A PC-based weather data bank for crop growth modelling

BSE Clemence
Grain Crops Research Institute, Private Bag X9059, Pietermaritzburg 3200, South Africa

Abstract

One of the problems existing for many crop modellers and researchers in water-related fields in South Africa has been the difficulty in day-to-day access of detailed, country-wide weather data, which, to date, have been stored almost exclusively on large mainframes. The new generation of more powerful personal computers (PCs) and the host of related software has made the option of weather data storage at microcomputer level distinctly feasible. With these considerations in mind, a weather data bank framework was created on an 80386 microcomputer using an intelligent data base development system (DATAEASE). This was designed as a simple, menu-based structure with ease-of-use being the prime objective.

Daily weather data for some 280 South African weather stations were downloaded from a Burroughs mainframe, screened for errors, and imported to the PC weather data bank. Modules created within the data bank included those functions necessary for searching and extracting weather data selectively and quickly. A weather station information directory was added, together with a built-in search-by-coordinate routine.

Importation of DOS ASCII files is one of the features used for updating the weather data bank. Exportation of selected records to a DOS file is another of the data bank functions. The normal range of standard data base procedures exist, such as those needed for maintenance and system configuration.

Several supporting computer programs were developed in PASCAL. These included a comprehensive weather data capture system, a program producing summary tables for any weather station on the bank, and a crop model data preparation program. The latter features routines for infilling of missing weather observations, as well as one for generating typical weather data for the last part of a growing season. Together with a suitable crop growth model, typical weather data can be used in the field of crop yield and maturity date forecasting.

Future enhancements to the system include the possible incorporation of a stochastic climate model for supplementing weather records, and the inclusion of a routine for the estimation of evaporation values for the same reason.

Introduction

With its inherent unpredictability, weather plays a central role in agricultural planning in our predominantly semi-arid country. As De Jager and Schulze (1977) point out, "Physiographic, technological and human factors aside, climate largely determines what crops can be grown, where they are best grown, when they should be grown and what reliability of yields can be expected". Answers to these questions can be sought by using a crop growth computer model. Weather data are hence an essential input to any of these programs. Apart from crop modelling applications, timely and efficient access to historical climate data would benefit farmers, businesses and individuals with operational decision-making such as timing of field work, pesticide applications and irrigation (Reinke and Taylor, 1991).

Considering her vast area and relatively sparse population density, South Africa is blessed with extensive weather data resources. Up till now, weather data here have been stored primarily on large mainframe computers, with the processing power and memory capabilities necessary for dealing with large amounts of daily data. The mainframe data banks include, to date, those of the Soil and Irrigation Research Institute (SIRI), the Weather Bureau and that of the Computing Centre for Water Research (CCWR). However, often the only way the so-called man in the street has had access to these data has been through someone with access to the relevant mainframe via a terminal, or by way of summary reports such as those published regularly by SIRI and the Weather Bureau. While summarised data certainly have their uses, there are many applications, such as crop modelling, which require easy access to the detailed daily data themselves.

Received 20 February 1992; accepted in revised form 4 May 1992.

The second drawback in having most of the weather data on mainframe computer only, is that many computer models are designed for use on microcomputers. Examples of these in the crop modelling field are NTRM (Shaffer and Larson, 1987), CORNF (Stapper and Arkin, 1980), PUTU (De Jager and Lourens, 1991), the IBSNAT family of models (IBSNAT Project, 1989) and AUSIM (McCown and Williams, 1989). The problem in using mainframe weather data as inputs to PC-based software is that the procedure for data extraction and transmission of files to PC hard disk is often relatively long and complicated. Communications software such as HANDSHAKE (a registered trademark of MasterLink Corp.) is required, together with the necessary hardware. This is an expensive and somewhat cumbersome option, and requires a fair amount of computer expertise.

The vast improvement in the processing power and storage capabilities of microcomputers in recent years has made the development of a PC-based weather data bank a feasible proposition. Storing weather data efficiently on a PC can have several advantages. Developing software on a microcomputer is relatively simple compared with many mainframes. Development environments such as TURBO PASCAL (a registered trademark of Borland International Inc.) are fast and easy to use (Radford and Haigh, 1986). Also, there are a wealth of "packages" for data analysis, graphics etc. available at the PC level. Furthermore, many millions of computer users are conversant with the PC operating system. PC processors are now very fast — an 80386 machine compares favourably with some mainframes. Finally, the advantage of using a PC to store, access and analyse weather data is that it is inexpensive to purchase, as well as to operate.

Bearing the above points in mind, it was decided to develop an efficient means of storage of daily weather data at PC level. The chief consideration in this exercise was that the system had to be very easy to use for updating, summarising and extracting the various data for modelling purposes and otherwise.

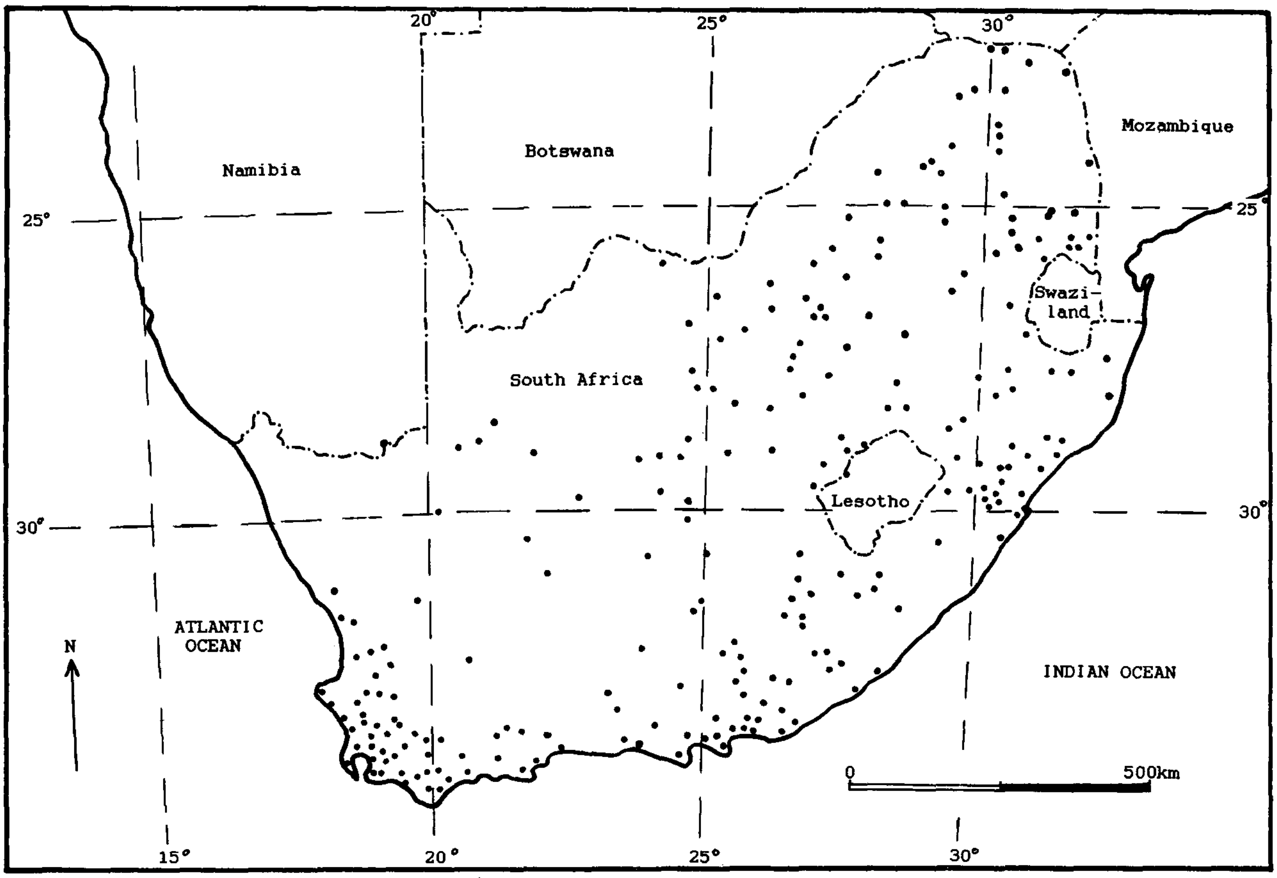

Figure 1
Spatial distribution of stations selected for the weather data bank

Procedure

System description

The computer on which the weather data base was developed was an 80386-based PC running at 33 MHz (an 80386-33 system). The mass storage device used was an 80 Mb Seagate hard disk drive, serviced by a Western Digital FASST controller card. The operating system used was MS-DOS version 4.0 (Microsoft Corporation, 1988). This was necessary for accessing disk space over 20 Mb.

The software used in the actual creation of the weather data base was DATAEASE 4.2 (DataEase International, 1990). This is a powerful relational data base development system. Unlike many other data base languages, DATAEASE operates as an "intelligent" environment, which enables one to develop applications extremely quickly. Its capabilities include being able to use up to 16 Mb of extended PC memory, as well as to access almost unlimited disk space. A single DATAEASE data base can hold up to 2 000 files, each of 2 billion records. In weather data terms, this is the equivalent of 33 million years of daily data per file (5 weather elements).

Supporting programs for data capture and summary, and for the infilling (patching) of weather observations were all developed using the TURBO PASCAL 3.0 compiler (Borland International Inc., 1985). The minimum hardware and software reuirements for using the weather data bank appear in Table 1.

Weather data set

The weather data originated from the Agrometeorological Data Bank of the Soil and Irrigation Research Institute (SIRI). This

TABLE 1
MINIMUM HARDWARE AND SOFTWARE REQUIREMENTS FOR USING THE WEATHER DATA BANK

Hardware requirements:	-IBM PC, XT, AT, PS/2, 80386, and most compatibles -640Kb RAM -1 floppy disk drive -40Mb hard disk drive
Software requirements:	-DOS 4.0 or higher -DATAEASE version 4.2

TABLE 2
CRITERIA USED FOR SCREENING DAILY WEATHER DATA USING THE WFILTER AND METCAP COMPUTER PROGRAMS. Tgraph REFERS TO THE THERMOHYGROGRAPH TEMPERATURE READING, TminH THE MINIMUM HUMIDITY READING FOR THE DAY OFF THE THERMOHYGROGRAPH AND Tmax THE MAXIMUM TEMPERATURE FOR A DAY. ONLY RAINFALL, AIR TEMPERATURE AND HOURS OF SUNSHINE DATA HAVE BEEN LOADED ON TO THE WEATHER DATA BANK

-Temperature: (and Tgraph)	ERROR if Tmax < Tmin or if Tmax/Tmin > 50,0 or < 0,0. WARNING issued if a Tmax record has no equiv. Tmin. WARNING issued when difference between Tmin and TminH > 1,0 C above mean deviation for that month. ERROR when difference between Tmin & TminH > 3 x the mean deviation between them for that month (same checks apply to Tmax and TmaxH as well).
-Rainfall :	ERROR when daily value is <0,0 mm or >500,0 mm...
-Evaporation:	ERROR when <0,0 mm or >25,0 mm, or when evap=0 when rainfall =0 on same day. WARNING issued if evap > 15 mm when Tmax<25,0 AND/OR windrun<250 km.
-SunShine :	ERROR when < 0 h or > max possible hrs for month. ERROR when=0 h with no rainfall on that day.
-WindRun :	ERROR when <7,5 km or >970,0 km in a day.........
-Radiation :	ERROR when < 1,0 $MJ/m^2 \cdot d$ or > 35 $MJ/m^2 \cdot d$.......
-Max Humid :	ERROR when <0,0% or > 100,0% or when < Min Humidity.
-Min Humid :	ERROR when <0,0% or > 100,0% or when > Max Humidity. WARNING issued if a MaxH record has no equiv MinH.

existed on the Department of Agricultural Development's Burroughs A17 mainframe. The daily data were extracted as a series of "element-months" for each station. An "element-month" is a single record of 36 fields in I4 format. Five of these fields are used as identifiers, namely, for the latitude and longitude of the station, the year, the month and the number of the particular element observed. The remaining 31 fields contain the actual observations themselves.

Weather data for over 5 000 stations were present on the SIRI data bank at the time of writing. Most crop growth simulation models require daily observations of rainfall, air temperature and solar radiation (or the more readily available hours of sunshine). It was decided therefore, to select only so-called "sunshine stations" from the total SIRI data set. These were to be stations with at least 5 years of sunshine data with a minimum of 10 years of air temperature and rainfall data at the same site. This criterion was lowered slightly for stations which were "open" (operating) at the time of data extraction. Roughly 280 stations on the SIRI bank conformed with these requirements, of which some 230 were still open. Many of the stations had rainfall-only records back to the early part of the century. These data were considered integral parts of the weather record and were extracted from the mainframe along with the other elements. The spatial distribution of the stations selected is shown in Fig. 1.

The records extracted from the SIRI data bank were then stored as identifiable files on the Burroughs mainframe. Using communications software (Masterlink Corporation, 1983) and hardware, the 280 odd files were transferred to the PC hard disk.

Next a PASCAL computer program, WFILTER, was written to sift the weather data for errors. The criteria used for the screening of the data are presented in Table 2. Every observation in all the files was then screened thoroughly for possible errors, and corrected where necessary. Many of the sunshine observations had values of zero when no rainfall had fallen on the same day. This error was prior to 1985 in all cases, and was corrected accordingly. Any observation in doubt was assigned the value of -999, as per the SIRI convention for missing data.

Creating the data base

As stated earlier, the primary aim in the design of the data base was for it to be user-friendly. The basic unit of storage for the weather data was a record in the element-month format. The records were stored in a structure known in DATAEASE terminology as a "form". A DATAEASE form comprises a number of fields each with a clearly specified set of attributes. The form includes the definition of the file structure, record entry, the data records themselves, and the separate DOS files that hold the rapid-access search "indices" defined for each file. The data base was divided into 5 related structures, each corresponding to 5 different "regions" of data, namely, Transvaal, Natal, Orange Free State, Cape and Eastern Cape. The boundaries of each region did not conform strictly to geographical boundaries. The entire data bank had used roughly 35 Mb of disk space at the time of writing.

The data base was designed as a menu-based structure within which the related applications programs were bound. These appear as options on the appropriate menus. The basic structure of the data base is depicted conceptually in Fig.2. It is centered around the so-called opening menu.

Options 1 to 5 on the opening menu are for accessing the "regional menus". Option 6 accesses the "station information

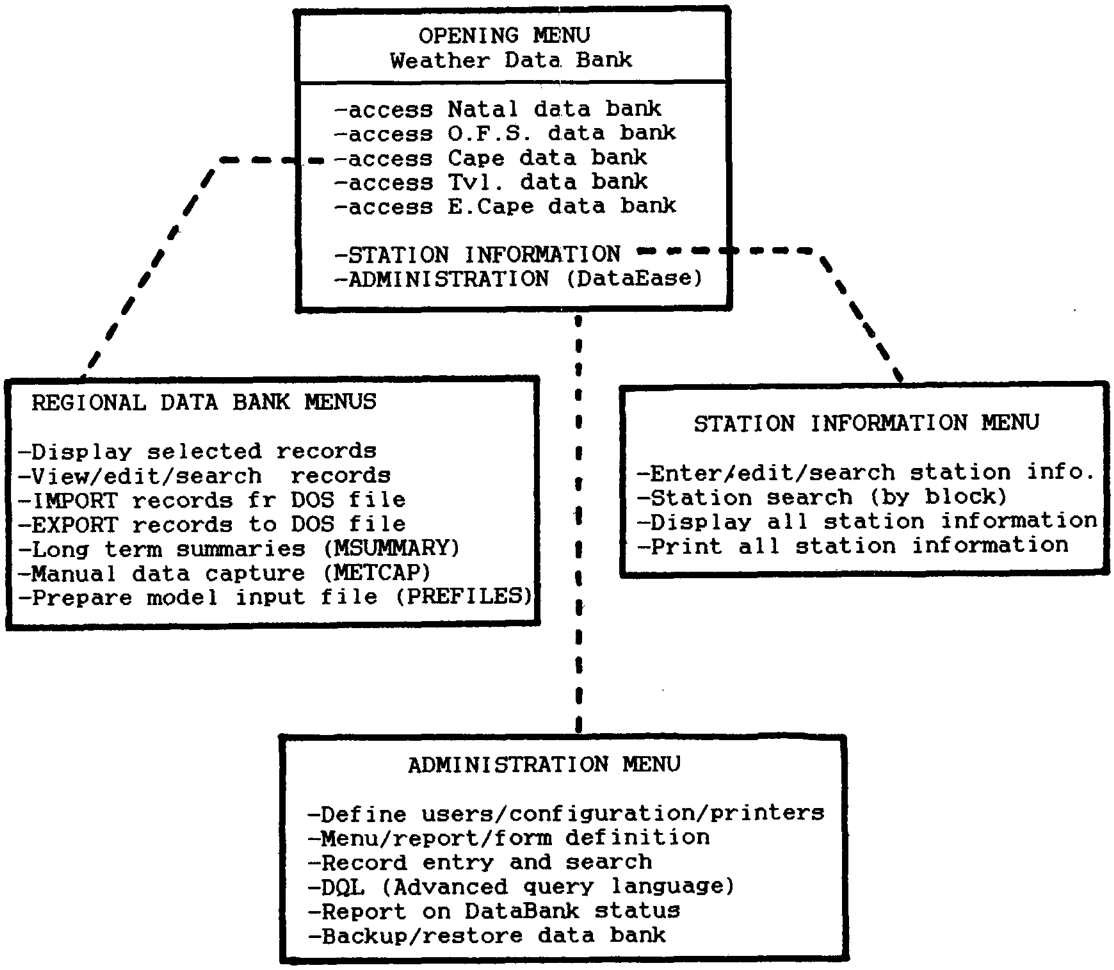

Figure 2
Conceptual structure of weather data bank

TABLE 3
SAMPLE OUTPUT FROM A BLOCK SEARCH FOR STATION INFORMATION. CO-ORDINATES USED WERE 27°00'S TO 30°00'S, AND 28°00'E TO 30°00'E. PSU REFERS TO THE SIRI STATION REFERENCE NUMBER, ALT IS THE ALTITUDE OF THE STATION IN METRES, LAT/LON ARE THE STATION CO-ORDINATES TO THE NEAREST MINUTE AND YEAR END IS THE DATE OF THE MOST RECENT OBSERVATION ON THE DATA BANK

PSU	Name	Region	LAT	LON	Alt(m)	Year end
1421	Geluksberg- GCRI research	NTL	2834	2919	1150	1991
1894	Ladysmith- WalkersHoek	NTL	2828	2940	1200	1991
1901	Tabamhlope- Agr.Res.Stn.	NTL	2902	2939	1450	1990
1948	Himeville- Cobham Forestry	NTL	2941	2925	1675	1991
2004	Villiers CO-OP	OFS	2702	2837	1493	1991
2008	Bethlehem- Loch Lomond	OFS	2810	2818	1631	1991
2068	Kestell- DORP	OFS	2819	2842	1707	1991
2922	Reitz- CO-OP	OFS	2748	2826	1615	1991
8910	Newcastle- Weather Bureau	NTL	2745	2955	1198	1972

menu" which allows the user a number of choices, such as searching for stations within user-specified latitude and longitude co-ordinates, and editing/entering station information. Table 3 is an example of output from such a search. The last option on the opening menu is for the so-called "station administration" function. Depending on the user's security status, there are a number of features available under this choice. These include report design, data base maintenance functions, setting the system configuration and defining user codes and user status.

From the opening menu most users will proceed directly to one of the regional menus. These contain features for entering/editing of records, powerful search procedures as well as functions for the import and export of selected records. Importing of data records from a DOS ASCII file can only be performed by users with relatively high data bank status. This is because it is relatively easy for existing, good data on the bank to be over-written or appended to by new, erroneous data which have not been screened programmatically. To import 10 years of daily weather observations (480 records) to the data bank takes about 11 s on an 80386-33 system. This option is used when updating the data bank, or when adding new stations or batches of weather data. Updating of the bank was being carried out twice a year at the time of writing. The update file with 6 months of data occupies approximately 50% of a 1,44 Mb floppy disk.

Using the data export module on the regional menu does not require any special user status whatsoever. The user inputs the co-ordinates of the station and the years and elements required, and the records are either displayed on the screen or written to a DOS ASCII file. To extract, sort and write 480 records of data (10 years) takes approximately 32 s on the 80386-33 computer.

The remaining options on each of the regional menus are for running the supporting applications programs discussed earlier.

Applications programs

METCAP is a comprehensive meteorological data capture program. Its features include modules for rapid entry of daily weather data with "on-line" error checking, data editing, and generation of month-by-month printed summaries. Before new data can be saved in the element-month format necessary for importation to the bank, they are forced through the same error-checks as those in the WFILTER program (Table 2).

The MSUMMARY support program summarises historical daily data into monthly means. Long-term means and standard deviations are calculated as well.

The PREFILES program was designed to prepare weather data in the various formats necessary for input to different crop models. Solar radiation is calculated from sunshine hours and geographical variables during data preparation. PREFILES contains routines for the automatic patching of weather data using wet and dry day monthly means. This method has been proved to be quite successful for infilling shorter periods of missing observations (Clemence, 1991).

Estimation of solar radiation using air temperature and simple geographical inputs is another feature included in PREFILES. Clemence (1992) reports R-squared values in the order of 0,75 when using a relatively simple regression equation for radiation estimates at a number of sites in South Africa. While this method cannot hope to replace measurements of radiation over extended periods, it has proved invaluable in the patching of missing observations in certain cases, during crop model input file preparation.

PREFILES includes a module for generating typical weather data for the last part of the growing season. When used in conjunction with a good crop model, this has obvious benefits for the forecasting of crop yields and maturity dates (Clemence and Mallett, 1988). Apart from the weather data functions, PREFILES contains a highly interactive set of subroutines for preparing genetic and soils-related input files for crop growth models.

Discussion and conclusions

The data bank system described in this paper allows crop modellers, meteorologists, agricultural advisors and the like to access and process large amounts of high quality weather data rapidly and inexpensively. In addition, the data can be used in conjunction with the vast array of PC level software currently available.

The data bank is relatively compact. Data for the 280 odd weather stations fit easily on to a standard 40 Mb hard disk. Another strength of the system described is the in-built crop model data preparation programs. The associated data-patching techniques have been tested extensively in separate exercises.

One of the big problems in dealing with vast quantities of weather data such as those decribed in this study is that they need to be processed quickly. During the updating process, data is added to the bank on a somewhat *ad hoc* basis and requires rapid sorting into chronological sequence before extraction. By using computer memory dynamically, DATAEASE 4.2 has proved to be entirely suitable for the task. The programs performing the secondary processing of the extracted weather data are all written in compiled PASCAL, an extremely fast and efficient computer language, even at PC level.

Although the weather data bank appears to be perfectly adequate for many applications as it stands, some enhancements are planned which will be incorporated in the long term. These include the following:

Patching of long periods of missing weather data. This is achieved at present using the wet/dry day method described earlier. Tests by Mallett and Clemence (1992) with the CERES-Maize crop growth model have shown that using wet/dry day patched air temperature and sunshine data for long periods can affect the simulations slightly. It is thought that this is due to the omission of extreme values of air temperature. It is planned to test, and possibly incorporate, the daily climate model developed by Zucchini and Brandao (1990). This is a technique of climatic sequencing which apparently emulates climatic means and extremes realistically.

Incorporation of input file data formats for other crop models. At present PREFILES prepares weather data in the IBSNAT, CERES-Maize 1.0 and NTRM model formats. It is a relatively simple task to add routines which output the data in formats for other crop growth models, depending on the demand.

Estimation of pan evaporation. It is common knowledge that consistently accurate measurements of evaporation using an evaporimeter are difficult to achieve, unless these are strictly monitored. Methods of estimating evaporation are improving all the time, and some of the more recent versions such as Hargreaves and Samani (1985) compare favourably with the classic Penman (1963) model which requires a number of sophisticated inputs (Cahoon et al., 1991). Testing of some of these methods is currently in progress. Soon it is hoped to integrate a suitable evaporation model into the PREFILES program which will compute the necessary values during the crop model data preparation process, if required.

Development of a local area network (LAN). DATAEASE supports most of the popular network operating systems (DataEase International, 1990). Should the need arise, the bank can be expanded relatively easily into a full multi-user system.

Acknowledgements

The author wishes to thank the Soil and Irrigation Research Institute of the Department of Agricultural Development, and the Weather Bureau of the Department of Environment Affairs for access to their daily weather data. Special thanks to Dr Johann Erasmus of SIRI for the use of his mainframe weather data extraction program.

References

BORLAND INTERNATIONAL INC. (1985) Turbo Pascal Version 3.01 Reference Manual. PO Box 660001, Scotts Valley, CA 95066-0001. 376pp.

CAHOON, JE, COSTELLO, TA and FERGUSON, JA (1991) Estimating pan evaporation using limited meteorological observations. *Agric. For. Meteor.* **55** 181-190.

CLEMENCE, BSE (1991) A simple method for supplementing long-term weather records in a maize-modelling exercise. *S. Afr. J. Plant Soil* **8**(1) 27-30.

CLEMENCE, BSE (1992) An attempt at estimating solar radiation at South African sites which measure air temperature only. *S. Afr. J. Plant Soil* **9**(1) 40-42.

CLEMENCE, BSE and MALLETT, JB (1988) Typifying late-season weather for maize yield predictions. *S. Afr. J. Plant Soil* **5**(1) 46-47.

DATAEASE INTERNATIONAL (1990) Reference Manual Vol. 1-4. DataEase International, Trumbull, CT 06611.

DE JAGER, JM and SCHULZE, RE (1977) The broad geographic distribution in Natal of climatological factors important in agricultural planning. *Agrochemophysica* **9** 81-91.

DE JAGER, JM and LOURENS, UW (1991) PUTU Maize Crop Growth Model. User Manual. Dept. of Agrometeorology, Univ. of the Orange Free State. 18pp.

HARGREAVES, GH and SAMANI, ZA (1985) Reference Crop Evapotranspiration from Ambient Air Temperature. ASAE Pap. No. 85-2517, ASAE, St. Joseph, MI.

IBSNAT PROJECT (1989) Decision Support System for Agrotechnology Transfer V2.1 (DSSAT V2.1). Dept. of Agronomy and Soil Science, College of Tropical Agriculture and Human Resources, University of Hawaii, Honolulu.

MALLETT, JB and CLEMENCE, BSE (1992) The importance of accurate site information in crop model application with particular reference to CERES-Maize v2. 1. *Proc. S.A.S.C.P. Congress*, Golden Gate.

MASTERLINK CORPORATION (1983) HANDSHAKE for the Personal Computer. Version 4.10. 82 Mitchell Blvd., San Rafael, CA 94903.

McCOWN, RL and WILLIAMS, J (1989) AUSIM: A cropping systems model for operational research. *Proc. SSA IMACS Bienn. Conf. on Modelling and Simulation.* Australian National University, Canberra. 7pp.

MICROSOFT CORPORATION (1988) Microsoft MS-DOS Version 4.00. Redmond, WA 98052-6399.

PENMAN, HL (1963) *Vegetation and Hydrology.* Tech. Commun. No. 53, Commonw.Bur. Soils. Harpenden, England. 125pp.

RADFORD, LE and HAIGH, RW (1986) Turbo Pascal for the IBM PC. PWS Computer Science, Boston, USA 478pp.

REINKE, BC and TAYLOR, SE (1991) Historical climate data management for agriculture: A microcomputer-based methodology. *Comput. Electron. Agric.* **5** 297-304.

SHAFFER, MJ and LARSON, WE (1987) NTRM, a soil-crop simulation model for nitrogen, tillage, and crop-residue management. US Depart. of Agriculture. *Conservation Research Report* **34**(1). 103pp.

STAPPER, M and ARKIN, GF (1980) CORNF: A Dynamic Growth and Development Model for Maize (*Zea mays* L.). Texas Agricultural Experiment Station, Blackland Research Center. Program and Model Documentation 80-2.

ZUCCHINI, W and BRANDAO, A de G (1990) A Stochastic Daily Climate Model for South African Conditions. Dept. of Mathematical Statistics, University of Cape Town, Cape Town. Water Research Commission Report No. 200/1/90.