18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

CLIMA: a weather generator framework

Donatelli, M.[1,3], Bellocchi, G.[1], Habyarimana, E.[1], Bregaglio, S.[2], Confalonieri, R.[2,3], Baruth, B.[3]

[1] *CRA - Research Center for Industrial Crops, Bologna, Italy E-Mail: marcello.donatelli@agsys.cra-cin.it*
[2] *University of Milan, Department of Crop Science, Milan, Italy*
[3] *European Commission - Joint Research Center, Ispra (VA), Italy*

Abstract: Weather generators (WG) can be defined as collections of models to estimate site specific weather data and derived variables. They are commonly used for providing inputs to a variety of biophysical models or for deriving weather indices. Also, using either global circulation models or local area models inputs, sets of parameters calculated from long term weather series specific to a site can be modified to reproduce via WG synthetic series representing climate change scenarios. Finally, models implemented in WG are used for estimating missing data and to perform quality control on data collected from sensors at weather stations.

The scientific foundation of models implemented in WG varies from purely empirical to physically based. Several models exist allowing either the estimate or the generation of specific weather variables, with different input requirements. New models are continuously being proposed, and, although some models to estimate specific variables are commonly accepted as reference methods, the lack of some inputs requires at times using alternate approaches. Currently available WG are applications which implement a predefined set of modelling options, in software implementations which do not allow for independent extensions by third parties.

The CLIMA weather generator is a component based application which consist of a set of reusable graphical user interface (GUI) components, and a set of extensible model components. The latter are subdivided into six namespaces to estimate variables related to air temperature, rainfall, solar radiation, evapotranspiration, wind, and leaf wetness. The temporal resolution of the estimated variables varies from a day to ten minutes. Another software library allows the estimation of climatic indices from one year of daily data at the time. The current implementation consists of more than 300 models.

Components are usable either via the CLIMA GUI, or via custom developed applications in a client-server architecture. The architecture of components is based on the composite and strategy as keystone design patterns. Models are implemented as single approaches (simple strategies), and as composite models (composite strategies) which are associated to models of finer granularity. Another type of model unit is represented by context strategies, which implement a logic to select within associated models. Finally, the GUI allows building composite models which can be saved as libraries, to be reused both within CLIMA for weather series generation, or independently by other applications.

The components are implemented as .NET libraries. They implement a test of pre- and post-conditions, and a scalable tracing via .NET listeners. All variables and parameters are documented by a description, a physical unit and default, maximum and minimum values. Components are extensible: i.e. new models can be added independently by third parties and detected by the CLIMA application, which can also use them for data generation via building new composite libraries. Each component is made available via a software development kit which includes the code of two sample projects, either to extend or to reuse the component. CLIMA and its model components are freely available for reuse in no-profit applications.

Keywords: *weather generation; component-oriented programming; re-usability; extensibility*

1. INTRODUCTION

Investigating the potential impact of climate on agro-ecosystems using simulation models is underpinned by the availability of climate data at the appropriate temporal scale (daily or higher temporal resolution). The production of artificial series of weather data can be achieved by a variety of alternative methods. These range from empirical functions where simple relationships between weather variables are used to estimate missing data from available data (e.g. global solar radiation from air temperature), to more sophisticated approaches using physically-based models (e.g. combined equations with radiative and aerodynamic terms to estimate crop evapotranspiration). Artificial weather series can be generated using the approaches mentioned above, preconditioned that sufficient measured data are available to extract statistical parameters representative of actual weather, and a generation method, preserving such statistical properties in the generated series, is available. Methods of various complexity have been developed that proved to be capable to provide good approximations of the variability of daily weather patterns. Such models are well suited to ecological and environmental applications, striking a balance between the complexity of the climate system and the goodness-of-fit between observed and simulated data (Hutchinson, 1987). Both estimation and generation methods have been implemented in weather generators and other software tools (e.g., WGEN, Richardson and Wright, 1984; Cligen, Nicks and Gander, 1994; USCLIMATE, Johnson et al., 1996; ClimGen, Stöckle et al., 2001; Climak, Danuso, 2002; RadEst, Donatelli et al., 2003). All such approaches illustrate from different perspectives that there is actually a wealth of well developed solutions to the basic problem of either estimating or generating weather data. Such modelling solutions are implemented as software units in a variety of ways. The weather generation and estimation problem and the frequent need to evaluate alternative approaches in a comparative fashion has suggested that such methods could usefully be made available as one comprehensive set of linkable components, targeting also the possibility of independent extension by third parties.

The advent of component-based programming has enabled the development of scalable, robust, large-scale applications in a variety of domains, including agro-ecological modelling (Argent, 2004). The concept of developing modular systems for biophysical simulation has led to the development of several modelling frameworks (e.g. Simile, ModCom, IMA, TIME, OpenMI, SME, OMS, as listed in Argent and Rizzoli, 2004) which make use of models as software components that can be linked together and composed according to the modelling purpose. However, little attention has been paid to the development of components which intrinsically promote re-usability, interchangeability, and extensibility of models (Donatelli et al., 2005). Components with such characteristics would have different use via custom developed clients and in different contexts, from end-user applications to web services.

The objectives of this paper are: 1) to describe the models implemented in the current development of the *CLIMA* framework, 2) to present the software design chosen to achieve reusability, interchangeability and extensibility of the components developed.

2. WEATHER MODELLING AT EARTH SURFACE LEVEL

Weather variables can be estimated via deterministic models and/or generated using stochastic-based approaches. Such models are made available in the *CLIMA* framework via implementation as separate software components. These model components are: *AirTemperature* (Donatelli et al., 2009b): air temperature generation; *EvapoTranspiration* (Donatelli et al., 2006b): reference evapotranspiration estimation; *SolarRadiation* (Donatelli et al., 2006): solar radiation estimation/generation; *Rain* (Carlini et al., 2006): precipitation generation; *Wind* (Donatelli et al., 2009): wind speed generation; *LeafWetness*: leaf wetness estimation. These key weather variables are modelled on either daily or sub-daily time steps. Most of the weather models implemented have been extracted from peer-reviewed sources, which provide references for model evaluation. The relevant information was collected, interlinked and uniformly formatted into navigable structures to grant easy access to readers. Basic literature on models is reported without full citation, which can be retrieved from the hypertext help files at the URL specified in the relevant paragraph for each component.

2.1. Air temperature

Air temperature modelling includes modelling approaches from the following sources:

Allen et al. (1998)
Bekele et al. (2007)
Campbell (1985)
Kenneth et al. (2003)
Kimball et al. (1997)
Linacre (1977)

Danuso (2002)
Dumortier (2002)
Ephrath et al. (1996)
Goudriaan and Van Laar (1994)
Gracia et al. (2003)
Hubbard et al. (2003)
Iribarne & Godson (1981)
Linacre (1992)
Meteotest (2003)
Porter et al. (2000)
Remund and Page (2002)
Richardson (1981)
Stöckle (2002)

The generation of daily maximum (T_{max}, °C) and minimum (T_{min}, °C) air temperatures is considered to be a continuous stochastic process with daily means and standard deviations, possibly conditioned by the precipitation status of the day (wet or dry). Three alternative methods are implemented for generating daily values of T_{max} and T_{min}. The multi-stage generation system is conditioned by the precipitation status for both the Richardson-type and Danuso-type approaches. Residuals for T_{max} and T_{min} are computed first, then daily values are generated independently (Richardson-type) or with dependence on T_{max} on T_{min} (Danuso-type). A third stage that adds an annual trend calculated from the Fourier series, is included in the Danuso-type generation. The Richardson-type approach accounts for the correlation of air temperature and global solar radiation. A third approach generates T_{max} and T_{min} independently in two stages (daily mean air temperature generation first, T_{max} and T_{min} next), making use of an auto-regressive process from mean air temperatures and solar radiation parameters. Daily values of T_{max} and T_{min} are used by six approaches to generate hourly air temperature values, according to alternative methods. Sinusoidal functions are largely used to represent the daily pattern of air temperature. Six approaches are used to generate hourly values from daily maximum and minimum temperatures. The Dumortier-type approach derives hourly air temperatures from the daily solar radiation profile. Mean daily values of dew point air temperature are estimated via empirical relationships with T_{max} and T_{min} and other variables. A diurnal pattern of dew point air temperature is also modelled via the methods Ephrath and Meteotest. http://agsys.cra-cin.it/tools/airtemperature/help/

2.2. Evapotranspiration

Evapotranspiration and related variables include modelling approaches from the following sources:

Allen et al. (1998)
ASAE Standards (1998)
Brunt (1939)
Díaz (1989)
Donatelli (2009)
German (2000)
Hargreaves & Samani (1985)
Harrison (1963)
Holtslag & Van Ulden (1983)
Jackson et al. (1988)
Kimberley (2008)
Penman (1948)
Pérez et al. (1994)
Smith et al. (1991)
Stanghellini (1987)
Steiner et al. (1991)
Stöckle et al. (1998)
Tetens (1930)
Thom & Oliver (1977)
Waichler & Wigmosta (2003)

Evapotranspiration for reference crops (ET0) is calculated from alternative inputs, conditions and time steps, using one-dimensional equations based on aerodynamic theory and energy balance. A standardized form of the FAO-56 Penman-Monteith equation is used to estimate daily or hourly ET0 for two reference surfaces (0.12-m height, cool-season extensive grass such as perennial fescue or ryegrass; 0.50-m height, similar to alfalfa). The Priestley-Taylor equation is useful for the calculation of daily ET0 for conditions where weather inputs for the aerodynamic term (relative humidity, wind speed) are unavailable. The aerodynamic term of the Penman-Monteith equation is replaced by a dimensionless empirical multiplier (alternative implementations). As an alternative when solar radiation data are missing, daily ET0 can be estimated using the Hargreaves & Samani equation. The Stanghellini model estimates hourly ET0 from a multi-layer canopy (well-developed tomato crop) grown in a single glass, Venlo-type greenhouse with hot-water pipe heating system. http://agsys.cra-cin.it/tools/evapotranspiration/help/

2.3. Precipitation

Rainfall modelling includes modelling approaches from the following sources:

Arnold & Williams (1989)
Connolly et al. (1998)
Geng et al. (1986)
Olsson (1998)
Selker & Haith (1990)
Srikanthan & Chiew (2003)

Mann et al. (1974)
Meteotest (2003)
Nicks et al. (1990)
Srikanthan & McMahon (1995)
Williams (1995)

The occurrence of wet or dry days is modelled as a stochastic process represented by a first-order Markov chain. The transition from one state (dry or wet) to the other (dry or wet) is governed by transition probabilities. According to multi-transition model , the daily precipitation amounts are divided into a number of up to seven states - dry, or wet - from 1 (lowest level of rainfall) through 6 (highest level of rainfall). On days when precipitation is determined to occur, the precipitation amount is generated by sampling from alternative probability distribution functions. Most approaches are based on the two-state transition for dry/wet days. The Gamma distribution is used to model precipitation amounts for the last state (highest level of rainfall) in the multi-state transition probability, while a linear distribution is applied for the other states. The pattern of Gamma plus linear distribution across various occurrence states, exhibits a combined J shaped function. Short-time rainfall data are generated by disaggregating daily rainfall into a number of events, then deriving the characteristics (amount, duration and start time) for each event. Four approaches have been implemented to dissect daily amounts into 6-hour or shorter resolution (as short as 10 minutes) amounts. The method described by Arnold and Williams allows half-hour time resolution and assumes that daily rainfall falls in only one event. The peak location is generated first according to a broken linear distribution. The other half-hourly amounts are generated from an exponential distribution and relocated on both sides of the peak. http://agsys.cra-cin.it/tools/rain/help/

2.4. Solar radiation

Solar radiation modelling include modelling approaches from the following sources:

Ångström (1924)
Bristow & Campbell (1984)
Chen et al. (1999)
Donatelli & Bellocchi (2001)
Donatelli & Bellocchi (1998)
Garcia y Garcia & Hoogenboom (2005)
Hargreaves & Samani (1982)
ESRI http://www.esri.com
Johnson et al. (1995)
Liu & Jordan (1960)
Prescott (1940)
Richardson (1981)
Ross (1975)
Stine & Harrigan (1985)
Supit & van der Goot (2003)
Supit & van Kappel (1998)
Thornton & Running (1999)
Winslow et al. (2001)
Woodward et al. (2001)

Solar radiation outside earth's atmosphere is calculated hourly based on routines derived from the solar geometry. Daily values are an integration of hourly values from sunrise to sunset. The upper bound for the transmission of global radiation through the earth's atmosphere (i.e., under conditions of cloudless sky), can be set to a site-specific constant or estimated daily by diverse methods. Broadband global solar radiation (about 0.3-3.0 μm wave length) striking daily horizontal earth's surfaces is estimated from alternative sets of weather inputs according to strategies based on either physical relationships or stochastic procedures. A sine-curve assumption is used to prescribe the hourly distribution of solar radiation from its daily value, assuming changes with solar elevation angle. The most simplified models relate diurnal temperature range to solar energy transmission through the earth's atmosphere. As clouds are one of the most important phenomena limiting solar radiation at the earth's surface, a cloud cover measure is incorporated in one of the models supplied to estimate transmissivity. The Winslow-type radiation model uses saturation vapour pressures at minimum and maximum air temperature as a measure of the atmospheric transmission of incident solar radiation.. The Ångström-Prescott model is the most common choice to estimate global solar radiation when sunshine measurements are available. As an alternative, an implementation of the Johnson-Woodward model is given. Stochastic generation is based on the dependency between daily maximum and minimum temperature, and solar radiation (Richardson-type). Such variables are reduced to time series of normally distributed residual with mean zero and variance of one. An autoregressive, weakly stationary multivariate process is used to generate the residual series. Daily values of global solar radiation are generated for dry and wet days as daily deviations above and below the monthly average value.. http://agsys.cra-cin.it/tools/solarradiation/help/

2.5. Wind speed

Wind speed modelling includes modelling approaches from the following sources:

Ephrath et al. (1996)
Goudriaan & Van Laar (1994)
Gregory et al. (1994)
Hoffmann (2002)
Meteotest (2003)
Mitchell et al. (2000)
Nicks et al. (1995)
Porter et al. (2000)
Swartzman & Kalzuny (1987)
Takle & Brown (1977)
Tatarko et al. (1997)

Daily mean wind speed is generated by sampling from several probability distribution functions. Following generation of daily mean wind speed, alternative approaches are available to estimate the maximum and minimum wind speeds for the day. Probability density functions and wave functions are used to distribute daily mean wind speed. http://agsys.cra-cin.it/tools/wind/help/

2.6. Leaf wetness

LeafWetness is a software library containing different models to estimate hourly values of leaf wetness according to various approaches. Leaf wetness duration is a driving variable in epidemiological models for simulating risk of yield losses due to plant diseases. It is highly related to the epidemiology of many important crops as it influences the germination of fungal pathogens, their penetration through the leaves, and the occurrence of primary and secondary infections during the same season. Its modelling is therefore essential in crop protection, particularly in forecasting the development rates of plant diseases. The *LeafWetness* component contains 5 models to estimate hourly values of leaf wetness according to alternative approaches, either empirical or mechanistic. These are:

SWEB (Surface Wetness Energy Balance)
LWR (Leaf Wetness Reference)
DP (Dew Parametrization)
CART (Classification And Regression Tree)
ET (Extended Threshold)
http://agsys.cra-cin.it/tools/leafwetness/help/

3. SOFTWARE DESIGN

CLIMA is an application made of a set of components written in C# of the Microsoft .NET, platform, most of which are deployable and re-usable outside the *CLIMA* framework. The *CLIMA* software (Fig. 1) is provided with a sample client endowed with a graphical user interface, including an application (*MCE*: Model Component Explorer) which makes it possible to discover inputs, parameters and outputs of alternative models, called strategies, in each of the six model components.

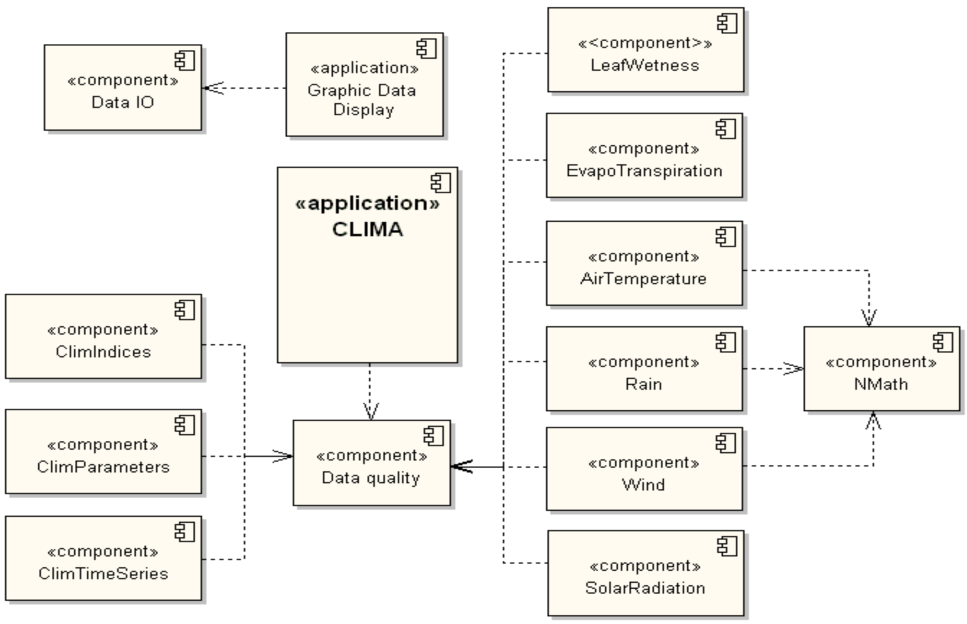

Figure 1. *CLIMA* component diagram. Main components are shown.

Besides the six model components described in the sections above, *CLIMA* also includes components to:

- perform data quality analysis (*DataQuality*)
- estimate climatic indices (*ClimIndices*)
- estimate parameters (*ClimParameters*)
- evaluate/estimate trend and outliers data in time series (*ClimTimeSeries* – under development)

Common features of model components are:

- full documentation as hypertext help files for modelling solutions, and of sample clients inclusive of source code
- possibility of extending models by adding new ones without recompiling the component and still using the same signature of the application programming interface call
- test of pre- and post-conditions
- unit tests for all public methods (data values reported in the documentation)

3.1. Model component architecture

All models can be called up explicitly with or without the test of pre- and post-conditions, following the design-by-contract approach (Meyer, 1997), which requires pre-conditions (in this case input values within a given range, and constraints conditions, e.g. $T_{max} > T_{min}$) to be respected in order to obtain results complying with post-conditions (in these implementations, output values within the given range). Although respecting pre-conditions is a responsibility of the client application, the components implement exception handling when input data violate pre-conditions. Implementing the test of pre- and post-conditions also forces developers of biophysical models to state the numerical conditions which need to be respected in order to use the model correctly. All logical components depend on a service component (Data Quality in Fig.1), which implements the test of pre- and post-conditions, and makes possible to obtain the output test on screen, file, and to .NET listeners. The Data quality component of Fig. 1 is also used for form validation into the *CLIMA* client.

Models are implemented in discrete units called strategies as simple, composite, and context strategies (Fig. 2). The Strategy and Composite design patterns are applied, hence all types of strategies implement the same interface. *CLIMA* (full package) and its components (independently packaged) are freely distributed at http://agsys.cra-cin.it/tools/. The installation includes compiled subcomponents; the installation of individual components includes the source code of sample applications showing how to extend and reuse components.

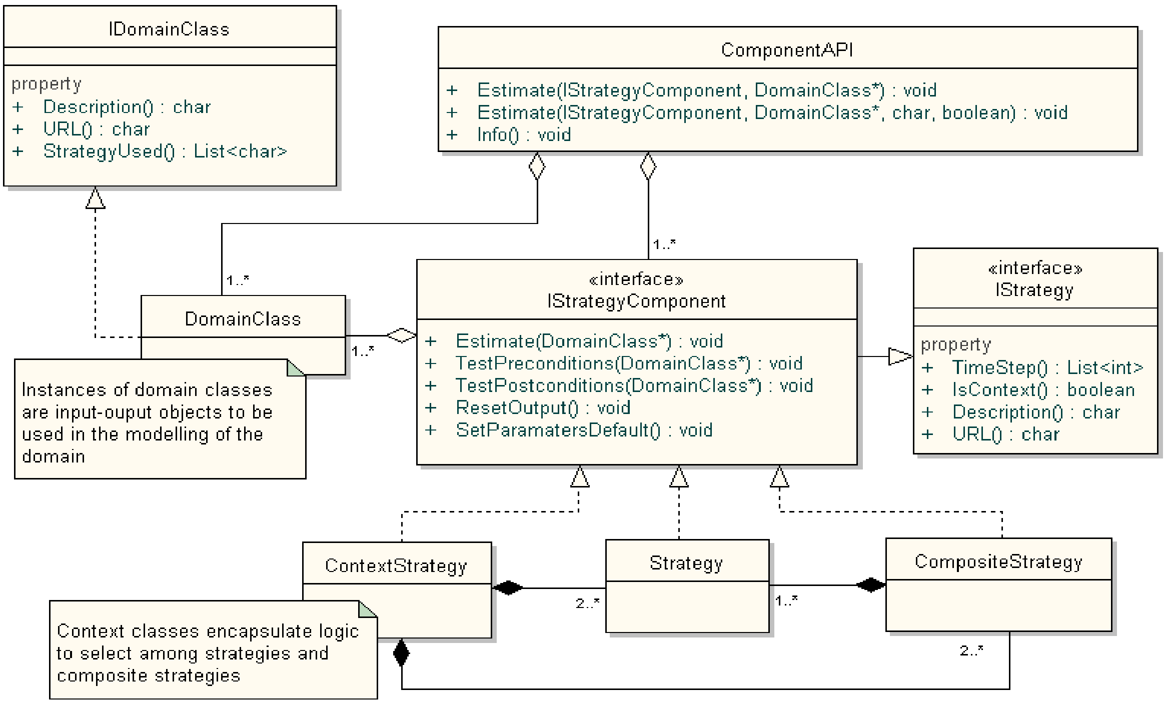

Figure 2. Class diagram of *CLIMA* model components. A third party can extend the components both by adding new models via realizations of the `IStrategyComponent` interface, and by extending the domain classes inheriting from the `DomainClass` data-type. Domain classes encapsulate attributes of variables.

3.2. Building composite strategies and executable configurations at run time

The *CLIMA* GUI allows building composite strategies using a wizard that allows using alternate modelling options. Such strategies are saved as discrete components (DLL), and are re-usable within *CLIMA* for data generation and independently by third parties. It also allows building configurations (selection of models and sequence of generation) to generate data. An executable file is built for use both within *CLIMA* for data generation and independently by third parties. The source code of the newly built either composite strategy or executable is also saved.

4. CONCLUSIONS

The goal of *CLIMA* development is to extend access to weather modelling features to multiple users, and to provide an architecture to ensure re-use and extension of coded models. *CLIMA* attempts to overcome some of the technical challenges that to date have limited the development of reusable weather generation and estimation capabilities. On account of their model documentation and the metadata associated with each variable, *CLIMA* model components are also a way to share modelling knowledge in an extensible way. New graphical metaphors are being developed, and a version based on data persistence on database via a generic data-layer is also available. Details on CLIMA are available at http://agsys.cra-cin.it/tools/clima/help/

ACKNOWLEDGMENTS

This publication was partly funded by the AgroScenari project granted by the Italian Ministry of Agriculture, Food and Forestry Policies.

REFERENCES

Argent, R.M., 2004. An overview of model integration for environmental applications - components, frameworks and semantics. Environmental Modelling and Software 19: 219 - 234.

Argent, R.M., Rizzoli, A.E., 2004. Development of multi-framework model components, iEMSs 2004 Conference, Osnabrück, Germany, 1, 365-370.

Carlini L., Bellocchi, G., Donatelli, M. 2006. Rain, a software component to generate synthetic precipitation data Agronomy Journal 98: 1312-1317.

Confalonieri R., Bellocchi G., Donatelli M. 2009. A software component to compute agro-meteorological indicators. Environmental Modelling & Software (in press) http://dx.doi.org/10.1016/j.envsoft.2008.11.007

Danuso, F., 2002. Climak: a stochastic model for weather data generation. Ital. J. Agron., 6:57-71.

Donatelli M., Bellocchi G., Carlini L., 2006 Sharing knowledge via software components: models on reference evapotranspiration European Journal of Agronomy 24: 186-192

Donatelli, M., Bellocchi, G., Carlini, L., 2006. A software component for estimating solar radiation. Environmental Modelling and Software, 21:411-416.

Donatelli, M., Bellocchi, G., Fontana, F., 2003. RadEst3.00: Software to estimate daily radiation data from commonly available meteorological variables. European Journal of Agronomy, 18: 363 - 367.

Donatelli M., Bellocchi G., Habyarimana E., Confalonieri R., Micale F. 2009. An extensible model library for generating wind speed data. Computers and Electronics in Agriculture (submitted).

Donatelli M., Bellocchi G., Habyarimana E., Bregaglio S., Baruth B. 2009b. AirTemperature, an extensible software library to generate air temperature data. Agronomy Journal (submitted).

Hutchinson, M.F., 1987. Methods for generation of weather sequences. In: Bunting A.H., Agricultural environments: characterisation, classification and mapping. CAB International, Wallingford, United Kingdom, 149 – 157.

Johnson, G.L., Hanson, C.L., Hardegree, S.P., and Ballard, E.B., 1996. Stochastic weather simulation: overview and analysis of two commonly used models. Journal of Applied Meteorology, 35: 1878 - 1896.

Meyer, B., 1997. Object-oriented software construction, 2nd edition. Prentice Hall, Upper Saddle River, NJ, USA.

Nicks, A.D., and Gander, G.A., 1994. Cligen: a weather generator for climate inputs to water resources and other models. 5th International Conference on Computer in Agriculture, Orlando, FL, USA, 3 - 94.

Richardson, C.W., and Wright, D.A., 1984. WGEN: a model for generating daily weather variables. U.S. Department of Agriculture, Agricultural Research Service, ARS-8, Washington, D.C, USA.

Rizzoli A.E., Donatelli M., Athanasiadis I.N., Villa F., Huber D. 2008. Semantic links in integrated frameworks. Mathematics and Computers in Simulation, 78: 412-423

Stöckle, C.O., Nelson, R.L., Donatelli, M., and Castellvì, F., 2001. ClimGen: a flexible weather generation program. 2nd ISMCS, Florence, Italy, 229-230.