SDSM 4.2 — A decision support tool for the assessment of regional climate change impacts

User Manual

Robert L. Wilby[1,2] and Christian W. Dawson[3]

August 2007

[1]Department of Geography, Lancaster University, UK

[2]Science Department, Environment Agency of England and Wales, UK

[3]Department of Computer Science, Loughborough University, UK

Preface

General Circulation Models (GCMs) suggest that rising concentrations of greenhouse gases will have significant implications for climate at global and regional scales. Less certain is the extent to which meteorological processes at individual sites will be affected. So–called "downscaling" techniques are used to bridge the spatial and temporal resolution gaps between what climate modellers are currently able to provide and what impact assessors require.

This manual describes a decision support tool for assessing local climate change impacts using a robust statistical downscaling technique. SDSM 4.2 (**S**tatistical **D**own**S**caling **M**odel) facilitates the rapid development of multiple, low–cost, single–site scenarios of daily surface weather variables under present and future climate forcing. Additionally, the software performs ancillary tasks of data quality control and transformation, predictor variable pre–screening, automatic model calibration, basic diagnostic testing, statistical analyses and graphing of climate data.

This manual also describes the UKSDSM archive, a set of daily predictor variables prepared for model calibration and downscaling at sites across the UK. The archive contains variables describing atmospheric circulation, thickness, stability and moisture content at several levels in the atmosphere, under climate conditions observed between 1961 and 1990. Equivalent predictor variables are provided for four GCM experiments of transient climate change between 1961 and 2099. Users seeking to apply SDSM to regions outside the UK may obtain predictor variables online by visiting: http://www.cics.uvic.ca/scenarios/index.cgi?Scenarios

Application of SDSM is illustrated with respect to the downscaling of daily maximum temperature and precipitation scenarios fat a hypothetical location (Blogsville) under present (1961–90) and future (2070–99) climate forcing.

Acknowledgements

SDSM Version 4.2 was supported by the Environment Agency of England and Wales as part of the Thames Estuary 2100 project.

SDSM Version 3.1 was supported by the Environment Agency of England and Wales as part of the *Climate Impacts and Adaptation Research Programme*. The UKSDSM data archive was updated by Ian Harris (Climate Research Unit) and now includes data kindly supplied by the UK Hadley Centre, CSIRO Atmospheric Research, and the Canadian Centre for Climate Modelling and Analysis.

SDSM Version 2.2 was sponsored by the Environment Agency through the National Centre for Risk Analysis and Options Appraisal.

SDSM Version 2.1 was supported by the Canadian Climate Impacts Scenarios (CCIS) Group though the Climate Change Action Fund. Assistance in kind was provided by A Consortium for the Application of Climate Impact Assessments (ACACIA) at the National Centre for Atmospheric Research (NCAR). NCAR is sponsored by the U.S. National Science Foundation.

NCEP Re–analysis data were provided by the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, USA, from their Web site at http://www.cdc.noaa.gov/

The Climate Impacts LINK Project is funded by the UK Department of the Environment, Transport and the Regions (Contract Reference EPG 1/1/124).

CONTENTS

0 TECHNICAL INFORMATION

SDSM version 4.2 runs on PC-based systems and has been tested on Windows 98/NT/2000/XP. Note, on older machines, some statistical analyses may take longer to perform and/or may exhaust available memory when large data sets are processed.

1. INTRODUCTION

Even if global climate models in the future are run at high resolution there will remain the need to 'downscale' the results from such models to individual sites or localities for impact studies (DOE, 1996; p34).

General Circulation Models (GCMs) indicate that rising concentrations of greenhouse gases will have significant implications for climate at global and regional scales. Unfortunately, GCMs are restricted in their usefulness for local impact studies by their coarse spatial resolution (typically of the order 50,000 km^2) and inability to resolve important sub–grid scale features such as clouds and topography.

As a consequence, two sets of techniques have emerged as a means of deriving local–scale surface weather from regional–scale atmospheric predictor variables (Figure 1.1). Firstly, statistical downscaling is analogous to the "model output statistics" (MOS) and "perfect prog" approaches used for short–range numerical weather prediction. Secondly, Regional Climate Models (RCMs) simulate sub–GCM grid scale climate features dynamically using time–varying atmospheric conditions supplied by a GCM bounding a specified domain. Both approaches will continue to play a significant role in the assessment of potential climate change impacts arising from future increases in greenhouse–gas concentrations.

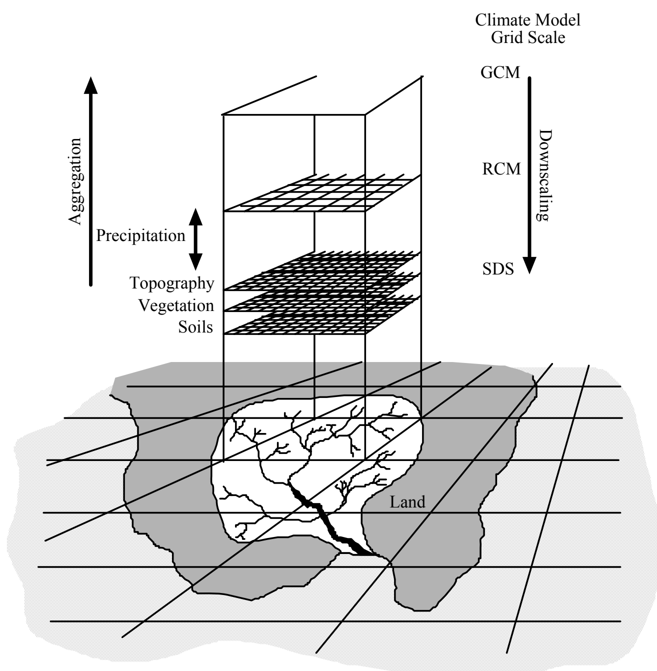


Figure 1.1 A schematic illustrating the general approach to downscaling.

Statistical downscaling methodologies have several practical advantages over dynamical downscaling approaches. In situations where low–cost, rapid assessments of localised climate change impacts are required, statistical downscaling (currently) represents the more promising option. In this manual we describe a software package, and accompanying statistical downscaling methodology, that enables the construction of climate change scenarios for individual sites at *daily* time–scales, using grid resolution GCM output. The software is named SDSM (Statistical DownScaling Model) and is coded in Visual Basic 6.0.

As far as the authors are aware, SDSM was the first tool of its type freely offered to the broader climate change impacts community. Most statistical downscaling models are generally restricted in their use to specialist researchers and/or research establishments. Other software, although more accessible, produces relatively coarse regional scenarios of climate change (both spatially and temporally). For example, SCENGEN blends and re–scales user–defined combinations of GCM experiments, and then interpolates monthly climate change scenarios onto a 5º latitude × 5º longitude global grid. "Weather generators" — such as WGEN, LARS–WG or CLIGEN (see bibliography) — are widely used in the hydrological and agricultural research communities, but do not directly employ GCM output in the scenario construction processes.

Following a brief overview of downscaling techniques, we describe the structure and operation of SDSM with respect to seven tasks: 1) quality control and data transformation; 2) screening of potential downscaling predictor variables; 3) model calibration; 4) generation of ensembles of present weather data using *observed* predictor variables; 5) statistical analysis of observed data and climate change scenarios; 6) graphing model output; 7) generation of ensembles of *future* weather data using GCM–derived predictor variables. The key functions of SDSM will be illustrated using observed and climate model data for a hypothetical station (Blogsville), comparing downscaled daily precipitation and temperature series for 1961-1990 with 2070–2099.

1.1 Downscaling techniques

The general theory, limitations and practice of downscaling have been discussed in detail elsewhere (see bibliography). Reviews typically group downscaling methodologies into four main types: a) dynamical climate modelling, b) synoptic weather typing, c) stochastic weather generation, or d) transfer-function approaches. Each family of techniques is briefly described below.

1.1.1 Dynamical

Dynamical downscaling involves the nesting of a higher resolution Regional Climate Model (RCM) within a coarser resolution GCM. The RCM uses the GCM to define time–varying atmospheric boundary conditions around a finite domain, within which the physical dynamics of the atmosphere are modelled using horizontal grid spacings of 20–50 km. The main limitation of RCMs is that they are as computationally demanding as GCMs (placing constraints on the feasible domain size, number of experiments and duration of simulations). The scenarios produced by RCMs are also sensitive to the choice of boundary conditions (such as soil moisture) used to initiate experiments. The main advantage of RCMs is that they can resolve smaller–scale atmospheric features such as orographic precipitation or low–level jets better than the host GCM. Furthermore, RCMs can be used to explore the relative significance of different external forcings such as terrestrial–ecosystem or atmospheric chemistry changes.

1.1.2 Weather typing

Weather typing approaches involve grouping local, meteorological data in relation to prevailing patterns of atmospheric circulation. Climate change scenarios are constructed, either by re–sampling from the observed data distributions (conditional on the circulation patterns produced by a GCM), or by generating synthetic sequences of weather patterns and then re–sampling from observed data. Weather pattern downscaling is founded on sensible linkages between climate on the large scale and weather at the local scale. The technique is also valid for a wide variety of environmental variables as well as multi–site applications. However, weather typing schemes ca be parochial, a poor basis for downscaling rare events, and entirely dependent on stationary circulation–to–surface climate relationships. Potentially, the most serious limitation is that precipitation changes produced by changes in the frequency of weather patterns are seldom consistent with the changes produced by the host GCM (unless additional predictors such as atmospheric humidity are employed).

1.1.3 Stochastic weather generators

Stochastic downscaling approaches typically involve modifying the parameters of conventional weather generators such as WGEN, LARS–WG or EARWIG. The WGEN model simulates precipitation occurrence using two–state, first order Markov chains: precipitation amounts on wet days using a gamma distribution; temperature and radiation components using first–order trivariate autoregression that is conditional on precipitation occurrence. Climate change scenarios are generated stochastically using revised parameter sets scaled in line with the outputs from a host GCM. The main advantage of the technique is that it can exactly reproduce many observed climate statistics and has been widely used, particularly for agricultural impact assessment. Furthermore, stochastic weather generators enable the efficient production of large ensembles of scenarios for risk analysis. The key disadvantages relate to the low skill at reproducing inter-annual to decadal climate variability, and to the unanticipated effects that changes to precipitation occurrence may have on secondary variables such as temperature.

1.1.4 Transfer functions

Transfer-function downscaling methods rely on empirical relationships between local scale predictands and regional scale predictor(s). Individual downscaling schemes differ according to the choice of mathematical transfer function, predictor variables or statistical fitting procedure. To date, linear and non–linear regression, artificial neural networks, canonical correlation and principal components analyses have all been used to derive predictor–predictand relationships. The main strength of transfer function downscaling is the relative ease of application, coupled with their use of observable trans–scale relationships. The main weakness is that the models often explain only a fraction of the observed climate variability (especially in precipitation series). In common with weather typing methods, transfer methods also assume validity of the model parameters under future climate conditions, and the downscaling is highly sensitive to the choice of predictor variables and statistical form (see below). Furthermore, downscaling future extreme events using regression methods is problematic since these phenomena, by definition, tend to lie at the limits or beyond the range of the calibration data set.

1.2 Relative skill of statistical and dynamical downscaling

The wide range of downscaling techniques (both dynamical and statistical) has prompted a growing number of model comparisons using generic data sets and diagnostics. Until recently, these studies were restricted to statistical–versus–statistical or dynamical–versus–dynamical model comparisons. However, some studies are now undertaking statistical–versus–dynamical model comparisons and Table 1.1 summarises the relative strengths and weaknesses that have emerged.

	Statistical downscaling	**Dynamical downscaling**
Strengths	• Station–scale climate information from GCM–scale output • Cheap, computationally undemanding and readily transferable • Ensembles of climate scenarios permit risk/ uncertainty analyses • Applicable to 'exotic' predictands such as air quality and wave heights	• 10–50 km resolution climate information from GCM–scale output • Respond in physically consistent ways to different external forcings • Resolve atmospheric processes such as orographic precipitation • Consistency with GCM
Weakness	• Dependent on the realism of GCM boundary forcing • Choice of domain size and location affects results • Requires high quality data for model calibration • Predictor–predictand relationships are often non–stationary • Choice of predictor variables affects results • Choice of empirical transfer scheme affects results • Low–frequency climate variability problematic • Always applied off-line, therefore, results do not feedback into the host GCM	• Dependent on the realism of GCM boundary forcing • Choice of domain size and location affects results • Requires significant computing resources • Ensembles of climate scenarios seldom produced • Initial boundary conditions affect results • Choice of cloud/ convection scheme affects (precipitation) results • Not readily transferred to new regions or domains • Typically applied off-line, therefore results do not always feedback into the host GCM

Table 1.1 Main strengths and weakness of statistical and dynamical downscaling.

The consensus of model inter–comparison studies is that dynamical and statistical methods have comparable skill at estimating surface weather variables under *present* climate conditions. However, because of recognised inter–variable biases in host GCMs, assessing the realism of *future* climate change scenarios produced by statistical downscaling methods is problematic. This is because uncertainties exist in *both* GCM and downscaled climate scenarios. For example, precipitation changes projected by the U.K. Met Office's coupled ocean–atmosphere model HadCM2, were found to be over–sensitive to future changes in atmospheric humidity. Overall, the greatest obstacle to the successful implementation of both statistical and dynamical downscaling is the realism of the GCM output used to drive the schemes.

However, because of the parsimony and "low–tech" advantages of statistical downscaling over dynamical downscaling (Table 1.1), a hybrid conditional weather generator method was chosen as the basis of the decision support tool, SDSM.

1.3 Manual outline

The rest of this manual is organised in seven main parts:

Section 2 provides a brief overview of the key operations of SDSM. For a complete description of the model specification, interested readers should refer to the articles listed in the Bibliography (see below). Descriptions of the UKSDSM and Canadian Climate Impacts Scenarios (CCIS) data archives and file nomenclature are also provided in Section 2.

Sections 3 to 12 provide guidance on the practical implementation of the key functions in SDSM for downscaling regional climate change scenarios. Application of SDSM is illustrated using a hypothetical case study for Blogsville.

Section 13 provides a few cautionary remarks concerning the limitations of SDSM and appropriate usage. Users are strongly recommended to consider the issues raised here, before developing local scenarios using SDSM.

Next, a comprehensive Bibliography is supplied. This provides a general overview of downscaling as well as more detailed discussions of the technical basis of SDSM, example applications and comparisons with other downscaling methods.

Enhancements to SDSM since version 3.1 are listed in Appendix 1.

A trouble–shooting guide and outline of the most common pitfalls is provided in the form of a Frequently Asked Questions (FAQs) section in Appendix 2.

Finally, definitions of commonly used technical terms related to statistical downscaling are provided in a Glossary.

2 OVERVIEW OF SDSM STRUCTURE AND UKSDSM ARCHIVE

Downscaling is justified whenever GCM (or RCM) simulations of variable(s) used for impacts modelling are unrealistic at the temporal and spatial scales of interest, either because the impact scales are below the climate model's resolution, or because of model deficiencies. Downscaling may also be used to generate scenarios for exotic variables (such as urban heat island intensity) that can not be obtained directly from GCMs and RCMs. However, the host GCM must have demonstrable skill for large–scale variables that are strongly correlated with local processes. In practice, the choice of downscaling technique is also governed by the availability of archived observational and GCM data because both are needed to produce future climate scenarios.

The SDSM software reduces the task of statistically downscaling daily weather series into seven discrete steps:

1) quality control and data transformation;
2) screening of predictor variables;
3) model calibration;
4) weather generation (using observed predictors);
5) statistical analyses;
6) graphing model output;
7) scenario generation (using climate model predictors).

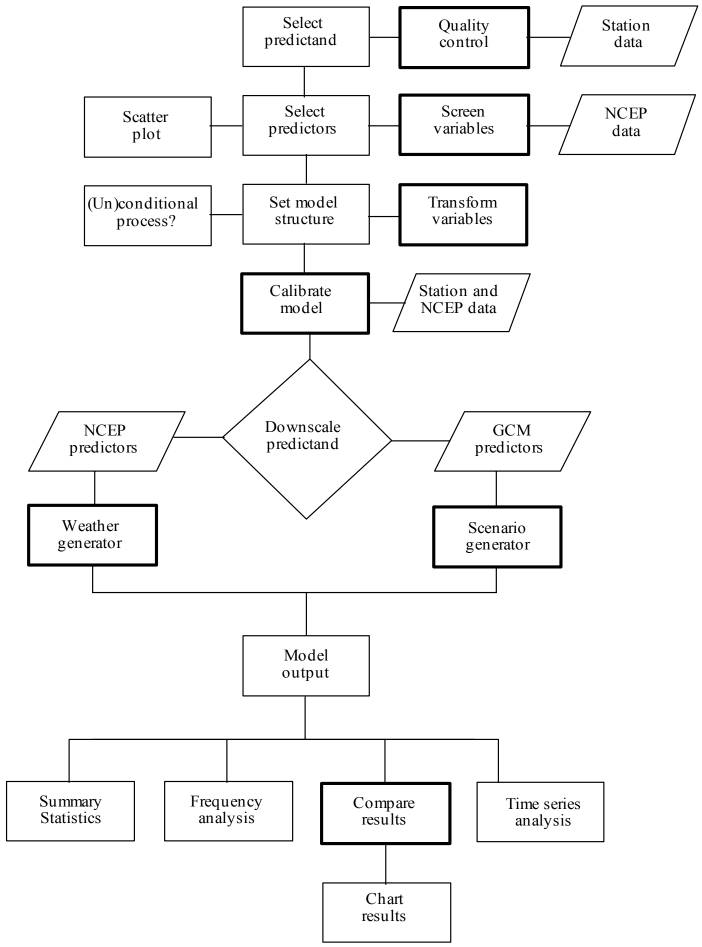


Figure 2.1 SDSM Version 4.2 climate scenario generation

Full technical details of SDSM (and downscaling prototypes) are provided in the Bibliography. Within the taxonomy of downscaling techniques, SDSM is best described as a hybrid of the stochastic weather generator and transfer function methods. This is because large–scale circulation patterns and atmospheric moisture variables are used to condition local–scale weather generator parameters (e.g., precipitation occurrence and intensity). Additionally, stochastic techniques are used to artificially inflate the variance of the downscaled daily time series to better accord with observations. To date, the downscaling algorithm of SDSM has been applied to a host of meteorological, hydrological and environmental assessments, as well as a range of geographical contexts including Africa, Europe, North America and Asia.

The following sections outline the software's seven core operations, along with the UKSDSM data archive and recommended file protocols.

2.1 Key functions of SDSM

As noted previously, SDSM performs seven key functions. The following paragraphs outline the purpose of each. Further technical explanation and User guidance are provided in Sections 3 to 12.

2.1.1 Quality control and data transformation

Few meteorological stations have 100% complete and/or fully accurate data sets. Handling of missing and imperfect data is necessary for most practical situations. Simple **Quality Control** checks in SDSM enable the identification of gross data errors, specification of missing data codes and outliers prior to model calibration.

In many instances it may be appropriate to transform predictors and/or the predictand prior to model calibration. The **Transform** facility takes chosen data files and applies selected transformations (e.g., logarithm, power, inverse, lag, binomial, etc).

2.1.2 Screening of downscaling predictor variables

Identifying empirical relationships between gridded predictors (such as mean sea level pressure) and single site predictands (such as station precipitation) is central to all statistical downscaling methods.

The main purpose of the **Screen Variables** operation is to assist the user in the selection of appropriate downscaling predictor variables. This is one of the most challenging stages in the development of any statistical downscaling model since the choice of predictors largely determines the character of the downscaled climate scenario. The decision process is also complicated by the fact that the explanatory power of individual predictor variables varies both spatially and temporally. **Screen Variables** facilitates the examination of seasonal variations in predictor skill.

2.1.3 Model calibration

The **Calibrate Model** operation takes a User–specified predictand along with a set of predictor variables, and computes the parameters of multiple regression equations via an optimisation algorithm (either dual simplex of ordinary least squares).

The User specifies the model structure: whether monthly, seasonal or annual sub–models are required; whether the process is unconditional or conditional. In unconditional models a direct link is assumed between the predictors and predictand (e.g., local wind speeds may be a function of regional airflow indices). In conditional models, there is an intermediate process between regional forcing and local weather (e.g., local precipitation amounts depend on the occurrence of wet–days, which in turn depend on regional–scale predictors such as humidity and atmospheric pressure).

2.1.4 Weather generator

The **Weather Generator** operation generates ensembles of synthetic daily weather series given <u>observed</u> (or NCEP re–analysis) atmospheric predictor variables. The procedure enables the verification of calibrated models (using independent data) and the synthesis of artificial time series for present climate conditions.

The User selects a calibrated model and SDSM automatically links all necessary predictors to model weights. The User must also specify the period of record to be synthesised as well as the desired number of ensemble members. Synthetic time series are written to specific output files for later statistical analysis, graphing and/or impacts modelling.

2.1.5 Data analysis

SDSM provides means of interrogating both downscaled scenarios and observed climate data with the **Summary Statistics** and **Frequency Analysis** screens.

In both cases, the User must specify the sub–period, output file name and chosen statistics. For model output, the ensemble member or mean, must also be specified. In return, SDSM displays a suite of diagnostics including monthly/ seasonal/ annual means, measures of dispersion, serial correlation and extremes.

2.1.6 Graphical analysis

Three options for graphical analysis are provided by SDSM 4.2 through the **Frequency Analysis**, **Compare Results**, and the **Time Series Analysis** screens.

The **Frequency Analysis** screen allows the User to plot extreme value statistics of the chosen data file(s). Analyses include Empirical, Gumbel, Stretched Exponential and Generalised Extreme Value distributions.

The **Compare Results** screen enables the User to plot monthly statistics produced by the **Summary Statistics** screen. Having specified the necessary input file, either bar or line charts may be chosen for display purposes. The graphing option allows simultaneous comparison of two data sets and hence rapid assessment of downscaled versus observed, or present versus future climate scenarios.

The **Time Series Analysis** screen allows the User to produce time series plots for up to a maximum of five variables. The data can be analysed as monthly, seasonal, annual or water year periods for statistics such as Sum, Mean, Maximum, Winter/Summer ratios, Partial Duration Series, Percentiles and Standardised Precipitation Index.

2.1.7 Scenario generation

Finally, the **Scenario Generator** operation produces ensembles of synthetic daily weather series given atmospheric predictor variables supplied by a climate model (either for present or future climate experiments), rather than observed predictors. This function is identical to that of the **Weather Generator** operation in all respects except that it may be necessary to specify a different convention for model dates and source directory for predictor variables. The input files for both the **Weather Generator** and **Scenario Generator** options need not be the same length as those used to obtain the model weights during the calibration phase.

2.2 UKSDSM data archive

As Figure 2.1 indicates, the SDSM procedure begins with the preparation of coincident predictor and predictand data sets. Although the predictand is typically an individual daily weather series, obtained from meteorological observations at single stations (e.g., daily precipitation, maximum or minimum temperature, hours of sunshine, wind speed, etc.), the methodology is applicable to other environmental variables (e.g., air quality parameters, wave heights, snow cover, etc.). In any event, *these data must be supplied by the User in SDSM format* (see also Section 2.4.2). This is single column, text only, data beginning 1st January 1961, if necessary padded with the **Missing Data Identifier**.

Assembly of the candidate predictor suite can be a far more involved process entailing data extraction, re–gridding and normalisation techniques. For this reason, SDSM is supplied with a prepared set of daily predictor variables for selected grid boxes covering the British Isles (Figure 2.2) and globally for all land areas via the web (Section 2.5). The User simply locates the required grid box and data source from the UKSDSM or online archive. As Figure 2.2 shows the UK is represented by nine grid boxes each measuring 2.5º latitude by 3.75º longitude, corresponding to the grid co–ordinate system of the Hadley Centre's coupled ocean–atmosphere GCMs (see below). Of the nine cells, six are land, and three are ocean. To obtain more realistic estimates of forcing over land areas that are represented by ocean grid boxes in the GCM, data from the two nearest land cells were averaged. For example, predictor variables for Southwest England (SW) are the average of data from the Wales (WA) and Southern England (SE) grid boxes.

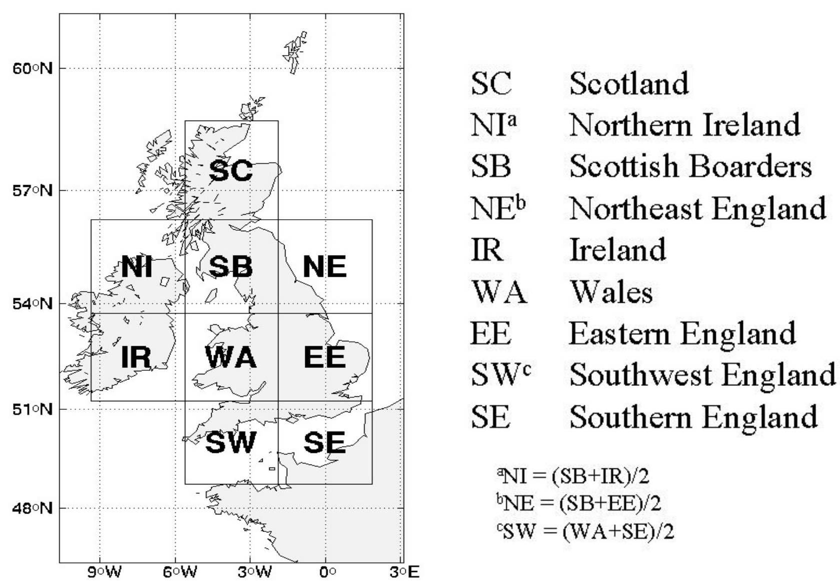


Figure 2.2 Location and nomenclature of the UK grid boxes in the SDSM archive.

For model calibration, the source is the National Centre for Environmental Prediction (NCEP) re–analysis data set. The data were re–gridded to conform to the grid system of HadCM3 (Figure 2.2). All predictors (with the exception of the geostrophic wind direction, see below) were normalised with respect to the 1961 to 1990 average. However, daily predictors are also supplied for the period 1961–2000.

For downscaling future climate scenarios four sets of GCM output are available: HadCM2, HadCM3, CGCM2, CSIRO. Three emission scenarios are: available: the greenhouse gas only experiment with CO_2 compounded annually by 1% per year (HadCM2 only), the two SRES scenarios A2 and B2 produced by greenhouse gas, sulphate aerosol, and solar forcing (HadCM3, CSIRO, CGCM2).

2.3 UKSDSM predictors

Table 2.1 lists the daily predictor variables held in the UKSDSM data archive. Ideally, candidate predictor variables should be physically and conceptually sensible with respect to the predictand, strongly and consistently correlated with the predictand, and realistically modelled by GCMs. For precipitation downscaling, it is also recommended that the predictor suite contain variables describing atmospheric circulation, thickness, stability and moisture content. In practise, the choice of predictor variables is often constrained by data availability from GCM archives. The predictors in Table 2.1, therefore, represent a compromise between maximum overlap between NCEP and GCM archives, as well as a range of choice for downscaling.

Daily variable	Code	NCEP 1961–2000	HadCM2 GG 1961–2099	HadCM3 SRES 1961–2099	CGCM2 SRES 1961–2099	CSIRO SRES 1961–2099
Precipitation (mm)	prec		×	×	×	×
Maximum temperature (ºK)	tmax		×	×	×	×
Minimum temperature (ºK)	tmin		×	×	×	×
Mean temperature	temp	×		×	×	×
Mean sea level pressure	mslp	×		×	×	×
500 hPa geopotential height	p500	×		×	×	×
850 hPa geopotential height	p850	×		×	×	×
Near surface relative humidity	rhum	×		×	×	×
Relative humidity at 500 hPa height	r500	×		×	×	×
Relative humidity at 850 hPa height	r850	×		×	×	×
Near surface specific humidity	shum	×		×	×	×
Geostrophic airflow velocity	**_f	×	×	×	×	×
Vorticity	**_z	×	×	×	×	×
Zonal velocity component	**_u	×	×	×	×	×
Meridional velocity component	**_v	×	×	×	×	×
Wind direction	**th	×	×	×	×	×
Divergence	**zh	×	×	×	×	×

Table 2.1 Daily variables held in the UKSDSM data archive (denoted by ×). **Bold** type indicates variables that have **not** been normalised and are provided for comparative purposes. *Italics* indicate secondary (airflow) variables derived from pressure fields (surface, 500 and 850 hPa).

2.4 SDSM file protocols

For convenience, the SDSM file protocol is described in two parts. Firstly, the file name system and file structure of the UKSDSM archive. Secondly, the meta–data and output files produced by SDSM more generally.

2.4.1 UKSDSM file structure and nomenclature

Figure 2.3 shows how the directory structure of the UKSDSM data archive relates to ancillary file systems in SDSM. The UKSDSM archive is organised into three levels. At the highest level are the data sources: presently NCEP, HadCM2, HadCM3, CSIRO or CGCM2. At the second level, are the nine cells shown in Figure 2.2. At the third level, are files containing individual predictor variables.

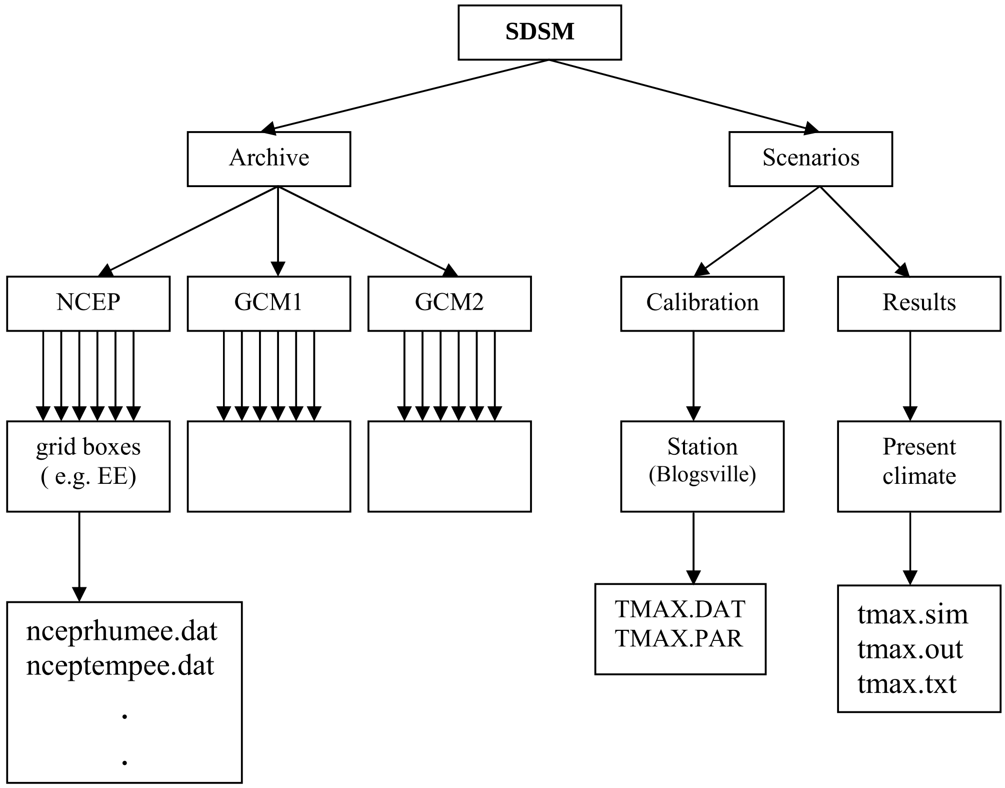


Figure 2.3 SDSM file structure with example file names
(see Table 2.2 for definitions of file name extension).

Each file in the archive complies with a generic nomenclature of the form

[*source*] [*variable*] [*grid box*] . dat

The **source** is denoted by characters 1–4, the **variable** name by characters 5–8, and the **grid box** by characters 9–10. All files have the extension .dat, for example, the file name

nceprhumee.dat

indicates that the source is NCEP [*ncep*], the variable is near surface relative humidity [*rhum*], and the grid box is Eastern England [*ee*]. Similarly, the file name

h3b2p8_zsw.dat

indicates that the source is HadCM3, SRES scenario B2 [*h3b2*], the variable is vorticity computed at the 850 hPa geopotential height [*p8_z*], and the grid box is Southwest England [*sw*]. Alternatively, the file name

h2ggp_thsb.dat

indicates that the source is HadCM2, greenhouse gas only experiment [*h2gg*], the variable is surface wind direction [*p_th*], and the grid box is Scottish Boarders [*sb*].

2.4.2 SDSM file name protocol

With the above prerequisites in mind, Table 2.2 lists the file name extensions employed by SDSM, and Figure 2.3 shows the associated directory structures.

All input and output files are text only format. Individual predictor and predictand files (one variable to each file, time series data only) are denoted by the extension ***.dat**. The ***.PAR*** file records meta–data associated with the model calibration, model weights, and measures of "goodness–of–fit" (percentage explained variance and standard error of the model). The ***.SIM*** file records meta–data associated with every downscaled scenario (e.g., number of predictor variables, ensemble size, period, etc.), and the ***.OUT*** file contains an array of daily downscaled values (one column for each ensemble member, and one row for each day of the scenario). Finally, ***.TXT*** files are created whenever statistical analyses are undertaken by SDSM. These files record summary statistics for individual ensemble members or for the ensemble mean, and are accessed by bar/line chart options. The data format also enables convenient export to other graphing software and spreadsheets.

Table 2.2 SDSM file names and recommended directory structure

Extension	Explanation	Directory
*.DAT	Observed daily predictor and predictand files employed by the **Calibrate** and **Weather Generator** operations (input).	SDSM/Scenarios/ Calibration
*.PAR	Meta–data and model parameter file produced by the **Calibrate** operation (output) and used by the **Weather Generator** and **Generate Scenario** operations (input).	SDSM/Scenarios/ Calibration
*.SIM	Meta–data produced by the **Weather Generator** and **Generate Scenario** operations (output).	SDSM/Scenarios/ Results
*.OUT	Daily predictand variable file produced by the **Weather Generator** and **Generate Scenario** operations (output).	SDSM/Scenarios/ Results
*.TXT	Information produced by the **Summary Statistics** and **Frequency Analysis** operations (output).	SDSM/Scenarios/ Results

2.5 Obtaining SDSM predictors online

SDSM predictors may be obtained for any global land area courtesy of a data portal maintained by the Canadian Climate Impacts Scenarios Group. The web-site is accessed from: http://www.cics.uvic.ca/scenarios/index.cgi?Scenarios

Having registered by e-mail address, the User then selects predictors from the available GCMs (currently HadCM3 and CGCM2), given the latitude and longitude of the nearest grid-box(es) to the study region. All data files, including NCEP predictors, may then be downloaded directly to Users' PC for immediate deployment by SDSM.

3. GETTING STARTED

To launch SDSM, click on the **Start** button on the Windows desktop, then on **Programs**, and then on **SDSM** (which will appear as a small rain cloud on the list of available programs). The following screen will appear:

Figure 3.1 The SDSM "splash" screen.

Click on **Start** to continue to the SDSM main menu (Figure 3.2). If you do not wish the splash screen to appear in future (ie the main menu screen will appear upon starting SDSM) click the tick box by 'Do not show this splash screen again'. If further information is required at any time, click on the **Help** button at the top of each screen (the User may then search the Help Contents by key word or task).

SDSM is navigated by selecting appropriate buttons from the bar at the top of each screen. These are arranged in the same logical order as key functions of SDSM.

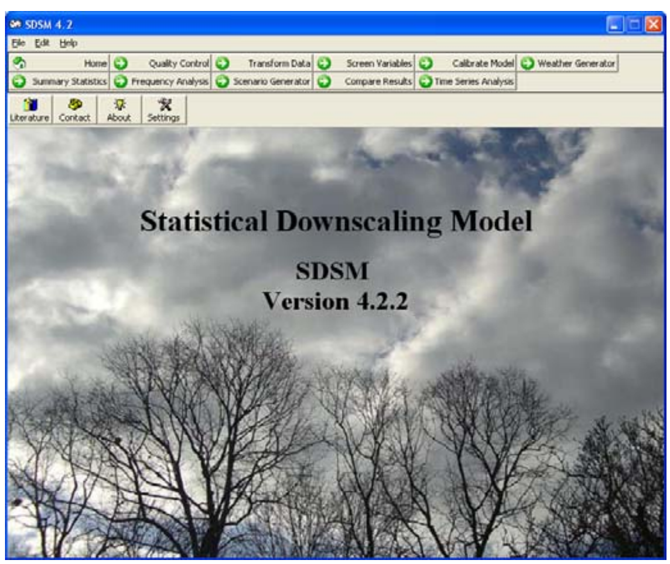


Figure 3.2 Main menu of SDSM 4.2

Before downscaling, the User should check the date ranges, type and integrity of all input data. To establish the working environment click on the spanner symbol at the top of the main menu (or at the top of any other screen) to access the **Settings** screen (Figure 3.3).

3.1 Settings

The **Settings** screen may be accessed throughout SDSM. The following global preferences are available:

Year Length: The default "Calendar (366)" allows 29 days in February every fourth year (i.e., leap years) and should be used with observed data. The alternatives allow for different numbers of days in GCM data. For example, CGCM2 and CSIRO have 365 days and no leap years, whereas HadCM2 and HadCM3 have model years consisting of 360 days. WARNING: Failure to set this parameter correctly can lead to system errors due to insufficient data or the production of non-sensible output.

Standard Start/End Date: Enter the global default start and end date for all input data. These dates will appear throughout the operation of SDSM, but may be updated from any screen.

Allow Negative Values: The default allows simulation of negative values by unconditional processes in the downscaling model (e.g., for minimum temperature); deselection truncates values at zero (e.g., for sunshine hours). Conditional processes (e.g., rainfall amounts) are unaffected by this button.

Event Threshold: For some variables it is necessary to specify an event threshold. For example, when calibrating daily precipitation models, the parameter might be set to 0.3 mm/day to treat trace rain days as dry days. Similarly, the threshold for sunny versus cloudy days might be set at 1.0 hours/day to discriminate between overcast and sunny conditions.

Missing Data Identifier: This is the code assigned to missing data in all input series. Whenever SDSM encounters this code the value will be skipped (e.g., during model calibration, or calculation of summary statistics). The default is –999.

Random Number Seed: Ensures that the random sequences produced by **Weather Generator** (Section 7) and **Scenario Generator** (Section 10) are different each time the model is run. If replicate experiments are preferred, the check box should be deselected.

Default File Directory: Allows the user to select a default directory that is accessed by all screens when first searching for files.

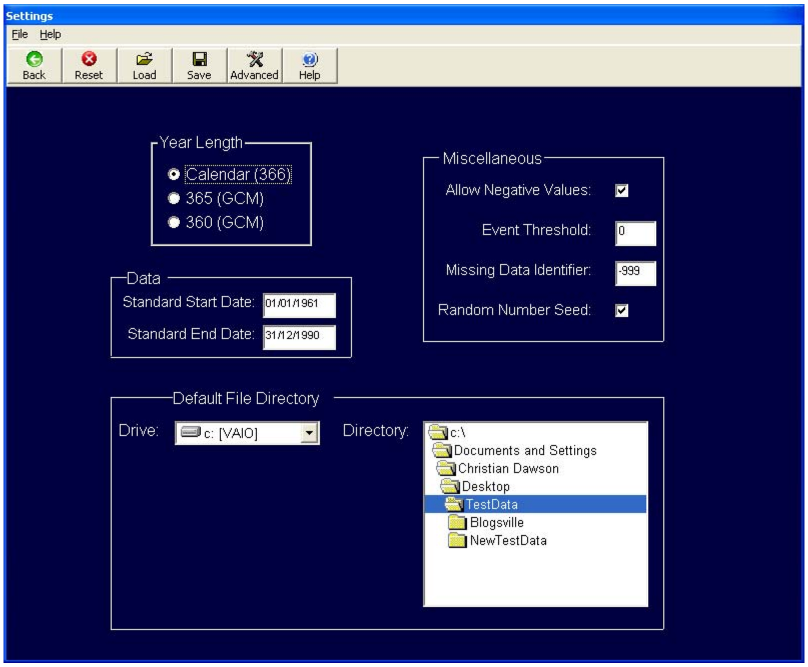


Figure 3.3 The **Settings** screen

3.2 Advanced settings

The advanced settings are accessed from the **Settings** screen by clicking on the **Advanced** button at the top of the screen. The **Advanced Settings** screen allows the User to change and save further downscaling model preferences (Figure 3.4):

Model Transformation: Specifies the transformation applied to the predictand in conditional models. The default (None) is used whenever the predictand is normally distributed (as is often the case for daily temperature). The alternatives (Fourth root, Natural log and Inverse Normal) are used whenever data are skewed (as in the case of daily precipitation). Note that the Inverse Normal transformation employs **conditional resampling** of the observed predictand (see Wilby et al. 2003). The transformation type is recorded in *.PAR and *.SIM files to ensure that data are consistently handled during subsequent scenario and data analysis routines.

Variance Inflation: Controls the magnitude of variance inflation in downscaled daily weather variables. This parameter changes the variance by adding/reducing the amount of "white noise" applied to model estimates of the local process. The default value produces approximately normal variance inflation (prior to any transformation). Larger values increase the variance of downscaled properties. Variance inflation is de–activated by setting the parameter to zero. Note that for Fourth root and Natural log **Model Transformation** (see above), this parameter also affects changes in the mean estimate.

Bias Correction: Compensates for any tendency to over– or under–estimate the mean of conditional processes by the downscaling model (e.g., mean daily rainfall totals). The default value is 1.0, indicating no bias correction.

Conditional Selection: Adjusts the way in which conditional processes (e.g., rainfall amounts) are sampled. The default (Stochastic) allows the outcome to be entirely based on chance. Fixed Threshold allows the User to increase the chance of a conditional event (by setting the threshold closer to zero), or reducing the chance (by setting closer to 1.0).

Optimisation Algorithm: SDSM 4.2 provides two means of optimising the model – Dual Simplex (as in earlier versions of SDSM) and Ordinary Least Squares. Although both approaches give comparable results, Ordinary Least Squares is much faster. The User can also select a Stepwise Regression model by ticking the appropriate box. Stepwise regression works by progressively including more variables and selecting the most parsimonious model of the predictand according to one of two metrics – either Akaike's Information Criterion (AIC) or the Bayesian Information Criterion (BIC).

Settings File: Locates standard and advanced settings held in a User defined reference file and directory. A new or updated settings file is created whenever the **Save** button is clicked at the top of the screen. The C:\SDSM.INI settings file is automatically loaded whenever SDSM starts up.

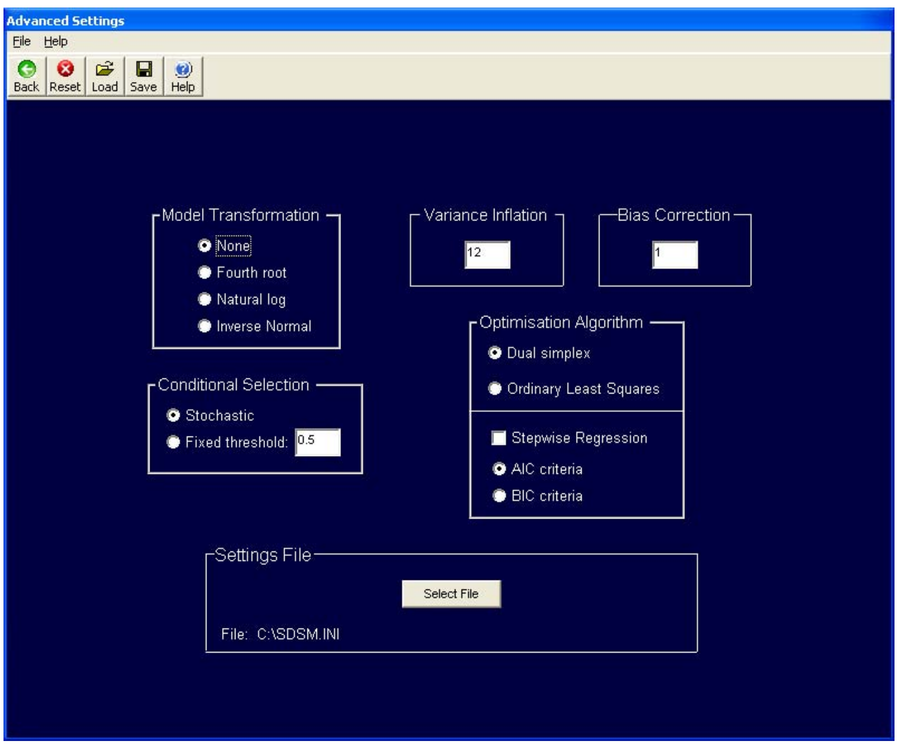


Figure 3.4 The **Advanced Settings** screen

Press **Reset** at any time to reload the original settings, or **Back** to return to the **Settings** screen, followed by **Back** again to return to the last open screen.

4. QUALITY CONTROL AND DATA TRANSFORMATION

Few meteorological stations have complete and/or fully accurate data sets. Handling of missing and imperfect data is necessary for most practical situations. In some cases it may also be necessary to transform data prior to model calibration. SDSM enables both quality control and data transformation.

4.1 Quality control

To check an input file for missing data and/or suspect values, click on the **Quality Control** button from any of the main screens. The following screen will appear:

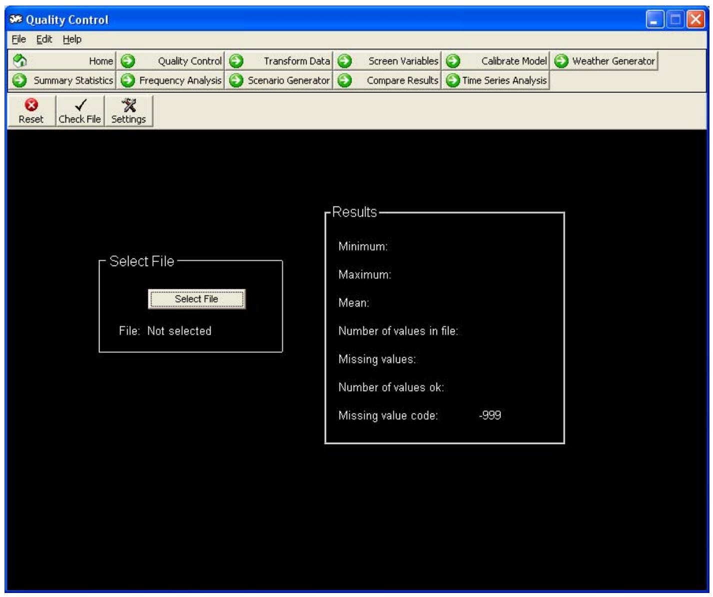


Figure 4.1 The **Quality Control** screen

Click on the **Select File** button. An **Open** file window will appear – browse through until you have located the directory and file to be checked – in this example the Blogsville maximum daily temperature, TMAX.DAT. Click on the required data file then on **Open**. To activate the quality control procedure, click on the **Check File** button at the top of the screen. The following confirmation will appear:

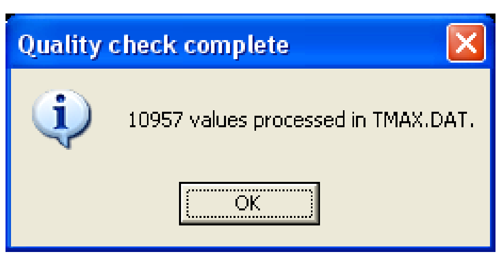


Figure 4.2 The **Quality check complete** dialogue box.

Click on the **OK** button to view the quality control information. In this example, there are 10957 values with no missing data (i.e., no missing value codes of

–999 were detected). The data range from –6.7 to 34.8 ºC, and have a mean of 13.1871ºC (see Figure 4.3). Click on the **Reset** button to clear the screen entries, or select another file to perform a new quality check.

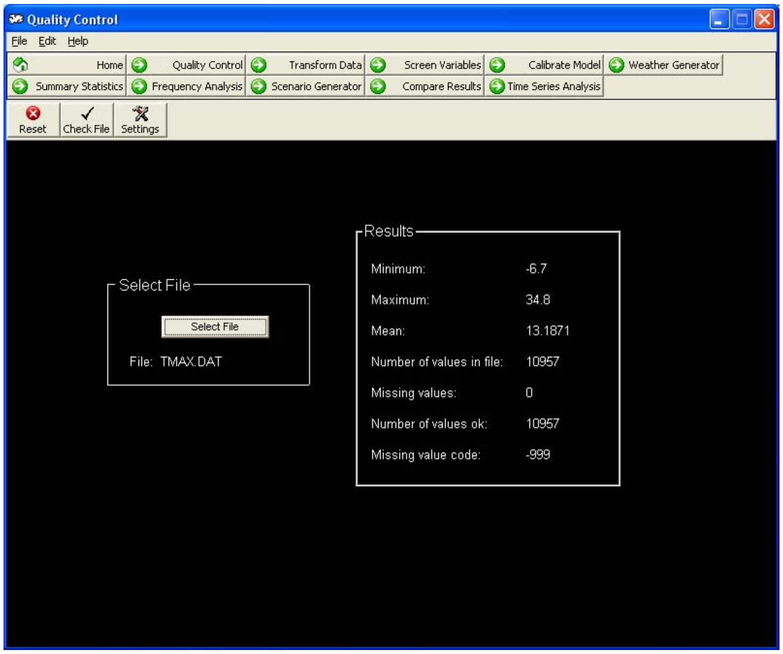


Figure 4.3 Results of the **Quality Control** check for TMAX.DAT

4.2 Data transformation

To transform data, click on the **Transform Data** button from any of the main screens. The following screen will appear:

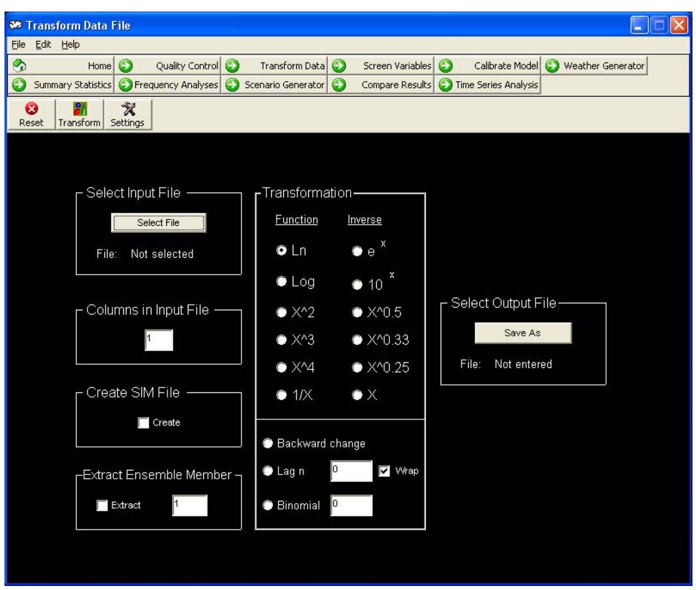


Figure 4.4 The **Transform Data File** screen

Click on the **Select Input File** button. An **Open** file window will appear. Browse through until you have located the directory and file to be transformed – for example the surface vorticity over Eastern England, *ncepp__zee.dat*. Click on the required file. If there is more than one column of input data (as in the case of an ensemble simulation produced by the **Weather Generator** or **Generate Scenario** functions, see Sections 7 and 10) enter the appropriate number in the **Columns in Input File** box. To enable transformed data with multiple columns to be handled by the **Analyse Data** function (Section 8), check the box under **Create SIM File**.

If the User wishes to extract a single ensemble member from a multi-column data file check the **Extract** box on this screen. Enter the number of the ensemble member required and the data will be written to the selected Save File. Note, in this case, no transformation is applied to the extracted member.

Select the **Transformation** by checking the appropriate button. Available transformations include: natural logarithms and log10, squares, cubes, fourth powers, inversion, lag interval and binomial, together with the inverse transformations of the above where appropriate. If **Wrap** is selected (for **Lag n**) the last value is used as the first value in the lag transformation; otherwise the **Missing Data Identifier** is inserted (note that a negative lag value will shift the data forward, a positive lag value will shift the data back). The **Backward change** button is used to compute differences between successive days. All transformations can be applied to standard predictor variables prior to **Model Calibration** (Section 6), to produce non–linear regression models (e.g., use power transformations for polynomial models).

For the Eastern England data, select **Lag n**, enter "–1" in the box, and check the **Wrap** box (which will produce a lag–1 series of the variable with no missing data). Click on the **Select Output File** button. An **Open** file window will appear – browse through until the required directory is located, enter the **Filename** for transformed data, in this case *ncepzlagee.dat* (i.e., vorticity on previous days), then click on **Save**. WARNIG: The name used for transformed files MUST comply fully with the protocol described in Section 2.4.1.

To activate the procedure, click on the **Transform** button at the top of the screen. The following confirmation will appear:

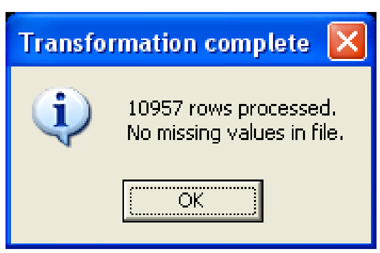


Figure 4.4 The **Transformation complete** dialogue box

Click on the **OK** button to return to the **Transform Data File** screen. Click on the **Reset** button to clear the screen entries, or to perform a new transformation.

5. SCREENING DOWNSCALING PREDICTOR VARIABLES

Identifying empirical relationships between gridded predictors (such as mean sea level pressure) and single site predictands (such as station precipitation) is central to all statistical downscaling methods and is often the most time consuming step in the process. The purpose of the **Screen Variables** option is to assist the User in the choice of appropriate downscaling predictor variables for model calibration (Section 6). SDSM performs three supporting tasks: seasonal correlation analysis, partial correlation analysis, and scatterplots. Ultimately, however, the User must decide whether or not the identified relationships are physically sensible for the site(s) and predictands in question. In this matter, there is no substitute for local knowledge.

To investigate potentially useful predictor–predictand relationships, click on the **Screen Variables** button from any of the main screens. The following screen will appear:

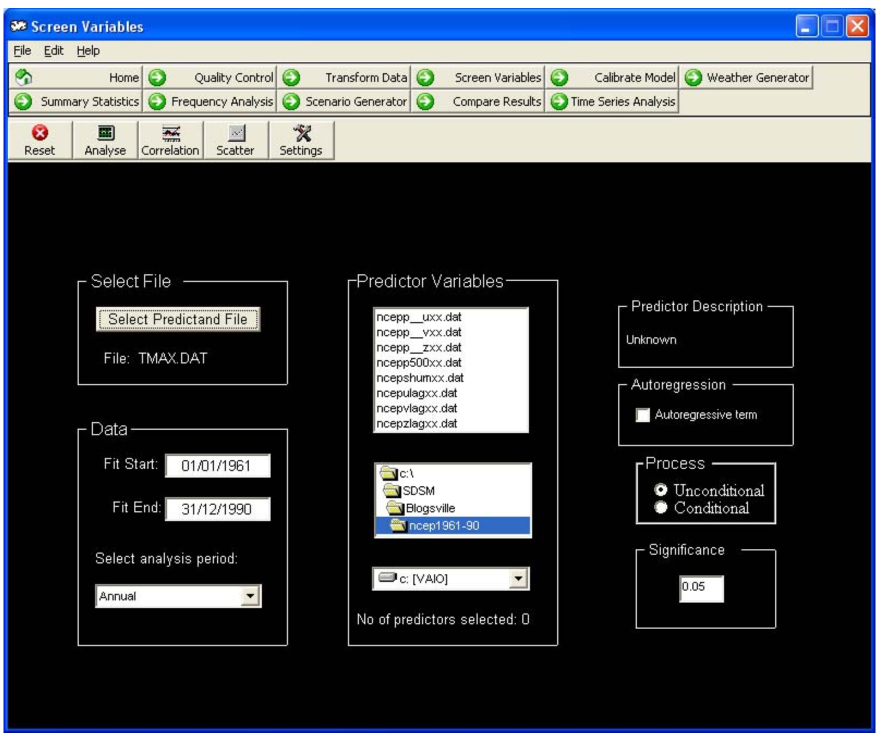


Figure 5.1 Illustration of the **Screen Variables** screen using daily maximum temperatures for Blogsville, 1961–1990.

5.1 Setup

The first step in the **Screen Variables** operation is the selection of the predictand and predictor files. The predictand file (e.g., observed daily maximum temperature, daily precipitation totals, etc.) must be supplied by the User, in SDSM format (see Section 2.4). Click on the **Select Predictand File** button. An **Open** file window will appear – browse through until the appropriate directory has been located. Click on the predictand data file – for example, the maximum daily temperature at Blogsville, *TMAX.DAT*, located in C:\SDSM\Blogsville\observed1961-90.

Follow a similar procedure, locate and select the desired **Predictor Variables** by choosing the correct drive from the pull down window in the centre of the screen. The directories available on this drive will then appear in the window directly above the drive window. Browse through again until the appropriate directory is located. All *.DAT files in this directory are then listed in the window above. To select a predictor, simply click on the file name – it will be highlighted in blue. A brief definition of the chosen variable is given in the **Predictor Description** window. To deselect a file, click on it again, and it will no longer be highlighted. The number of predictor variables chosen is shown beneath this window (up to a maximum of 12).

The **Data** menu on the left–hand side of the **Screen Variables** screen allows the start and end dates of the analysis period to be changed. The default dates are held in the **Settings** screen (see Section 3.1), in this case 1961–1990. If the start and end dates lie outside the permissible range, the User will be prompted to enter new values. The User must also choose the seasonal subset from the pull down window under **Select analysis period**. The available options are Annual (no seasonal sub–setting), Winter (December–February), Spring (March–May), Summer (June–August), Autumn (September–November), and individual months.

Three more actions are necessary before the analysis can take place. Firstly, the type of **Process** must be specified. If the predictor–predictand process is <u>not</u> regulated by an intermediate process (as in the case of maximum temperature) then click on **Unconditional**, otherwise select **Conditional** (as with precipitation where amounts depend on wet–day occurrence). Secondly, amend the **Significance Level** as required. This value is used to test the significance of predictor–predictand correlations. The default is p<0.05 (5%). Finally, if the User wants an autoregressive term to be included in the calculations the Autoregression option should be selected.

Once the above have been specified, SDSM is ready to analyse the chosen predictor–predictand relationship(s), for specified sub–period(s).

5.2 Temporal variations in predictor strength

The **Analyse** button is used to investigate the percentage of variance explained by specific predictand–predictor pairs. The strength of individual predictors often varies markedly on a month by month basis (see Figure 5.2). The User should, therefore, be judicious concerning the most appropriate combination(s) of predictor(s) for a given season and predictand. As stated above, the local knowledge base is also invaluable when determining sensible combinations of predictors.

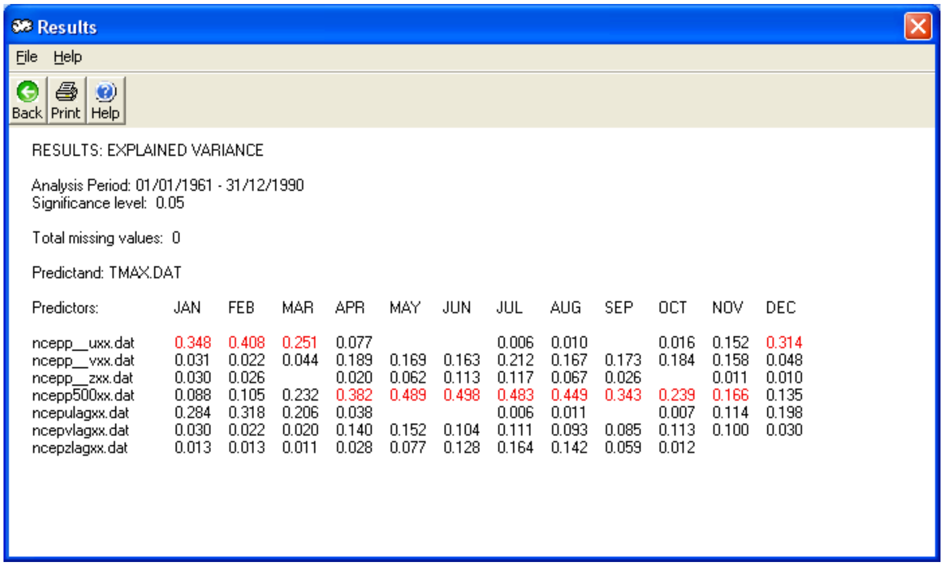


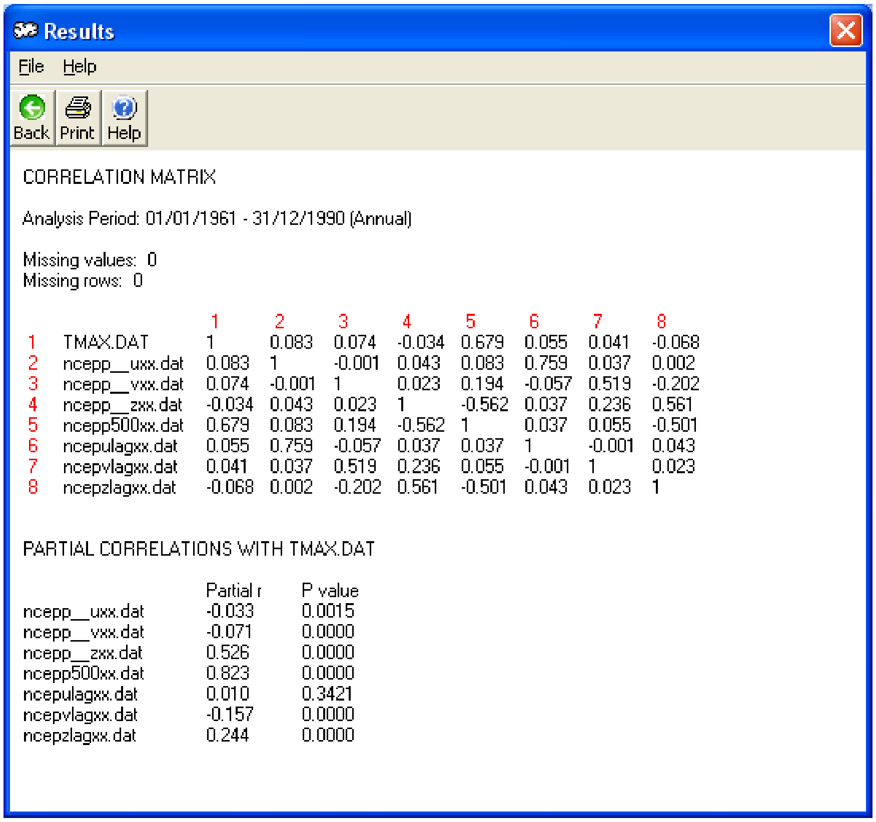
Figure 5.2 The **Results** screen for the Blogsville example. The strongest correlation in each month is shown in red, indicating that the relationship between maximum temperature and *p500* and *p__u* are most important. Blanks represent insignificant relationships at the chosen **Significance Level**

For the Blogsville example, select maximum daily temperatures as the predictand (*TMAX*), and the following predictor files: *p500, p__u, p__v, and p__z* (see Table 2.1). In addition, use the **Transform** facility (Section 4.2) to create lagged values (one day i.e., lag the data by -1) for the surface airflow indices. The predictand does not depend on an intermediate occurrence process, so **Unconditional** is checked under the **Process** option. Use the default dates for the **Data** option (i.e., 1961–1990), and choose **Annual** under **Select Analysis Period**. Use the default **Significance Level** (i.e., 0.05) then click on the **Analyse** button at the top of the **Screen Variables** menu. The results in Figure 5.2 would suggest that *p500* is a potentially useful predictor for April through October maximum temperature, and *p__u* for December through March.

5.3 Correlation matrix

The **Correlation** button is used to investigate inter–variable correlations for specified sub–periods (annual, seasonal or monthly). SDSM also reports partial correlations between the selected predictors and predictand. These statistics help to identify the amount of explanatory power that is unique to each predictor.

For the Blogsville example, use the same predictand and predictors as in Section 5.2, but for Select analysis period choose February from the drop down list. Then click on the **Correlation** button at the top of the **Screen Variables** menu. The results are shown in Figure 5.3.



Results

File Help

Back Print Help

CORRELATION MATRIX

Analysis Period: 01/01/1961 - 31/12/1990 (Annual)

Missing values: 0
Missing rows: 0

		1	2	3	4	5	6	7	8
1	TMAX.DAT	1	0.083	0.074	-0.034	0.679	0.055	0.041	-0.068
2	ncepp__uxx.dat	0.083	1	-0.001	0.043	0.083	0.759	0.037	0.002
3	ncepp__vxx.dat	0.074	-0.001	1	0.023	0.194	-0.057	0.519	-0.202
4	ncepp__zxx.dat	-0.034	0.043	0.023	1	-0.562	0.037	0.236	0.561
5	ncepp500xx.dat	0.679	0.083	0.194	-0.562	1	0.037	0.055	-0.501
6	ncepulagxx.dat	0.055	0.759	-0.057	0.037	0.037	1	-0.001	0.043
7	ncepvlagxx.dat	0.041	0.037	0.519	0.236	0.055	-0.001	1	0.023
8	ncepzlagxx.dat	-0.068	0.002	-0.202	0.561	-0.501	0.043	0.023	1

PARTIAL CORRELATIONS WITH TMAX.DAT

	Partial r	P value
ncepp__uxx.dat	-0.033	0.0015
ncepp__vxx.dat	-0.071	0.0000
ncepp__zxx.dat	0.526	0.0000
ncepp500xx.dat	0.823	0.0000
ncepulagxx.dat	0.010	0.3421
ncepvlagxx.dat	-0.157	0.0000
ncepzlagxx.dat	0.244	0.0000

Figure 5.3 The **Results** screen for the Blogsville example. Partial correlations indicate that *p500* and *p__z* have the strongest association with *TMAX* once the influence of all other predictors has been removed.

5.4 Scatterplot

The **Scatter** button is used for visual inspections of inter–variable behaviour for specified sub–periods (annual, seasonal or monthly). The resultant scatterplot(s) indicate the nature of the association (linear, non–linear, etc.), whether or not data transformation(s) may be needed, and the importance of outliers.

For the Blogsville example, select *TMAX* as the predictand, *p__u* as the predictor file, and February under **Select analysis period** (following the results in Figure 5.2). Check that all other predictors have been deselected, and that **Unconditional** is selected under **Process**. (Note that if **Conditional** is selected, all values less than or equal to the **Event Threshold** in **Settings** are excluded from the plot). Click on the **Scatter** button at the top of the **Screen Variables** menu. The results are shown in Figure 5.4.

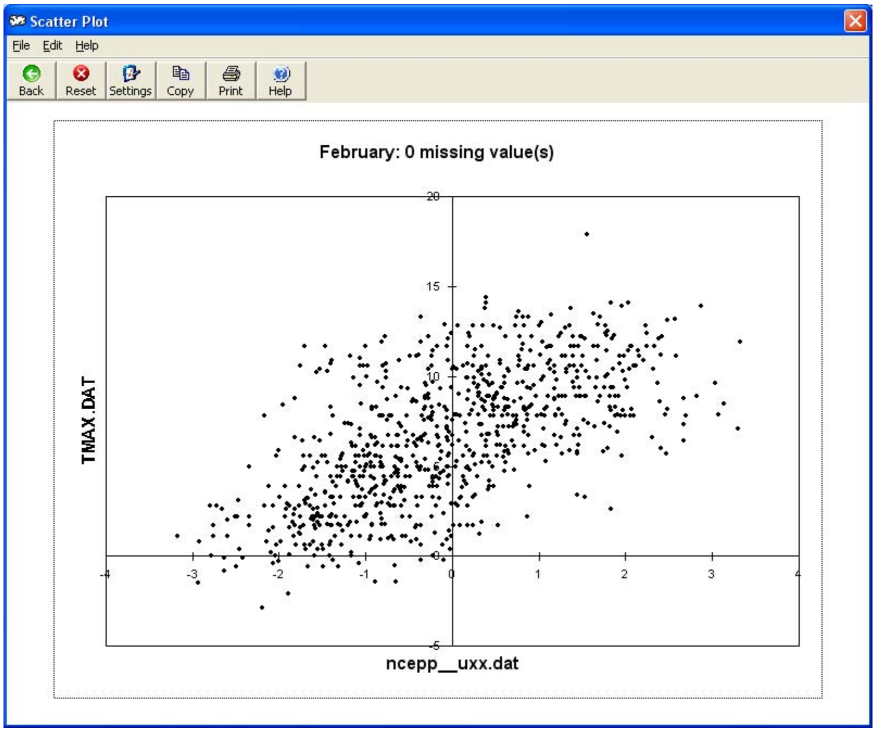


Figure 5.4 The **Scatterplot** for the Blogsville example, showing the association between *TMAX* and *p__u* in February. The results suggest that during February, higher maximum daily temperatures are associated with stronger westerly airflows.

The presentation quality of the **Scatterplot** may be customized, as required, by doubling clicking on any of the axis legends, titles or data points. Additional windows enable changes to be made to chart font, style, size, colour, etc.

To incorporate the **Scatterplot** in a Word document, first use the **Copy** button at the top of the screen, then in Word use Paste Special (Picture).

6. MODEL CALIBRATION

The **Calibrate Model** process constructs downscaling models based on multiple regression equations, given daily weather data (the predictand) and regional–scale, atmospheric (predictor) variables. The parameters of the regression model are written to a standard format file with the extension *.PAR, along with meta–data recording details of the calibration period, model type, predictors used, etc (see Section 6.4). SDSM optimises the model using either dual simplex or ordinary least squares optimisation (see **Advanced Settings** in Section 3.2).

The User specifies the model structure: whether monthly, seasonal or annual sub–models are required; whether the process is unconditional or conditional. In unconditional models a direct link is assumed between the predictors and predictand (e.g., local wind speeds may be a function of regional airflow indices). In conditional models, there is an intermediate process between regional forcing and local weather (e.g., local precipitation amounts depend on wet–/dry–day occurrence, which in turn depend on regional–scale predictors such as humidity and atmospheric pressure). Furthermore, it is possible to apply standard transformations to the predictand in conditional models (see Section 3.2), and/or to specific predictors (see Section 4.2).

To access the model building facility, click the **Calibrate Model** button at the top of any screen. The following screen will appear:

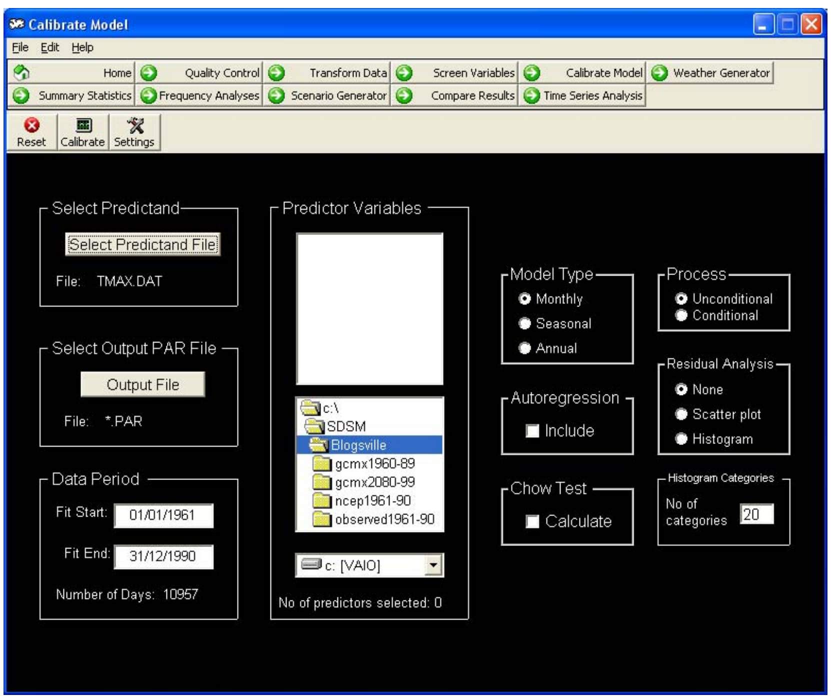


Figure 6.1 The **Calibrate Model** screen

6.1 File handling

To begin model building, click on the **Select Predictand File** button in the top left–hand corner. An **Open** file window appears; browse through until the correct directory and file are reached, then click on the appropriate file name – for example, the maximum daily temperature at Blogsville, *TMAX.DAT*. The name of the file will then appear beneath the button.

Follow a similar procedure to locate and select the desired predictor variables by choosing the correct drive from the pull down window in the centre of the screen. The directories available on this drive will then appear in the window directly above the drive window. For example, locate the C:\SDSM\Blogsville\NCEP1961-90 directory. All *.DAT files in this directory are then listed in the window above. To select a predictor, simply click on the file name – it will be highlighted in blue. To deselect a file, click on it again, and it will no longer be highlighted. The number of predictor variables chosen is shown beneath this window.

The **Data Period** menu on the left–hand side of the **Calibrate Model** screen allows the start and end dates of the analysis period to be changed. The default dates are held in the **Settings** screen (Section 3.1), in this case 1961–1990. If the start and end dates lie outside the permissible range, the User will be prompted to enter new values. Ideally, the model should be calibrated using part of the available data, withholding the remainder for independent model validation (see Sections 7 and 8).

To specify the name of the output parameter (*.PAR) file, click on the **Output File** button. An **Output PAR file** window appears. For maximum convenience, make sure that the parameter file is saved in the same directory as the predictand files, in this case, C:\SDSM\Blogsville\Observed1961-90. Enter an appropriate file name in the **File name** box then click on the **Save** button. The name of the parameter file will then be displayed beneath the **Output File** button, for example, TMAX61-75.PAR if data from 1961 – 75 are used for calibration.

6.2 Model type

To determine the temporal resolution of the downscaling model check either **Monthly**, **Seasonal** or **Annual** under the **Model Type** box. In **Monthly** models, different model parameters are derived for each month. In **Seasonal** models, all months in the same season (e.g., December, January and February for winter) have the same model parameters. In **Annual** models, all months have the same parameters (i.e., there is no attempt to specify intra–annual variations in parameter values).

Next, indicate whether the downscaling process should be **Unconditional** or **Conditional** by checking the appropriate option in the **Process** box. Note that for conditional processes in which the distribution of predictand values is skewed, it is possible to apply one of several transformations in **Advanced Settings** (see Section 3.2). For example, the Fourth root might be selected for daily precipitation amounts.

If an autoregressive component is required in the model (i.e., a lagged predictand is used as a predictor), the User should select Include within the **Autoregression** box.

SDSM 4.2 can calculate residual statistics and display these on either a scatter diagram or in a histogram. The scatter diagram plots the residuals against the modelled predictor while the histogram shows the distribution of the residuals. These two charts are generated after the summary statistics of the modelling are presented to the User. The number of bars in the histogram can be adjusted by altering the value in the **Histogram Categories** box.

The User can also view the Chow test statistics (for model stationarity) by checking the appropriate box. The Chow test is an optional test as it can slow down the modelling process significantly, particularly if a Dual Simplex optimisation is selected. Finally, click the **Calibrate** button at the top of the screen.

6.3 Blogsville example

For the Blogsville example, five predictor files (*p__u*, *p__z*, *p500*, *vlag* and *zlag*) might be selected to downscale daily maximum temperatures, *TMAX* (see Figure 5.3). There is clearly a seasonal cycle in the regional forcing (Figure 5.2), so **Monthly** is checked in the **Model Type** box. The **Unconditional** option is checked in the **Process** box because a direct link is assumed to exist between the regional–scale predictors and local temperature. The date range in the **Data** menu is set at 1961 to 1975, ensuring that the second half of the data (i.e., 1976 to 1990) is retained for model validation. Select *Calculate* to derive the Chow statistics for the model and in **Advanced Settings** the optimisation algorithm is set to *Ordinary Least Squares*. Save the output results to TMAX61-75.PAR.

Once the appropriate selections have been made, click on the **Calibrate** button. The process may take several seconds and on completion a summary screen will appear (**Calibration Results**, see Figure 6.2) reporting the percentage of explained variance (R–squared value), the Standard Error for the model, the Chow statistic and Durbin-Watson statistic for each month.

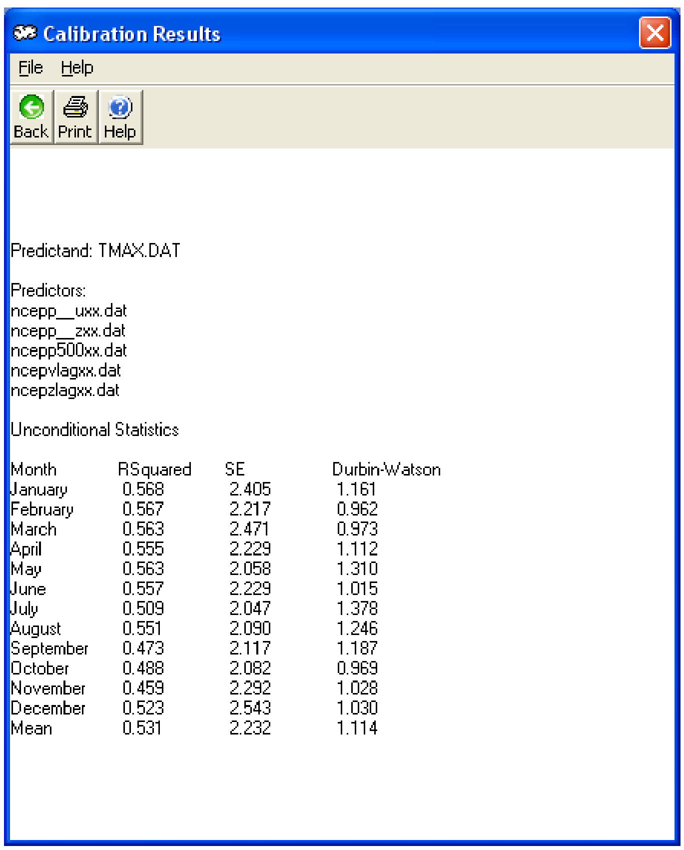


Figure 6.2. The **Calibration Results** screen

Click on the **Back** button and a **Scatter Plot** will be displayed. The example in Figure 6.3 shows an even spread of residuals across all values of the modelled predictand that is desirable.

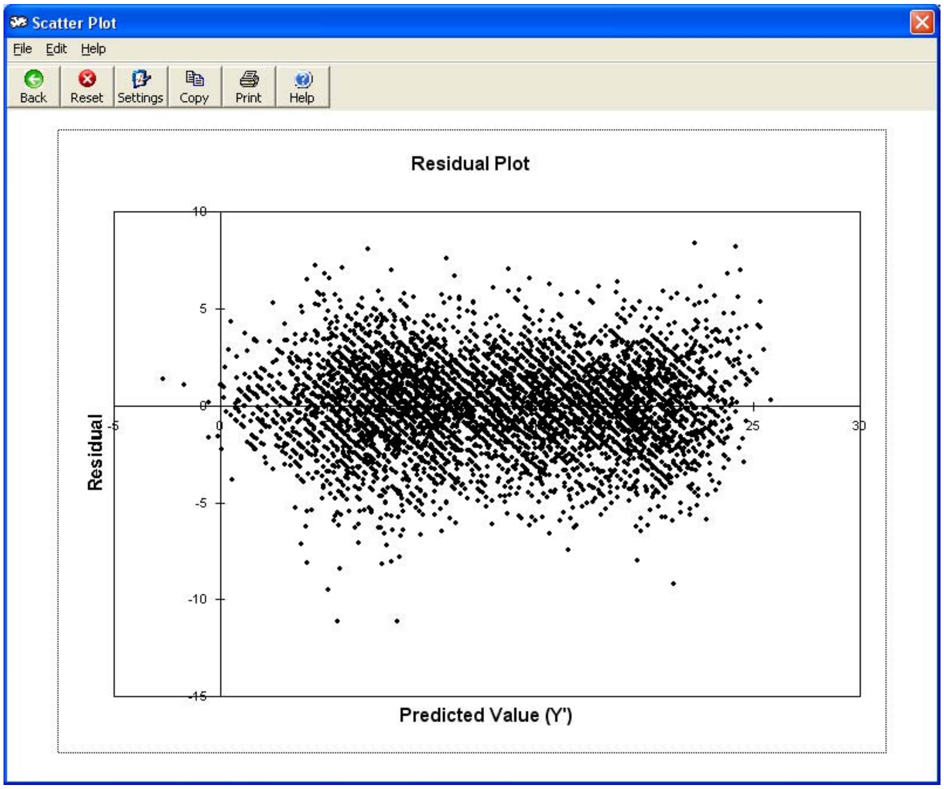


Figure 6.3 The **Scatter Plot** screen

Click on the **Back** button to return to the **Calibrate Model** screen.

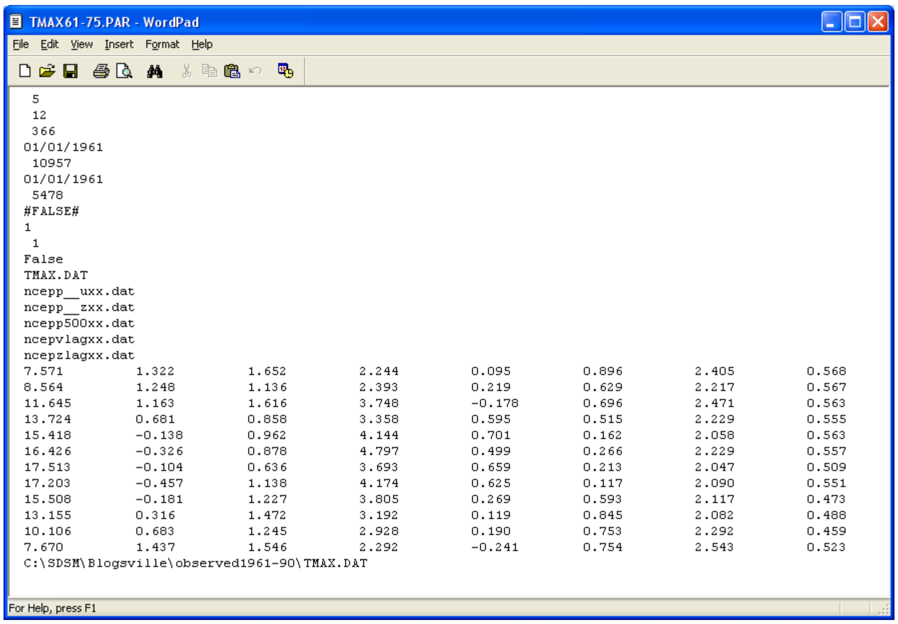
6.4 The *.PAR file

During model calibration a *.PAR file is generated that stores various parameters relating to the structure of the model. NOTE: Information held in the *.PAR file can often be used to diagnose the cause of any unexpected model results or behaviour. Figure 6.4 provides an example of such a file – produced using the Blogsville data set. In this file the data are stored in line order as follows:

[1] The number of predictors
[2] The season code (12 = months, 4 = seasons, 1 = annual model)
[3] The year length indicator (366, 365, or 360)
[4] Record start date
[5] Record length (days)
[6] Model fitting start date
[7] Number of days used in the model fitting
[8] Whether the model is conditional (True) or unconditional (False)
[9] Transformation (1 = none, 2 = fourth root, 3 = natural log, 4 = lognormal)
[10] Ensemble size
[11] Autoregression indicator (True or False)
[12] Predictand file name
[13-17] Predictor filenames (in this case five)

[18-29] Model parameters; the first 6 columns in this example are the parameters (including the intercept), the last two columns are the SE and r-squared statistic

[30] The root directory of the predictand file




```
 5
 12
 366
01/01/1961
 10957
01/01/1961
 5478
#FALSE#
1
 1
False
TMAX.DAT
ncepp__uxx.dat
ncepp__zxx.dat
ncepp500xx.dat
ncepvlagxx.dat
ncepzlagxx.dat
7.571    1.322    1.652    2.244    0.095    0.896    2.405    0.568
8.564    1.248    1.136    2.393    0.219    0.629    2.217    0.567
11.645   1.163    1.616    3.748    -0.178   0.696    2.471    0.563
13.724   0.681    0.858    3.358    0.595    0.515    2.229    0.555
15.418   -0.138   0.962    4.144    0.701    0.162    2.058    0.563
16.426   -0.326   0.878    4.797    0.499    0.266    2.229    0.557
17.513   -0.104   0.636    3.693    0.659    0.213    2.047    0.509
17.203   -0.457   1.138    4.174    0.625    0.117    2.090    0.551
15.508   -0.181   1.227    3.805    0.269    0.593    2.117    0.473
13.155   0.316    1.472    3.192    0.119    0.845    2.082    0.488
10.106   0.683    1.245    2.928    0.190    0.753    2.292    0.459
7.670    1.437    1.546    2.292    -0.241   0.754    2.543    0.523
C:\SDSM\Blogsville\observed1961-90\TMAX.DAT
```

Figure 6.4 The *.PAR file produced by the **Calibrate Model** screen

7. WEATHER GENERATOR

The **Weather Generator** operation produces ensembles of synthetic daily weather series given observed (or NCEP re–analysis) atmospheric predictor variables and regression model weights produced by the **Calibrate Model** operation (see Section 6). The **Weather Generator** enables the verification of calibrated models (assuming the availability of independent data) as well as the synthesis of artificial time series representative of present climate conditions. The **Weather Generator** can also be used to reconstruct predictands or to infill missing data.

To access this facility, click the **Weather Generator** button from any of the main screens. The following screen will appear:

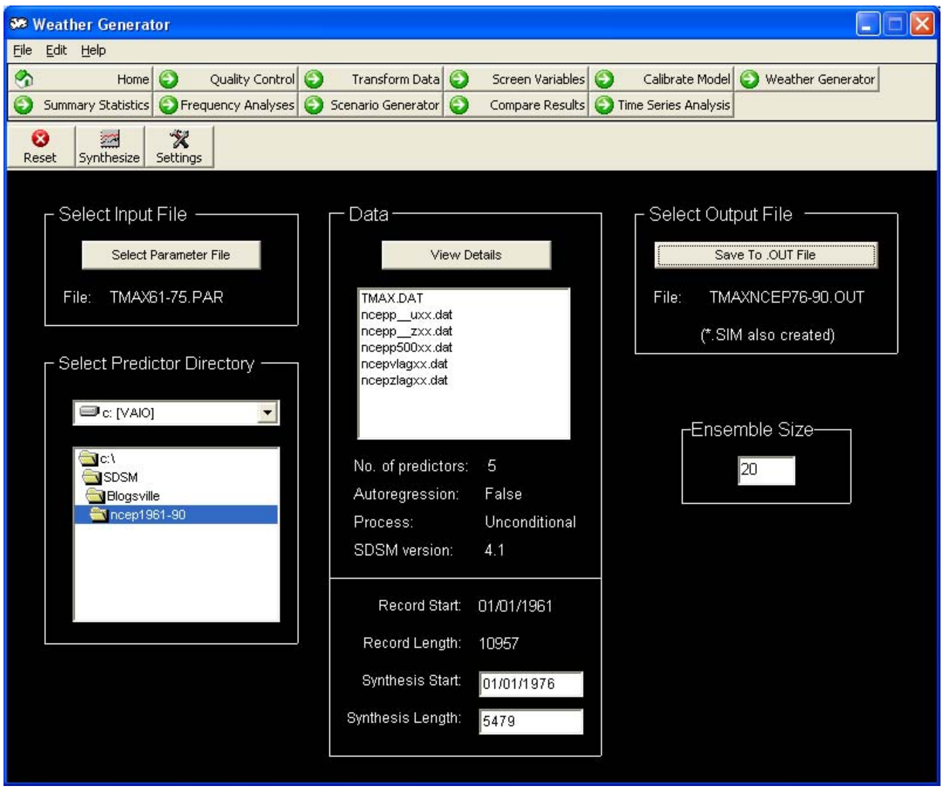


Figure 7.1 The **Weather Generator** screen.

7.1 File handling

The first step in the synthesis process is the selection of the appropriate model parameter file. Click on the **Select Parameter File** button, in the top left–hand corner. An **Open** file window appears; browse through until the correct directory and file are reached, then click on the appropriate file name – for example, the parameters for maximum daily temperature at Blogsville, are held in TMAX61-75.PAR. The name of the file will then appear beneath the button.

Next, specify the location of the predictor variable files by choosing the correct drive and directory from the window in the bottom left–hand corner of the screen under **Select Predictor Directory**.

To write synthetic data to a results file, it is necessary to select an appropriate directory and output file name. Click on the **Save To .OUT File** button in the top right–hand corner. An **Open** file window appears; browse through until the correct directory is reached, then enter a suitable file name – for example, TMAXNCEP76-90.OUT. The name of the file will then appear beneath the button.

Click on the **View Details** button, and the (predictand followed by predictor) files used in model calibration are listed in the window below. The **Record Start** date and available **Record Length** (number of days) are also displayed. NOTE: <u>*PAR files generated by earlier versions of SDSM can still be handled</u>. The SDSM version number of the PAR file is displayed along with the process type (conditional or unconditional) and whether autoregression was selected. The User must specify the (sub–) period required for weather generation using the **Synthesis Start** and **Synthesis Length** boxes, respectively. In this case the synthesis spans 1976-90 (so **Synthesis Start** is 1/1/76 and **Synthesis Length** is 5479 days).

The default values for **Synthesis Start** and **Synthesis Length** are used to simulate the period of record used for model calibration. If, however, model verification is to be undertaken using a set of independent data withheld from the calibration process, then the two values should be amended accordingly. If simulation of observed data based on the complete predictor record is needed, then the **Record Start** and **Record Length** values should be used.

7.2 Ensemble size

Decide how many ensemble members are needed, up to a maximum of 100, and enter the appropriate value in the **Ensemble Size** box at the bottom right–hand corner of the screen (the default is 20). Individual ensemble members are considered equally plausible local climate scenarios realised by a common set of regional–scale predictors. The extent to which ensemble members differ depends on the relative significance of the deterministic and stochastic components of the regression models used for downscaling. For example, local temperatures are largely determined by regional forcing whereas precipitation series display more "noise" arising from local factors. The former will yield similar looking ensemble members; the latter, larger differences between individual members.

Once the above selections have been completed, click the **Synthesize** button at the top of the menu. After a few seconds, the follow dialogue box will appear:

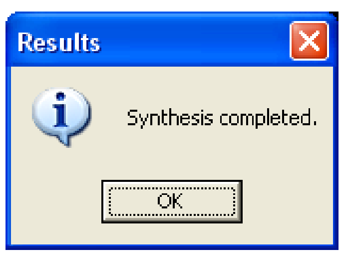


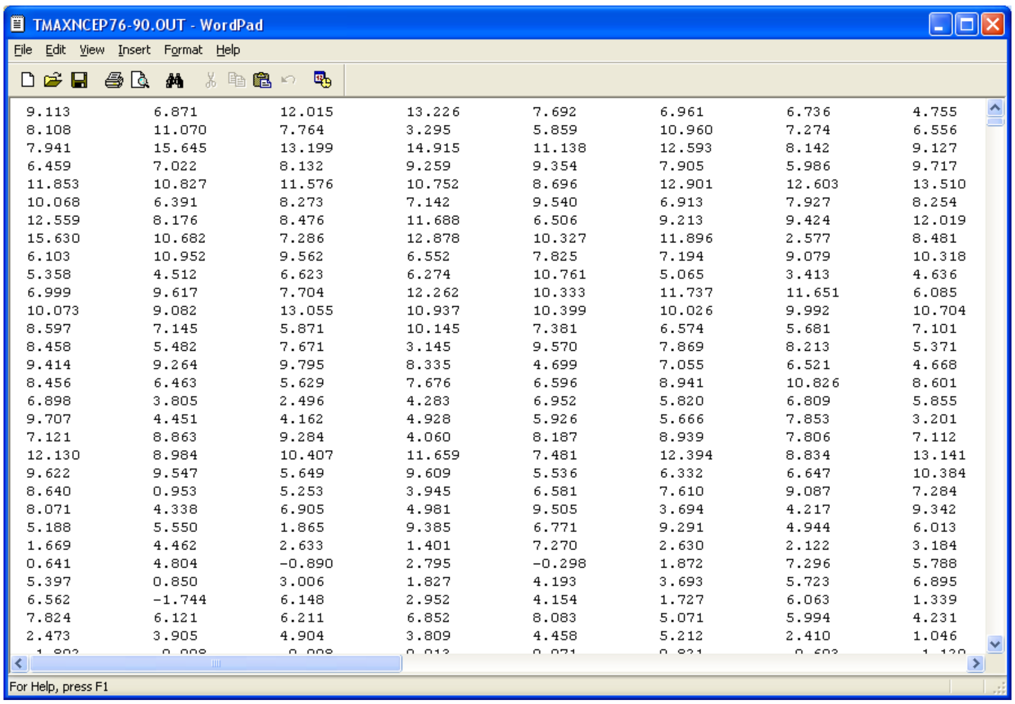
Figure 7.2 The synthesis completed dialogue box.

Click on **OK** to return to the **Weather Generator** screen.

7.3 Blogsville example

Selections for Blogsville are illustrated in Figure 7.1. In this example, the **Weather Generator** synthesized 20 runs of 15 years daily maximum temperature, using five regional–scale predictors. The data were synthesized using independent predictors withheld from model calibration (i.e., for the period 1976–1990; 5479 values).

Figure 7.3 shows the first few values of 12 ensemble members held in TMAXNCEP76-90.OUT. Figure 7.4 shows the corresponding TMAXNCEP76-90.SIM file, which contains meta–data associated with the synthesis. In both cases, the files have been opened in WordPad.



9.113	6.871	12.015	13.226	7.692	6.961	6.736	4.755
8.108	11.070	7.764	3.295	5.859	10.960	7.274	6.556
7.941	15.645	13.199	14.915	11.138	12.593	8.142	9.127
6.459	7.022	8.132	9.259	9.354	7.905	5.986	9.717
11.853	10.827	11.576	10.752	8.696	12.901	12.603	13.510
10.068	6.391	8.273	7.142	9.540	6.913	7.927	8.254
12.559	8.176	8.476	11.688	6.506	9.213	9.424	12.019
15.630	10.682	7.286	12.878	10.327	11.896	2.577	8.481
6.103	10.952	9.562	6.552	7.825	7.194	9.079	10.318
5.358	4.512	6.623	6.274	10.761	5.065	3.413	4.636
6.999	9.617	7.704	12.262	10.333	11.737	11.651	6.085
10.073	9.082	13.055	10.937	10.399	10.026	9.992	10.704
8.597	7.145	5.871	10.145	7.381	6.574	5.681	7.101
8.458	5.482	7.671	3.145	9.570	7.869	8.213	5.371
9.414	9.264	9.795	8.335	4.699	7.055	6.521	4.668
8.456	6.463	5.629	7.676	6.596	8.941	10.826	8.601
6.898	3.805	2.496	4.283	6.952	5.820	6.809	5.855
9.707	4.451	4.162	4.928	5.926	5.666	7.853	3.201
7.121	8.863	9.284	4.060	8.187	8.939	7.806	7.112
12.130	8.984	10.407	11.659	7.481	12.394	8.834	13.141
9.622	9.547	5.649	9.609	5.536	6.332	6.647	10.384
8.640	0.953	5.253	3.945	6.581	7.610	9.087	7.284
8.071	4.338	6.905	4.981	9.505	3.694	4.217	9.342
5.188	5.550	1.865	9.385	6.771	9.291	4.944	6.013
1.669	4.462	2.633	1.401	7.270	2.630	2.122	3.184
0.641	4.804	-0.890	2.795	-0.298	1.872	7.296	5.788
5.397	0.850	3.006	1.827	4.193	3.693	5.723	6.895
6.562	-1.744	6.148	2.952	4.154	1.727	6.063	1.339
7.824	6.121	6.211	6.852	8.083	5.071	5.994	4.231
2.473	3.905	4.904	3.809	4.458	5.212	2.410	1.046

Figure 7.3 An example of the format of the simulated data file (*.OUT).

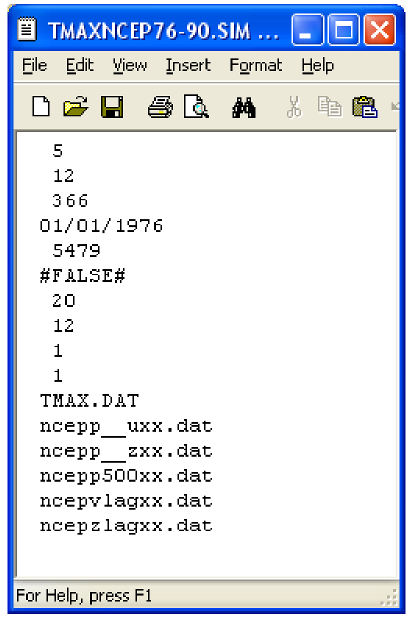


Figure 7.4 The *.SIM file produced by the **Weather Generator** operation for the Blogsville example. The output is (in line order): [1] the number of predictor variables; [2] the number of regression models used (1=annual, 4=seasonal, 12=monthly); [3] the maximum number of days in a year (here a calendar year is used, so there are up to 366 days in leap years); [4] the start date of the data used for model calibration; [5] the number of days simulated; [6] whether or not the predictand is a conditional (#TRUE#) or unconditional (#FALSE#) variable; [7] the number of ensemble members; [8] the variance inflation parameter (see **Advanced Settings**); [9] the transformation code for conditional variables (1=none, 2=fourth root, 3=natural log, 4=inverse normal); [10] the bias correction parameter (see **Advanced Settings**); [11] the predictand file name; [12 onward] the predictor file name(s).

8 ANALYSIS OF OBSERVED AND DOWNSCALED DATA

8.1 Overview

Statistical analyses of observed and downscaled weather data are handled in slightly different ways by SDSM but both are performed in the **Summary Statistics** screen. Common diagnostic tests are available for both observed and synthetic data. These statistics include the variable mean, maximum, minimum, variance, peaks above/below thresholds, percentiles, percent wet–days, and wet–/dry–day spell–lengths, computed on a calendar month, seasonal or annual basis.

To evaluate either downscaled data or observed data, click on the **Summary Statistics** button at the top of any main screen. The following screen will appear:

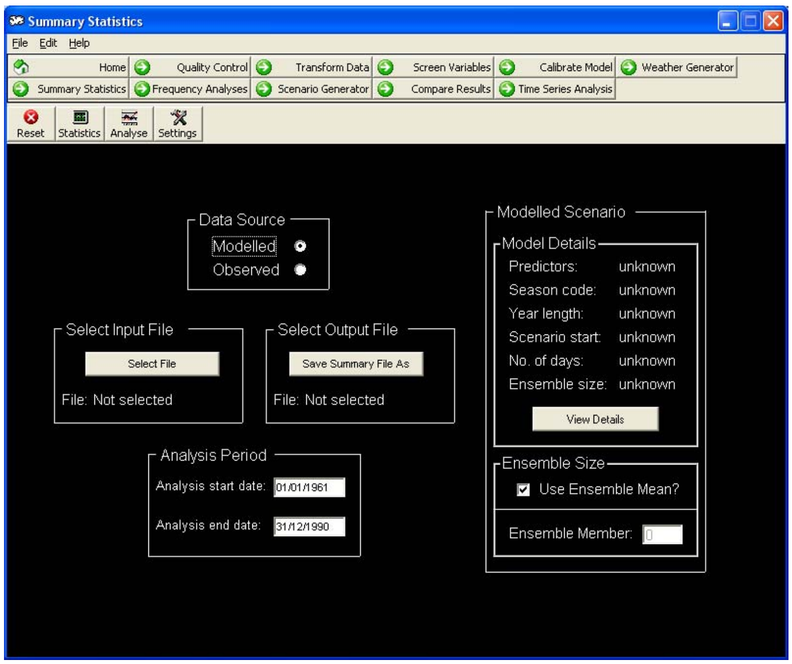


Figure 8.1 The **Summary Statistics** screen

The first step in the analysis is to select the **Data Source** – click on either Modelled, for downscaled data analysis, or Observed for observed data analysis.

The second step is the selection of an appropriate data file. Click on the **Select Input File** button, on the left–hand side. An **Open** file window appears; browse through until the correct directory and file are reached, then click on the appropriate file name – for example, actual maximum daily temperatures at Blogsville, are held in TMAX.DAT. The name of the file will then appear beneath the button. If using Modelled output click on **View Details** to check basic information about the downscaling experiment (such as the number of predictors, start date, etc.)

Next, specify the (sub–) period required for analysis using the **Analysis start date** and **Analysis end date** windows, under the **Analysis Period** header. The default values are the **Standard Start Date** and **Standard End Date** held in the global **Settings** (Section 3.1). The default **Use Ensemble Mean?** box produces mean diagnostics for all ensemble members and the standard deviation of the ensembles (see Figure 8.4). However, diagnostics for individual members may be extracted by deselecting **Use Ensemble Mean?** and by entering the required **Ensemble Member** (in this case, integers 1 to 100) in the **Ensemble Member** box.

To save the analysis results, it is necessary to select an appropriate directory and file name. Click on the **Save Summary File As** button on the right–hand side. An **Open** file window appears; browse through until the correct directory is reached, then enter a suitable file name – for example, TMAXOBS76-90.TXT. The name of the file will then appear beneath the button.

The final step is to select the required diagnostics. Click on the **Statistics** button at the top of the menu. The following screen will appear:

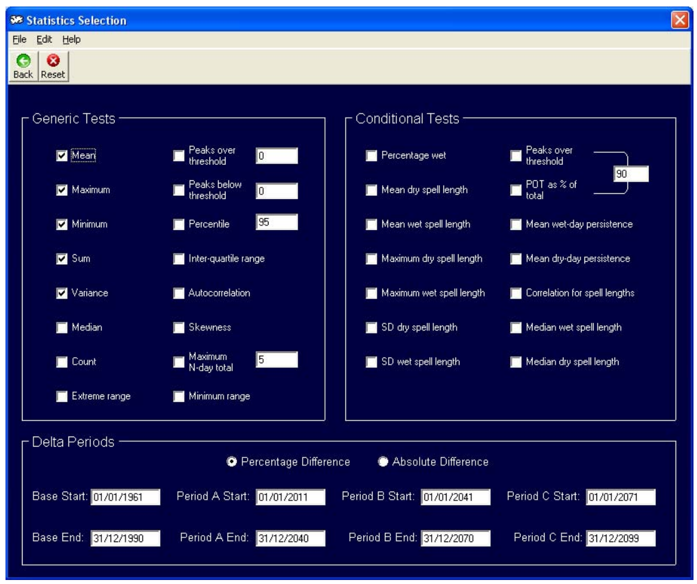
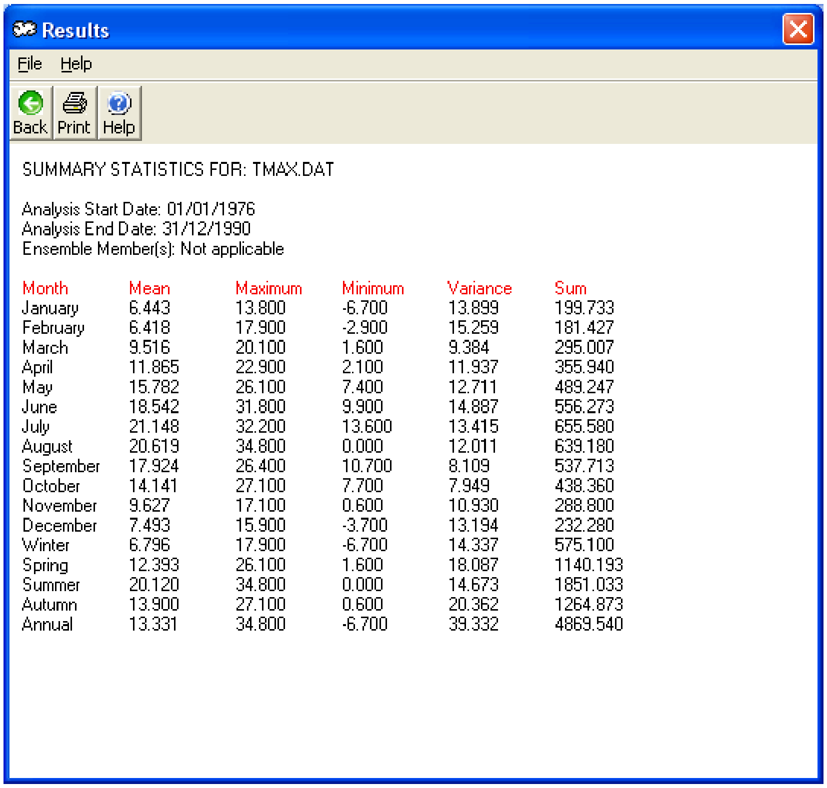


Figure 8.2 The **Statistics Selection** screen.

The screen is divided into three areas. The first, headed **Generic Tests**, lists statistics that may be applied to any variable (mean, maximum, minimum, sum, variance, median, count, peaks over/below threshold, percentile, inter–quartile range, autocorrelation, skewness and maximum N-day total, etc.). The second, headed **Conditional Tests**, lists statistics that are only applicable to daily conditional series (percentage wet, mean dry–/wet–spell length, maximum dry–/wet–spell length, standard deviation of dry-/wet-spell, peaks over threshold as a percentile, peaks over threshold as a percentage of total rainfall, etc.). Note that the definition of a wet–day can be adjusted using the **Event Threshold** under **Settings**. The default is zero (i.e., wet–days are defined as all days with non–zero precipitation totals). Note, the Sum is averaged by the number of years in the data set providing the monthly/seasonal/annual mean sum. See section 8.3 for an explanation of the statistics available. The third, headed Delta Periods is where the Delta time periods are entered and the type of Delta Statistic is selected (see below).

By checking the appropriate boxes, the User can select up to eight statistics for analysis. The defaults are the mean, maximum, minimum, sum, and variance. Click on **Back** to return to the **Summary Statistics** screen.

Once all the above selections have been completed, click on the **Analyse** button at the top of the menu. After a few seconds, the **Results** screen will appear:



SUMMARY STATISTICS FOR: TMAX.DAT

Analysis Start Date: 01/01/1976
Analysis End Date: 31/12/1990
Ensemble Member(s): Not applicable

Month	Mean	Maximum	Minimum	Variance	Sum
January	6.443	13.800	-6.700	13.899	199.733
February	6.418	17.900	-2.900	15.259	181.427
March	9.516	20.100	1.600	9.384	295.007
April	11.865	22.900	2.100	11.937	355.940
May	15.782	26.100	7.400	12.711	489.247
June	18.542	31.800	9.900	14.887	556.273
July	21.148	32.200	13.600	13.415	655.580
August	20.619	34.800	0.000	12.011	639.180
September	17.924	26.400	10.700	8.109	537.713
October	14.141	27.100	7.700	7.949	438.360
November	9.627	17.100	0.600	10.930	288.800
December	7.493	15.900	-3.700	13.194	232.280
Winter	6.796	17.900	-6.700	14.337	575.100
Spring	12.393	26.100	1.600	18.087	1140.193
Summer	20.120	34.800	0.000	14.673	1851.033
Autumn	13.900	27.100	0.600	20.362	1264.873
Annual	13.331	34.800	-6.700	39.332	4869.540

Figure 8.3 Default statistics for observed daily maximum temperatures at Blogsville during the validation period 1976–1990

The **Results** screen lists the name of the input file, along with the start and end dates of the analysis. Monthly, seasonal and annual mean statistics are listed beneath for the chosen tests. Comparison of the results obtained from the **Weather Generator** (see below) gives an indication of model skill. See Section 11 for graphical comparisons of monthly statistics.

Figure 8.4 shows the summary statistics for the modelled maximum daily temperatures of Blogsville during the validation period 1976-1990. The summary results of these statistics are saved to TMAXNCEP76-90.TXT.

Results

File Help

Back Print Help

SUMMARY STATISTICS FOR: TMAXNCEP76-90.OUT

Analysis Start Date: 01/01/1976
Analysis End Date: 31/12/1990
Ensemble Member(s): ALL

Month	Mean	Maximum	Minimum	Variance	POT
January	6.851	17.879	-5.319	13.999	0.000
February	6.850	16.377	-3.060	12.878	0.000
March	10.060	20.541	-1.008	13.623	0.000
April	12.034	21.264	2.775	10.394	0.000
May	15.777	25.061	4.632	12.259	0.650
June	19.025	29.755	8.329	14.247	25.850
July	20.699	28.302	12.132	8.175	27.450
August	20.675	29.245	11.757	8.661	31.600
September	18.057	26.347	9.359	8.073	2.800
October	14.288	23.047	6.323	7.721	0.050
November	9.723	18.615	-0.017	10.517	0.000
December	7.004	17.880	-4.363	14.309	0.000
Winter	6.903	18.647	-5.979	13.768	0.000
Spring	12.630	25.061	-1.008	17.797	0.650
Summer	20.145	30.094	8.326	10.930	84.900
Autumn	14.026	26.351	-0.017	20.249	2.850
Annual	13.455	30.094	-5.979	37.818	88.400

Standard Deviations of Results

January	0.105	1.517	1.736	0.902	0.000
February	0.114	0.841	0.895	0.740	0.000
March	0.114	1.155	1.045	0.860	0.000
April	0.111	1.324	0.924	0.525	0.000
May	0.099	0.832	1.596	0.502	0.792
June	0.070	0.956	1.169	0.652	3.595
July	0.074	0.700	0.646	0.317	3.801
August	0.104	1.029	0.957	0.575	5.083
September	0.102	0.948	0.963	0.502	1.503
October	0.081	1.188	0.906	0.557	0.218
November	0.108	0.753	1.263	0.451	0.000
December	0.110	1.113	1.443	0.628	0.000
Winter	0.062	1.315	1.553	0.504	0.000
Spring	0.073	0.832	1.045	0.464	0.792
Summer	0.053	0.766	1.165	0.319	7.543
Autumn	0.060	0.944	1.263	0.555	1.459
Annual	0.037	0.766	1.553	0.393	7.618

Figure 8.4 Example output of **Summary Statistics** (Modelled) showing the mean and standard deviation of diagnostics for a 20 member ensemble

8.2 Delta Statistics

Click on the **Delta Stats** button to calculate Delta Statistics. Delta statistics take the form:

$$\Delta 2020s = \frac{(V_{2020s} - V_{base})*100}{V_{base}}$$

$$\Delta 2050s = \frac{(V_{2050s} - V_{base})*100}{V_{base}}$$

$$\Delta 2080s = \frac{(V_{2080s} - V_{base})*100}{V_{base}}$$

if Percentage Difference is selected in **Statistics**, or:

$$\Delta 2020s = V_{2020s} - V_{base}$$

$$\Delta 2050s = V_{2050s} - V_{base}$$

$$\Delta 2080s = V_{2080s} - V_{base}$$

if Absolute Difference is selected.

V_{base} is the mean of all ensembles (or a specific ensemble if selected) for each statistic for the base period. Likewise, V_{2020s} is the mean of all ensembles (or a specific ensemble) for each statistic for period A, and so on for V_{2050s} and V_{2080s}.

8.3 The Statistics

The diagnostics that can be produced by the **Summary Statistics** screen are derived for each time period (i.e., month, season, annual) as follows:

Generic Tests

Mean	Average of all values.
Maximum	Largest of all values.
Minimum	Smallest of all values.
Sum	Total sum of all values.
Variance	Variance of all values in each time period.
Median	Median of all values in each time period.
Count	Count of the total number of values.
Extreme range	Maximum range of values within a given period.
Minimum range	Minimum range of values within a given period.
Peaks over threshold	Number of values greater than or equal to the User specified threshold.
Peaks below threshold	Number of values less than or equal to the User specified threshold.
Percentile	Value of the User specified percentile.
Inter-quartile range	Difference between the 25^{th} and 75^{th} percentiles.
Autocorrelation	Correlation coefficient for successive days.
Skewness	Skewness of the data.
Maximum N-day total	Maximum total accumulated over N-days.

Conditional Tests

Percentage wet	Percentage of days that exceed the threshold.

Mean dry spell length	Average length of spells with amounts less than the wet-day threshold.
Mean wet spell length	Average length of spells with amounts greater than or equal to the wet-day threshold.
Maximum dry spell length	Longest spell with amounts less than the wet-day threshold.
Maximum wet spell length	Longest spell with amounts greater than or equal to the wet-day threshold.
SD dry spell length	Standard deviation of spells with amounts less than the wet-day threshold.
SD wet spell length	Standard deviation of spells with amounts greater than or equal to the wet-day threshold.
Peaks over threshold	Count of peaks over User specified threshold (defined as a percentile of all data).
POT as % of total	Ratio of the sum of all values over the User specified threshold (defined as a percentile of all values) to the sum of all values.
Mean dry-day persistence	Total number of consecutive dry days divided by total number of dry days.
Mean wet-day persistence	Total number of consecutive wet days divided by total number of wet days.
Correlation for spell lengths	Overall persistent of spells, both wet and dry
Median dry spell length	Median length of spells with amounts less than the wet-day threshold.
Median wet spell length	Median length of spells with amounts greater than or equal to the wet-day threshold.

9 FREQUENCY ANALYSIS

The **Frequency Analysis** option allows the User to plot various distribution diagnostics for both modelled (ensemble members) and observed data. To access this facility select **Frequency Analysis** from any of the main screens. The following screen appears:

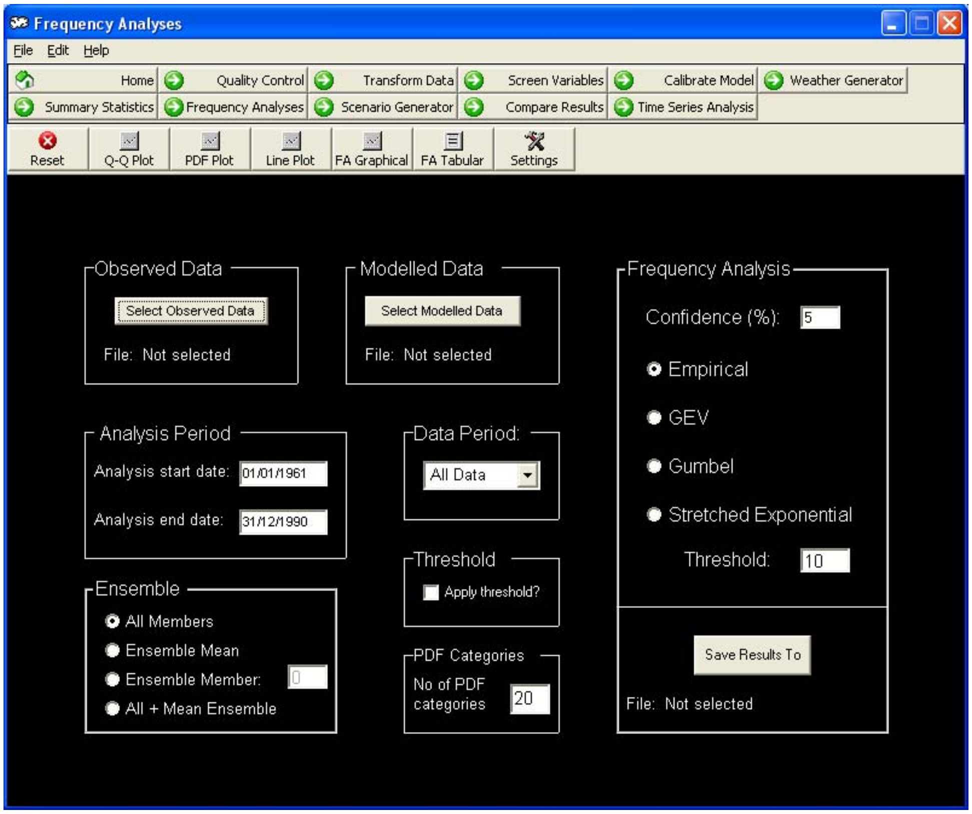


Figure 9.1 The **Frequency Analysis** screen

9.1 Setup

The first stage in the process is to select an observed data file and/or modelled data file to analyse. By clicking on the appropriate selection button (for example **Select Observed Data**), an input dialogue window appears in which the desired files can be selected. The second stage of the process is to enter the analysis period by entering the appropriate start and end dates in the **Analysis Period** box. In the **Data Period** box the User can select the time period for the analysis – all the data, individual months, or seasons. If modelled (ensemble) data are being analysed, the User can select which part of the ensemble to include in the analysis by selecting either All Members, Ensemble Mean, Ensemble Member, or All + Ensemble Member, in the Ensemble box. If the **Apply threshold?** box is ticked, only data which are above the global threshold (see **Settings**) are included in the analysis. If the User wishes to plot a Probability Density Function (PDF) as part of the analysis, the number of categories can be entered in the **PDF Categories** box – the default is 20.

9.2 Diagnostics and plots

Following the initial set up process, the User can perform a number of diagnostics on the data. These diagnostics are discussed in turn.

Quantile-Quantile (Q-Q) Plot

A Quantile-Quantile plot is used to compare a modelled data set with an observed data file. The procedure works by sorting each of the data files into order and calculating the percentiles (1 to 99). These are then plotted against one another on a scatter chart with observed data on the y-axis and modelled data on the x-axis. Note that it is assumed that observed data are always based on calendar years, while the length of the modelled data year is set using the **Settings** screen.

Figure 9.2 provides an example of a Quantile-Quantile plot. In this case TMAX.DAT has been selected as the observed data, TMAXCCF61-90.OUT as the modelled data (maximum daily temperature at Blogsville downscaled using HadCM3 output) for the period 1961-1990. In this case all the data have been analysed and the ensemble mean has been chosen to represent the modelled data.

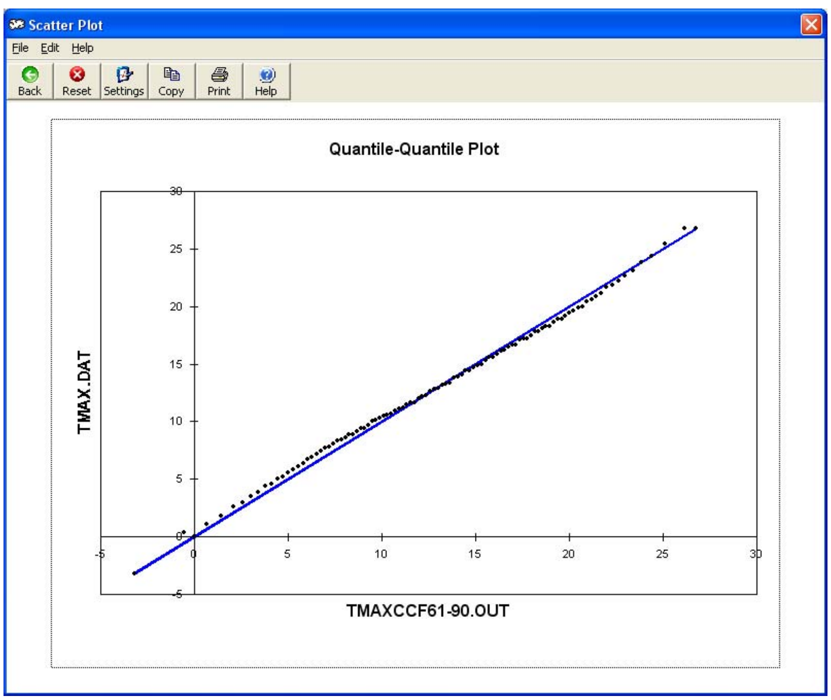


Figure 9.2 Quantile-Quantile plot of maximum daily temperature at Blogsville versus the ensemble mean downscaled from HadCM3 for the period 1961-90

Chart settings can be adjusted selecting the **Settings** button at the top of the screen. Refer to Section 11.3 for information on customising charts in this way.

PDF Plot

The PDF plot provides a Probability Density Function of the selected data files. The data are first sorted into order, then into categories (as defined by the User – the default being 20). A count is made of the number of data points in each category. The resultant density is plotted on a line chart as shown, for example, in Figure 9.3. In this case the TMAX.DAT data have been distributed into 20 categories. Figure 9.4 shows a graph of the same data, this time with only ten categories.

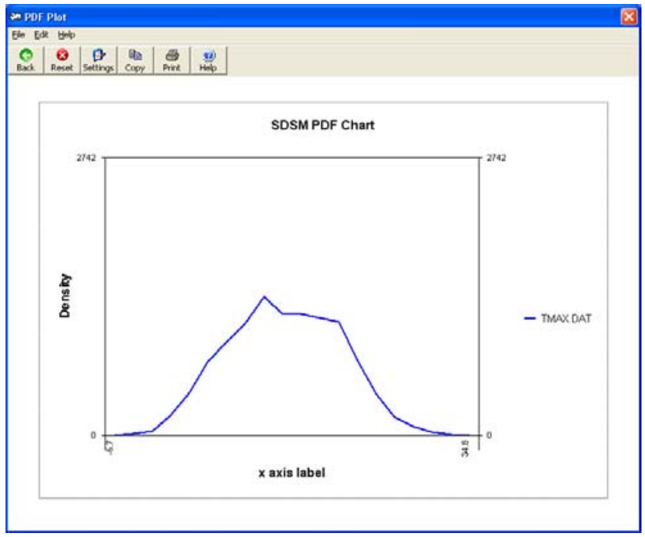


Figure 9.3 PDF plot of observed maximum daily temperature at Blogsville for the period 1961-1990 (20 categories)

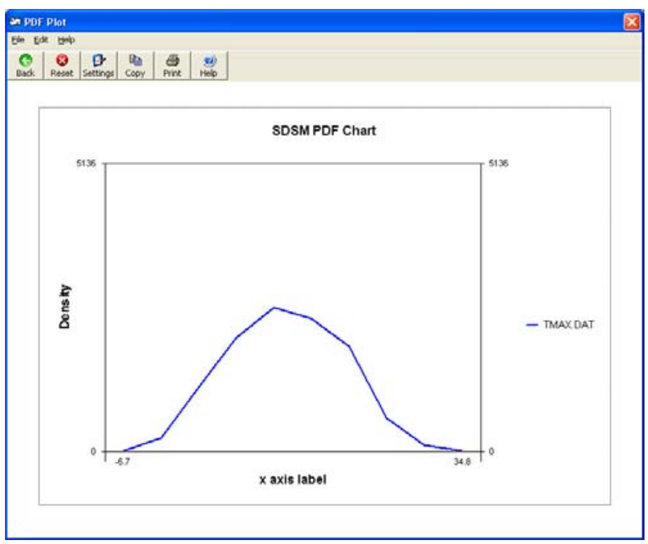


Figure 9.4 PDF plot of observed maximum daily temperature at Blogsville for the period 1961-1990 (10 categories)

Line Plot

This produces a simple time-series chart of the selected data. Note that only a maximum of ten year's of data can be plotted using this option. Figure 9.5 provides an example of such a plot. In this case, the ensemble mean of the maximum temperature (TMAXNCEP61-90) downscaled from NCEP has been plotted against the observed maximum temperature for Blogsville for the period 1961-1970.

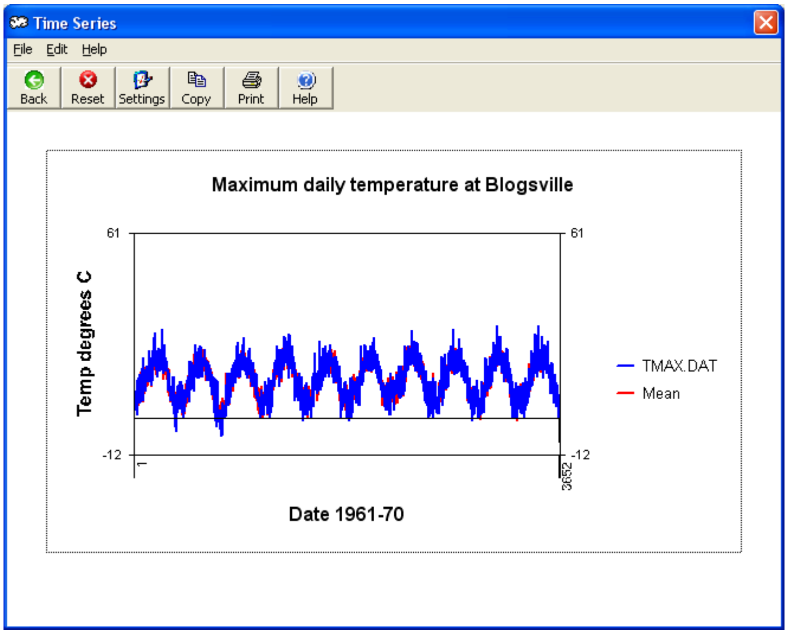


Figure 9.5 Time series plots of maximum daily temperature at Blogsville – observed data (blue line) and ensemble mean downscaled from NCEP (red line)

9.3 Extreme value analysis

The remaining four statistical measures allow the User to fit distributions to observed and downscaled data (as either a whole data set or by isolating particular seasons or months) in order to interpret extreme events. The available distributions are: Generalized Extreme Value (GEV), stretched exponential, empirical and Gumbel. Results can be viewed in either tabular format by selecting **FA Tabular**, or as line charts by selecting **FA Graphical** from the menu buttons at the top of the screen.

Empirical
This option fits a simple empirical distribution to the data by sorting the annual maximums into ascending order and plotting these according to the return period. Figure 9.6 provides an example of an Empirical line plot while Figure 9.7 shows the same results presented in a tabular format (FA Tabular). In this case, the maximum temperatures downscaled from NCEP (TMAXNCEP61-90.OUT) clearly under-estimate the intensity of the observed (TNAX.DAT) hot-days with return periods exceeding 3-years.

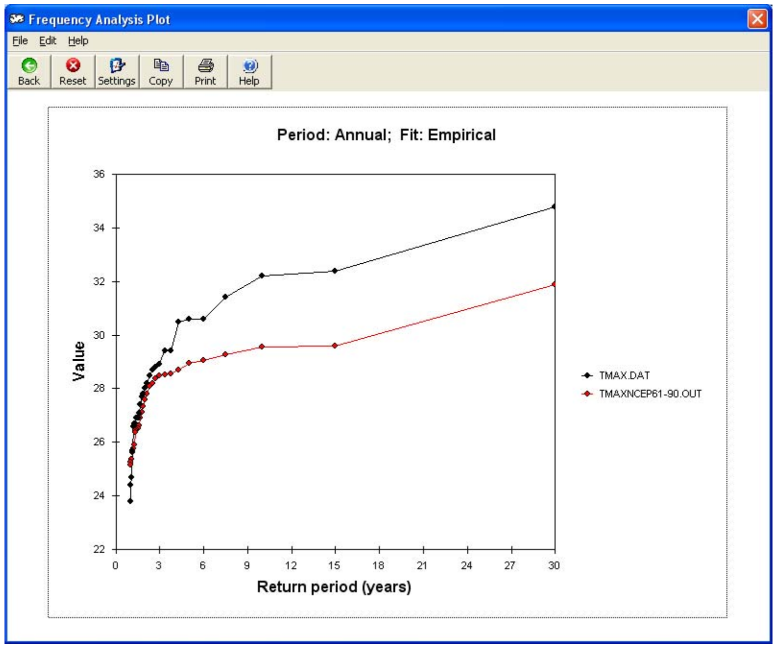


Figure 9.6 Example of an empirical fit to observed and downscaled maximum daily temperature at Blogsville for the period 1961-1990

Frequency Analysis Results

File Help

Back Print Help

Season: Annual
Fit: Empirical
Observed data: TMAX.DAT
Modelled data: TMAXNCEP61-90.OUT

Return Period	Obs	Mod
1.0	23.800	25.161
1.0	24.400	25.274
1.1	24.700	25.351
1.1	25.600	25.625
1.2	25.700	25.693
1.2	26.600	25.773
1.3	26.700	25.908
1.3	26.700	26.382
1.4	26.700	26.432
1.4	26.900	26.479
1.5	26.900	26.514
1.6	27.100	26.626
1.7	27.400	26.911
1.8	27.700	27.134
1.9	27.800	27.323
2.0	28.000	27.577
2.1	28.200	27.790
2.3	28.500	28.100
2.5	28.700	28.189
2.7	28.800	28.370
3.0	28.900	28.487
3.3	29.400	28.528
3.8	29.400	28.564
4.3	30.500	28.702
5.0	30.600	28.936
6.0	30.600	29.041
7.5	31.400	29.273
10.0	32.200	29.553
15.0	32.400	29.591
30.0	34.800	31.891

Figure 9.7 Same results as Figure 9.5 presented in a tabular format

GEV

This fits a three-parameter (ξ, β, k) Generalised Extreme Value (GEV) distribution to the data of the form:

$$F(x) = \exp(-(1 - k(\frac{x-\xi}{\beta})^{\frac{1}{k}}))$$

The parameters (ξ, β, k) are estimated using the method of L moments in which the first three L moments (l_1, l_2, l_3) are estimated from the data (see Kysely, 2002). The parameters are then calculated according to:

$$k = 7.8590z + 2.955z^2$$

$$\beta = \frac{l_2 k}{(1-2^{-k})\Gamma(1+k)}$$

$$\xi = l_1 + \beta \frac{\Gamma(1+k)-1}{k}$$

In which:

$$z = \left(\frac{2}{3 + \frac{l_3}{l_2}} - \frac{\ln 2}{\ln 3} \right)$$

The results are plotted up to a return period of 100 years. Figure 9.8 shows the GEV plot using the same data as in Figure 9.6.

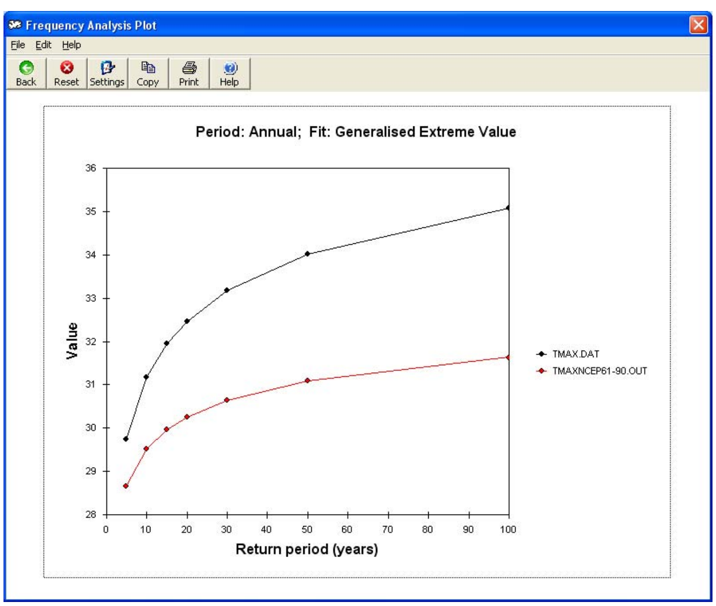


Figure 9.8 Generalised Extreme Value plot of maximum and downscaled temperature for Blogsville for the period 1961-1990

For values of *k* approaching zero (-0.005<*k*<0.005 in SDSM) the two-parameter Gumbel distribution is applied using the following equations (see Kysely, 2002):

$$F(x) = \exp(-\exp(-\frac{x-\xi}{\beta}))$$

where

$$\beta = \frac{l_2}{\ln 2}$$

$$\xi = l_1 - \gamma \frac{l_2}{\ln 2}$$

γ is the Euler constant (0.577215655)

Gumbel

Fits a Gumbel Type 1 distribution to the data using the annual maximum series after the method of Shaw (1994):

$$F(x) = 1 - e^{-e^{-(x-\mu)/\sigma}}$$

Thus, the annual maximum for a return period of T-years can be calculated from:

$$Q_T = \overline{Q} + K(T)S_Q$$

$$K(T) = -\frac{\sqrt{6}}{\pi}\left(\gamma + \ln\ln\left[\frac{T(X)}{T(X)-1}\right]\right)$$

In which $\overline{Q}$ is the mean of the annual maximums, S_Q is the standard deviation of these maximums, *K(T)* is a frequency factor, *T(X)* is the return period in years, and γ is the Euler constant (0.577215655).

Figure 9.8 shows an example of a Gumbel plot using the same data as in Figure 9.6.

Stretched Exponential

Fits the data to a Stretched Exponential distribution of the form:

$$P(R > r) = \exp(1(\frac{r}{R0})^c)$$

It is used to calculate the probability that an event is greater than a threshold, *r*. *R*0 is the mean of all events, and *c* is determined from the data fitting. The data are truncated according to the User specified threshold value. Figure 9.10 provides an example of a Stretched Exponential plot using the same data as in Figure 9.6.

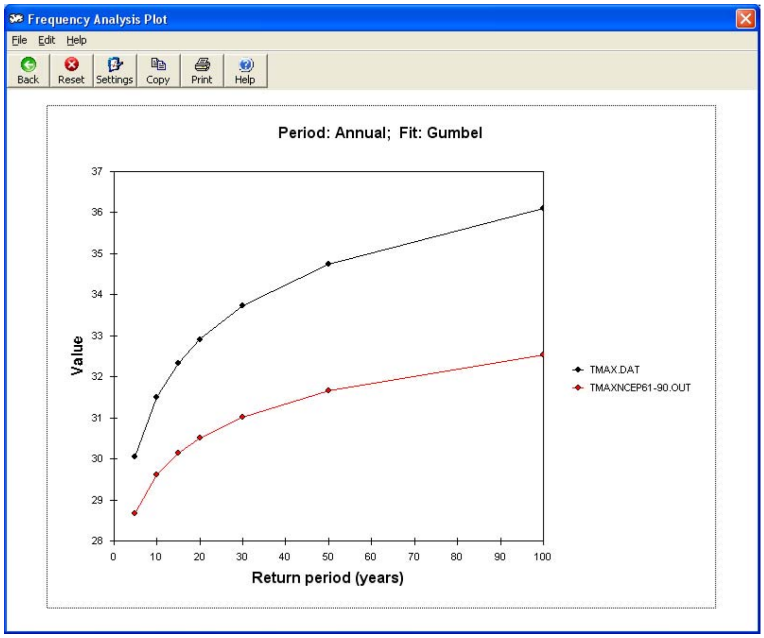


Figure 9.9 Gumbel plot of maximum and downscaled temperature for Blogsville for the period 1961-1990

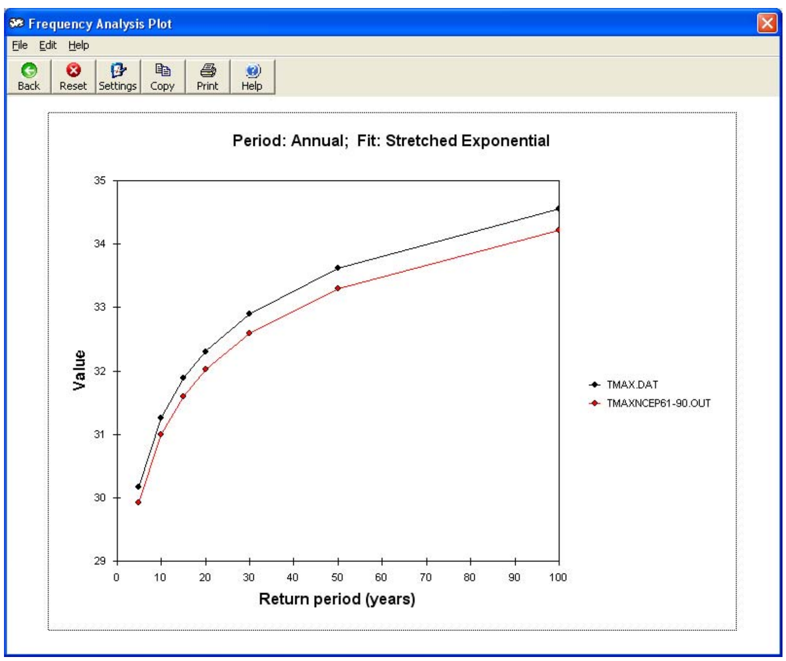


Figure 9.10 Stretched Exponential plot of maximum and downscaled temperature for Blogsville for the period 1961-1990

10 SCENARIO GENERATION

The **Scenario Generator** operation produces ensembles of synthetic daily weather series given daily atmospheric predictor variables supplied by a GCM (either under present or future greenhouse gas forcing). The GCM predictor variables must be normalised with respect to a reference period (or control run) and available for all variables used in model calibration (see Section 2.2).

The procedure is identical to that of the **Weather Generator** operation in all respects except that it may be necessary to specify a different convention for model dates and source directory for predictor variables. As in the case of the **Weather Generator** (see Section 7), input files for the **Scenario Generator** need <u>not</u> be the same length as those used to obtain the regression weights during calibration.

To access this facility select **Scenario Generator** from any of the main screens. The following screen appears:

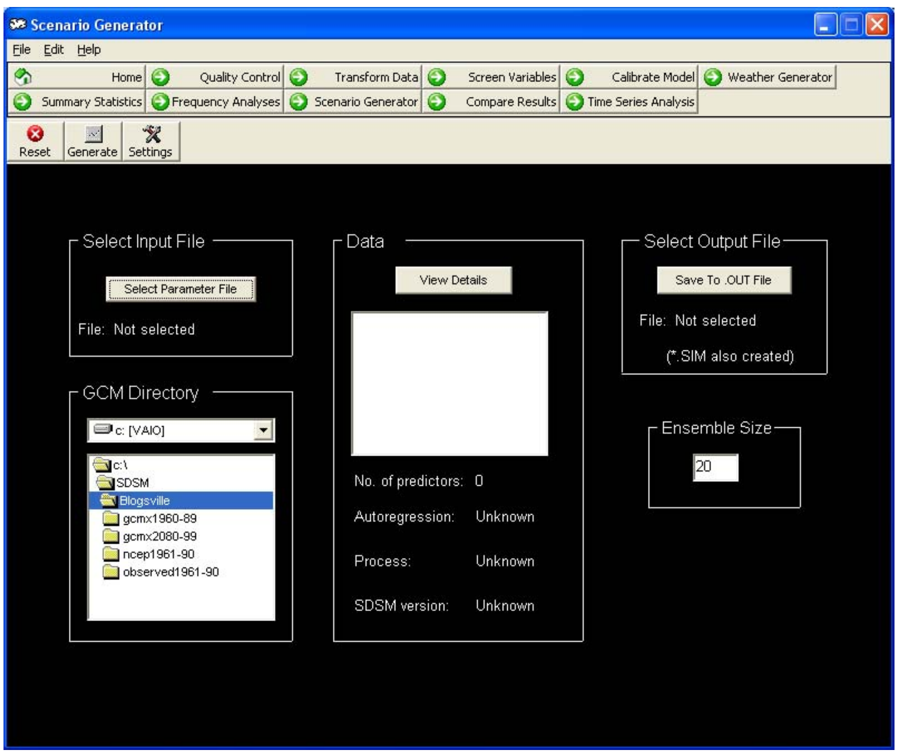


Figure 10.1 The **Scenario Generator** screen

10.1 Check settings

Before starting scenario generation, it may be necessary to change some of the options in the **Settings** menu. Click on the **Settings** button at the top of the screen and check the appropriate **Year Length** box. Also, amend the **Standard Start/End Date** in line with the GCM data time–slices. For example, HadCM2 and HadCM3 have year lengths of 360 days, and for the Blogsville example, the period 1961-1990 was used to represent present climate forcing (10800 values). Once necessary changes have been made to the **Settings**, click on **Back** to return to the **Scenario Generator** screen.

10.2 Setup

The first step in scenario generation is the selection of the appropriate downscaling model parameter file. Click on the **Select Parameter File** button, in the top left–hand corner. An **Open** file window appears; browse through until the correct directory and file are reached, then click on the appropriate file name – for example, the parameters for maximum daily temperature at Blogsville, are held in TMAX61-75.PAR. The name of the file will then appear beneath the button.

Next, click on the **View Details** button, and the (predictand followed by predictor) files used in model calibration are listed in the window below. In addition, information on the number of predictors, autoregression, process type and SDSM version are also presented.

Next, select the appropriate drive location and directory for the GCM predictors under the **GCM Directory** header. For best practice, GCM predictors originating from different experiments or time–slices (e.g., 1961–1990 or 2070–2099) should be held in separate folders. This is because SDSM will load only files with the same predictor names (i.e., characters 5 to 8) as those used in model calibration (see Table 2.1). Note that in order to proceed with the Blogsville example two lagged files (of one day) will need to be created using the **Transform Data** screen in the Blogsville\gcmx1961-90 and the Blogsville\gcmx2070-99 directories; gcmxp__vxx.dat to produce gcmxvlagxx.dat and gcmxp__zxx.dat to produce gcmzlagxx.dat.

As in the **Weather Generator** (Section 7), decide how many ensembles members are needed, up to a maximum of 100, and enter the appropriate integer in the **Ensemble Size** box on the right–hand side of the screen (the default is 20).

Finally, to save the scenario data to a results file, it is necessary to select an appropriate directory and file name. Click on the **Select Output File** button in the top right–hand corner. An **Open** file window appears; browse through until the correct directory is reached, then enter a suitable file name – for example, TMAXCCF61-90.OUT (maximum temperature, present climate forcing, 1961-1990). The name of the file will then appear beneath the button.

Once the above selections have been completed, click on the **Generate** button at the top of the screen. After a short while, a dialogue box will appear (Figure 10.2). Click on **OK** to return to the **Generate Scenario** screen.

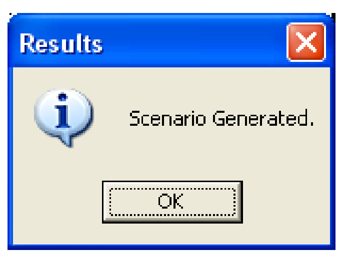


Figure 10.2 The Scenario Generated dialogue box

10.3 Blogsville example (temperature)

For the Blogsville example, the **Scenario Generator** operation was applied twice. First predictors from the HadCM3 experiment for the period 1961–1990 were used to downscale present climate forcing. Figure 10.3 shows the **Results** screen for this scenario, using the **Summary Statistics** operation (see Section 8).

Results

File Help

Back Print Help

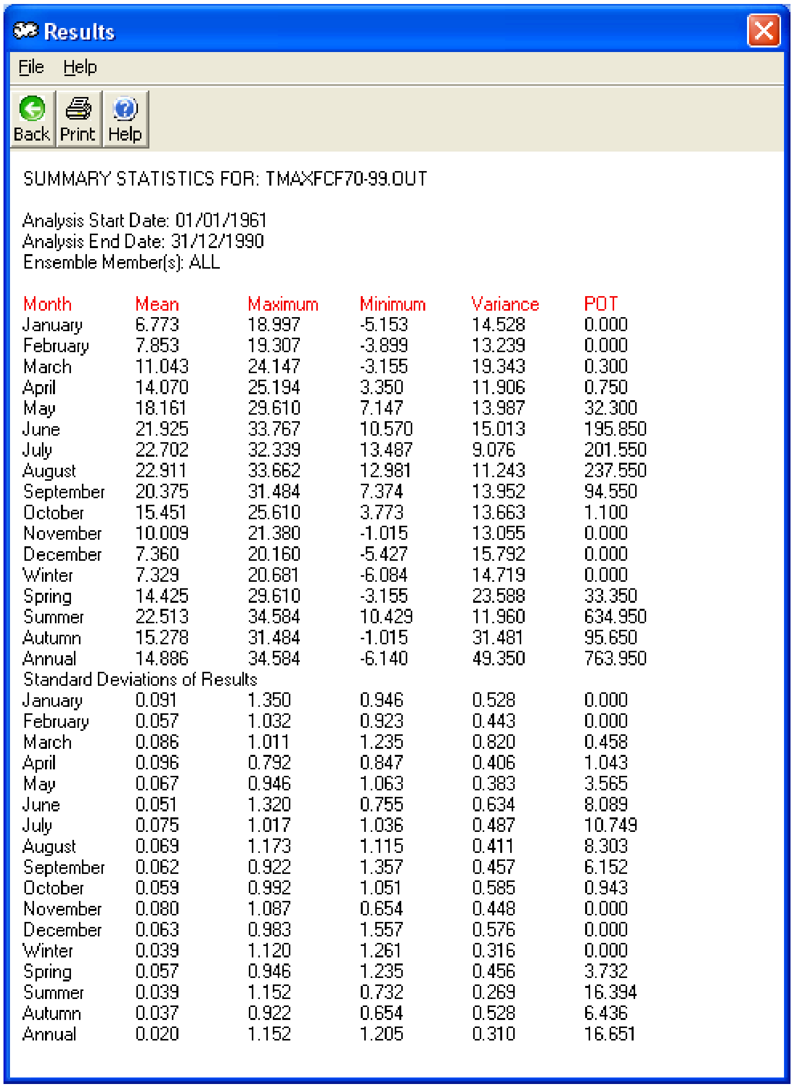
SUMMARY STATISTICS FOR: TMAXCCF61-90.OUT

Analysis Start Date: 01/01/1961
Analysis End Date: 31/12/1990
Ensemble Member(s): ALL

Month	Mean	Maximum	Minimum	Variance	POT
January	5.955	18.117	-6.995	15.287	0.000
February	6.710	18.464	-4.076	13.280	0.000
March	9.078	21.566	-5.422	20.089	0.000
April	12.475	23.133	1.203	12.286	0.050
May	15.934	26.839	4.511	12.038	3.900
June	19.729	31.363	8.519	12.673	63.950
July	20.701	30.477	12.284	8.548	69.200
August	20.194	30.450	9.802	11.794	74.600
September	18.152	28.373	5.173	13.756	22.600
October	13.390	23.705	1.890	13.754	0.100
November	8.910	20.561	-2.602	13.391	0.000
December	6.749	18.310	-5.606	15.262	0.000
Winter	6.471	19.602	-7.427	14.747	0.000
Spring	12.495	26.839	-5.422	22.643	3.950
Summer	20.208	31.951	8.403	11.166	207.750
Autumn	13.484	28.373	-2.602	27.875	22.700
Annual	13.165	31.951	-7.430	42.849	234.400
Standard Deviations of Results					
January	0.093	1.178	1.583	0.486	0.000
February	0.073	1.561	0.953	0.548	0.000
March	0.092	0.998	1.133	0.683	0.000
April	0.064	1.053	1.118	0.486	0.218
May	0.064	1.028	1.710	0.411	1.670
June	0.080	1.452	1.029	0.361	5.912
July	0.062	1.075	0.659	0.334	5.501
August	0.066	0.748	1.103	0.441	6.184
September	0.065	0.664	1.258	0.552	4.236
October	0.078	0.990	1.042	0.356	0.300
November	0.070	1.153	1.083	0.514	0.000
December	0.080	1.036	1.464	0.554	0.000
Winter	0.060	1.184	1.400	0.240	0.000
Spring	0.039	1.028	1.133	0.393	1.687
Summer	0.042	1.111	0.934	0.211	9.889
Autumn	0.049	0.664	1.083	0.247	4.291
Annual	0.029	1.111	1.398	0.225	11.868

Figure 10.3 Example results for Blogsville using GCM predictors (1961–1990)

Second, predictors from the HadCM3 experiment for the period 2070–2099 were used to downscale future climate forcing. Note that as the Blogsville data set contains future GCM data for 2070-99 it was necessary in the **Summary Statistics** screen to set the analysis period to 1961-90 to cover the thirty year period of the 2070-99 data set. (Ordinarily, SDSM predictors would be supplied for the full period 1961-2100 so this "tweak" to the dates would not be needed). Figure 10.4 shows the **Results** screen for this scenario, using the **Summary Statistics** operation.



SUMMARY STATISTICS FOR: TMAXFCF70-99.OUT

Analysis Start Date: 01/01/1961
Analysis End Date: 31/12/1990
Ensemble Member(s): ALL

Month	Mean	Maximum	Minimum	Variance	POT
January	6.773	18.997	-5.153	14.528	0.000
February	7.853	19.307	-3.899	13.239	0.000
March	11.043	24.147	-3.155	19.343	0.300
April	14.070	25.194	3.350	11.906	0.750
May	18.161	29.610	7.147	13.987	32.300
June	21.925	33.767	10.570	15.013	195.850
July	22.702	32.339	13.487	9.076	201.550
August	22.911	33.662	12.981	11.243	237.550
September	20.375	31.484	7.374	13.952	94.550
October	15.451	25.610	3.773	13.663	1.100
November	10.009	21.380	-1.015	13.055	0.000
December	7.360	20.160	-5.427	15.792	0.000
Winter	7.329	20.681	-6.084	14.719	0.000
Spring	14.425	29.610	-3.155	23.588	33.350
Summer	22.513	34.584	10.429	11.960	634.950
Autumn	15.278	31.484	-1.015	31.481	95.650
Annual	14.886	34.584	-6.140	49.350	763.950
Standard Deviations of Results					
January	0.091	1.350	0.946	0.528	0.000
February	0.057	1.032	0.923	0.443	0.000
March	0.086	1.011	1.235	0.820	0.458
April	0.096	0.792	0.847	0.406	1.043
May	0.067	0.946	1.063	0.383	3.565
June	0.051	1.320	0.755	0.634	8.089
July	0.075	1.017	1.036	0.487	10.749
August	0.069	1.173	1.115	0.411	8.303
September	0.062	0.922	1.357	0.457	6.152
October	0.059	0.992	1.051	0.585	0.943
November	0.080	1.087	0.654	0.448	0.000
December	0.063	0.983	1.557	0.576	0.000
Winter	0.039	1.120	1.261	0.316	0.000
Spring	0.057	0.946	1.235	0.456	3.732
Summer	0.039	1.152	0.732	0.269	16.394
Autumn	0.037	0.922	0.654	0.528	6.436
Annual	0.020	1.152	1.205	0.310	16.651

Figure 10.4 Example results for the Blogsville using GCM predictors (2070–2099)

Using the **Compare Results** operation (see Section 11), it is possible to compare the frequency of "hot" days at Blogsville downscaled using observed (NCEP) and GCM (HadCM3) predictor variables. For example, Figure 10.5 shows the respective monthly mean frequencies produced by each set of predictors with an ensemble size of 20. This was achieved by comparing peaks over threshold (POT) statistics for TMAXNCEP61-90.TXT and TMAXCCF61-90.TXT.

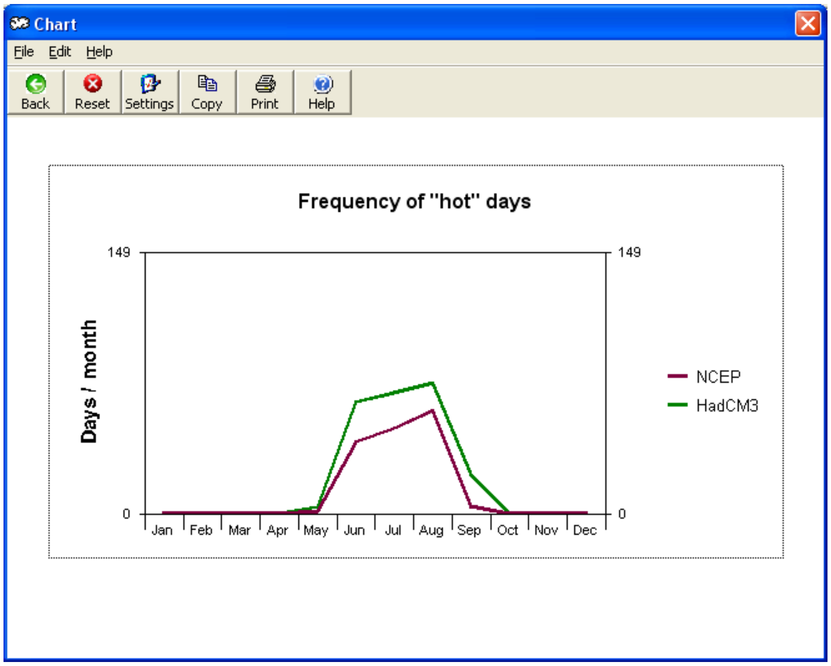


Figure 10.5 Monthly frequency of “hot” days (>25ºC) at Blogsville for the present climate downscaled using observed (NCEP) predictors (1961–1990) and GCM (HadCM3) predictors (1961–1990)

By using the **Compare Results** operation again, it is possible to compare the frequency of “hot” days at Blogsville under present (1961–1990) and future (2080–2099) climate forcing. This was achieved by comparing POT statistics for TMAXCCF61-90.TXT with TMAXFCF70-99.TXT. For example, Figure 10.6 shows a significant increase in the frequency of hot–days in summer by the end of the 21st century. The downscaling also indicates that hot-days could begin to appear as early as May by the end of the 21st century.

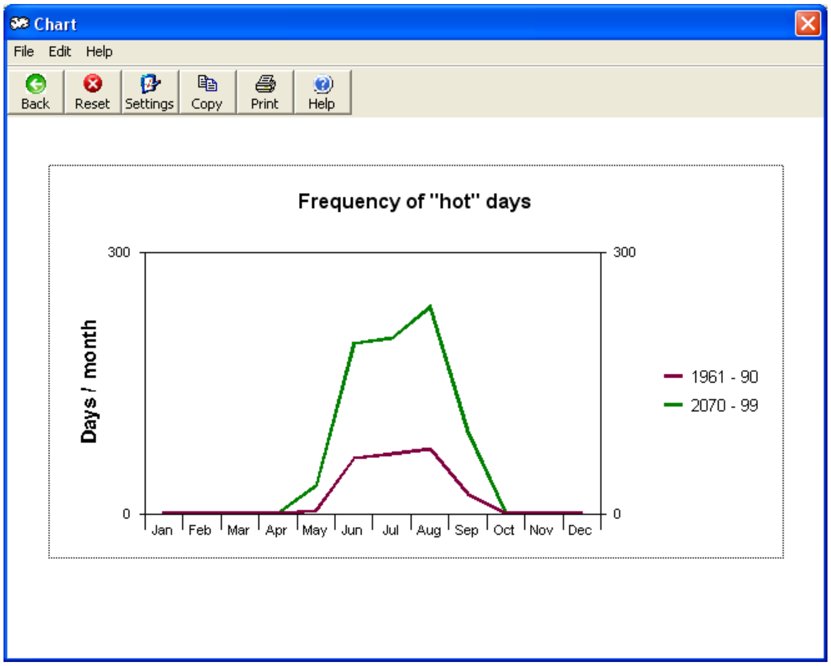
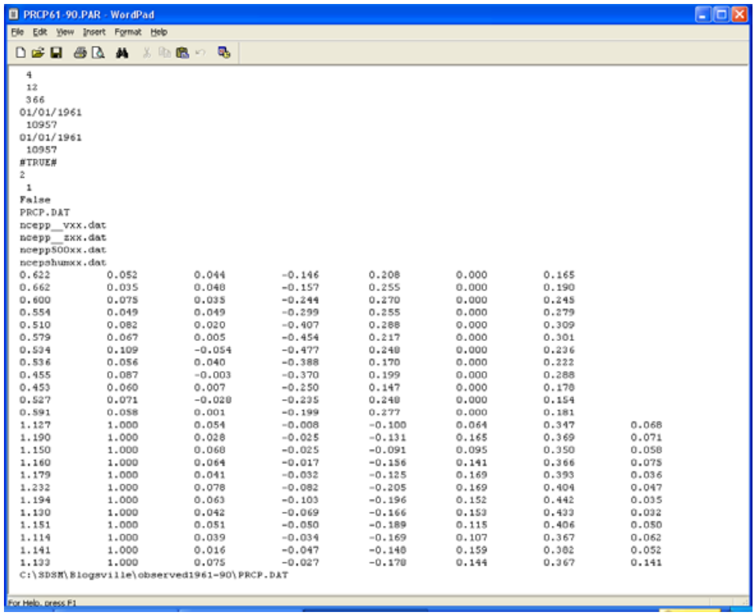


Figure 10.6 Monthly frequency of "hot" days (>25ºC) at Blogsville downscaled using HadCM3 predictors under present (1961–1990) and future (2070–2099) forcing

10.4 Blogsville example (precipitation)

Precipitation downscaling is necessarily more problematic than temperature, because daily precipitation amounts at individual sites are relatively poorly resolved by regional–scale predictors, and because precipitation is a conditional process (i.e., both the occurrence and amount processes must be specified). Figure 10.7 shows the *.PAR file (PRCP61-90.PAR) generated when a precipitation model was generated using the following parameters; 4 predictors (ncepp__vxx.dat, ncepp__zxx.dat, ncepp500xx.dat and ncepshumxx.dat), monthly model, using NCEP data from 1961-1990, with a fourth root transformation of the predictand.



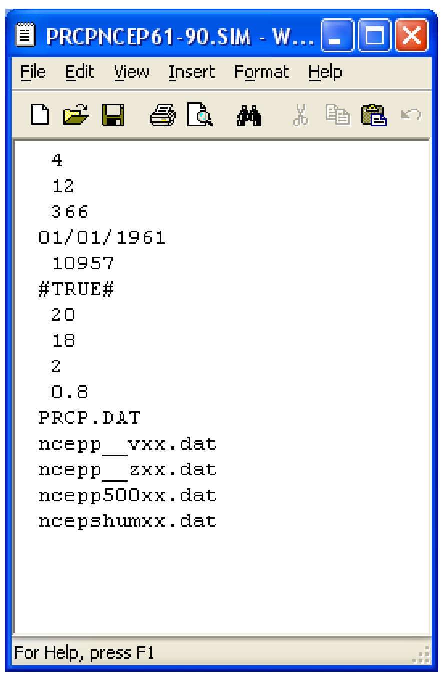
```
PRCP61-90.PAR - WordPad
File Edit View Insert Format Help

 4
 12
 366
01/01/1961
 10957
01/01/1961
 10957
#TRUE#
2
 1
False
PRCP.DAT
ncepp__vxx.dat
ncepp__zxx.dat
ncepp500xx.dat
ncepshumxx.dat
0.622    0.052    0.044    -0.146   0.208    0.000    0.165
0.662    0.035    0.048    -0.157   0.255    0.000    0.190
0.600    0.075    0.035    -0.244   0.270    0.000    0.245
0.554    0.049    0.049    -0.299   0.255    0.000    0.279
0.510    0.082    0.020    -0.407   0.288    0.000    0.309
0.579    0.067    0.005    -0.454   0.217    0.000    0.301
0.534    0.109    -0.054   -0.477   0.248    0.000    0.236
0.536    0.056    0.040    -0.388   0.170    0.000    0.222
0.455    0.087    -0.003   -0.370   0.199    0.000    0.288
0.453    0.060    0.007    -0.250   0.147    0.000    0.178
0.527    0.071    -0.028   -0.235   0.248    0.000    0.154
0.591    0.058    0.001    -0.199   0.277    0.000    0.181
1.127    1.000    0.054    -0.008   -0.100   0.064    0.347    0.068
1.190    1.000    0.028    -0.025   -0.131   0.165    0.369    0.071
1.150    1.000    0.068    -0.025   -0.091   0.095    0.350    0.058
1.160    1.000    0.064    -0.017   -0.156   0.141    0.366    0.075
1.179    1.000    0.041    -0.032   -0.125   0.169    0.393    0.036
1.232    1.000    0.078    -0.082   -0.205   0.169    0.404    0.047
1.194    1.000    0.063    -0.103   -0.196   0.152    0.442    0.035
1.130    1.000    0.042    -0.069   -0.166   0.153    0.433    0.032
1.151    1.000    0.051    -0.050   -0.189   0.115    0.406    0.050
1.114    1.000    0.039    -0.034   -0.169   0.107    0.367    0.062
1.141    1.000    0.016    -0.047   -0.148   0.159    0.382    0.052
1.133    1.000    0.075    -0.027   -0.178   0.144    0.367    0.141
C:\SDSM\Blogsville\observed1961-90\PRCP.DAT

For Help, press F1
```

Figure 10.7 shows a *.PAR file for a precipitation model

Figure 10.8 shows a *.SIM file used to downscale daily precipitation from observed (NCEP) predictors using the PRCP61-90.PAR model.



```
PRCPNCEP61-90.SIM - W...
File Edit View Insert Format Help

 4
 12
 366
01/01/1961
 10957
#TRUE#
 20
 18
 2
 0.8
PRCP.DAT
ncepp__vxx.dat
ncepp__zxx.dat
ncepp500xx.dat
ncepshumxx.dat

For Help, press F1
```

Figure 10.8 The *.SIM file for downscaling precipitation at Blogsville 1961–1990.

Figure 10.8 shows that four predictors were employed (line 1), to simulate 12 months (line 2), using calendar years (line 3), beginning in 01/01/1961 (line 4) and lasting 10957 days (line 5). The model was conditional (#TRUE#, line 6), had 20 ensemble members (line 7), variance inflation (line 8), a fourth root transformation of the predictand (line 9) and bias correction of 0.8 (line 10). The predictand file was *PRCP.DAT* and the four predictors were *p__v, p__z, p500*, and *shum* (lines 12 onwards).

With the above specifications, the **Weather Generator** was used to downscale observed (NCEP) predictors, and **Scenario Generator** to downscale GCM (HadCM3) predictors representing the present climate (saved as PRCPNCEP61-90.OUT and PRCPCCF61-90.OUT respectively). (Note that a **Year Length** of 366 days should be checked in **Settings** when working with NCEP, and 360 when using HadCM3 predictors). Downscaled scenarios were evaluated, firstly using the **Summary Statistics**, and then **Compare Results**. Figure 10.9 shows the summary statistics for the downscaling using NCEP predictors and Figure 10.10 shows the equivalent results for the downscaling using GCM predictors.

Results

File Help

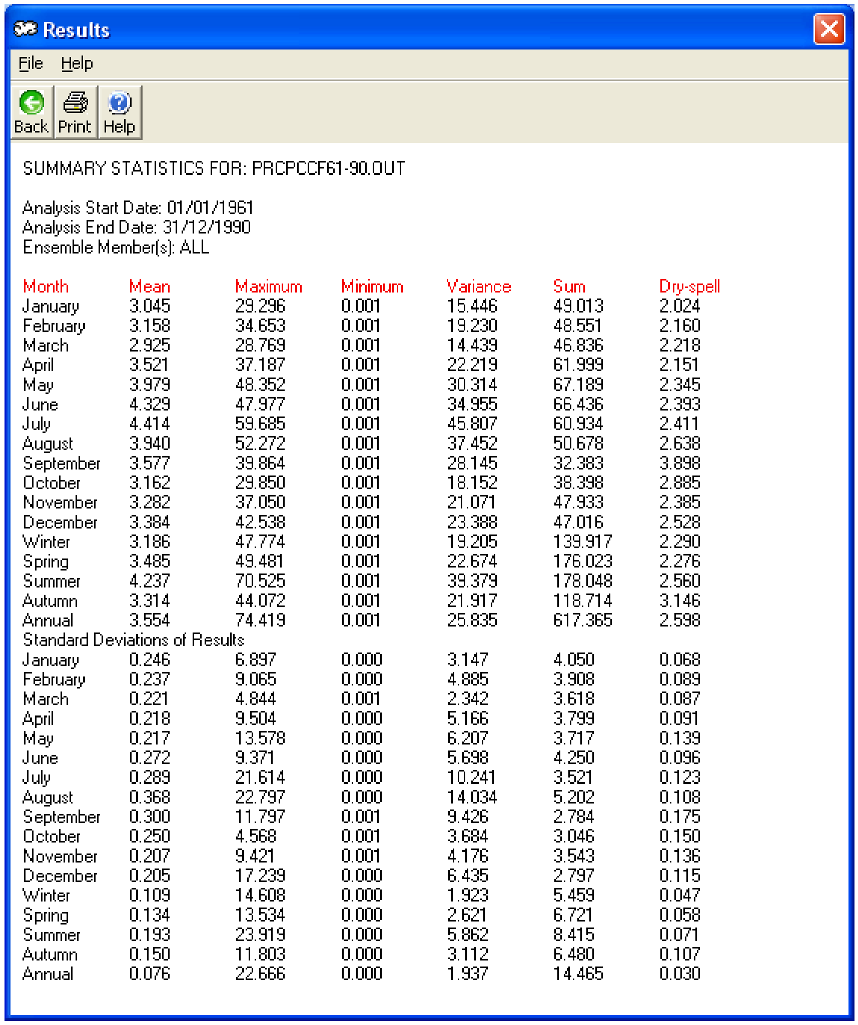
Back Print Help

SUMMARY STATISTICS FOR: PRCPNCEP61-90.OUT

Analysis Start Date: 01/01/1961
Analysis End Date: 31/12/1990
Ensemble Member(s): ALL

Month	Mean	Maximum	Minimum	Variance	Sum	Dry-spell
January	2.958	30.887	0.001	15.054	49.684	2.009
February	3.041	33.451	0.001	18.197	43.055	2.256
March	2.654	27.855	0.001	12.647	42.008	2.300
April	3.038	35.336	0.001	17.751	43.606	2.537
May	3.413	40.340	0.001	24.804	49.815	2.578
June	4.072	50.800	0.001	34.787	52.503	2.823
July	4.348	48.217	0.001	41.302	50.664	3.021
August	4.168	56.362	0.001	41.588	57.456	2.516
September	3.750	40.965	0.001	28.262	46.117	2.864
October	3.080	34.071	0.001	17.324	43.990	2.388
November	3.186	37.392	0.002	19.918	49.407	2.121
December	3.282	35.233	0.001	20.626	50.504	2.241
Winter	3.090	39.488	0.001	17.917	138.622	2.221
Spring	3.024	43.743	0.001	18.371	135.429	2.544
Summer	4.191	64.568	0.001	39.286	160.623	2.900
Autumn	3.315	45.188	0.001	21.598	139.513	2.518
Annual	3.374	65.833	0.001	23.922	578.808	2.575
Standard Deviations of Results						
January	0.150	6.110	0.001	2.560	2.511	0.106
February	0.236	7.738	0.001	3.719	3.673	0.075
March	0.118	5.199	0.000	1.753	2.192	0.131
April	0.151	8.829	0.000	2.789	2.369	0.106
May	0.209	10.826	0.000	5.119	3.062	0.089
June	0.279	14.049	0.001	8.536	3.488	0.107
July	0.231	13.791	0.000	8.678	2.944	0.143
August	0.341	15.253	0.001	10.446	5.030	0.133
September	0.276	8.306	0.000	5.904	4.040	0.167
October	0.155	8.923	0.000	3.664	2.601	0.067
November	0.164	7.946	0.001	2.602	3.238	0.085
December	0.212	6.692	0.000	4.002	3.222	0.084
Winter	0.116	6.728	0.000	1.819	5.653	0.056
Spring	0.070	9.800	0.000	2.149	3.442	0.070
Summer	0.186	12.154	0.000	5.613	6.697	0.085
Autumn	0.113	7.197	0.000	2.388	5.599	0.074
Annual	0.055	10.472	0.000	1.525	8.560	0.035

Figure 10.9 Summary statistics for downscaled precipitation using observed (NCEP) predictors



SUMMARY STATISTICS FOR: PRCPCCF61-90.OUT

Analysis Start Date: 01/01/1961
Analysis End Date: 31/12/1990
Ensemble Member(s): ALL

Month	Mean	Maximum	Minimum	Variance	Sum	Dry-spell
January	3.045	29.296	0.001	15.446	49.013	2.024
February	3.158	34.653	0.001	19.230	48.551	2.160
March	2.925	28.769	0.001	14.439	46.836	2.218
April	3.521	37.187	0.001	22.219	61.999	2.151
May	3.979	48.352	0.001	30.314	67.189	2.345
June	4.329	47.977	0.001	34.955	66.436	2.393
July	4.414	59.685	0.001	45.807	60.934	2.411
August	3.940	52.272	0.001	37.452	50.678	2.638
September	3.577	39.864	0.001	28.145	32.383	3.898
October	3.162	29.850	0.001	18.152	38.398	2.885
November	3.282	37.050	0.001	21.071	47.933	2.385
December	3.384	42.538	0.001	23.388	47.016	2.528
Winter	3.186	47.774	0.001	19.205	139.917	2.290
Spring	3.485	49.481	0.001	22.674	176.023	2.276
Summer	4.237	70.525	0.001	39.379	178.048	2.560
Autumn	3.314	44.072	0.001	21.917	118.714	3.146
Annual	3.554	74.419	0.001	25.835	617.365	2.598
Standard Deviations of Results						
January	0.246	6.897	0.000	3.147	4.050	0.068
February	0.237	9.065	0.000	4.885	3.908	0.089
March	0.221	4.844	0.001	2.342	3.618	0.087
April	0.218	9.504	0.000	5.166	3.799	0.091
May	0.217	13.578	0.000	6.207	3.717	0.139
June	0.272	9.371	0.000	5.698	4.250	0.096
July	0.289	21.614	0.000	10.241	3.521	0.123
August	0.368	22.797	0.000	14.034	5.202	0.108
September	0.300	11.797	0.001	9.426	2.784	0.175
October	0.250	4.568	0.001	3.684	3.046	0.150
November	0.207	9.421	0.001	4.176	3.543	0.136
December	0.205	17.239	0.000	6.435	2.797	0.115
Winter	0.109	14.608	0.000	1.923	5.459	0.047
Spring	0.134	13.534	0.000	2.621	6.721	0.058
Summer	0.193	23.919	0.000	5.862	8.415	0.071
Autumn	0.150	11.803	0.000	3.112	6.480	0.107
Annual	0.076	22.666	0.000	1.937	14.465	0.030

Figure 10.10 Summary statistics for downscaled precipitation using GCM predictors

Using Compare Results (Section 11), Figure 10.11 shows, for example, that the downscaling produced similar monthly mean daily totals under observed (NCEP) and GCM (HadCM2) forcing for the present climate.

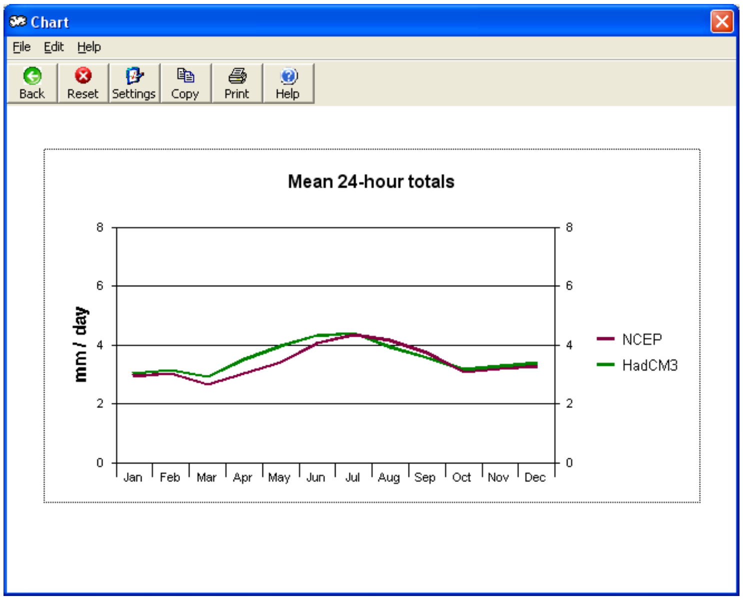


Figure 10.11 Monthly mean daily precipitation totals at Blogsville for the present climate downscaled using observed (NCEP) predictors (1961–1990) and GCM (HadCM3) predictors (1961–1990)

The **Scenario Generator** operation was implemented for a second time using HadCM3 predictors under present (1961–1990) and future (2070–2099) climate forcing. Figure 10.12 shows dry-spell lengths plotted by the **Compare Results** option. The results signal a shift to longer dry-spells in late summer and autumn.

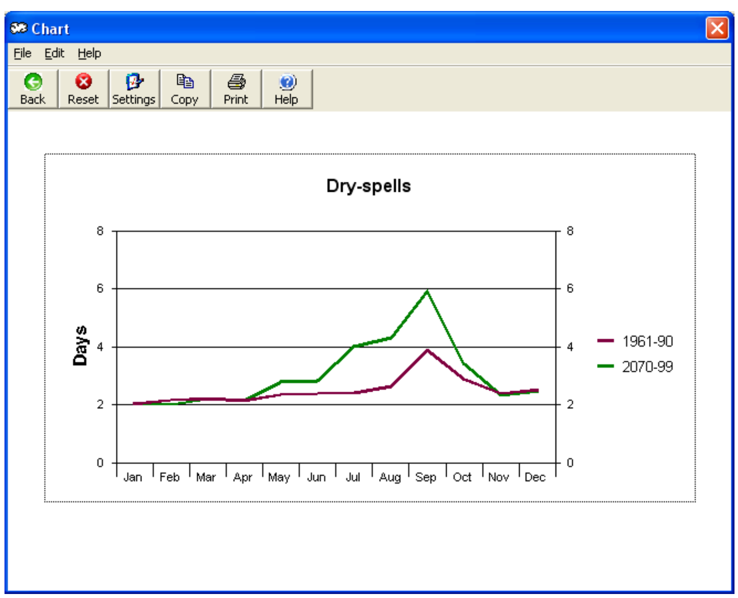


Figure 10.13 Monthly mean dry–spell lengths at Blogsville downscaled using HadCM3 predictors under present (1961–1990) and future (2070–2099) forcing.

11 GRAPHING MONTHLY STATISTICS

The **Compare Results** operation enables the User to plot monthly statistics produced by the **Summary Statistics** screen (Section 8). Graphing options allow the comparison of two sets of results and hence rapid assessment of downscaled versus observed, or present versus future climate scenarios.

To access this facility, click the **Compare Results** button at the top of any main screen. The following screen will appear:

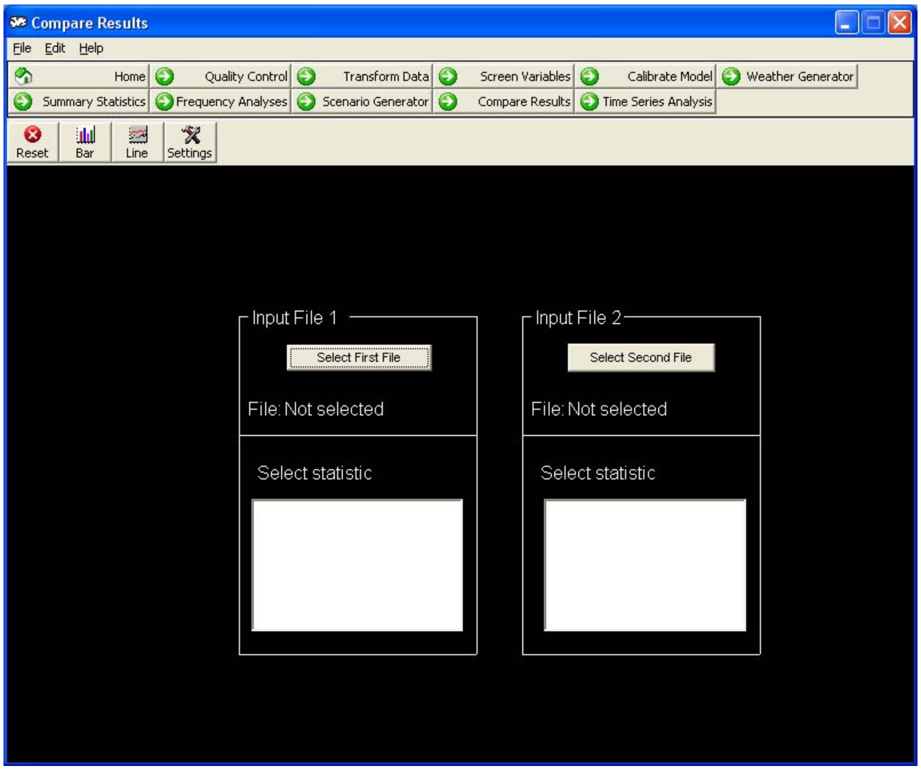


Figure 11.1 The **Compare Results** screen

11.1 Line chart

To choose a results (*.TXT) file, click on **Select First File** button. An **Open** file window appears; browse through until the correct directory and file are reached, then click on the appropriate file name – for example, observed statistics for maximum daily temperature at Blogsville, might have been stored in TMAXOBS60-91.TXT. The name of the file will then appear beneath the button, along with a list of available statistics. Repeat the process by clicking on the **Select Second File** button. Then click on the required statistic listed under **Select Statistic**. Finally, to show a line chart click on the **Line** button at the top of the screen:

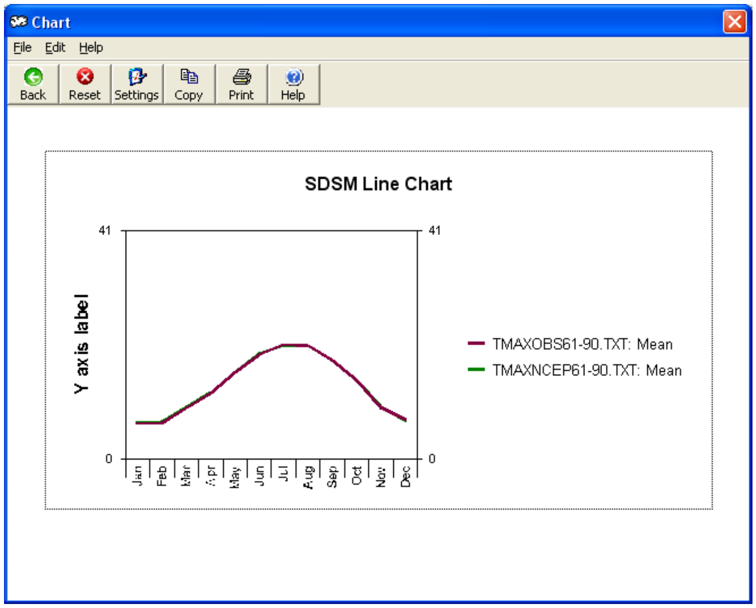


Figure 11.2 Example of the **Line** chart using observed (TMAXOBS61-90) and NCEP downscaled (TMAXNCEP61-90) monthly mean maximum daily temperatures at Blogsville 1976–1990.

11.2 Bar chart

Alternatively, having selected the required files and statistics from each list (as in Section 11.1), click on the **Bar** button at the top of the **Compare Results** screen to produce a bar chart:

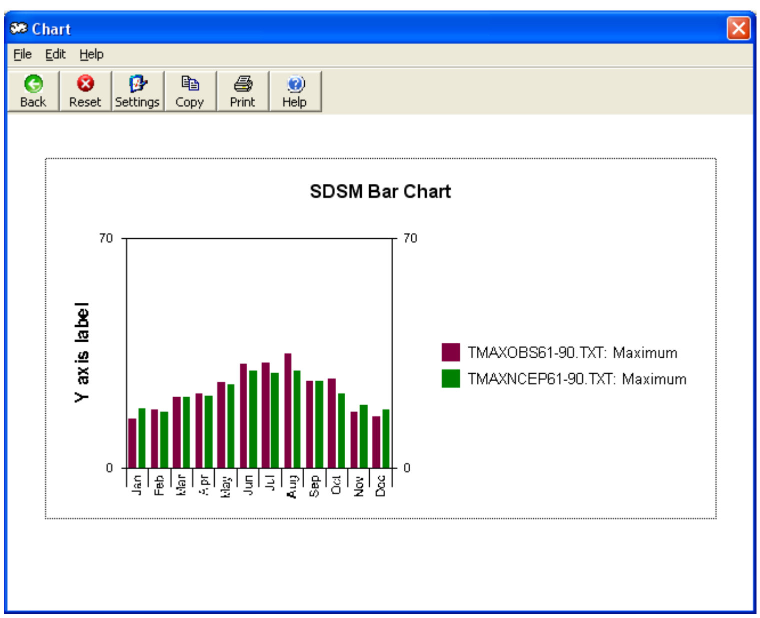


Figure 11.3 Example of the **Bar** chart using observed (TMAXOBS61-90) and downscaled (TMAXNCEP61-90) monthly absolute maximum daily temperatures at Blogsville 1976–1990

11.3 Customizing charts

To change (or remove) tick marks, y-axis labels, chart titles or y-axis maximum/minimum, in either the **Line** or **Bar** chart, click on the **Settings** button at the top of the screen. The following screen will appear:

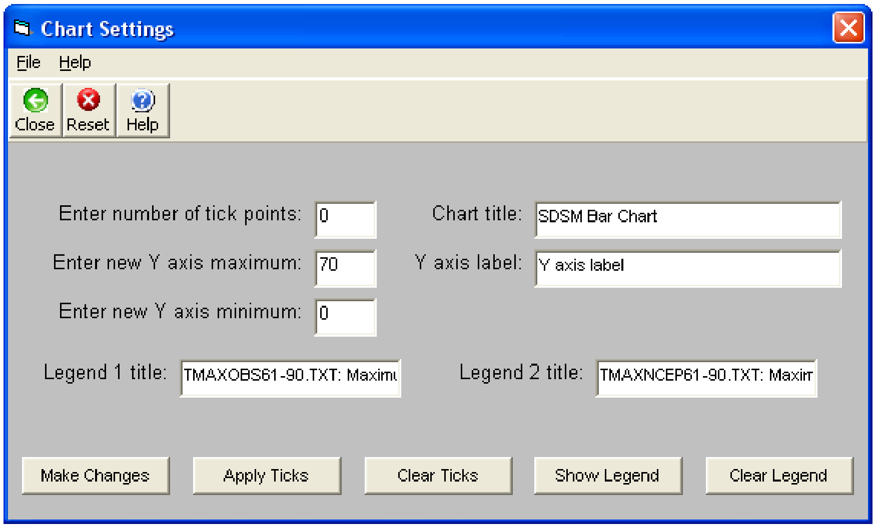


Figure 11.4 An illustration of the **Chart Settings** screen

Enter the required values/text then click on the **Make Changes** button (to change text) and/or click on the **Apply Ticks** button to change tick marks. Similarly, click on the **Clear Ticks**, **Show Legend**, or **Clear Legend** buttons as required. Then click on the **Back** button to return to the plot.

To change the colour scheme of the lines or bars, double click on the object required. A **Colour** palette will appear. Select the desired colour, then **OK** to return to the chart. Similarly, to change the **Font**, double click on title and/or y-axis title. To change the position of the title, single click then drag to the required location on the chart.

By applying the following design preferences, in Figure 11.5, it is possible to customize the bar chart in Figure 11.3 to that shown in Figure 11.6.

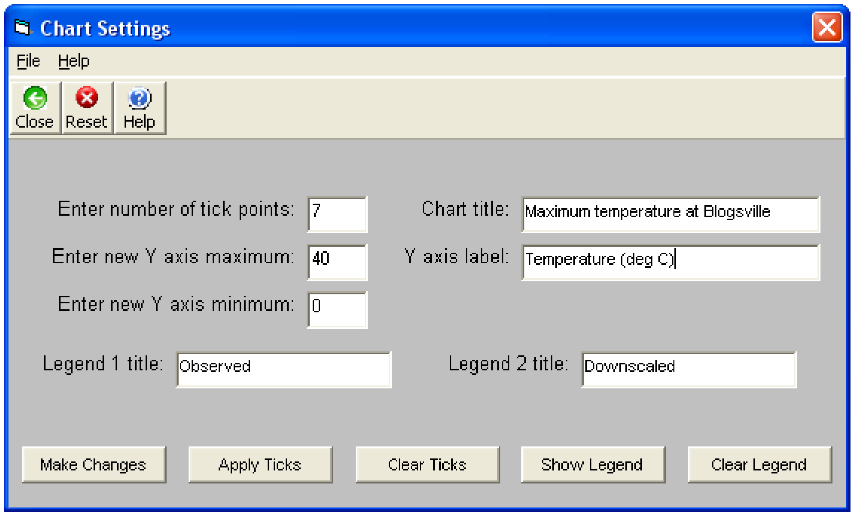


Figure 11.5 Design preferences entered into Chart Settings screen

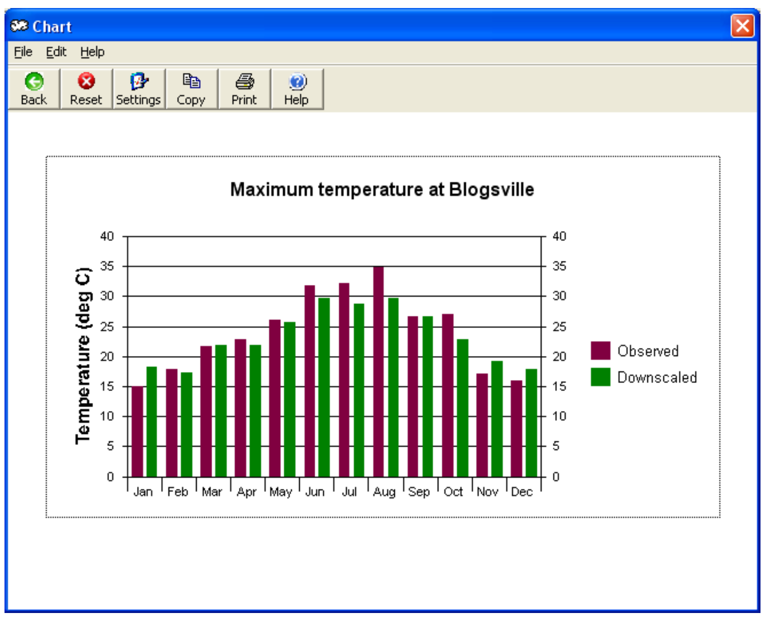


Figure 11.6 The same as Figure 11.3 but with customised bar colour scheme, tick marks, scale and legend.

Finally, to incorporate **Line** or **Bar** charts in a Word document, first use the **Copy** button at the top of the screen, then in Word use Paste Special (Picture).

12 TIME SERIES ANALYSIS

12.1 Time series chart

The **Time Series Analysis** screen allows the User to produce a time series plot of chosen data file(s). Up to a maximum of five files can be plotted simultaneously on the same chart. This screen employs files that contain a single column of data, so when using downscaled ensemble output some prior data handling must be undertaken using the **Frequency Analysis** (see Section 9) screen to extract individual members. To access the plotting facility click on the **Time Series Analysis** button at the top of any main screen. The following screen will appear:

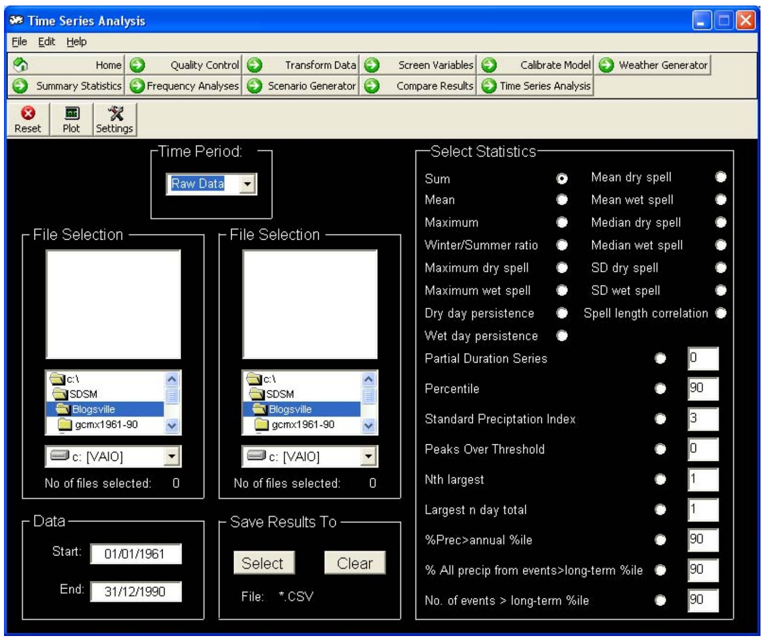


Figure 12.1 The **Time Series Analysis** screen

File Selection

Using the Drive, Directory and **File Selection** boxes, the User can select up to five files to plot. Note that two **File Selection** windows are provided to allow the User to select files from different directories. Only a maximum of five files from the two **File Selection** windows can be selected in total.

Data

Allows the User to specify the required time period. Note that if an attempt is made to plot a period longer than the available data set (as defined in the global **Settings**), an error message will appear.

Save Results To

The User can choose to save a summary of the calculated results to a data file. The option will not work if plotting **Raw Data** as no summary statistics are calculated in this case. The default format text file is comma separated (*.CSV) so data can be opened in a spreadsheet for further analysis. Clicking on the **Clear** button deselects the selected file.

Time Period

Allows the User to select from **Raw Data**, **Month**, **Season**, **Annual** or **Water Year**. **Raw Data** simply graphs the data from the chosen file(s) as a daily time series plot for the selected period (set by the User under **Data Start** and **Data End**). No statistics are derived for **Raw Data**. Note, the **Water Year** runs from October to September and is referred to by the year in which it starts. **Season** is referred to by the year in which it ends. Winter (December, January, February) is referred to by the year in which the January, February fall

When selecting a **Month**, **Season**, **Annual** or **Water Year Time Period**, SDSM calculates the chosen metric (from the list below **Select Statistics**) for the specified period and plots them on a line graph. For example, if the User chooses a **Time Period** of January, and selects **Sum**, SDSM will plot the annual series of January Sums for the selected fit period (i.e., the sum for January 1961, sum for January 1962, and so on) as a line chart.

Select Statistics

The User selects the summary statistics to be plotted by clicking the appropriate check button in this section (default is **Sum**). SDSM calculates the chosen statistic for the selected **Time Period**, repeated across the range of the fit period, and plots these as a time series chart. A number of these statistics are based on the widely used STARDEX indices (see: Goodess et al., 2007)

Sum, Mean, Maximum: are self-explanatory measures for the selected time period.

Winter/Summer ratio: is calculated as the sum of the winter data (December, January, February), divided by the sum of the following summer data (June, July, August). The metric is referenced to the year in which the summer period falls.

Maximum dry (wet) spell: the maximum dry (wet) spell length in days for the given time period.

Dry (wet) day persistence: the total number of consecutive dry (wet) days for a period divided by the total number of dry (wet) days in that period.

Mean dry (wet) spell: mean dry (wet) spell length for the period.

Median dry (wet) spell: median dry (wet) spell length for the period.

SD dry (wet) spell: standard deviation of dry (wet) spell length for the period.

Spell length correlation: a measure of the combined persistence of wet- and dry-spells for the period.

Partial Duration Series: is calculated as the sum of data values less than or equal to the chosen threshold for the selected **Time Period**. The default is the threshold value held in the main **Settings** screen. This value can be adjusted by entering the required threshold in the text box (this will not affect the global threshold value set in the main **Settings** screen and applied elsewhere).

Percentile calculates the specified percentile for the chosen **Time Period**. The default is 90% but this can be adjusted by entering the required value in the text box.

Standard Precipitation Index (SPI): This is calculated for monthly time series only, so the Time Period selection is ignored when SPI is chosen. The SPI is derived by first calculating the monthly sums of the data, then calculating a moving average of these monthly sums (smoothing) across the time period entered by the User in the adjacent text box. The default moving average period is 3 months. The smoothed data are then normalised by subtracting the mean of all the data in the fit range and dividing by the standard deviation of the smoothed data for each month.

Peaks Over Threshold (POT). This counts the number of events greater than the user specified threshold for the chosen time period.

Nth largest: determines the nth largest value when the data are sorted into descending order for a given time period.

Largest n day total: calculates the n day total for all possible windows in the time period and presents the largest value.

%Prec>annual %ile: Percentage of total precipitation above the specified annual percentile.

% All precip from events>long-term %ile: Percentage of all precipitation from events that are greater than the specified long term percentile.

No. of events > long-term %ile: Count of the total number of events in the time period that are greater than the specified long-term percentile.

Plot

By clicking the **Plot** button the selected statistics are displayed as a time series graph, as in Figure 12.2. In this case the raw data from PRCP.DAT are plotted as a line chart covering the period 1961-1970 (the maximum number of years that can be plotted as a line chart is 10 years).

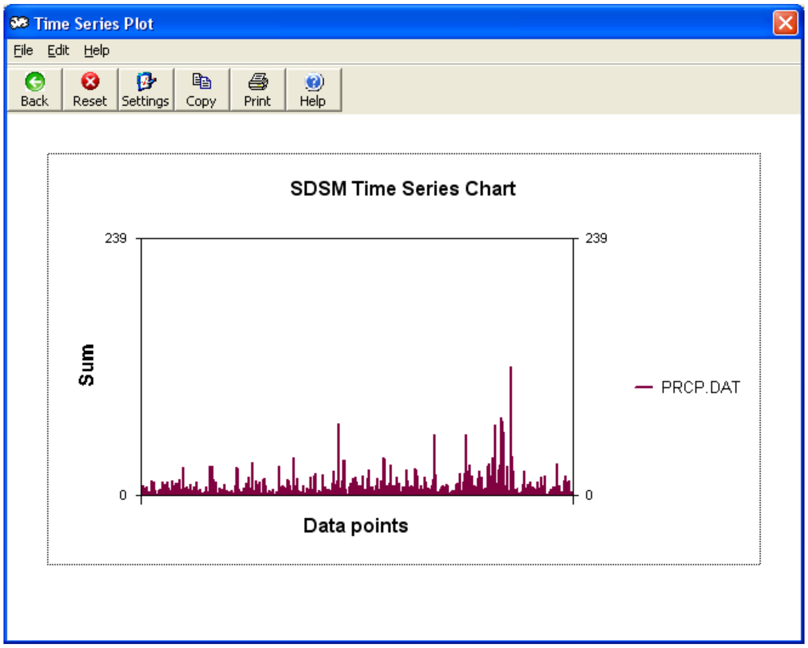


Figure 12.2 Time series plot of raw data from PRCP.DAT file

12.2 Adjusting chart appearance

The appearance of the time series chart can be adjusted in several ways. For example, by clicking on the **Settings** button, the User is presented with a settings form that allows various adjustments to be made (Figure 12.3).

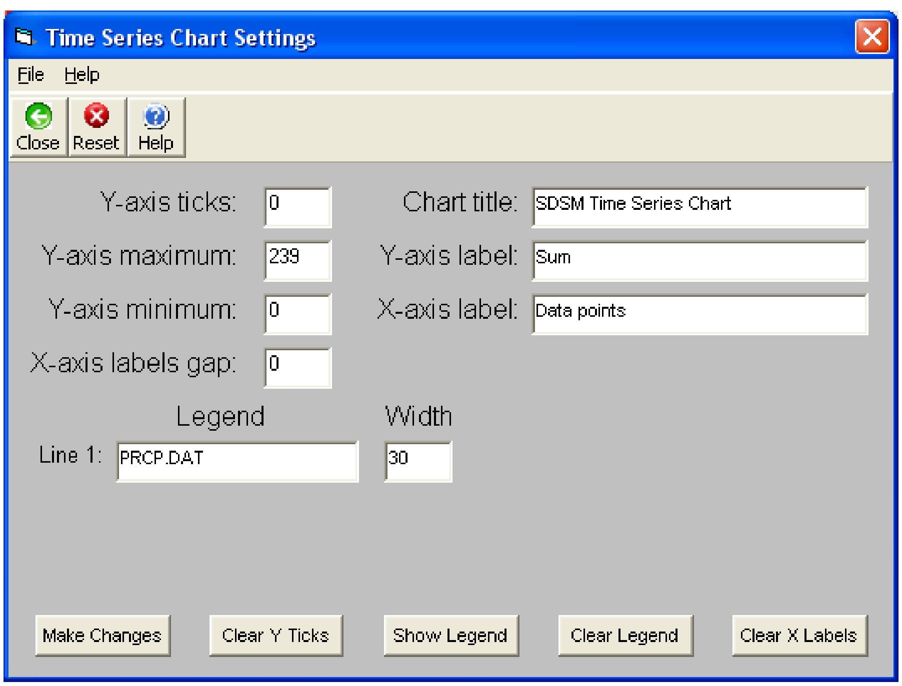


Figure 12.3 An example of the **Time Series Chart Settings** form

The following points explain how each of these options work:

Lines

The User can adjust the width and legend text of each line in the chart. After making the required changes the User must click the **Make Changes** button to apply the changes to the chart.

Legend

The User can choose to show the legend on the chart or remove the legend by clicking the appropriate buttons at the bottom of this screen.

Y-axis ticks

These refer to the tick lines drawn across the chart on the y-axis. The default is no tick lines (except for the line y=0). If the User wishes to apply y-axis tick lines, enter the number required in the text box here and click the **Make Changes** button. Clicking the **Clear Y Ticks** button removes these tick lines.

Y-axis range

The User can adjust the extent of the Y-axis by entering appropriate minimum and maximum values on this page and clicking the **Make Changes** button.

X-axis labels

For analysed data year markers are shown on the X-axis. For raw data it is possible to apply a number of data markers (a counter) on the X-axis. X-axis label spacing can be specified by entering an appropriate value in the **X-axis labels gap** text box. This specifies the interval between successive X-axis labels/markers. The default for analysed data is 1 (i.e. markers appear every year). The default for raw data is 0 (i.e. no markers appear). Note that when plotting SPI data series the X-axis labels are determined by the total number of months available. In this case it may be better to remove the labels entirely to avoid overcrowding on the X-axis. The User implements chart settings by clicking the **Make Changes** button. X-axis labels are removed by clicking the **Clear X Labels** button.

Labels

The User can adjust the text appearing on the X- and Y-axis and also the chart title by typing in the appropriate text on this screen and clicking the **Make Changes** button. The User can also make adjustments to the chart directly. For example, by double-clicking on the lines the User can adjust their colour. Double clicking on the title and axis labels allows the User to change the text font. The axis labels, chart title or legend are removed by clicking on them and hitting delete or backspace. To incorporate the **Chart** in a Word document, first use the **Copy** button at the top of the window then, in Word, use Paste Special (Picture).

Figure 12.4 below provides an example of a time series plot generated using **Largest n day total** statistic.

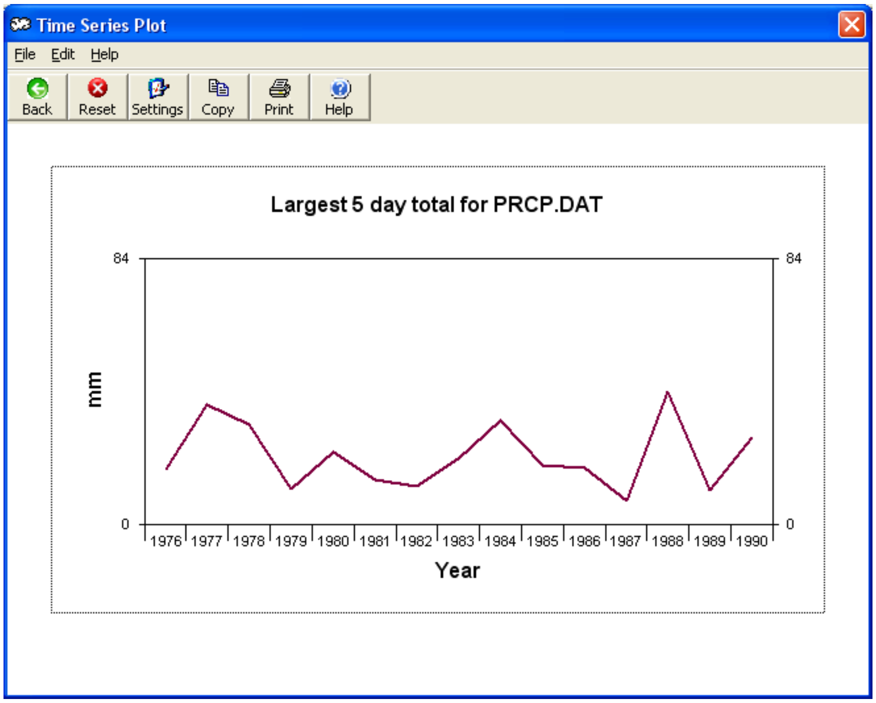


Figure 12.4 An example of an annual time series plot of the largest 5-day total rainfalls at Blogsville, 1976-1990

13 FINAL CAUTIONARY REMARKS

SDSM is a Windows–based decision support tool for the rapid development of single–site, ensemble scenarios of daily weather variables under present and future regional climate forcing. Version 4.2 performs the tasks required to statistically downscale climate model output, namely: quality control of input data; screening of candidate predictor variables; model calibration; synthesis of present weather data; generation of future climate scenarios; basic statistical and time series analyses; and graphing results. SDSM provides a robust and parsimonious technique of scenario construction that complements other methods (e.g., direct use of climate model output, dynamical downscaling, sensitivity analysis, etc.). Prospective Users should, however, consider the relative strengths and weaknesses of each category of downscaling to determine whether SDSM is most appropriate for the task in hand.

The authors strongly caution that the software should **not** be used uncritically as a “black box”. This is a very real danger when employing regression–based modelling techniques. Rather, the downscaling should be based upon physically sensible linkages between large–scale forcing and local meteorological response. Therefore, good practice demands rigorous evaluation of candidate predictor–predictand relationships using independent data. Furthermore, the local knowledge base is an invaluable source of information when determining sensible combinations of predictors.

Daily precipitation amounts at individual stations continue to be the most problematic variable to downscale, and research is ongoing to address this limitation. This arises because of the generally low predictability of daily precipitation amounts at local scales by regional forcing factors. The unexplained behaviour is currently modelled stochastically within SDSM by artificially inflating the variance of the downscaled series to accord better with daily observations. Even so, the model can produce unrealistic behaviour if the stochastic component is not properly handled. This, again, underlines the importance of independent testing of all model parameters against data withheld from model calibration.

Ultimately, however, the plausibility of all SDSM scenarios depends on the realism of the climate model forcing. Systematic biases in the mean and variance of GCM predictors can be reduced through normalisation with respect to a control period (as in the case of all pre-prepared SDSM predictors). Biases in large–scale patterns of atmospheric circulation in GCMs (e.g. shifts in the dominant storm track relative to observed data) or unrealistic inter–variable relationships are much harder to accommodate. Where possible, Users should not, therefore, restrict themselves to the use of a single GCM or emission scenario for downscaling. By applying multiple forcing scenarios (via different GCMs, ensemble members, time–slices, or emission pathways) better insight may be gained into the magnitude of these uncertainties.

Finally, the authors welcome constructive suggestions about the design or application of SDSM, particularly from the wider climate change impacts community. The authors would also appreciate copies of any publications or reports arising from the use of SDSM. This helps share experience with other Users, and adds to the knowledge base of projected climate changes in different regions.

BIBLIOGRAPHY

This Bibliography cites papers containing full technical details of SDSM, followed by example case studies. Additional overview material is also recommended for other downscaling methods, as well as selected review papers in which various downscaling methods have been compared. Note that an *IPCC-TGCIA Guidance Document on Statistical Downscaling* is available via the Data Distribution Centre.

Technical basis of SDSM

Conway, D., Wilby, R.L. and Jones, P.D. 1996. Precipitation and air flow indices over the British Isles. *Climate Research*, **7**, 169–183.

Hassan, H., Aramaki, T., Hanaki, K., Matsuo, T. and Wilby, R.L. 1998. Lake stratification and temperature profiles simulated using downscaled GCM output. *Journal of Water Science and Technology*, **38**, 217–226.

Narula, S.C. and Wellington, J.F. 1977. An algorithm for linear regression with minimum sum of absolute errors. *Applied Statistics*, **26**, 106–111.

Rubinstein, R.Y. 1981. *Simulation and the Monte Carlo method*. Wiley, New York.

Wilby, R.L., Dawson, C.W. and Barrow, E.M. 2001. SDSM – a decision support tool for the assessment of regional climate change impacts. *Environmental and Modelling Software*, **17**, 145–157.

Wilby, R.L., Tomlinson, O.J. and Dawson, C.W. 2003. Multi-site simulation of precipitation by conditional resampling. *Climate Research*, **23**, 183-194.

Example applications of SDSM

Abraham, L.Z. 2006. Climate change impact on Lake Ziway watershed water availability, Ethiopia. Unpublished MSc Thesis, University of Applied Sciences, Cologne, pp123.

Aspen Global Change Institute (AGCI), 2006. *Climate Change and Aspen: An Assessment of Impacts and Potential Responses*. Appendix B, p107-111. Aspen Global Change Institute, Colorado, pp147.

Bootsma, A., Gameda, S. and McKenney, D.W. 2005. Impacts of potential climate change on selected agroclimatic indices in Atlantic Canada. *Canadian Journal of Soil Science*, **85**, 329-343.

Crawford, T., Betts, N.L. and Favis-Mortlock, D.T. 2007. Issues of GCM grid box choice and predictor selection associated with statistical downscaling of daily precipitation over Northern Ireland. *Climate Research*, under review.

Diaz-Nieto, J. and Wilby, R.L. 2005. A comparison of statistical downscaling and climate change factor methods: impacts on low flows in the River Thames, United Kingdom. *Climatic Change*, **69**, 245-268.

Fealy, R. 2006. An assessment of the relationship between glacier mass balance and synoptic climate in Norway: Likely future implications of climate change. Unpublished PhD Thesis, University of Maynooth, Ireland.

Goodess, C.M., Anagnostopoulo, C., Bardossy, A., Frei, C., Harpham, C., Haylock, M.R., Hundecha, Y., Maheras, P., Ribalaygua, J., Schmidli, J., Schmith, T., Tolika, K., Tomozeiu, R. and Wilby, R.L. 2007. An intercomparison of statistical downscaling methods for Europe and European regions - assessing their performance with respect to extreme temperature and precipitation events. *Climatic Change*, in press.

Guangul, S.G. 2003. *Modelling the effect of climate and land-use changes on hydrological processes: An integrated GIS and distributed modelling approach*. Published PhD Thesis, Vrije Universiteit, Brussels, Belgium.

Harpham, C. and Wilby, R.L. 2005. Multi-site downscaling of heavy daily precipitation occurrence and amounts. *Journal of Hydrology*, **312**, 235-255.

Haylock, M.R., Cawley, G.C., Harpham, C., Wilby, R.L. and Goodess, C.M. 2006. Downscaling heavy precipitation over the UK: a comparison of dynamical and statistical methods and their future scenarios. *International Journal of Climatology*, **26**, 1397-1415.

Khan, M.S., Coulibaly, P. and Dibike, Y. 2006. Uncertainty analysis of statistical downscaling methods. *Journal of Hydrology*, **319**, 357-382.

Khan, M.S., Coulibaly, P. and Dibike, Y. 2006. Uncertainty analysis of statistical downscaling methods using Canadian Global Climate Model predictors. *Hydrological Process*, **20**, 3085-3104.

Lines, G.S. and Pancura, M. 2005. Building climate change scenarios of temperature and precipitation in Atlantic Canada using the Statistical DownScaling Model (SDSM). Meteorological Service of Canada, Atlantic Region. Science Report series 2005-9, Dartmouth, Canada, pp41.

London Climate Change Partnership, 2002. *A climate change impacts in London evaluation study*. Final Technical Report, Entec UK Ltd.

MacDonald, O. 2004. Coupling glacier mass balance and meltwater yields in the European Alps with future climate change: downscaling from integrations of the HadCM model. Unpublished PhD Thesis, University of Salford, UK.

Reynard, N., Crooks, S., Wilby, R.L. and Kay, A. 2004. Climate change and flood frequency in the UK. *Proceedings of the 39th Defra Flood and Coastal Management Conference*, University of York, UK.

Scibek, J. and Allen, D.M. 2006. Modeled impacts of predicted climate change on recharge and groundwater levels. *Water Resources Research*, **42**, W11405.

Wetterhall, F., Bárdossy, A., Chen, D., Halldin, S., and Xu, C. 2007. Daily precipitation-downscaling techniques in three Chinese regions. *Water Resources Research*, **42**, W11423, doi:10.1029/2005WR004573.

Wetterhall, F., Halldin, S. and Xu, C.Y. 2007. Seasonality properties of four statistical-downscaling methods in central Sweden. *Theoretical and Applied Climatology*, **87**, 123-137.

Whitehead, P.G., Futter, M. and Wilby, R.L. 2006. Impacts of climate change on hydrology, nitrogen and carbon in upland and lowland streams: Assessment of adaptation strategies to meet Water Framework Directive objectives. *Proceedings of the British Hydrological Society Conference*, Durham, UK.

Whitehead, P.G., Wilby, R.L., Butterfield, D., and Wade, A.J. 2006. Impacts of climate change on nitrogen in a lowland chalk stream: An appraisal of adaptation strategies. *Science of the Total Environment*, **365**, 260-273.

Wilby, R.L. 2003. Past and projected trends in London's urban heat island. *Weather*, **58**, 251-260.

Wilby, R.L. 2005. Constructing wet season precipitation scenarios for a site in the Anti Atlas Mountains, Morocco. Proceedings of the *Conference on Optimising Land and Water Resources in Arid Environments*, Agadir, Morocco.

Wilby, R.L. 2007. Constructing climate change scenarios of urban heat island intensity and air quality. *Environment and Planning B: Planning and Design*, under review.

Wilby, R.L. and Dettinger, M.D. 2000. Streamflow changes in the Sierra Nevada, CA simulated using a statistically downscaled General Circulation Model scenario of climate change. In: McLaren, S.J. and Kniveton, D.R. (Eds.), *Linking Climate Change to Land Surface Change*, Kluwer Academic Publishers, Netherlands, pp. 99–121.

Wilby, R.L., Hassan, H. and Hanaki, K. 1998b. Statistical downscaling of hydrometeorological variables using general circulation model output. *Journal of Hydrology*, **205**, 1-19.

Wilby, R.L., Whitehead, P.G., Wade, A.J., Butterfield, D., Davis, R. and Watts, G. 2006. Integrated modelling of climate change impacts on the water resources and quality in a lowland catchment: River Kennet, UK. *Journal of Hydrology*, **330**, 204-220.

Wilby, R.L.and Harris, I. 2006. A framework for assessing uncertainties in climate change impacts: low flow scenarios for the River Thames, UK. *Water Resources Research*, **42**, W02419, doi:10.1029/2005WR004065.

Downscaling overviews and general guidance

Christensen, J.H. and Hewitson, B.C. 2007. *Regional climate projections*. IPCC WG1 Fourth Assessment Report, Chapter 11, forthcoming.

Fowler, H., Blenkinsop, S. and Tebaldi, C. 2007. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology,* in press.

Goodess, C., Osborn, T. and Hulme, M. 2003. The identification and evaluation of suitable scenario development methods for the estimation of future probabilities of extreme weather events. Tyndall Centre for Climate Change Research, Technical Report 4. http://www.tyndall.ac.uk/research/theme3/final_reports/it1_16.pdf

Leung, L.R., Mearns, L.O., Giorgi, F. and Wilby, R.L. 2003. Regional climate research: needs and opportunities. *Bulletin of the American Meteorological Society*, **84**, 89-95.

Lu, X. 2006. Guidance on the Development of Climate Scenarios within the Framework of National Communications from Parties not Included in Annex I (NAI) to the United Nations Framework Convention on Climate Change (UNFCCC). National Communications Support Programme (NCSP), UNDP-UNEP-GEF, in press.

Mearns, L.O., Giorgi, F., Whetton, P., Pabon, D., Hulme, M. and Lal, M. 2003. *Guidelines for use of climate scenarios developed from Regional Climate Model experiments.* IPCC Task Group on Scenarios for Climate Impact Assessment (TGCIA).

Prudhomme, C., Reynard, N. and Crooks, S. 2002. Downscaling of global climate models for flood frequency analysis: Where are we now? *Hydrological Processes* **16**, 1137-1150.

Wilby, R.L. and Wigley, T.M.L. 1997. Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography*, **21**, 530-548.

Wilby, R.L. and Wigley, T.M.L. 2000. Precipitation predictors for downscaling: observed and General Circulation Model relationships. *International Journal of Climatology*, **20**, 641-661.

Wilby, R.L., Charles, S., Mearns, L.O., Whetton, P., Zorito, E. and Timbal, B. 2004. *Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling Methods.* IPCC Task Group on Scenarios for Climate Impact Assessment (TGCIA).

Wilks, D.S. and Wilby, R.L. 1999. The weather generation game: a review of stochastic weather models. *Progress in Physical Geography* **23**, 329-357.

Xu, C.–Y. 1999. From GCMs to river flow: a review of downscaling methods and hydrologic modelling approaches. *Progress in Physical Geography*, **23**, 229–249.

NCEP re–analysis and Hadley Centre experiments

Gordon, C., Cooper, C, Senior, C., Banks, H., Gregory, J., Johns, T., Mitchell, J. and Wood, R. 2000. The simulation of SST, sea–ice extents and ocean heat transport in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, **16**, 147–168.

Johns, T.C., Carnell, R.E., Crossley, J.F., Gregory, J.M., Mitchell, J.F.B., Senior, C.A., Tett, S.F.B. and Wood, R.A. 1997. The Second Hadley Centre coupled ocean–atmosphere GCM: Model description, spinup and validation. *Climate Dynamics*, **13**, 103–134.

Kalnay, E., Kanamitsu, M., Kistler, R. *et al.* 1996. The NCEP/NCAR 40–year reanalysis project. *Bulletin of the American Meteorological Society*, **77**, 437–471.

Mitchell, J.F.B. and Johns, T.C. 1997. On modification of global warming by sulphate aerosols. *Journal of Climate*, **10**, 245–267.

Mitchell, J.F.B., Johns, T.C., Gregory, J.M. and Tett, S. 1995. Climate response to increasing levels of greenhouse gases and sulphate aerosols. *Nature,* **376**, 501–504.

Tett, S.F.B, Johns, T.C. and Mitchell, J.F.B. 1997. Global and regional variability in a coupled AOGCM. *Climate Dynamics*, **13**, 303–323.

Statistical downscaling methods

Bárdossy A., Bogardi, I. and Matyasovszky, I. 2005. Fuzzy rule-based downscaling of precipitation. *Theoretical and Applied Climatology* **82**, 119–129.

Bardossy, A. and Plate, E.J. 1992. Space–time model for daily rainfall using atmospheric circulation patterns. *Water Resources Research*, **28**, 1247–1259.

Beersma, J.J. and Buishand, T.A. 2003. Multi-site simulation of daily precipitation and temperature conditional on the atmospheric circulation. *Climate Research* **25**, 121–133.

Benestad, R.E. 2004. Tentative probabilistic temperature scenarios for northern Europe. *Tellus Series A – Dynamic Meteorology and Oceanography* **56**, 89-101.

Burger, G. 1996. Expanded downscaling for generating local weather scenarios. *Climate Research*, **7**, 111–128.

Bürger, G. and Chen, Y. 2005. Regression-based downscaling of spatial variability for hydrologic applications. *Journal of Hydrology* **311**, 299-317.

Cavazos, T. and Hewitson, B.C. 2005. Performance of NCEP-NCAR reanalysis variables in statistical downscaling of daily precipitation. *Climate Research* **28**, 95-107.

Charles, S.P., Bates, B.C., Smith, I.N., Hughes, J.P. 2004. Statistical downscaling of daily precipitation from observed and modelled atmospheric fields. *Hydrological Processes* **18**, 1373-1394.

Crane, R.G. and Hewitson, B.C. 1998. Doubled CO2 precipitation changes for the Susquehanna Basin: downscaling from the GENESIS general circulation model. *International Journal of Climatology*, **18**, 65–76.

Hay, L.E., McCabe, G.J., Wolock, D.M. and Ayers, M.A. 1991. Simulation of precipitation by weather type analysis. *Water Resources Research*, **27**, 493–501.

Huth, R. 1999. Statistical downscaling in central Europe: evaluation of methods and potential predictors. *Climate Research*, **13**, 91–101.

Huth, R. 2005. Downscaling of humidity variables: A search for suitable predictors and predictands. *International Journal of Climatology* **25**, 243-250.

McCabe, G.J. and Dettinger, M.D. 1995. Relations between winter precipitation and atmospheric circulation simulated by the Geophysical Fluid Dynamic Laboratory General Circulation Model. *International Journal of Climatology*, **15**, 625-638.

Schmidli, J., Frei, C. and Vidale, P.L. 2006. Downscaling from GCM precipitation: A benchmark for dynamical and statistical downscaling methods. *International Journal of Climatology*, **26**, 679-689.

Schubert, S. and Henderson-Sellers, A. 1997. A statistical model to downscale local daily temperature extremes from synoptic-scale atmospheric circulation patterns in the Australian region. *Climate Dynamics*, **13**, 223-234.

Wilby, R.L. 1997. Non–stationarity in daily precipitation series: implications for GCM downscaling using atmospheric circulation indices. *International Journal of Climatology*, **17**, 439–454.

Wilby, R.L. and Wigley, T.M.L. 2000. Precipitation predictors for downscaling: Observed and general circulation model relationships. *International Journal of Climatology* **20**, 641-661.

Weather generators

Katz, R.W. and Parlange, M.B. 1998. Overdispersion phenomenon in stochastic modeling of precipitation. *Journal of Climate*, **11**, 591-601

Kilsby, C.G., Cowpertwait, P.S.P., O'Connell, P.E. and Jones, P.D. 1998. Predicting rainfall statistics in England and Wales using atmospheric circulation variables. *International Journal of Climatology*, **18**, 523-539.

Kilsby, C.G., Jones, P.D., Burton, A., Ford, A.C., Fowler, H.J., Harpham, C., James, P., Smith, A. and Wilby, R.L. 2007. A daily weather generator for use in climate change studies. *Environmental Modelling and Software*, in press.

Palutikof, J.P., Goodess, C.M., Watkins, S.J. and Holt, T. 2002. Generating rainfall and temperature scenarios at multiple sites: Examples from the Mediterranean. *Journal of Climate* **15**, 3529-3548.

Qian, B., Hayhoe, H. and Gameda, S. 2005. Evaluation of the stochastic weather generators LARS-WG and AAFC-WG for climate change impact studies. *Climate Research* **29**, 3-21.

Richardson, C.W. 1981. Stochastic simulation of daily precipitation, temperature and solar radiation. *Water Resources Research*, **17**, 182–190.

Semenov, M.A. and Barrow, E.M. 1997. Use of a stochastic weather generator in the development of climate change scenarios. *Climatic Change*, **35**, 397–414.

Wilks, D.S. 1992. Adapting stochastic weather generation algorithms for climate change studies. *Climatic Change*, **22**, 67–84.

Wilks, D.S. 1999. Multisite downscaling of daily precipitation with a stochastic weather generator. *Climate Research*, **11**, 125–136.

Wilks, D.S. and Wilby, R.L. 1999. The weather generation game: a review of stochastic weather models. *Progress in Physical Geography*, **23**, 329–357.

Downscaling comparisons

Charles, S.P., Bates, B.C., Whetton, P.H. and Hughes, J.P. 1999. Validation of downscaling models for changed climate conditions: case study of southwestern Australia. *Climate Research*, **12**, 1–14.

Christensen, J.H. 2007. Evaluating the performance and utility of regional climate models in climate change research: Reducing uncertainties in climate change projections – the PRUDENCE approach. *Climatic Change, in press.*

Dibike, Y. B. and Coulibaly, P. 2005. Hydrologic impact of climate change in the Saguenay watershed: comparison of downscaling methods and hydrologic models. *Journal of Hydrology* **307**, 145-163.

Frei, C., Schöll, R., Fukutome, S., Schmidli, J. and Vidale, P.L. 2006. Future change of precipitation extremes in Europe: An intercomparison of scenarios from regional climate models. *Journal of Geophysical Research-Atmospheres* **111***,* D06105, doi:10.1029/2005JD005965.

Hanssen-Bauer, I., Achberger, C., Benestad, R.E., Chen, D. and Forland, E.J. 2005. Statistical downscaling of climate scenarios over Scandinavia. *Climate Research* **29**, 255-268.

Hay, L.E. and Clark, M.P. 2003. Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western United States. *Journal of Hydrology* **282**, 56-75.

Hay, L.E., Wilby, R.L. and Leavesley, G.H. 2000. A comparison of delta change and downscaled GCM scenarios for three mountainous basins in the United States. *Journal of the American Water Resources Association*, **36**, 387–397.

Hewitson, B.C. and Crane, R.G. 2006. Consensus between GCM climate change projections with empirical downscaling: Precipitation downscaling over South Africa. *International Journal of Climatology* **26**, 1315-1337.

Huth, R. 1999. Statistical downscaling in central Europe: evaluation of methods and potential predictors. *Climate Research* **13**, 91-101.

Kidson, J. W. and Thompson, C.S. 1998. A comparison of statistical and model–based downscaling techniques for estimating local climate variations. *Journal of Climate*, **11**, 735–753.

Mearns, L. O., Bogardi, I., Giorgi, F., Matayasovsky, I. and Palecki, M. 1999. Comparison of climate change scenarios generated daily temperature and precipitation from regional climate model experiments and statistical downscaling. *Journal of Geophysical Research* **104**, 6603-6621.

Mearns, L.O., Mavromatis, T., Tsvetsinskaya, E., Hays, C. and Easterling, W. 1999b. Comparative responses of EPIC and CERES crop models to high and low spatial resolution climate change scenarios. *Journal of Geophysical Research*, **104**, 6623–6646.

Murphy, J. 1999. An evaluation of statistical and dynamical techniques for downscaling local climate. *Journal of Climate*, **12**, 2256–2284.

Murphy, J. 2000. Predictions of climate change over Europe using statistical and dynamical downscaling techniques. *International Journal of Climatology* **20,** 489-501.

Salathé, E.P. 2003. Comparison of various precipitation downscaling methods for the simulation of streamflow in a rainshadow river basin. *International Journal of Climatology*, **23**, 887-901.

Schoof, J.T. and Pryor, S.C. 2001. Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks. *International Journal of Climatology* **21**, 773-790.

Takle, E.S., Gutowski, Jr., W.J., Arritt, R.W., Pan, Z., Anderson, C.J., Silva, R., Caya, D., Chen, S.-C., Christensen, J.H., Hong, S.-Y., Juang, H.-M. H., Katzfey, J.J., Lapenta, W.M., Laprise, R., Lopez, P., McGregor, J. and Roads, J.O. 1999. Project to Intercompare Regional Climate Simulations (PIRCS): Description and initial results. *Journal of Geophysical Research*, **104**, 19,443-19,462.

Wilby, R.L., Hay, L.E. and Leavesley, G.H. 1999. A comparison of downscaled and raw GCM output: implications for climate change scenarios in the San Juan River basin, Colorado. *Journal of Hydrology* **225**, 67-91.

Wilby, R.L., Hay, L.E., Gutowski, W.J., Arritt, R.W., Takle, E.S., Pan, Z., Leavesley, G.H. and Clark, M.P. 2000. Hydrological responses to dynamically and statistically downscaled climate model output. *Geophysical Research Letters*, **27**, 1199–1202.

Wilby, R.L., Wigley, T.M.L., Conway, D., Jones, P.D., Hewitson, B.C., Main, J. and Wilks, D.S. 1998.. Statistical downscaling of general circulation model output: A comparison of methods. *Water Resources Research* **34**, 2995-3008.

Winkler, J.A., Palutikof, J.P., Andresen, J.A. and Goodess, C.M. 1997. The simulation of daily temperature series from GCM output. Part II: Sensitivity analysis of an empirical transfer function methodology. *Journal of Climate*, **10**, 2514-2532.

Wood, A.W., Leung, L.R., Sridhar, V. and Lettenmaier, D.P. 2004. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change* **62**, 189-216.

Zorita, E. and von Storch, H. 1999. The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. *Journal of Climate* **12**, 2474-2489.

Miscellaneous

Kysely, J. 2002. Probability estimates of extreme temperature events: stochastic modelling approach vs extreme value distributions. *Stud. Geophys. Geod*. **46**, 93 – 112.

Shaw, E. 1994. Hydrology in Practice (3rd Edition), Chapman & Hall, London.

APPENDIX 1: ENHANCEMENTS SINCE SDSM VERSION 3.1

SDSM 4.2 includes a number of enhancements to version 3.1, sponsored by the Environment Agency of England and Wales.

Frequency analysis for extremes

- Allows the User to fit distributions to observed and downscaled data (as either a whole data set or by isolating particular seasons or months): GEVs, stretched exponential, Empirical and Gumbel distributions. Results can be viewed in either tabular form or as line charts.
- User can also plot PDFs of observed and modelled data and Quantile-Quantile plots (settings allow all charts to be changed).
- The user can save these analysed results to a text file and a threshold can be applied.
- A line plot can be made – allowing the user to compare observed data with ensembles (either as means, all ensembles or individual ensembles).

Step-wise regression

- Examines all possible combinations of predictors. Analyses models using either AIC or BIC criteria which user can select in advanced settings.

Optimisation Algorithm

- In addition to the dual simplex algorithm of SDSM 3.1, an ordinary least squares algorithm has been implemented. This is much quicker and efficient. It can be selected in Advanced Settings.

Screen Variables

- The User can now apply an autoregression component alongside other predictors.

Calibrate Model

- An autoregressive term can now be included in the model.
- Residual analysis has been added so that following calibration SDSM allows the user to plot residuals of the model either as a scatter diagram or a histogram (both of which can be amended through additional settings).
- The Chow test has been added so the user can also now assess the calibrated model for stationarity.

Weather Generator

- Additional information is captured within the *.PAR file (i.e., SDSM version, auto regression and process)

Scenario Generator

- Additional information is provided on the model before generation begins.

Summary Statistics (replaces Analyse Data screen)

- A raft of new statistics have been added; Extreme Range, Minimum Range, Maximum N-day Total, Mean Wet-Day Persistence, Mean Dry-Day Persistence, Correlation for Spell Lengths, Median Wet-Spell Length, Median Dry-Spell Length

Time Series Analysis

- Includes a raft of additional STARDEX indices for analysis: Mean dry spell, Mean wet spell, Median dry spell, Median wet spell, SD dry spell, SD wet spell, Spell length correlation, Dry day persistence, Wet day persistence, Maximum dry spell, Maximum wet spell, Nth largest value, Largest n day total, Percentage of precipitation above annual percentile, Percentage of all precipitation from events greater than long-term percentile, Number of events greater than long term percentile (the User can enter their own thresholds and percentile values).

Miscellaneous improvements

- Default file directory established in settings to ensure that every screen searches in the same directory for files each time.
- Improved interface so that it is now easier to move between stages of the process, with bigger screens, and improved colour schemes.
- Soft reset when error occurs so that User settings are not reset if a problem occurs.
- Splash screen changed (can now be removed).
- Advanced Settings enables fixed or stochastic threshold for conditional processes.
- Error trapping and efficiency improved throughout.
- Help files and User manual updated accordingly.

APPENDIX 2: FREQUENTLY ASKED QUESTIONS

The following generic and specific questions are arranged in the order in which they might typically be encountered during a downscaling procedure.

Q. Do I need to perform any re–gridding or normalisation of the predictor variables?

No. These tasks have already been performed for the UKSDSM data set released with the software and available to non-profit organisations on request. All UK data have been re–gridded to a standard co–ordinate system (2.5º latitude × 3.75º longitude), and normalised with respect to the 1961–1990 climatology. The User must simply select the nearest grid box(es) to the site in question. For all other regions (including the UK), gridded predictor variables are available online courtesy of the Canadian Climate Impacts Scenarios Group. The web-site is accessed from: http://www.cics.uvic.ca/scenarios/index.cgi?Scenarios

Q. Can I use observational data that lie outside the standard period 1961 to 2000?

No. Observed predictor variables for SDSM archives are obtained from NCEP and normalised only for the period 1961 to 2000. Station meteorological data prior to 1st January 1961 or after 31st December 2000 will have no pre-prepared predictor variables. The software also assumes that meteorological data provided by the User commences on 1st January 1961 (i.e., has the same start date as the predictors). If this is not the case, the User should pad the station data with the **Missing Data Identifier**.

Q. How important is the selection of predictor variables?

Identifying sensible predictor–predictand relationships is the most critical procedure in all statistical downscaling methods. The **Screen Variables** screen is designed to assist the User in the choice of appropriate downscaling predictor variables for model calibration via seasonal correlation analysis, partial correlation analysis, and scatterplots. Ultimately, however, the User must decide whether or not the identified relationships are physically sensible for the site(s) in question.

Q. How can I determine if I have chosen the correct predictor variables for the predictands that I require?

The correlation statistics and P values indicate the strength of the *association* between two variables. Higher correlation values imply a higher degree of association. Smaller P values indicates that this association is less likely to have occurred by chance. A P value <0.05 is routinely used as the cut-off, so a P value of 0.37 would indicate that the predictor–predictand correlation is likely to be due to chance. However, even if P <0.05 the result can be statistically significant but *not* be of practical significance — there's a difference!

Even if a high correlation and low P value is returned, the **Scatterplot** indicates whether this result is due to a few outliers, or is a potentially useful downscaling relationship. The Scatterplot may also reveal that one (or both) of the variables should by modified using the **Transform** operation, to linearise the relationship.

Q. What does the Event Threshold parameter (in Settings) do?

The **Event Threshold** parameter specifies the boundary between the two states in a **Conditional** process model. For example, if the Conditional process is precipitation, changing the Event Threshold from 0 to 0.3 will result in more "dry" days and fewer "wet" days (a simple way of classifying "trace" rainfall days as dry days). Therefore, different values for the Event Theshold will yield different results in **Screen Variables** (correlation values and scatterplots are both affected), will produce different parameters in **Calibrate Model**, and different results from the two **Analyse Data** operations. Note, however, that the **Weather Generator** and **Scenario Generator** operations will still produce values in the range 0 to 0.3 even if the threshold is set at 0.3.

Q. What are the advantages and disadvantages of using the monthly, seasonal or annual Model Type in Calibrate Model?

The **Model Type** button in **Calibrate Model** determines whether individual downscaling models will be calibrated for each calendar month, climatological season or entire year. The monthly button should be selected whenever the predictand is known to have a strong seasonal cycle, noting that even the annual button can produce the same result *provided* that one or more predictor variables have strong seaonality. Annual models are more parsimonious in the sense that they have only one set of regression weights instead of twelve in the case of the monthly models. Seasonal models might be used in situations where data are too sparse at the monthly level for model calibration, for example, a low incidence of precipitation in semi-arid regions.

Q. I am trying to model precipitation and have chosen the fourth root transformation in Advanced Settings. What else must I do?

Nothing! The fourth root button in **Advanced Settings** tells the software that this transformation is to be used throughout (including calibration, weather generation and scenario generation). If checked, there's no need to apply any further transformations as this is all backed out automatically. So when calibrating the model with fourth root checked, you should supply the model with untransformed rainfall data, making sure that the **Conditional** process button is checked in the **Calibrate Model** screen.

Q. Is it OK to model precipitation as an unconditional process?

As a general rule, precipitation should be modelled as a **Conditional** process. It does not make much sense to neglect the occurrence process (i.e., sequences of wet or dry days are first modelled, then the amounts if it is a wet day). If you are being swayed by higher R-sq values of an unconditional model during calibration, beware, the result is probably seriously biased by the large number of zero values entered in the multiple regression. Remember, daily precipitation amount is the most problematic daily variable to downscale.

Q. When I use the Weather Generator I get unrealistically large maximum daily (precipitation) values. What's going wrong?

Unrealistically large values generally imply that the variance inflation and/or bias correction in **Advanced Settings** are too high.

Q. Why do I get slightly different results every time I run the Weather Generator (with the same inputs)?

Even with the same inputs (i.e., *.PAR file, **Settings** and data period) the **Weather Generator** (and **Scenario Generator**) operation is not expected to produce identical results if the **Random Number Seed** is checked in **Settings**. This is because of the stochastic (random) component that is applied to each downscaled series to compensate for the fact that the deterministic component of the model (due to the chosen predictor variables) does not explain all of the observed variance. Differences between individual runs and/or **Ensemble Members** is likely to be greater for poorly determined predictands such as precipitation than in better constrained predictands such as temperature.

Q. Does SDSM produce realistic results for multiple sites? Also, what if I'm interested in preserving relationships between variables?

Both of these questions are the subject of ongoing research. However, results from previous studies suggest that regression–based downscaling does preserve some of the observed inter–site correlations *provided* that models calibrated on a site by site basis are forced by a common set of predictors. In other words, inter–site correlations are implicitly reproduced by virtue of correlated predictor variables, rather than by the model structure. Alternatively, inter-site behaviour may be reproduced by employing a conditional resampling approach in which case SDSM is used to downscale a predictand at a benchmark site. This series is, in turn, used to resample observations at dependant locations using events occurring on the same date (see Wilby et al., 2003 for more details).

Preliminary tests of inter–variable correlations produced by SDSM (e.g., between downscaled precipitation and temperature series) indicate that inter–annual variations in the strength of relationships are preserved, but there can be differences between the model and observations in individual months. Once again, it is suspected that inter–variable relationships are implicitly preserved by virtue of commonality in the predictor variables used to downscale each predictand.

However, if required, it is relatively straightforward to explicitly condition one predictand on another (e.g., daily precipitation occurrence might be used to condition maximum temperatures). In this case, the conditioning variable (precipitation occurrence) would be entered as a predictor during model calibration.

Q. I've calibrated my model. How do I now produce values of PRCP, TMAX or TMIN using GCM data?

Provided you have produced a *.PAR file via **Calibrate Model**, the software will automatically know what predictors are needed. Of course you may need to transform some of the GCM files if this was done for calibration. For example, if Z.DAT was transformed to ZSQUARED.DAT and then used to train the model, the same transformation should be applied to the equivalent GCM file (i.e., Z.GCM to ZAQUARED.GCM). In which case, be sure to maintain the same nomenclature of the file but with the *.GCM extension.

Q. Why do I keep getting an error message when I use GCM data?

The most likely explanation is that the **Year Length** in **Settings** has not been set correctly with respect to the number of days in the GCM simulation. For example, HadCM2 and HadCM3 have year lengths of 360 days, whereas CGCM1 has 365 days in every year (i.e., no leap years). Version 4.2 prompts the User to double-check the number of days before proceeding.

Q. What's the best way of handling SDSM files outside the software?

All SDSM output files are written in ASCII format and, therefore, accessible by any word processor. Model results (*.OUT files) are tab–delimited if the number of **Ensemble Members** is greater than one, and, as such, can be imported into commercial spreadsheets for further analysis or graphing.

Q. I've looked at the predictor variable files and the values only range between +/-5. Is there something wrong with the data?

No. All predictor variables (NCEP and GCM) are normalised using their respective 1961–1990 means and standared deviations. The result is that each predictor variable is dimensionless, and will typically vary between –5 and +5.

GLOSSARY

Where appropriate, the following definitions were drawn from the Glossary of terms in the Summary for Policymakers, A Report of Working Group I of the Intergovernmental Panel on Climate Change, and the Technical Summary of the Working Group I Report.

Terms in *italics* are found elsewhere in this Glossary.

Aerosols Airborne solid or liquid particles, with a typical size between 0.01 and 10μm that reside in the atmosphere for at least several hours. Aerosols influence the *climate* directly through scattering and absorbing radiation, and indirectly through the formation and optical properties of clouds.

Akaike's Information Criterion (AIC) A measure used to distinguish between two competing statistical models that takes into account the goodness-of-fit of the model, whilst penalising models with larger numbers of parameters. See *BIC*.

Airflow (index) Trigonometric measures of atmospheric circulation obtained from surface pressure or geopotential height fields. Commonly derived indices include *vorticity*, *zonal flow*, *meridional flow*, and *divergence*. Certain indices have been used to replicate subjective classifications of daily *weather patterns*, or as predictor variables in statistical *downscaling* schemes.

Anthropogenic Resulting from, or produced by, human beings.

Atmosphere The gaseous envelope surrounding the Earth, comprising almost entirely of nitrogen (78.1%) and oxygen (20.9%), together with several trace gases, such as argon (0.93%) and *greenhouse gases* such as carbon dioxide (0.03%).

Autocorrelation A measure of the linear association between two separate values of the same random variable. The values may be separated in either space or time. For time series, the autocorrelation measures the strength of association between events separated by a fixed interval or lag. The autocorrelation coefficient varies between –1 and +1, with unrelated instances having a value of zero. For example, temperatures on successive days tend to be positively autocorrelated.

Bayesian Information Criterion (BIC) A measure used to distinguish between two competing statistical models that takes into account the goodness-of-fit of the model, whilst penalising models with larger numbers of parameters. The BIC also depends on the number of data points and tends to favour simpler models compared with the *AIC*.

Black box Describes a system or model for which the inputs and outputs are known, but intermediate processes are either unknown or unprescribed. See *regression*.

Climate The "average weather" described in terms of the mean and variability of relevant quantities over a period of time ranging from months to thousands or millions of years. The classical period is 30 years, as defined by the World Meteorological Organisation (WMO).

Climate change Statistically significant variation in either the mean state of the *climate*, or in its variability, persisting for an extended period (typically decades or longer). Climate change may be due to natural internal processes or to *external forcings*, or to persistent *anthropogenic* changes in the composition of the atmosphere or in land use.

Climate model A numerical representation of the climate system based on the physical, chemical and biological properties of its components, their interactions and feedback processes, and accounting for all or some its known properties.

Climate prediction An attempt to produce a most likely description or estimate of the actual evolution of the climate in the future, e.g. at seasonal, inter–annual or long–term time scales.

Climate projection A projection of the response of the climate system to emission or concentration scenarios of *greenhouse gases* and *aerosols*, or *radiative forcing* scenarios, often based on simulations by *climate models*. As such climate projections are based on assumptions concerning future socio–economic and technological developments.

Climate scenario A plausible and often simplified representation of the future climate, based on an internally consistent set of climatological relationships, that has been constructed for explicit use in investigating the potential consequences of anthropogenic *climate change*.

Climate variability Variations in the mean state and other statistics (such as standard deviations, the occurrence of extremes, etc.) of the climate on all temporal and spatial scales beyond that of individual weather events.

Conditional process A mechanism in which an intermediate state variable governs the relationship between regional forcing and local weather. For example, local precipitation amounts are conditional on wet–day occurrence (the state variable), which in turn depends on regional–scale predictors such as atmospheric humidity and pressure.

Deterministic A process, physical law or model that returns the same predictable outcome from repeat experiments when presented with the same initial and boundary conditions, in contrast to *stochastic* processes.

Domain A fixed region of the Earth's surface and over-lying atmosphere represented by a *Regional Climate Model*. Also, denotes the grid box(es) used for statistical *downscaling*. In both cases, the downscaling is accomplished using pressure, wind, temperature or vapour information supplied by a host GCM.

Divergence If a constant volume of fluid has its horizontal dimensions increased it experiences divergence and, by conservation of mass, its vertical dimension must decrease.

Downscaling The development of climate data for a point or small area from regional climate information. The regional climate data may originate either from a *climate model* or from observations. Downscaling models may relate processes operating across different time and/or space scales.

Dynamical See *Regional Climate Model*.

Emission scenario A plausible representation of the future development of emissions of substances that are potentially radiatively active (e.g. *greenhouse gases*, *aerosols*), based on a coherent and internally consistent set of assumptions about driving forces and their key relationships.

Ensemble (member) A set of simulations (members) in which a deterministic *climate model* is run for multiple *climate projections*, each with minor differences in the initial or boundary conditions. Conversely, *weather generator* ensemble members differ by virtue of random outcomes of successive model simulations. In either case, ensemble solutions can be grouped and then compared with the ensemble mean to provide a guide to the *uncertainty* associated with specific aspects of the simulation.

External forcing A set of factors that influence the evolution of the climate system in time (and excluding natural internal dynamics of the system). Examples of external forcing include volcanic eruptions, solar variations and human–induced forcings such as changing the composition of the atmosphere and land use change.

Extreme weather event An event that is rare within its statistical reference distribution at a particular place. Definitions of "rare" vary from place to place (and from time to time), but an extreme event would normally be as rare or rarer than the 10th or 90th percentile.

General Circulation Model (GCM) A three–dimensional representation of the Earth's atmosphere using four primary equations describing the flow of energy (first law of thermodynamics) and momentum (Newton's second law of motion), along with the conservation of mass (continuity equation) and water vapour (ideal gas law). Each equation is solved at discrete points on the Earth's surface at fixed time intervals (typically 10–30 minutes), for several layers in the atmosphere defined by a regular *grid* (of about 200km resolution). Couple ocean–atmosphere general circulation models (O/AGCMs) also include ocean, land–surface and sea–ice components. See *climate model*.

Geopotential height The work done when raising a body of unit mass against gravity (i.e., acceleration due to gravity at a given level in the atmosphere multiplied by distance) divided by the value of gravity at the Earth's surface.

Greenhouse gas Gaseous constituents of the atmosphere, both natural and anthropogenic, that absorb and emit radiation at specific wavelengths within the spectrum of infrared radiation emitted by the Earth's surface, the atmosphere and clouds. The primary greenhouse gases are water vapour (H_2O), carbon dioxide (CO_2), nitrous oxide (N_2O), methane (CH_4), and ozone (O_3).

Grid The co–ordinate system employed by *GCM* or *RCM* to compute three–dimensional fields of atmospheric mass, energy flux, momentum and water vapour. The grid spacing determines the smallest features that can be realistically resolved by the model. Typical resolutions for GCMs are 200km, and for RCMs 20–50km.

Meridional flow An atmospheric circulation in which the dominant flow of air is from north to south, or from south to north, across the parallels of latitude, in contrast to *zonal flow*.

NCEP The acronym for the National Center for Environmental Prediction. The source of re–analysis (climate model assimilated) data widely used for dynamical and statistical *downscaling* of the present climate.

Normalisation A statistical procedure involving the standardisation of a data set (by subtraction of the mean and division by the standard deviation) with respect to a predefined control period. The technique is widely used in statistical *downscaling* to reduce systematic biases in the mean and variance of climate model output.

Parameter A numerical value representing a process or attribute in a model. Some parameters are readily measurable climate properties; others are known to vary but are not specifically related to measurable features. Parameters are also used in climate models to represent processes that poorly understood or resolved.

Partial Duration Series Events above a defined threshold that are recorded as a time series or as a frequency distribution. Essentially a 'peaks over threshold' approach to describing the occurrence of extreme events.

Predictand A variable that may be inferred through knowledge of the behaviour of one or more *predictor* variables.

Predictor A variable that is assumed to have predictive skill for another variable of interest, the *predictand*. For example, day–to–day variations in atmospheric pressure may be a useful predictor of daily rainfall occurrence.

Probability Density Function (PDF) A distribution describing the probability of an outcome for a given value for a variable. For example, the PDF of daily temperatures often approximates a normal distribution about the mean, with small probabilities for very high or low temperatures.

Radiative forcing The change in net vertical irradiance (expressed as Watts per square metre) at the *tropopause* due to an internal change or a change in the *external forcing* of the climate system, such as, for example, a change in the concentration of carbon dioxide, or the output of the Sun.

Random See *stochastic*.

Re–gridding A statistical technique used to project one co–ordinate system onto another, and typically involving the interpolation of climate variables. A necessary pre–requisite to most statistical *downscaling*, because observed and climate model data are seldom archived using the same *grid* system.

Regional Climate Model (RCM) A three–dimensional, mathematical model that simulates regional scale climate features (of 20–50 km resolution) given time–varying, atmospheric properties modelled by a General Circulation Model. The RCM *domain* is typically "nested" within the three–dimensional *grid* used by a GCM to simulate large–scale fields (e.g. surface pressure, wind, temperature and vapour).

Regression A statistical technique for constructing empirical relationships between a dependent (*predictand*) and set of independent (*predictor*) variables. See also *black box, transfer function*.

Relative humidity A relative measure of the amount of moisture in the air to the amount needed to saturate the air at the same temperature expressed as a percentage.

Resolution The *grid* separation of a climate model determining the smallest physical feature that can be realistically simulated.

Scenario A plausible and often simplified description of how the future may develop based on a coherent and internally consistent set of assumptions about driving forces and key relationships. Scenarios may be derived from projections, but are often based on additional information from other sources, sometimes combined with a "narrative story–line".

Specific humidity The ratio of the mass of water vapour (in grams) to the mass of moist air (in kilograms) in a given volume of air.

Station The individual site at which meteorological measurements are systematically observed and recorded.

Stochastic A process or model that returns different outcomes from repeat experiments even when presented with the same initial and boundary conditions, in contrast to *deterministic* processes. See *weather generator*.

Transfer function A mathematical equation that relates a *predictor*, or set of predictor variables, to a target variable, the *predictand*. The predictor(s) and predictand represent processes operating at different temporal and/or spatial scales. In this case, the transfer function provides a means of *downscaling* information from coarse to finer resolutions.

Tropopause The boundary between the lowest part of the atmosphere, known as the troposphere, and the highly stratified region of the atmosphere, known as the stratosphere. The tropopause is typically located 10km above the Earth's surface.

Uncertainty An expression of the degree to which a value (e.g. the future state of the climate system) is unknown. Uncertainty can result from a lack of information or from disagreement about what is known or knowable. It can also arise from poorly resolved climate model parameters or boundary conditions.

Unconditional process A mechanism involving direct physical or statistical link(s) between a set of *predictors* and the *predictand*. For example, local wind speeds may be a function of regional airflow strength and *vorticity*.

Vorticity Twice the angular velocity of a fluid particle about a local axis through the particle. In other words, a measure of rotation of an air mass.

Weather generator A model whose stochastic (random) behaviour statistically resembles daily weather data at single or multiple sites. Unlike *deterministic* weather forecasting models, weather generators are not expected to duplicate a particular weather sequence at a given time in either the past or the future. Most weather generators assume a link between the precipitation process and secondary weather variables such as temperature, solar radiation and humidity.

Weather pattern An objectively or subjectively classified distribution of surface (and/or upper atmosphere) meteorological variables, typically daily mean sea level pressure. Each atmospheric circulation pattern should have distinctive meteorological properties (e.g. chance of rainfall, sunshine hours, wind direction, air quality, etc). Examples of subjective circulation typing schemes include the European Grosswetterlagen, and the British Isles Lamb Weather Types.

Zonal flow An atmospheric circulation in which the dominant flow of air follows the lines of latitude (e.g. the westerlies), in contrast to *meridional flow*.