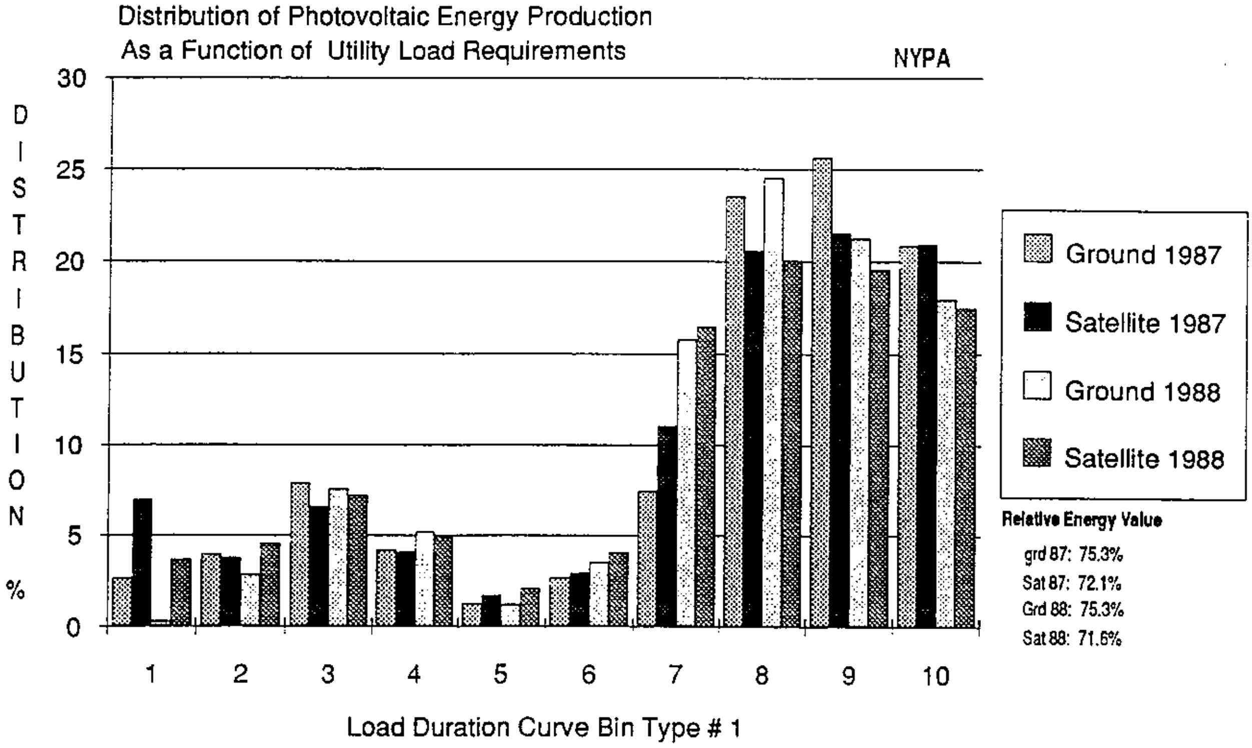


Figure 6.2.25: Distribution of the energy produced by a tracking PV array as a function of NYPA-SENY's load requirements (time axis bins on Fig. 11) in 1987 and 1988 as calculated from ground and satellite data

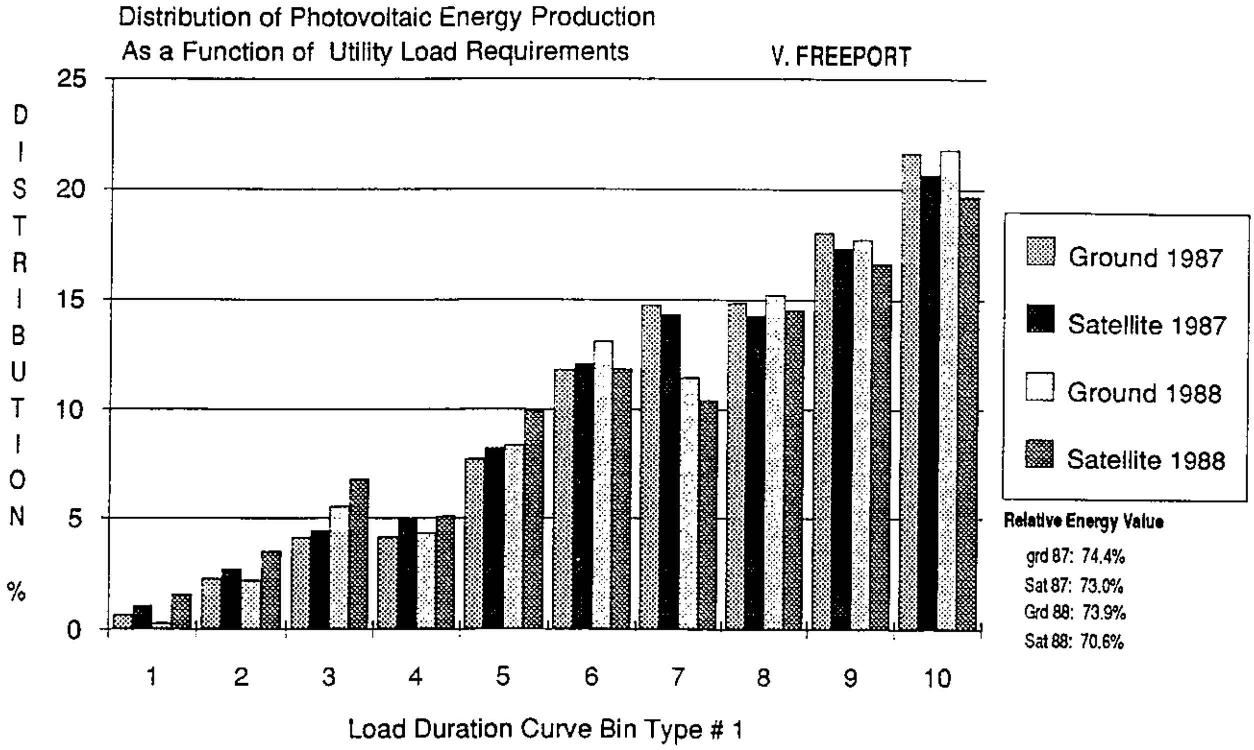


Figure 6.2.26: Distribution of the energy produced by a tracking PV array as a function of V. Freeport's load requirements (time axis bins on Fig. 11) in 1987 and 1988 as calculated from ground and satellite data.

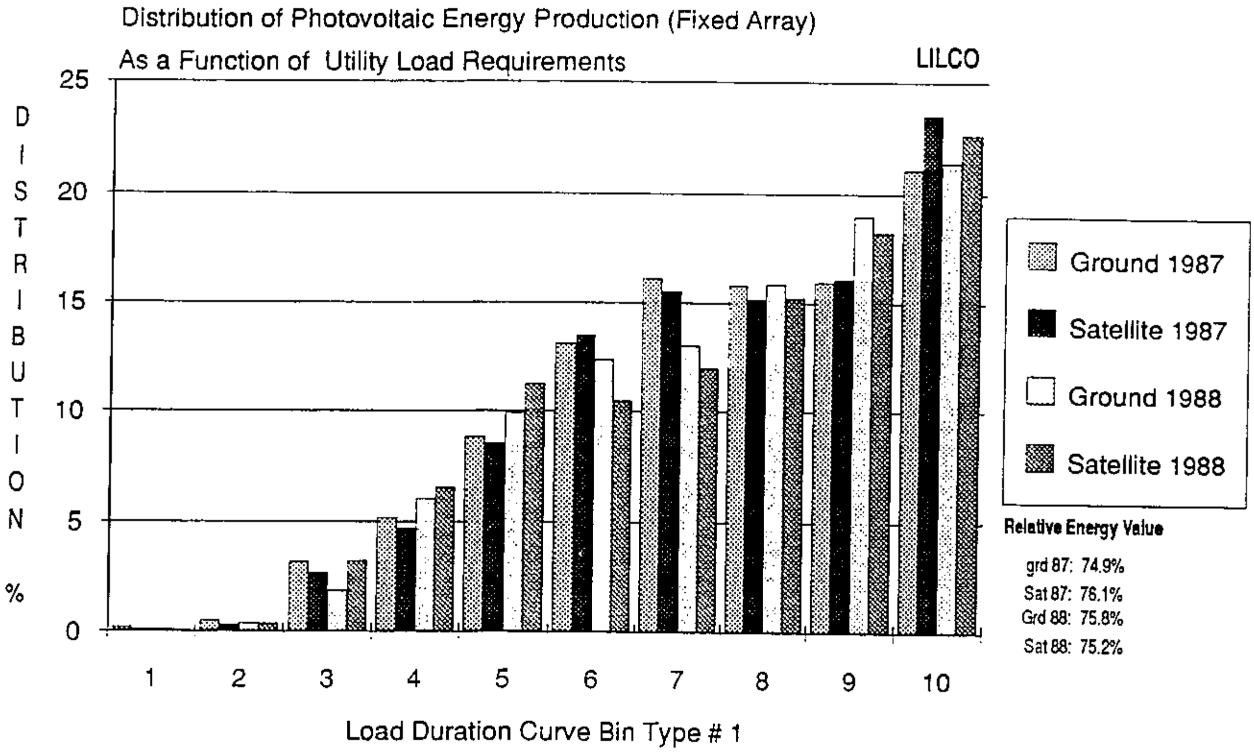


Figure 6.2.27: Distribution of the energy produced by an optimized fixed PV array as a function of LILCO's load requirements (time axis bins on Fig. 11) in 1987 and 1988 as calculated from ground and satellite data

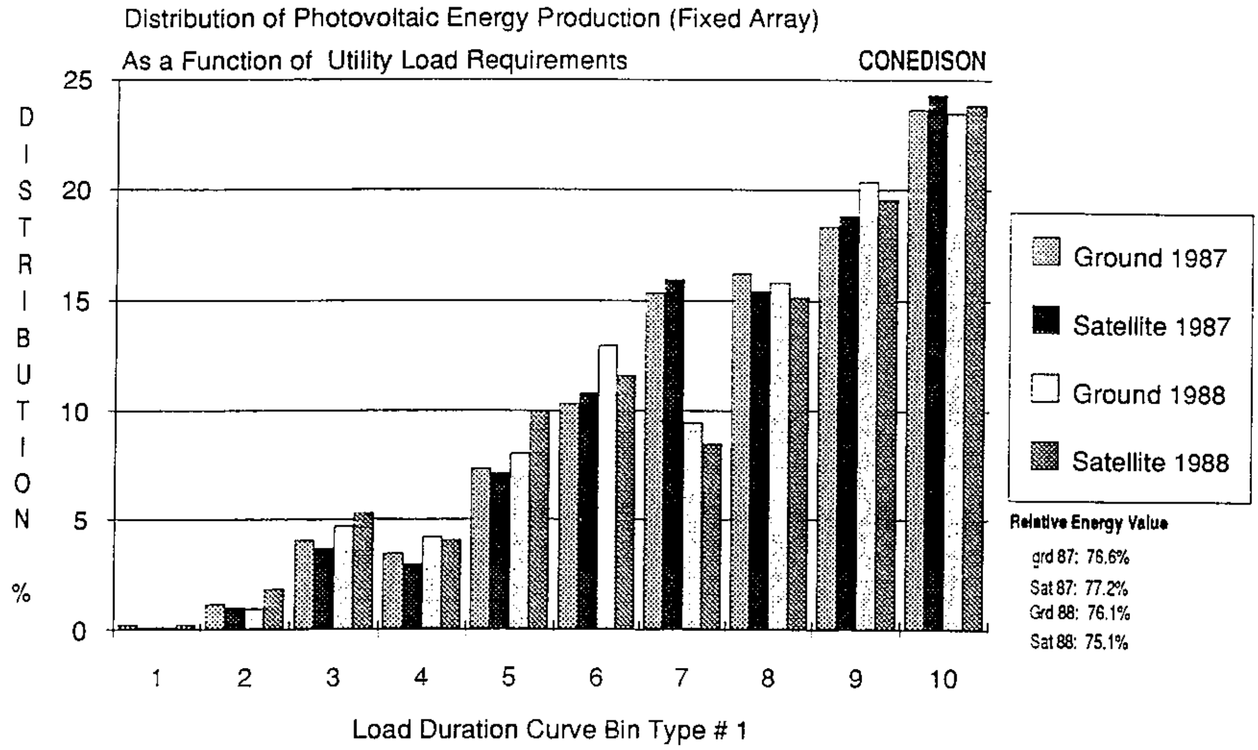


Figure 6.2.28: Distribution of the energy produced by an optimized fixed PV array as a function of ConEd's load requirements (time axis bins on Fig. 11) in 1987 and 1988 as calculated from ground and satellite data.

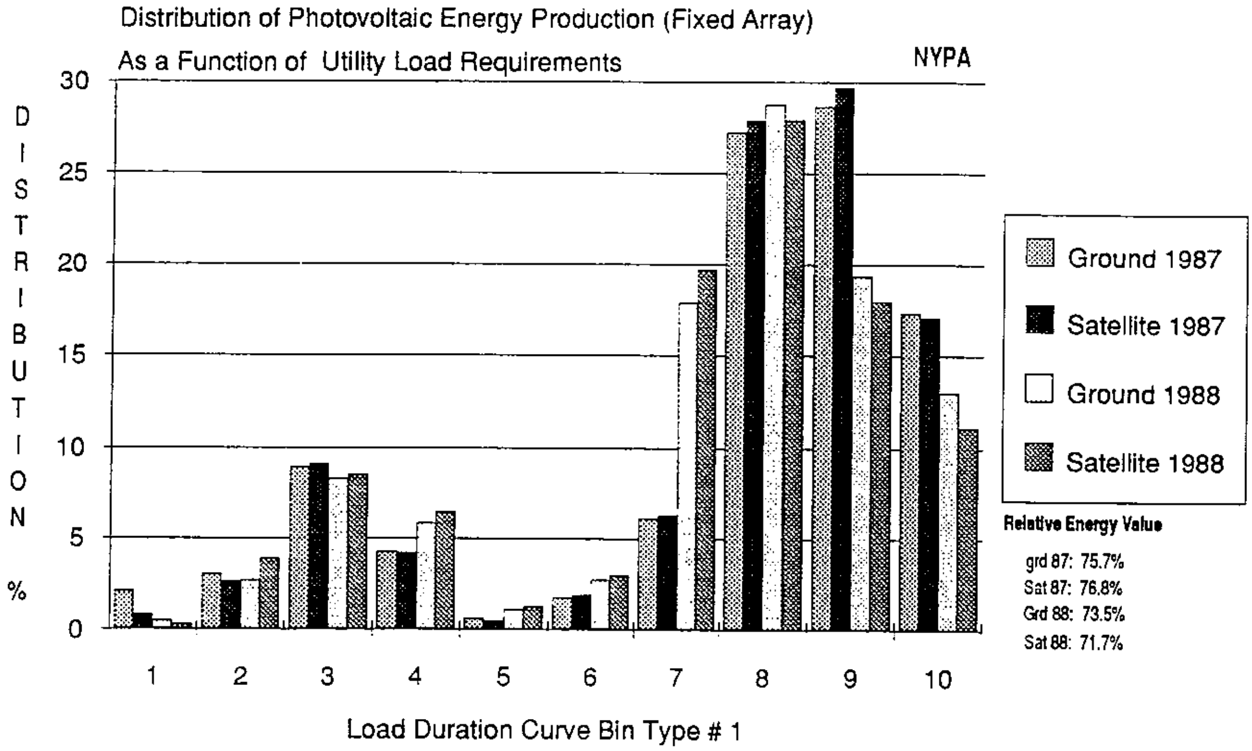


Figure 6.2.29: Distribution of the energy produced by an optimized fixed PV array as a function of NYPA-SENY's load requirements (time axis bins on Fig. 11) in 1987 and 1988 as calculated from ground and satellite data

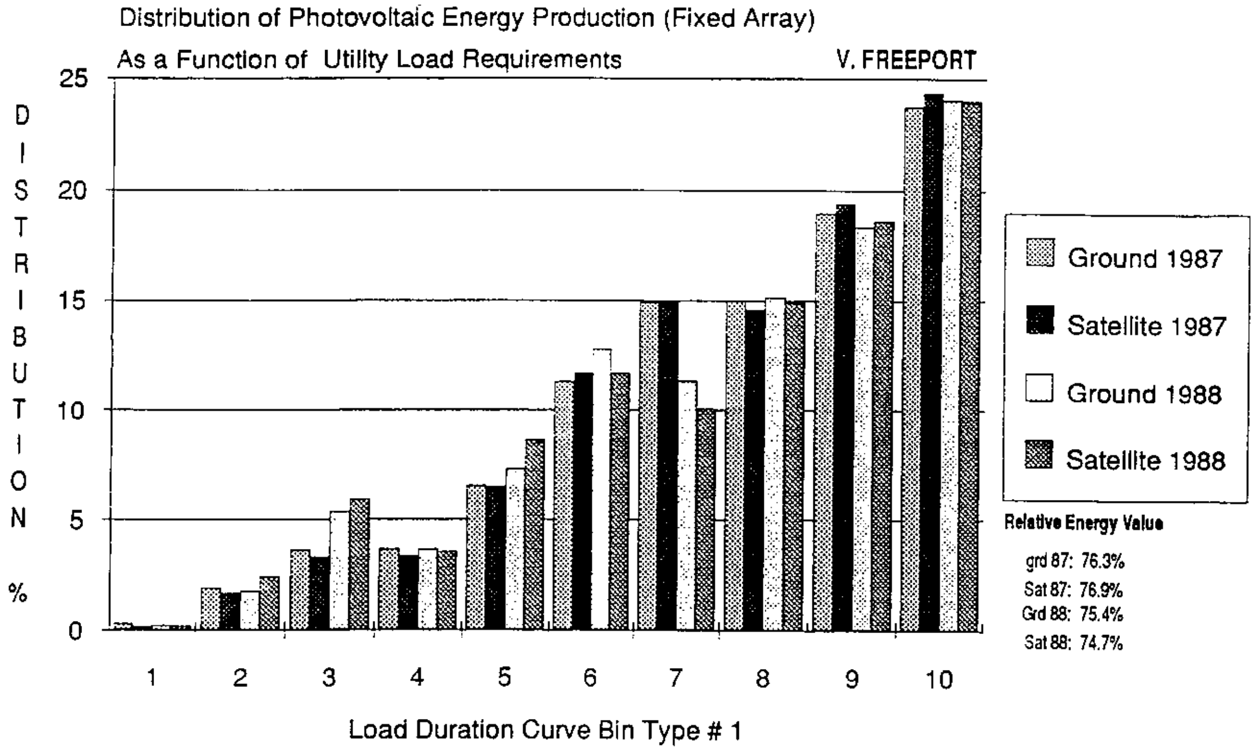
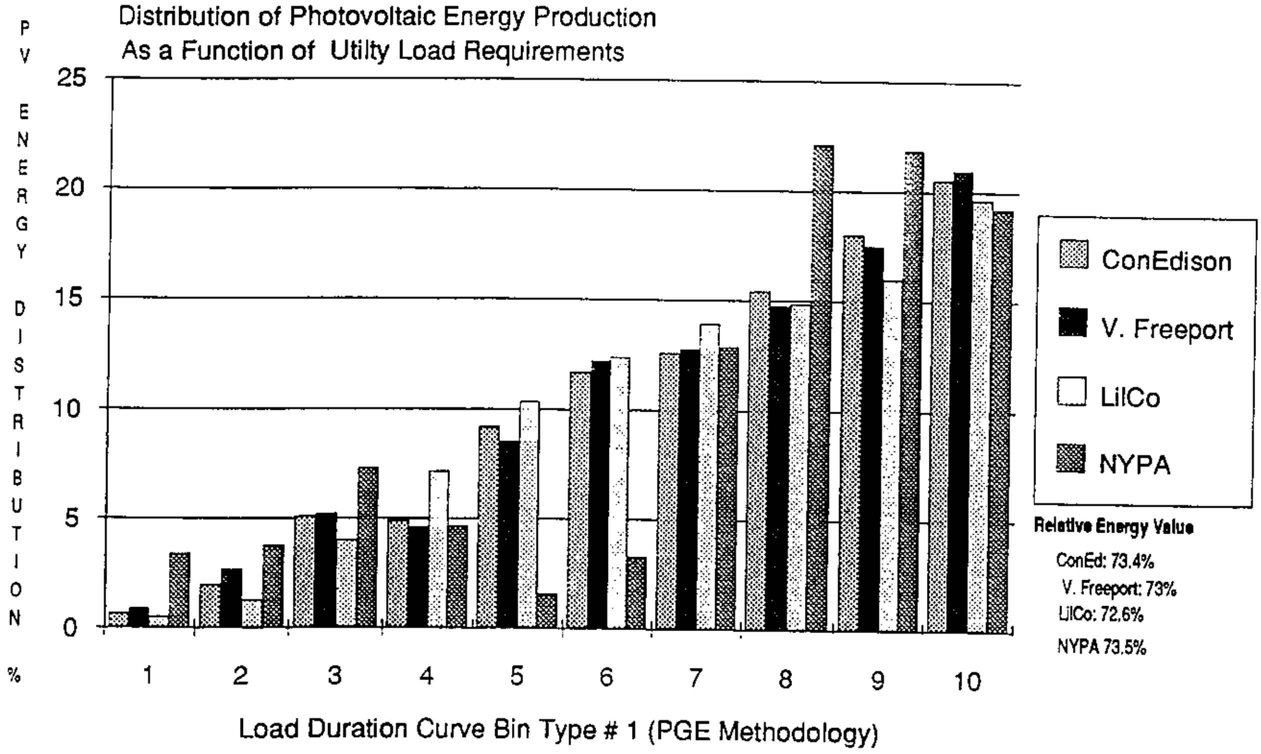


Figure 6.2.30: Distribution of the energy produced by an optimized fixed PV array as a function of V. Freeport's load requirements (time axis bins on Fig. 11) in 1987 and 1988 as calculated from ground and satellite data.



Relative Energy Value
ConEd: 73.4%
V. Freeport: 73%
LilCo: 72.6%
NYPA 73.5%

Figure 6.2.31: Comparative distribution of the energy produced by a tracking PV array as a function four utilities' load requirements (time axis bins on Fig. 11) -- 1987 and 1988 data

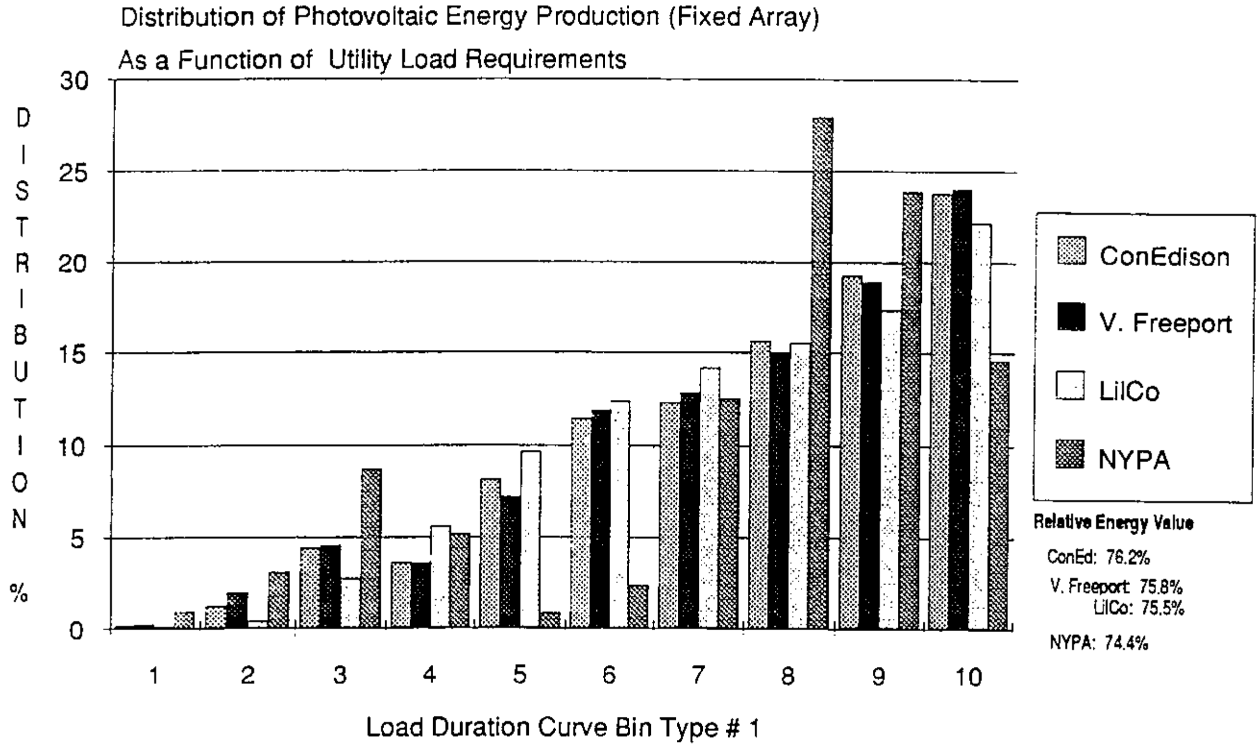


Figure 6.2.32: Comparative distribution of the energy produced by an optimized fixed PV array as a function four utilities' load requirements (time axis bins on Fig. 11) -- 1987 and 1988 data

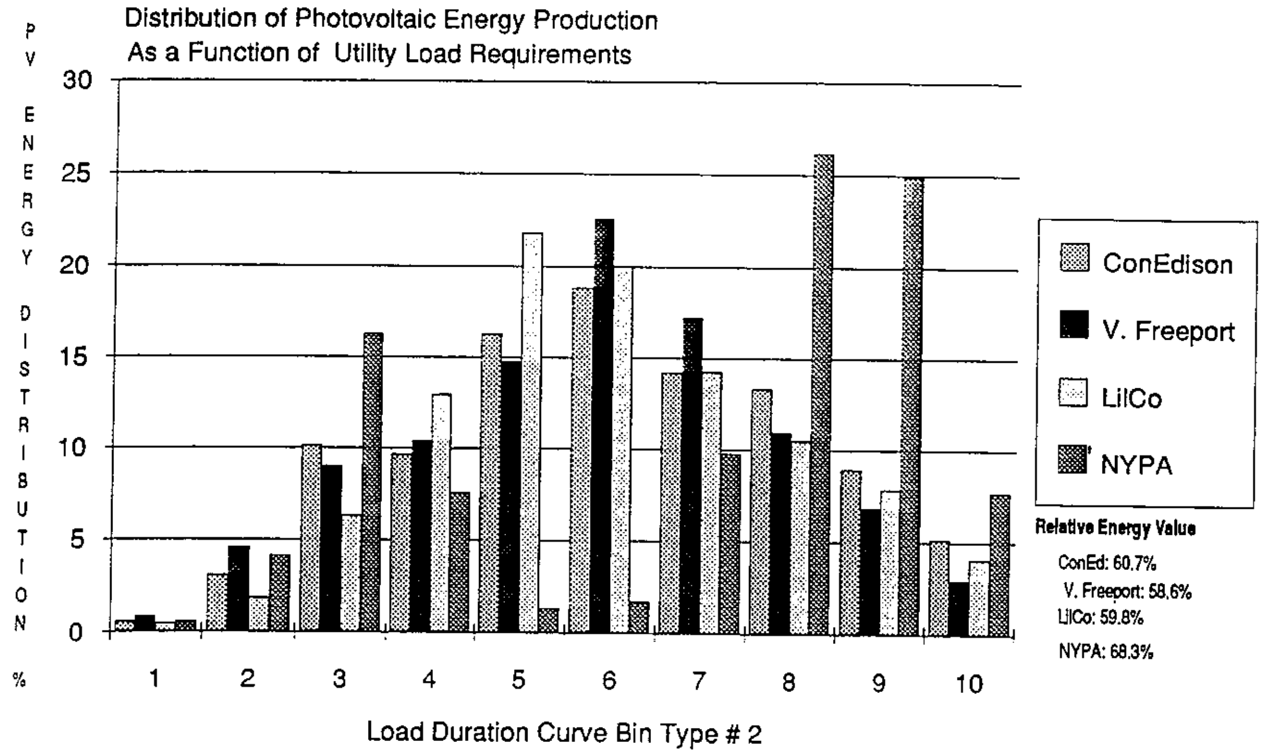


Figure 6.2.33: Comparative distribution of the energy produced by a tracking PV array as a function four utilities' load requirements (load axis bins on Fig. 11) -- 1987 and 1988 data

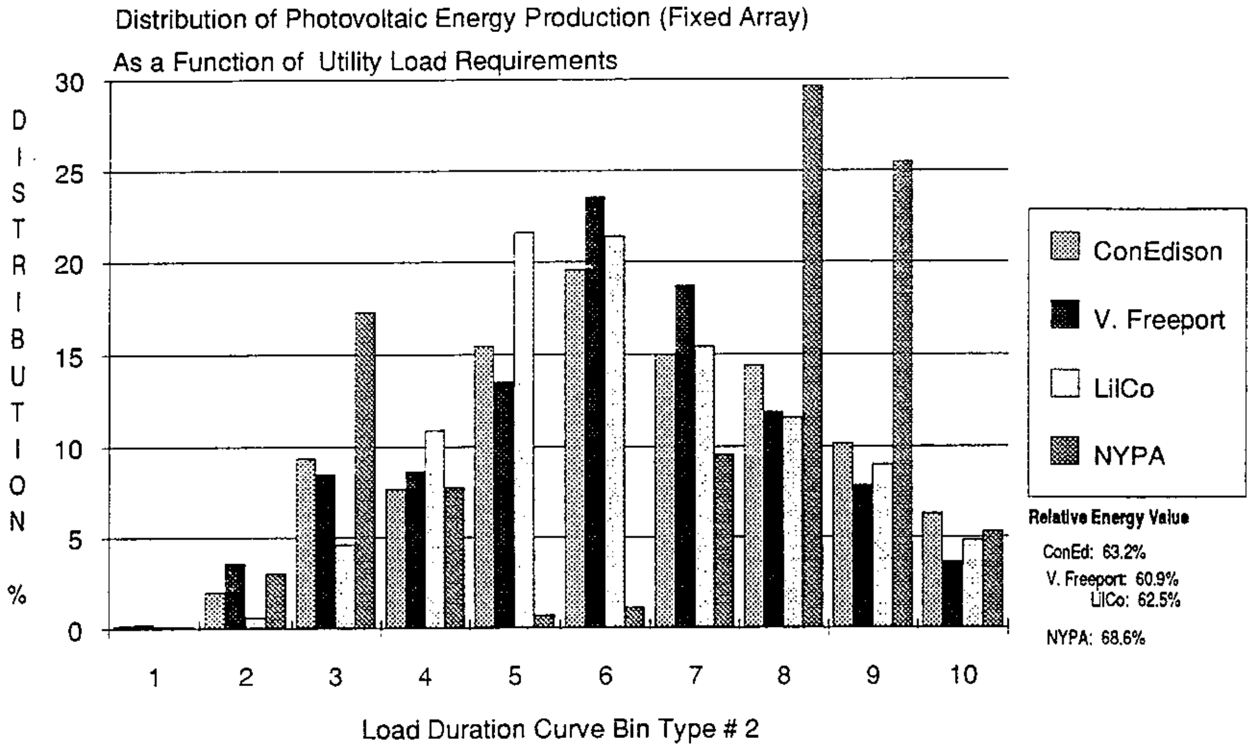
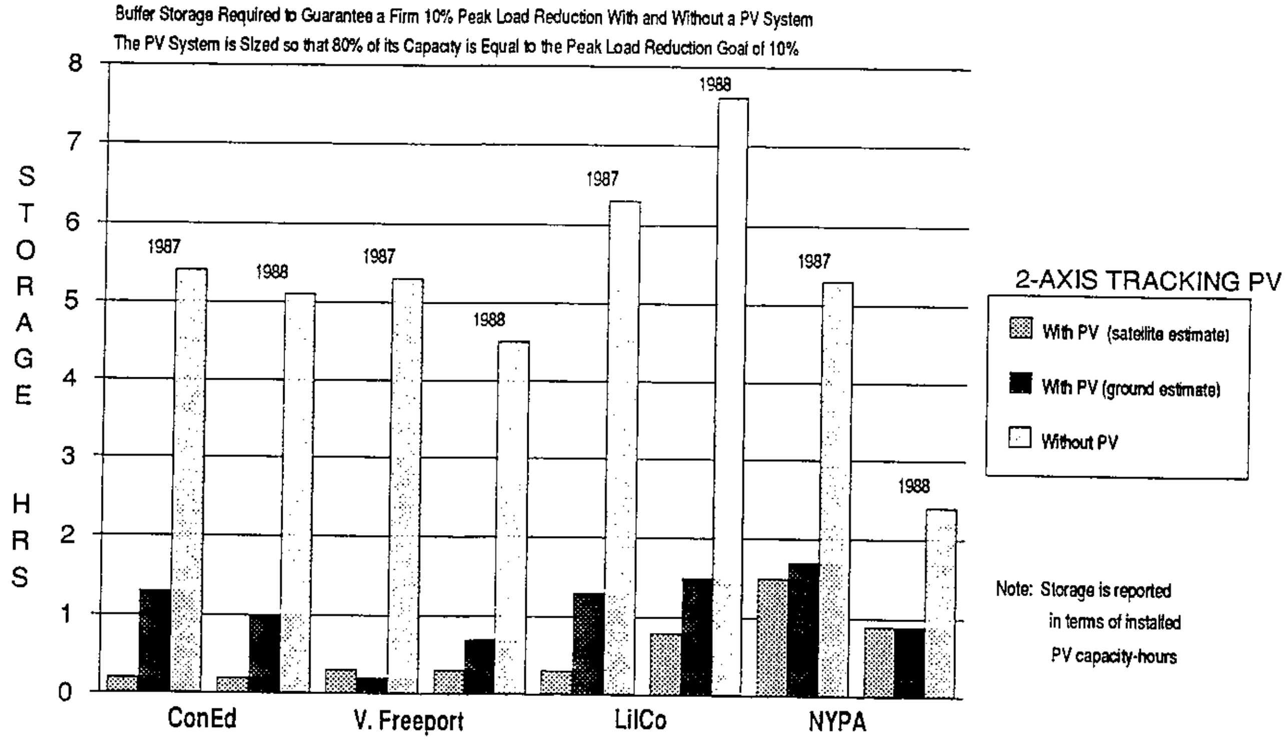


Figure 6.2.34: Comparative distribution of the energy produced by an optimized fixed PV array as a function four utilities' load requirements (load axis bins on Fig. 11) -- 1987 and 1988 data



Note: Storage is reported in terms of installed PV capacity-hours

Figure 6.2.35: Minimum storage required to guarantee a firm 10% peak load reduction with and without a PV system (Tracking). The PV system is sized so that 80% of its capacity is equal to the peak load reduction requirement of 10%. Storage values computed from ground measurements are compared to values derived from satellite data

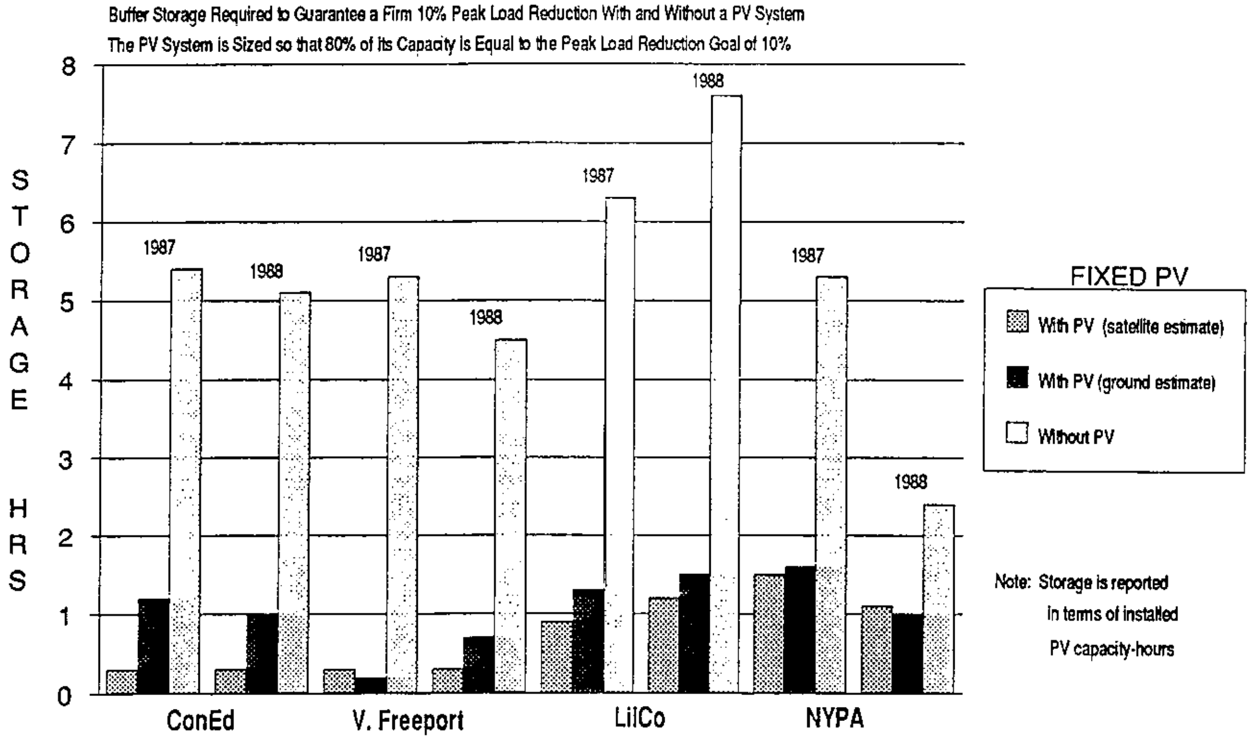
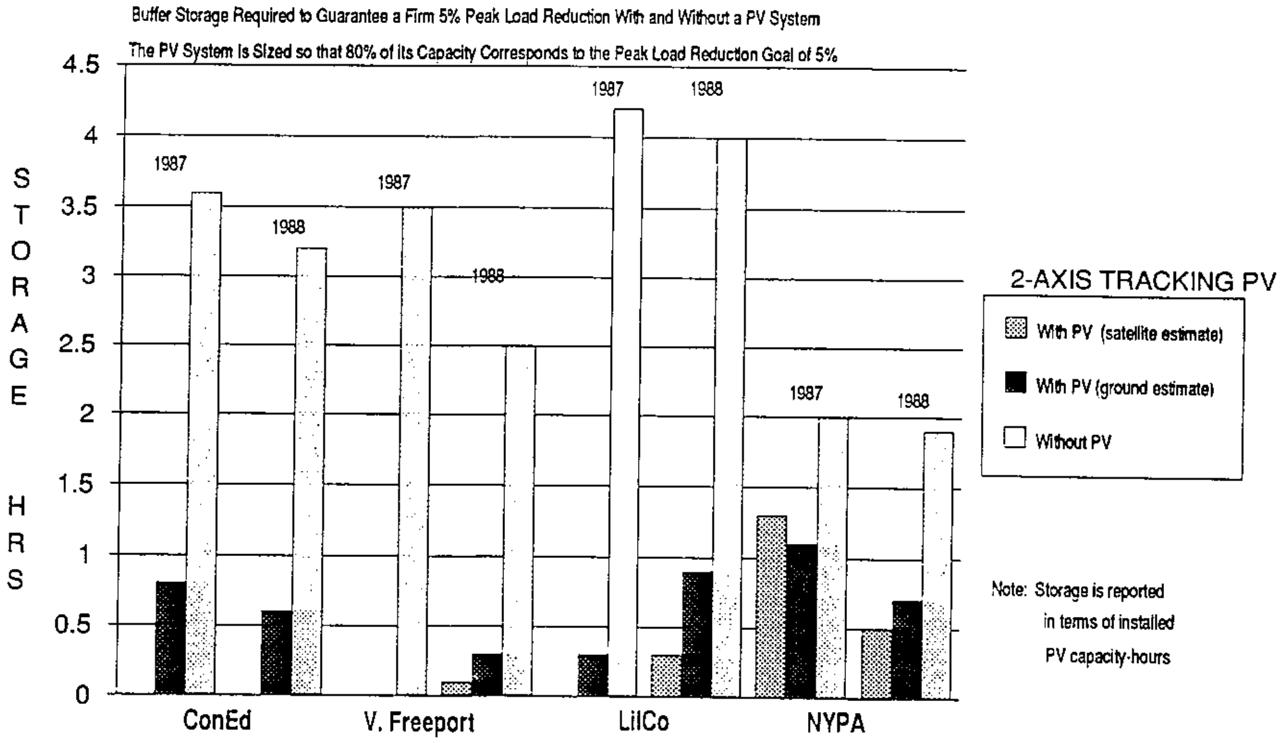


Figure 6.2.36: same as Fig. 35, but optimized fixed PV system


Buffer Storage Required to Guarantee a Firm 5% Peak Load Reduction With and Without a PV System

The PV System Is Sized so that 80% of its Capacity Corresponds to the Peak Load Reduction Goal of 5%

Note: Storage is reported in terms of installed PV capacity-hours

Figure 6.2.37: Same as Fig. 35, but 5% peak load reduction goal

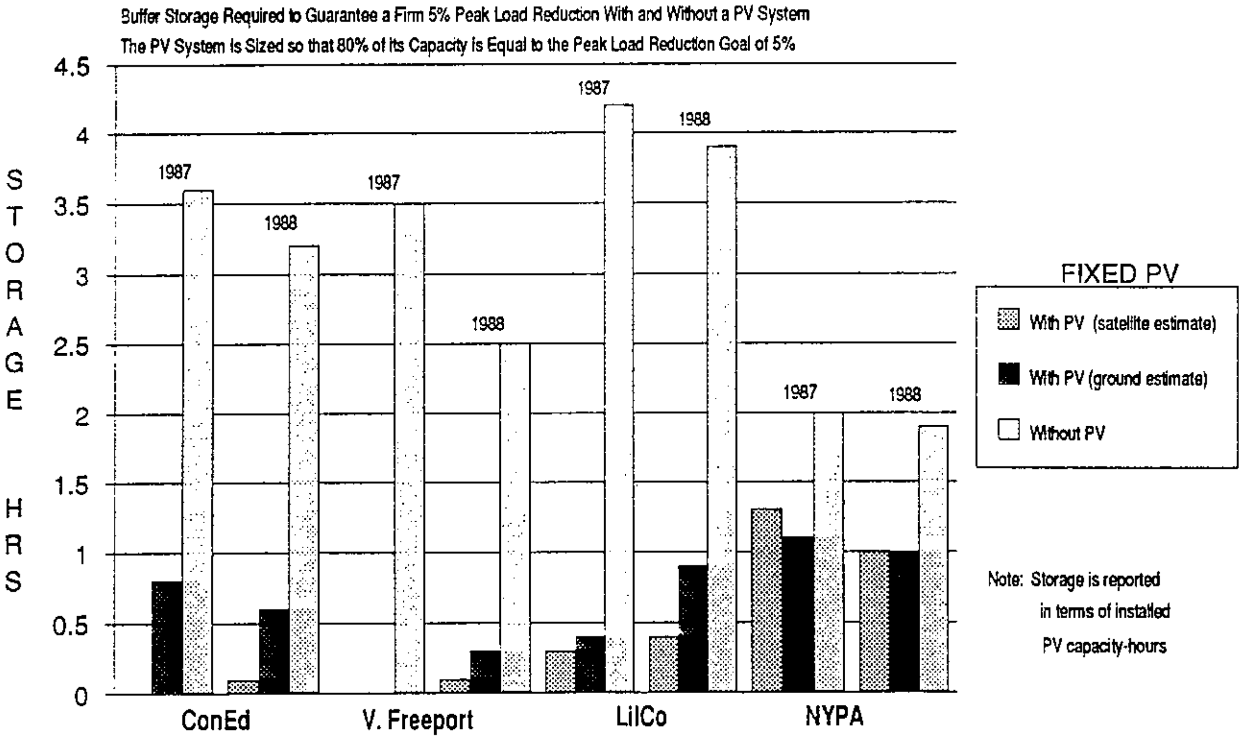


Figure 6.2.38: same as Fig. 37, but optimized fixed PV system

CHAPTER 7

7. Summary of validation results

Summary prepared by: *Antoine Zelenka*
Swiss Meteorological Institute

A synopsis of the methods whose performance has been assessed is attempted in Fig. 7.1. Table 7.1 recapitulates the overall cross-validation and validation results respectively, network by network, for all stations and all days which could be accessed by a given method at a given time. Sample sizes markedly vary between ground-based and satellite-based techniques for the DWD and ANETZ networks because the METEOSAT images have not been treated for each observation day. The kriging sample is very limited for the USNEU network. Therefore, Table 7.2 reports the results for congruent sample sizes, the days now being determined by the method with the most limited validation period. Entries in the tables have the standard format from Volume 3 (Results):

MM (ND)
MBE / MBE%
RMSE / RMSE%
MAX% (MON)

where MM: meas. mean in $MJm^{-2}d^{-1}$; ND: Number of observed daily totals;
MBE, resp. RMSE: absolute MBE, resp. RMSE in $MJm^{-2}d^{-1}$;
MBE%, resp. RMSE%: relative MBE, resp. RMSE in % of MM.
MAX%: Maximum monthly RMSE% for an individual site and month (MON).

Modelling based on interpolated cloudiness in the analysis for numerical weather predictions (NWP) exists for the SMHI network only. Validation of the principal components method for the DWD network cannot be reported here because, according to the German national project, the analysis has been performed for monthly averages and not for sequences of individual days. At this level of monthly averages, maximum and minimum validation RMSE% for the 29 sites of the DWD network read respectively: 15.2% (Jan.) and 4.2% (May) for weighted averages, 16.9% (Dec.) and 4.6% (June) for EOFs in conjunction with land-use differentiation, 9.6% (Dec.) and 2.9% (June) for kriging [7.1]. It should, however, be noted that the method is applicable at the level of individual daily totals, as some interpolation with respect to time is always possible for the four principal components considered (see, e.g., Fig. 3.5.3).

The RMSE% behaviour of extrapolation (nearest neighbor approximation) versus extrapolation distance has been determined as

$$\text{RMSE\%} = 2.2 + 2.5\sqrt{d_{km}}\ [\%], \qquad 4\text{km} \le d_{km} \le 165\text{km}, \tag{7.1}$$

while that of gravity interpolation (weighted averages) versus mean, generalized distance within the search radius was found to follow the relation

$$\text{RMSE\%} = 2.6 + 1.7\sqrt{d_{km}}\ [\%], \qquad 9\text{ km} \le d_{km} \le 196\text{ km}. \tag{7.2}$$

Table 7.1

Summary of overall (cross-) validations

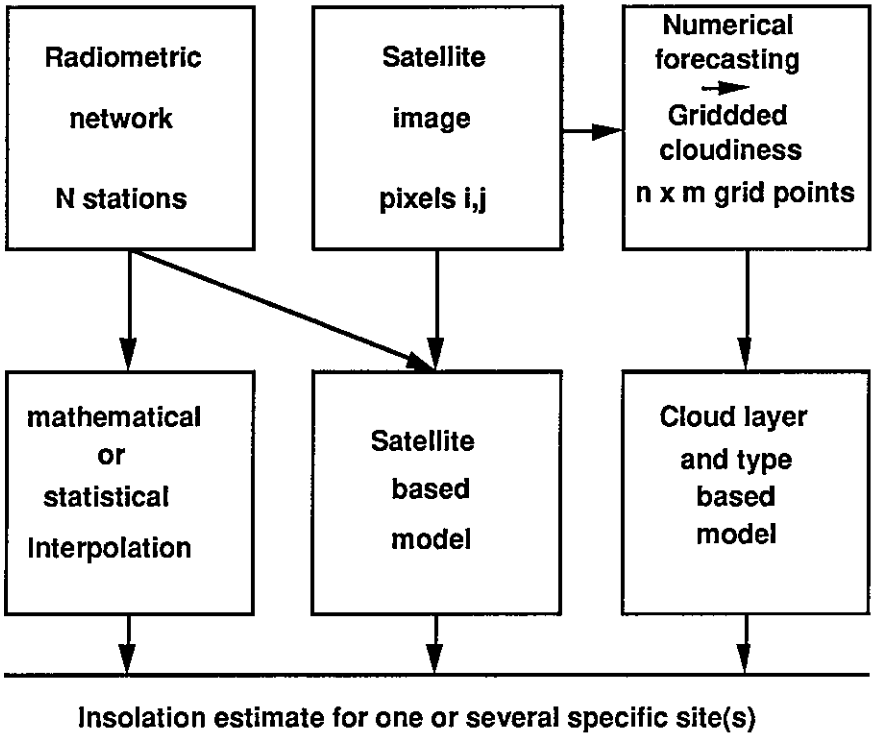
Network	Extrapolation	Gravity Interpolation	Kriging	Satellite	Meso-β NWP
PNW	14.36 (2993) -0.23 / -2 3.87 / 27 80 DEC	14.36 (2993) -0.16 / -1 3.26 / 23 72 DEC		13.94 (2739) / -7 / 20	
USWN	19.62 (4335) -0.30 / -2 2.65 / 14 41 MAY	19.62 (4335) -0.27 / -1 2.05 / 10 31 MAY			
USNEU	12.59 (938) -0.01 / 0 2.59 / 21 81 DEC	12.59 (938) 0.09 / 1 2.40 / 19 84 DEC	19.26 (287) 0.08 / 0 2.91 / 15	12.59 (938) / 2 / 21	
DWD	9.80 (10538) -0.06 / -1 2.64 / 27 154 DEC	9.80 (10538) -0.07 / -1 2.11 / 22 126 NOV		12.64 (4239) / -4 / 17	
SMHI	15.81 (2573) -0.05 / 0 2.18 / 14 27 JUL	15.81 (2573) -0.07 / 0 1.96 / 12 27 JUL			15.41 (3530) -0.16 / -1 2.18 / 14
ANETZ	12.23 (12687) -0.11 / -1 3.11 / 25 115 DEC	12.23 (12687) -0.07 / -1 2.30 / 19 69 JAN	12.23 (12687) 0.00 / 0 2.02 / 17 49 JAN	15.20 (5632) / -1 / 21	

Table 7.2

Summary of (cross-) validations for congruent sample sizes

Network	Extrapolation	Gravity Interpolation	Kriging	Satellite
PNW	13.94 (2739) -0.23 / -2 3.78 / 27 80 DEC	13.94 (2739) -0.16 / -1 3.20 / 23 72 DEC		13.94 (2739) / -7 / 20
USWN	19.62 (4335) -0.30 / -2 2.65 / 14 41 MAY	19.62 (4335) -0.27 / -1 2.05 / 10 31 MAY		
USNEU	19.26 (287) -0.08 / 0 3.28 / 17	19.26 (287) 0.07 / 0 2.94 / 15	19.26 (287) 0.08 / 0 2.91 / 15	19.26 (287) 1.33 / 7 3.15 / 16
DWD	12.64 (4239) -0.09 / -1 3.18 / 25 (58 OCT)	12.64 (4239) -0.12 / -1 2.54 / 20 (47 OCT)		12.64 (4239) / -4 / 17
SMHI	15.81 (2573) -0.05 / 0 2.18 / 14 27 JUL	15.81 (2573) -0.07 / 0 1.96 / 12 27 JUL		
ANETZ	15.20 (5632) -0.13 / -1 3.43 / 23 (93 OCT)	15.20 (5632) -0.04 / 0 2.52 / 17 (39 JUN)	15.20 (5632) 0.01 / 0 2.23 / 15 (35 OCT)	15.20 (5632) / -1 / 21

The maximum monthly RMSE% at one site is set in parenthesis when the sample does not encompass a whole year.



Extrapolation (next neighbour)	**Cologne/DWD/ METEOSAT**	**SMHI meso-β model**
Weighted averages (gravity)	**Tarpley/NESDIS/ GOES**	**SERI mesoscale insolation estimates**
Geostatistics (kriging)	**Héliosat/CTAMN/ METEOSAT**	
Principal components (EOFs)		

Fig. 7.1: Synopsis of the validated methods. The left column is described in Chapter 3, the median in Chapter 4 and the right one in Chapters 5 and 6 respectively. Note that analysis for numerical weather prediction relys heavily on observations by meteorological networks, as well as on optimum interpolation (which is identical to ordinary kriging).

It would have been of greatest interest to derive the corresponding relation for kriging, where definition of a mean distance is equally possible because the software packages allow access to the (optimum) weights. However, complete results exist only for one network and this is not enough to ensure the required generality (recall that the scatter observed in Figs. 3.1.15 and 3.3.16 originates more from different behaviors among the networks than from inside the networks themselves).

Two dense networks, the U.S. South West and the Swedish experimental, obviously subtend more homogeneous insolation climates than the other networks. In both cases, the prediction precision achieved by interpolation (RMSE% around 10%) may be considered as sufficient. As to the Swedish network, it should not be forgotten that the sample covers Summer conditions only.

A conspicuous homogeneity concerning "critical" distances to observing sites emerges from the different analyses of different networks. In Section 6.2.4, relating to the USNEU network, it is concluded that "Concerning geographically extended solar resource assessment, the satellite was found to be a better source of data than a ground station if the distance from the ground station exceeds about 65 km. For a network, the satellite's precision would exceed that of the interpolated ground measurements if the considered point features a mean interpolation distance exceeding 95 km." In Section 5.4.6, relating to the Swedish experimental dense network in Summer, it is concluded that "Use of the meso-β model is definitely the best method for areas where the measurement stations are more than 100 km away and also in situations where the climate is very different from the neighboring stations. Also, in the range 50-100 km the meso-β model will produce values as good as those retrieved with the other methods. Although there is an indication that interpolation methods may perform better in the range 50-100 km, it certainly is more practical to use one method, i.e. meso-β, for all distances larger than 50 km." Similar conclusions would be arrived at for the other networks. This directly proceeds from Eqs. (7.1) and (7.2) and from the RMSE% of satellite-based, respectively of NWP-based estimates which all lie around 15 to 20%. In this context, Wilson's 1979 statement that "The Hay and Suckling (1978) relationship for example suggests that extrapolation should not exceed 60 km in order to maintain the error within ±15%" [3.1.7] acquires an almost prophetic character.

The 15 to 20% range discussed here applies to the RMSE% of yearly sequences of daily totals *averaged over all sites of a network.* Scrutiny of the detailed Tables in Volume 3 reveals that definitely less favorable conditions prevail for several sites and months, as also evidenced by the worst cases reported in Table 7.1. At the level of individual daily totals, estimations based on measurements from even relatively close sites can be erroneous by factors of two to five! In these cases, nonsense can be avoided only when additional, related information is available with spatial resolution equivalent to the distance of the coherence decay (the range) of the insolation field. Recognition of this situation is no less important than the other results of our investigations. Also note that Chapters 5 and 6 suggest several useful sources of such additional information.

References

[7.1] G. Czeplak, 1992. "Abschlussbericht DWD-Projekt 69". Deutscher Wetterdienst, Meteorologisches Observatorium Hamburg.

CHAPTER 8

8. Conlusions and recommendations

General

The choice among different methods for estimating global irradiation ("insolation" in the following) at sites where it is not measured will depend upon the available data, the required time scale, and the desired precision.

The available data may be (1) pyranometric measurements from more or less dense monitoring networks, (2) calculated values using models, (3) ancillary measurements which are input to the models (sunshine duration, meteorological observations and cloud opacities, satellite data etc.), and (4) ancillary information, as rastered orography or land-use, which directly or indirectly influences the distributions of the variables under consideration.

The achievable precision will be determined by the spatial distribution of the data (density and climatic representation), as well as the measurement and modelling uncertainties. Ground-measured insolation will generally be more accurate than insolation estimated from satellite data, because the latter is the outcome of a model. Therefore, when applied to ground-measured data, a given method (e.g., weighted averages) is expected to yield more precise estimations than when it is applied to satellite-derived data (provided, of course, that the spatial distribution is the same in both cases). Direct quantitative evaluation of the uncertainties generated by the input data is only possible with kriging, because the kriging variance is modified by introduction or removal of a nugget effect equal to the measurement's RMSE.

The time scale influences the choice of the method because increasing averaging time is expected to increase the smoothness of the resulting insolation field. So, for a given method and in most cases, interpolation of hourly values will be less accurate than interpolation of daily values, which, in turn, will be less accurate than interpolation of monthly averages. Or, for a fixed precision, more powerful (and time consuming) methods will be required at the level of hourly values than at the level of monthly averaged daily values. Spatial and time scales are somewhat coupled by the fact that sequences of hourly or daily values (for simulations etc.) are usually required for one specific site or for a limited area, while long term means are best for gridded information for state-wide mapping purposes.

Specific

All numbers stated below refer to yearly sequences of daily sums, but there are no reasons why the hierarchy that they define among the various techniques should change as the integration time becomes shorter or longer [3.3.7]. Evidence supporting this expectation can be found in Chapter 7 for monthly daily averages and in Section 6.2 for sequences of hourly sums.

A. Univariate techniques

In this case, the sole source of information is a discrete, more or less scattered and irregular distribution of sites with known (measured or physically modelled) insolation.

Table 7.1 demonstrates that extrapolation (direct substitution of data from a nearby site) can be used only at locations with a slowly varying (low spatial gradient) insolation field. This condition is found, e.g., in the south of Sweden during the summer. It is recommended that solar radiation experts be consulted if no previous experience exists with respect to the spatial gradient of the insolation field (for average inter-site distances) in the region of interest. Without considering gradients, use of data from the closest measurement site could be greatly in error (RMSE greater than 100% for an individual site and month are possible), especially at coastal and mountainous locations. Note that at many locations in the Great Plains of the United States, extrapolation over a distance greater than 200 km could provide a MBE of 5% for monthly means of daily totals. Even in the Great Plains, however, familiarity with the insolation gradients is vital, since the isolines trend North-South in the summer and East-West in the Winter.

As far as can be determined from the somewhat sparse comparisons given in Tables 7.1 and 7.2, "gravity" interpolation with adequately selected vertical scale factors is about as efficient as kriging and "optimum interpolation" respectively. In some obvious situations, as coasts or valley systems, directional factors may also be required. It can be recommended for application everywhere where network data are easily accessible for sufficiently long time intervals to allow a safe determination of the scale factors and search radii. Given that trustworthy scale factors and search radii are *a priori* available, then "gravity" interpolation (and some other weighted averages) ought to be selected as the most efficient (accuracy vs. cost) technique.

If, however, adequate knowledge in the form of (3-D) directional scale factors cannot be acquired (e.g., due to a lack of pertinent data) then use of geostatistics (e.g., kriging) becomes mandatory. The use of geostatistics is also required when prediction alone is not sufficient, that is, when information about the achieved prediction precision is as important as the prediction itself. Commercial and public domain geostatistical software packages for PCs are available, but the user should expect to undertake a minimum training effort for operating them succesfully. This effort becomes important when all possible configurations (including non-stationarity) of the insolation field should be mastered. The solar user community could help accelerate the development of expert systems in this domain.

When the average inter-site distances become of the order of 50 to 100 km (the exact distance depends on the spatial gradient of the insolation field, which may vary seasonally) then use of estimates extracted from the images of geostationary (meteorological) satellites is recommended. These limiting distances are not only the horizontal ones but also the generalised ones (i.e., with directional scale factors).

Satellite-based estimates may be spatially "continuous" in the sense of contiguous pixel values valid for areas between 1 and 8 km on a side. The estimates are still not easy to obtain because much effort must be invested in data acquisition system management, keeping track of and adapting to changing spacecraft operations, and in image processing. If the estimates can be obtained from primary data *via* national weather services, as is the case in Germany or in the United States, then the data are in the format prescribed by the service, at the service's convenience. If one individually operates a modest system with secondary data input to a desktop system as we have also done for this investigation, then increased independence and flexibility must be paid for in terms of precision losses.

Numerical weather prediction models are powerful tools in meteorology. Calculating a weather forecast (i.e., integrating the model's equations) requires 3-D gridded initial conditions, the so-called analysis. This analysis relies on interpolation of meteorological observations and on remotely sensed data. It is possible to extract cloud opacities from the analysis and, thus, to calculate insolation for each grid point of interest. High resolution models have mesh widths of some 20 km and (good) meteorological models for insolation achieve better than 20% RMSE. Thus, this data source is about equivalent to that provided by the satellite-based techniques. Furthermore, it has an eminently important development potential: to provide *insolation forecasting*. The procedure remains unchanged for this step, but it is now applied to opacities from the integration (i.e., from the forecast clouds) rather than to opacities from the analysis (i.e., from interpolated observations).

B. Multivariate techniques

For the fairly varied (yet not extreme) configurations considered in the study, the supplementing techniques based on only one data source have been found to operate with RMSE%s around 17 to 20% for periods of one year. Also, estimates from individual days can be in error by factors as large as five at certain sites. If we wish to lower the annual or seasonal RMSE%, or limit the individual daily site specific errors to physically meaningful values, then optimal use of any available additional information is mandatory.

For example, sunshine duration correlates well with insolation at the monthly average level. It is easier to measure than insolation itself and is, therefore, still a key parameter in climatology with corresponding network density. Monthly average insolation network data can, thus, be sensibly supplemented if the sunshine duration sites are reasonably clustered (Sections 5.2 and 10.4). Alternatively, sunshine duration could assume the role of a covariable in a cokriging process whose principal variable would be the network's insolation. Meteorological (cloud) observations are a similar source of information (Section 6.1).

There is another, major reason for preaching intensive use of ancillary information. The evaluation process will never render spatial features of the insolation field whose scale is below the minimum distance separating two neighboring stations in the network. In a certain sense, these features slip through the meshes of the estimation's sieve. The only possibility to retreive them is to use some related quantity which is available with the relevant resolution.

From experience gathered in the course of this study, multivariate geostatistics (cokriging) appear to be the ideal method for the data fusion process, where precise but undersampled (networks can never be dense enough) insolation is supplemented with less precise, yet sufficiently well sampled satellite-based estimates (Section 5.3). Alternatively, the covariable could be the insolation modelled from the highly resolved NWP analysis (Section 5.4). And even digital terrain models or land-use rasters available in the majority of industrialized countries will help in many cases (Sections 5.2 and 6.1).

Summary

Each situation in each country is unique and requires study to select the best method for site specific and grid (raster) analyses. Features which deserve consideration in any case are:

- Unique climates: Delineating regions with similar spatial gradients (insolation) can save time by selecting appropriate methods for different regions.
- Different network densities: The degree of sophistication required can decrease with increasing network density.
- Different ancillary data: Always search for correlated, ancillary information with optimum spatial resolution (sunshine duration, meteorological observations, remotely sensed (satellite) data, digital terrain models, etc.)
- Different measurement uncertainties: Assess their importance in the final estimates (adding noise to the analyses will degrade rather than upgrade the results).
- Different applications: Some applications are less sensitive to RMSE scatter but more sensitive to biases or *vice-versa*. Clear definition of impact of prediction uncertainties on the end-use product may help save useless efforts.

In the final analysis, we recommend the appropriate and combined use of all methods supported by the data available and the requirements for a given study.

Definition or supplementing of monitoring networks, or of targets for satellite-based estimates when continuous imaging is not foreseen, should always be accompanied (or, better preceeded) by geostatistical analyses which provide criteria for optimum siting.

Any insolation prediction must be based on measurements (ground or satellite) and knowledge formalized in a computer code. But there is often some unused "knowledge" which usually generates uncertain information or an empirical description of facts. Within this wider and more complex framework it is necessary to define a strategy (rather than a model) which is able to integrate data, models, uncertain information, and "experiences". Therefore, as the ultimate conclusion of this report, we urge development of *expert systems* which would not only *support* (as Geographical Information Systems (GIS) already do), but also *guide* the data fusion process.

CHAPTER 9

9. References and textbooks

Instead of an epilogue, the reader will find here references to investigations or to textbooks of general interest. The last section is intended to provide access to recent advances in the field of models for insolation (and daylight availablity) on sloping surfaces for which the horizontal insolation treated here is a fundamental input. The selection is by no means exhaustive: listed are those titles which were often valuable guides to our work.

9.1 Overview of previous work

9.1.1 Spatial interpolation of insolation

J.E. Hay, 1983: "Structure of the Global Radiation Field in Canada". Rept. to Canadian Climate Centre, Atmospheric Environment Service, Downsview, Ontario M3H 5T4.

J.E. Hay and K.J. Hanson, 1985: "Evaluating the solar resource: a review of problems resulting from temporal, spatial and angular variations". *Solar Energy* **34**, 151.

P. Allerup and K. Frydendahl, 1986: "Variability in Space of Solar Radiation". EC Solar Energy Programme, Final Rept.. Danish Meteorological Institute, DK-2100 Copenhaguen, Denmark.

F. Neuwirth and D.C. McKay, 1987: "Small Scale Time and Space Variability of Solar Radiation". Final Rept. (Draft) for IEA Solar Heating and Cooling Programme Task 9, Phase I, Subtask A. Central Institute for Meteorology and Geodynamics, A-1190 Vienna, Austria.

R. Trapp and F. Kasten, 1988: "Kleinskalige Variabilität der Sonnenstrahlung". Berichte des DWD, Deutscher Wetterdienst, D-6050 Offenbach a.M., FR Germany.

9.1.2 Satellite-based methods

R.J. Bahm ed., 1981: "Satellites and forecasting of solar radiation". Proc. 1st Workshop on Terrestrial Solar Resource Forecasting and on the Use of Satellites for Terrestrial Solar Resource Assessment, Newark, DE. ASES 111pp.

C.J. Riordan and R.L. Hulstrom, 1983: "A review of potential satellite techniques for mesoscale mapping of insolation". Solar Energy Research Institute TR-215-1824, 1617 Cole Boulevard, Golden, CO 80401.

C. Raphael, 1983: "Models for estimating solar irradiance at the Earth's surface from satellite data: an initial assessment". Rept. No. 83-1, Canadian Climate Centre, Atmospheric Environment Service, Downsview, Ontario M3H 5T4.

C. Raphael and J.E. Hay, 1984: "An assessment of models which use satellite data to estimate solar irradiance at the Earth's surface". *J. Climate & Applied Meteor.* **23**, 832.

N. Beriot, 1984: "Rapport sur l'évaluation de la ressource énergétique solaire à l'aide de satellites météorologiques". WCP-80, WMO, CH-1211 Geneva 20, Switzerland.

C.J. Riordan, 1985: "Solar radiation modeling using satellite data". *INTERSOL'85*, Proc. ISES Biennial Congress, Montreal, Vol.**4**, pp. 2543-2547, Pergamon 1986.

A. W. Grüter, H. Guillard, W. Möser, J.M. Monget, W. Palz, E. Raschke, R.E. Reinhardt, P. Schwarzmann, and L. Wald, 1986: "Solar Radiation Data fron Satellite Images". Solar Energy R&D in the European Community, Series F, Vol. **4**, D. Reidel.

WCP-136, 1987: "Report of the Joint Scientific Committee on radiative flux measurements". WMO/TD-NO. 76, CH-1211 Geneva 20.

C. Delorme, 1988: "Le gisement solaire". G.D.T.A. 18, av. E. Belin, F-31055 Toulouse, France.

J. Schmetz, 1989: "Towards a surface radiation climatology. Retrieval of downward irradiances from satellites". *Atm. Res.* **23**, 287-321.

M. Noia, 1990: "Valutazione della radiazione solare al suolo a partire da immagini da satelliti geostazionari". Thesis Phys. Dept. University Genova, I-16146 Genova, Italy.

9.2 General readings

B.D. Ripley, 1981: "Spatial Statistics". Wiley & Sons.

WMO - No. 557, 1981: "Meteorological Aspects of the Utilization of Solar Radiation as an Energy Source". WMO Technical Note No. 172, CH-1211 Geneva 20.

G.L. Gaile and C.J. Willmott Eds., 1984, "Spatial Statistics and Models". D. Reidel.

R. Guzzi and C.G. Justus Eds., 1988: "Physical Climatology for Solar and Wind Energy". World Scientific.

9.3 Related references

J.A. Davies, D.C. McKay, G. Luciani, and M. Abdel-Wahab, 1988: "Validation of models for estimating solar radiation on horizontal surfaces". Final Rept. for IEA Solar Heating and Cooling Programme Task 9, Phase I, Subtask B. Canadian Climate Centre, Atmospheric Environment Service, Downsview, Ontario M3H 5T4.

J.E. Hay and D.C. McKay: 1988, "Calculation of solar irradiances for inclined surfaces: verification of models which use hourly and daily data". Final Rept. for IEA Solar Heating and Cooling Programme Task 9, Phase I, Subtask B. Canadian Climate Centre, Atmospheric Environment Service, Downsview, Ontario M3H 5T4.

R. Perez, P. Ineichen, R. Seals, J. Michalsky, and R. Stewart: 1990, "Modeling daylight availability and irradiance components from direct and global irradiance". *Solar Energy* **44**, 271-289.

R. Perez, P. Ineichen, E. Maxwell, R. Seals, and A. Zelenka, 1991: "Dynamic models for hourly global-to-direct irradiance conversion". *1991 Solar World Congress*, Proc. 1991 Biennial Congres ISES, Arden/Burley/Coleman Eds., Vol.1/Part II, pp. 951-956, Pergamon 1991.

Appendix

List of Participating Experts

Gerhard Czeplak
Deutscher Wetterdienst, Meteorologisches Observatorium Hamburg
Frahmredder 95
D - 22393 Hamburg 65
Germany

Vito D'Agostino
SASIAM, Tecnopolis CSATA Novus Ortus
P.O. Box 775
I - 70010 Valenzano / Bari
Italy

Weine Josefsson
Swedish Meteorological and Hydrological Institute
S - 60176 Norrköping
Sweden

Eugene Maxwell
National Renewable Energy Laboratory
1617 Cole Boulevard
Gloden, CO 80401-3393
U.S.A.

Richard Perez
Atmospheric Siences Research Center
State University of New York at Albany
100 Fuller Road
Albany, NY 12205
U.S.A.

Antoine Zelenka (Subtask Leader)
Swiss Meteorological Institute
Krähbülstrasse 58
CH - 8044 Zürich
Switzerland

Temporary Expert:

Corrado Ratto
Università di Genova
Dipartimento di Física
Via Dodecaneso 33
I - 16146 Genova
Italy