Environmental Modelling & Software 72 (2015) 402–417

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

# Analysis and classification of data sets for calibration and validation of agro-ecosystem models☆

K.C. Kersebaum [a,*], K.J. Boote [b], J.S. Jorgenson [c], C. Nendel [a], M. Bindi [d], C. Frühauf [e], T. Gaiser [f], G. Hoogenboom [g], C. Kollas [a], J.E. Olesen [h], R.P. Rötter [i], F. Ruget [j], P.J. Thorburn [k], M. Trnka [l], M. Wegehenkel [a]

[a] Leibniz Centre for Agricultural Landscape Research, Institute of Landscape Systems Analysis, D-15374 Müncheberg, Germany
[b] University of Florida, Dept. of Agronomy, Gainesville, FL 32611-0500, USA
[c] University of Reading, School of Agriculture, Policy & Development, Reading, RG6 6AR, UK
[d] University of Florence, Department of Agri-food Production and Environmental Sciences, I-50144 Florence, Italy
[e] Deutscher Wetterdienst DWD, Centre for Agrometeorological Research, D-38116 Braunschweig, Germany
[f] University of Bonn, Institute of Crop Science and Resource Conservation, D-53115 Bonn, Germany
[g] Washington State University, AgWeatherNet, Prosser, WA 99350-8694, USA
[h] Aarhus University, Dept. of Agroecology, DK-8830 Tjele, Denmark
[i] Natural Resources Institute Finland (Luke), FI-50100 Mikkeli, Finland
[j] INRA, UMR1114 Environnement Méditerranéen et Modélisation des Agro-Hydrosystèmes, F-84914 Avignon Cedex 9, France
[k] CSIRO Ecosystem Sciences, Queensland 4102, Australia
[l] Mendel University Brno, Institute of Agriculture Systems and Bioclimatology, CZ-61300 Brno, Czech Republic

A R T I C L E I N F O

*Article history:*
Received 5 February 2014
Received in revised form
31 March 2015
Accepted 26 May 2015
Available online 13 June 2015

*Keywords:*
Field experiments
Data quality
Crop modelling
Data requirement
Minimum data
Software

A B S T R A C T

Experimental field data are used at different levels of complexity to calibrate, validate and improve agro-ecosystem models to enhance their reliability for regional impact assessment. A methodological framework and software are presented to evaluate and classify data sets into four classes regarding their suitability for different modelling purposes. Weighting of inputs and variables for testing was set from the aspect of crop modelling. The software allows users to adjust weights according to their specific requirements. Background information is given for the variables with respect to their relevance for modelling and possible uncertainties. Examples are given for data sets of the different classes. The framework helps to assemble high quality data bases, to select data from data bases according to modellers requirements and gives guidelines to experimentalists for experimental design and decide on the most effective measurements to improve the usefulness of their data for modelling, statistical analysis and data assimilation.



## Software availability

Name of the software: Dataset classification software
Developer: Jason Scott Jorgenson, Kurt-Christian Kersebaum, Chris Kollas
License: GPL v2
Year first available: 2014
Contact: ckersebaum@zalf.de
Hardware requirements: Intel/AMD PC, 4 GB RAM
Software requirements: Microsoft Windows operating system
Availability: ftp://tran.zalf.de/pub/out/lsa/kersebaum/DatasetRanker.zip

## 1. Introduction

Soil–crop–atmosphere interactions play a central role in the multiple functions of agro-ecosystems and rural landscapes such as

☆ Special Issue on Agricultural Systems Modeling & Software.

* Corresponding author. Leibniz Centre for Agricultural Landscape Research, Institute of Landscape Systems Analysis, Eberswalder Str. 84, D-15374 Müncheberg, Germany. Tel.: +49 33432 82 394; fax: +49 33432 82 334.

*E-mail addresses:* ckersebaum@zalf.de (K.C. Kersebaum), kjboote@ufl.edu (K.J. Boote), j.s.jorgenson@reading.ac.uk (J.S. Jorgenson), nendel@zalf.de (C. Nendel), marco.bindi@unifi.it (M. Bindi), Cathleen.Fruehauf@dwd.de (C. Frühauf), tgaiser@uni-bonn.de (T. Gaiser), gerrit.hoogenboom@wsu.edu (G. Hoogenboom), chris.kollas@zalf.de (C. Kollas), JorgenE.Olesen@agrsci.dk (J.E. Olesen), reimund.rotter@luke.fi (R.P. Rötter), francoise.ruget@avignon.inra.fr (F. Ruget), Peter.Thorburn@csiro.au (P.J. Thorburn), mirek_trnka@yahoo.com (M. Trnka), mwegehenkel@zalf.de (M. Wegehenkel).

http://dx.doi.org/10.1016/j.envsoft.2015.05.009
1364-8152/

food and energy production, carbon sequestration, soil properties, biodiversity or conservation of water resources. Emissions from agriculture are also seen as a threat to the global climate system, which makes agriculture one of the key handles for climate change mitigation. There is an increasing need to better understand these complex systems, and to develop and utilize reliable process-based models for scenario analyses as a basis for policy and management decisions. Agro-ecosystem models are increasingly applied beyond the point and field scales to support decision-making (van Ittersum et al., 2003; Jones et al., 2003; Brisson et al., 2003; Stöckle et al., 2003; Keating et al., 2003), assess the impact of climate change (Holzworth et al., 2015, position paper of thematic issue), and to derive adaptation and mitigation strategies for the sustainable use and management of land and other natural resources (Hammer et al., 2002; White et al., 2011). Integrated Assessment and Modelling as suggested by Parker et al. (2002) requires the integration of dispersed data sources in a consistent and spatially and temporarily complete data set to provide necessary model inputs for decision making (Janssen et al., 2009) and to transfer site-based knowledge to regions and continents. With increasing size of the area under investigation, input data tend to become more uncertain relative to the point data of experimental sites, which were the original basis of development for the majority of agro-ecosystem models. Hence, model uncertainty also increases with the area under investigation since data of relevant state variables for testing and evaluation are not commonly available.

Critical to the evaluation, improvement, and use of crop models is the availability of high quality data from field observations. There is a mismatch between the rising demand by users for tested models and research budgets for suitable experimental research and monitoring, which tend to be decreasing (Rötter et al., 2011).

Since field experimental data sets are usually not recorded for modelling purposes, their level of detail, quality of records, variables considered as well as their number of spatial and temporal replicates vary enormously (Nix, 1985; Groot and Verberne, 1991) Therefore, their suitability for modelling is often insufficient for different reasons. White et al. (2013) proposed a standard approach for describing and identifying variables of management, environmental conditions, soils, and crop measurements, all for the purpose of developing, testing, and applying crop simulation models. In general, datasets used for model calibration and validation consist of data describing a) the initial soil conditions, b) the crop-specific management and c) the seasonal weather conditions (Palosuo et al., 2011; Rötter et al., 2012; White et al., 2013). Additionally, data on phenology of the crop, yields and nutrient contents from intermediate harvests, intra-seasonal soil conditions and measurements of fluxes of energy, water and $CO_2$ may be provided.

The international community of agricultural system modellers, e.g., in the Agricultural Model Intercomparison and Improvement Project AgMIP (Rosenzweig et al., 2013) or the European MACSUR (**M**odelling European **A**griculture with **C**limate Change for Food **S**ec**UR**ity) project (Rötter et al., 2013) are currently building harmonized data bases for the purpose of model testing and improvement including the opportunity to create model-specific interfaces for various models (Porter et al., 2014). In order to find suitable experimental data for specific applications in the context of modelling out of the vast offer of available data sets, a transparent method of screening and pre-selection is demanded, which highlights specific positive and negative features of a data set with respect to the intended application. To evaluate and select data sets, Rosenzweig et al. (2013) proposed different classes of data for so-called "Sentinel Sites", which represent specific sites with experimental data suitable for different levels of model testing and improvement. However, specific for a transparent classification were not provided. A joint community effort lead to the development of a qualitative (Boote et al., 2015) and a quantitative framework (this publication) to evaluate the quality of field experimental data sets for crop modelling according to robust and accepted criteria.

The aim of this paper is to provide a quantitative classification framework by which the consistency and quality of agricultural datasets can be evaluated. Variables under consideration are weighted according to both their importance and their quality, and justified by literature describing variance and errors of the different state variables and measurement methods. The objective of such a classification framework of data evaluation and labelling is (i) to allow data base managers to pre-check the quality of data sets before integrating them into their data base, (ii) to support the creation and use of international publicly available benchmarked data sets for model evaluation, inter-comparison and improvement, (iii) to enable modellers to select appropriate data according to their requirements, (iv) to give guidance to experimentalists for designing their experiments with respect to aspects that go beyond their primary research question, allowing for a broader use of experimental data for systems analysis and modelling.

## 2. Definitions and terminology

*Parameterisation* means the estimation of fixed model parameter values (e.g., diffusion coefficient of a substance in water) for single processes under controlled conditions.

*Calibration* means the adjustment of values of model parameters outside the model code (e.g. thermal sums for phenological development in external parameter files) to fit their output to a set of measured state variables or fluxes (Penning de Vries and van Laar, 1982). According to Van Keulen (1976), the main purpose of calibration is to adapt weak or unknown parameters or relations. Parameter values should preferably have a real and measurable background and their values should be adjusted within a reasonable range. Calibrated parameter values are valid only for the model configuration that was used for the calibration. Introducing new processes or algorithms usually requires re-calibration of at least parts of the parameter set.

*Validation* (or *falsification* if model application is beyond its limits) is the examination whether a model derived from analyses of some systems is capable of describing other systems, or simply, the test of a calibrated model against an independent data set that has not been used for calibration (De Wit, 1982).

*State variables* represent the status of a specific dynamic system variable (e.g. soil water content) at a particular time and location or compartment. The variable can be expressed as a total amount or concentration in a pre-defined compartment (e.g. soil layer or crop organ).

*Fluxes* are defined as a transition of matter or energy across a defined compartment border. They can be observed cumulatively over a specific time period and compared to corresponding simulations.

## 3. Data requirements for model calibration and validation

Application of a model in a new geographic/climatic environment or for a new crop requires new parameterisation and eventually modifications of the model, e.g. by consideration of additional processes; otherwise parameter adjustment to fit observed data becomes a pure tuning or curve fitting exercise (De Wit, 1982). Such extension of a model requires suitable data to identify and parameterise processes and it sometimes requires a re-calibration of parameters of other processes or modules if processes interact strongly.