

Models, Datasets and Indicators

Models, datasets and indicators allow for standardised interpretation and efficient re-use of scientific insight and efforts. Policy and decision support based on models, datasets and indicators is embedded in software and webservice or incorporated in reports and policy briefs. Documentation on scientific considerations, operational efficiency and user interpretation remain hidden.

Perspectives on quality

We present a checklist for the quality of models, datasets and indicators from three perspectives: scientific (topics 1 - 4), development (topics 5 & 6) and end use (topic 7). These perspectives relate to scientific rigor, efficiency of operation and the interpretation of results.

Requirements and criteria

Within the 7 topics 22 requirements are defined. For each requirement an elaborate formulation of criteria to be met is available. By providing references to the documentation related to these criteria, the checklist provides a reading guide to the model, dataset or indicator.



WAGENINGEN
UNIVERSITY &
RESEARCH

A checklist for the Quality of Models, Datasets and Indicators to be used in policy and decision support

2020

Checklist

1. *The model/dataset/indicator description*
 - a. There is a general description
 - b. The conceptual and formal model are described
2. *The technical implementation*
 - a. The implementation is documented
 - b. The technical environment is described
 - c. The implementation is tested
3. *Parameters, variables, inputs and output*
 - a. The parameters and variables are documented
 - b. Calibration of parameters is described
 - c. Input and output is described
 - d. The provenance of input data is described
4. *Evaluation*
 - a. A sensitivity analysis is performed
 - b. An uncertainty analysis is performed
 - c. The model / dataset / indicator is validated
 - d. The use is monitored
 - e. There is a general assessment of quality
5. *Planned development*
 - a. There is a development plan
 - b. A version control system is in place
6. *The organisation*
 - a. The metadata is available
 - b. There is a management plan
 - c. Dependencies are discussed
 - d. External use is formalised
7. *User documentation*
 - a. Interpretation guidance is provided
 - b. There is a user manual

Further information

Detailed description of the requirements and information on the use of this checklist in the quality assessment of models, datasets and indicators at Wageningen Research can be found in the accompanying leaflet or by contacting Geerten.Hengeveld@wur.nl

QUALITY SYSTEM



INDICATORS



MODELS



DATASETS

A



WR Quality criteria for models and datasets. Categorised at 2 quality levels in 22 Requirements within 7 themes across 3 perspectives on quality.

For more information see <https://intranet.wur.nl/Project/WRModellingToolbox>, or contact Geerten.Hengeveld@wur.nl

Perspective: Science & Technology

ST.1 The model/dataset is described

1.1 There is a general description of the model/dataset

A *purpose * area of application * theoretical framework * paradigms*

The general description includes statements on the purpose/goal/aim in developing the model/dataset. It provides a delineation of the area of application both spatial (generic (e.g., field, country) or specific (e.g., Netherlands, Wageningen) and conceptual (e.g., under constant climate). With the theoretical framework the world as addressed within the model/dataset is framed according to some basic paradigms and core assumptions.

1.2 The conceptual and formal model are documented

A *explicitly documented * assumptions * simplifications * embedded in literature*

The theoretical framework is worked out into a conceptual model (most relevant components and their relationships/dependencies) and subsequently into a formal model (mathematical model, decision tree, database structure etc.). The assumptions and simplifications made at each step are presented along with their justification (e.g., a table with assumptions). The formal model is explicitly documented (e.g., a full set of formulae, decisions, measurements). This is embedded in scientific literature as illustrated by targeted references.

AA *motivated complexity * peer-reviewed scientific publication*

From the idea that the model should be as simple as possible, but not more simple than that, a reflection is included on the relationship between the complexity of the conceptual and formal model (e.g., number of variables included, statistical design, number of equations, precision reached) and the foreseen use (e.g., application, data availability, evaluation options, past experience). The model/dataset has been published in a peer reviewed scientific journal.

ST.2 The technical implementation of the model/dataset is documented

2.1 The implementation is documented

A *Basic structure * flow diagram*

The documentation of the implementation should support the tech-savvy reader in the interpretation of the computer code, scripts etc.. A flow diagram highlighting the main modules of the program, or a database scheme is provided. It is considered good practice to explicitly provide a link between the elements in the diagram and the text.

AA *Code commenting * motivated (modular) design * code review*

The structure of the program is motivated, consistent, and modular where relevant. Code is commented. Discussing choices for algorithms and model structure facilitates interpretation of the logic of the program or database. A code review (external to the development team) is performed to

get feedback on best-practices, consistencies and potential (minor) programming errors. This review is to be organised by the development team. The feedback from the code-review is used in refining the development plan.

2.2 The technical environment is documented

A *Language * IDE * settings * limitations*

The development environment is documented, both in general and specific terms. This includes the computer language, IDE, versions used and critical settings. Technical limitations due the implementation (e.g., numeric precision, memory or multi-processor use) are mentioned.

2.3 The model/dataset is tested

A *Tests documented * protocol * untested components named*

There is a clear list/table (protocol) of tests that can / are relevant / need to be performed on the model/dataset to ensure correct technical implementation of the model/dataset (e.g. unit tests, order of magnitude tests, checksums, algebraic or numerical recalculation of simplified or extreme cases). In the documentation the performance on these tests is noted. Deviations from expected test results are discussed. If components of the model/dataset are not tested (yet), these components are identified and a motivation is provided.

AA *Tests on schedule * periodic evaluation on completeness*

During model/dataset implementation tests are performed on a regular basis to ensure correct implementation. Periodically the protocol is evaluated for missing components of the model/dataset.

ST.3 The parameters, variables, inputs to and output of the model/dataset are described

3.1 The parameters and variables of the model/dataset are documented

A *Quantities * units * default values * default source * description*

A full list of parameters and variables is provided. For each of these the quantities and units of measurement are given and a description of the interpretation is provided. When available, default values are given with their source.

AA *Range * Precision*

The documentation of the parameters and variables is extended with the range of possible, or likely, values, uncertainty range for default values and the associated precision is provided (i.e., number of decimals). N.B. this does not refer to the floating point limit of the implementation.

3.2 Calibration of parameters is described

A *Procedure * results discussed*

Calibration is defined as the deviation of parameter values based on (partial) model output in order to reach a pre-defined or desired output value. Depending on the model, calibration can constitute formal (mathematical) calibration, expert judgement or 'tuning' of parameters based on literature review. Calibration is generally applied to estimate values for parameters for which no default values can be derived based on first principles or experiments. The (preferred or applied) procedure for calibration is described. Impact of calibration is discussed.

3.3 The input and output is described

A *Structure * format * quantities * units * precision * description * link variables & parameters * version echo*

The structure and format of the input and output are described, including quantities, units used and the precision of values. From the description the link between input/output and the variables and parameters in the model is clear (linked to formal model description and implementation). The output should include a reference (echo) to the version number of the model/dataset that generated this specific output.

AA *(Inter)national standards * input echo & timestamp*

The choice for input/output in a specific format, or using specific units or projections has implications for re-use and interoperability. Discuss which international standards are used and why (not). The output should include a reference (echo) to the input (query) and parameters settings used and time of execution, as to enable tracking the source of the specific output. Echoing is also important for extractions of data from (dynamic) databases.

3.4 The origin of input data is described

A *Data preparation pipeline * source * scripts tested*

Preparation of data into the format required for operation in the model, or inclusion into the dataset is described. This can include a clearly delineated data preparation pipeline from source to input. Scripts used should adhere to the same principles of technical documentation and testing described.

AA *Protocol for acquisition * periodically updated*

A protocol for acquisition of (raw) data is mostly appropriate for (dynamic) datasets or periodically performed standard calculations. Such a protocol includes source of data, measurement protocol, and contact persons for institutional sources. This protocol is periodically updated.

ST.4 The functioning of the model/dataset is evaluated

4.1 A sensitivity analysis is performed

A *Tailored to model/dataset type * documented * discussed*

The sensitivity of the model to variation in parameter values and initial conditions is analysed. For different types of models, different methods are appropriate (e.g., one at a time, monte carlo simulation, analytical sensitivity analysis). Sensitivity of different components of the output to the same input can be different. The sensitivity analyses performed are documented, motivated and the results are discussed, e.g., with respect to model performance, precision or accuracy of measurements and input data.

AA *Protocol * evaluated*

The sensitivity analysis is performed according to a protocol that prompts repeated sensitivity analysis for subsequent (sub)versions of the model. N.B. changes in one part of the model can impact the sensitivity to other parts of the model. Periodically this protocol is evaluated for completeness.

4.2 An uncertainty analysis is performed

A *Qualitative discussion*

The uncertainties underlying the assumptions, structure and data sources are analysed in a qualitative way, naming both known and unknown uncertainties. The documentation includes a brief

description of the methods used. The results of the analysis are interpreted and discussed, e.g. with respect to model performance and reliability of the output/dataset. For datasets, this includes an analysis of the measurement error and sample design.

AA Quantitative analysis * evaluated

With the quantitative uncertainty analysis, the impact of known and quantifiable uncertainties in the input on the output is analysed. Various methods exist for different types of models, the choice for the method used should be motivated. The extent of the uncertainty analysis, both with respect to the quantification of uncertainties in the input, as with respect to the model components and outputs considered is periodically evaluated.

4.3 The model/dataset is validated

A Discussed * non-validated components named

Through validation one judges the validity of model output or dataset content based on external information. This information can come from various sources (e.g., measurements, literature or expert judgement). The method of validation used is documented. Components that are not validated (yet) are named. Validation status of the model/dataset is discussed, e.g., with respect to the interpretation and reliability of the model output or dataset content.

AA Protocol * evaluated

As with the sensitivity and uncertainty analyses, validation should be performed according to a protocol for each new version of the model/dataset. The extent of the validation, with respect to model components and outputs is periodically evaluated, e.g., considering new data, new model components etc..

4.4 The use of the model/dataset is monitored

A Example studies listed

Monitoring of the use of the model implies that one is aware of the use of the model inside and outside of the development team, this can range from tracking citations to evaluation of the application of the model/dataset. From this monitoring, example studies are drawn and referred to.

AA Use & use(r) experience tracked * evaluated * documented

Options for the evaluation of the use and use(r) experience are model/dataset specific. Core is that the feedback from users (active or passive) is reflected upon, documented and feeds into the development plan.

4.5 There is a general assessment of model/dataset quality

A Relate goal to: test * sensitivity * uncertainty * validation * use

The performance of the model/dataset (as reflected in the five evaluations: tests, sensitivity, uncertainty, validation and use) is related to the specifications (goal, intended area of application) to reflect 'fitness for purpose'. The fitness for purpose is documented.

AA Include reliability * precision * data used * external review

The general assessment of the quality is extended to include the quality (reliability, accuracy, precision) of the data used. During an external scientific review a number of scientific peers is asked to provide a fresh view on the fitness for purpose and relate assumptions to the relevant scientific state of the art. The aim of this review is to provide input for future developments (constructive) rather than to provide a binary judgement.

Development & Organisation

DO.5 The development of the model/dataset is planned

5.1 There is a development plan

A *List of plans * progress reported * based on evaluation*

There is a point-wise list of planned developments. The progress of these planned developments is periodically reported. The evaluations of model performance (partly) motivate these plans.

AA *Further evaluation * periodically updated*

A timeline for the planned developments is provided. To assure continuous development, further - extended- model evaluations are planned. The development plan, that includes a motivation for the planned developments, is periodically updated. The time horizon will be model/dataset specific, but should be mentioned.

5.2 A version control system is in place

A *Documented * acceptance criteria * (WUR) central archiving*

The method for keeping track of versions is documented, including what is, and what is not under version control. Version control should encompass both the development versioning (subversions during implementation) and production versioning (versions released for use). Version acceptance criteria are documented explicitly. Differences between versions are reported and discussed. The version control makes use of (WUR) central archiving.

AA *Protocol for documenting & code-commenting*

The protocol for version control includes a protocol for documentation and code-commenting (e.g., who produces text for what type of documentation, during development, and how and at which level of detail code is commented).

DO.6 The organisation around the model/dataset is planned

6.1 The metadata of the model/dataset is available

A *Domain appropriate format*

The metadata is provided in a domain appropriate format. The metadata reported should at least include items included in the WR modellibrary metadata format, that is available at <https://intranet.wur.nl/Project/WRModellingToolbox>.

AA *FAIR*

Up to date metadata is publicly provided and according to FAIR principles. FAIR principles state that (meta)data should be Findable, Accessible, Interoperable and Reusable.

6.2 There is a management plan

A *Responsibilities: content * technical * next-in-line * ownership * financial cover*

The management plan minimally lists the distribution of core responsibilities; content, technical development & maintenance, next-in-line responsibility and ownership. (Un)availability of funds to cover planned developments should be mentioned explicitly.

AA *Vision on future * periodically updated*

The management plan is further extended with a vision on the future use and development of the model/dataset, encompasses current and potential use, anticipated internal and external developments (e.g., personnel, technical) and how these could provide opportunities or pose threats to future operation. The management plan is periodically updated.

6.3 Dependencies are discussed

A *Datasources * (third-party) use*

A list is provided with the input and output dependencies of the model/dataset (e.g., updates in source data, monitoring networks, (third-party) use of model output). With dependencies we aim at (dynamic) data sources on which future model/database use is dependent, and at (third party) users that rely on (future) model output or dataset versions.

AA *Tracked * Obligations * liabilities*

Because these dependencies could pose threats for continuation or might provide opportunities for shared future development, the continued availability or planned demands from these dependencies should be tracked. Risks of losing input sources (by lack of alternatives) and explicit obligations for future operation should be highlighted.

6.4 External use is formalised

A *Conditions for use * User support*

Conditions for use outside the development group are defined. The responsibility for user support is named.

AA *User agreement * legally checked * financial paragraph*

A user agreement is available that is legally checked and in line with the ownership of the model/dataset. This user agreement contains a financial paragraph, even if no fees are charged.

Interpretation & Use

IU.7 User documentation is provided

7.1 Interpretation guidance is provided

A *Goal * area of application * theoretic framework * summary of evaluations * general public*

Interpretation of model output or dataset contents is in general not trivial. The interpretation implies understanding of the theoretic framework, conceptualisations and formalisations of the model - i.e., the assumptions and simplifications - and of the outcome of model evaluations. Guidance should be supplied on what these mean for the value of the outcome and on when (not) to use the model/dataset. This interpretation guidance should be readable for more general public than the scientific community.

AA *Reflection on goal, area of application, structure, complexity*

The interpretation guidance is extended with a reflection on the tension between goal, area of application and complexity of the model, limitations in the implementation, data quality and availability and the realised model performance.

7.2 There is a user manual

A *Operation instructions * installation guide * summary of technical documentation * minimal system requirement * format of input & output * contact information*

The user manual includes clear operation instructions, an installation guide, a summary of the technical documentation, listing the minimum system requirements, and clear documentation of the format, structure and content of the user-relevant input and output files. Contact information for user support is provided.