

Project Report No. 410

March 2007

Price: £13.00



Developing a framework to provide local agro-meteorological data in the U.K. ('INTERMET')

by

M Hims¹, N Hardwick¹, S Parker¹, M Taylor¹, W Luo¹, D Stott¹,
S Elcock¹, N Stuart², M Mineter², C Jarvis³, R Palmer⁴, T Mayr⁴,
P Bellamy⁴, J Hanam⁴, C Parker⁵, K Jaggard⁶, A Qi⁶

¹Central Science Laboratory, Sand Hutton, York YO41 1LZ

²University of Edinburgh, Kings Buildings, West Mains Road, Edinburgh EH9 3JH

³University of Leicester, University Road, Leicester LE1 7RH

⁴National Soil Research Institute, Cranfield University, Silsoe, Beds. MK45 4DT

⁵University of Loughborough, Loughborough, Leicestershire LE11 3TU

⁶Brooms Barn Research, Higham, Bury St Edmunds, Suffolk IP28 6NP

This is the final report of a thirty-six month project which started in July 2002. The project was sponsored by Defra with £105,993 through the Foresight LINK project and NERC with £187,968. Industrial funders were HGCA (£190,137, Project No. 2200), Delta-T Devices Ltd (£20,000), ESRI UK Ltd (£15,000), Farming On-Line Ltd (£5,500), HDC (£30,000), PGRO (£6,000), Sun Microsystems Ltd (£43,002), Technicrop Ltd (£4,500) and Vaisala TMI Ltd (£43,000), making a total of £657,100.

The Home-Grown Cereals Authority (HGCA) has provided funding for this project but has not conducted the research or written this report. While the authors have worked on the best information available to them, neither HGCA nor the authors shall in any event be liable for any loss, damage or injury howsoever suffered directly or indirectly in relation to the report or the research on which it is based.

Reference herein to trade names and proprietary products without stating that they are protected does not imply that they may be regarded as unprotected and thus free for general use. No endorsement of named products is intended nor is it any criticism implied of other alternative, but unnamed, products.

CONTENTS

1	ABSTRACT	1
1.2	PROJECT SUMMARY	2
2.	A BRITISH FRAMEWORK TO DELIVER AGRO-METEOROLOGICAL DATA	11
2.1	OUTLINE	11
2.2	INTRODUCTION	11
2.3	DATA REQUIREMENTS FOR AGRICULTURAL AND HORTICULTURAL MODELS	11
2.4	AVAILABILITY OF LOCALLY RELEVANT MODEL INPUT VARIABLES FOR DECISION SUPPORT MODELS	14
2.4.1	Government meteorological networks.....	14
2.4.2	Gridded weather data as model inputs	14
2.4.3	On-farm or locally-networked sensor data.....	15
2.4.4	Communication of agro-meteorological data to users of decision support systems.....	15
2.4.5	GIS-based agro-meteorological data, products and services on the Internet ...	16
2.5	A BRITISH FRAMEWORK FOR ENHANCED VALUE AGRO-METEOROLOGICAL DATA.....	18
3.	USER NEEDS	21
3.1	OUTLINE	21
3.2	BACKGROUND.....	21
3.3	APPROACH.....	21
3.4	FORECAST METEOROLOGICAL DATA.....	21
3.5	ACTUAL METEOROLOGICAL DATA.....	22
3.6	COMMERCIAL AVAILABILITY OF INTERPOLATED DATA	23
4.	INTERPOLATION OF METEOROLOGICAL VARIABLES	24
4.1	OUTLINE	24
4.2	TEMPERATURE	24
4.2.1	Approach	24
4.3	PRECIPITATION	25
4.3.1	Spline interpolation for estimation of precipitation	25
4.3.2	Data.....	26
4.3.3	Approach	27
4.3.4	Comparison of the interpolated surfaces	28
4.3.5	Conclusion	30
4.4	RELATIVE HUMIDITY	30
4.4.1	Network & data description.....	31
4.4.2	Daily mean Relative Humidity.....	32
4.4.3	Relative humidity thresholds	34
4.4.4	Interpolation of outputs from a decision support system (DSS)	35
4.5	WIND.....	36
4.5.1	Wind Speed	38
4.5.2	Wind Direction	40
4.6	SOLAR RADIATION AND SUNSHINE HOURS	42
5.	INTERPOLATION OF SOIL MOISTURE AND TEMPERATURE	47
5.1	OUTLINE	47
5.2	CHOICE OF MODELS	47
5.3	MOSES DEVELOPMENT	47
5.4	INITIAL SENSITIVITY TESTING OF THE NSRI-MODIFIED MOSES MODEL	48

5.4.1	Batcombe series	49
5.4.2	Quorndon series	51
5.5	VALIDATION OF NSRI-MODIFIED MOSES OUTPUT	53
5.6	CORRELATION OF RESULTS FROM NSRI SOIL TEMPERATURE PROBES AND LONG-TERM UKMO SOIL TEMPERATURE RECORDERS	53
5.7	INTEGRATION OF NSRI-MODIFIED MOSES INTO THE INTERNET MODEL INTEGRATION FRAMEWORK (MIF)	56
5.8	REVIEW OF PREDICTIONS OF SOIL MOISTURE AND SOIL TEMPERATURE FROM NSRI-MODIFIED MOSES AND MEASURED DATA AT CAMBORNE, WATNALL AND ROTHAMSTED	57
5.9	CONCLUSIONS	62
5.9.1	Soil moisture assessment	62
5.9.2	NSRI-modified MOSES model issues with soil moisture using site-specific weather and soil data	62
5.9.3	Soil temperature issues	63
5.9.4	NSRI-modified MOSES model issues with soil temperature using site-specific weather and soil data	63
5.10	RECOMMENDATIONS	63
5.10.1	Model improvements	63
5.10.2	Compatibility with MIF	64
5.10.3	New weather surfaces	64
5.10.4	Implementation trial area	64
5.10.5	Validation requirements	64
6.	DISTANCE BETWEEN FIELDS AND WEATHER STATIONS	66
6.1	OUTLINE	66
6.2	APPROACH	66
6.3	RAINFALL	66
6.4	TEMPERATURE	67
6.5	IDEAL LOCATIONS FOR RECORDING STATIONS	68
7.	THE USE OF INTERPOLATED VARIABLES IN DECISION SUPPORT	73
7.1	OUTLINE	73
7.2	INTERPOLATED VARIABLES FOR PREDICTING POTATO BLIGHT RISK ..	73
7.2.1	Approach	73
7.2.2	Date of first Smith Period	77
7.2.3	Seasonal total of blitecast severity units	77
7.2.4	Impact of interpolation on risk warnings	79
7.3	TEMPERATURE SUMS	80
7.3.1	Approach	80
7.4	LODGING RISK	83
8.	MODELLING AND INTERPOLATION FRAMEWORK (MIF)	86
8.1	OUTLINE	86
8.2	BACKGROUND	86
8.2.1	The MIF Framework	87
8.2.2	Context	87
8.3	SOFTWARE REQUIREMENTS	88
8.4	SOFTWARE DEVELOPMENT	89
8.4.1	System Platform and development Language	89
9.	INTERNET SERVICE ARCHITECTURE	92
9.1	OUTLINE	92
9.2	INFRASTRUCTURE DESIGN	92
9.3	DATA FLOW	93

9.3.1	Overview	93
9.3.2	Detail of Data Input Processes.....	94
9.3.3	Email to database importing script	94
9.3.4	MIF Control Script.....	94
9.3.5	Return interpolated data to the GIS Application server	95
9.3.6	Transferring MIF output data into the ArcSDE “geodatabase”	96
9.4	DATA OUTPUT PROCESSES.....	97
9.4.1	Data selection	98
9.4.2	ArcIMS map.....	98
9.4.3	ArcIMS and ArcSDE	100
9.4.4	Returning the map image to the client PC	100
9.5	WEBSITE DESIGN AND FUNCTIONALITY	101
9.5.1	Available data	101
9.5.2	Source data.....	101
9.5.3	Ease of accessibility features	103
9.5.4	Data exporting	104
9.5.5	Results data	104
9.6	FUTURE DEVELOPMENTS	106
9.6.1	Extracting time series from a point in interpolated space	106
9.6.2	End-user alerting	106
9.6.3	Displaying interpolated temp for the given area	106
9.7	CONCLUSION.....	107
10.	BUSINESS PLAN FOR INTERNET	108
10.1	OUTLINE	108
10.2	DESCRIPTION OF THE PROBLEM	108
10.3	BENEFITS	109
10.3.1	Contributors of data	109
10.3.2	Clients purchasing data	110
10.3.3	Sponsors of the infrastructure	110
10.4	CRITERIA FOR SUCCESS.....	110
10.5	RISKS AND OPPORTUNITIES.....	110
10.6	IMPLEMENTATION IN YEAR 1	112
10.7	LOCATION OF CORE NETWORK STATIONS.....	112
10.8	RESOURCES AND COSTS.....	112
10.8.1	Capital cost of meteorological equipment for core network	112
10.8.2	Operational & Staff Costs.....	113
10.8.3	Funding beyond the establishment year	114
10.9	SUPPORT FOR THE CONCEPT	116
	Public sector research	116
	Levy body.....	116
	Private sector research.....	116
	Crop Consultants.....	116
	Growers.....	116
	Marketing.....	116
11.	GLOSSARY	117
12.	PUBLICATIONS	120
	Refereed Journal Articles	120
	Conference Proceedings.....	121
	Chapters in Books	121
13	REFERENCES	123

ACKNOWLEDGMENTS.....	126
APPENDIX 1.....	127
Weather parameters for NSRI-modified MOSES.....	127
Preparation of MOSES weather data for Camborne, Rothamsted and Watnall	128
APPENDIX 2.....	130
Soil physical property input data for NSRI-modified MOSES.....	130
APPENDIX 3.....	136
Sensitivity analysis of the NSRI-modified MOSES model; the effects of changing soil properties.....	136
Output files.....	137
Results.....	138
APPENDIX 4.....	140
Soil profile descriptions at Meteorological Stations	140
APPENDIX 5.....	150
Analysis of soil temperature data measured by INTERNET instruments and UKMO equipment at Camborne, Watnall and Rothamsted	150
APPENDIX 6.....	162
A review of NSRI measured soil temperature and soil moisture data vs NSRI-modified MOSES modelled output.....	162

1 ABSTRACT

Agricultural enterprises across the UK frequently face important environmental management decisions which will have an impact of their businesses. With regard to the management of pests and diseases, growers aim to adopt a holistic approach to factors which they can control (e.g., seed sources, previous season residue disposal) with those they cannot – mainly related to the weather. As development and severity of attack by crop pests is closely related to weather conditions, management decisions can be guided or improved if timely and relevant weather information is available to growers. Existing monitoring networks across the UK were not developed to service agri-environment needs but to help make long range weather forecasts. Their data quality can be inconsistent, their locations skewed towards airfields, the coast and hill tops and they don't usually record parameters which are vital for pathogen development, such as leaf wetness duration. In order to investigate how the situation could be improved this project was commissioned with the aim to use data from those existing networks and spatial interpolation to provide more localised weather parameters. A world wide web based delivery system which would enable interested parties to access data relevant to them in near real time and at reasonable cost was also developed.

A mixture of daily and hourly weather observations of temperature (air and soil), rainfall, wind speed, wind direction, sunshine hours, incident solar radiation and relative humidity during the period 1998 to 2002 were acquired from the UK Met. Office. These data were supplemented in 2002 with observations from a network of roadside weather stations operated by Vaisala Group. The number of observing stations recording each parameter ranged from as few as 30 for solar radiation to over 3000 for daily rainfall. After discarding stations with incomplete records or those that reported retrospectively, the number of stations available to interpolate each parameter was usually less than 250. Different methods of spatial interpolation were employed to produce continuous surfaces of the parameters depending which gave the most consistent result. The resolution of the output surfaces depended on the nature of the parameter interpolated and the density of the network used but was usually a grid of either 2km or 5km squares.

Surveys of end users carried out during an early stage of the project revealed that job type and crop sector had an influence on which weather parameters were perceived as the most important. Rainfall, temperatures and wind speed were regarded generally as the most useful while leaf wetness, sunshine hours and wind direction were of lesser concern. Forecasts of weather conditions were seen as being more useful than historical data which would have more of a role within Decision Support Systems (DSSs). Daily records were preferred to hourly and the favourite delivery method was either email or some web based interface.

To produce a delivery system that appears transparent to customers the various parts of the InterMet project must be integrated using a wide variety of physical computers, software applications, linking scripts and network storage/transfer components. A web site was developed which allowed a user to select an interpolated parameter (temperature, relative humidity, forecasting model etc.) and display an interactive map which could be zoomed in or out. Daily and/or hourly data could also be displayed for any weather station appearing within the current view of the map. As the map is zoomed the amount background detail changes to allow the user to orientate themselves with features that are increasingly local.

As a final step, a business plan was developed that would enable InterMet to move from a demonstrable concept to a publicly available product albeit on a non-profit basis.

1.2 PROJECT SUMMARY

A BRITISH FRAMEWORK TO DELIVER AGRO-METEOROLOGICAL DATA

A number of agricultural decision support system (DSS) are presently being developed in Britain and elsewhere, which aim to provide more accurate predictions of risks and therefore assist farmers and growers with control actions. The intention is to provide both economic and environmental benefit, within a move towards more sustainable agriculture. However, many operational decisions in agriculture are still often based on data from a few remote point samples. Additionally, in the still rare cases where continuous surfaces of weather data are used in decision making the resolution of such surfaces remains crude. This suggests that there is considerable scope to use methods of geographical analysis to derive more comprehensive and locally reliable estimates of agro-meteorological parameters.

At present, weather related inputs for use in agricultural decision support systems may be obtained in three main forms. Firstly, point data such as that available from governmental, primarily synoptic, networks. Secondly, such data may have been pre-processed to form gridded surfaces, for example the UK MORECS data of 10-daily rainfall and evapo-transpiration observations at 40 km resolution. Thirdly, farmers, advisors and researchers may use on-farm or local co-operative sensor data to drive their agricultural models.

Even where agro-meteorological data are available from external bodies at appropriate temporal and geographical scales, a number of further issues arise relating to their timely communication, updating and maintenance.

Firstly, in Britain there is no single place where spatially referenced meteorological data are consistently collated and updated. The overhead of having to receive data from multiple sources and dealing with how to merge these data into a consistent data set is likely to create a significant barrier to the use of distributed spatial data within decision support systems. Secondly, data needs to be communicated in a timely manner so it is available for the current day's decision making. The use of telemetrics for distributing point weather data for agricultural/horticultural modelling purposes is particularly well developed in Scandinavia, illustrating the potential for official meteorological organisations. Centralised networks of commercial sensors are also common. In the majority of those cases however raw data is communicated with minimal pre-processing, by fax or e-mail. While these services are valuable, it is preferable for designers of decision support systems to hide the details surrounding the provision of such model input data from their users if the systems are to become an efficient tool within the workplace. Services that explicitly target the users of agricultural DSS, and that provide data in formats for automatic downloading to and updating of users' systems, are therefore required. Services over the World Wide Web are increasing and are a potential solution.

It is possible in principle to use advances in geographical information systems (GIS) and sophisticated methods of spatial interpolation to gain improvements in the accuracy and to some degree the resolution with which the main agro-meteorological variables can be produced in gridded format required for DSS. The degree of improvement achievable will be dependent firstly on the inherent spatial variability of the meteorological variables themselves (that will constrain the ability of even sophisticated interpolation techniques) and secondly and most importantly, upon the time taken for the network of recording stations to provide a collated, quality controlled and sufficient set of point data for gridding. It is also evident that the present UK synoptic network is not adequate for providing the required volume of agro-meteorological data, in a time frame of 1-2 days, which is necessary for immediate decision support across the entire cropped area of Britain.

Based on the preceding review a framework of methods is advanced, which will allow agro-meteorological data to be collated and delivered swiftly to agricultural models and DSS. The framework envisions the co-operative use of spatial databases; methods of spatial interpolation that can be guided using further GIS data sets; a linked suite of biological models of pest development and web mapping to serve the resultant gridded maps of risk rapidly using the internet.

The framework proposed for collating and distributing agro-climatic data is aimed at supporting the monitoring of meso- rather than micro-scale climatic processes and their effects on agriculture and horticulture.

Currently, agricultural decision support systems both in Britain and elsewhere are largely sited on on-farm personal computers or those of advisors. For the agricultural advisor running phenological models for a number of locations on behalf of growers, the supply of standard agro-meteorological variables without the difficulty of collecting, merging, formatting, and error checking data, is likely to be the most cost-effective means of obtaining the inputs required by the next generation of agricultural DSS. The framework will allow sequences of interpolated meteorological results over time to be produced at given points for transmission to users. Locally relevant gridded data from appropriately merged networks would also provide a uniform base upon which decisions may be made nationwide. By avoiding the cost overhead of complex sensors, the use of decision support systems becomes more viable to small and larger enterprises alike.

INTERPOLATION OF METEOROLOGICAL VARIABLES

Temperature

A comprehensive set of continuous topographic and land cover related variables were examined to test their suitability for guiding the interpolation of daily maximum and minimum temperatures over England and Wales, for an entire annual cycle to a resolution of 1 km. The work draws on and updates historical topoclimatic modelling through use of digital elevation data and land cover data, using the modelling capabilities of GIS. Once variables were included to guide the interpolators, differences in estimation accuracy between partial thin plate splines, ordinary kriging and inverse distance weighting results were not significant, although the performance of trend surface analysis was poorer. Best accuracies were achieved using partial thin plate splines, with jack-knife cross-validation root mean square (RMS) errors for an annual series of daily maximum temperatures of 0.8 °C and 1.14 °C for daily minimum temperatures.

Precipitation

The relative accuracy of interpolating precipitation amount using a series of different partial thin-plate splines, which have previously produced accurate estimates for monthly precipitation totals in areas of variable relief, were compared. The approach adopted is primarily interpolative, using rain-gauge data as the main data source. Two ways to improve the accuracy of daily estimates nationwide are to incorporate, within the existing network, additional rain-gauges that report rapidly in areas of interest, or to use more sophisticated methods to interpolate the rainfall field between the existing stations. When interpolating from point observations for example, some methods allow additional collateral data sets to be included, such as known variations in the surface topography and elevation.

A data set of one month's daily rainfall observations from England and Wales was used to explore the accuracy of different spline interpolations for estimating daily precipitation at a resolution of 5 km. For a relatively dense network of 820 rain-gauges, a 2-D partial thin-plate spline on (x,y) with elevation as a single linear covariate robustly estimated daily precipitation across England and Wales with an RMSE averaged over all stations and dates of ~1.9 mm. The models were considered unreliable for up to 4 days in the month. For a sparser network of 110 points, the RMSE was ~2.3 mm and models were unreliable on up to 7 days in the month.

Relative Humidity (RH)

The daily mean RH was estimated using only those stations that had at least 12 hourly readings during each day. This number was usually around 170 during 1998 and 1999 but increased to almost 250 from during 2000 onwards. There was not a large difference between any of the interpolation methods used. Five different variants of spline interpolation were tested but the tension spline always produced the lowest RMS. The advantage of the spline method is that there is no requirement to fit a model function to an experimental

variogram which is the cornerstone of all kriging techniques. Variograms for mean daily RH were specific for the day that they were fitted and predictive performance for other days was poor. This was expected because RH is dependent on other weather factors, especially temperature. As a consequence, an InterMet service will need to fit unique variograms each day to produce an RH surface.

Wind

Air movements affect the development of plant disease epidemics by transporting pathogen inoculum and vectors between locations. And can similarly affect dispersal of insect pests and weed species. These effects may operate at a number of scales; from within crop canopies up to continental levels. Unlike relative humidity and temperature, wind speed and direction are likely to vary hugely over short distances particularly close to the ground, but the monitoring network has not been designed to collect readings at such a scale.

Seven interpolation methods for estimating wind speed surfaces were compared, using the RMS error to judge their accuracy and precision as before. The minimum and maximum wind speeds predicted by the methods was also used to check how well the different methods maintained the range of the input data. The RMS errors showed some variation but all methods produced a result of less than 4 knots but the maximum and minimum estimation was more variable. The cokriging method (again using elevation as the covariate) resulted in the lowest overall error.

Unlike simple wind speed, the direction from which the wind blows has two components, i.e., orientation and strength, making it a vector rather than scalar quantity. As direction should be biased in favour of stronger winds an approach was developed to decompose the direction and simultaneous speed measurements into separate Cartesian components – an x and a y direction. Each component was interpolated separately by the usual methods and then recombined.

Solar Radiation and Sunshine Hours

Far fewer weather stations (36) monitor global solar radiation (GSR) than other weather variables, such as rainfall or air temperature, mainly due to the high cost of the required equipment. Sunshine recorders are more common than radiation monitors (180), so an attempt was made to relate hours of sunshine to GSR using multiple regression. Sunshine hours alone were only weakly related to the GSR recorded at the same site. However, the precision of the relationship was improved to an R^2 of 93% by accounting for other factors, specifically; station latitude, calculated day length and the proportion of each day that the sun shone.

INTERPOLATION OF SOIL MOISTURE AND TEMPERATURE

Soil temperature and moisture are important to crop production via their direct effects on plant growth and development, and also indirectly by their effects on pest and pathogen populations. Generally however, crop managers have at best crude estimates of soil conditions. Better knowledge of the soil environment is therefore to the aspiration of predicting crop performance more accurately and to the goal of improved crop management.

A sensitivity analysis using a modified version of the Meteorological Office Surface Exchange System (MOSES) model on 12 contrasting soils and a standard 6-month weather dataset showed that the model differentiated between all soils (with regard to moisture) and the differences were statistically significant in all cases. Generally very good correlations were obtained between the existing UKMO soil temperature probes at Camborne, Watnall and Rothamsted and the Delta-T probes installed during this project. However, lower layers at all sites indicated different relationships between the two soil temperature measurements when the soil was on a warming curve as opposed to when cooling down. Overall, predictions of soil temperature were better than those for soil moisture. The modelling of topsoil temperature variations is reasonably good and the bimodal distribution of measured temperatures is matched, albeit less strongly, in the modelled output. In subsurface layers

the modelled soil temperature values show a stronger relationship with measured data than was apparent in the top soil.

Soil temperature and soil moisture predictions from MOSES require further rigorous validation. In terms of soil moisture, further research questions have arisen regarding how MOSES can be engineered to deal with the broad group of soils suffering at varying degrees from waterlogging by groundwater. Some form of blocking mechanism will have to be incorporated into the model for susceptible soils whereby through-flow is strictly limited at certain times of the year or maybe the blocking mechanism is progressively triggered by certain levels of antecedent rainfall. This research will require further soil moisture monitoring sites across soil exhibiting a range of groundwater wetness.

THE USE OF INTERPOLATED VARIABLES IN DECISION SUPPORT

Weather data are a crucial requirement for most Decision Support Systems to aid crop management. Most of these have been developed with the expectation of using local records from existing meteorological networks, or from independent, farm specific recording stations. The impacts from using interpolated data are generally untested so DS tools, at three levels of complexity, were evaluated using interpolated surfaces in comparison to a run on data from the nearest weather station.

Predicting Risk From Potato Late Blight

The effect of using interpolated data was tested using two established forecasting schemes for the disease, the Smith Period and BLITECAST.

Based on industry reports during the period 1998 to 2002 a series of maps of the date of first occurrence of potato late blight were generated. These were subtracted from an identical map of the date of the first Smith Period using either nearest weather station or interpolated weather data in each of the five years. The interpolated data showed a tendency to give later warnings, i.e., they occurred nearer in time to the actual appearance of the disease, but there was no statistically significant change.

The BLITECAST scheme was compared using the season long (1 June to 30 September) total of severity units which are calculated each day depending on the duration and temperature of periods of high humidity. A pair of maps for each year were generated, one based on polygons and the other using spline interpolation. The area of potatoes within five severity total size groups (<25, 25 – 50, 50 – 75, 75 – 100, 100 – 125; lower numbers are lower risk) was calculated and compared for years 1998 – 2002. Interpolation caused a statistically significant shift from the first severity total category (<25 or lowest risk) into the second category (25 – 50). This meant that the interpolation resulted in more infection warnings under the BLITECAST scheme and therefore probably more recommendations to apply fungicides.

Temperature accumulation

Over the period of one year (1976) the maximum air temperature at a resolution of 1km² was interpolated by either Thiessen polygons (assigning the value of each pixel to the nearest weather station) or partial thin plate splines. Comparing accuracy by cross validating each station on each day of the year for all 365 days showed that the spline method produced root mean square errors that were three times less than the Thiessen polygons. The InterMet technology can also be used to explore the geography of daily agro-meteorological conditions for specific regions. A detailed analysis of degree-day accumulations within the Vale of York over a year and the rapid changes in topography across an extent of 50 km leads to a wide range in the degree-days accumulated in different locations. The values calculated for valley floor locations amount to roughly twice those computed on the nearby moors. For tactical crop management decisions, for example about the control of pests, farmers also need to know how the conditions favouring risk alter from day to day. For the

same area of Yorkshire, differences in the day in the year (Julian Date) at which a certain number of degree-days have been accumulated again show a marked geographical difference in when these dates occur. The warmer valley floor areas reach a predetermined threshold value after about 160 days (early June), whereas some areas up on the moors do not reach the same threshold before 190 days (early July).

Lodging risk

Previous work has shown that the risk of in wheat crops is related to four criteria based on rainfall and wind speed all recorded during June, July and August.

Using the InterMet database of daily rainfall (over 2500 locations across the UK) and wind speed observations (over 500 UK locations) continuous surfaces at 2Km resolution were created. As four years were available (1998 – 2001) the results for each criteria were averaged and normalised so that the final images had values close to zero for low lodging

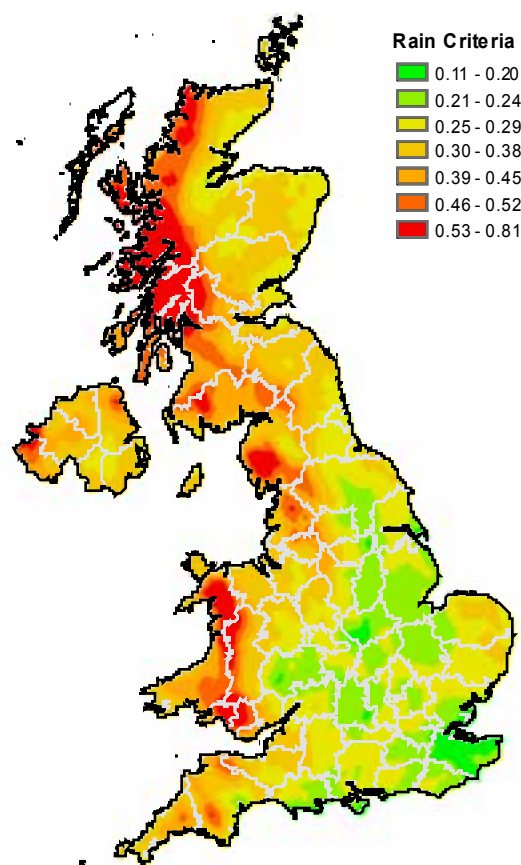


Figure 1 – Combined rain criteria for lodging risk. Scale normalised, 0 - 1 is lower to higher risk.

risk to nearer one for higher lodging risk (Figure 1).

THE DISTANCE BETWEEN FIELDS AND WEATHER STATIONS

InterMet will deliver access to meteorological records; by providing direct measurements from the recording stations and also, estimates for unmeasured locations that are interpolated by reference to the recording network. To examine the distances likely between unmeasured sites and the nearest recording station, the range of distances between randomly chosen fields and the nearest weather station were calculated. Field locations

were provided by the 419 samples collected by the annual wheat disease survey of 2000. The survey stratifies sampling according to the area of wheat within each region, so that the number of samples with each region depends on the area of wheat grown. The weather recording network comprised all of the stations supplied for the InterMet project where daily rainfall totals were measured from 1 January 1998 to 31 December 2001 at about 2300 stations across England and Wales. To coincide with the wheat survey data, only stations that had a continuous daily record of rainfall between 20 April and 20 June 2000 were used which revealed that nearly 90% of the survey locations were within 9 km of a rainfall monitoring station. Temperature was measured over the same period as rainfall at about 450 stations across England and Wales. Of these about 50 stations had infrequent records and more than 230 stations measured the temperature only once a day, only approximately 150 stations had a continuous hourly record over the four years. For the purposes of calculating distances to the nearest weather station, only those which had at least eight hourly observations on each day between 1 November and 30 November 2000 were included. Nearly half of the sites were within 15 km of the nearest recording station and most were within 30 km.

MODELLING AND INTERPOLATION FRAMEWORK (MIF)

The wider uptake of DSS depends upon the availability of a reliable supply of standard agro-meteorological variables that shields the end users from the difficulty of collecting, merging, formatting, and error checking data. The need to provide such model input data, and indeed spatially referenced model results, is the core task of InterMet.

The modelling and interpolation framework was developed for InterMet, by the University of Edinburgh Geography Department. This bespoke software capitalised on existing code for modelling and interpolation to make advances in provision of a flexible environment for developers of decision support tools. Despite the number of time-dependent point-based agricultural models and a range of interpolation software, a flexible generic integrated interpolation and modelling environment to provide both services has not previously been available. Examples exist that demonstrate how geospatial technologies are being used to construct spatially continuous weather fields but focusing in particular upon the provision of timely, locally relevant meteorological data to decision support modules, these methods are rarely coupled within a modelling environment. Prior to the current project many agricultural modelling tasks were carried out within a loosely coupled environment where the user provided the interface between multiple packages for data management, interpolation and modelling. Especially in the case where models are run over long periods at a daily time step, managing the potentially large volumes of spatial data creates a high overhead and requires large volumes of disk storage in the case where data are obtainable. Where data are not obtainable without significant delay, or disk storage space is restricted, the applications modeller must also both understand and implement the interpolation process and the final model. Neither option is tenable when considering interactive agricultural modelling by client users for day-to-day decision support.

The MIF provides sequences of interpolated meteorological data to point-based models, runs these models and produces results referenced over time and space. It has been designed and implemented around a central framework, which allows the inter-working of model and interpolation algorithms via compliance with stated interfacing protocols. The framework is structured in a modular way, allowing its components to be installed only where and when needed and which offers a route for efficient development and expansion of its capabilities. Users may interact flexibly with the software as developers, or applied users, through multiple levels of configurable parameter files. In this sense, the design of MIF aims to facilitate a shift away from using of stand-alone programs and towards the greater adoption of interacting modules so that multiple consortium members could interact in software development and its application.

INTERMET SERVICE ARCHITECTURE

The InterMet system is made up of a series of distinct modules, each designed to run on a specific hardware and software platform. The major components of the InterMet system are:

- **Microsoft SQL Server.** A full enterprise relational database engine running on commodity hardware and Windows.
- **MIF.** Bespoke interpolation software running on Sun Ultra Sparc hardware and Solaris, with the Sun Pro Fortran software.
- **ESRI ArcObjects 8.3.** GIS object library running on commodity hardware and Windows.
- **ESRI ArcSDE 8.3.** GIS database interface software running on commodity hardware and Windows
- **Macromedia Coldfusion MX and Apache.** Web application and server software running on commodity hardware and Linux.
- **ESRI ArcIMS 9.** GIS map generation software running on commodity hardware and Windows

In order to optimize performance and reliability of the InterMet system, open standards were used wherever possible to integrate these modules. When feasible, the individual components and modules were designed to be run either interactively or automatically to provide maximum flexibility.

Overview of data flow

The flow of data through the InterMet system can be summarised as two distinct stages, data input and data output.

Data Input:

- o A processing script inserts new data into the 'source' database
- o A control script extracts appropriate data from the 'source' database and executes MIF with it
- o Once MIF execution has completed, the control script sends the interpolated data to the GIS Application server
- o A COM based program, making heavy use of the ESRI ArcObjects library, extracts the interpolated data from the MIF files and inserts it into the ArcSDE geodatabase 'results'

Data Output:

- o A Coldfusion web application provides users with a route to browse the available data and to request data of specific interest
- o Coldfusion requests ArcIMS to draw a map containing the specific data requested by the user
- o ArcIMS requests the data from the ArcSDE-enabled 'results' database
- o ArcIMS receives the data and creates an image of the map
- o Coldfusion receives the map image and sends it back to the user

As implemented users are able to zoom and pan to any area of interest on a map which displays the interpolated weather variable of their choice (Figure 2). Weather stations within the current map boundaries are highlighted (Figure 3) and data from these can be displayed as a time series over a user defined period.

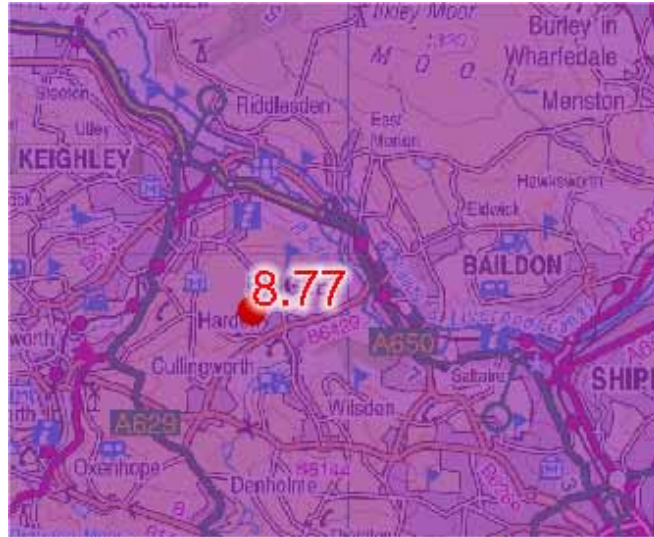


Figure 2 – Illustration of interpolated maximum temperature map including an identified pixel value

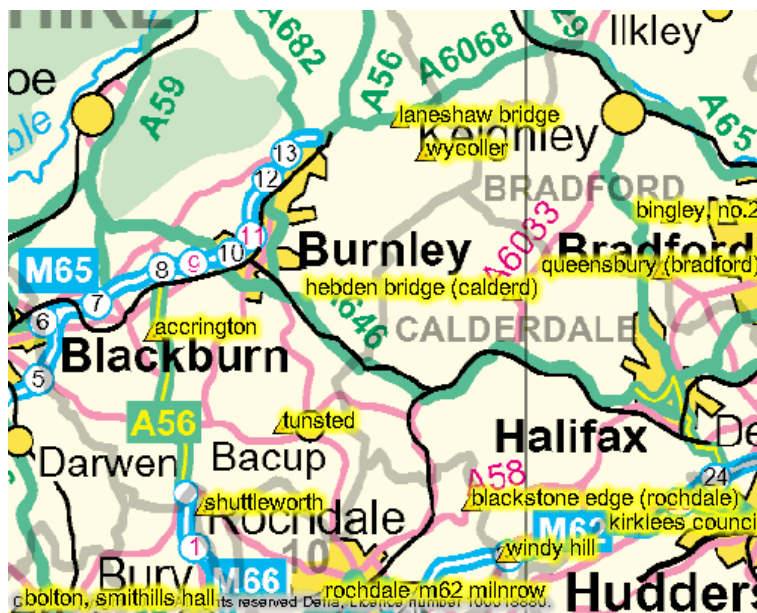


Figure 3 – Example of weather station map

BUSINESS PLAN FOR INTERMET

The InterMet project has striven to develop technologies to collate, store, enhance by interpolation and distribute weather data, predominantly for the farming industry. Whilst the project has demonstrated proof of concept for the technologies employed, it has encountered two significant barriers to implementing a commercial service using the model envisaged at the start of the work.

1. The availability and uptake of DS tools has not grown as expected, so a market place for the sale of weather data is not yet established. Moreover the industry remains broadly sceptical of the benefits of such systems.
2. The cost of weather data from existing meteorological networks is much greater than could be recouped by a resale service, which aims to add value using the methods developed for InterMet.

These barriers are likely to prevent the achievement of a sustainable service using the commercial model envisaged originally for InterMet. In addition, the work has also identified a number of other issues expected to compromise the quality of a service based on the existing networks:

- o Existing meteorological networks were not developed to service agri-environment needs but to make long range weather forecasts.
- o Many of the recording stations within existing networks do not report quickly enough to service the needs of agri-environment applications, nor do they record all of the variables that are often needed.
- o Data quality from existing networks is highly variable (at individual stations and across sites). The best data tend to come from those sites needed for national weather forecasting and include airfield, coastal and hill top locations. However, these key stations are generally located away from areas that are important for more specific agri-environment applications such as, for example, the development and implementation of decision support systems designed to facilitate more sustainable agri-environment management practices in arable and other crops.

An alternative business model has been formulated that aims to address the problems identified above, so that the technologies and approaches developed for InterMet can be exploited widely and cost effectively for farming and environmental applications. The alternative model plans to:

- o Establish a self-financing network of electronic weather observing stations concentrated across the major arable areas of England and Wales.
- o Exploit existing database and Internet technologies developed by CSL; for collecting, managing, summarising and delivering meteorological data.
- o Retrieve observations from the network daily by automatic methods, check for errors and store data centrally to allow web-based access for stakeholders, data providers and customers.
- o Expand the network by inviting third parties with existing weather stations to add their data in return for free access to processed summaries and/or to the wider network at reduced cost.
- o Provide access to the weather data, priced at a level that ensures cost recovery for maintaining the network. Any surplus will be invested in improving the network.

In order to achieve these goals the business plan aims to:

- Deliver a network of weather stations that provide hourly records of important variables within 12 hours of collection with access to raw data and summarised information as required;
- Provide an opportunity for the agricultural industry to evolve a large network of weather stations that is operated on a cooperative principle;
- Use existing infrastructures and technologies, developed by InterMet, to host the storage and dissemination of the data;
- Integrate into the network a database for all meteorological sensors and data-logging equipment purchased by sponsors of the core network.

The proposed business model is innovative in providing a cost effective, demand-led route to establish a bespoke agri-environment meteorological network for England and Wales. For UK farming to achieve the objectives of sustainability and practical environmental management, will require strong leadership and commitment to ensuring that affordable tools are in place to support research, uptake and delivery.

2. A BRITISH FRAMEWORK TO DELIVER AGRO-METEOROLOGICAL DATA

2.1 OUTLINE

This section reviews the data required to support the next generation of agricultural decision support systems, differentiating between longer-term strategic uses and shorter term uses for supporting tactical actions. In particular, the complex balance of requirements are explored for modelling at different spatial and temporal resolutions. The density and timeliness of the main types of agro-meteorological observations presently available from the national UKMO station network are examined, in derived as gridded data sets, and from local on-farm sensor networks. Consideration is also given to variations in the speed with which different types of observations become available for analysis. Based on these observations, a conceptual framework of methods is advanced that should allow agro-meteorological data to be collated and delivered swiftly to agricultural models and DSS. Work towards developing this framework is described in the succeeding sections. An extended version of this section is available in the publication: Jarvis, Stuart and Hims (2002) *Applied Geography*, **22**, 157-174.

2.2 INTRODUCTION

A number of agricultural decision support system (DSS) are presently being developed in Britain (e.g. DESSAC, now rebranded ArableDS (Brooks, 1998), MORPH (Walton, 1998)), and elsewhere, which aim to provide more accurate predictions of risks and therefore assist farmers and growers with control actions. The intention is to provide both economic and environmental benefit, within a move towards more sustainable agriculture. However, both the uptake and the success of such systems has been disappointing (Way & van Emden, 2000). Parker & Campion (1997) observe that one of the main reasons for agricultural DSS to fail is '*the use of inputs that the grower cannot easily provide*'. Many operational decisions in agriculture are still often based on data from a few remote point samples (e.g. Barrie, Jonson & Gordon, 2000; Parker & Turner, 1996). Additionally, in the still rare cases where continuous surfaces of weather data are used in decision making, for example within Danish [Pl@ntInfo](#) system (Jensen, Thyssen, Boll, Hansen, Secher & Juhl, 1997), the resolution of such surfaces remains crude. This suggests that there is considerable scope to use methods of geographical analysis to derive more comprehensive and locally reliable estimates of agro-meteorological parameters.

2.3 DATA REQUIREMENTS FOR AGRICULTURAL AND HORTICULTURAL MODELS

The range of models used within the agricultural/horticultural arena is broad. Table 2.1 lists the probable longer-term requirements of fully integrated agricultural decision support systems involving both air and edaphic environments. This suggests that agricultural DSS will require the incorporation of variables not commonly provided by major data providers or by many on-farm sensors, such as relative humidity or leaf wetness.

As Way and van Emden (2000) note, work has largely been focused upon developing strategic, rather than tactical, modelling frameworks and these have considerably different scaling requirements in space and in time, as Figure 2.4 demonstrates. On the temporal axis, most process-based models in use within the crop and pest modelling community for tactical purposes explicitly require daily data as input. In the case of strategic models, especially those used for example to model climate change scenarios, monthly data are more commonly required. It is important to identify the time step at which the model computes estimates, as the volume of spatial data that is available for modelling reduces for smaller time steps. Similarly, the structure of models, used for monthly and for daily estimation, typically vary in their sensitivity to short-term fluctuations in weather patterns.

Table 2.1. Summary of meteorological data requirements for agricultural decision support systems (Essential elements in capital letters)

Daily values	Crop growth/yield	Pests	Diseases
Max & min temperature	Yes	Yes	Yes
Rainfall	Yes	Yes	Yes
Wind speed		Yes	Yes
Wind run			Yes
Total radiation	Yes	Yes	
Relative humidity	Yes		Yes
Soil temperature	Yes	Yes	
Evapotranspiration	Yes	Yes	
Ground frost frequency	Yes		

Whilst the temporal resolution of data has been considered by the agricultural modelling community, to date it seems there has been little work on how to assess the *spatial* resolutions of data required to support agricultural modelling, particularly in relation to insect ecology and development (phenology). Knowledge of the spatial scale over which the underlying physical process can be considered to operate provides some guidance about the representative spatial unit for modelling (Grayson, Bloschl, Barling & Moore, 1993). Additionally, it has been proposed that meteorological processes can be characterised according to their extents in space and time (e.g. Wieringa, 1998). Table 2.2 matches meteorological processes occurring over a daily time step as corresponding with a spatial extent from 3km to 100km. The table suggests that consideration should be given to variables influencing meso-and topo-climates if it is necessary to interpolate agro-meteorological data for modelling to scales between 1-5km. It also suggests that it may be impractical to expect synoptic data alone to derive estimates of agrometeorological variables for grid sizes to the lower end of this range with suitable quality.

Table 2.2 places within-field or within-crop variations in climatic conditions firmly in the microscale, suggesting that models constructed using data from in-field sensors are representing conditions operating at temporal scales of the order of hours. This microscale sampling, even if it is undertaken with sensors of the same design as at the synoptic stations, is measuring a different scale of processes; it follows that models calibrated to run using in-field sensor data will consequently be incompatible with models calibrated using synoptic data.

Table 2.2. Scales in time and space, after Wieringa (1998)

Scale	Time scale	Spatial scale	
		Horizontal	Vertical
Microscale	0.1 seconds to 1 minute	<1 mm to 100mm	<1 mm to 3m
Toposcale	3 seconds to 30 minutes	10m to 3km	1m to 100m
Mesoscale	1 minute to 3 hours	300m to 30km	10m to 1km
Synoptic scale	1 hour to 1 day	3km to 1000km	100m to 10km
Macroscale	½ day to 1 week	30km to 10 ⁴ km	1km to 20km
Globalscale	3 days and longer	300km to globe	1km to 100km

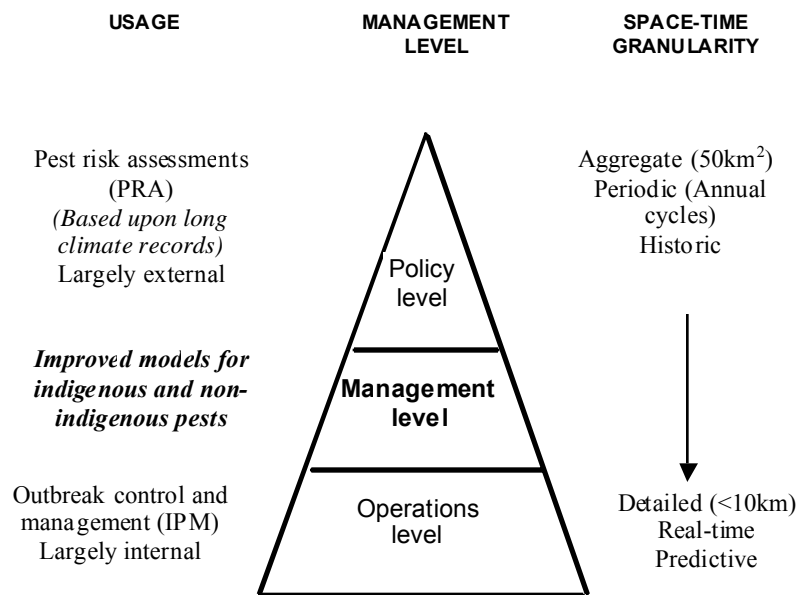


Figure 2.4. Range of requirements for decision support in crop pest management and the corresponding geographical and temporal scales at which data are required

2.4 AVAILABILITY OF LOCALLY RELEVANT MODEL INPUT VARIABLES FOR DECISION SUPPORT MODELS

At present, weather related inputs for use in agricultural decision support systems may be obtained in three main forms. Firstly, point data such as that available from governmental, primarily synoptic, networks. Secondly, such data may have been pre-processed to form gridded surfaces for example the UK MORECS data (Hough & Jones, 1997) of 10-daily rainfall and evapo-transpiration observations at 40 km resolution. Thirdly, farmers, advisors and researchers may use on-farm or local co-operative sensor data to drive their agricultural models.

2.4.1 Government meteorological networks

The primary role of governmental run meteorological networks is to provide data for synoptic weather forecasting purposes, requiring a consistent number of points from day to day. The instrumentation used to collect these records is well-maintained, and sited at 'standard exposure' allowing comparisons to be made between locations and over time. As a consequence, these are the data upon which the majority of agricultural DSS and their underlying models are based (e.g. Finch, Collier & Phelps 1996; Morgan & Solomon, 1993). However, the fact that data are collected primarily for the purpose of weather forecasting means that recording sites are predominantly placed in coastal locations and are not 'representative' of the variety of landscape nationwide. Moreover, as Table 2.3 identifies, while the overall UKMO network is large, some variables, such as rainfall are observed at a much larger number of sites than others, such as solar radiation. Other data relevant for agro-meteorology, but not necessarily for weather forecasting, such as soil temperature, humidity and moisture status, have been collected on a sparser network by a mixture of governmental and non-governmental organisations. For each agrometeorological variable, there is often a time lag of up to one month before the majority of station observations are received centrally by the UKMO.

Table 2.3. Volumes of the main agro-meteorological observations, by time of transmission

Variable	Updated ~monthly (UKMO)	Updated daily (UKMO)	Updated daily (Additional commercial sensor data)
Daily air temperatures	392	187	610
Daily soil temperatures	218	156	-
Daily rainfall	3162	410	610
Radiation	27	26	610
Wet bulb (potential evaporation).	386	353	610

Source: Analysis of data from the UKMO network in 1999, available through British Atmospheric Data Centre (BADC)

2.4.2 Gridded weather data as model inputs

A variety of alternatives to traditional point meteorological and edaphic data are becoming more commonly used especially within environmental modelling, from meso-scale climate models (e.g. Kimball, Running & Nemani, 1997) to satellite data (e.g. Legates, 2000) and surfaces interpolated from point observations (e.g. Hutchinson, 1991). Interpolation in particular provides a practical and efficient means of constructing spatial input data across the multiple variables required by decision support systems operating daily at a mesoscale within Britain.

Only rarely are the more sophisticated methods for creating gridded meteorological surfaces found within an applied geographical modelling context. Unusually, Russo, Liebhold & Kelley (1993) drive an applied insect phenology model using geographically distributed daily temperature data at a resolution of 1 km resolution. Further afield, the Danish [Pl@ntInfo](#) system uses data at a daily time step, which is then aggregated regionally (Jensen, Thyssen, Boll, Hansen, Secher & Juhl, 1997). For crop modelling, daily sets of gridded meteorological variables are used (Hoogenboom, 2000) that tend to be constructed using simple methods such as inverse distance weighting with a constant power function for all days (e.g. Supit, 1997) or time invariant kriging (e.g. Landau et al., 1998).

For the user or developer seeking 'off-the-shelf' data to complement an existing decision support system, the resolution of commercially available gridded data is more limited. The current UKMO 'MORECS' service provides agricultural users with national data for a limited number of variables at a resolution of 40 km in space, every ten days (Hough & Jones, 1997). This service is valued because of its consistency, but does not provide individual daily grids for modelling purposes, rather a temporally-amalgamated product.

2.4.3 On-farm or locally-networked sensors or data

The increasing popularity of on-farm weather recorders provides some indication of the value attributed by some farmers and advisors to the latest and most localised data with which to run models, such as those for pest phenology or crop disease. These models, by offering more localised estimates of when crops in certain areas may be at risk of pest attack or disease, allow growers to focus their control actions more tightly to certain locations and at critical periods of time. This information can be used particularly by larger businesses to save money and avoid unnecessary and wasteful applications on dates when the risk is estimated to be low. It is likely that data from these on-farm sensors will produce more accurate estimates of the localised, microscale meteorological conditions (at resolutions below 1 km) than the gridded estimates (essentially of the mesoscale conditions) produced by interpolation from the synoptic network.

Although investing in data collection from a local sensor may be worthwhile for some agribusinesses, Phelps, Collier & Finch (1993) suggest in a British context that, to be of practical use to the majority of farmers and growers, phenological models must be capable of using standard UKMO meteorological data. This is because models run using data from local on-farm sensors with different exposures may give unsatisfactory results, or at least require the model to be re-calibrated for each specific location. Many users of decision support systems will not have the resources to purchase and maintain an 'on-farm' automatic weather station, and those that do may not wish to operate independent software to compute estimates of derived variables such as evapo-transpiration or solar radiation required by many models. In summary, whilst certain business users may find particular benefits from maintaining on-farm sensors, many growers with dispersed or small acreages may prefer to receive risk assessments based on a regional network of synoptic observations maintained by recognised providers of agro-meteorological data.

2.4.4 Communication of agro-meteorological data to users of decision support systems

Even where agro-meteorological data are available from external bodies at appropriate temporal and geographical scales, a number of further issues arise relating to their timely communication, updating and maintenance.

Firstly, in Britain there is no single place where spatially referenced meteorological and edaphic data are consistently collated and updated. The overhead of having to receive data from multiple sources and dealing with how to merge these data into a consistent agro-

meteorological data set is likely to create a significant barrier to the use of distributed spatial data within decision support systems. For many of the farmers and growers who may be reluctant to change from their traditional decision-making processes, the agrometeorological data should be supplied already collated and ready for loading into the DSS.

Secondly, data needs to be communicated in a timely manner so it is available for the current day's decision making. The use of telemetrics for distributing point weather data for agricultural/horticultural modelling purposes is particularly well developed in Scandinavia (e.g. Magnus Lingaarden & Munthe, 1993) and Germany (e.g. Kleinhenz, Jörg, Gutsche, Kluge & Rossberg, 1996), illustrating the potential for official meteorological organisations. Centralised networks of commercial sensors are also common (e.g. Denzer, 1996, for Austria). In the majority of those cases however raw data is communicated with minimal pre-processing, by fax or e-mail.

While fax or e-mail data services are valuable, it is preferable for designers of decision support systems to hide the details surrounding the provision of such model input data from their users if the systems are to become an efficient tool within the workplace. Services that explicitly target the users of agricultural DSS, and that provide data in formats for automatic downloading to and updating of users' systems, are therefore required. Services over the World Wide Web are increasing (Weiss, van Crowder & Bernadi, 2000), and are a potential solution.

2.4.5 GIS-based agro-meteorological data, products and services on the Internet

Review of weather data for operational agricultural decision support systems has identified the need for a new data framework addressing the two interconnected issues of:

- Improved availability and accuracy of locally relevant agro-meteorological variables for agricultural/horticultural models;
- The timely communication of weather data, that are immediately compatible with a range of current decision support systems.

Improving the availability and accuracy of locally relevant meteorological input variables

Because accuracy may be improved either by using more sophisticated analysis methods, or by increasing the amount of underlying data, a dual approach, to achieving improvement is proposed that:

- Based upon the present UKMO network, uses knowledge of topoclimatology and more sophisticated interpolation methods developed within geographical information science to construct more accurate and locally relevant agro-meteorological data to finer resolutions;
- Augments standard governmental meteorological recording networks with additional sensor networks, to increase the currently insufficient volume of certain agro-meteorological observations available on a daily basis for nationwide decision support purposes.

Deriving local agro-meteorological data using sophisticated interpolators

Many techniques well known within meteorological and climatological circles are available but they remain under used within an integrated multi-disciplinary environment. While much recent progress has been made with downscaling climate process models and satellite and radar imaging of meteorological variables. However, practical means to create spatial coverage of current and historical weather data could perhaps be achieved more efficiently using interpolation techniques. This approach is computationally less intensive than using a complete atmospheric/land process model, and uses standard UKMO data with which many of the models in present decision support systems have been field-tested (Finch, Collier & Phelps, 1996; Morgan & Solomon, 1993). Empirical knowledge of how agro-meteorological

variables such as temperature, rainfall etc. are influenced by topographic and other variables, can be used to derive data sets of these topographic, edaphic and land cover influences (e.g. Jarvis & Stuart, 2001a). These data sets are then used with sophisticated methods of spatial interpolation, to improve estimates of each agro-meteorological variable (e.g. Daly, Neilson & Phillips, 1994; Hutchinson, 1991; Jarvis & Stuart, 2001b; Rigol, Jarvis & Stuart, 2001).

It will be important to measure the effects of using such interpolated input data on the accuracies of models proposed for agricultural DSS.

The need to augment the 'real-time' network for agro-meteorological data

As examined in Table 2.3, the volume of data available through synoptic networks in 'real time', in Britain as elsewhere, is limited relative to that of the overall governmental network. Efforts to date suggest that the scope for interpolating variables such as daily rainfall and relative humidity to resolutions in the mesoscale range from 300m - 30km with acceptable accuracies may be severely constrained by the volume and distribution of daily data that are available.

The adequacy of sample data to provide gridded estimates at different resolutions may be explored using a number of geostatistical measures. One measure of the degree to which a splining interpolator was able to produce a satisfactory estimate from an underlying data sample, is given by the trace statistic (Hutchinson & Gessler, 1994). By examining the value of the trace statistic for each surface, users may determine the number of days that the spline estimate is of poor quality and the resulting surface is likely to contain large errors. This is particularly important as a diagnostic tool, since a method of checking error and rejecting surfaces that are too error-prone is needed to control the quality and reliability of results produced. An analogous, but arguably more strict, diagnostic for interpolation by kriging would be the inability to construct an acceptable variogram on a particular day.

The diagnostic trace statistic, was used to measure the number of days during the year of 1976 when the resulting surface interpolated by splining each agro-meteorological variable could not be considered statistically valid (i.e. was too erroneous), when using data from the UKMO synoptic network alone.

Table 2.4 details the percentage of days of invalid estimations, for each of the major agro-meteorological variables. High percentage errors, particularly for rainfall and global solar radiation and some soil temperatures underline the insufficiency of data available from the UKMO real-time network for estimating most of the primary agro-meteorological variables.

Expansion of the data recording network could be achieved firstly, by increasing the number of variables recorded at the existing stations since, as Table 2.3 identified, many agro-meteorological variables are not recorded at all stations. Secondly however, it may be more economic to further expand the base of observations through the integration of data from additional networks. For example, commercial networks of meteorological stations are now used in assessing nationwide road conditions, whilst some local farming co-operatives operate their own automatic weather stations to supplement the UKMO synoptic data. In this way, value may be added both to on-farm sensor data and to national governmental and private networks. Principally, the advantages of this strategy would be to improve the timely delivery of detailed agro-meteorological requirements through the increase in additional real-time sensor data and to provide, in due course, nested models to account for within field or crop variations in climate. Vulnerability from the possible inaccuracy of on-farm sensors instead of quality controlled UKMO station data could be minimised through a merging and quality assurance process where unusual patterns or outliers in data may be identified and

corrected to ensure data consistency (e.g. Meek & Hatfield, 1994). The time of day at which recordings are made would also need to be considered in such a data merging and harmonisation operation (e.g. Karl, Williams, Young & Wendland, 1986). For network merging, close attention will be needed within the modelling procedures to account for the differing exposure between sensor types (e.g. Jacobs & Raatz, 1996) using land cover/use and topographic data. In practical terms, support is growing for a mixed network of data sensors (e.g. Brown, 1996) as a means to infill presently scarce data.

Table 2.4. Number of days in 1999 on which the computed partial thin-plate spline surface is statistically invalid, by network

Variable	Synoptic data network: percentage of statistically invalid surfaces (1976)	Synoptic and climatic data network: percentage of statistically invalid surfaces (1976)
Minimum daily air temperature	4%	4%
Maximum daily air temperature	3%	3%
Daily rainfall	23%	-
Soil temperature (100cm)	30%	22%
Soil temperature (30cm)	39%	38%
Soil temperature (10cm)	9%	6%
Global solar radiation	29%	69%

Timely communication of weather data with decision support systems

Many factors point to the use of a central service provider, processing and disseminating requests for model input data from users. These include:

- Shared ownership of output while protecting the rights of input data owners;
- Avoiding the need for clients to have considerable local processing capabilities;
- Harmonisation between networks to provide a combined data set that is consistent and of high quality;
- Using a central service to overcome present incompatibilities between the multiple computer platforms of potential user communities, and their varying processing capabilities;
- Ease of updating and maintaining the services, software and data.

Strand (2000) emphasises that *'the local culture ... must be taken into account to assure that information on techniques will reach the right participants.'* Within an European context, familiarity is rapidly gaining with Internet-based communications, particularly e-mail and the World Wide Web (WWW). Developments in spatial data handling using the WWW suggest the development of an Internet service to provide a modern solution for the query and distribution of agro-meteorological data.

2.5 A BRITISH FRAMEWORK FOR ENHANCED VALUE AGRO-METEOROLOGICAL DATA

It is possible in principle to use advances in GIS and sophisticated methods of spatial interpolation to gain improvements in the accuracy and to some degree the resolution with which the main agro-meteorological variables can be produced in gridded format required for DSS. The degree of improvement achievable will be dependent firstly on the inherent spatial variability of the meteorological variables themselves (that will constrain the ability of even sophisticated interpolation techniques) and secondly and most importantly, upon the time taken for the network of recording stations to provide a collated, quality controlled and sufficient set of point data for gridding. It is also evident that the present UK synoptic network is not adequate for providing the required volume of agro-meteorological data, in a time

frame of 1-2 days, which is necessary for immediate decision support across the entire cropped area of Britain.

Based on the preceding review a framework of methods is advanced, which will allow agro-meteorological data to be collated and delivered swiftly to agricultural models and DSS. The framework envisions the co-operative use of spatial databases; methods of spatial interpolation that can be guided using further GIS data sets; a linked suite of biological models of pest development and web mapping to serve the resultant gridded maps of risk rapidly using the internet (Figure 2.5). It is intended that this framework could also incorporate the best gridded national forecast estimates available on a day-by-day basis.

The framework proposed for collating and distributing agro-climatic data is aimed at supporting the monitoring of meso- rather than micro-scale climatic processes and their effects on agriculture and horticulture. Whilst in-crop sensors may provide useful operational information to specific farms, the concepts and framework described by the work of this project aim to exploit that sensor technology to densify the daily national network, so that spatially continuous daily input data can be made available for DSS running models designed with mesoscale input data. Potentially erroneous results are likely to occur should micro-scale data be used to drive models developed using mesoscale data, and vice versa, and given the recent proliferation of meteorological data of both types it is vital that both model developers and data providers are scale-explicit.

Currently, agricultural decision support systems both in Britain (e.g. Brooks, 1998) and elsewhere (e.g. Régnière, Cook & Bergeron, 1996) are largely sited on on-farm personal computers or those of advisors. For the agricultural advisor running phenological models for a number of locations on behalf of growers, the supply of standard agro-meteorological variables without the difficulty of collecting, merging, formatting, and error checking data, is likely to be the most cost-effective means of obtaining the inputs required by the next generation of agricultural DSS. The framework will allow sequences of interpolated meteorological results over time to be produced at given points for transmission to users (Figure 2.5, light shading). Locally relevant gridded data from appropriately merged networks would also provide a uniform base upon which decisions may be made nationwide. By avoiding the cost overhead of complex sensors, the use of decision support systems becomes more viable to small and larger enterprises alike. Looking forward, it is likely that there will be greater use of the WWW as a base for decision support systems (e.g. MORPH, Walton, 1998) using interpolated data (Figure 2.5, darker shading). This would allow fuller use of the integrated modelling software to demonstrate gridded model results, together with estimates of error for individual surfaces or locations.

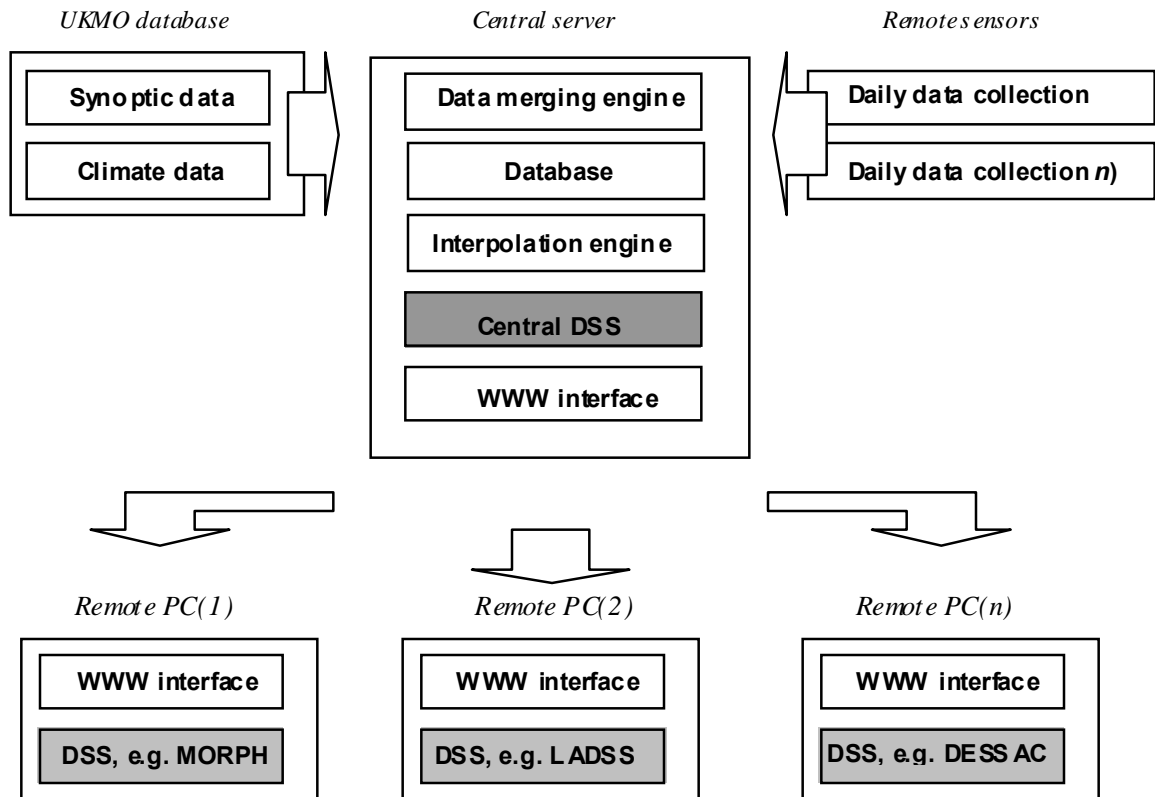


Figure 2.5. Proposed framework for data collation, processing and dissemination from WWW interface to DSS on remote PCs (light shading) and from a central DSS via a WWW interface to individual PCs (dark shading)

3. USER NEEDS

3.1 OUTLINE

This section outlines work done to examine the requirements of potential users, in agriculture and horticulture, for meteorological data. Full records of the User Surveys can be found at the InterMet site (www.intermet.co.uk).

3.2 BACKGROUND

The aspiration of developing computer based decision support and knowledge management systems for agriculture has been held virtually since the advent of computing technology. Until recently, it has been rare for developers to examine how such tools are best designed, so that the end users will use and get the full benefits from them. However, over the past decade in the UK, the development of DS has been influenced strongly by more thorough examination of user requirements.

A problem for the present study is that the technology it develops is intended to service DS systems, some of which are presently at the concept stage. Moreover there is considerable debate about the delivery mechanisms and complexity of modelling, which will be acceptable to decision makers in agriculture and horticulture.

The user need for supplying data seamlessly into systems is a relatively straightforward prescription of the data formats, accuracy and frequency of updating that are required. However, this does not recognize the more *ad hoc* data supply that might be used to inform decision makers. The surveys described in this section are focussed primarily on defining the likely requirements of those users.

3.3 APPROACH

The study used two questionnaires: the first employed in an initial small scale survey, and a second published and distributed via the Internet. The results from both were broadly in agreement. This section details the key findings, full details of the surveys can be found at www.intermet.co.uk.

The surveys obtained responses from across the main crop sectors. Most of the respondents identified themselves as farmers or farm managers. A wide range of cropping area was reported by the respondents with around half of the responses coming from people responsible for 10ha or less and 20% for areas over 500 ha. Referenced against the weather stations reported to be local to respondents this land appears to be widely spread over England and Wales, but with less representation from Scotland and N. Ireland. However, nearly a third of respondents did not know where their nearest station was and some of the missing regions may be within this set. There were a wide range of distances reported for nearest weather stations with a large proportion at the limits of the distance range *i.e.* 11-20 miles away and over 51 miles away (both 21% of respondents).

3.4 FORECAST METEOROLOGICAL DATA

The most valuable forecast weather variables appear to be rainfall, temperature (minimum and maximum), wind speed and wind direction; with rainfall identified as the most useful by far. On the basis of information about job type and crop sector, rainfall and minimum and maximum temperature are useful to all users. Whereas wind speed is more useful to farmers/growers (90.5%) than to consultants (60%). This is logical, because information on wind speed is important to immediate tactical decisions *e.g.* deciding whether conditions permit spraying.

Rainfall is of high value to managers of all crop types. Wind speed appears to be more useful to those in the wheat, barley, OSR, sugar beet, potatoes and field vegetables crop groups. Leaf wetness is of higher importance to field vegetable growers (64%) and top fruit growers (53.3) than indicated by the overall figures. This may be due to the historical availability of models employing leaf wetness in these crops and the associated publicity and education. Relative humidity is also identified to be important to potato growers (68.4%), presumably because of its importance in determining the risk of potato blight. Forecast average temperature was rated as useful or very useful by 60% and sunshine hours by 76% of bedding group respondents in contrast to other sectors and to the initial survey where interest was low. Over 60% of people in the wheat, barley, OSR, potato, field vegetable sectors were thought forecast soil moisture and temperature was useful or very useful and over 60% of soft fruit and sugar beet respondents were thought soil moisture was useful or very useful.

The most important forecast variable is rainfall and that this should be given priority for provision to the industry. The other key variables, minimum and maximum temperature and wind speed are more important for some crop types than others. Although none of these crop specific figures are enormously different from the general trend, they do indicate clear crop specificity in the demand for some weather data. These preferences will need to be considered when deciding the priorities for data provision (it is unlikely that all variables could be provided at the launch of an InterMet service). It is possible that increased availability of DSS will raise awareness about the importance of some weather variables, for example in predicting pest and disease problems, and that this will alter the perceived importance of some measurements.

3.5 ACTUAL METEOROLOGICAL DATA

Generally, actual meteorological data were perceived to be less useful than forecast data although again rainfall was the most useful across crop types. The least useful historical weather variables were leaf wetness, solar radiation, relative humidity, wind direction and sunshine hours.

Forecast data are currently perceived to be more important than historical records, presumably because they inform day-to-day planning activities on farms. The value of historical data is largely confined to its use within DSSs, which are not used by industry routinely at present.

Most respondents required daily data and were less interested in hourly records. Rainfall and wind speed were the only variables required on an hourly basis. Therefore, ignoring any specific DSS requirements, rainfall and wind speed are the only variables that might need to be offered for delivery on an hourly scale.

The majority of respondents were content with a grid size of 3-5 km or greater for both forecast and historical data. Around 20% of respondents would like field level data on forecast leaf wetness, soil moisture and wind speed, with 15% also wanting this level of detail for rainfall and soil moisture.

The most popular choices for delivery of both forecast and historical weather data were email and the Internet and this should therefore be the main form of access for individual users, DSS requirements may be different.

No clear preference was found for the display of most weather variables, so the simplest to produce can therefore be selected. For wind data, however, the design should adopt the 'value at a point' style. When data is presented over time the clear choice is to provide this in a graphical format.

3.6 COMMERCIAL AVAILABILITY OF INTERPOLATED DATA

About 40% of respondents indicated that they would be likely to buy interpolated data if it became available. This is a very positive response, because it is unusual to get any indication of enthusiasm when this type of question is posed. Moreover, the very low numbers of respondents who indicated they would be unlikely purchase data is perhaps a more useful indicator of interest. It appears that the vegetable and top fruit sectors are most interested in purchasing interpolated data.

4. INTERPOLATION OF METEOROLOGICAL VARIABLES

4.1 OUTLINE

This section details research to examine the feasibility and develop methodologies for interpolating weather variables likely to be important for decision making in agriculture and horticulture. The examination includes variables that are not currently utilized for crop management, but that are likely to become more important if practicable DS tools become available and used more widely.

4.2 TEMPERATURE

Temperature can have effects on almost every biological component of the crop system. It is a key determinant of the growth and development of most plants, pathogens and pests. As a consequence, temperature is measured, almost universally, in experiments aiming to understand crop health and production and is often correlated with biological responses. Decision support systems designed to aid crop production are therefore often very dependent upon temperature measurements.

4.2.1 Approach

A comprehensive set of continuous topographic and land cover related variables were examined to test their suitability for guiding the interpolation of daily maximum and minimum temperatures over England and Wales, for an entire annual cycle to a resolution of 1 km. The work draws on and updates historical topoclimatic modelling through use of digital elevation data and land cover data, using the modelling capabilities of geographical information systems (GIS).

The influential guiding variables under a variety of dominant weather patterns were identified and used to assist with the interpolation of an annual sequence of daily maxima and minima for 1976. Northing, elevation, coastal and urban effects were found to be particularly significant variables in explaining the variation in UK daily minimum temperature. Urban factors have not previously been thoroughly investigated, despite the high density of population in England and Wales. Analysis of the residuals from data withheld from the partial thin plate spline interpolation suggests that the incorporation of coastal shape and situation, land cover and soils data might further improve the modelling of local-scale influences on maximum and minimum temperature. They also suggest that the results achieved (r.m.s. errors of 0.8 °C for maxima and 1.14 °C for minima) may be close to the limits of accuracies achievable at 1 km resolution given the density of temperature observation data and standard exposure of the observing network used.

In a comparative experiment, the sequence of daily maximum and minimum temperatures for the year 1976 were interpolated over England and Wales to a resolution of 1 km using partial thin plate splines, ordinary kriging, trend surface and an automatic inverse distance weighted method of interpolation. A 'level playing field' for comparing the estimation accuracies was established through the incorporation of a consistent set of guiding variables in all interpolators.

Once variables were included to guide the interpolators, differences in estimation accuracy between partial thin plate splines, ordinary kriging and inverse distance weighting results were not significant although the performance of trend surface analysis was poorer. Best accuracies were achieved using partial thin plate splines, with jack-knife cross-validation root mean square (rms) errors for an annual series of daily maximum temperatures of 0.8 °C and 1.14 °C for daily minimum temperatures.

The results from this study suggest that sole reliance on the selection of guiding variables can be a less efficient means of achieving the required accuracies than the placing of greater reliance on empirical techniques of interpolation that can account for known autocorrelation in the temperature data. The use of guiding variables narrows the gap between performance of the different interpolation methods. Generally, however, more sophisticated interpolators such as kriging and splining require fewer guiding covariates in order to achieve similar estimation accuracies. Day-to-day variability in the interpolation accuracies confirms the need for increased adaptability in the manner in which the guiding variables are incorporated in the interpolation process.

4.3 PRECIPITATION

Attempts to estimate daily precipitation across Britain at fine spatial scales, from observations that are available within 24 h of collection, reveal certain regions that are poorly covered in near real time by the present network of stations operated by the UK Meteorological Office (UKMO). Two ways to improve the accuracy of daily estimates nationwide are to incorporate, within the existing network, additional rain-gauges that report rapidly in areas of interest, or to use more sophisticated methods to interpolate the rainfall field between the existing stations. When interpolating from point observations for example, some methods allow additional collateral data sets to be included, such as known variations in the surface topography and elevation. Including these geographical influences may lead to a more realistic estimate of variations in the rainfall field between observing stations. Densification of the existing network of observing stations is presently occurring, with data providers beginning to pool data to improve daily estimation. For example, the Environment Agency of England and Wales now provides data to the UKMO from 300 rain-gauges that report daily to a national flood warning system.

The relative accuracy of interpolating precipitation amount using a series of different partial thin-plate splines, which have previously produced accurate estimates for monthly precipitation totals in areas of variable relief, were compared. The approach adopted is primarily interpolative, using rain-gauge data as the main data source. The following section briefly reviews previous work using splines to estimate precipitation and introduces some of the practical advantages that partial thin-plate splines provide for estimating surfaces from combinations of point rain-gauges and other gridded collateral data.

4.3.1 Spline interpolation for estimation of precipitation

There are many variant forms of splining, but two particular sub-types - the regularized spline with tension and the partial thin plate spline have been used most often for interpolation of precipitation. In this study, the analysis was limited to partial thin-plate spline methods of interpolation, so that differences in estimation accuracy due to differences in the number of data points and the handling of collateral data sets could be compared directly.

As summarized by Hutchinson (2000) the observational model for a partial thin-plate spline with two independent variables (x,y) is that n data values z_i are given by:

$$z_i = f(x_i, y_i) + \sum_{j=1}^p \beta_j \Psi_{ij} + \varepsilon_i \quad (i = 1, \dots, n)$$

where f is an unknown smooth bivariate function, the β_j are a set of unknown parameters, x_i, y_i are the independent spline variables, each Ψ_j is a linear covariate and the ε_i are independent random errors with zero mean and variance $d_i \sigma^2$. The d_i are known weights while σ^2 may or may not be known. The spline function f and the parameters β_j for the

parametric linear sub-models are estimated by minimizing:

$$\sum_{i=1}^n [(z_i - f(x_i, y_i) - \sum_{j=1}^p \beta_j \Psi_{ij}) / d_i]^2 + \lambda J_m(f)$$

where $J_m(f)$ is a measure of the roughness of f defined in terms of m -th order derivatives of f . And λ is a positive number called the smoothing parameter, which determines a trade off between data fidelity and surface roughness. The value of λ is computed automatically by the ANUSPLIN software by minimizing the generalized cross validation (GCV). While the order of the derivative m may theoretically be found by optimization, for operational simplicity the derivative is set beforehand in ANUSPLIN. After finding that higher order derivatives offered no improvement to predictive accuracy, we set $m = 2$ for all the reported spline experiments.

ANUSPLIN 4.1 was used to create the spline equations (Hutchinson, 2000). This software allows collateral data sets such as topographic data sets or radar grids to be included as a series of linear covariates Ψ_j , or as a third independent variable forming a trivariate spline function $f(x_i, y_i, z_i)$ that can be solved similarly to the bivariate form described above.

In a direct comparison of various methods for interpolating precipitation conducted on a subset of 100 rain-gauge observations taken in Switzerland on 8 May 1986, Hutchinson (1998) found that a bivariate partial thin plate spline on (x, y) was able to predict precipitation at 367 withheld gauge locations with lower RMSE (5.6 mm) than ordinary (5.97 mm) or indicator kriging (6 mm). This finding was confirmed by Hofierka *et al.* (2002) whose 2-D regularized spline on the same data set gave an RMSE of 5.89 mm.

Splines do not require a preliminary determination of the range or distance over which point observations of precipitation may be spatially autocorrelated, as this covariance structure is computed as part of deriving the spline equation. This feature makes splines more robust than kriging for low data volumes. Nevertheless, especially for small data volumes there is a risk that splines may “overfit” the data, giving an exact match at the data points but a poor surface fit elsewhere. For this reason, a considerable number of widely distributed rain-gauges were used to form an independent data set to rigorously test the splined estimates at many apparently “unsampled” locations. Diagnostic functions such as the signal of the spline were used to identify days when the surface may be unreliable (Hutchinson, 1998).

4.3.2 Data

For reasons described earlier, the primary input data for this study is the location and daily values of rainfall observed at rain-gauges from the UK Meteorological Office (UKMO) station networks. The data consisted of 30 values at each station observed for each day during March 2002. The relative accuracy and the robustness of various spline methods were evaluated for two volumes of observing station data; firstly when 820 rain-gauges are available and secondly when as few as 110 gauges provide data. The small data set was chosen to represent the most data-limited case of making an estimate from only a subset of the UKMO rain-gauge network that reports daily totals within 24 h. The larger data set represents the most data-rich scenario foreseeable in the next 5–10 years, based on an assumption that the UKMO network may be supplemented with meteorological data through partnerships with other public and private data providers. For the purposes of these experiments, the data sets were created as subsets of the UKMO climate station network, pre-processed to remove any stations that had null values for rainfall on any day in March 2002. This avoids the problem of applying infilling methods before interpolation can be carried out. An independent data set of a further 288 points, drawn separately from the climate station network, was used to validate all the splined surfaces constructed from both the large and small set of input data.

The splining method allows flexible selection of which other data sets are used in the estimation and specifically whether elevation is treated as a third independent variable in a tri-variate spline equation or as a linear covariate submodel in a bivariate spline. In addition to elevation, the improvement on estimation accuracy from incorporating other linear covariates was tested. Three covariates used in this study were: (a) elevation; (b) degree of coastal influence; and (c) variability of local relief. These were derived by processing GIS data layers at a resolution of 5 km. Elevation is the most common type of collateral data used in estimating precipitation, because of known orographic influences in many locations (Chua & Bras 1982). The ratio of land to coast computed using a 5×5 window around the 5-km grid square in question seeks to represent the distance from the coast, which can influence rainfall both during cyclonic conditions (westerly frontal rain) and anticyclonic conditions (easterly fog). The third covariate describes the relative variability in relief. The standard deviation of elevation over a 10×10 window identifies areas of rapidly and slowly changing relief, which may be related to areas of relative orographic uplift and rain shadow.

Additionally, $4 \times 24 \times 31$ files of 15-min rainfall accumulation data were available from the NIMROD data archive for March 2002. These files had been checked by the UKMO to remove spurious echoes, corrected for “bright band” errors and converted to rainfall amounts using a calibration factor derived from gauges near to each radar (Golding, 1998; Harrison *et al.*, 2000). Following the convention that a rain-gauge total for “day 1” covers the period from 09:00 h on the first of the month to 08:59 h on the second of the month, 24 h accumulations of radar rainfall were formed from 09:00h on the first to 08:00 h on the second etc., producing 30 complete daily grids of radar data for 1–30 March. This was done to allow consistent comparison of the data types.

Whilst radar rainfall may seem an ideal source of gridded daily precipitation data at 5 km resolution, present problems with calibrating radar data and the fact that many environmental models are originally “at-a-point” models calibrated to local meteorological station data, means that it is conceptually and operationally often most practical to use the rain-gauge network as the primary source of input data and to use interpolation to create gridded surfaces that closely honour the values observed at gauge locations. Rain radar provides valuable information about the extent and variance of the rainfall pattern over spatial scales from 2 to 5 km. For this reason the benefit from including gridded radar data as an additional surface covariate was tested, particularly to guide the values estimated by the spline function in areas remote from any points of observation.

4.3.3 Approach

Various spline approaches, summarized in Table 4.5, were compared for their accuracy in interpolating precipitation surfaces. Experiments 1–3 and 5 differ only in the choice of which of the three collateral variables are included in each spline and whether elevation is included as a third independent variable that the spline may treat in a non-linear way, or as a covariate that is assumed to be related linearly with precipitation. Experiments 4 and 6 first apply a square-root transform to the rainfall values before interpolating them, as some workers have suggested that this corrects for the usual tendency of rainfall data distributions to be skewed towards smaller amounts (Hutchinson, 1998). In experiments 7 and 8, a trivariate and a bivariate spline are each supplemented with a grid of radar data for each day, to assess if this can improve the accuracy of the estimation more or less than by using the collateral topographic data.

Table 4.5. Spline models used in the experiments.

Experiment	Independent spline variables	linear covariate(s) in sub models	Transform on rainfall data?
1	$f(x_i, y_i, z_i)$	Pcoast25; stdev 50	no
2	$f(x_i, y_i)$	Elev; Pcoast25; stdev 50	no
3	$f(x_i, y_i)$	Elev;	no
4	$f(x_i, y_i)$	Elev;	Yes, $\sqrt{\quad}$
5	$f(x_i, y_i, z_i)$	None	no
6	$f(x_i, y_i, z_i)$	None	Yes, $\sqrt{\quad}$
7	$f(x_i, y_i, z_i)$	Daily radar grid	no
8	$f(x_i, y_i)$	Daily radar grid	no

Spline models were developed for each experiment for both the 820 and 110 point data sets. Accuracy for the splines was evaluated against the independent validation data set described earlier. The maximum and minimum values of precipitation found in any of the 30 daily precipitation surfaces were reported, together with the root mean square error (RMSE) in mm, the mean error (bias) in mm and the mean absolute error (MAE) in mm. These were computed each day at the validation points (which were identical for both data sets) and then averaged over 30 days, using the standard equations below.

$$RMSE = \sqrt{1/n \sum_{i=1}^n (P_i - O_i)^2}$$

$$ME(bias) = 1/n \sum (P_i - O_i)$$

$$MAE = 1/n \sum |P_i - O_i|$$

Where P_i is the precipitation estimated by the spline function and O_i the value observed at the i -th rain-gauge withheld in the constant validation data set. ($n = 288$).

In addition, diagnostic information about the signal of the spline, derived from the trace of the influence matrix in a manner described by Hutchinson (1998) was used to identify the number of the 30 daily surfaces where the surface fit was judged to be extremely unsatisfactory.

4.3.4 Comparison of the interpolated surfaces

Table 4.6 summarises the results from the experiments. For the data set of 820 points, models 2 to 6 produced results that were broadly similar and satisfactory with an estimation accuracy averaged over the month (RMSE) of around 1.9 mm. Although there is relatively little difference between the estimations, experiments 2 and 3 may be judged the best in terms of lowest overall RMSE and biases closest to zero. Model 2 treated elevation as a linear covariate along with two further collateral topographic variables. Model 3, which used elevation as the only linear covariate, gave almost identical results and suggests that the additional topographic indices provided little gain. Including elevation as a third independent (non linear) variable without other covariates (model 5) resulted in similar if slightly poorer accuracy.

Table 4.6. Results of spline interpolations.

Splines on 820 data points, validated at 288 points								
Experiment	1	2	3	4	5	6	7	8
RMSE(mm)	4.08	1.87	1.86	1.92	1.91	1.94	5.02	5.05
ME(mm)	-0.02	-0.04	-0.03	-0.24	-0.03	-0.25	-0.02	-0.02
MAX(mm)	21.81	20.28	20.01	17.03	19.87	17.28	29.63	34.26
MIN(mm)	-53.32	-48.65	-48.65	-48.72	-48.60	-49.04	-64.81	-64.66
MAE(mm)	0.04	0.02	0.02	0.02	0.02	0.02	0.05	0.05
Invalid fits	4	4	4	2	4	2	1	2
Splines on 110 data points, validated at 288 points								
Experiment	1	2	3	4	5	6	7	8
RMSE(mm)	2.26	2.38	2.28	2.25	2.24	2.22	5.04	5.10
ME (mm)	0.01	0.01	0.01	-0.20	0	-0.22	0.02	-0.01
MAX (mm)	24.22	24.12	24.00	22.93	24.21	23.04	33.78	37.91
MIN (mm)	-53.72	-54.29	-53.53	-52.37	-53.53	-53.58	-64.70	-64.73
MAE (mm)	0.02	0.03	0.02	0.02	0.02	0.02	0.05	0.05
Invalid fits	7	3	6	2	5	2	7	5

Interpolating the square-root transform of rainfall (4, 6) did not improve estimation accuracy but reduced the number of invalid surfaces, which is a benefit for operational use. Including elevation as a third independent (nonlinear) variable and using two guiding topographic covariates gave unexpectedly poorer estimation accuracy (RMSE ~4 mm) and a wider range of estimated rainfall values. Also surprisingly, the two spline models which used radar rainfall data had the poorest estimation accuracy (RMSE ~5 mm) and the widest range of predicted rainfall amounts (+ 34.26 mm to - 64.66 mm); however these surfaces had near-zero biases and failed the 'robustness' criteria least often (only one or two invalid surfaces).

Whilst the smaller data set was a little more difficult to interpolate, with higher estimation error and more daily surfaces found to be invalid, the average estimation error using any of the spline models 1-6 was increased by only about 0.5 mm up to 2.4 mm. Given the 8-fold reduction in the number of data points, this deterioration in accuracy is relatively minor and less than might be expected from a much sparser data set. Given the similar estimation accuracies from models 1-6, the preferred choice would probably be between models 2, 4, or 6, although any would probably suffice. Model 2, which treated elevation as a linear covariate along with two collateral topographic variables has a slightly higher estimation error (RMSE = 2.38 mm) but has zero bias and is reliable on all but 3 days. Models 4 and 6 used the square-root transformation and again have the most consistently reliable fit throughout the month. These models however have a slight negative bias and lower maximum precipitation estimates. The other models (1, 3, 5) also had similar estimation accuracy and zero biases, but had a greater number of days when the fit was considered unreliable. This suggests that modelling elevation as an independent non-linear (models 1, 5) or linear covariate (model 3) makes little difference to the estimation accuracy for this volume of input data. The additional topographic variables did not give model 1 an advantage.

Again, when radar data was included, although the bias in the splined estimates remained close to zero, the RMSE doubled to over 5 mm and the maximum predicted values of precipitation rose by around 10 mm. Evidence from other studies where a collateral data set was used to further guide estimation of precipitation, suggests that the collateral variable needs to exhibit a strong correlation with the primary variable in order for estimation accuracy to be improved by its inclusion. For example, in studying the influence of elevation on the accuracy of estimating daily

precipitation, Goovaerts (2000) found no additional benefit of including this collateral data when the daily strength of correlation was below 0.75. In this study, the daily correlation between the gridded rainfall radar accumulations and the totals observed at coincident rain gauges was only above 0.75 for 4 days and only above 0.6 on 10 days during the one month data period.

4.3.5 Conclusion

A data set of one month's daily rainfall observations from England and Wales has been used to explore the accuracy of different spline interpolations for estimating daily precipitation at a resolution of 5 km. For a relatively dense network of 820 rain-gauges, a 2-D partial thin-plate spline on (x,y) with elevation as a single linear covariate robustly estimated daily precipitation across England and Wales with an RMSE averaged over all stations and dates of ~1.9 mm. The models were considered unreliable for up to 4 days in the month. For a sparser network of 110 points, the RMSE was ~2.3 mm and models were unreliable on up to 7 days in the month. Experiments to model precipitation as a non-linear function of elevation or to add further collateral variables such as topographic data, did not improve the estimation of precipitation amount considerably with this volume of input observations. Inclusion of rainfall radar data surprisingly led to poorer estimation accuracy by all spline models.

Whilst still preliminary, these results suggest that the densification of the near real-time rain-gauge networks in the UK from 150 to upwards of 500 stations should improve the accuracy of estimating daily rainfall surfaces by about 0.5 mm on average, although the improvements may be greater in some places. Although radar estimates of rainfall provide useful information on the extent of the rainfall area, it seems these data may only improve estimation of actual rain amount by splining when there is a strong correlation between the rain-gauge measurement and the rain radar accumulation values at corresponding locations.

4.4 RELATIVE HUMIDITY

Atmospheric water can have a large effect on pests, pathogens, vectors and host plants, so a measure of its variability in space and time is needed by many Decision Support Systems intended for use in agriculture and horticulture.

At any temperature there is a limit to the density of water vapour in the air and hence a consequent upper limit to the vapour pressure. Above this upper limit vapour can no longer enter the air, so that condensation equivalent to the additional amount of vapour is produced. Evaporated water enters the atmosphere as individual energetic water vapour molecules. The resulting moisture content of the air can be expressed in numerous ways. For example, as the absolute humidity ($g\ m^{-3}$), which is the total mass of water in a given volume of air. The mass-mixing ratio (e) is the mass of water vapour in grams per kilogram of dry air. Specific humidity (q), is nearly equivalent to (x), and is defined as the mass of vapour per kilogram of air including its moisture. Most commonly however, for agricultural purposes the air moisture content is measured by Relative Humidity (RH).

Relative humidity is the ratio of the actual mixing ratio to the saturation mixing ratio and is usually expressed as a percentage.

$$RH = \frac{x}{x_s} \times 100 \approx \frac{q}{q_s} \times 100 \approx \frac{e}{e_s} \times 100$$

where the subscript s represents the respective saturation values at the same temperature (Barry & Chorley, 1987).

Because vapour pressure is functionally dependent upon temperature, the relative humidity of air also depends upon temperature. Without any change in moisture content, RH will

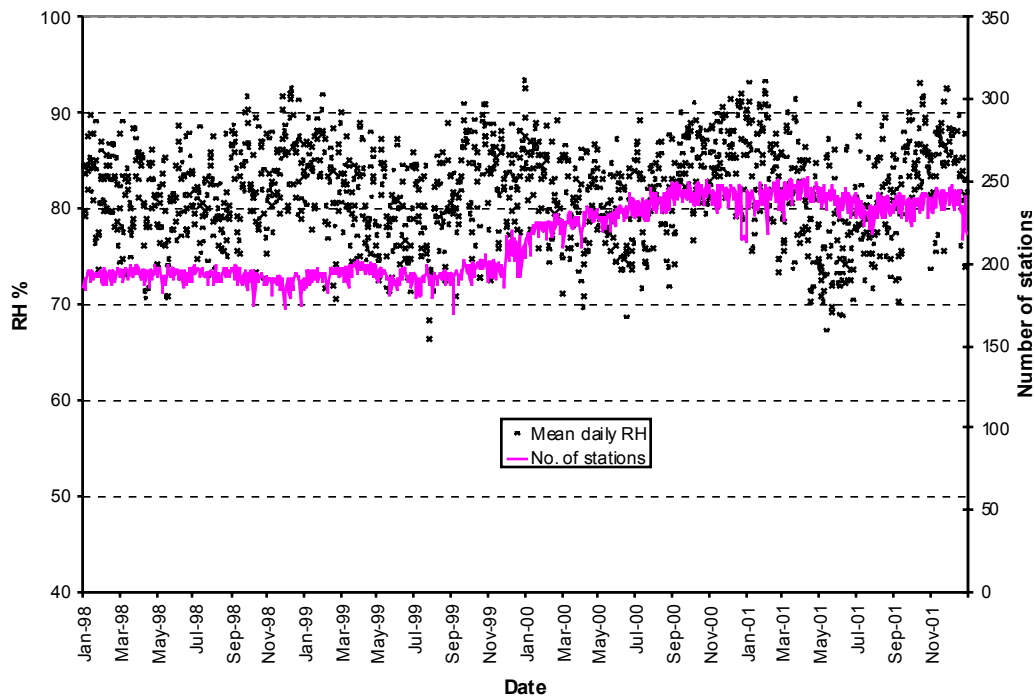
therefore alter whenever the temperature changes. As a consequence, unless it is also reported with the respective air temperature RH is only useful in a qualitative way (Rosenberg *et al.*, 1983).

The present study investigated the extent to which RH could be estimated for unsampled locations guided by measurements from a sparse network of recording stations.

4.4.1 Network & data description

The daily mean RH was estimated using only those stations that had at least 12 hourly readings during each day. This number was usually around 170 during 1998 and 1999 but increased to almost 250 from during 2000 onwards (Figure 4.6). The location of these stations across the UK (Figure 4.7) reflects their primary purpose as a monitoring network for weather forecasting, *i.e.*, many are either on the coast, located at airfields or out at sea.

Figure 4.6 – Daily mean relative humidity from January 1998 through December 2001.



Examination of the measurements recorded hourly by these stations shows that daily mean RH varies by around only 20% (Figure 4.6).

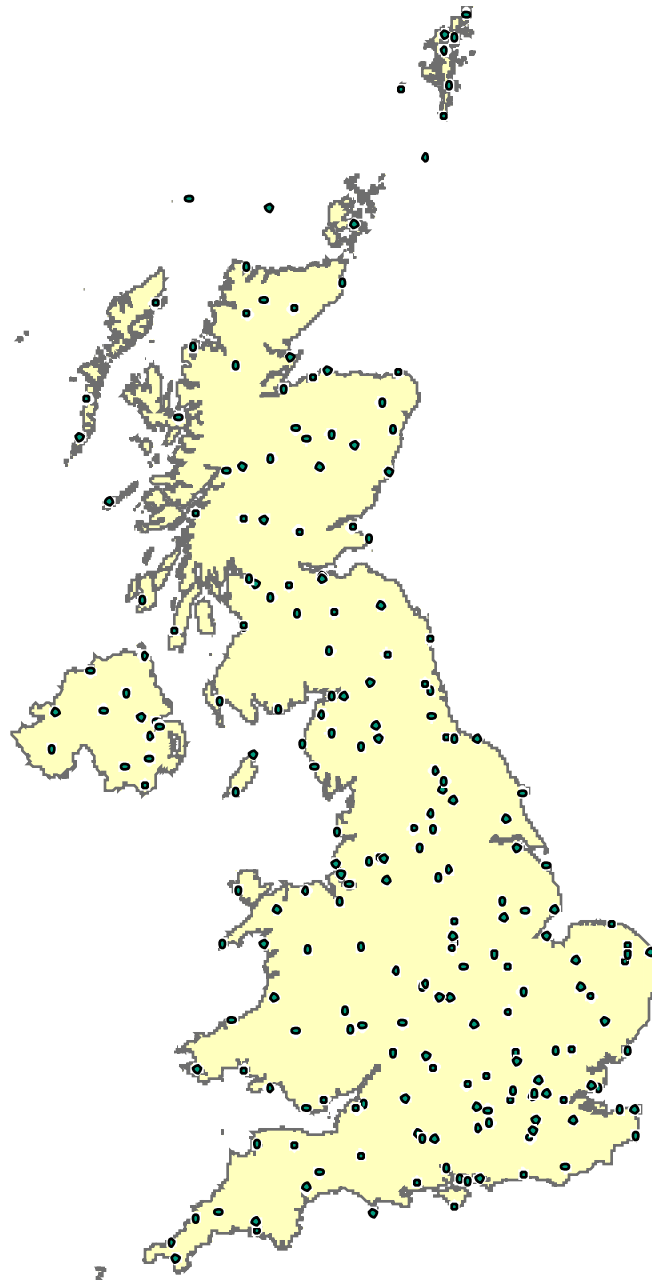


Figure 4.7 – Location of weather monitoring stations across the UK with hourly records of relative humidity adequate for informing spatial interpolation

4.4.2 Daily mean Relative Humidity

The relative accuracy of interpolation methods was compared (Table 4.7). The comparison criteria used is the root mean square error of prediction (RMS) which has the same units as the interpolated phenomenon. Accurate and precise interpolations are therefore indicated by small RMS values, ideally close to zero.

Table 4.7 – Comparison of interpolation methods for daily RH values on two selected days.

Interpolation Method	RMS	
	1 June 2000	20 May 2001
Ordinary Kriging	4.394	5.096
Ordinary Cokriging	4.362	5.053
Spline with tension	4.361	5.079

There is not a large difference between any of the interpolation methods used. The additional variable used for cokriging was elevation of the monitoring stations which was only weakly correlated (coefficient of correlation = 0.2). Five different variants of spline interpolation were tested but the tension spline always produced the lowest RMS. The advantage of the spline method is that there is no requirement to fit a model function to an experimental variogram.

Variograms for mean daily RH were specific for the day that they were fitted. Predictive performance for other days was poor. This was expected because RH is dependent on other weather factors, especially temperature. As a consequence, an InterMet service will need to fit unique variograms each day for daily mean RH (and any of its derivatives). One alternative to this operational constraint, not examined here, might be to include temperature as a covariate within cokriging. Further development of InterMet would examine this possibility.

Figure 4.8 is an example from a day that had above average relative humidity. The kriging method allows the standard error of the interpolation to be mapped as well, which illustrates where the fitted variogram deviates from the input data.

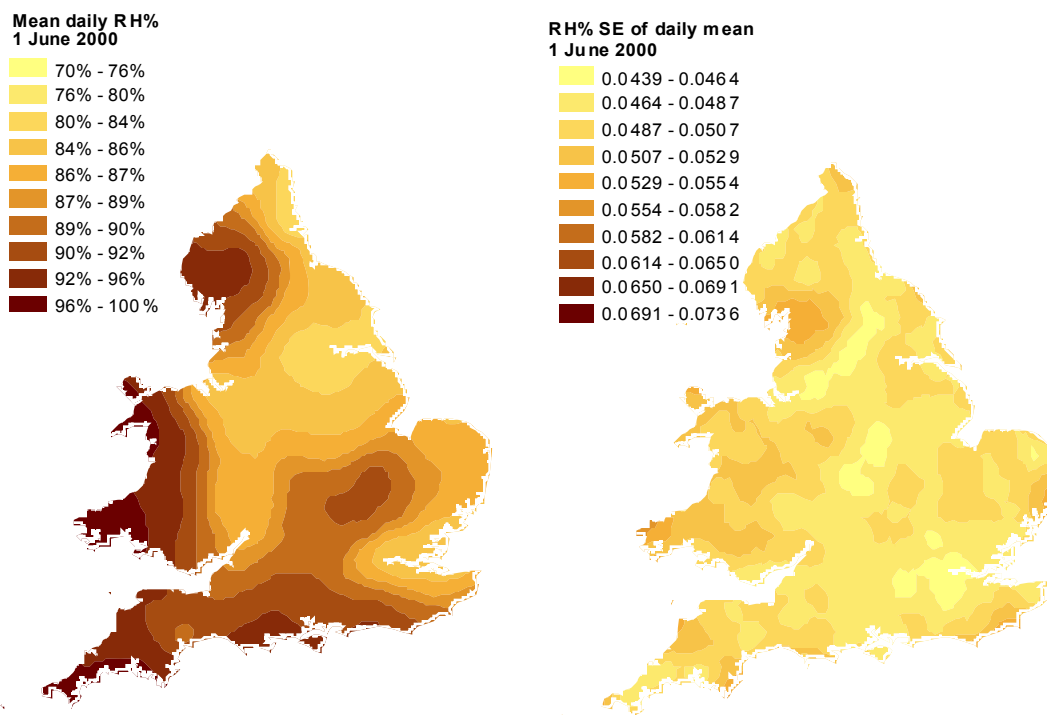


Figure 4.8 – Spatial interpolation example of relative humidity (daily mean) and standard error of the interpolation method (co-kriging).

4.4.3 Relative humidity thresholds

The techniques applied to estimate mean daily RH can be used to interpolate weather variables that are more directly meaningful for agri-meteorological applications. For example, to decide whether conditions favour the development of a plant pathogen, it may be useful to know how many hours each day the relative humidity exceeded a critical value such as 90% (Figure 4.9). The different interpolation methodologies again produced fairly similar results. However, the kriging technique can be extended to yield a probability that a defined event will occur. To exemplify this, the surface interpolated and shown in Figure 4.4a, was expressed to show the probability of more than 11 hours when RH was 90% or greater during the day (Figure 4.4b).

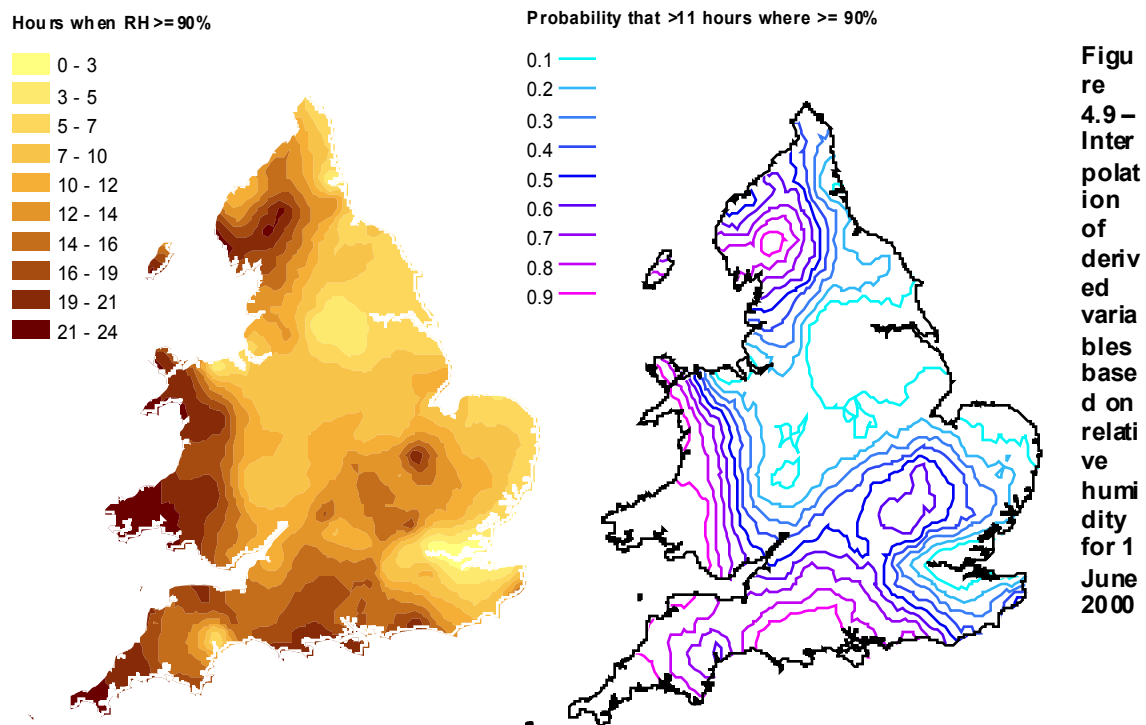


Figure 4.9 – Interpolation of derived variables based on relative humidity for 1 June 2000

4.4.4 Interpolation of outputs from a decision support system (DSS)

Decision Support Systems based either wholly or in part on weather data can be run through the Internet database and the results (as a date) interpolated. Using the Smith Period¹ warning system aimed at potato late blight disease the date of the first period and the seasonal total of periods can be interpolated (Figure 4.10). Detailed examination of the impacts of this approach on risk warnings is presented section [Section 7.2]

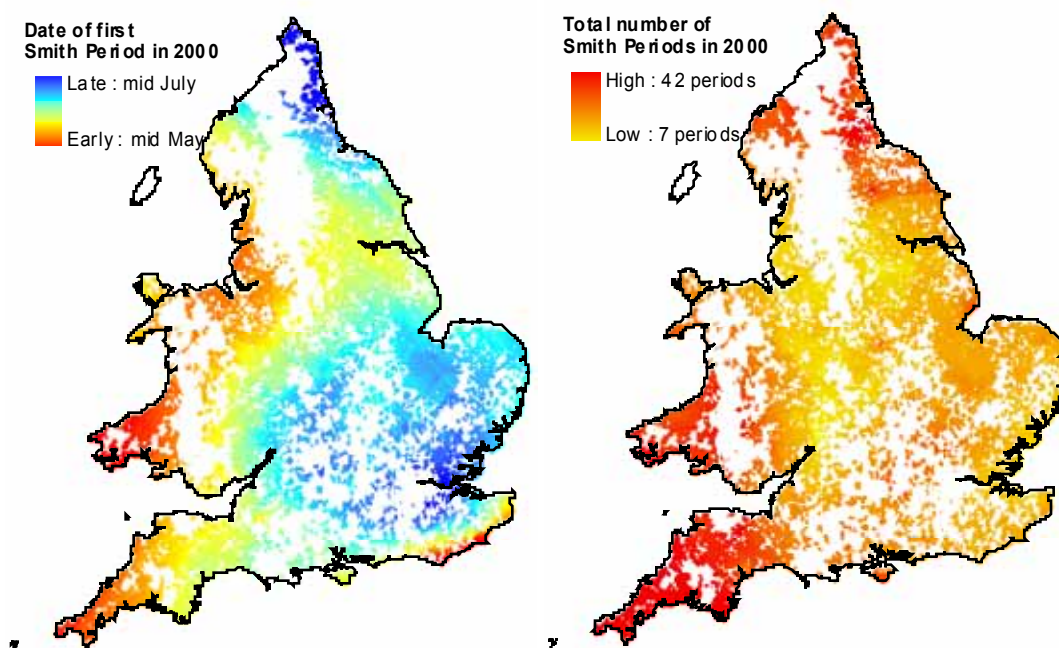


Figure 4.10 – Interpolated Smith Periods, pixels based on 2km x 2km grid squares that grew potatoes in 1997

¹ The Smith Period (Smith, LP; *Plant Pathology* 5, 83-87, 1956) is defined as two consecutive days ending at 0900 when the minimum temperature never falls below 10°C and the RH is above 90% for at least 11 hours on each day.

4.5 WIND

Air movements affect the development of plant disease epidemics by transporting pathogen inoculum and vectors between locations. And can similarly affect dispersal of insect pests and weed species. These effects may operate at a number of scales; from within crop canopies upto continental levels. In addition, the application of pesticide sprays is only permitted below certain wind speeds.

The movement of air can be described by its speed direction and turbulence. Whilst turbulence is important to plant pathogen, vector and pest movements, it is not commonly quantified. Certainly its influences are unlikely to be incorporated within practicable Decision Support applications in the foreseeable future. This study therefore concentrated on examining the potential to interpolate wind speed and direction.

The monitoring stations of the UKMO sample wind at spatial and temporal scales designed to gather a general description of wind variability to aid to weather forecasting. Although around 750 stations monitor wind speed many of these only make one or two observations each day at a prescribed time (usually 0900 hrs). To avoid using such readings, which could bias the interpolation, only stations with at least 12 readings each day were used to generate surfaces for daily average wind speed and direction. The stipulation reduces the number of available data points to around 230 – 250 on any particular day (Figure 4.11).

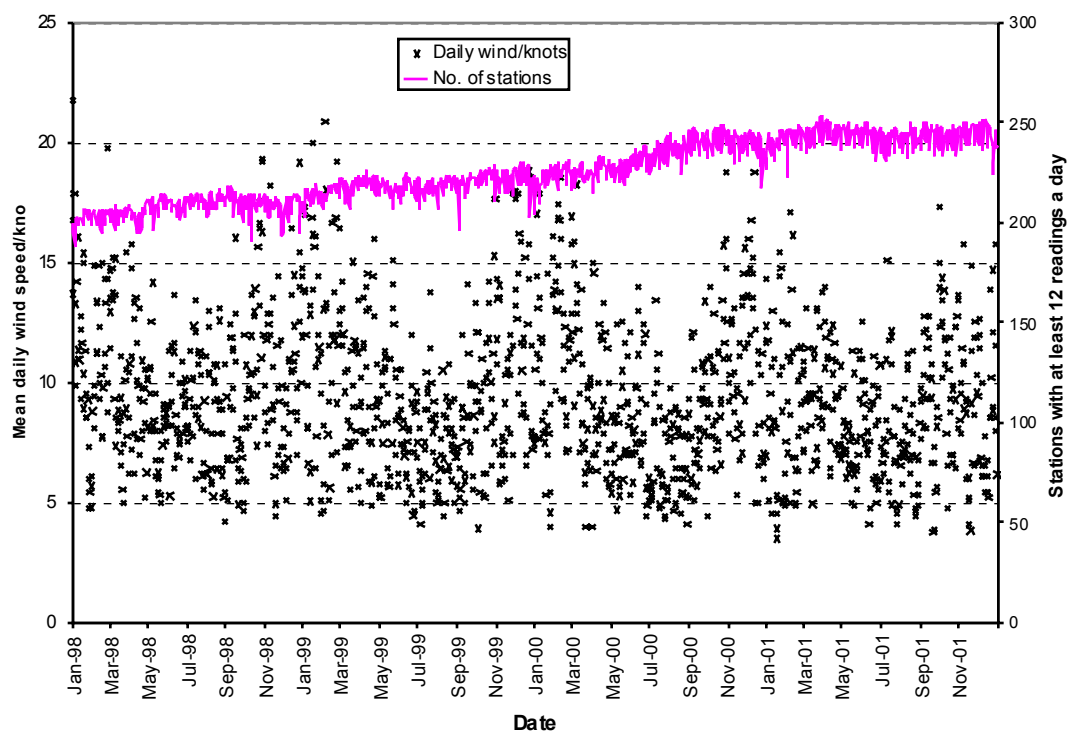


Figure 4.11. Mean daily wind speed (knots) and number of recording stations with at least 12 readings each day between 1 January 1998 and 31 December 2001

Unlike relative humidity and temperature, wind speed and direction are likely to vary hugely over short distances particularly close to the ground, but the monitoring network has not been designed to collect readings at such a scale (Figure 4.12).

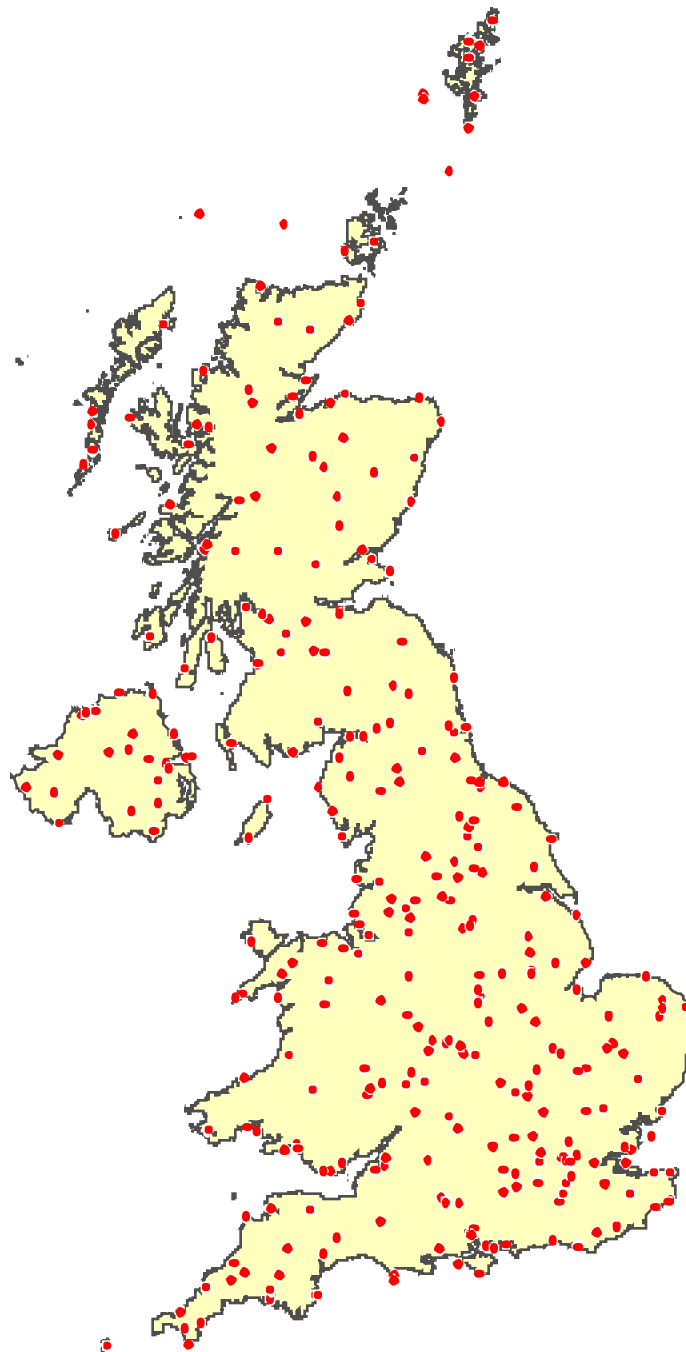


Figure 4.12. Location of wind monitoring stations with at least 12 readings each day

4.5.1 Wind Speed

Seven interpolation methods for estimating wind speed surfaces were compared, using the RMS error to judge their accuracy and precision as before. The minimum and maximum wind speeds predicted by the methods was also used to check how well the different methods maintained the range of the input data (Table 4.8). Wind measurements from the 27 March 2001 were used for the evaluation.

Table 4.8. Comparison of interpolation methods for daily mean wind speed (knots)

Method	Min. estimate	Max. estimate	RMSE
Trend Surface Analysis	8.99	18.11	3.80
Inverse Distance Weighted	6.09	23.80	3.42
Local polynomial	0.29	23.00	3.33
Spline	5.11	25.27	3.36
Ordinary kriging	5.19	26.43	3.18
Universal kriging	6.32	23.14	3.36
Cokriging	5.19	26.62	3.06

The RMS errors show some variation but all methods produced a result of less than 4 knots, the maximum and minimum estimation was more variable (the range of the input data was 0 – 35 knots). The cokriging method (again using elevation as the covariate) resulted in the lowest overall error (figure 4.1.3).

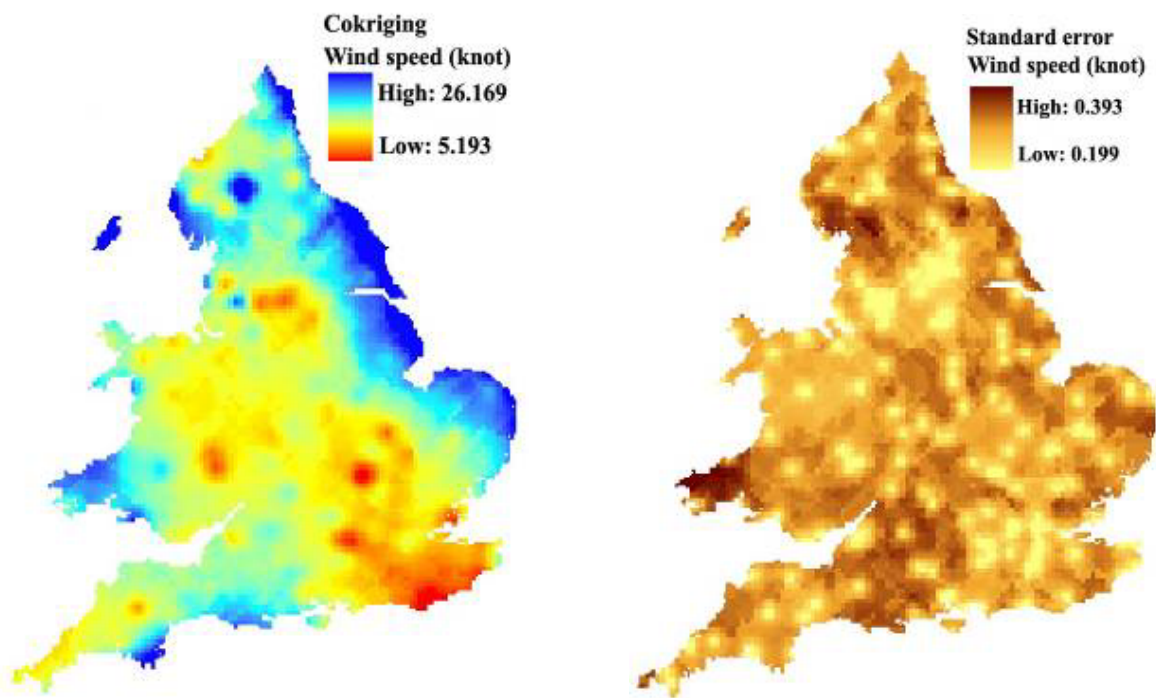


Figure 4.13. Daily mean wind speed and standard error maps for 27 March 2001

4.5.2 Wind Direction

Unlike simple wind speed the direction from which the wind blows has two components i.e., orientation and strength, making it a vector rather than scalar quantity. As direction should be biased in favour of stronger winds an approach was developed to decompose the direction and simultaneous speed measurements into separate Cartesian components – an x and a y direction. Each component was interpolated separately by the usual methods and then recombined. The final image of wind direction can be difficult to visualise, as the data is circular, so a palette was defined which arranged all possible values into 45° sectors. A similar colour was used for winds in the 316° - 360° and the 1° - 45° ranges (Figure 4.14). The values in the interpolated map covered over 130° (easterly to south westerly), so were almost completely contained within three sectors.

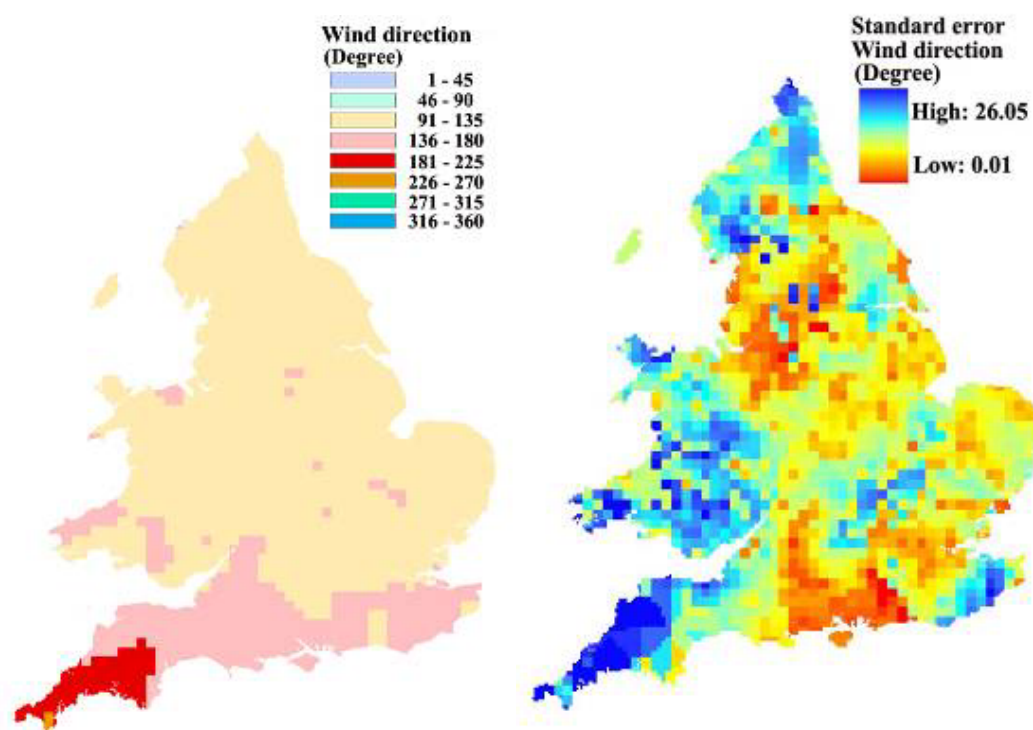


Figure 4.14. Average wind direction and associated standard error using ordinary kriging for 27 March 2001.

4.6 SOLAR RADIATION AND SUNSHINE HOURS

Solar radiation is the primary energy source of many physical and biological processes. New DS systems that model crop growth, especially those estimating yield components, are likely to require some estimate of energy receipts by the crop system.

Far fewer weather stations monitor global solar radiation (GSR) than other weather variables, such as rainfall or air temperature, mainly due to the high cost of the required equipment. As the maximum amount of radiation falling on the Earth's surface is related to latitude and time of year it would be possible to allocate any field a GSR value based on what was recorded at its nearest monitoring station. Such Voronoi polygons ensure that the land surface is as efficiently covered by the data points as possible. For example, the 36 stations available to InterMet that record global solar radiation can be used to delimit GSR estimates for unsampled areas (Figure 4.15).

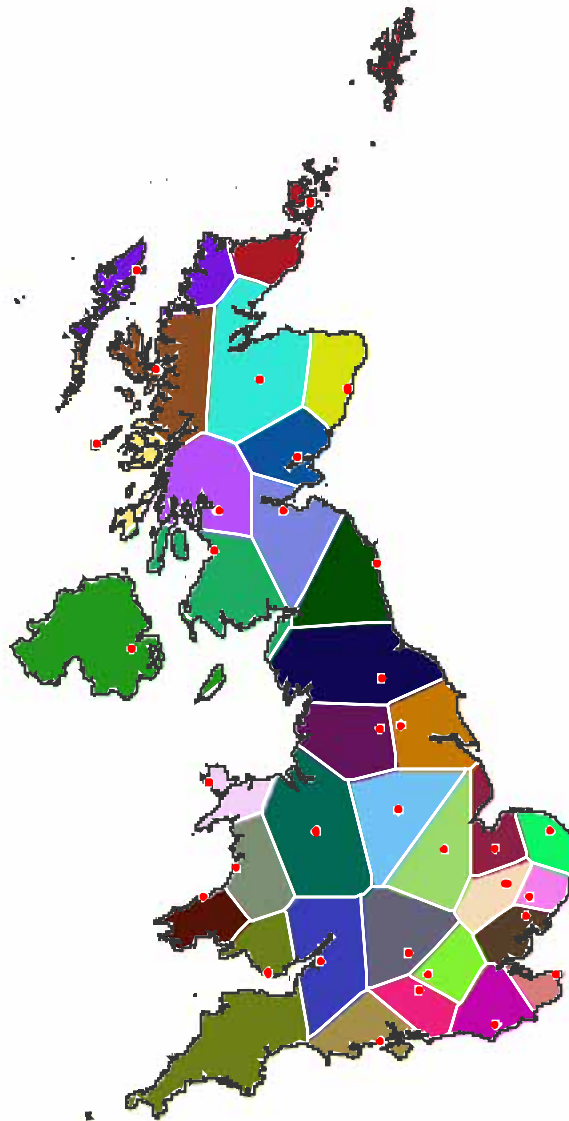


Figure 4.15. Allocation of all land areas to the nearest solar radiation recording site

Cloud cover, depending upon its thickness and area can create a significant barrier to radiation receipts at the ground. The proportion of incident radiation reflected is termed the

albedo (expressed as a fraction or percentage) is affected by cloud type, so the albedo for overcast conditions can vary from around 50% for cirrostratus to 90% for cumulonimbus. As a consequence, over the large areas derived from such a sparse network of measurement stations, estimates of GSR are likely to be poorly quantified – especially those that contain coastal boundaries.

Sunshine recorders are more common than radiation monitors (Figure 4.16), so an attempt was made to relate hours of sunshine to GSR using multiple regression.

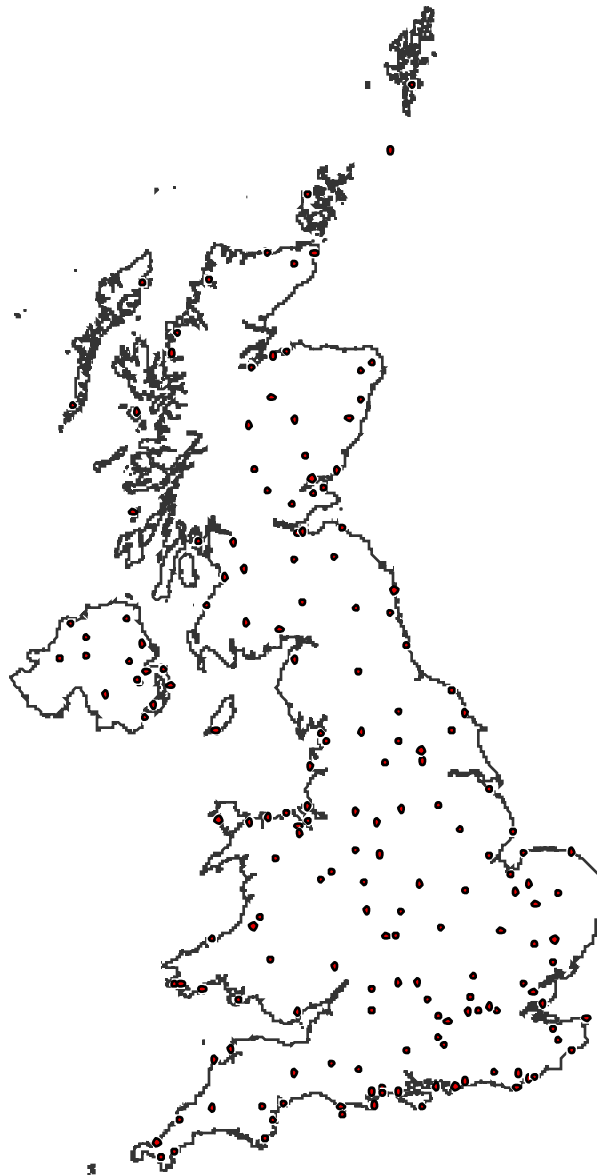


Figure 4.16. Location of sunshine recording stations

Sunshine hours alone are only weakly related to the GSR recorded at the same site (Figure 4.17). However, the precision of the relationship was improved by accounting for other factors, specifically; station latitude, calculated day length and the proportion of each day that the sun shone (Figure 4.18). The day length calculation and therefore the proportion of the day that the sun can shine is dependant on the latitude and time of year. The variation in day length between locations depends on their relative north to south position and time of the year. For example, on 17 October the difference in day length between Exeter (latitude 50.87°) and Carlisle (latitude 54.62°) is 27.5 minutes, while on 20 June the difference is 59.75 minutes.

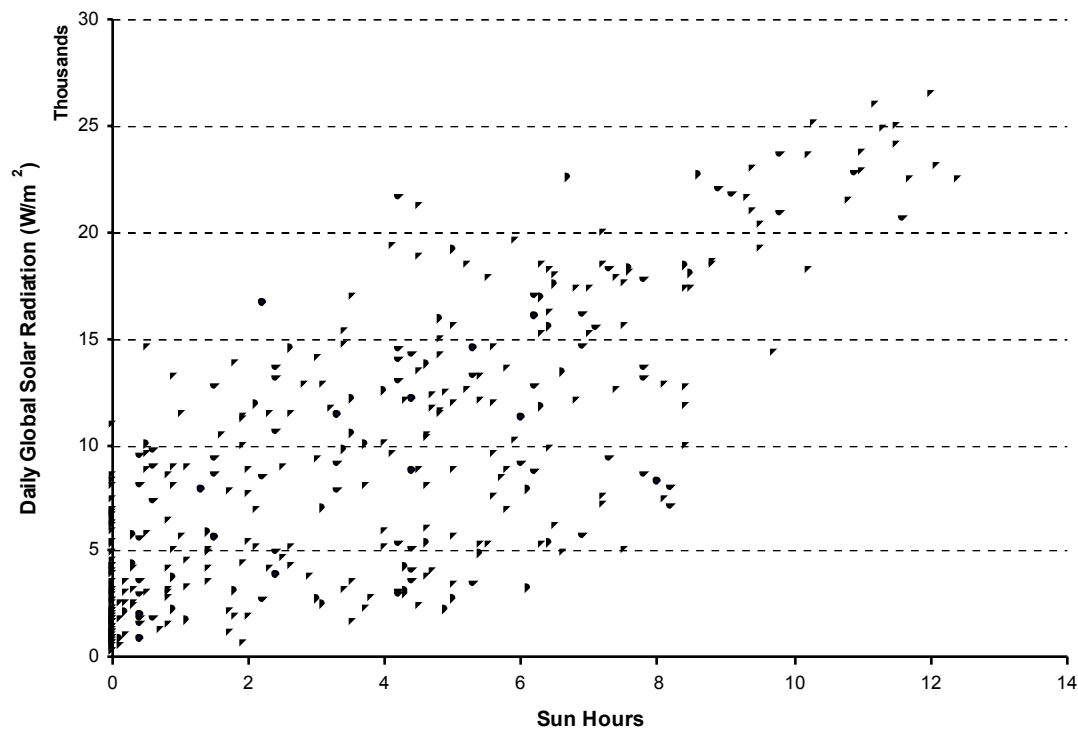


Figure 4.17. Relationship between sunshine hours and daily global radiation at Wallingford, Oxfordshire, throughout 1998

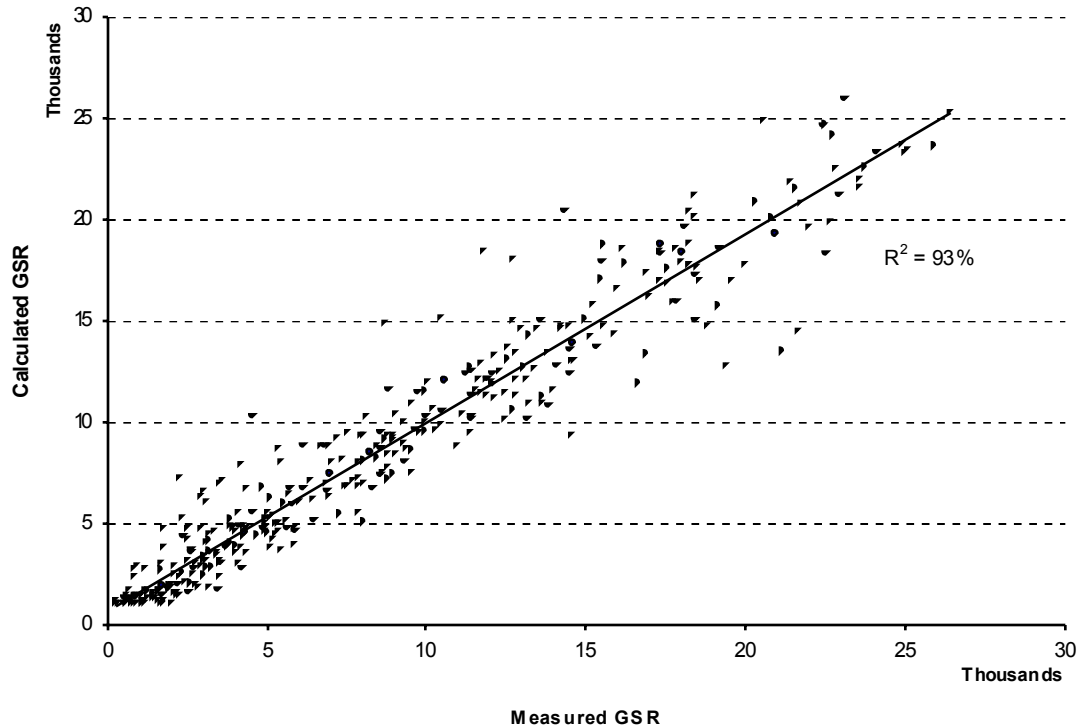


Figure 4.18. Multiple regression result using the Wallingford data to estimate GSR from sunshine hours, the proportion of the day that the sun shone and the calculated day length.

The precise values of the coefficients in the multiple regression equation varied by location, so data from five places was used to produce a generic equation. A comparison between the recorded GSR values and those calculated using the regression was carried out using a paired t-test but they were not significantly different ($P < 0.05$). The general applicability of the regression equation was tested further by using another location that was not used to develop the regression equation, but which had both GSR and sunshine duration data. Calculated GSR values from this site were again not significantly different ($P < 0.05$) from the measured values. The sunshine hours dataset (which numbered over 180 locations) were used to interpolate a continuous map of GSR using the generic multiple regression model (Figure 4.19).

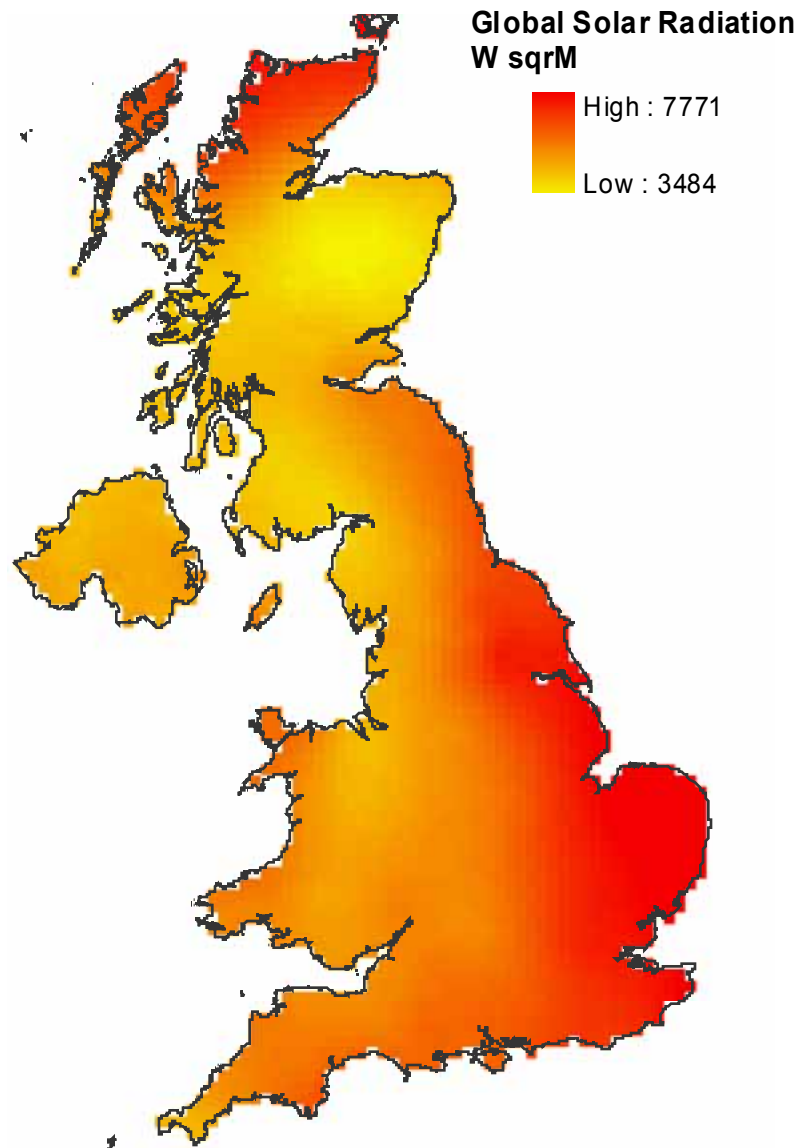


Figure 4.19. Estimated GSR for 17 October 2001 based on multiple regression of data for sun hours

5. INTERPOLATION OF SOIL MOISTURE AND TEMPERATURE

5.1 OUTLINE

Soil temperature and moisture are important to crop production via their direct effects on plant growth and development, and also indirectly by their effects on pest and pathogen populations. Generally however, crop managers have at best crude estimates of soil conditions. Better knowledge of the soil environment is therefore to the aspiration of predicting crop performance more accurately and to the goal of improved crop management.

This section describes work to implement the use of the van Genuchten approach to modelling soil moisture (van Genuchten, 1980) and the incorporation of multiple layers to describe soil within an existing UKMO modelling framework.

5.2 CHOICE OF MODELS

In 2002, and for many years before this, the UKMO reported soil moisture deficit (SMD) as one of the variables available within its sophisticated meteorological reporting package – Meteorological Office Rainfall and Evaporation Calculation System (MORECS). MORECS reported on a weekly basis and data were available for 40 x 40 km blocks across the UK. This package reported SMDs for a range of common crops using water balance calculations based on crop specific root density distribution. Initially it had been intended to use a modified version of MORECS within InterMet.

During 2003, however, all development work on MORECS was stopped by the UKMO, because it was superseded by a new system; Meteorological Office Surface Exchange System (MOSES). Following this UKMO decision MORECS could no longer be used for InterMet as the model would have required further development such as allowing a 15 x 15 km grid (or maybe at a later date even denser grids) to be imposed. Hence, the use of MOSES became the only option. This increased the model development workload for InterMet (specifically, scientists at NSRI), because some of the soil elements of the MORECS system are not available within MOSES.

An alternative option was considered at this time, that was to use the simple water balance models developed for an EU funded research project; IMPEL 3. These models were rejected because they only considered arable crops, required detailed cropping information and raised complex intellectual property issues. MOSES on the other hand, provided a more generic approach which was deemed to be much more appropriate for InterMet.

5.3 MOSES DEVELOPMENT

In the standard MOSES model, although the soil profile was described by 4 horizons (layers) of various depths, the same soil physical parameters are used to describe each layer. The program was unable to respond either to soil profiles with more or less than 4 horizons, or where specific and different physical properties were available for each layer. In effect soil was treated as one layer and therefore variations in soil properties with depth had negligible influence on moisture modelling. The MOSES model was modified for InterMet, to implement the use of the van Genuchten approach to modelling soil moisture and the incorporation of multiple layers to describe soil.

In order to incorporate MOSES into the Model Integration Framework (MIF), soil property information for each grid cell was required. This was accommodated by restructuring the MOSES code, so that all soil related information can be read from an external file and allowing the specification of up to 9 soil horizons.

In addition, it was necessary to integrate MOSES into the MIF framework and to write computer code to initially store and then extract the relevant soil physical property information required by the NSRI-modified MOSES model on a series by series and a layer by layer basis

In order that development work on the integration of the NSRI-modified MOSES and MIF could be done at NSRI, a stand-alone Mini-MIF framework was supplied by EUGD. This allowed NSRI to develop optimum methods of integration and eliminate, or at least greatly reduce, any future difficulties in final model integration. Programs have been written to extract the relevant soil information required by MOSES on the basis of soil series. In addition, horizon-specific soil moisture and temperature results produced by MOSES were transferred to the Mini-MIF to be linked to produce grid-based result files which can be used in their own rights or as inputs to further models. The integration of MOSES and the mini-MIF system has been completed.

5.4 INITIAL SENSITIVITY TESTING OF THE NSRI-MODIFIED MOSES MODEL

Soil physical data for individual layers of eleven contrasting soils were assembled from NSRI soil archives (LandIS) and from soil moisture release data collected specifically for this project. Hydraulic parameters were derived from pedo-transfer functions and an input file built up as shown in Table 5.1. Details of the soil parameters calculated for each of the trial soils are given in Appendix 2.

Weather data for Cardington, Bedfordshire, was provided by UKMO at 15-minute intervals over a six month time period (January – June, 2001) giving a total of 17,472 time steps. These weather data were supplied in the correct format for direct input to MOSES and included the parameters shown in

Table 5.2. An example of this dataset is given in Appendix 1.

Table 5.1. Hydrological parameters required for NSRI-modified MOSES soil data input files

Parameter	Description	Units
NSOIL	Number of soil layers	-/-
DZSOIL	Soil thickness	m
B_EXP	Exponent used in calculation of soil water suction and hydraulic	-/-
SATCON	Saturated hydrological conductivity of the soil	$\text{kg m}^{-2} \text{s}^{-1}$
SATHH	Saturated soil water suction	-/-
V_CRIT	Volumetric soil moisture content at saturation	$\text{m}^3 \text{m}^{-3} \text{soil}$
V_SAT	Volumetric soil moisture content at the critical point	$\text{m}^3 \text{m}^{-3} \text{soil}$
V_WILT	Volumetric soil moisture content at the wilting point	$\text{m}^3 \text{m}^{-3} \text{soil}$
HCAP	Soil heat capacity	$\text{J K}^{-1} \text{m}^{-3}$
HCON	Soil thermal conductivity	$\text{W m}^{-1} \text{K}^{-1}$

Table 5.2. Weather data parameters supplied for Cardington as input files for NSRI-modified MOSES

Original data	Code	Units
Surface downward shortwave radiation	SW	W m^{-2}
Surface downward longwave radiation	LW	W m^{-2}
Rainfall rate	RAIN	$\text{kg m}^{-2} \text{ s}^{-1}$
Snowfall rate	SNOW	$\text{kg m}^{-2} \text{ s}^{-1}$
Air temperature	TA	K
Westerly wind component	U	m s^{-1}
Southerly wind component	V	m s^{-1}
Surface Pressure*	PSTAR	Pa
Specific humidity	QA	kg kg^{-1}

The NSRI-modified MOSES model was run for each soil and a time series of soil moisture and soil temperature data was generated for each layer (up to 5) in each soil series. The results of these runs with the Cardington weather data are shown in Appendix 3, and two examples for Batcombe and Quomdon series are described below.

5.4.1 Batcombe series

This is a silty clay loam over clay soil that suffers from seasonal surface wetness caused by slowly permeable subsoils. The moisture content (Figure 5.1) of each layer was arbitrarily set at field capacity moisture content at the beginning of the sequence (1 January, 2001). This equates to 45 per cent by volume for the topsoil layer— sm1; 40 per cent for the subsurface layer— sm2 etc. In the topsoil (sm1) and subsurface layer (sm2) moisture contents rise to 55 per cent and 49 per cent respectively by time step 3,000 (end of January). These moisture contents, as shown by the data tables given in Appendix 2, p.3 are equivalent to volumetric water content at saturation (total pore space). Moisture contents remain within 2-3 per cent of these levels until the time step 12,000 (end April). Topsoil moisture content should have the capacity to fluctuate over a 48 hour period between 55.8 per cent at saturation and about 45 per cent at field capacity as coarse pores fill and empty following rainfall and subsequent drainage. Similarly the subsurface layer (sm2) should fluctuate between 49 per cent (saturation) moisture content and 40 per cent (field capacity). In fact, it remains static until drying commences about time step 12,000 (end of April) when transpiration begins to draw water from this layer and declines gradually to about 35 per cent at the end of the time sequence. The model does not allow realistic drainage through its coarse pore space due to gravity between rainfall events from these upper layers.

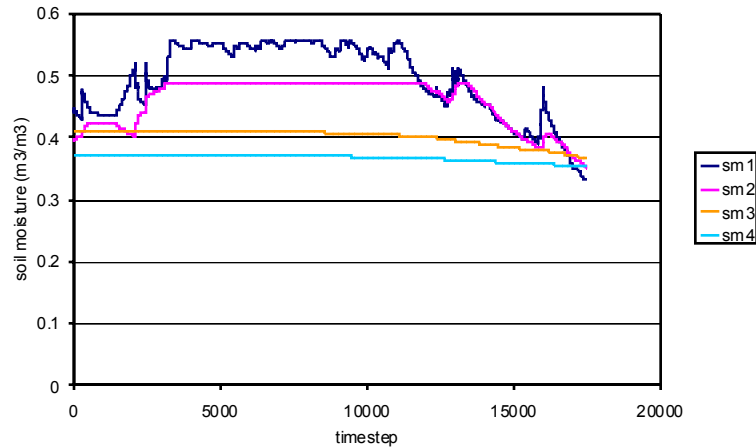


Figure 5.1. Soil moisture in Batcombe series (January – June 2001)

The modelled soil moisture data predicts uniform soil moisture contents at 41 and 38 per cent in the dense clay-rich subsoil layers (sm 3 and sm4 respectively) through the January to April period with a subsequent slight decline in moisture content during May and June. These levels are below saturation (46 and 41 per cent respectively) further demonstrating that the model does not allow sufficient drainage of excess water from surface layers.

Soil temperatures predicted by MOSES for the Batcombe series are given in Figure 5.20. Topsoil and immediate subsurface layers show appreciable diurnal variation in temperatures but as would be expected temperatures change only slowly in response to gradual increases in average daily temperatures in dense subsoils with uniform moisture contents.

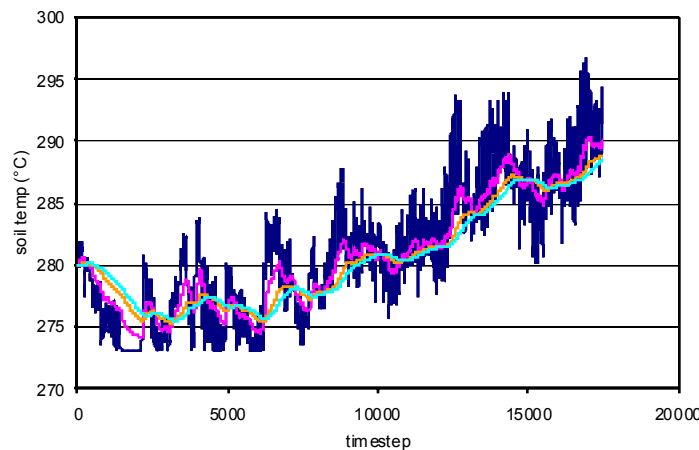


Figure 5.20. Soil temperature in Batcombe series (January – June 2001)

Soil temperatures predicted by MOSES for the Batcombe series are given in Figure 5.20. Topsoil and immediate subsurface layers show appreciable diurnal variation in temperatures but as would be expected temperatures change only slowly in response to gradual increases in average daily temperatures in dense subsoils with uniform moisture contents.

5.4.2 Quorndon series

This permeable sandy loam soil is affected by fluctuating groundwater during the winter half of the year. For the first 10,000 time steps (up until early April) topsoil moisture content (sm1 in Figure 5.3) is predicted to fluctuate between 27 and 35 per cent in response to rainfall, which is at or around field capacity moisture content (33 per cent).

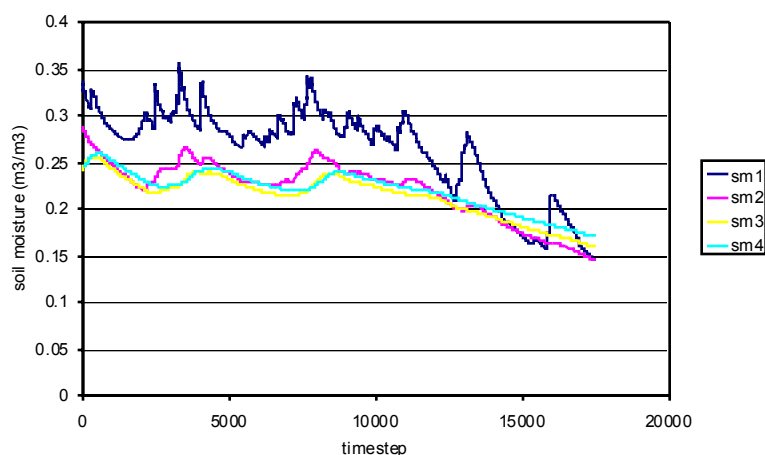


Figure 5.3. Soil moisture in Quorndon series (January – June 2001)

The hydraulic conductivities of Quorndon soil layers are at least ten times greater than in the Batcombe series and the model gives more realistic results for these coarse textured soils. However, the coarse porosity (drainable pore space) of Quorndon topsoils is between 15 and 20 per cent, so again the model underestimates the rapid drainage of coarse pores after rainfall. Moisture contents in the other three horizons (sm2, sm3 and sm4) were predicted to be much less variable than in the topsoil, but nevertheless still respond slowly to peak rainfall events.

Predicted soil temperatures for Quorndon series are given in Figure 5.4.

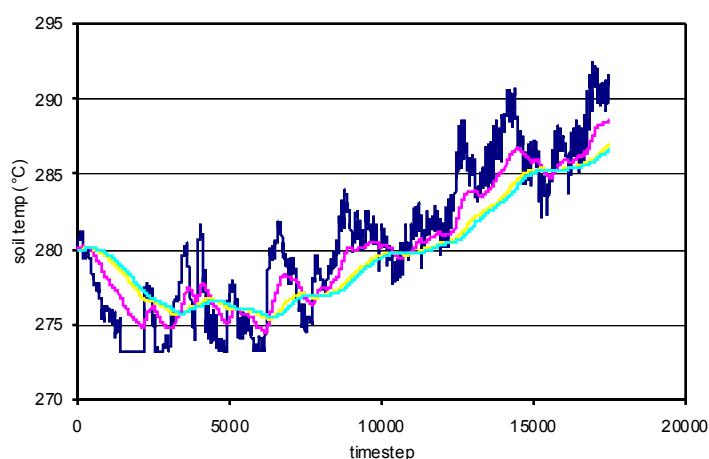


Figure 5.4. Soil temperature in Quorndon series (January – June 2001)

The NSRI-modified MOSES output appeared to show marked differences in temporal soil moisture content across the range of soil profiles tested with the Cardington weather data. A statistical analysis was undertaken to check for any statistical significance in the apparent differences. The results of this analysis for topsoils and deep subsoil layers for each soil are discussed below. In order that the patterns of change within each time series of data could be analysed, each time series was standardized, so that each sequence of data began at zero soil moisture or zero soil temperature. *Repeated measures analysis* was used to examine the data, because every time step was generated from identical weather data and, therefore, any variation that was identified would provide information on inherent differences between the soils.

When the eleven topsoils were compared there were statistically significant differences between each soil type in terms of moisture content (Appendix 3). The mean step change was different in all eleven soil series (Figure 5.5). In particular, the Andover and Blackwood series had consistently large mean differences and Batcombe series had the least differences in mean soil moisture. Because of the very large dataset involved (174,700 entries) the horizontal bars denoting the 0.95 confidence interval are too close to the plotted points to show on Figure 5.5.

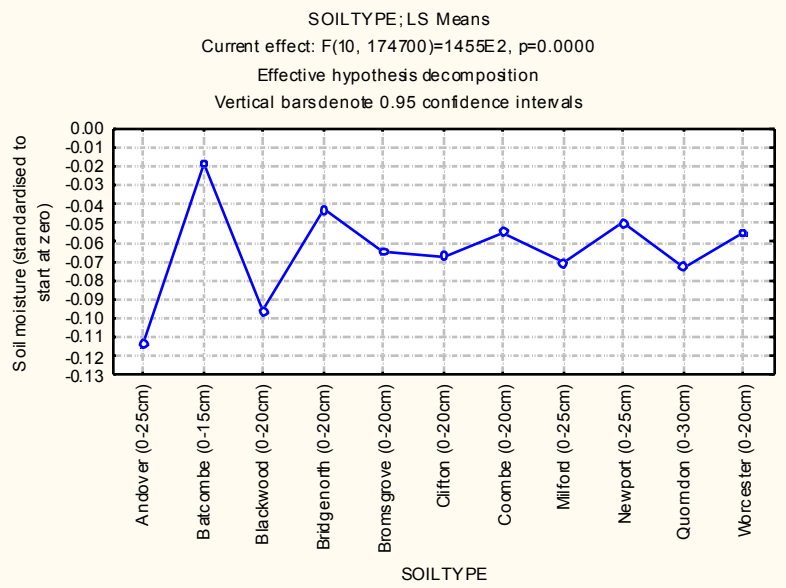


Figure 5.5. Mean changes in soil moisture predicted by MOSES for topsoils of test soil series

When the deep subsoil layers were considered all soil types were again statistically significantly different. In this case there were only 9 soil series represented as both the Andover and Bridgnorth series are developed over hard rock within 100 cm depth. The well drained, sandy Newport series gave the largest changes in mean values followed by Clifton series (Figure 5.6).

Both these soils were dominated by increases in soil moisture at this depth, whilst all other soils had mean values that indicated drying out to various degrees. It should be expected that drying would be dominant in a time series running from January to June.

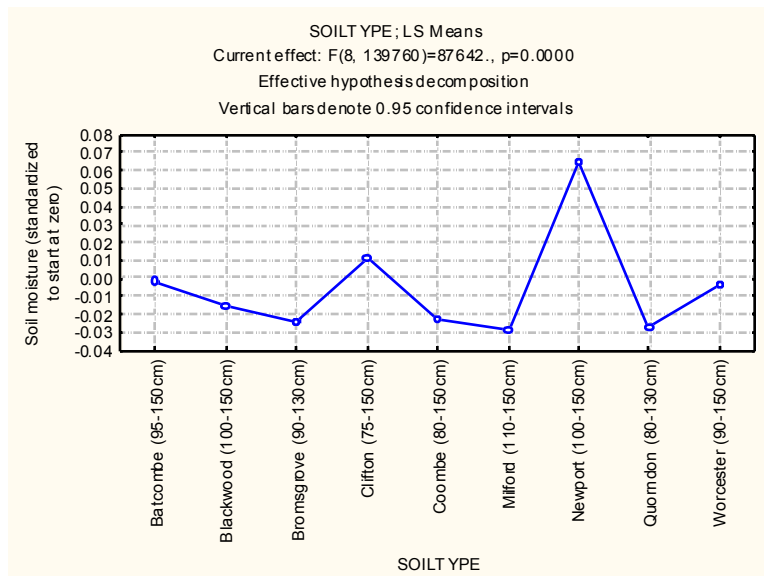


Figure 5.6. Mean changes in soil moisture predicted by MOSES for deep subsoils of test soil series

5.5 VALIDATION OF NSRI-MODIFIED MOSES OUTPUT

It was planned that MOSES output would be calibrated against measured data obtained from a range of sites across England and Wales. Three fully instrumented experimental sites were set up during the project to measure soil temperature and soil moisture at a range of depths. These sites were located alongside long-term soil temperature measuring stations. Only a limited number of sites were identified to be suitable for the experiment because of the need for: space within or adjacent to existing equipment; security for the equipment; a national spread of sites; and, finally, the facility to excavate, describe and sample the soil profile. UKMO sites at Camborne in Cornwall and Watnall in Nottinghamshire were chosen together with the long-term weather station at Rothamsted Experimental Station in Hertfordshire.

At each experimental site, six Delta-T soil probes were installed to monitor the temporal variation in soil moisture and temperature at 5, 20, 40, 60, 80 and 100 cm depth. In addition, a rain gauge was installed at each location. Detailed soil profile descriptions have been made and the soil profiles characterised by laboratory analysis for particle size class, density and soil water release characteristics (see Appendix 4). All sites are equipped with remote access facilities for downloading data using mobile phone technology. Results have been received from these sites from March 2004 to present.

Twelve further study sites were chosen, where soil temperatures are routinely measured and reported to UKMO. Here the soil profile adjacent to the soil temperature probes was carefully described with each horizon (soil layer) assessed and recorded for texture, stoniness and density. An overall assessment of the soil water regime of the site was also made. In this way information was collected about the physical soil properties of the materials in which soil temperatures were measured. This information could then be linked to the National Soil Map (NATMAP) to derive spatial extents of these materials. The results of these assessments are given in Appendix 4

5.6 CORRELATION OF RESULTS FROM NSRI SOIL TEMPERATURE PROBES AND LONG-TERM UKMO SOIL TEMPERATURE RECORDERS

Soil temperature probes were installed by NSRI at 3 meteorological stations as close as possible to existing long-term soil temperature reporting facilities. The existing stations

measured soil temperature at 10, 30 and 100 cm depths. The NSRI probes were 10 cm long and installed at the surface (0-10 cm depth), 20 (15-25 cm depth), 40 (35-45 cm), 60 (55-65 cm depth), 80 (75-85 cm depth) and 100 (95-105 cm depth). The differences in the measurement depths complicate the direct comparison of results. In some cases a difference of 10 cm in depth can mean that 'equivalent' probes were installed in different soil layers. A statistical analysis of the NSRI and UKMO soil temperature data is provided in Appendix 5, a short review of the results from Camborne are given below.

At Camborne results were initially compared for the NSRI 0-10 cm 'surface' temperature and the UKMO data for 10 cm depth (unfortunately at the boundary between the Ah layer at 0-10 cm depth and the Bw1 at 10-25 cm depth and hence incorporating data from both). A good correlation was achieved (Figure 5.7) with a correlation coefficient of 0.964.

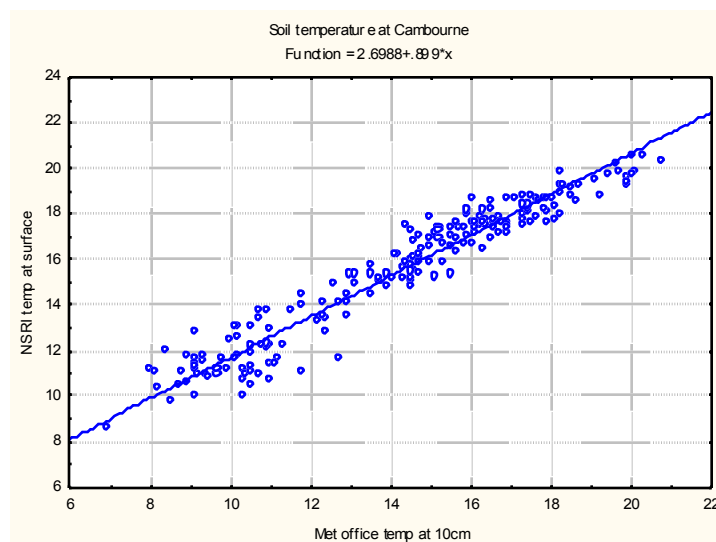


Figure 5.7. Comparison of soil temperatures at NSRI 'surface' and UKMO 10 cm depth

NSRI soil temperatures were slightly higher than UKMO values except for about 20 days, which occurred mainly when temperatures were above 17°C.

When the UKMO probe at 10 cm depth was compared with the NSRI probe at 20 cm depth (Figure 5.8), there was again a good correlation with a correlation coefficient of 0.958. Again NSRI temperatures are generally higher than UKMO values by up to 2°C but above 16°C there is an increasing trend for this relationship to be reversed.

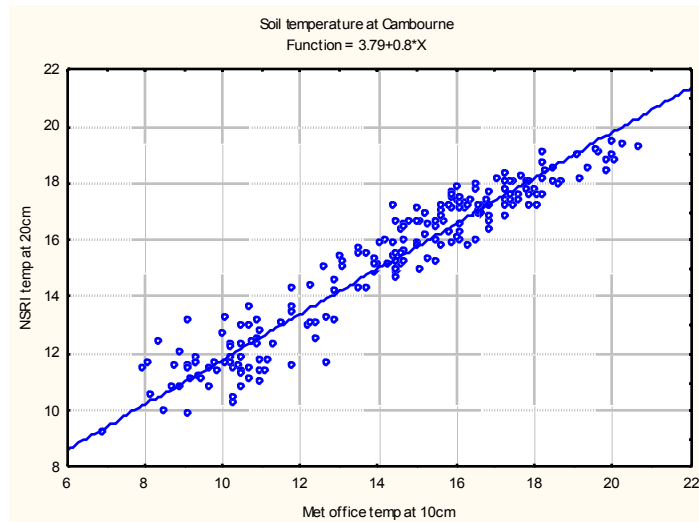


Figure 5.8. Comparison of soil temperatures at NSRI 20 cm and UKMO 10 cm depths

The comparison of NSRI 40 cm depth probe and the UKMO probe at 30 cm depth gave no significant difference between the two with a correlation coefficient of 0.991 (Figure 5.9).

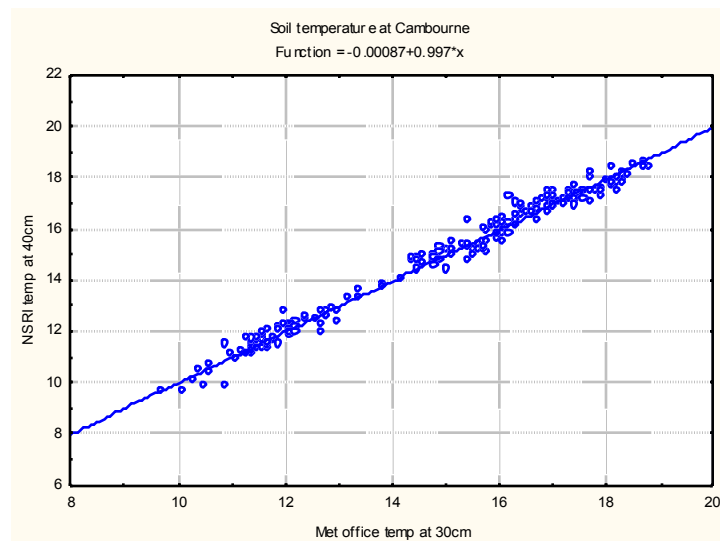


Figure 5.9. Comparison of soil temperatures at NSRI 40 cm and UKMO 30 cm depths

The comparison of temperatures at 100 cm depth gave surprising results. An overall correlation coefficient of 0.993 was achieved but there appears to be two relationships within the measurements; one when the soil is warming up and another when the soil is cooling (Figure 5.10). The change occurs on 11 August 2004. When the data are divided at this date to separate the warming phase from the cooling phase the correlation coefficient for the warming cycle is 0.998 and 0.999 for the cooling phase.

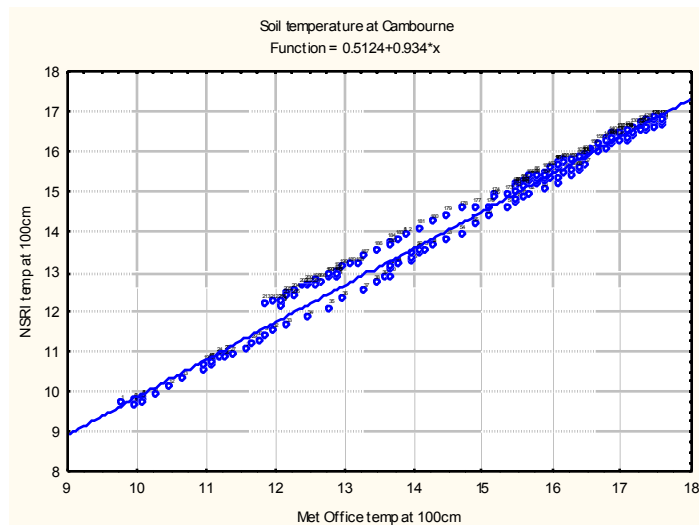


Figure 5.10. Comparison of soil temperatures at NSRI 100 cm and UKMO 100 cm depths

Overall there was a good correlation between soil temperatures measured by NSRI probes and the UKMO recorded data. Records from all 'equivalent' depths were statistically correlated and all correlation coefficients were above 0.90.

5.7 INTEGRATION OF NSRI-MODIFIED MOSES INTO THE INTERMET MODEL INTEGRATION FRAMEWORK (MIF)

To facilitate project development work at Silsoe a stripped down version of the Model Integration Framework (Mini-MIF) was provided by Mike Minter of UEGD for stand-alone work at Silsoe. Mini-MIF incorporated the functions that allow the extraction of necessary meteorological variables from the modelled surfaces within MIF that had been produced by other parts of the InterMet project (UEGD and CSL inputs) to run the soil moisture and soil temperature predictive model (MOSES).

The original project aim was to link the predictive soil model (NSRI-modified MOSES) directly to MIF to demonstrate the feasibility of producing a nation-wide surface of soil moisture and soil temperature. Soil properties had been verified *in-situ* at 12 UKMO soil temperature measurement sites spread across England and Wales (Appendix 4). These 12 sites should have formed a national validation network for reviewing the quality of soil moisture and soil temperature outputs from MIF via the MOSES model.

MIF was developed using FORTRAN programming, as a research tool at USGD, at an early stage of the InterMet project through NERC funding. Several meteorological data surfaces were developed at a range of grid sizes and installed on MIF (rainfall, temperature etc). Following the completion of the USGD input to InterMet, MIF and associated meteorological surfaces were transferred to CSL, York to be run alongside the large dataset of met data assembled for the InterMet project.

Since it has been housed at CSL, MIF has been updated and following that update meteorological surfaces can no longer be extracted to run on the NSRI version of Mini-MIF.

In order to facilitate a limited but rigorous appraisal of the soil moisture and soil temperature data predicted from the NSRI-modified MOSES model, three weather datasets, one for each of the three research sites (Camborne, Watnall and Rothamsted) were purchased from the

UKMO. These have been used as surrogates for weather surfaces that would be produced within MIF. The results of this appraisal are discussed below.

5.8 REVIEW OF PREDICTIONS OF SOIL MOISTURE AND SOIL TEMPERATURE FROM NSRI-MODIFIED MOSES AND MEASURED DATA AT CAMBORNE, WATNALL AND ROTHAMSTED

All weather variables that are required to run the NSRI-modified MOSES model were purchased from UKMO for each of the Camborne, Watnall and Rothamsted experimental sites in April 2005 to cover the period of soil moisture and soil temperature monitoring (March 2004-March 2005). On delivery these data, with a 15 minute time step, were transposed into 3 separate MOSES input files, one for each site according to the methods outlined in Appendix 1. The soil physical properties measured at the three sites were manipulated through pedo-transfer functions to provide specific soil input files for each site (Appendix 2).

NSRI-modified MOSES was then run for all three sites using the site specific weather and soil input data files. The soil moisture/temperature probes provide hourly records. For the purpose of this review one drying and wetting cycle was used with a subset of data taken from the on-site installation (30th March 2004 at Watnall, 2nd April at Rothamsted and 14th April at Camborne) up until 12th November 2004. A full analysis of the results is given in Appendix 6. A short review of the results for Camborne is given below.

Soil moisture measurements from the soil moisture probe at 5 cm depth (inserted at 0-10 cm depth) and the predicted values for 0-10 cm depth from MOSES are given in Figure 5.11 and Figure 5.12.

The soil data to run the MOSES model is generated from the analytical results of tin samples taken at 2-7 cm depth. Figure 5. plots the two datasets from 14th April to 11th November 2004. The measured soil moisture initially shows variation around field capacity moisture content (50 per cent volume) reflecting heavy rain immediately after installation and subsequent early spring rainfall events. There is rapid drying in early May as transpiration increases and wilting point (15 per cent volume) is achieved at several times in the summer.

There are two peaks in the histogram of measured data (Figure 5.12, right hand side) reflecting field capacity and wilting point. These peaks are not evident in the modelled data (Figure 5.12, top histogram). The modelled data fails to drain below about 40 per cent moisture content even though the marked variations in moisture content shown by the measured data are reflected in a subdued form in the modelled data.

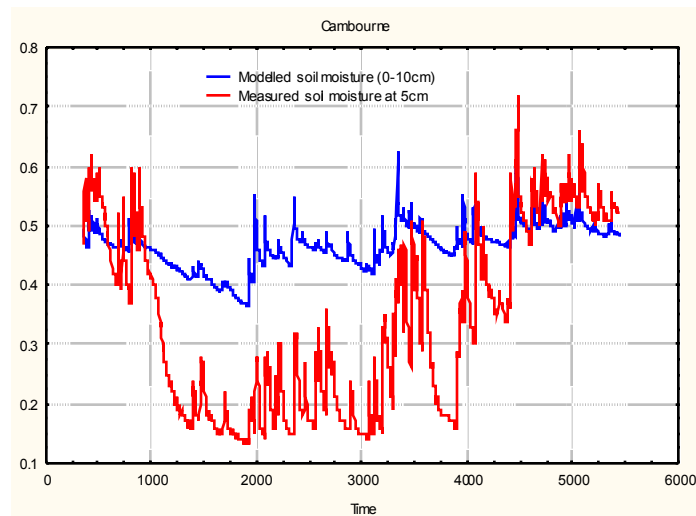


Figure 5.11. Measured and modelled volumetric soil moisture contents in Ah horizon (0-10 cm depth) at Camborne

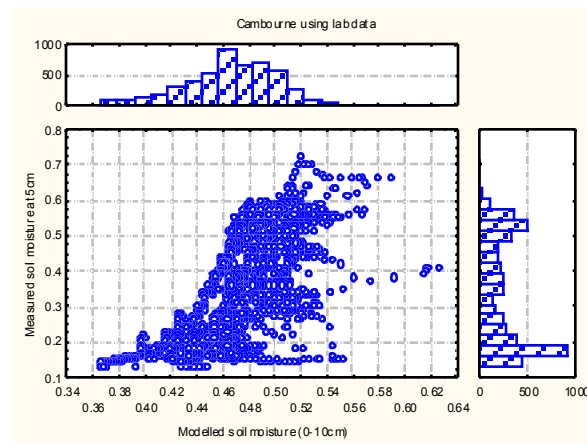


Figure 5.12. Measured vs modelled soil moisture contents in Ah horizon at Camborne

The next horizon, Bw1 at 10-25 cm depth, is represented by soil moisture predictions based on MOSES using soil information derived from tin samples taken at 11-16 cm depth to represent the whole horizon and soil moisture measurements at 20 cm depth (probe inserted at 15-25 cm depth). Figure 5.13 and Figure 5.14 demonstrate a better correlation between modelled and measured data, although the modelled data are again less responsive to changes in moisture content and fail to predict the short periods of saturation (up to 50-55 per cent volume) or the longer periods when the soils approach wilting point moisture contents (15-20 per cent volume) in late June and early August (around time step 2000 and 3000 respectively).

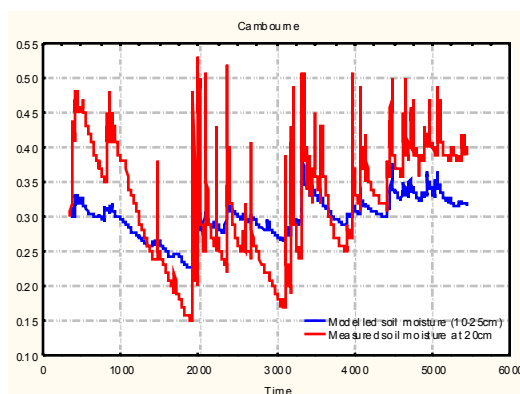


Figure 5.13. Measured and modelled volumetric soil moisture contents in Bw1 horizon (10-25 cm depth) at Camborne

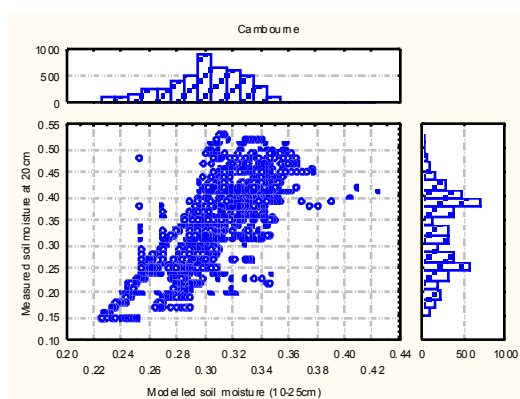


Figure 5.14. Measured vs modelled soil moisture contents in Bw horizon at Camborne

Figure 5.13 demonstrates a lack of sensitivity in the modelled data immediately after installation. Heavy rainfall in the days following installation increases moisture content by 50 per cent (30 per cent moisture content on installation to 45 per cent a few days later), while the modelled data only predicts a 20 per cent increase (30 to 36 per cent moisture content). This restricted response to rainfall is probably due to an in-built barrier within MOSES that artificially restricts the amount of rapid water movement down through coarse pores within the soil profile. Some programming adjustments are required to overcome this defect.

A different set of modelling errors are demonstrated by the examination of results from deep in the soil profile (Figure 5.15 and Figure 5.16). Here soil information to run MOSES is derived from a sample taken from the Cu horizon at the base of the profile at 78 to 83 cm, whilst measurements were obtained from a probe at 80 cm depth (75-85 cm)

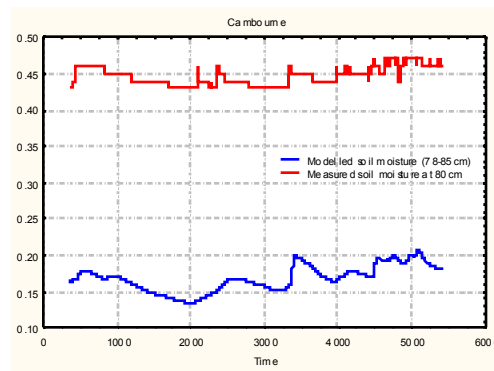


Figure 5.15. Measured and modelled volumetric soil moisture contents in Cu horizon (78-85 cm depth) at Camborne

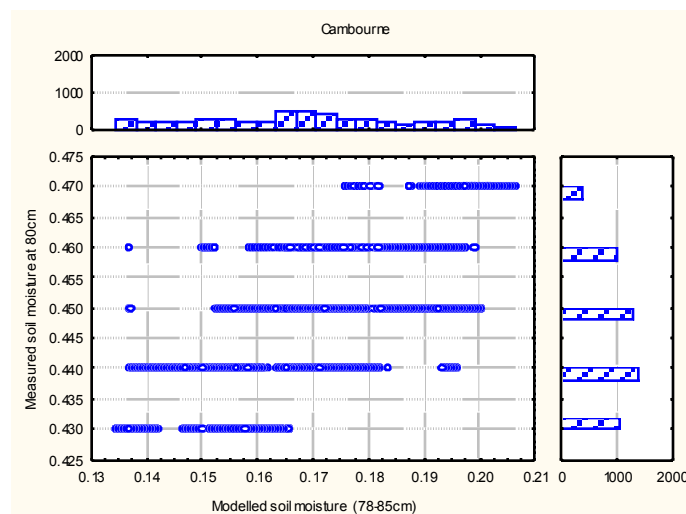


Figure 5.16. Measured vs modelled soil moisture contents in Cu horizon at Camborne

In both plots in Figure 5.15 the range of variation is small (4 per cent in the measured and 7 per cent in the modelled moisture contents). Furthermore the rises and falls in moisture content correlate well. However, there is marked discrepancy in the overall volumes of water predicted by the model and measured in the field. The general rule whereby all model runs were started at field capacity moisture content does not work well in these materials as it leads to a greater than 20 per cent discrepancy between the modelled and measured data throughout the observation period. This is probably for one of two reasons or maybe a mixture of both. The soil at this depth was described as very moist and extremely stony (more than 75 per cent stones and hence less than 25 per cent fine earth between the stones). It is very difficult to take small undisturbed sample tins in this type of material and large errors in the assessment of field capacity moisture content can be made. Allowance should be made for the water release characteristics of the stones in this layer. Furthermore it is reported that the shattered rock layers at the base of these soils in Cornwall can act as a conduit for groundwater, which is perched above the hard impermeable slate bedrock (Harrod, 1975). If groundwater is present in these soils the MOSES model does not at present take this source of water into consideration in its predictions. MOSES integrates rainfall as the sole water input to its moisture content calculations. Following further monitoring at this site, a best fit starting point should be estimated for the moisture content of this layer. From current data this is likely to be around 45 per cent. If this value had been

taken for starting the modelling the correlation with measured probe data would have been very good.

Short comings have been identified in the estimation of soil water by the MOSES model approach. Recommendations for overcoming these are given later in this report. However, estimates of soil temperature from MOSES are much better and relatively good correlations are achieved between measured and modelled temperature data (Figure 5.17 - Figure 5.19).

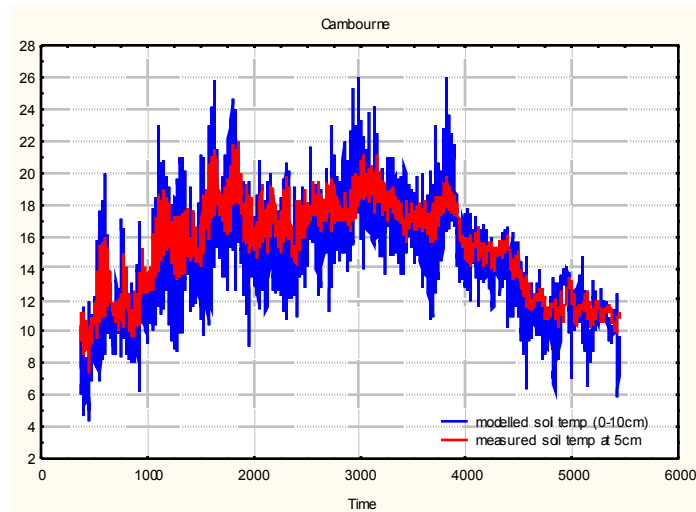


Figure 5.17. Measured and modelled soil temperatures in Ah horizon (0-10 cm depth) at Camborne

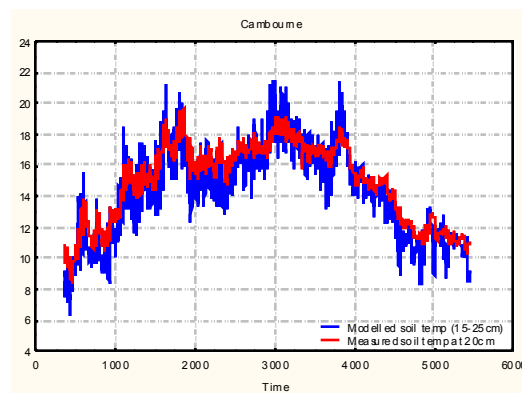


Figure 5.18. Measured and modelled soil temperatures in Bw horizon (10-25 cm depth) at Camborne

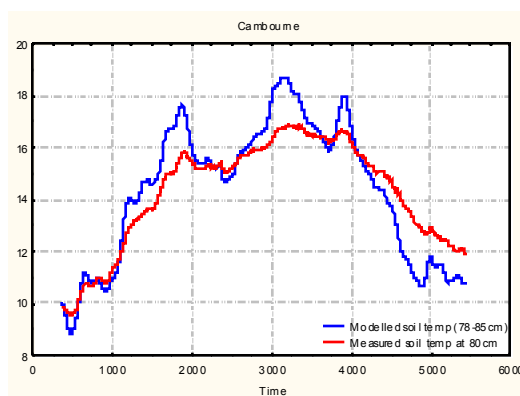


Figure 5.19. Measured and modelled soil temperatures in Cu horizon (78-85 cm depth) at Camborne

5.9 CONCLUSIONS

5.9.1 Soil moisture assessment

A key challenge when MOSES was selected as the preferred model was to ensure that its coding and/or input files were sufficiently flexible to allow the model to be used with the full range of soil physical properties that occur across England and Wales. A **sensitivity analysis** of the NSRI-modified MOSES model was carried out using twelve contrasting soils, which had between 1 and 5 input layers (soil horizons), a range of soil textures (loamy sand to clay) and hence a range of hydraulic properties. NSRI-modified MOSES was run for each of these soils using a standard 6-month weather dataset for Cardington, Bedfordshire. The sensitivity analysis showed that the model differentiated between all soils and the differences were statistically significant in all cases.

5.9.2 NSRI-modified MOSES model issues with soil moisture using site-specific weather and soil data

When running MOSES for subsoil layers there can be uncertainty as to the level of soil moisture with which to start the model running. In this review MOSES was consistently started at the field capacity moisture content relating to the soil layer under consideration. This generally worked well as the model runs started in late March or early April (within the field capacity period of the sites). However in very stony subsoil layers at Camborne there is some uncertainty about the level of field capacity moisture content. The values calculated from the analysis of tin samples from the site were considerably lower than the moisture levels measured on site by the moisture probes. The discrepancy resulted from the inherent variability often found in very stony samples. Soil specific protocols are required for assessing optimum soil moisture content for initiating MOSES model runs.

Overall soil moisture predictions from MOSES are less responsive to rainfall or drying than the measured values. During the field capacity period, data from the soil probes demonstrate clearly that there are large fluctuations in moisture levels associated with rainfall events when air-filled coarse pores fill with water only to drain by gravity over the 24-48 hour period following the rainfall event. The MOSES modelled data do not show the same level of variation. Adjustments to the internal programming of MOSES or to the input files will be required to more accurately reflect the different rates of water movement in coarse pores (where gravity can act) compared to fine pores.

Soil moisture predictions are better in well drained stoneless soils than is the case in stony soils affected by fluctuating groundwater. MOSES does not recognise the potential influence of groundwater sources augmenting the water content of some subsoil layers and additional programming, either to MOSES internal code or to input files is required to implement this.

5.9.3 Soil temperature issues

The initial *sensitivity analysis* of the NSRI-modified MOSES model indicated that statistically significant differences in soil temperature were being generated by the model when it was applied at a range of soil types based on weather data from Cardington.

Generally very good correlations were obtained between the existing *UKMO soil temperature probes* at Camborne, Watnall and Rothamsted and the *Delta-T probes* installed during this project. However, lower layers at all sites indicated different relationships between the two soil temperature measurements when the soil was on a warming curve as opposed to when cooling down. Further evidence gathered from visits to the network of 12 soil temperature validation sites, suggests that different methods of soil probe insertion may have been used over the last 50-100 years. In very stony or rocky subsoils the practice seems to have been to drill a larger hole than necessary, insert the temperature probe and then back-fill with sand to produce a close fit. The NSRI technique in very stony materials at Camborne was to drill as small a hole as possible, so that the probe is held as tight as possible from the outset and to only back-fill around the probe with material taken from the same layer. In this way the probe is taking measurements within material from the depth of interest rather than from within a column of sand which may act as a preferential conduit for heat when it is waterlogged and act as a thermal blanket when dry. Further research is needed into the methods of insertion of the deep (1 m) soil probes, especially where hard or shattered rock is likely.

5.9.4 NSRI-modified MOSES model issues with soil temperature using site-specific weather and soil data

Overall, predictions of soil temperature were better than those for soil moisture. Using Camborne as an example, the modelling of topsoil temperature variations is reasonably good and the bimodal distribution of measured temperatures is matched, albeit less strongly, in the modelled output. In subsurface layers the modelled soil temperature values show a stronger relationship with measured data than was apparent in the topsoil (see Appendix 6, p.13-15).

5.10 RECOMMENDATIONS

5.10.1 Model improvements

There are several parts of the NSRI-modified MOSES model that should be refined in order that the MOSES soil moisture predictions are improved. At present MOSES predictions are unresponsive to temporary increases in water content above field capacity. This occurs throughout the winter when coarse pores fill with water after rain and then drain under gravity within a few hours or at most 1-2 days. Some sandy soils can be made up of 20 per cent by volume of this size of pore ($>60\mu$). It is likely that some form of dual porosity process will have to be incorporated into MOSES, either as internal code or through the soil input files.

A robust mechanism is needed to identify the moisture content at which to start the model running, especially in subsoil layers. Further work is required on the effect of stones of varying type on soil moisture release characteristics. Pragmatic methodologies are required to provide optimum moisture contents with which to initialise the model runs.

The validation of the NSRI-modified MOSES model was carried out using site-specific soil property data which incorporated soil hydraulic properties measured for individual horizons described in detail from Camborne. These local data should also have been available for Rothamsted and Watnall. However, laboratory technical difficulties meant these data were

not available in time for the validation exercise and LandIS soil data were used to characterise the Bromsgrove and Batcombe soil series identified at Watnall and Rothamsted respectively. Further validation of the NSRI-modified MOSES model should be undertaken using the local soil property data when available.

5.10.2 Compatibility with MIF

The NSRI mini-MIF system (including the re-engineered MOSES program) must be made compatible with the new CSL project framework (MIF2) that has been developed since MIF moved to Sand Hutton. This requires similar re-programming that was undertaken when MIF was converted to MIF2.

Although the NSRI-modified MOSES model appears to be working well with the UKMO-supplied weather data for Camborne, Watnall and Rothamsted, it still has to be linked to weather surfaces generated by InterMet research and held in MIF2. A comparison of running NSRI-modified MOSES with UKMO-derived data and MIF2 surfaces is required and this comparison should be extended to incorporate a review of MOSES run on weather data generated by the new UKMO NIMROD system.

5.10.3 New weather surfaces

Before detailed comparisons can be drawn additional weather surfaces will have to be generated to cover the needs of MOSES for cloud cover. InterMet has some extensive runs of radar weather data which could be used for this purpose. Interpolated values for solar radiation will also be required.

5.10.4 Implementation trial area

The next stage of incorporating soil moisture and soil temperature into an operational system is to carry out a feasibility study for a trial area (100 x 100 km block) such as Kent as identified in the original project specification. This would enable NSRI to test the use of the 1:250,000 broad-scale national soil map as well as small areas of detailed 1:10,000 scale soil data. A system would require:

- Generation of soil information for MOSES soil input files on a grid basis – at 10, 5 and maybe 1 km intervals where detailed soil information is available.
- Generating land use data on similar grids to above.
- Reviewing the feasibility of letting operators select soils and land use from a menu of background soils and land uses in order to satisfy individual needs. The system would then be run using their selected site-specific data. There is an ever increasing need for more detailed information, and this would provide a bonus for land owners who have invested in collecting information about their land and soils.

Any operational systems should be built around and integrated with the MIF2 at CSL.

5.10.5 Validation requirements

Soil temperature and soil moisture predictions from MOSES require further rigorous validation. In terms of soil moisture, further research questions have arisen regarding how MOSES can be engineered to deal with the broad group of soils suffering at varying degrees from waterlogging by groundwater. Some form of blocking mechanism will have to be incorporated into the model for susceptible soils whereby through-flow is strictly limited at certain times of the year or maybe the blocking mechanism is progressively triggered by certain levels of antecedent rainfall. This research will require further soil moisture monitoring sites across soils exhibiting a range of groundwater wetness.

The network of twelve characterised soil temperature sites is sufficient to give a national spread but if a trial operational system is attempted a group of UKMO soil temperature sites, within the trial area, should have their soils characterised for future validation purposes.

If an operational system is developed and soil moisture or soil temperature data from MIF are used to run external models, such as some of the disease models held by CSL, then there will be a need to validate the results from these models in terms of the soil moisture and temperature inputs.

6. DISTANCE BETWEEN FIELDS AND WEATHER STATIONS

6.1 OUTLINE

This section examines two important considerations for the proposed InterMet service. More specifically, the quality of the service will depend greatly upon the location for which a client requests data:

- And the distance to the nearest network station that can supply the requested measurement (or derived, but not interpolated, value)
- The density of network stations that around the location for which an interpolated measurement is requested.

6.2 APPROACH

InterMet will deliver access to meteorological records; by providing direct measurements from the recording stations and also, estimates for unmeasured locations that are interpolated by reference to the recording network. To examine the distances likely between unmeasured sites and the nearest recording station, the range of distances between randomly chosen fields and the nearest weather station were calculated. Field locations were provided by the 419 samples collected by the annual wheat disease survey (www.croponitor.co.uk) of 2000. The survey stratifies sampling according to the area of wheat within each region, so that the number of samples with each region depends on the area of wheat grown. The recording network comprised all of the weather stations supplied for the InterMet project.

6.3 RAINFALL

In the InterMet database daily rainfall totals were measured from 1 January 1998 to 31 December 2001 at about 2300 stations across England and Wales. To coincide with the wheat survey data, only stations that had a continuous daily record of rainfall between 20 April and 20 June 2000 were used (Table 6.1).

Table 6.1 - Shortest distance between rainfall recording stations and annual disease survey sample fields for different regions

Distance Num. & Percentage	Distance range (0.1 Km to 22.5 Km)				
	Within 3 km	3 – 6 km	6 – 10 km	>10 km	Total
East	40	45	18	0	103
Midlands	31	64	25	2	122
North	13	27	23	18	81
South East	14	17	5	0	36
South West	21	23	2	0	46
Wales	7	18	6	0	31
England and Wales	126	194	79	20	419
Percentage of total	30.1	46.3	18.8	4.8	100

The national situation reveals that nearly 90% of the survey locations were within 9 km of a rainfall monitoring station (Figure 6.1).

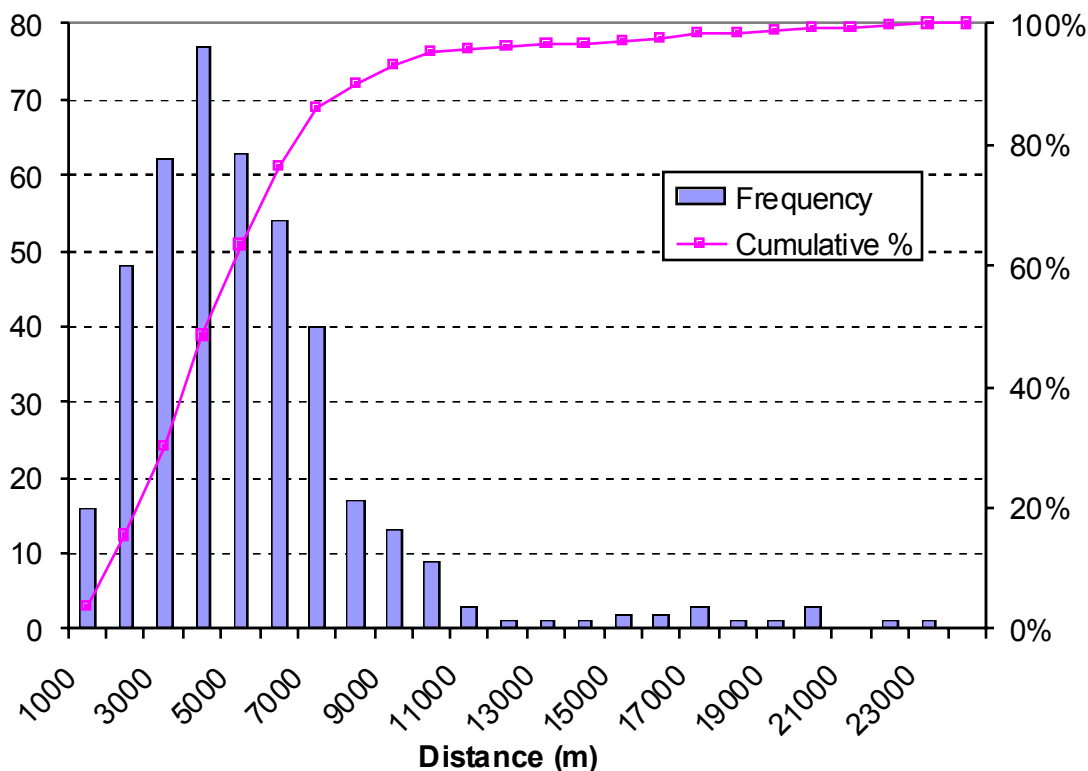


Figure 6.1 - Histogram of the number of weather stations recording rainfall in increasing distance groups across England and Wales

6.4 TEMPERATURE

Temperature was measured over the same period as rainfall at about 450 stations across England and Wales. Of these about 50 stations had infrequent records and more than 230 stations measured the temperature only once a day, only approximately 150 stations had a continuous hourly record over the four years. For the purposes of calculating distances to the nearest weather station, only those which had at least eight hourly observations on each day between 1 November and 30 November 2000 were included (Table 6.2). Nearly half of the sites were within 15 km of the nearest recording station and most were within 30 km.

Table 6.2 - Shortest distance between temperature recording stations and annual disease survey sample fields for different regions

Distance Num. & Percentage	Distance range (0.9 Km to 40.6 Km)			
	Within 15 km	15 – 30 km	30 – 40 km	Total
East	48	50	5	103
Midlands	48	66	8	122
North	41	35	5	81
South East	19	16	1	36
South West	20	24	2	46
Wales	18	12	1	31
England and Wales	194	203	22	419
Percentage of total	46.3	48.4	5.3	100

6.5 IDEAL LOCATIONS FOR RECORDING STATIONS

The current network of weather stations available to the ArableDS decision support system comprises around 100 locations from across England and Wales (a further 30 stations lie in Scotland and Northern Ireland). The stations are not selected with any reference to the major wheat growing locations, so many have low potential influence that does not justify the cost of supplying their data from the UKMO. Using the 107 registered users of ArableDS in 2004 an attempt was made to find the ideal locations for weather stations that would minimise the distance to DSS users while requiring as few stations as possible.

Over 80% of the ArableDS users were within 40 Km of their nearest UKMO station of those supplied to support the DSS. Using this distance as the ideal maximum, buffer zones of 40 Km were defined (Figure 6.2) around the ArableDS users and scaled to between one and zero (from most to least favourable).

The same procedure was used for the locations of the weather stations currently supplied to ArableDS by the UKMO therefore using the existing station locations as a starting point for the potential optimised network.

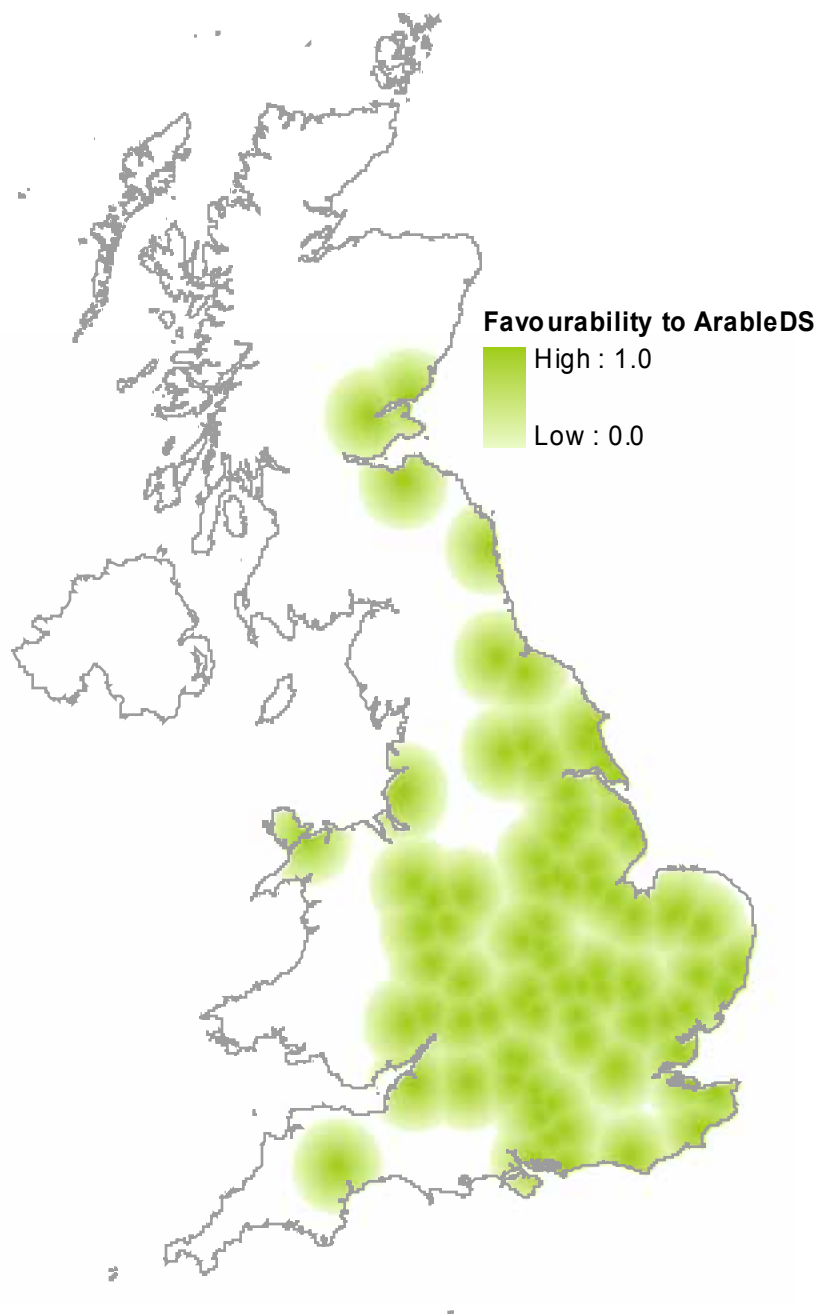


Figure 6.2 – Favourability of locations centred on existing users of ArableDS in 2004, up to a maximum range of 40Km

Multiplying the two images together produces a map which highlights the optimal locations for weather stations based simply on proximity to existing DSS users and their current sources of weather data (Figure 6.3).

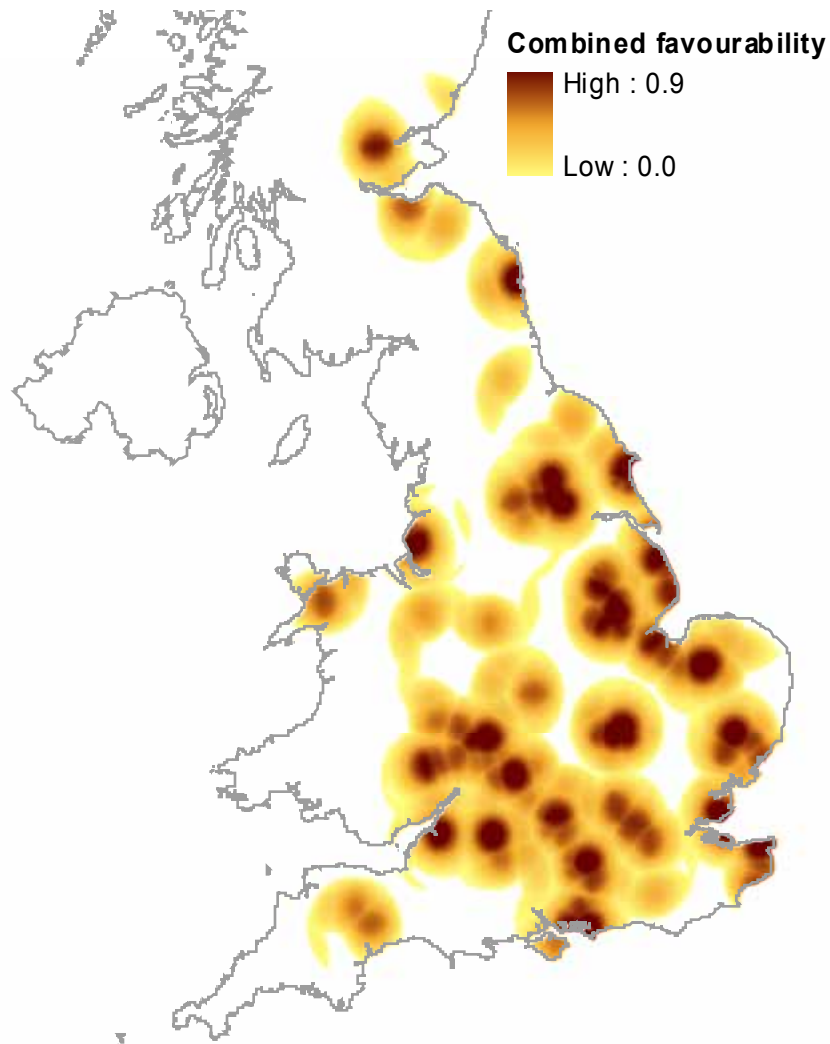


Figure 6.3 Position of optimized locations based on the combined parameters of distance to existing ArableDS users and UKMO weather sources

Using the darkest areas as a guide a total of 31 weather stations were located across England and Wales (Figure 6.4).

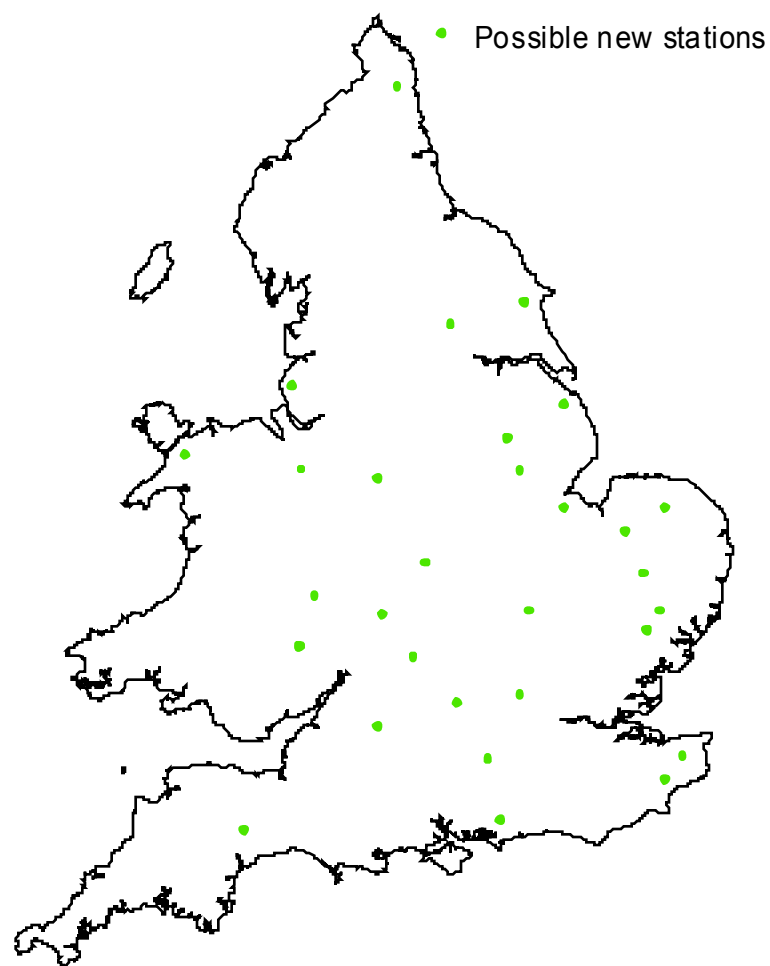


Figure 6.4. Possible locations for weather station network to independently support the registered users of ArableDS

The distance between the registered ArableDS users and the positions of these new stations was calculated. The more targeted network uses 70 fewer stations than the UKMO network but puts more users closer to a source of weather information (Figure 6.5). As the number of registered users of ArableDS expanded, the supporting network of weather stations could be periodically re-evaluated to optimise their locations.

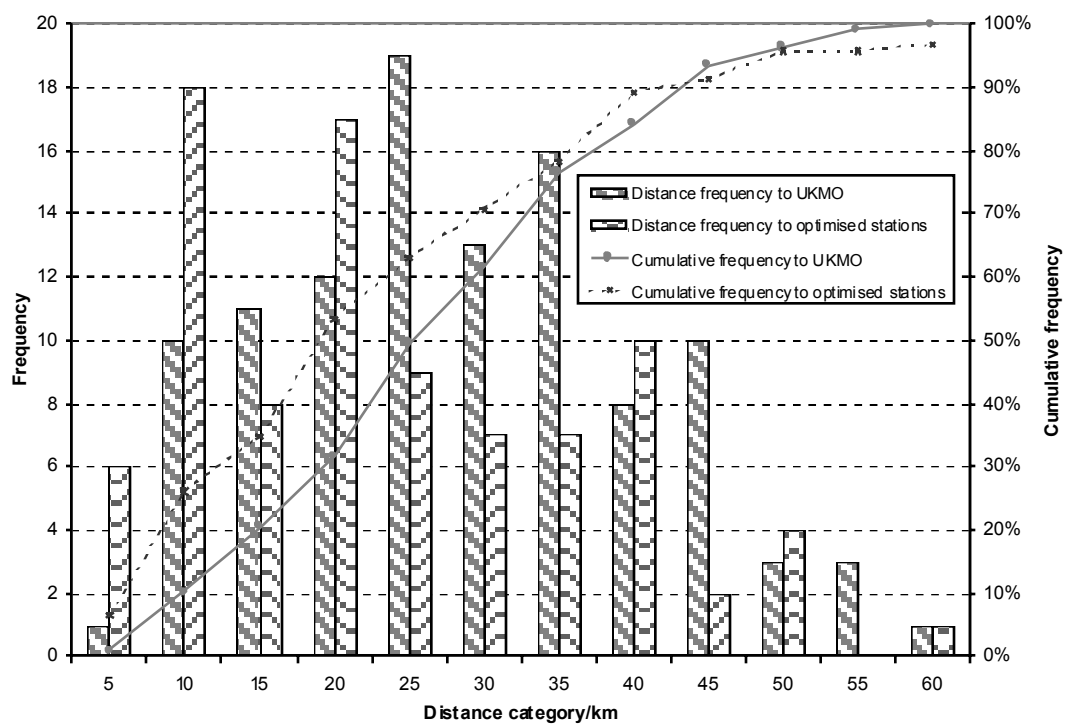


Figure 6.5. Summary of distances between registered ArableDS users and their UKMO stations compared with a more targeted network

7. THE USE OF INTERPOLATED VARIABLES IN DECISION SUPPORT

7.1 OUTLINE

Weather data are a crucial requirement for most Decision Support Systems to aid crop management. Most of these have been developed with the expectation of using local records from existing meteorological networks, or from independent, farm specific recording stations. The impacts from using interpolated data are generally untested.

This section examines how the use of interpolated meteorological variables can be used to inform decision support systems. The work evaluates DS tools at three levels of complexity:

- o formal systems, Smith Periods and BLITECAST, designed to measure risk from potato blight
- o a simple accumulated temperature model, typically applied, for example, to estimate the harvest dates for produce
- o national summaries of long-term regional risks, in this case to wheat lodging

7.2 INTERPOLATED VARIABLES FOR PREDICTING POTATO BLIGHT RISK

Many potato growers are encouraged to use decision support (DS) tools when scheduling fungicide applications to control late blight. Such tools may save unnecessary sprays in some seasons and they can help to justify fungicide inputs to show adherence to a production protocol. Late blight infection is strongly dependant on the weather, so DS tools need a source of meteorological data on which to base their advice. Farm based weather stations have become widely available in recent years, but it is impracticable to establish and maintain one in every potato field. Alternatively, observations from an existing, professionally maintained, network could be used for blight warnings, but these will be at varying distances from potato crops. This study aimed to compare how blight warnings based on the latter type of network might change if the DS results were spatially interpolated, rather than simply taking the DS result from the nearest weather station.

7.2.1 Approach

Hourly weather data from over 300 observing stations was obtained from the UK Meteorological Office (UKMO). Only those 220 stations that had a near complete set of hourly readings between 1998 and 2002 were used (Figure 7.1). The temperature and relative humidity (RH) data from these locations were used to find the date of the first Smith Period (Smith, 1956) in each of the five years.

An additional data set detailing the date and location of initial observations of late blight infection in fields across England & Wales was obtained for each of the five years. This was used to produce five maps of apparent infection date using a spline based spatial interpolation method (Hutchinson, 1995), as an example Figure 7.2 is the map for 2002.

Using the date of the first Smith Period results from the weather station data two further maps can be produced, each of the forecasted outbreak date. One is produced using a similar method of spatial interpolation as the actual outbreak map. The spline method was preferred as it makes no assumptions about the distribution of the input data which was usually skewed. The second map was produced by allocating every point the same value as the nearest weather station. This was effectively a simple method of interpolation known as Thiessen Polygons where every location within each polygon has the same value (Figure 7.3). In both cases final maps were constrained by the potato cropping area (Figure 7.). Each of these maps can be subtracted from the actual outbreak map to give a third map which in each case shows the size of the warning, in days, that the Smith Period forecasting scheme gave for that year. As a final step each warning map was reclassified into three

categories; warning too late (<10 days), an ideal warning (10 – 21 days) and a warning which was too early (>21 days, Figure 7.5). Using the potato hectare map (Figure 7.) the area of potatoes within each warning category can be calculated and compared with that produced by the opposite method over the five year period (Figure 7.6).

To investigate the influence of interpolation over a longer period the weather station data was used to calculate a season long (1 June to 30 September) total of severity units defined by the BLITECAST forecasting scheme (Krause *et al.*, 1975). The average temperature during a period when the relative humidity rises above 90% defines five classes (0 – 4) of increasing favour for late blight infection (Table 7.). Once again a pair of maps for each year were generated, one based on polygons and the other using spline interpolation. The area of potatoes within five severity total size groups (<25, 25 – 50, 50 – 75, 75 – 100, 100 – 125) was calculated and compared for years 1998 -2002 (Figure 7.7).



Figure 7.1. Location of UKMO weather stations with near complete hourly records between 1998 and 2002

Late Blight Outbreak Date 2002

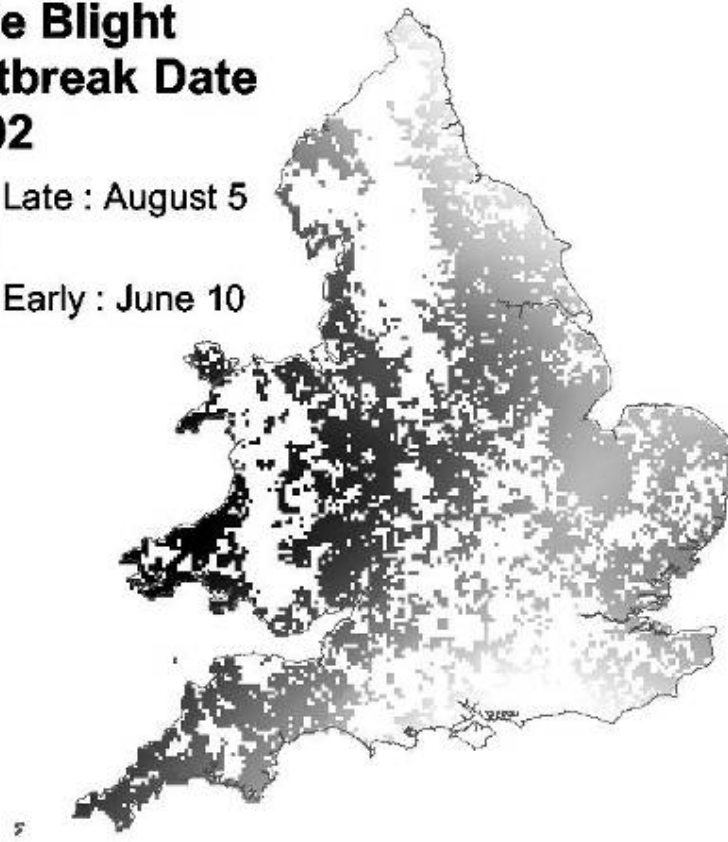
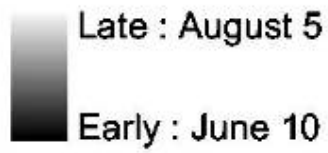


Figure 7.2. Example of late blight infection date generated from about 20 field reports and smoothed with the thin plate spline spatial interpolation technique.

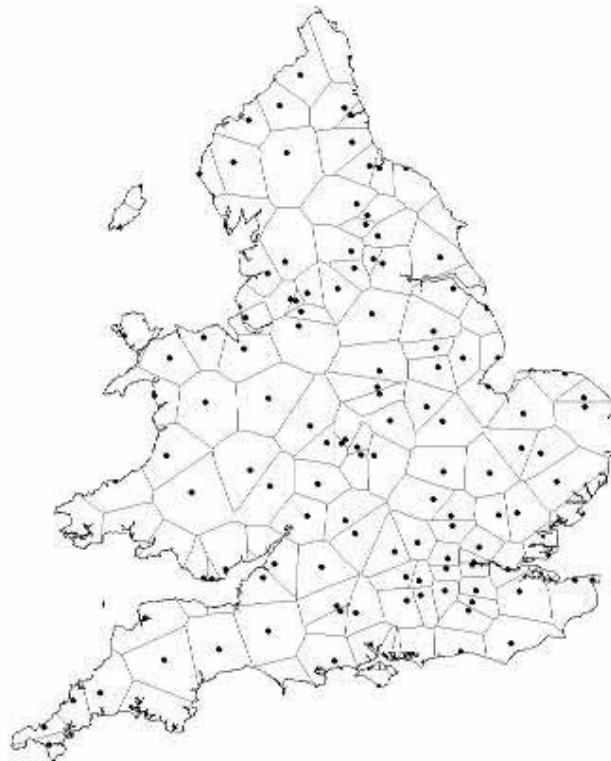


Figure 7.3. Allocation polygons of all areas to the value of the nearest weather station.

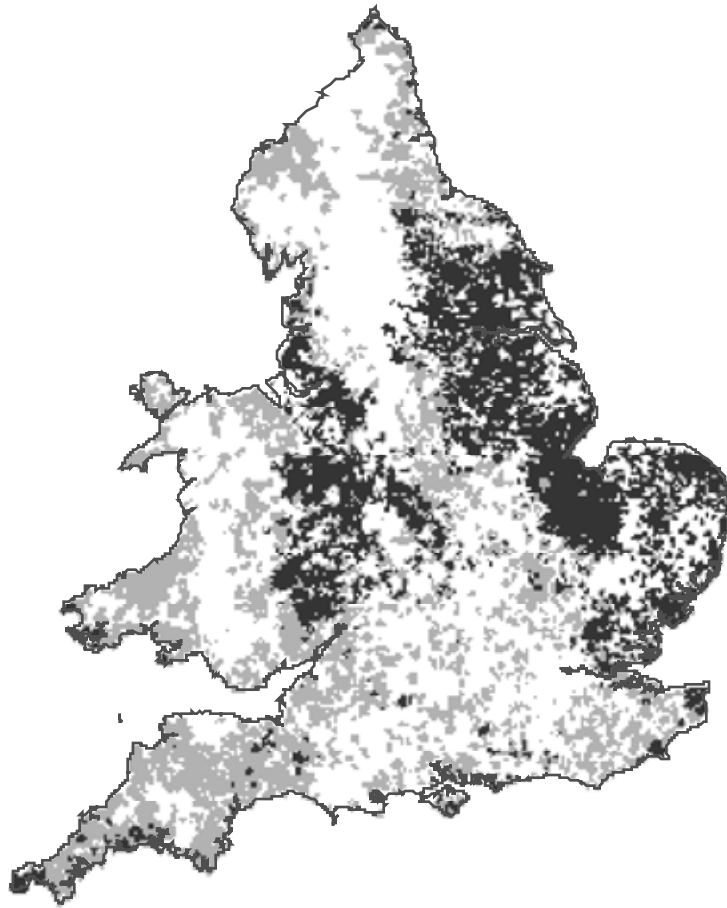


Figure 7.4. Potato growing areas in England & Wales, 1997. Grid is defined by 2km x 2km squares, light shaded areas have <5 hectares of potatoes in 4 square km, dark areas have >5 hectares in 4 square km.

Smith Period Infection Warning

- Too Late (<10 days)
- Ideal (10 - 21 days)
- Too Early (>21 days)

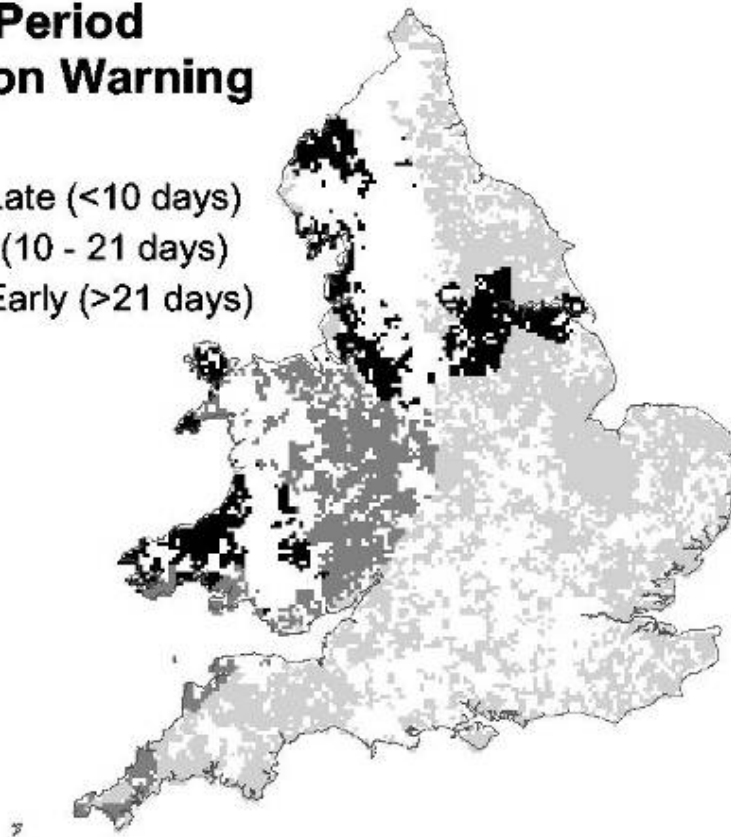


Figure 7.5. Infection warning categories based on actual outbreak date minus forecast outbreak date and reclassified.

7.2.2 Date of first Smith Period

There were no statistically significant changes in the area of potatoes within each warning category when spatial interpolation was used (Figure 7.6). However, the year to year variation was large resulting in a wide standard error range, the tendency was for spatial interpolation to give fewer early warnings but more later warnings.

7.2.3 Seasonal total of blitecast severity units

Spatial interpolation caused a statistically significant shift from the first severity total category (<25 or lowest risk) into the second category (25 – 50). This meant that the interpolation resulted in more infection warnings under the BLITECAST scheme and therefore probably more recommendations to apply fungicides (Figure 7.7).

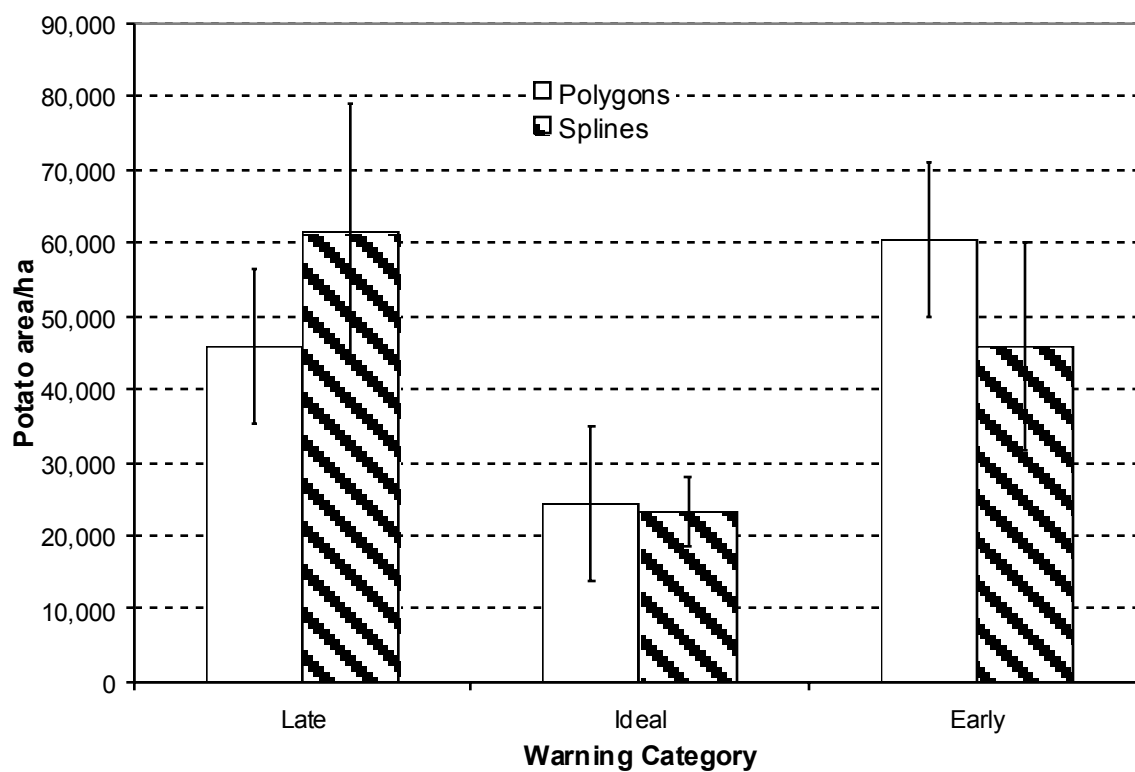


Figure 7.6. Effect of spatial interpolation on the areas of potatoes falling within three warning categories based on the Smith Period forecasting scheme, 1998 – 2002

Table 7.1. BLITECAST severity units (0 – 4) defined by mean temperature during high humidity periods.

Mean temperature range/°C	Severity Class				
	Hours ≥90% Relative humidity				
	0	1	2	3	4
7.2-11.6	15	16 – 18	19 – 21	22 – 24	25+
11.7-15.0	12	13 – 15	16 – 18	19 – 21	22+
15.1-26.6	9	10 – 12	13 – 15	16 – 18	19+

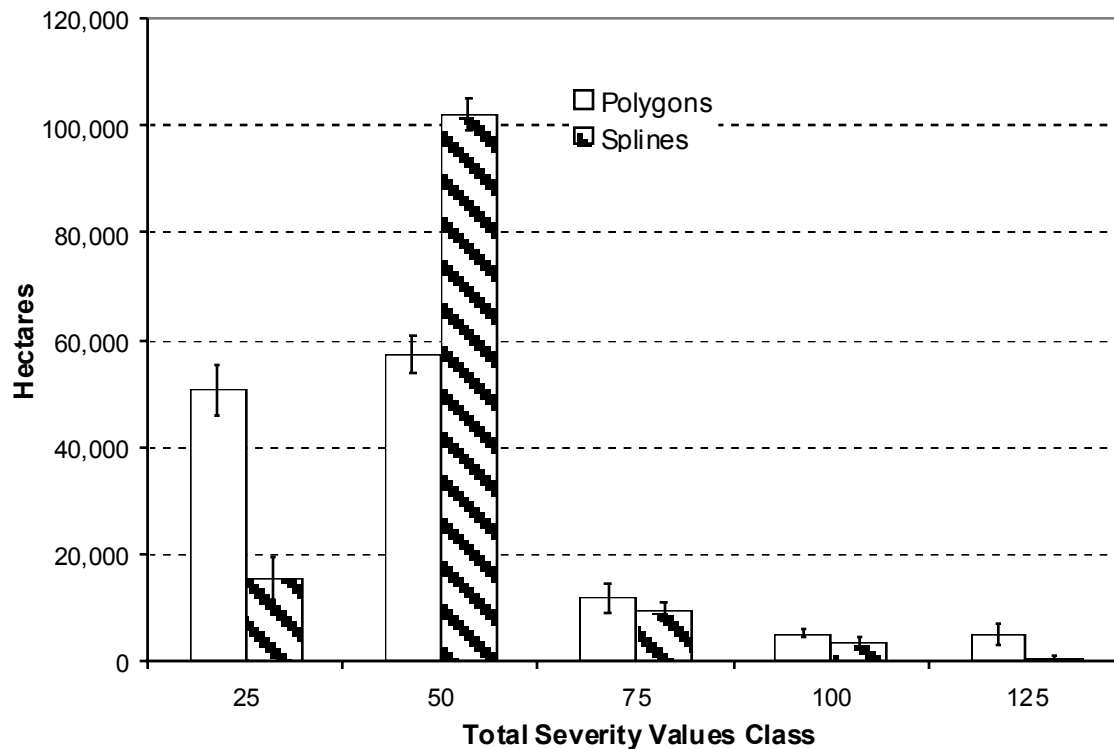


Figure 7.7. Effect of interpolation on the seasonal total of BLITECAST severity units, 1998 – 2002.

7.2.4 Impact of interpolation on risk warnings

The result for the comparison on the date of the initial Smith Period was confounded by a poorly performing spatial interpolation in the majority of the five years. Although the average differences between the 'Too Early' and 'Too Late' warning categories was large, so was the annual variation. The BLITECAST severity unit data was more amenable to interpolation. This was probably because it was derived from a longer series of readings (122 days) at each weather station rather than the initial Smith Period date which was a single event. However, it was more instructive to compare these two different types of derived variable.

The result for the seasonal total of BLITECAST severity units had a much lower amount of annual variability and the spline method of spatial interpolation worked more consistently. The interpolation will honour the input data values at the known locations but acts like stretching a rubber sheet between points so producing a smoother output than the simple polygons. The net result is a move towards the average value of the input data, in Figure 7.7 this is seen as the greater normality in the distribution of the spline interpolated data in comparison to the polygon data.

Spatial interpolation is a complex field in terms of the mathematics that can be employed on the problem and absolute accuracy can be hard to define. The weather data available to this project was used to derive two abstract variables related to late blight infection risk. In both cases the non-normal distribution of the input data restricted the choice of interpolation method, in particular it would have been inappropriate to use the geostatistical technique of Kriging (Webster, 1996). However, the raw weather data could be analysed by geostatistical methods even though it would considerably lengthen the processing time required. A mean temperature and an 'hours above 90% RH' surface at 4 km² resolution would need to be generated for each day of each season and the two forecasting schemes run for every pixel

of the potato growing area – 20,244 cells each day. Such an approach may lead to a more consistent interpolation performance and a clearer answer to what difference spatial interpolation, rather than simple allocation, could make.

The low frequency of confirmed reports on the location of late blight infection during the five years for which weather data was available was a weak spot for these analyses. There was no network in place to collect rigorous information on the appearance of the disease across England and Wales. The estimates of pathogen infection date were based on 20 – 25 reports each year some of which were quite vaguely located. In contrast a scouting network instigated by the British Potato Council for the season in 2003 resulted in 104 positive identifications of late blight with excellent location information. Unfortunately, the weather data available to the project does not stretch far enough into 2003 to be able to use the disease outbreak data.

7.3 TEMPERATURE SUMS

In 1735 Réaumur suggested that the biological development of an organism proceeds only when a certain number of heat units have been accumulated. The general veracity of that observation has been demonstrated for many biological processes in the intervening years. As a consequence, simple models of temperature accumulation are often proposed as the basis for a wide range of crop management decisions. Almost certainly at current, this type of simple temperature sum algorithm is the most commonly used type of formalized model in crop management. As a consequence, it is important to understand the impacts of interpolation methodologies on the estimates of temperature sums.

7.3.1 Approach

InterMet software (Jarvis 2001) is structured to allow multi-temporal environmental models to be closely coupled with the input grids of weather data. As a simple example, this combined interpolation and modelling approach was used to link interpolated daily maximum and minimum temperatures with an accumulated temperature model (Anonymous, 1969). This simple measure of the aggregate warmth or coolness of areas has been the starting point for many climatological and biological modelling studies.

The particular method of interpolation used in this example was a partial thin-plate spline function (Hutchinson, 1991), which incorporated factors influencing topoclimate such as the elevation, distance from coast and degree of urban coverage, at a resolution of 1 km. From the values interpolated at each 1 km cell of the landscape the accumulated temperature, or number of degrees by which the maximum value exceeds a base temperature of 8.5°, was calculated for each day of the year 1976. The sum of these exceedances throughout the year is known as the annual degree-day total. Figure 7.8 illustrates the geographical and temporal differences in this model result over the national extent of England and Wales. A strong north-south gradient is evident in the values for both years shown, whereas the hot summer of 1976 resulted in much larger totals for accumulated degree-days than the cooler year of 1986.

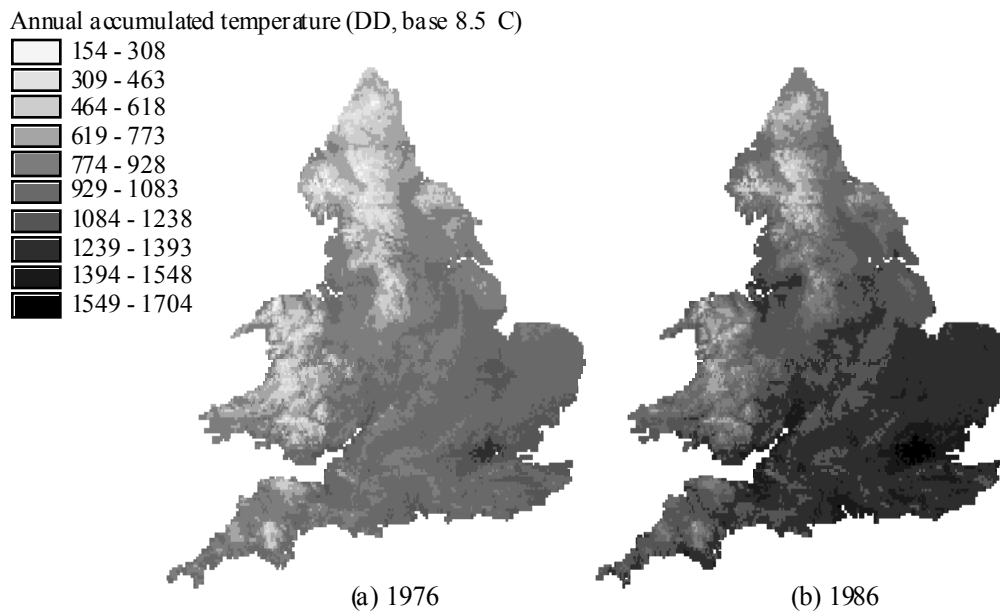


Figure 7.8. Estimated accumulation of temperatures (°C) using UK Meteorological Office method of degree-days (Anon, 1969) over a base temperature of 8.5°C in 1976 and 1986, England and Wales

Importantly in the context of the argument for wider incorporation of spatially referenced meteorological data in agricultural DSS, Figure 7.9 shows the results of performing a cross-validated assessment of the error in the daily maps of degree days that were used in the calculation of the annual totals of Figure 7.8. Error here is reported as the root-mean-square error (RMS), first summed by day, then totalled for all days in the year. Figure 7.9 illustrates the differences in accuracy that result from using two different types of interpolator - the first is the simple Voronoi/Theissen Polygon method of partitioning space, where the value estimated at any location is simply the value at the nearest observation point. This mimics the process of using data from the nearest meteorological station, when trying to extend estimation to other locations. The second method is the more sophisticated spline method. For a range of base temperatures used in the degree-day model computation, Figure 7.9 shows that the spline method produces errors approximately three times smaller than the voronoi method. In some cases, using data from the nearest station only will lead to RMS errors of over 200 degree days, which referred to Figure 7.8 is more than 10% of the estimated annual totals.

Whilst a national scope is necessary for certain forms of risk assessment, the InterMet technology can also be used to explore the geography of daily agro-meteorological conditions for specific regions. Figure 7.10 shows a detailed analysis of degree-day accumulations within the Vale of York over a year. The rapid changes in topography across an extent of 50 km leads to a wide range in the degree-days accumulated in different locations, with the values calculated for valley floor locations roughly twice those computed on the nearby moors. For tactical crop management decisions, for example about the control of pests, farmers also need to know how the conditions favouring risk alter from day to day. Figure 7.11 illustrates, for the same area of Yorkshire, differences in the day in the year (Julian Date) at which a certain number of degree-days have been accumulated. Again, marked geographical differences in these dates occur, even within this relatively small region; the warmer valley floor areas reach the threshold value after about 160 days (early June), whereas some areas up on the moors do not reach the same threshold before 190 days (early July). A knowledge of these geographical differences could be used to give

advice to landowners on when to apply control measures much more locally specific, helping to reduce ineffective and unnecessary applications

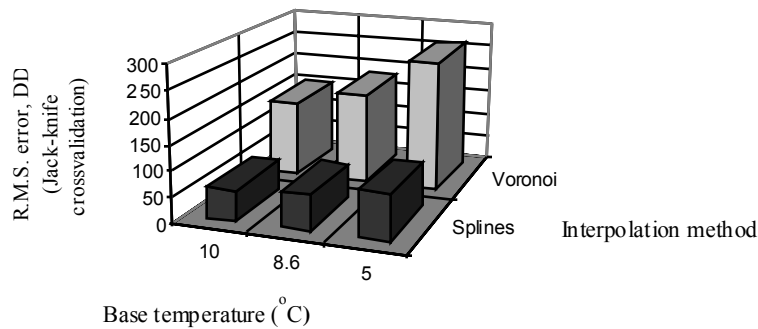


Figure 7.9. Residual Mean Square Errors (Degree days,) in grids of accumulated temperatures computed using Voronoi and partial thin plate spline methods of interpolation. Degree days computed for calendar year 1976 over base temperatures of 5°, 8.5° and 10°C respect

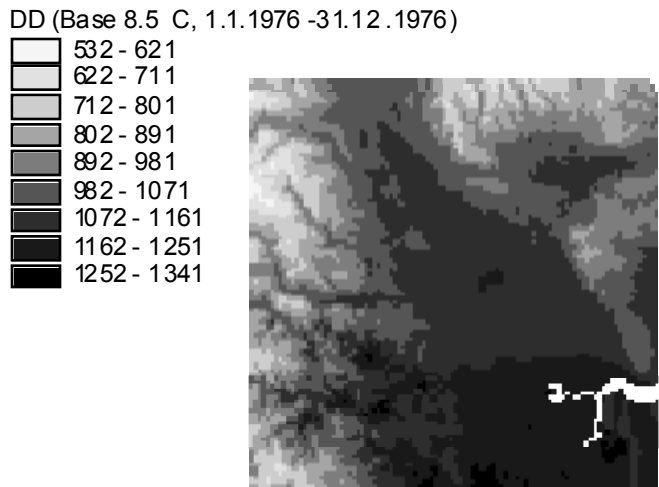


Figure 7.10. Estimated accumulation of temperatures (°C) using UK Meteorological Office method of degree-days (Anon, 1969) over a base temperature of 8.5°C in 1976, Vale of York.

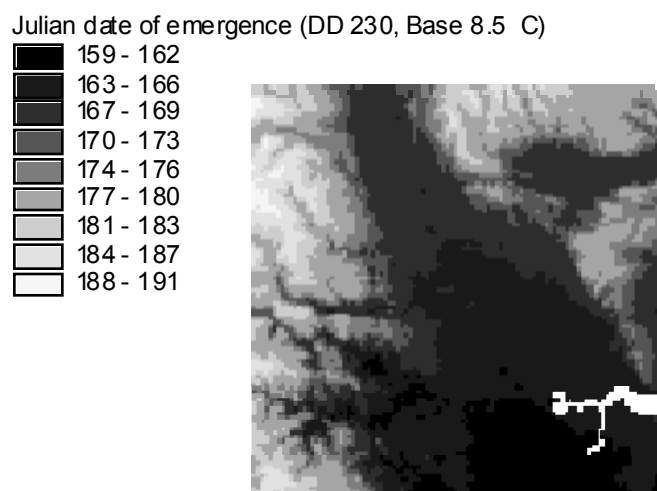


Figure 7.11. Geographical variation in the date by which the 230 accumulated Day-Degree threshold (Base 8.5°C) estimated to be reached throughout the Vale of York, 1976

7.4 LODGING RISK

Previous work has shown that the risk of lodging² in wheat crops is related to four criteria based on rainfall and wind speed all recorded during June, July and August.

The rainfall criteria are:

- a) The number of days which experience more than 7 mm of precipitation
- b) The total amount of rainfall recorded for the three months

The wind speed criteria are:

- c) Maximum gust recorded during the three months
- d) Average wind speed during the same time interval

Using the InterMet database of daily rainfall (over 2500 locations across the UK) and wind speed observations (over 500 UK locations) continuous surfaces at 2Km resolution were created. The greater availability of rainfall data enabled the use of the universal cokriging interpolation method with station elevation. The wind speed data was interpolated using a radial basis function, which gave more accurate results for these criteria (Figure 7.12). As four years were available (1998 – 2001) the results for each criteria were averaged and normalised so that the final images had values close to zero for low lodging risk to nearer one for higher lodging risk.

² Avoiding Lodging in Winter Wheat - practical guidelines. Home Grown Cereals Authority 2005

**Max wind speed kph
1 June - 31 August
Averaged 1998 - 2001**

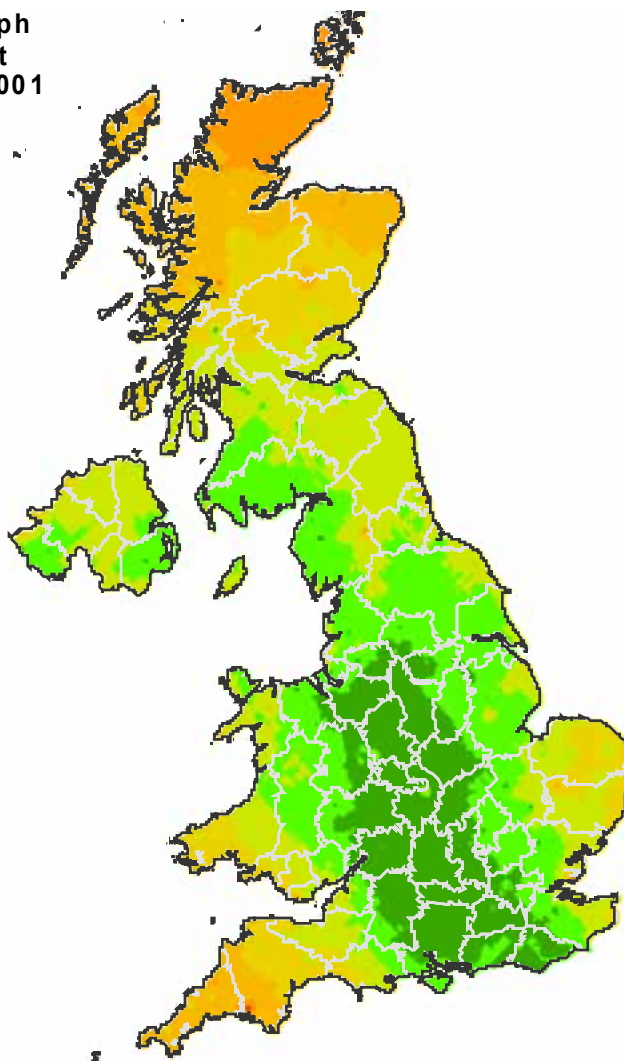
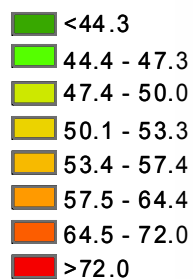


Figure 7.12. Lodging risk based on maximum wind speed gust during June, July and August (criteria c) before normalisation.

Each of the scaled images for the wind speed and rainfall criteria were combined to create two final images of lodging risk – one based on rainfall and the other on wind speed (Figure 7.13).

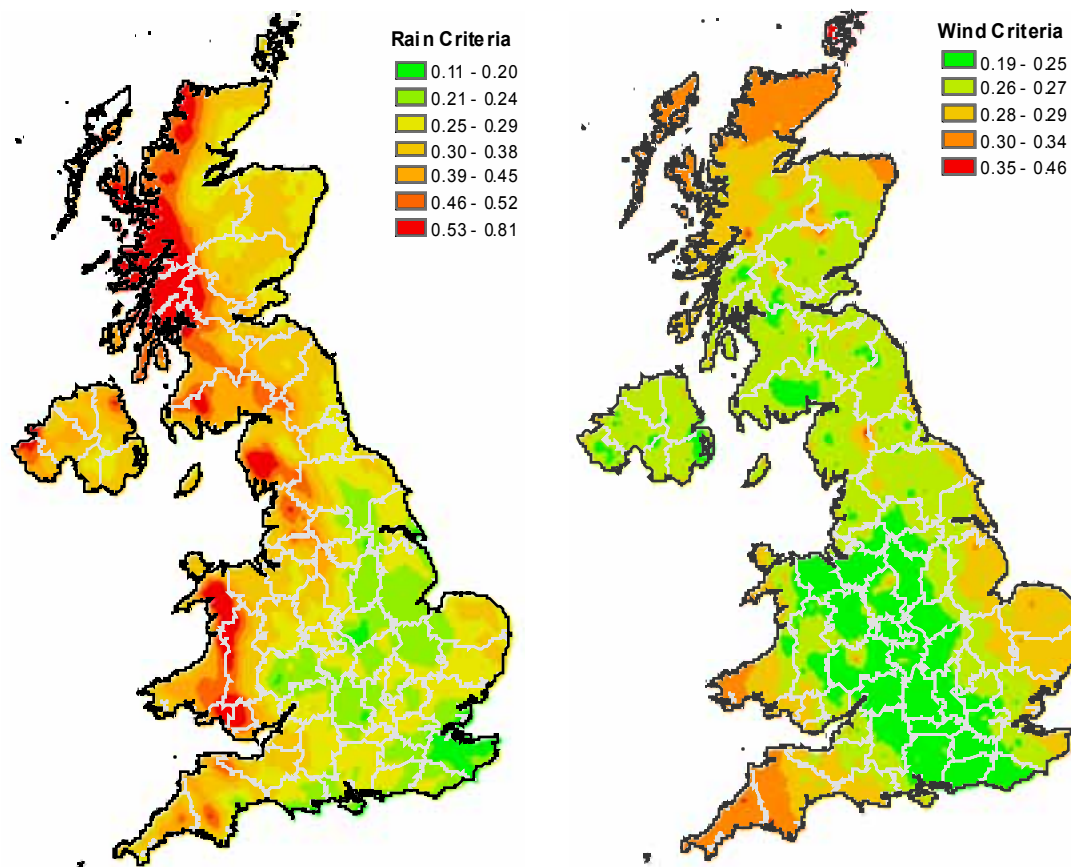


Figure 7.13. Combined rain and wind criteria for lodging risk. Lower values (green) relate to lower risk

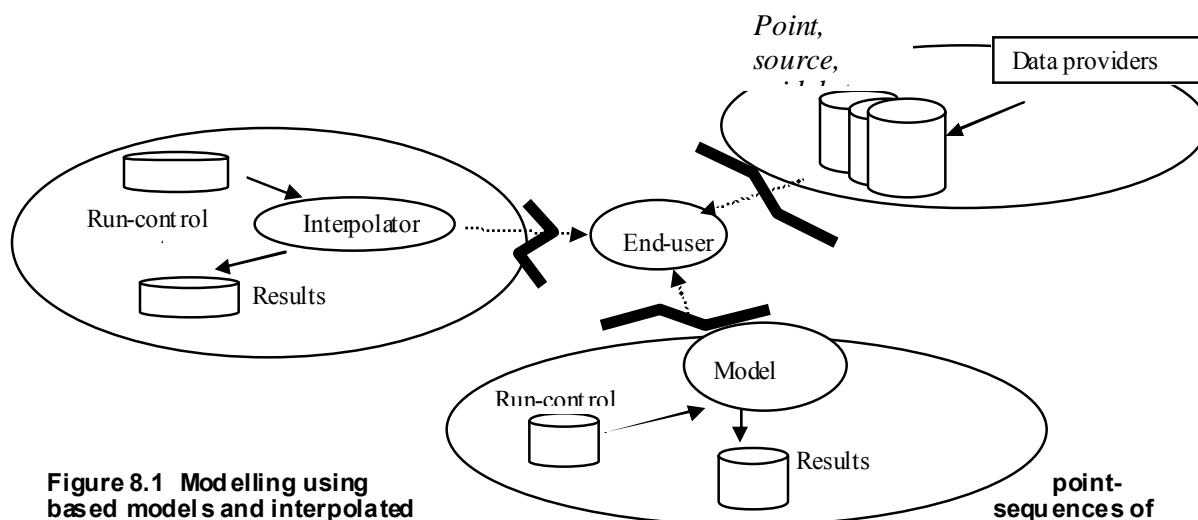
8. MODELLING AND INTERPOLATION FRAMEWORK (MIF)

8.1 OUTLINE

This section describes bespoke software for space-time modelling; the '*Modelling and Interpolation Framework*' (MIF), which was developed for InterMet, by the University of Edinburgh Geography Department. Development of this software capitalised on existing code for modelling and interpolation and made advances in provision of a flexible environment for developers of decision support tools. The software is a key part of the InterMet delivery architecture (described in Section 9). Comprehensive detail of the MIF software is available at from the InterMet web site (www.InterMet.co.uk).

8.2 BACKGROUND

Despite the number of time-dependent point-based agricultural models and a range of interpolation software, a flexible generic integrated interpolation and modelling environment has not previously been available, which is able to provide both services for a potentially wide range of agricultural applications. Many examples exist that demonstrate how geospatial technologies are being used to construct spatially continuous weather fields (e.g. Cornford, 1999; Hutchinson, 1998; Patterson, 1998). However, focusing in particular upon the provision of timely, locally relevant meteorological data to decision support modules, these methods are rarely coupled within a modelling environment. Prior to GEO-BUG (Jarvis, 2001) many agricultural modelling tasks were carried out within a loosely coupled environment where the user provided the interface between multiple packages for data management, interpolation and modelling (Figure 8.1). Especially in the case where models are run over long periods at a daily time step, managing the potentially large volumes of spatial data creates a high overhead and requires large volumes of disk storage in the case where data are obtainable. Where data are not obtainable without significant delay, or disk storage space is restricted, the applications modeller must also both understand and implement the interpolation process and the final model. Neither option is tenable when considering interactive agricultural modelling by client users for day-to-day decision support.



8.2.1 The MIF Framework

The Modelling and Interpolation Framework (MIF) provides sequences of interpolated meteorological data to point-based models, runs these models, and produces model results referenced over time and space. It has been designed and implemented around a central framework, which allows the inter-working of model and interpolation algorithms via compliance with stated interfacing protocols. The framework is structured in a modular way, allowing its components to be installed only where and when needed and which offers a route for efficient development and expansion of its capabilities. Users may interact flexibly with the software as developers, or applied users, through multiple levels of configurable parameter files. In this sense, the design of MIF aims to facilitate a shift away from using of stand-alone programs and towards the greater adoption of interacting modules (Figure 8.2), so that multiple consortium members could interact in software development and its application (Mineter *et al.*, 2003).

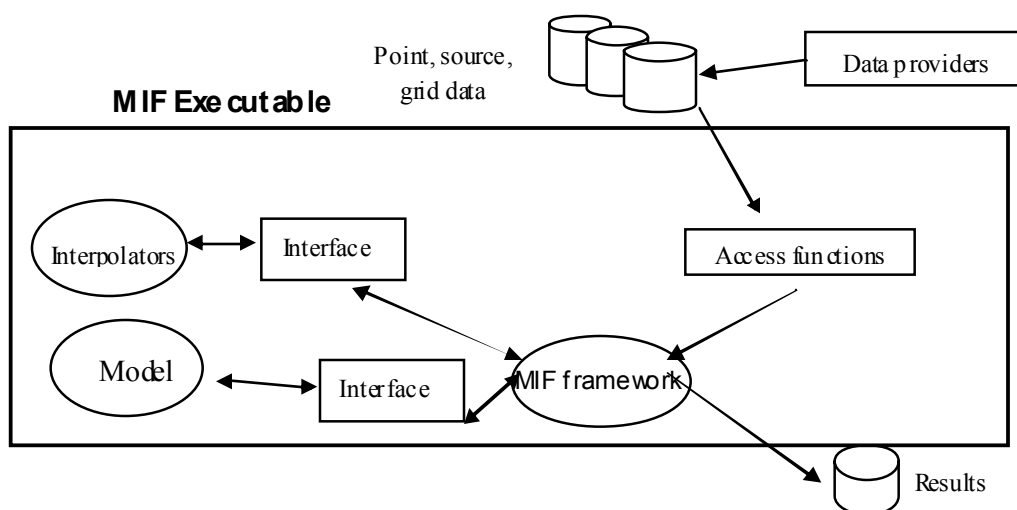


Figure 8.2 From stand-alone programs to interacting modules: a cultural shift whereby interpolation and model components are linked together in one executable image

8.2.2 Context

The wider uptake of DSS depends upon the availability of a reliable supply of standard agro-meteorological variables that shields the end users from the difficulty of collecting, merging, formatting, and error checking data. The need to provide such model input data, and indeed spatially referenced model results, is the core task of InterMet.

It is intended that MIF will allow sequences of interpolated meteorological results over time to be produced at given points for transmission to users. Locally relevant gridded data from appropriately merged networks would also provide a uniform base upon which decisions may be made nationwide. By avoiding the cost overhead of complex sensors, the use of a variety of decision support systems has the potential to become more viable to small and larger enterprises alike. The MIF framework is designed to construct and access these meteorological data. Grids for visualisation, on a daily basis if desired, may be exported using utilities to ArcGISTM, Arc-InfoTM/ArcviewTM grid format (suitable for Internet MapperTM). Alternatively, MIF facilitates the computation of meteorological on the basis of grid reference. These data might then be converted within the Internet interface software to the formats required by particular agricultural DSS.

Maintaining MIF on a centralised server is crucial, owing to the volume of underlying meteorological data and the complexity and size of the interpolation framework.

8.3 SOFTWARE REQUIREMENTS

From the overview and context above (8.2), a list of requirements for MIF emerges.

Output capabilities :

Grids: the facility should exist to create both model and input data as grids over time and space domains

Point runs: sequences of interpolated inputs and model results should be obtainable at individual points across the landscape for which no original meteorological records exist;

Assessments of error resulting from the interpolation process are needed.

Processing abilities :

Data collation:

Software should handle data for multiple meteorological variables that may not be observed at the same station locations. (e.g. temperature observations may be read from other stations, not just those observing rainfall)

When the interpolation technique uses data from fixed grids to refine the estimation equations (e.g. regression, partial spline with covariate submodel) corresponding collateral information is needed for all locations at which data are to be interpolated and for which models are to be run.

Data management:

Because the number of stations reporting data within 24-48 hours in the UK is relatively sparse, maximum use of available station data is needed. For each day, all stations should be inspected and valid records extracted from sites that may also have null records on some or many days.

Data flow:

Models require time series of observations for each input meteorological variable, on a cell-by-cell basis, on every day. The options for managing this have ramifications for data storage, and run-time: the framework must be able to link and support the flow of large volume of data between interpolation and modelling components

More than two variables might be interpolated and passed to a model;

Interpolation requirement:

The incorporation of collateral data (such as elevation) to 'guide' the interpolation process (Jarvis & Stuart, 2001a), for example through linear regression or neural networks. The framework should not need any recompilation or redevelopment to incorporate a new collateral variable.

Multiple interpolators (ordinary kriging, partial thin plate splines, inverse distance weighting, trend surface analysis) should be implemented so that the accuracy and robustness of methods can be compared;

Each meteorological variable needed as input to a model should be capable of being interpolated using a different method;

Each meteorological variable needed should be capable of being interpolated using a different set of guiding variables (collateral data);

The possibility of automatic or semi-automatic parameter setting for interpolation methods should be investigated, given the large volume of data processing involved in annual runs

General requirements

Modularity between units of the software such that modellers and interpolators within the consortium can progress independently of each other;

Straightforward linking with point process models (or multiple linked point-process models) should be enabled so that new models, or a choice of models, can be linked with the framework.

8.4 SOFTWARE DEVELOPMENT

MIF, the modelling and interpolation framework designed for InterMet, was developed from earlier code, Geo-Bug, for integrating interpolation and modelling over time. This earlier code was used to link pest models with sequences of interpolated daily minimum and maximum temperature data, and output gridded, point sequence and cross-validated error results for both interpolated model inputs and model results (Jarvis, 2001). In relation to the requirements specified above (Section 8.3), InterMet required similar outputs, but a considerable extension of the interpolation capabilities. Of these, additional flexibility (to manage multiple input variables and maximise data available for use) and modularity (to integrate code developed by multiple parties) were paramount.

Assessing the Geo-Bug code, and learning from the experience gained in its use, led to the identification of the requirements for InterMet. In order to achieve the extra processing flexibility efficiently, a conceptual re-design of the software was required. Figure 8.3 illustrates how, in general, when software is revised to increase functionality, beyond a certain point it becomes more efficient to re-design the underlying approach: indeed the prototyping approach to software design is based upon this concept. Furthermore, the MIF code, being developed in the context of a LINK project will appropriately be used to identify further requirements, some will be within and some beyond the scope of its present architecture.

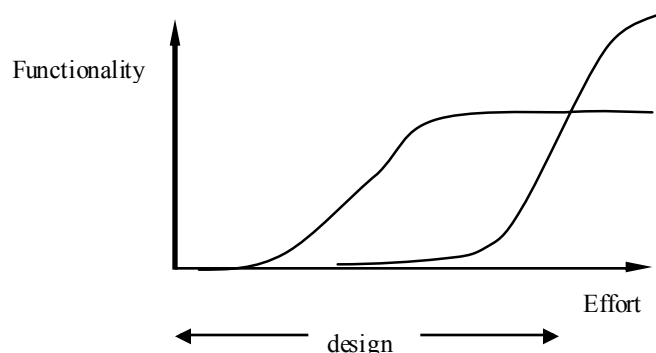


Figure 8.3. The need to re-engineer software

8.4.1 System Platform and development Language

MIF was written in FORTRAN and developed to run on a Sun system platform. Sun is a member of the Foresight LINK consortium. The software language was chosen because it allowed exploitation of a large amount of existing, working FORTRAN code written by Edinburgh and also from other libraries. Doing otherwise would have increased the need for on-going software support effort, and was ruled out by timescales and cost; because it was likely to cause confusion and delay from excessive software testing. In most cases it would have increased the effort inherent in linking a new model to MIF. While the Java language has many merits, the benefits from translating the current MIF code into Java, to facilitate integration with the Internet service, are not sufficient to justify the redevelopment work the task would incur. Moreover, such development could lead to a single, unwieldy software system and duplication of existing functional software (see section 9 for further detail).

MIF Software design and structure

Figure 8.4 provides a schematic overview of MIF, focussing in particular on **data** and **processing** components.

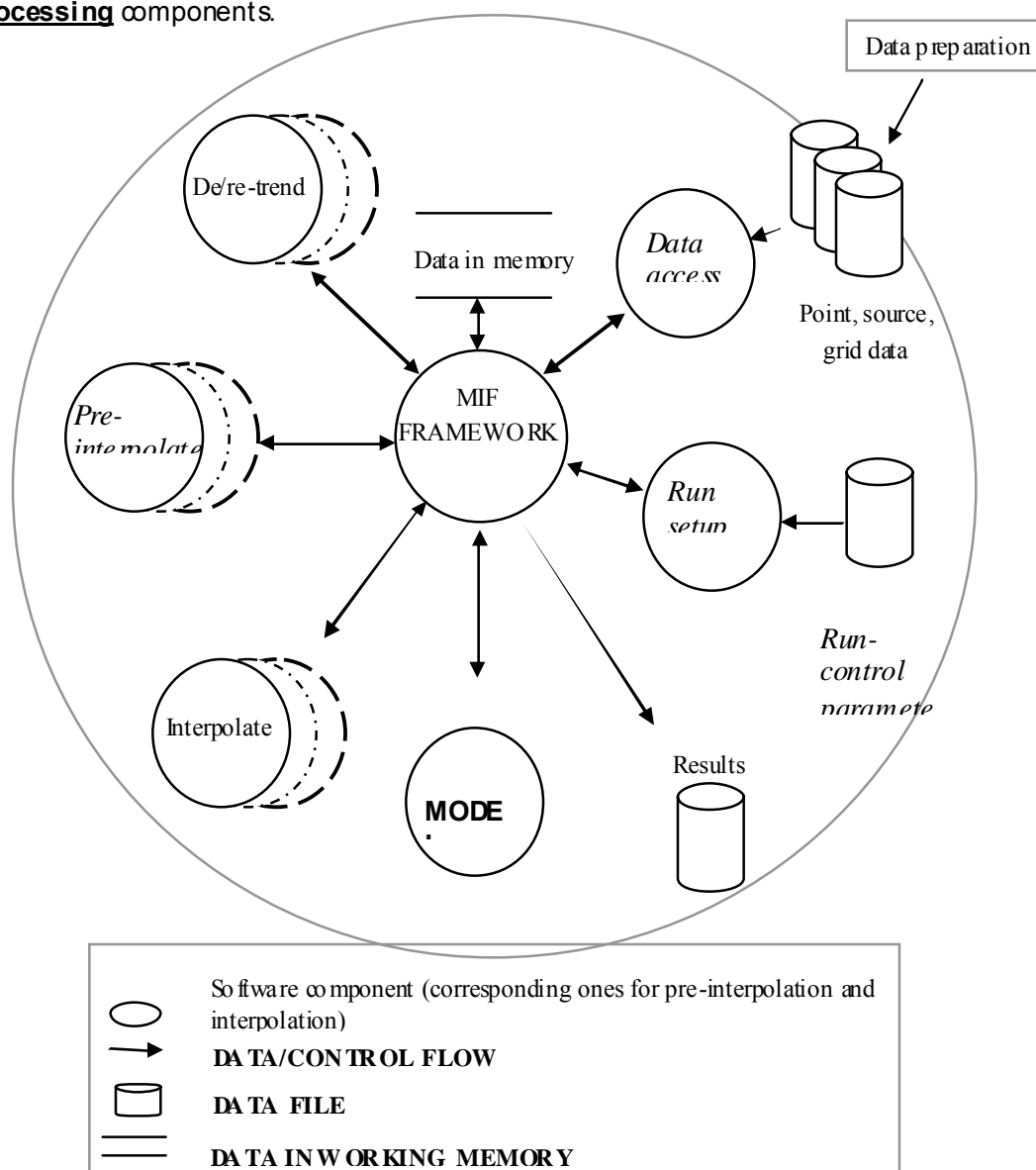


Figure 8.4. Schematic of the Modelling and Interpolation Framework (MIF) software

The contents of these basic software components are summarized below.

Data components

The major data used in the software comprise run-invariant data components (point data and fixed grid data collections) and run-specific data (configuration data and results data).

Run-invariant data

Point data: inputs to interpolation (possibly inputs to several methods – e.g. rainfall might influence tempMin interpolation in future);

Fixed grid data: used only in row-by-row interpolation routines.

Run-specific data

Configuration data:

Run Control Parameters (RCPs) define interpolation methods, guiding variables, dates of run, etc.

Run-Time Parameters (RTPs) held on disk file: specify exact options for each component;

Interpolated data: used directly as inputs to a model, or written out for research or archival purposes.

Model results can be produced as complete grids for specified dates, or time series for specified point (cell) locations.

Processing components

The software comprises:

Routines that create and/or use data:

Read in fixed grid data (elevation etc)

Read in point data

De-trending (regression, for example) and re-trending.

Pre-interpolation analysis, e.g. variogram generation for some methods. Produces data specific to both the method and to the variable being interpolated. Used in interpolation.

Several interpolation components, one per method (kriging etc.)

One model per executable, but with makefile options for linking different models.

Write out results for subsequent display/analysis.

Framework that glues the components together:

Reads run-control parameters

Determines processing required from the RCPs: what variable is created/read from where, used where etc.

Reads point and source data as part of initialisation.

Reads fixed grid data (e.g. elevation of cells) on row-by-row basis as needed.

Invokes other components' routines, including their reading of run-time parameters.

collates/writes out results

Access functions, used by the read/write components.

To read/write data from/to a 'database'

Miscellaneous functions

For example, to manipulate and validate dates.

9. INTERNET SERVICE ARCHITECTURE

9.1 OUTLINE

The InterMet system is necessarily large and complex. This section summarises the system implementation at CSL and gives detail of the technology and methodology used in each of the stages. The system is described from the perspective of the data flow used for the MIF interpolation scheme, which is the most complex scenario. Functionality of the website is described from the point of view of maximising accessibility of the data.

9.2 INFRASTRUCTURE DESIGN

The InterMet system is made up of a series of distinct modules, each designed to run on a specific hardware and software platform (Figure 9.1). The major components of the InterMet system are:

- **Microsoft SQL Server.** A full enterprise relational database engine running on commodity hardware and Windows.
- **MIF.** Bespoke interpolation software running on Sun Ultra Sparc hardware and Solaris, with the Sun Pro Fortran software.
- **ESRI ArcObjects 8.3.** GIS object library running on commodity hardware and Windows.
- **ESRI ArcSDE 8.3.** GIS database interface software running on commodity hardware and Windows.
- **Macromedia Coldfusion MX and Apache.** Web application and server software running on commodity hardware and Linux.
- **ESRI ArcIMS 9.** GIS map generation software running on commodity hardware and Windows.

In order to optimize performance and reliability of the InterMet system, open standards were used wherever possible to integrate these modules. When feasible, the individual components and modules were designed to be run either interactively or automatically to provide maximum flexibility.

Important data are stored permanently in a central database, which reduces the amount of storage required on each of the various servers and provides a single copy of the data for access in different scenarios and by multiple concurrent users.

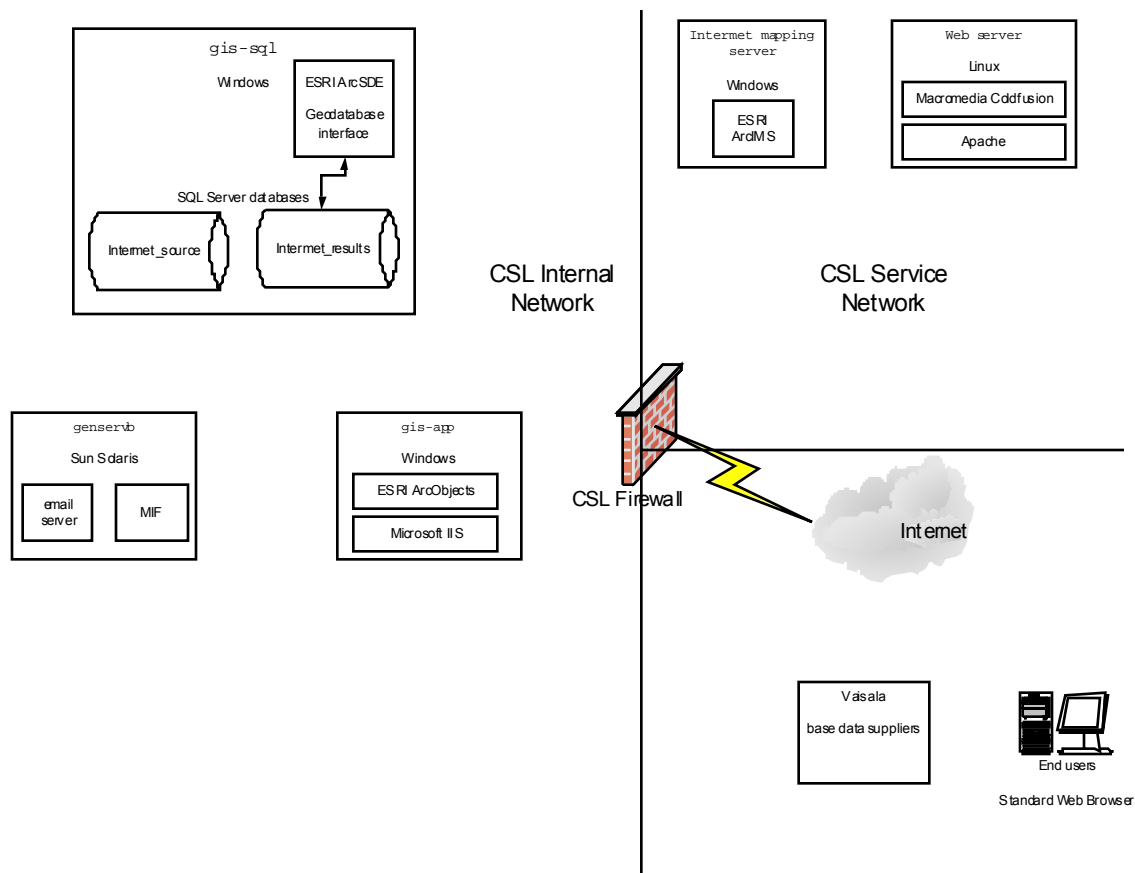


Figure 9.1. CSL InterMet Infrastructure Diagram

9.3 DATA FLOW

9.3.1 Overview

The flow of data through the InterMet system can be summarised as two distinct stages, data input (Figure 9.2) and data output (Figure 9.3).

Data Input:

- Vaisala send an email to vaisala.datain@cs.gov.uk with the data for 'today'
- A processing script inserts the new data into the 'source' database
- A control script extracts appropriate data from the 'source' database and executes MIF with it
- Once MIF execution has completed, the control script sends the interpolated data to the GIS Application server
- A COM based program, making heavy use of the ESRI ArcObjects library, extracts the interpolated data from the MIF files and inserts it into the ArcSDE geodatabase 'results'

Data Output:

- A Coldfusion web application provides users with a route to browse the available data and to request data of specific interest
- Coldfusion requests ArcIMS to draw a map containing the specific data requested by the user
- ArcIMS requests the data from the ArcSDE-enabled 'results' database

- ArcIMS receives the data and creates an image of the map
- Coldfusion receives the map image and sends it back to the user

9.3.2 Detail of Data Input Processes

Internet needs to obtain, collate and store data of different types and from disparate sources.

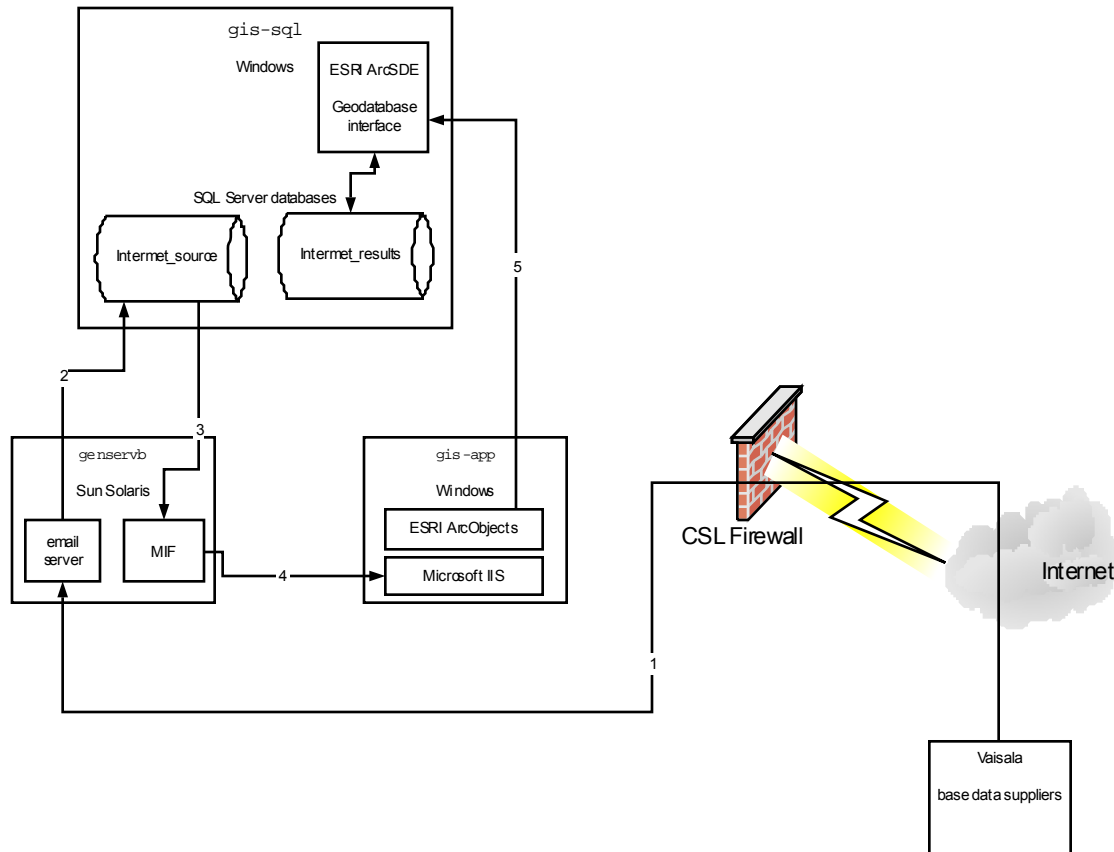


Figure 9.2. Data Input process diagram

9.3.3 Email to database importing script

CSL's email system is run on a fully standard and open Unix system. This makes it very straightforward to set up an email address, for which all receipts are directed automatically to a program.

This process is controlled by a conventional Bourne shell script and invokes a variety of other programs (such as `unpack`, `nawk`, `curl`, etc.) to receive and process the MIME-encoded email and then import any attached text files into the database.

9.3.4 MIF Control Script

The MIF Control Script was written using PHP programming language, which provides several important benefits. Specifically, that the programming language

- is modern and readily accessible
- is platform neutral
- has a huge variety of program libraries available

- requires no extra effort to invoke it either interactively on the server or remotely from another server

MIF is controlled by writing human readable configuration files, then executing a program that uses these configuration files. The configuration files are largely static, the only variables that change are the dates for the data of interest and the filenames to hold the output, so it is straightforward to dynamically create these configuration files automatically.

MIF stores its data using an efficient, but inaccessible, little-endian binary format, which means that significant efforts were required to translate data into the format for MIF to read. A number of possible solutions were explored but the simplest and most successful approach was to write a small program in Java to read the data directly from the database, using a standard JDBC interface, and write it straight into MIF format files. Java was chosen because it allows the programmer complete control (unhindered by unexpected interface layers), provides a rich variety of programming libraries available, and is platform independent.

The input variables to the MIF Control Script consist of a starting date and the duration for which interpolations are requested. Other information, such as output file names and server locations have been hard coded. To make the scenario as simple as possible and because there was no reason to arbitrarily limit the model, the data was always interpolated over the full GB mainland and used all 1180 weather stations as inputs.

For routine operation, the duration is clearly 1 day, but during implementation an initial backlog of data required interpolating in a batch. Experience has repeatedly shown that attempting to process a full extent of data in a single run to be a high risk and poorly scalable method. The PHP script therefore splits up the requested date range into sections of a given interval; 5 days by default. Once split into batches, it is straightforward to execute multiple interpolation runs simultaneously, up to the limit of the server hardware.

9.3.5 Return interpolated data to the GIS Application server

Once a particular model run has completed, the data must be returned to the database for structured and accessible storage. Attempting to use another server to 'pull' the files directly from the filesystem on the Solaris server would have been unnecessarily challenging and would also have presented a classic signalling problem: informing the requesting system that new data are available.

Because PHP on Unix is used to control the execution of the interpolation, there were a wide variety of options available to transfer the files across. The most appropriate standard available for transferring a file between servers of differing architecture is HTTP, commonly used to transfer arbitrary files from servers to web browsing software. The freely available extension "HTTP_Request" from the PEAR repository has been used by InterMet, to provide an easy-to-use interface to the HTTP protocol.

Given that binary data is base64 encoded prior to transit using HTTP and that the server executing the PHP script has obsolete hardware by modern standards, it was important to minimise the file sizes. The MIF output files are readily compressible by the common 'zip' algorithm, achieving ratios of approximately 5 to 1. Sending smaller files over the network reduces the risk of unforeseen limitations being encountered, such as smaller than expected POST buffers on the receiving HTTP server. It also minimises the server memory required to process the request, enhancing scalability.

9.3.6 Transferring MIF output data into the ArcSDE “geodatabase”

The MIF output files have a little-endian binary format, which cannot be directly read by conventional Data Transform and Load tools. As part of the MIF delivery package, CSL received a series of macros that ran within the ESRI ArcMap 8.x environment. These macros were written in Visual Basic for Applications (VBA) and translated the MIF output files into a layer within ArcMap for visualisation. By their very nature, these macros were designed to be run interactively, so work was needed to automate the process.

As they share a common syntax, it was easiest to use Visual Basic 6 as a platform to automate the process. The objective was an easy to use, collection of functions that reliably extract and transform the data into the ArcSDE database. Visual Basic 6 will readily compile a program into a Microsoft COM compliant library without any additional interface programming, this in turn is easily invocable from many scripting environments including VBScript and Active Server Pages (ASP).

The VBA macros were written in broadly a modular form, making it possible to take certain functions directly and put them into the new environment as the basis for new development.

All the GIS specific software being used was provided by the same vendor, ESRI. This provided an inherent degree of integration between the various components, which would otherwise have resulted in a significant amount of development work.

CSL's ownership of an ESRI ArcInfo licence permits InterMet to use the same ArcObjects based environment as the ArcMap scripts, which allows the various ESRI tools to write data directly into ArcSDE. This is not possible with the more common ArcView licences, use of which would have required a more indirect method to be developed for returning results to the geodatabase.

During development, the ArcObjects library was found to be remarkably unstable and contained many undocumented and seemingly arbitrary limitations. If possible, it is recommended that a different method, calling the ArcSDE interface libraries directly, should be used for a commercial InterMet service. For implementation of the InterMet prototype, reliability was maximised by keeping the amount of data (number of days) to a minimum each time ArcObjects is invoked, and by resetting ArcObjects after each run.

The data files were sent from the MIF control script to the ArcObjects server using HTTP, so a straightforward ASP script was written to receive the transmitted files to the application server's filesystem. The ASP environment has limited built-in functionality, the majority of implementations use it in conjunction with third-party components to provide features that developers on other platforms take for granted. After some research and evaluation, the StrongCube Upload Component (from strongcube.com) was chosen to handle HTTP file uploads to the ASP server: because it worked consistently and did not incur a licence fee.

Once the data files are transferred, the COM-based importer program is used to read and interpret the MIF files and export them to the geodatabase as a full ArcSDE Raster Layer.

A simple web page was created to allow interactive access to the ASP based script if required.

One restriction to using the ESRI ArcGIS infrastructure is that all of the parts of the system must be from the same version, for example, ArcObjects 8.3 will not write to ArcSDE 9. Some time after the 2002 temperature data was interpolated through MIF and imported into the ArcSDE geodatabase, CSL, as a matter of course, upgraded its GIS software to version 9.0 and then 9.1. The VBA code that successfully transformed the MIF output into ArcGIS

Raster Layers using ArcObjects version 8.x, did not work with version 9, despite claims of backward compatibility from ESRI. Some effort was devoted to debugging this new scenario, but no solution was found within the resource available. However, a workable solution is expected to be achievable during any future development of the InterMet system to provide a commercial service.

9.4 DATA OUTPUT PROCESSES

The provision of a practicable user interface was a significant requirement part of the InterMet project. Whilst having a well integrated repository of data covering the whole of Great Britain for a wide time span is valuable, much of the potential is lost the data are not readily accessible.

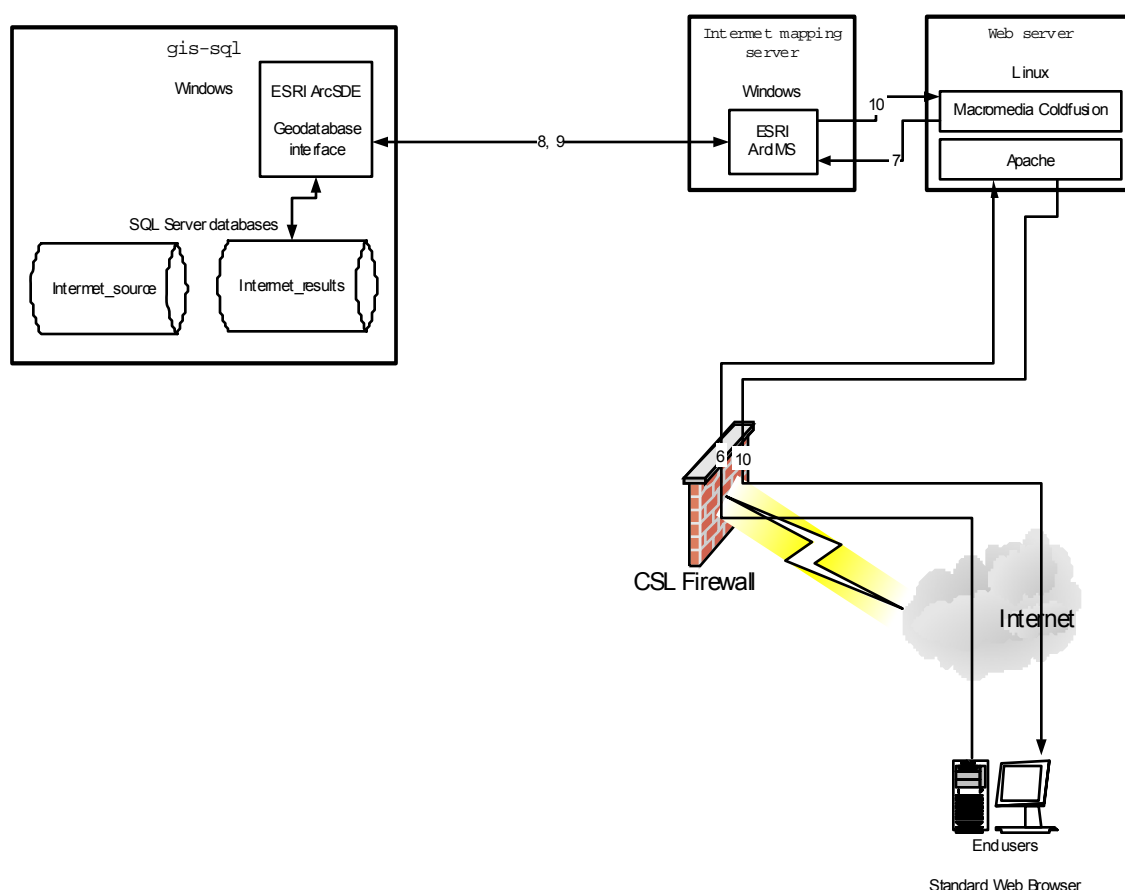


Figure 9.3. Data output process diagram

The development team at CSL has experience, expertise and a well established infrastructure for creating interactive web systems using databases and Macromedia's Coldfusion application server. Design of InterMet website involved acquisition of various data types from disparate sources and integrating them into a readily accessible system. The following sub sections describe the systems and infrastructure required to visualise the InterMet data on a website, the detail of the website functionality is covered elsewhere. For the purposes of this description, it is assumed that the end user is browsing for interpolated data for a location of their interest.

9.4.1 Data selection

The Coldfusion application server has full access to the geodatabase, so it is straightforward for the InterMet website to determine what data is available for perusal and to offer the user appropriate choices.

The InterMet source data are held as rows in conventional tables, but the interpolated data are held in a more complex structure. The ArcSDE Raster data structure is reasonably well defined in terms of which tables hold which data, but the actual pixel values are concealed from ad-hoc queries. This means that all requests for interpolated data must be sent through the full suite of ESRI software.

9.4.2 ArcIMS map

A map was developed to show appropriate background data from the CSL repository to give context to the interpolated data. The resolution of the MIF interpolation is 1km, so it was decided that the closest zoom level (minimum scale) would be the Ordnance Survey 1:50000 map data. The farthest zoom level (maximum scale), and indeed the default, is the full extent of Great Britain.

Visualising data using a map is always an exercise in balancing information density: maps containing too much data difficult to comprehend and may also affect the speed of InterMet processes. ArcIMS is not good at dynamically scaling some raster images beyond a certain level (due to the resampling method used). Specifically the GB outline image became interrupted and unpresentable when drawn as a 300 by 400 pixel image showing the full GB extent. This is illustrated below in Figure 9.4.



Figure 9.4. Suboptimal resampling method

To work around this limitation, two new raster layers were developed specifically for the InterMet map to optimally represent Great Britain at large scales (Figure 9.5). The logical alternative to this approach is to use a Vector outline of the coast, which would display maximum detail at arbitrary scale levels, but this has severe performance implications, taking somewhere between 1 and 1.5 seconds longer to draw a map than with the raster. Given that the default is to show the whole of Great Britain, it can be inferred that this will be a commonly viewed image and should therefore be cached as much as possible.

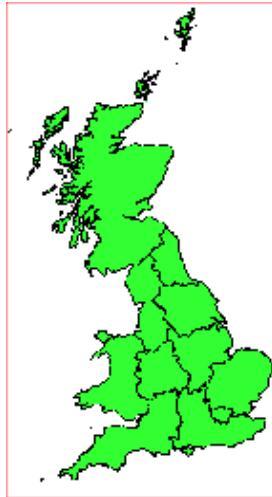


Figure 9.5. Pre-rendered small image

Rendering a raster image with real data values for each pixel requires more consideration than drawing a raster image of a background map. The most popular method of rendering a range of values for visual interpretation is to use a colour ramp, for temperature data red and blue are conventionally recognisable as representing hot and cold, as illustrated by Figure 9.6.

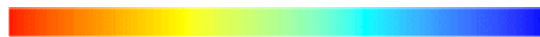


Figure 9.6. Red to Blue colour ramp

The biggest problem with this method of displaying data, is that of mapping numeric temperature data values to colours from the ramp. Probably the ideal approach to this is the Stretch Renderer method used by ESRI ArcMap in conjunction with ArcSDE. Ultimately, the goal is to display data with maximum clarity, which requires the visual contrast between areas of different temperature should be maximized. This is achieved in ArcMap by taking some precompiled statistics from ArcSDE and using a distribution to map the values to the ramp (Figure 9.7).

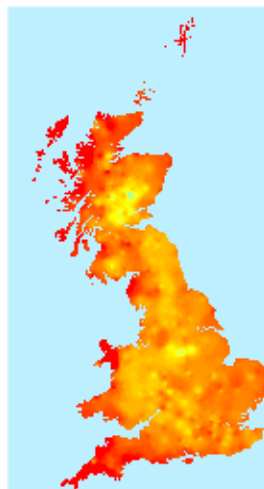


Figure 9.7. Stretch rendered maximum temperature, 3 Jan 2002

The chief problem with this approach is that it is not consistent between different datasets, the colour that represents one temperature for one day will not necessarily be the same temperature for another day. A more fundamental problem with this approach is that it isn't actually possible to define in the ArcIMS XML configuration language, the only options are leaving it as a default greyscale ramp, or to specify which temperature values correspond to which particular colour.

A method was developed which linearly mapped a range of temperature values against the colour ramp, this mapping was then statically defined within the ArcIMS XML map configuration. The downside to this approach is that colour contrast is not great, mostly because, on a given day, interpolated temperature does not differ greatly across Great Britain.

The result from using the linear colour scheme to show the data displayed in Figure 9.7 is illustrated by Figure 9.8.

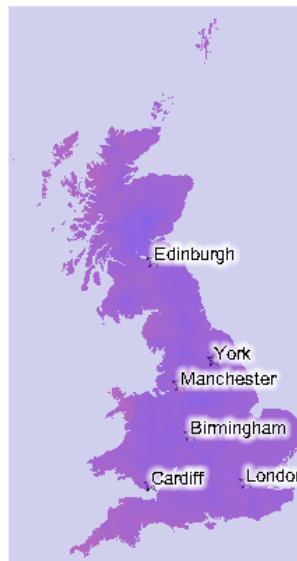


Figure 9.8. Linearly rendered max temperature, 3 Jan 2002

9.4.3 ArcIMS and ArcSDE

Communication between the ArcIMS application server and the ArcSDE database server is a fundamental piece of functionality of the ESRI software. The primary problem for developing links between these crucial components of the system was the optimisation of the number of database connections; in order to minimise resources consumed and therefore maximise scalability.

The InterMet map draws upon data held in a number of different databases within the ArcSDE geodatabase to provide the various layers. ArcIMS is structured in such a way that each map service opens a number of concurrent database connections, as far as ArcSDE is concerned, these need to be minimised as there is an upper limit. After some research and testing, the chosen method was to specify a single database connection for the ArcIMS map and then take advantage of the SQL Server ability to query between different databases.

9.4.4 Returning the map image to the client PC

After selecting the required data, the ArcIMS server renders the map to a standard 8-bit PNG image file. This image file can either be returned to Coldfusion so that the main CSL web servers can serve the image file to the end user, or it can save it to the ArcIMS server's filesystem and return a URL to Coldfusion. This latter scenario, whilst potentially confusing to

receive the image from a different server than the main InterMet website, was chosen because it takes some load from the main CSL web servers.

9.5 WEBSITE DESIGN AND FUNCTIONALITY

The InterMet website is the primary user interface for the data generated by InterMet and is intended to showcase various methods of presenting weather data in accessible formats

Design of the interactive website was influenced by two main considerations:

- presentation of the available data
- providing the facility for users to identify geographic areas of interest

9.5.1 Available data

The InterMet system has two distinct types of data available for use. Time series at fixed points in space held in the “source” database, and continuous surfaces for a fixed point in time, derived by interpolation and held in the “results” database.

9.5.2 Source data

The source data is an amalgamation of the original data from the InterMet consortium providers, which combines into a dense coverage of potential data, although not all weather stations had data present for the purposes of this project. Figure 9.9 illustrates the number of available weather stations across Great Britain:

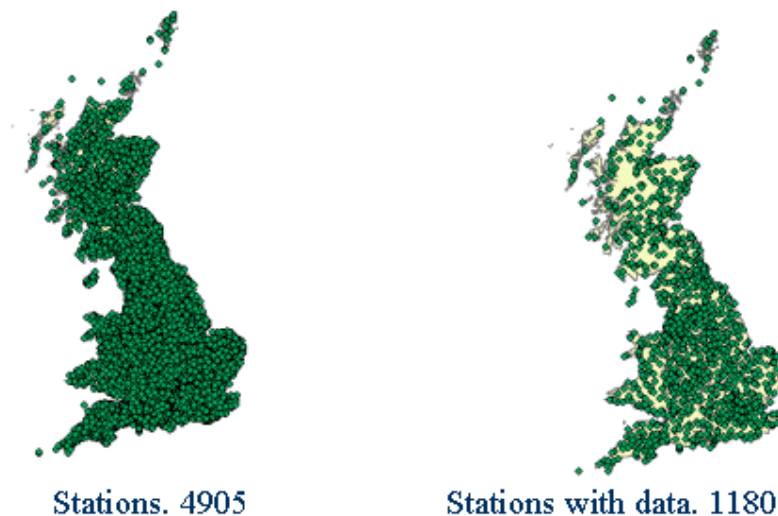


Figure 9.9. Weather station distribution

It is apparent from Figure 9.9 that some care is needed to allow users to select weather stations of interest, because the stations with data are too numerous to select efficiently by name or ID number alone.

It was decided that an interactive map would provide an efficient and logical method of choosing weather stations from the available selection. Coldfusion code was developed, using the ArcIMS Java API, to allow users to pan and zoom the InterMet map. The ArcIMS map used for InterMet has already been discussed, but some additional research was required to come up with an optimum method of displaying the weather station dots on the map.

Ultimately, a maximum scale range was required for the weather stations to start appearing on the map. Ideally, this should be based upon a density function of the number of stations within the currently viewed area. However, implementation of this solution would have had severe performance implications so an acceptable fixed scale was imposed.

This is illustrated in Figure 9.10 for an area between Lancashire and West Yorkshire with 14 visible stations:

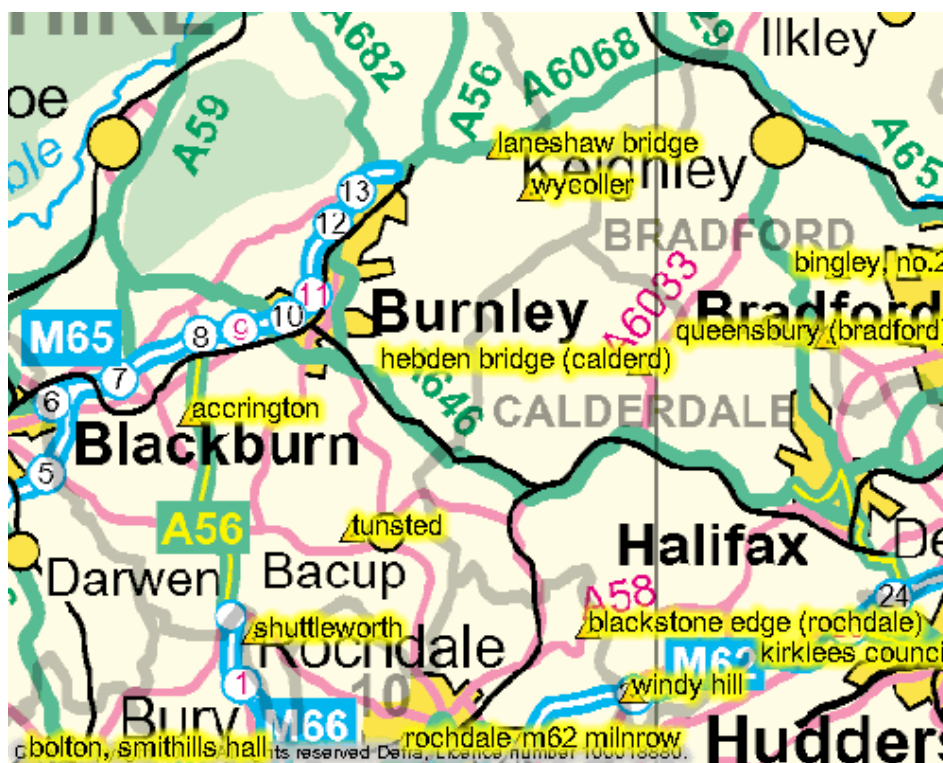


Figure 9.10. Example of weather station map

The detail available through the map is probably sufficient to physically identify the weather station on the ground, so consideration must be given to the Data Protection Act. The primary risk is that an individual might be personally identifiable by the geographic location of a weather station, for example in their back garden. InterMet has mitigated this risk by giving a maximum data accuracy as illustrated in Figure 9.11.

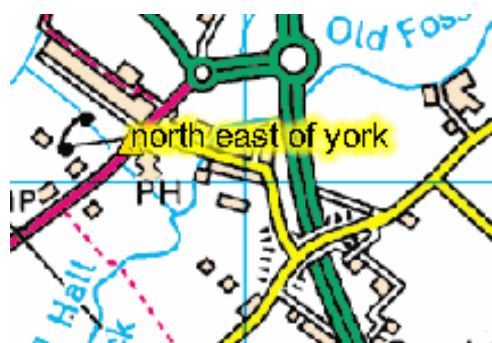


Figure 9.11. Example of maximum detail level

In addition to panning and zooming the map to browse to a location of interest, there is a postcode lookup facility for a user to jump straight to a specific area. Pradice has shown that combination of these two navigation tools provide the quickest means of identifying an area of interest.

Once the user has navigated to an area showing the locations of some weather stations, they need to be able to select a specific one of interest. The obvious choice for this interface is to click onto a station, represented by a yellow triangular symbol labelled with its name (Figure 9.10), but this raises a number of questions:

- which part of the identifier does the user click on
- how accurately must their click be? For example could they click, say, 5 pixels away from the yellow triangle
- what happens if a 5 pixel radius from a click encompasses more than one station?
- what if a user, understandably, tries to click on a label rather than a triangle? There is no way using ArcIMS to relate the label to the object as the layout of the labels is done dynamically each time an image is rendered.

To minimise development effort, this complex issue was avoided completely by simply listing all the stations in the current map in another area of the website and providing conventional links for the user to select from.

Selecting a weather station takes the user to a screen that graphically represents available variables for a period of time. The user is given the choice to specify a period of time of interest to them.

9.5.3 Ease of accessibility features

A number of features were designed into the website to increase its efficiency for users. The need to repeat a series of steps to get to a known outcome is particularly annoying, so some effort was dedicated to remembering user choices and recent selections.

A list of the five stations viewed most recently is maintained for each user session on the website. The function is anonymous, so that users do not have to identify themselves to the system to obtain the benefit. This facility allows a user to rapidly switch between stations of interest without having to browse to the geographic location of them.

The rest of the 'remembering' features are more permanent, in that they persist information in database tables. For this reason, the decision was taken to require individual user registration.

Each registered user is required to log in to the website with an email address and password. The email address is verified during the registration process.

A registered user has the option to mark a station as being a Favourite, this adds it to a list in the same way as the recently viewed stations but has the added advantage that it will be available between browser sessions. The envisaged scenario for this feature was that a user will initially log into the website and mark stations of interest. The next time they visit the site, they would be able to revisit those stations quickly, in order to obtain updated data.

Finally, the interactive map screen has an option for a registered user to save a map location for future reference. This is most useful for the interpolated data, but is equally valid for any map location. This would allow a user to immediately return to an area of interest, perhaps to choose from one of many stations.

9.5.4 Data exporting

One of the original goals of the InterMet website was to provide a central coordination facility for the input and output of data. Output of source data for human interpretation has already been discussed, but consideration needs to be given to exporting data from the central repository to arbitrary computer models.

Coldfusion is primarily used to generate HTML for interactive web pages, but it is equally good at outputting in an arbitrary, character-based format, such as XML, CSV or PDF. Bearing this in mind, it follows that it is straightforward to develop a small piece of code that outputs some data in a pre-specified format suitable for import into a software model.

For the proof of concept website, two formats are available for the user to download source data: Microsoft Excel worksheets and the ArableDS shell, which makes data available to DS tools such as the Wheat Disease Manager.

To provide access control to potentially large amounts of data, this download facility is restricted currently to registered users and could be restricted further if required.

9.5.5 Results data

The interpolated data varies by geography rather than by time, so it is easiest to visualise using a map. Indeed, because the underlying data is stored as an ArcSDE Raster layer it can only be visualised using a map.

An interactive map with varying levels of cartographic detail that allows panning, zooming and jumping to postcodes has already been developed to allow the selection of weather stations.

The interpolated data for a single variable for a single point in time is represented as a map layer, which for the production system would have numbered in the thousands. For this reason, it is impractical to define a map service for each layer, rather the website should dynamically display the appropriate layer based upon the user selection.

ArcIMS has the facility to add arbitrary layers to a given map service when an image is requested, so the map developed for the weather stations was used as the basis to display the interpolated data.

This was a reasonably good approach for general visualization of the data, but using this approach it was difficult to determine the numeric temperature at a given point. This was mostly due to the lack of contrast between adjacent colours and the difficulty of matching a particular shade to the legend by visual inspection, as illustrated by Figure 9.12.

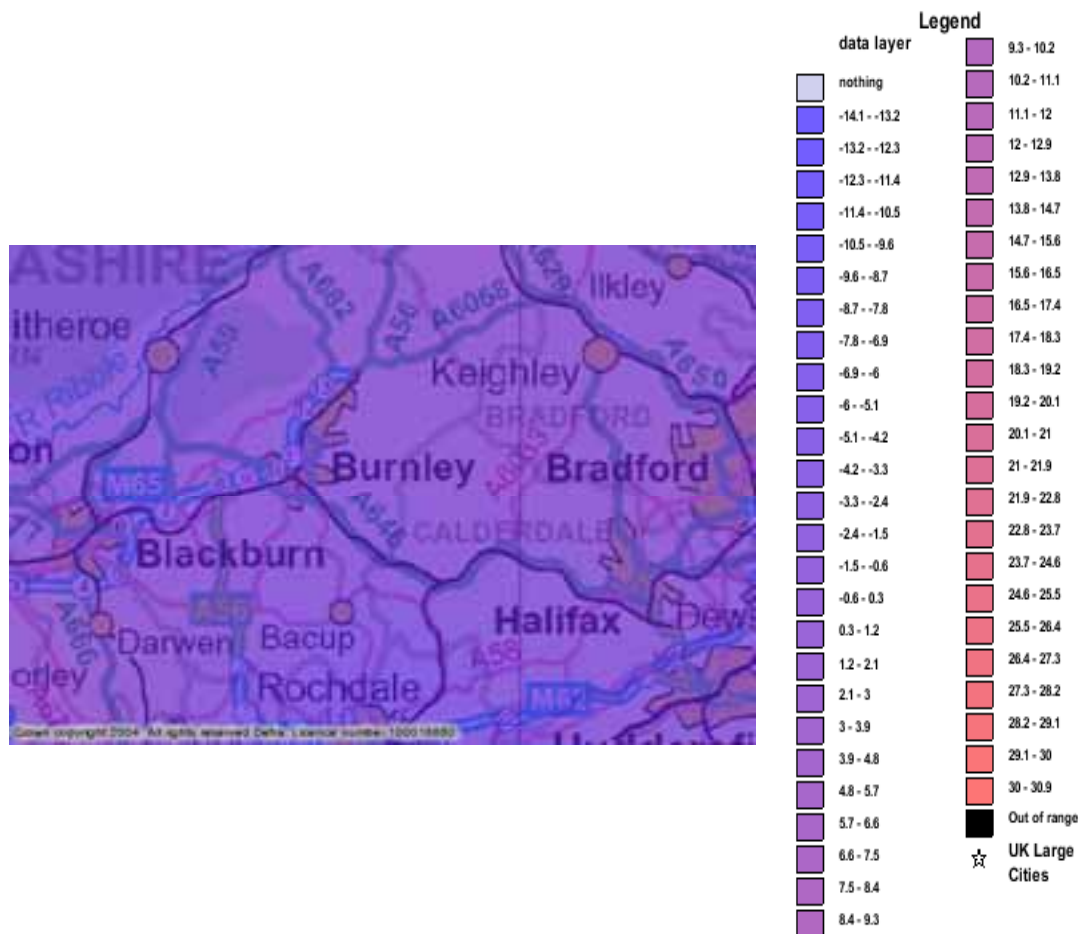


Figure 9.12 Sub optimal data clarity

To solve this problem, an “identify” tool was developed, which displays the numeric values for a given area of interest (Figure 9.13). ArcIMS does not provide this functionality and there was insufficient time available to directly query the ArcSDE database.

ArcIMS does have the ability to return the numeric value from a raster for one pixel per request, but it can not perform this against a layer that was dynamically added to the map. This required that the InterMet website dynamically create an ArcIMS map service that included the appropriate raster layer. There are several negative implications to taking this approach, the first is one of scalability. It takes a noticeable amount of time, in the order of several seconds, to recreate a map service, which has to occur each time the user selects a different variable or time period to view. Due to the way that ArcIMS works, all other InterMet maps are unavailable during this recreation process, which severely limits the number of concurrent users.

The other main concern of having the InterMet website administer the ArcIMS server is one of security. The ArcIMS security model is not particularly granular, it can allow access to certain map services on a per username basis but creating and administering the server is limited to a single administrative username. Whilst security is always a concern when developing any computer system, extra care was needed to secure access to the ArcIMS administrative functions.

The final solution to this issue was to combine both; dynamically overlaying the data and dynamically creating a map service for the data. This is represented to the end user with the interface of enabling and disabling the “identify” tool, which defaults to disabled. This seems to work reasonably well, but is not considered to be an optimal solution.



Figure 9.13. Illustration of an identified max temperature

9.6 FUTURE DEVELOPMENTS

During the development process, several good ideas had to be discarded due to lack of time and resource. They are all technically possible and would all add value to the InterMet system, whether improving an existing interface to the data or creating a new interface.

9.6.1 Extracting time series from a point in interpolated space

The full circle for the centrally stored InterMet data would be to use the values from the interpolated surface to create time series of data for an arbitrary point in space, which could then be exported to computer models in the same way as data from the weather stations. For this to be practical, a process would need to be developed for Coldfusion to systematically query the ArcSDE raster data directly and save that to new tables. To avoid unnecessary quantities of data being stored, this would have to be an offline process whereby a user specifies a new point of interest and asks the system to prepare the data.

9.6.2 End-user alerting

If this system were to be used in a fully live environment, with data routinely arriving into the database, end users would want to know when new data of interest becomes available. Sending an email is completely straightforward, but using mobile phone text messages should also be considered as an option for users.

9.6.3 Displaying interpolated temp for the given area.

The current colour based representation of temperature for a given area is not ideally suited for all cases. It gives a good illustration for large geographic areas, but loses accuracy in smaller areas. The current point identify tool is useful, but a numeric table or grid of data should be considered for the future. This would also require Coldfusion to directly query the ArcSDE raster data and could be displayed as a continually changing number as a user moves the pointer over the map.

9.7 CONCLUSION

The InterMet system as developed so far has proved that delivery, of weather data needed to inform decision support systems, via a web-based portal is technically possible and could achieve the stated aims.

The development outlined above relies heavily upon the established proprietary GIS infrastructure and datasets available at CSL. Whilst it would have been possible to develop a new infrastructure especially for the InterMet project, it would have taken much longer to implement. The overall cost might have been comparable, as there are many open-source and free GIS software packages available (<http://openmap.bbn.com>, <http://www.opengis.org/>), which would replace the expensive ESRI applications. However, other costs would be similar, because the servers that manage the system are based upon commodity hardware.

10. BUSINESS PLAN FOR INTERMET

10.1 OUTLINE

This section proposes the establishment of a cooperative network of meteorological stations to service agri-environment needs. This is necessary because the supply of data from existing networks is prohibitively expensive for many research users, or for those who wish to account for weather variation in their decision making, for example when deciding whether pesticide treatments are needed. The section outlines a business case predicated on the idea of using the database architectures developed for InterMet to provide low cost access to weather data. Initially the service would only provide measurements from the meteorological stations forming the network. Once the cooperative network is established and growing, the interpolation methodologies could be introduced: to offer data at user specified points by surface interpolation at appropriate grid scales (determined in Sections 4-6). We assume that the cost of establishing the core network will be borne by a key stakeholder, perhaps most obviously by Defra. However, recent recommendations from the *“Review of the Agricultural and Horticultural Levy Bodies”* published by Defra (11 November, 2005) highlight the need to promote common working across the levy bodies. The *ad hoc* collection, purchase and use of weather data for farming and environmental applications would clearly benefit from more coordinated management and it might be argued that this fits neatly with the remit of the levy bodies (or the proposed SectorCos). Support and adoption of the proposed network would reduce duplication and thus costs and lead to wider availability and accessibility of high quality weather data for farming. The justification for this is detailed below.

10.2 DESCRIPTION OF THE PROBLEM

The InterMet project has striven to develop technologies to collate, store, enhance by interpolation and distribute weather data, predominantly for the farming industry. Whilst the project has demonstrated proof of concept for the technologies employed, it has encountered two significant barriers to implementing a commercial service using the model envisaged at the start of the work.

1. The availability and uptake of DS tools has not grown as expected, so a market place for the sale of weather data is not yet established. Moreover the industry remains broadly sceptical of the benefits of such systems.
2. The cost of weather data from existing meteorological networks is much greater than could be recouped by a resale service, which aims to add value using the methods developed for InterMet.

These barriers are likely to prevent the achievement of a sustainable service using the commercial model envisaged originally for InterMet. In addition, the work has also identified a number of other issues expected to compromise the quality of a service based on the existing networks:

- Existing meteorological networks were not developed to service agri-environment needs but to make long range weather forecasts
- Many of the recording stations within existing networks do not report quickly enough to service the needs of agri-environment applications, nor do they record all of the variables that are often needed.
- Data quality from existing networks is highly variable (at individual stations and across sites). The best data tend to come from those sites needed for national weather forecasting and include airfield, coastal and hill top locations. However, these key stations are generally located away from areas that are important for more specific agri-environment applications such as, for example, the development and implementation of decision support systems designed to facilitate more sustainable agri-environment management practices in arable and other crops

An alternative business model has been formulated that aims to address the problems identified above, so that the technologies and approaches developed for InterMet can be exploited widely and cost effectively for farming and environmental applications. The alternative model plans to:

- Establish a self-financing network of electronic weather observing stations concentrated across the major arable areas of England and Wales.
- Exploit existing database and Internet technologies developed by CSL; for collecting, managing, summarising and delivering meteorological data.
- Retrieve observations from the network daily by automatic methods, check for errors and store data centrally to allow web-based access for stakeholders, data providers and customers.
- Expand the network by inviting third parties with existing weather stations to add their data in return for free access to processed summaries and/or to the wider network at reduced cost.
- Provide access to the weather data, priced at a level that ensures cost recovery for maintaining the network. Any surplus will be invested in improving the network.

In order to achieve these goals the business plan aims to:

- Deliver a network of weather stations that provide hourly records of important variables within 12 hours of collection with access to raw data and summarised information as required;
- Provide an opportunity for the agricultural industry to evolve a large network of weather stations that is operated on a cooperative principle;
- Use existing infrastructures and technologies, developed by InterMet, to host the storage and dissemination of the data;
- Integrate into the network a database for all meteorological sensors and data-logging equipment purchased by sponsors of the core network.

10.3 BENEFITS

- Purchasing data from existing commercial networks is prohibitively expensive for many research and crop management applications. For example, CSL has been quoted £15k per annum for access to **daily summaries** of a limited number of variables from a network of around 100 UK Met. Office stations. These costs more than double if any commercial use of the data is planned. Similar data obtained from the Vaisala roadside network costs approximately £40k per year. In addition, none of the existing commercially accessible or government-funded meteorological networks were developed to service agri-environment needs.
- All users of meteorological records will benefit from the supply of accurate, up to date weather records, collected to service agri-environment applications, at low cost. A wide range of Departments and Agencies of Defra (including CSL, VLA and Cefas), farm businesses, Universities, agricultural and environmental research organisations, would benefit from contributing to the network and using the data. Participation in the network will provide the following benefits:

10.3.1 Contributors of data

- Professional management and archiving of data from their station(s). This includes immediate error checking routines which will provide timely identification and warning of malfunctioning equipment;
- Access to data from nominated stations within the network (depending upon the amount of data contributed). Apart from its application to decision making, this data will provide a valuable backup for failures in their own equipment;

- Rapid provision of data derived from their weather station measurement, for example weekly accumulated temperatures;
- Provision of data summarized and formatted for use within decision support systems;

10.3.2 Clients purchasing data

- Greater availability of meteorological data that is affordable and relevant to agri-environment applications such as decision support, soil erosion, nitrate leaching and pesticide run-off;

10.3.3 Sponsors of the infrastructure

More efficient coordination of their collection, purchase and use of weather data within the sponsoring organizations and across the projects it funds with all contractors;

Provision of appropriate weather data to support research and monitoring activities;

Demonstrating leadership and commitment to ensuring that infrastructures are in place to support achievement of policy objectives;

Ensuring that appropriate data are available to support decision-making, which quantifies and responds to risk appropriately;

Ownership of the core network and the data collected will comprise a valuable business and research resource for the sponsors;

Tracking new and existing weather equipment through their lifespan, using the proposed database, will reduce wasteful purchase of duplicate equipment;

Efficiency gains from storing, managing and accessing weather data through a single dedicated system.

10.4 CRITERIA FOR SUCCESS

Development and growth of the network could be monitored against six primary criteria for success

1. Initial weather station network automatically reports hourly observations each day;
2. Industry stakeholders join the network with their own stations;
3. Data supplied by the network used across the activities of Defra and its Agencies;
4. Data being used by research groups for crop and environmental science studies;
5. Data being used by the farming industry to run decision support systems;
6. Revenue sufficient to maintain and improve the network in subsequent years

10.5 RISKS AND OPPORTUNITIES

Complete success in developing a national cooperative network of weather stations depends upon the agri-environment industry embracing the idea that it can supply its growing need for weather data by sharing and exploiting existing resources. Facilitating the development of this new cooperative infrastructure will require strong leadership and evidence of commitment from at least one major sponsor. Barriers to fulfilment of the project vision and our plans for managing them are examined in Table 10.11.

Table 10.1. Risks to the establishment of a sustainable cooperative network of weather stations

Risk	Likelihood	How can we prevent it?	What will we do if it happens?
Other organisations will not participate.	Medium	We have already approached key decision makers in potentially interested organisations (e.g., Scottish Crop Research Institute, KG Fruit Ltd., Plantsystems Ltd.) and all have pledged their support in principle (See Section 10).	The value of the network increases with density. However, the core network will be focussed on areas of significant importance for arable crop research and will have value at that scale (see Annex).
A high intellectual property (IP) value is placed on the data 'offered' by the other organisations or individuals.	Medium	Without large quantities of data the value from each station is small. If the network is successful contributors will benefit from opportunities for research and services that will generate greater revenues than any notional IP value given to data from any individual (or group) of stations.	If a significant proportion of potential contributors demand payment to participate, we will review the proposed price structure. And also consider assuming responsibility for all basic maintenance of all stations in the network. However, making significant payments to data providers would reduce our opportunities for reinvestment in the network.
Operators of large existing networks view the cooperative initiative as a threat and introduce an aggressive pricing strategy.	Medium	The cooperative network is targeted to serve a sector that is not currently seen as an important market for existing networks. We will emphasize our intention to deliver data that are appropriate for agri-environment applications.	Reduction in the price of weather data by existing suppliers would be a positive outcome of this initiative. The cooperative network would still have relevance and value because of its provision and frequent reporting and of variables not currently available.
The network grows too quickly and data storage/ access becomes a problem.	Low	Plan for phased increments in network size with sites located optimally at each stage. Follow good project planning principles.	If requests to join the network are enormously above the current infrastructure capabilities we will analyse which sectors of industry are represented and determine whether supporting infrastructure costs could be paid by associated levy bodies, NGOs or commercial organizations.
Weather stations are damaged or stop working.	Low	One of the new stations will be kept as a spare. Failures are most likely to be single sensors rather than the entire station. A regular maintenance cycle will be established.	Replace faulty station/sensors as quickly as possible with spare component. Repair and maintenance costs are estimated in the annual budget.
Storage and access systems at CSL fail.	Low	All incoming data are stored on RAID (automatic copying across multiple hard disks) enabled storage system so hard disk failure will not result in data loss.	Backup system will function as normal until primary system can be replaced.

10.6 IMPLEMENTATION IN YEAR 1

Month 01:	Complete project plan with milestones and gate reviews for presentation to the Challenge Fund team.
Month 02:	Procure and begin installation of weather station equipment at rate of five stations each month.
Month 06:	Appoint a steering committee of industry stakeholders. Data from first weather stations received. Error checking, database design and access system development.
Month 07:	Build up number of core stations and begin attracting additional members.
Month 09:	Launch web site to access data.
Month 12:	Demonstrate effectiveness of network to potential partners

10.7 LOCATION OF CORE NETWORK STATIONS

Using the locations of growers/crop consultants who used the Wheat Disease Manager Module of ArableDS³ in 2004, a density analysis was used to identify the most efficient locations for weather stations (see Section 6.5 for more detail). This group of early-adopters are also likely to use the other DS tools under development for arable production, which are planned for launch during the next 12 – 36 months. The ArableDS community provide an ideal group of customers to target in the first phase of network development. The density analysis shows that these users are located across the main areas for cereal, oilseed rape and potato production. Therefore the core network will have immediate value to crop research and management on these crops. Members of the Association of Independent Crop Consultants (AICC) will be approached to recommend growers who would be willing to host a weather station on their land in return for free access to the data it recorded.

10.8 RESOURCES AND COSTS

The total costs for establishing the basic network is £241,752 (assuming establishment during 2005/06).

These costs comprise the capital cost of the core network plus the operational and staff costs.

10.8.1 Capital cost of meteorological equipment for core network

We suggest developing the core network by purchasing 25 new stations (Table 10.2) and upgrading 25 existing stations (Table 10.3). Upgraded stations will have a minimum set of meteorological sensors; allowing measurement of wind speed, rainfall, air temperature, relative humidity and total solar energy.

³ The Wheat Disease Manager (WDM) is a decision support tool to help crop managers account appropriately to disease risk when making decisions about pesticide treatment. Uptake of the system would lead to an overall reduction in the amount of pesticides applied to wheat. Up to date weather data are needed to run WDM. The system was developed through funding from Defra and the Home-Grown Cereals Authority.

Table 10.2. Estimated cost for 25 stations from preferred supplier (Delta-T Devices, Cambridge) discounted at 17.5%

Description	Unit price (£)	Net Amount (£)
Dataloggers, internal memory expansion, station masts, secure enclosures, solar energy rechargeable battery power system.	3,081	59,979
Self contained, solar powered GSM communications system.	440	9,075
Meteorological sensors: wind direction, wind speed, rainfall, total solar energy, photosynthetic active radiation, soil temperature (10 & 30cm depth), air temperature, relative humidity and surface wetness.	2,594	53,502
Sub-total		122,556

Table 10.3. Cost of upgrading 25 existing Delta-T weather stations to remote data access via GSM cellular phone network, also discounted at 17.5%.

Description	Unit price (£)	Net Amount (£)
Solar powered GSM communication system, rechargeable battery, secure enclosure.	1,315	27,122
Sub-total (Table 10.2 + 10.3)		149,678
VAT @17.5%		26,194
Total for new equipment		175,872

10.8.2 Operational & Staff Costs

The costs required to develop and operate the cooperative network in the first year are shown in Table 10.4.

Table 10.4. Operational and staff costs for Challenge Fund financial year 2005/06.

Network establishment, software development and web design	Days	Cost (£)
Contract and IP issues arrangement	12	7,588
Project management, marketing and publicity	35	12,429
Scripting software to get data from loggers into database	14	5,243
Sufficient disk storage hardware		2,000
Average installation costs for new stations including T&S @ £400 per station		10,000
Average installation costs for upgraded stations including T&S @ £200 per station		5,000
Develop web interface	15	3,135
Online payment system	4	1,838
Monthly charge for 50 SIMM cards @ £268.50 each		3,222
Daily call charges assuming 2 minutes per weather station @ 4.5p minute which equates to £4.50 per day		1,643
Network monitoring and software systems maintenance	30	13,782
Total		65,880

10.8.3 Funding beyond the establishment year

The network will continue to expand and be self-funding through revenue obtained:

- o From annual subscription fee for users supplying data into the system
- o The sale of data to users that do not contribute any data

The subscription fee is based on a sliding scale depending on the amount of data contributed (Table 10.5). Thus, the annual fee for owners of smaller networks is proportional to how much data they wish to contribute. For larger organisations the annual fee reduces as more data are contributed to act as an inducement to join the cooperative with more than the minimum number of stations required to gain access to the entire network.

Table 10.5. Proposed charging scheme

User Group	Annual Subscription	Meteorological Stations Contributed	Stations Accessible	Total no. of stations available to network members
'Small' < 6 contributed stations	£80 ^a £50 £100 £150 £200 £250	0 1 2 3 4 5	Nom inated Stns. 2 4 6 8 10	As required 3 6 9 12 15
'Large' > 5 contributed stations	£1,000 £600 £400	Up to 10 Up to 15 Up to 20		Entire network

^aThe cost for each nominated station is £80 per annum for non contributors.

Growers, advisors, researchers and other interested parties who are unable to contribute data to the network will be able to buy readings from as many stations as they need. In the first year we expect to attract at least 200 data access payments based on interest expressed by the arable, horticultural, research and consultant communities. This would generate revenue of £16,000 to be set aside as a contingency fund for unforeseen costs. In subsequent years, we expect to have at least 400 data access payments generating around £30,000 per annum. These monies are sufficient to cover staff running costs (Table 10.6) and network maintenance (Table 10.7).

Table 10.6. Annual staff costs for monitoring and advertising the network

Activity	Days	Cost £
Project management	3	1,379
Integration of different weather station types	3	1,124
Network monitoring and software systems maintenance	6	2,757
Marketing & publicity	6	1,662
Total		6,922

Weather station maintenance and sensor recalibration costs are based on advice from Delta-T and our experience with existing equipment. Warranty period is one year from station installation.

Table 10.7. Costs for maintaining the core network

Description	Unit price/£	Net Amount/£
Three year cycle for each station's sensor recalibration, eight stations each year.	1,220.00	9,760
Average dismantling, transportation and re-installation costs per station.	500.00	4,000
Sub total		13,760
Total		20,682

The spare weather station will act as a replacement during recalibration and maintenance works on the core network stations, estimated to take 7 – 10 days per station. Organisations and individuals contributing data, from their weather stations into the cooperative network, will continue to pay their servicing and sensor recalibration charges as normal.

InterMet would dedicate effort to engaging other UK funding bodies to participate in the development of the network. In particular, a key target would be to present and demonstrate the concept to key individuals in the Welsh Assembly, DARDNI, SASA and the Scottish executive.

The proposed business model is innovative in providing a cost effective, demand-led route to establish a bespoke agri-environment meteorological network for England and Wales. Demonstration of success in the first year would be used to foster wider participation to produce a truly national agri-environment meteorological network that is self-financing and that evolves and grows according to need. Funding this initiative will establish a route for sponsoring stakeholders to take control of their potential resources, to underpin the delivery of policy objectives and to foster a new paradigm for data provision for themselves and across the wider industry. No other sources of equivalent weather information are available for agri-environment applications. Hence the project supports Defra's capacity to deliver by:

- o Demonstrating leadership;
- o Re-shaping the management and delivery of a key data resource;
- o Focussing that delivery to internal and external customers and partners.

The technologies needed to deliver the vision are in place at CSL, exploiting these will improve efficiency and value for money by:

- o Using technology to do more;
- o Developing a new corporate service that delivers access to high quality weather data;
- o Reducing the cost of weather data.

Stakeholder sponsorship is needed to develop the proposed business model, establish the minimum infrastructure needed, and provide impetus for growing the cooperative business with industry.

Knowledge of weather variation is crucial to many of the strategic priorities common to research funders: climate change & energy, sustainable production & consumption, natural resource protection, and a sustainable farming and food sector. The business model proposed here recognises the importance of obtaining high quality, relevant weather data for these tasks and sets out a vision to deliver this by partnership with other stakeholders. The establishment of a dedicated agri-environment focussed weather-monitoring network would be a bold and popular achievement. For UK farming to achieve

the objectives of sustainability and practical environmental management, will require strong leadership and commitment to ensuring that affordable tools are in place to support research, uptake and delivery.

10.9 SUPPORT FOR THE CONCEPT

To inform the development of the business model opinion has been canvassed from across the agri-environment industry (research, levy bodies, growers, crop consultants, supply chain). There is wide agreement about the scale of the problem and the merit of the solution planned. Those consulted and who have pledged support in principle include:

Public sector research

- o Scottish Crop Research Institute

Levy body

- o The Home-Grown Cereals Authority

Private sector research

- o Velcourt Ltd
- o The Arable Group

Crop Consultants

- o The Association of Independent Crop Consultants
- o Plantsystems Ltd.

Growers

- o H Duncalfe, Maltmas Farm, Wisbech

Marketing

- o KG Fruits Ltd.

11. GLOSSARY

Apache

A freely available and open-source web server. <http://httpd.apache.org/>

ArcGIS

A powerful and open ended Desktop GIS software, written by ESRI. Available in one of three distinct tiers of functionality, ArcView, ArcEditor and ArcInfo.

ArcIMS

Server software aimed at making GIS data readily available to internet clients

ArcInfo

The highest functionality tier of ArcGIS Desktop and the most expensive.

ArcMap

One part of the ArcGIS Desktop software suite, allows the user to interactively visualise data on a map.

ArcObjects

The underlying software library that the majority of the ESRI Arc software family is based upon.

ArcSDE

The ESRI Spatial Database Engine. This acts as an interface between a standard relational database engine and the ESRI Arc software family.

ArcView

The lowest functionality tier of ArcGIS Desktop, the cheapest and therefore the most common.

ASP

Active Server Pages. A Microsoft technology for writing script based web pages, works on a principle of minimal built-in functionality so is commonly used with a variety of third party software.

Bourne Shell

A fully standard scripting language, commonly used for automating tasks on UNIX.

Coldfusion

A richly functional web scripting language published by Macromedia. Allows the rapid development of fully featured web applications and is readily extensible using Java.

COM

A component software architecture from Microsoft, which defines a structure for building program routines (objects) that can be called up and executed in a Windows environment.

ESRI

Environmental Systems Research Institute. A commercial provider of GIS software. <http://www.esri.com/>

Field Capacity

The amount of water left in the soil after it has been saturated and allowed to drain by gravity for 24 hours.

GIS

Geographic Information System. An umbrella term for any system that manages data in a geographic context.

HTTP

Hypertext Transfer Protocol. A protocol used to request and transmit files, especially web pages and web page components, over the Internet or other computer network.
(answers.com)

IIS

Internet Information Services. The Microsoft web server software.

Java

A platform independent object-oriented programming language developed and controlled by Sun Microsystems. <http://java.sun.com/>

JDBC

Java DataBase Connection. A standard for connecting to relational database engines within the Java environment.

Layer

The visual representation of a geographic dataset in any digital map environment. Conceptually, a layer is a slice or stratum of the geographic reality in a particular area, and is more or less equivalent to a legend item on a paper map. On a road map, for example, roads, national parks, political boundaries and rivers are examples of different layers.

MIF

Modelling and Interpolation Framework. Fortran based interpolation software that runs on Sparc-based UNIX computers from Sun Microsystems. Provided by Institute of Geography, University of Edinburgh as part of the Foresight Link project.

Meteorological Office Rainfall and Evaporation Calculation System - MORECS

A long-standing UKMO service providing reliable UK-wide assessment of general soil moisture status at daily, 40 km resolution.

Meteorological Office Surface Exchange System - MOSES

Real time UKMO soil surface modelling system producing estimates of soil moisture deficit and run-off at hourly, 5 km resolution.

NSRI

National Soil Resources Institute

PHP

PHP Hypertext Pre-processor. A widely-used general-purpose cross-platform scripting language that is especially suited for Web development. <http://www.php.net/>

Raster

A spatial data model that defines space as an array of equally sized cells arranged in rows and columns. Each cell contains an attribute value and location coordinates.

A good example of a Raster map is the Ordnance Survey Land ranger series.

Smith Period

Two consecutive days ending at 0900 when the minimum temperature never falls below 10°C and the RH is above 90% for at least 11 hours on each day. Used for predicting whether conditions favour development of potato blight epidemics. Smith LP, *Plant Pathology* 5, 83-87, 1956.

Soil Moisture

The total amount of water, including the water vapour, in an unsaturated soil.

Soil Moisture Deficit

The difference between the amount of water present in the soil and the maximum amount of water that the soil can hold (the field capacity).

Soil Series

The basic soil types mapped on NSRI soil maps, there are 725 currently defined across England and Wales.

SQL Server

A standard relational database engine, provided by Microsoft
(<http://www.microsoft.com/sql/>)

UKMO

United Kingdom Meteorological Office.

VBA

Visual Basic for Applications A subset of Visual Basic that provides a common language for customizing Microsoft applications.

Vector

A coordinate-based data model that represents geographic features as points, lines, and polygons. Attributes are associated with each feature, as opposed to a raster data model, which associates attributes with grid cells.

Visual Basic

A general purpose programming language developed by Microsoft.

Wheat Disease Manager (WDM)

A decision support tool to help crop managers account appropriately to disease risk when making decisions about pesticide treatment. Uptake of the system would lead to an overall reduction in the amount of pesticides applied to wheat. Up to date weather data are needed to run WDM. The system was developed through funding from Defra and the Home-Grown Cereals Authority.

12. PUBLICATIONS

Publications by researchers associated with the project, including continuing or further related research beyond the period of the project.

Refereed Journal Articles

- Baker, R.H.A., Sansford, C.E., Jarvis, C.H., Cannon, R.J.C., MacLeod, A Walters, K.F.A (2000) The role of climate mapping in predicting the potential distribution of non-indigenous pests under current and future climates. *Agriculture, Ecosystems and Environment* 82, 57-71
- Jarvis, C.H. and Baker, R.H.A (2001a) Risk assessment for non-indigenous pests: 1 Mapping the outputs of phenology models to assess the likelihood of establishment. *Diversity and Distributions* 7, 223-235.
- Jarvis, C.H. and Baker, R.H.A (2001b) Risk assessment for non-indigenous pests: 2 Accounting for inter-year climate variability. *Diversity and Distributions* 7, 237-248.
- Jarvis, C.H. and Collier, R.H. (2002) Evaluating an interpolation approach for modelling spatial variability in pest development. *Bulletin of Entomological Research*, 92, 219-232.
- Jarvis, C.H. and Stuart, N., (2001a) A comparison between strategies for interpolating maximum and minimum daily air temperatures: a. The selection of guiding topographic and land cover variables. *Journal of Applied Meteorology* 40, 1060-1074.
- Jarvis, C.H. and Stuart, N., (2001b) A comparison between strategies for interpolating maximum and minimum daily air temperatures: b. The interaction between number of guiding variables and the type of interpolation method. *Journal of Applied Meteorology* 40, 1075-1084.
- Jarvis, C.H. and Stuart, N., (2001c) Accounting for error when modelling with time-series data: estimating the risk of crop pests throughout the year. *Transactions in GIS* 5, 327-343.
- Jarvis, C.H., (2001) GEO_BUG: A geographical modelling environment for assessing the likelihood of pest development. *Environmental Modelling and Software* 16, 739-751.
- Jarvis, C.H., Baker, R.H.A. and Morgan, D. (2003) Assessing the impact of interpolated daily temperature data on landscape-wide predictions of pest phenology. *Agriculture, Ecosystems and Environment*, 94, 169-181.
- Jarvis, C.H., Stuart, N. and Hims, M., (2002) Towards a British framework for enhanced availability and value of agrometeorological data. *Applied Geography*, 22, 157-174.
- Jarvis, C.H., Stuart, N., Cooper, W. (2003) Informetric and statistical diagnostics to provide artificially-intelligent support for spatial analysis: the example of interpolation, *International Journal of Geographical Information Science*, (6) 495-516.
- Luo, W., Taylor, M. C., Parker, S. R. A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed wind speed and direction data from the United Kingdom. *Agricultural and Forest Meteorology in press*.
- Miner, M.J., Jarvis, C.H. and Dowers, S. (2003) From stand-alone programs towards services and components: a case study in agricultural modelling with interpolated climate data. *Environmental Modelling and Software*, 18, 379-391.
- Rigol, J.P., Jarvis, C.H. and Stuart, N. (2001) Artificial neural networks as a tool for spatial interpolation. *International Journal of Geographical Information Science* 15, 323-343
- Stuart, N., Jarvis, C.H. (2004) Estimating daily precipitation for England and Wales using splines with collateral topographic and radar data sets, In *Proceedings of the International Conference on GIS and Remote Sensing in Hydrology, Water Resources & Environment*, Yichang, China. Institute of Hydrological Sciences, Publication No. 289, IAHS Press, 378-385.

Conference Proceedings

- Colley, M., Jarvis, C.H. (2002) Conditional simulation of monthly temperature surfaces using simulated annealing. In Proceedings of GISRUK'2002, April 2002, University of Sheffield.
- Cooper, W., Jarvis, C.H. (2002) A Java-based intelligent advisor for selecting a context-appropriate spatial interpolation algorithm, Proceedings of GISRUK'2002, Sheffield.
- Jarvis, C.H. (2002) Evaluating visualisations of spatial error: a taxonomic approach, 5th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Melbourne, Australia; 357-364.
- Jarvis, C.H. (2002) Insect phenology: a spatial perspective, Proceedings of the workshop 'Modelling phenology and seasonality', Freising, Germany. European Phenology Network
- Jarvis, C.H. (2003) Building bridges In Interdisciplinary agro-climatological research: the possibilities and challenges of e-science technologies for sharing data and knowledge. Proceedings of the workshop 'Agriculture, climate change and economic consequences'
- Jarvis, C.H. and Stuart, N., (2000) Uncertainties in modelling with time series data: estimating the risks posed by crop pests. 4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4), Banff, Canada: CD-ROM.
- Jarvis, C.H., (2000) Towards more integrated agricultural decision support systems: streamlining the processing and availability of meteorological data through the use of GIS and Internet technologies. 3rd European Conference on Applied Climatology, PISA, Italy; CD-ROM.
- Jarvis, C.H., Cooper, W. and Stuart, N., (2001) Infometric and statistical diagnostics to support an intelligent approach to interpolation. 4th International Conference of GeoComputation, University of Queensland, Brisbane: CD-ROM.
- Jarvis, C.H., Minster, M.J. and Stuart, N. (2001) Computational science strategies to support interactive agricultural modelling for day-to-day decision support. 3rd International Conference on Geospatial Information in Agriculture and Forestry, Denver, Colorado: CD-ROM.
- Stuart, N., Jarvis, C.H. (2003) Estimating daily precipitation for England and Wales using splines with collateral topographic and radar data sets, In Proceedings of the International Conference on GIS and Remote Sensing in Hydrology, Water Resources & Environment, Yichang, China
- Taylor, M. C. (2004) Weather and forecasts of late blight from your local weather station: is it worth doing fancy interpolations or can you just use the nearest one for guidance on your spray decisions? Proceedings of the eighth workshop of a European network for development of an integrated control strategy of potato late blight, Jersey. Applied Plant Research, 35-43.
- Taylor, M. C., Hardwick, N. V., Bradshaw, N. J. (2002) Spatially interpolated Smith Periods and blight outbreak dates in the UK, 1998-2002. Proceedings of the Seventh Workshop of a European network for development of an integrated control strategy of potato late blight, Poznan. Applied Plant Research, 105-120.

Chapters in Books

- Donaldson, C.J. (2005) Interpolating daily rainfall: a comparison of geostatistical approaches. Unpubl. Dissertation submitted for the MSc in Geographical Information Science, Institute of Geography, University of Edinburgh.
- Glover, S. (2002) User Centered Design: Communicating the Output of Spatial Models to Support Pest Management over the World Wide Web, Unpubl. Dissertation submitted for the MSc in Geographical Information Science, Department of Geography, University of Edinburgh.
- Jarvis, C.H., Stuart, N., Morgan, D., Baker, R.H.A (1999). To interpolate and thence to model, or viceversa? Chapter 18 in Gittings, B.M. (ed) Integrating Information

- Infrastructures with Geographical Information Technology. London, Taylor & Francis. Reports and Theses
- Luo, W. (2004) Spatial interpolation for wind data in England and Wales. Unpubl. Dissertation submitted for MSc in Data Analysis, Networks and non-linear Dynamics, Department of Mathematics, University of York.
- Miner, M. Jarvis, C.H., and Stuart N (2004) The Internet Foresight LINK. Modelling and interpolation framework (MIF) (2.0)

13 REFERENCES

- Anonymous, 1969. Tables for the evaluation of daily values of accumulated temperature above and below 42F from daily values of maximum and minimum temperature. 10, Meteorological Office, Bracknell.
- Barrie, A. I., Jonson, C. A. & Gordon, S. C., 2000. *A review of the UK raspberry cane midge protection system*. In: CNR-IATA (Editor), 3rd European Conference on Applied Climatology. CNR-IATA, Pisa, pp. CD-ROM.
- Barrow, E. M., Hulme, M. & Jiang, T., 1993. A 1961-90 baseline climatology and future climate change scenarios for Great Britain and Europe. Part I: 1961-90 Great Britain baseline climatology, Climate Research Unit, University of East Anglia, Norwich.
- Brooks, D. H., 1998. *Decision Support System for Arable Crops (DESSAC): an integrated approach to decision support*. In: B.C.P. Council (Editor), 1998 Brighton Conference: Pests and Diseases. BCPC, Brighton, UK, pp. 239-246.
- Brown, B. G., (1996). Improving forecasting performance by combining forecasts: the example of road-surface temperature forecasts. *Meteorological Applications* 3, 257-265.
- Daly, C., Neilson, R. P. & Phillips, D. L., (1994). A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology* 33, 140-158.
- Denzer, H., (1996). Establishing a plant protection advisory service in Austrian vineyards and apple orchards, based on electronic climate stations. *Bulletin OEPP/EPPO Bulletin* 26, 469-473.
- Deutsch, C.V. and Journel, A.G., (1998) *GSLIB Geostatistical Software Library and User's Guide* pp369, Oxford University Press, Oxford,
- Finch, S., Collier, R. H. & Phelps, K., (1996). A review of work done to forecast pest insect attacks in UK horticultural crops. *Crop Protection* 15, 353-357.
- Grayson, R. B., Blöschl, G., Barling, R. D. & Moore, I. D., 1993. *Progress, scale and constraints to hydrological modelling in GIS*. In: K. Kovar & H.P. Nachtnebel (Editors), *Application of Geographical Information Systems in Hydrology and Water Resources Management (HydroGIS 1993)*. IAHS.
- Hoogenboom, G., (2000). Contribution of agrometeorology to the simulation of crop production and its applications. *Agricultural and Forest Meteorology* 103, 137-157.
- Hough, M. N. & Jones, R. J. A., (1997). The United Kingdom Meteorological Office rainfall and evaporation calculation system: MORECS version 2.0 - an overview. *Hydrology and Earth Systems Science* 1, 227-239.
- Hutchinson, M.F., (2000) *ANUSPLIN Version 4.1 User Guide*, Version 4.1, Centre for Resource and Environmental Studies, Australian National University, Canberra.
- Hutchinson MH, Gessler PE, 1994. Splines more than just a smooth interpolator. *Geoderma* 62, 45-67.
- Hutchinson, M. F., 1991. *The application of thin plate smoothing splines to continent-wide data assimilation*. In J.D. Jasper (Ed.), *Data Assimilation Systems*, BMRC Research Report No. 27. Bureau of Meteorology, Melbourne, pp. 104-113.
- Hutchinson, M. F. & Gessler, P. E., (1994). Splines - More Than Just a Smooth Interpolator. *Geoderma* 62, 45-67.
- Jacobs, W. & Raatz, W. E., (1996). Forecasting road-surface temperatures for different site characteristics. *Meteorological Applications* 3, 243-256.
- Jarvis C. H., (2001). GEO_BUG: A geographical modelling environment for assessing the likelihood of pest development. *Environmental Modelling and Software* 16, 739-751.
- Jarvis C. H. & Stuart, N., (2001a). A comparison between strategies for interpolating maximum and minimum daily air temperatures: a. The selection of 'guiding' topographic and land cover variables. *Journal of Applied Meteorology* 40, 1060-1074.

- Jarvis, C. H. & Stuart, N., (2001b). A comparison between strategies for interpolating maximum and minimum daily air temperatures: b. The interaction between number of guiding variables and the type of interpolation method. *Journal of Applied Meteorology* 40, 1075-1084.
- Jensen, A. L., Thyssen, I., Boll, P. S., Hansen, J. G., Secher, B. J. M. & Juhl, O., 1997. *PI@ntelInfo - Using the internet for customized crop information*, 1st European Conference for Information Technology in Agriculture, Copenhagen.
- Karl, T. R., Williams, C. N., Young, P. J. & Wendland, W. M., (1986). A model to estimate the time of observation bias associated with monthly mean maximum, minimum and mean temperatures for United States locations. *Journal of Climatology and Applied Meteorology* 25, 145-160.
- Kimball, J. S., Running, S. W. & Nemani, R., (1997). An improved method for estimating surface humidity from daily minimum temperature. *Agricultural and Forest Meteorology* 85, 87-98.
- Kleinhenz, B., Jörg, E., Gutsche, V., Kluge, E. & Rossberg, D., (1996). PASO - computer aided models for decision making in plant protection. *Bulletin OEPP/EPPO Bulletin* 26, 461-468.
- Krause RA, Massie LB, Hyre RA, 1975. BLITECAST: a computerized forecast of potato late blight. *Plant Disease Reporter* 59, 95-98.
- Landau, S. et al., (1998). Testing winter wheat simulation models' prediction against observed UK grain yields. *Agricultural and Forest Meteorology* 89, 85-99.
- Lee, M., Hollis, D. & Spackman, E., 2000. *From raw data to the Internet - producing quality climatological services*. In: CNR-IATA (Editor), Proceedings of 3rd European Conference on Applied Climatology. CNR-IATA, Pisa, pp. CD-ROM.
- Legates, D. R., (2000). Real-time calibration of radar precipitation estimates. *Professional Geographer* 52, 235-246.
- Lennon, J. J. & Turner, J. R. G., (1995). Predicting the spatial distribution of climate: temperature in Great Britain. *Journal of Animal Ecology* 64, 370-392.
- Magnus, H. A., Langaarden, Å. & Munthe, K., (1993). An integrated PC environment for the dissemination of plant protection warnings in Norway, together with weather reports and forecasts. *Bulletin OEPP/EPPO Bulletin* 23, 669-671.
- Meek, D. W. & Hatfield, J. L., (1994). Data quality checking for single station meteorological databases. *Agricultural and Forest Meteorology* 69, 85-109.
- Minister, M.J., Jarvis, C.H. Dowers S., (2003) From stand-alone programs towards GRID-aware services and components: a case study in agricultural modelling with interpolated climate data, *Environmental Modelling and Software*, 18,(4), 379-391
- Morgan, D. & Solomon, M. G., (1993). PEST-MAN: a forecasting system for apple and pear pests. *Bulletin OEPP/EPPO Bulletin* 23, 601-605.
- Parker, C. G. & Campion, S., 1997. *Improving the uptake of decision support systems in agriculture*, Proceedings of the First Conference for Information Technology in Agriculture, Copenhagen.
- Parker, W. E. & Turner, S. T. D., (1996). Application of GIS modelling to pest forecasting and pest distribution studies at different spatial scales. *Aspects of Applied Biology* 46, 223-230.
- Phelps, K., Collier, R. H., Reader, R. J. & Finch, S., (1993). Monte Carlo simulation method for forecasting the timing of insect pest attacks. *Crop Protection* 12, 335-342.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P., (1992) *Numerical recipes in Fortran 77* pp933, Cambridge University Press,
- Régnière, J., Cook, B. & Bergeron, V., 1996. BioSIM: a computer-based decision support tool for seasonal planning of pest management activities. User's Manual. LAU-X-155, Canadian Forest Service Information Report.
- Rigol, J. P., Jarvis, C. H. & Stuart, N., (2001). Artificial neural networks as a tool for spatial interpolation. *International Journal of Geographical Information Science* 15, 323-343.

- Russo, J. M., Liebhold, A. M. & Kelley, J. G. W., (1993). Mesoscale weather data as input to a gypsy moth (Lepidoptera: Lymantriidae) phenology model. *Journal of Economic Entomology* 86, 838-844.
- Smith LP, 1956. Potato blight forecasting by 90 per cent humidity criteria. *Plant Pathology* 5, 83-87.
- Strand, J. F., (2000). Some agrometeorological aspects of pest and disease management for the 21st century. *Agricultural and Forest Meteorology* 103, 73-82.
- Supit, I., (1997). Predicting national wheat yields using a crop simulation and trend models. *Agricultural and Forest Meteorology* 88, 199-214.
- Walton, S. B., 1998. *MORPH: expediting the production and distribution of decision support systems to the horticultural industry*. In: B.C.P. Council (Editor), 1998 Brighton Conference: Pests and Diseases. BCPC, Brighton, pp. 823-828.
- Way, M. J. & van Emden, H. F., (2000). *Integrated pest management in practice - pathways towards successful application*. *Crop Protection* 19, 81-103.
- Webster R, 1996. *What is Kriging?* In: White E, ed. *Aspects of Applied Biology* 46, *Modelling in Applied Biology: Spatial Aspects*. Wellesbourne: The Association of Applied Biologists, 57-65.
- Weiss, A., van Crowder, L. & Bemadi, M., (2000). Communicating agrometeorological information to farming communities. *Agricultural and Forest Meteorology* 103, 185-196.
- Wieringa, J., 1998. *Climatological use of station network data*, Proceedings of Cost Action 79 seminar on Data Spatial Distribution in Meteorology and Climatology. Office for the Official Publications of the European Communities, Volterra, pp. 17-28.

ACKNOWLEDGMENTS

NSRI is very grateful to Martin Best (UKMO) and Peter Cox (formerly UKMO) for extremely helpful advice on the UKMO's MOSES and MORECS models and for modifying the MOSES model to serve the purposes of the InterMet project. NSRI are also grateful to Mike Minster (USDG) for supplying a mini-MIF version of the project model integration framework in order that progress on model integration could be made at Silsoe.

The fieldwork could only have been carried out with the co-operation of UKMO, in particular Philip Wills (UKMO Surface Networks Co-ordinator). We also thank Dave Drew (UKMO) at Cambridge, Dave Yeoman (Rothamsted Research) at Rothamsted and various staff at Watnall for allowing the soil moisture and temperature monitoring probes to be installed.