# The importance of simulation configuration to crop model development

**Jonathan J. Ojeda**[1], Hamish E. Brown[2], Neil Huth[3], Dean Holzworth[3], Rubí Raymundo[4], Robert F. Zyskowski[2], Sarah Sinton[2], Alexandre Michel[2]

[1] Regrow Ag, Brisbane, QLD 4101, Email: jonathan@regrow.ag, www.jojeda.com
[2] The New Zealand Institute for Plant and Food Research, Private Bag 4704, Christchurch, New Zealand
[3] CSIRO Agriculture and Food, P.O. Box 102, 4350, Toowoomba, QLD,
[4] Colorado State University, Fort Collins, CO 80521, United States

## Abstract
APSIM (Agricultural Production system SIMulator) Next Generation has in part been developed to accelerate the collation of test sets for crop model improvements. However, as far as we know, there is a lack of systemic testing of the quality of simulation configurations in such test sets. In this paper we suggest using all observations (i.e. measured datasets) available to scrutinize the test set and guide subsequent model calibration. We describe a simple but robust approach for scrutinizing simulation configuration. An exemplar crop model – potato (*Solanum tuberosum* L.) is used to identify the main sources of uncertainty during simulation configuration to accelerate crop model improvements. 426 experiments (44 cultivars from 55 locations across 19 countries) were run using APSIM. Model inputs and outputs were plotted using an automatic script. Based on these plots, we conducted a detailed interrogation of simulation configuration and sources of uncertainty, looking for systematic effects related to location, cultivar and crop management parameters. Sources of uncertainty were mainly associated with model input configuration (crop management>soil>climate) and inconsistencies in measured data. This study highlights the importance of high levels of rigor for configuring test sets and thorough consideration of the model performance so that subsequent model improvement can be effectively targeted.

## Keywords
Visualisation tools, model uncertainty, model improvement, APSIM Next Generation, Python, Metadata

## Introduction
The number of crop models and model users is increasing. Although these models have been developed to capture complex interactions between genotype, management and environment, generally, error propagation during simulation configuration is not tested, which creates difficulties to identify sources of uncertainty and, hence, priorities for model improvements. A lack of quality control during simulation configuration generates uncertainty in crop model outputs (Confalonieri et al., 2016). The main sources of uncertainty in crop model outputs are due to human errors (Palosuo et al., 2011), model structure and model parameters (Palosuo et al., 2011; Thorburn, 2017). The minimisation of the latter two is the focus of APSIM Next Generation development activities - using test sets including a wide range of field experiments (orange panels in Figure 1) for further model applications (green panels in Figure 1). Quality control in collation of model inputs needs to be addressed in the process of building a robust test set to reduce sources of uncertainties. This should involve identifying errors and outliers in model inputs (blue panels in Figure 1, the focus of this paper) so they can be minimized to ensure they do not confound the process of model development.
Previous efforts have been focused on the assessment of model input uncertainty due to data sources and data resolution. However, the uncertainties derived by simulation configuration (i.e. all steps required to set up input data for the model before the model is run) have been not widely investigated. This is a methodological paper describing a robust approach for scrutinizing simulation configuration. An exemplar crop model – potato (*Solanum tuberosum* L.) is used to identify the main sources of uncertainty during simulation configuration.



© Proceedings of the 20th Agronomy Australia Conference, 2022 Toowoomba Qld www.agronomyaustraliaproceedings.org

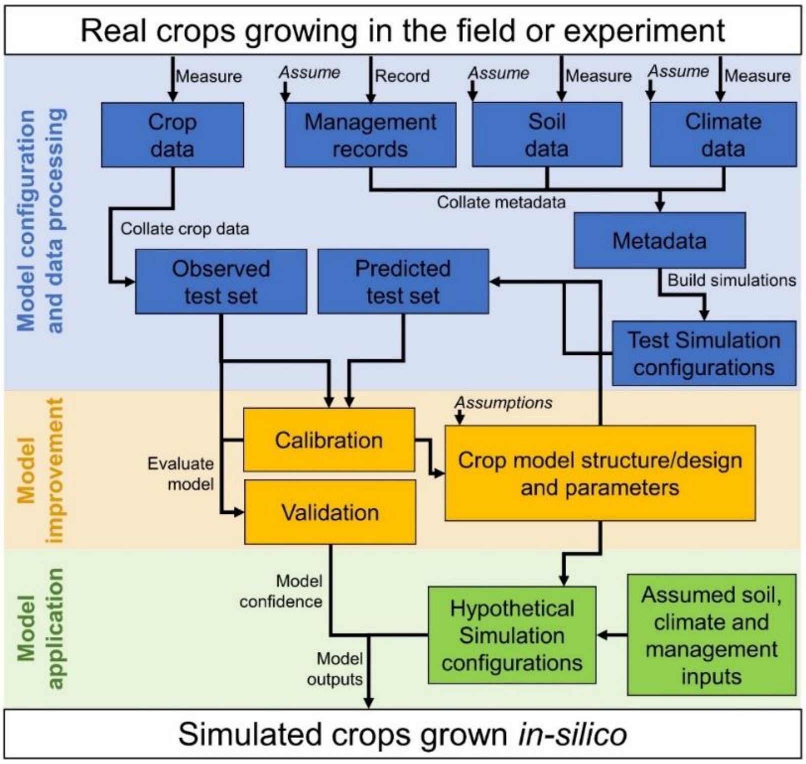


**Figure 1. Modelling workflow from real world observations to in-silico predictions. Different colours indicate steps in the modelling activities from model configuration and data processing (blue panels), model improvement and development (orange panels) and model applications (green panels). Note that this paper only focuses on the blue panels of this figure.**

## Methods

*Datasets and simulation setup*

The experimental test data chosen for testing the APSIM Next Generation (Holzworth et al., 2018) potato model were collected from 1970 to 2019 across 19 countries (55 locations), consisting of 44 cultivars. The experiments have been carried out to study dry matter allocation; yield response to various treatments, including nitrogen (N) fertiliser rates and timing, irrigation rates, time of planting, population, the adaptability of cultivars across locations and years and the effect of increased atmospheric $CO_2$ concentrations on crop development. We analysed 27 measured variables which included crop (e.g. leaf, stem and tuber) and soil (e.g. soil water) observations. The model was based on the existing potato model in APSIM implemented in the Plant Modelling Framework (Brown et al., 2014). The model inputs were daily climate data, soil profile parameters at a layer level, cultivar parameters, and crop management information. For methods see Ojeda et al. (2021).

*Visualisation tools*

To assess simulation configuration, model simulations were run, and model inputs and outputs (from three simulation reports [initial, harvest and daily]) were written to a database file (Brown et al., 2018a). This file was read by a Python script which generated a series of plots. Based on these plots, we conducted a detailed interrogation of simulation configuration and sources of uncertainty, looking for systematic effects related to location, cultivar and crop management parameters.

## Results

Several problems with configuration were quickly identified and corrected during the process of setting these tools up and all the issues fixed have been listed in Table 1 to demonstrate the type and extent of configuration errors that would typically go undetected in an evaluation set. The number of errors associated with observations (crop data) and configuration of crop management practices (management records) was high in comparison with climate and soil configurations (Table 1). The crop management data used to set up the complete set of simulations is showed in Figure 2. This plot allowed us to identify outliers and irregularities associated with the crop management parameters.

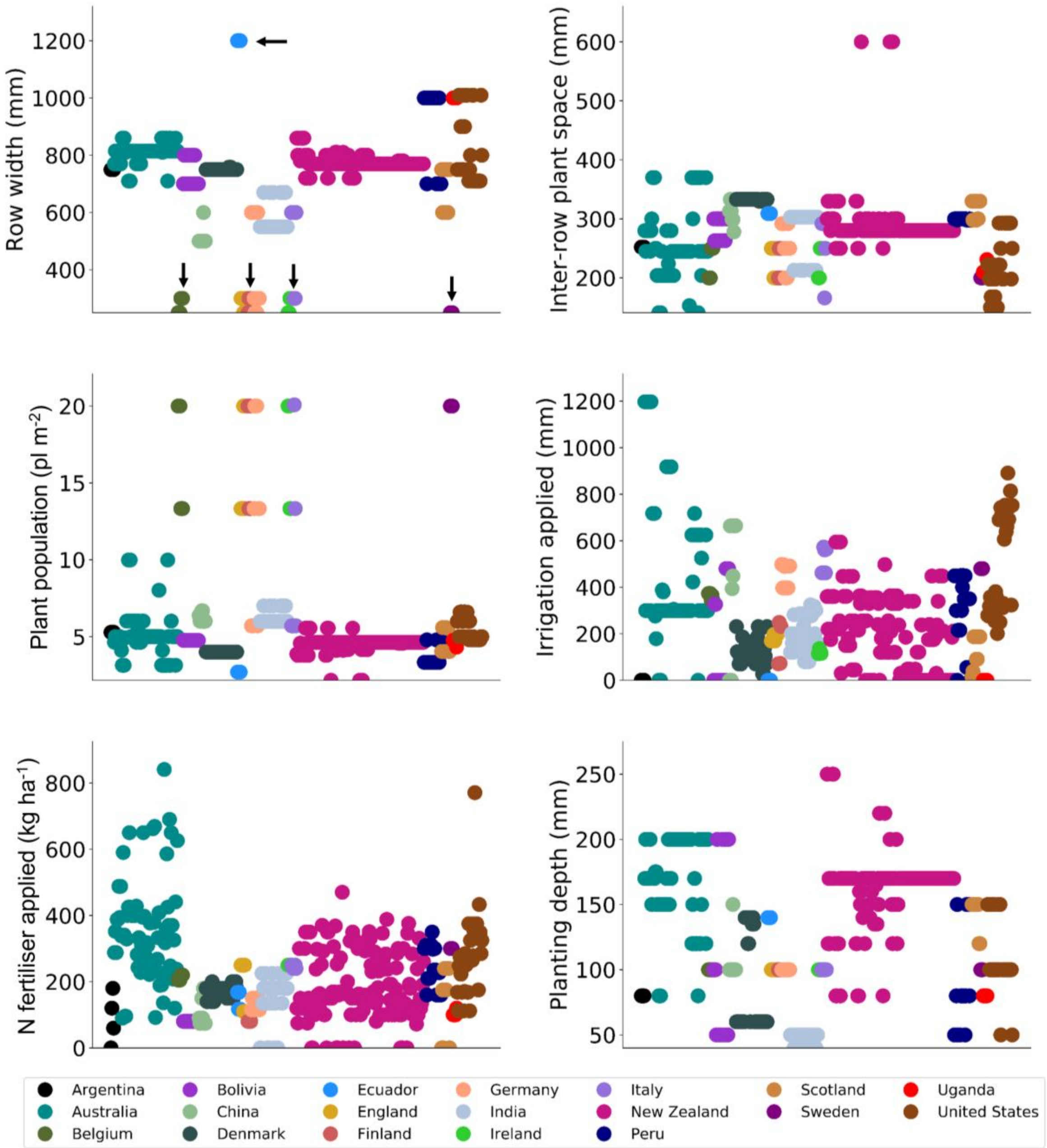


**Figure 2. Crop management variables (spacing between rows [Row width], spacing between plant within rows [Inter-row plant space], plant population [Plant population], irrigation applied during the growing season [Irrigation applied], N fertiliser applied during the growing season [N fertiliser applied], planting depth [Planting depth]) vs. country. Black arrows indicate locations with outliers of row width values for some locations.**

**Table 1. List of sources of uncertainties found during the simulation configuration process.**

| |
|---|
| *Crop data* |
| - Leaf nitrogen concentration values were extremely high because the unit conversion was calculated dividing twice a value for a given experiment.<br>- Mismatching between the observed and predicted days of tuber initiation.<br>- Issues matching harvesting dates.<br>- Inverted months and days of measurements. This error caused these experiments were not harvested on the correct date, so harvest observations did not match predicted values. |
| *Climate data* |
| - Longitude was in middle of see because it was a positive value when should be negative.<br>- Latitude was positive when should be negative. |
| *Soil data* |
| - Inconsistent methods for setting XF indicated by different patterns between data sources.<br>- Copying or guessing where parameters are not known. |

| *Management records* |
|---|
| - Planting depth was entered in cm when model needs mm.<br>- Row width and row spacing inverted.<br>- Incorrect dates, month and day inverted.<br>- Simulations not harvesting on the correct date so harvest observations not being matched.<br>- Nomenclature issues, e.g. row spacing was unclear if it was spacing between rows or between plants within rows. |
| *Others* |
| - Duplicated simulations from two data sources.<br>- Same location name in two countries causing ambiguity in analysis of results. |

## Conclusion

This study is the first demonstration of a model configuration that was tested including (i) high-standard and contrasting observed datasets, (ii) several model outputs and (iii) a palette of visualisation tools to identify sources of uncertainty during the configuration process. The value of these tools is to identify errors that could later undermine parameter fitting efforts. This approach moved the bulk of the effort from fitting model parameters to setting up a broad model testing and a detailed interrogation of all the model results to identify current gaps for further model improvement using APSIM Next Generation or other crop models. Our study also provides insights into the quality of the dataset by identifying gaps in the input data space.

**Acknowledgements.** This paper is a short version of the article published in Ecological Modelling 458, 109703.

## References

Brown HE, et al. (2014). Plant modelling framework: software for building and running crop models on the APSIM platform. Environmental Modelling and Software. 62, 385–398. (doi.org/10.1016/j.envsoft.2014.09.005)

Confalonieri R, et al. (2016). Quantifying uncertainty in crop model predictions due to the uncertainty in the observations used for calibration. Ecological Modelling. 328, 72–77. (doi.org/10.1016/j.ecolmodel.2016.02.013)

Holzworth D, et al. (2018). APSIM Next Generation: Overcoming challenges in modernising a farming systems model. Environmental Modelling and Software. 103, 43–51. (doi.org/10.1016/J.ENVSOFT.2018.02.002)

Ojeda, JJ, et al. (2021). Assessing errors during simulation configuration in crop models–A global case study using APSIM-Potato. Ecological Modelling 458, 109703.

Palosuo T, et al. (2011). Simulation of winter wheat yield and its variability in different climates of Europe: a comparison of eight crop growth models. European Journal of Agronomy. 35, 103–114. (doi.org/10.1016/j.eja.2011.05.001)

Thorburn PJ, (2017). Estimating uncertainty in crop model predictions: Current situation and future prospects. European Journal of Agronomy. 88, A1–A7. (doi.org/10.1016/J.EJA.2017.06.001)