

How to define model performance criteria?



Elke I. Zimmer¹, Starrlight Augustine², André Gergs³, Thomas G. Preuss⁴,
Virginie Ducrot⁴

¹ ibacon GmbH, Arheilger Weg 17, 64380 Roßdorf, Germany

² Akvaplan-niva, Fram – High North Research Centre for Climate and the Environment, 9296 Tromsø, Norway

³ Research Institute gaiaac, Kackertstrasse 10, 52072 Aachen, Germany

⁴ Bayer AG, Crop Science Division, Research & Development, 40789 Monheim, Germany

Elke.Zimmer@ibacon.com

Context

- Mechanistic effect models can be used to understand the effects of plant protection products (PPPs).
- The EFSA Scientific Opinion on Good modelling practice (EFSA, 2014) gives guidance for modelers on how to test and document the models.
- The document also defines evaluation criteria that should be used to evaluate the models upon submission.
- Though many of the evaluation criteria are defined in much detail, there are some in need for clarification, e.g. the **model performance criteria**.

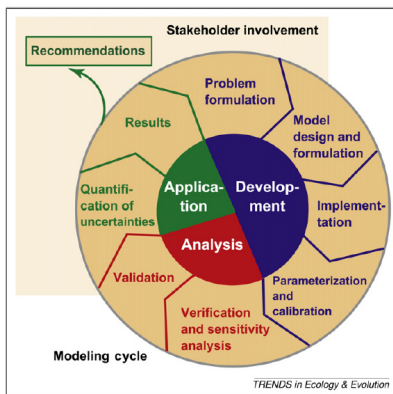
What are model performance criteria?

- **Aim:** to define **how qualitatively / quantitatively** accurate model predictions should be so that they are deemed reliable enough for the support of risk assessment.
- **In practice:** what are our expectations with regard to the quality (e.g. precision, accuracy) of the model prediction for the purpose of being used in a regulatory context?
- **Requirement by EFSA:** Criteria should be defined before starting the modelling exercise, in order to help the modeler evaluate the model along the modelling cycle (when is the model good enough?) and after completion (is the model output fit for the purpose of supporting the refinement of a risk assessment) .

Can we make a difference between qualitative and quantitative model performance?

- For model outputs that give mechanistic insight on biological responses of the animal but are not directly used for support to the risk assessment (e.g. respiration rate), is it enough to place only **qualitative expectations** (as derived from pattern oriented modelling) i.e. the model should be able to adequately predict the pattern seen in the data?
 - For model outputs that will be directly used as support to the risk assessment (e.g. length, cumulated number of offspring), the quantitative performance of the model needs to be addressed
- For example: If the mean model prediction lies within the natural variability and measurement error of the data as collected during standard tests, the performance of the model for predicting a specific data is considered to be sufficiently accurate
- How does it work out in practice?

How do performance criteria fit into the modelling cycle?



- During each step of model development, models are compared to data. The modelling cycle is represented in the scheme to the left.
- We need data for model calibration and validation at different stages of the model development.
- When the model does not fit to the data, the modeller needs to identify the cause for the deviation.
- Causes can e.g. be a missing process in the model, or sometimes a misinterpretation of data.
- The exact structure of the final model and processes included can usually not be defined beforehand.

Modeling cycle for ecological models in decision support (Schmolke, 2010). The elements of the cycle correspond to the elements of the TRACE documentation format, which are grouped in Model Development (blue), Model Testing and Analysis (red), and Model Application (green). Ecological models are developed by several iterations of the modeling process or parts of it. [...]

How to deal with variability in data?

- How much data (e.g. growth curves/weight) do we take to define the experimental variability range ?
- Do we need different data for defining the natural variability / measurement error than for model calibration?
- It is likely that the variability in the standard data is less than that in the calibration data set (where we mix data from standard and non-standard studies): this may create a bias toward declaring that model predictions have a greater uncertainty than experimental data variability. How to handle that?

Example: a physiological model that should be used to predict effects on growth

Context:

- We are parametrizing a Dynamic Energy Budget model that should be used e.g. to predict the effects on growth in a early life stage (ELS) test for fish.
- The model is a mechanistic effect model that allows for predicting effects on multiple endpoints simultaneously

Based on what type of data should we define the model performance criteria?

Rationale 1: the model is as accurate as the standard in vivo experiment we are focused on in the problem formulation for the risk assessment: reference variability is thus one fish experiment.

Rationale 2: the model is as accurate as a set of e.g. 10 standard in vivo experiment of the same type as the one we are focused on: reference variability is thus 10 fish experiment conducted in the same laboratory over the years.

Rationale 3: the model is as accurate as a set of ring-test data (where available): reference variability is thus 4 to 10 fish experiment conducted in various laboratories around the world.

If predictions for growth at the end of a standard test are a performance criteria, can we use the same type of data for model calibration?

When using the model to predict effects on other endpoints, will the defined performance criteria be different?

Increase in variability

Acknowledgements: This project is funded by Bayer AG

References:

EFSA (2014). Scientific Opinion on good modelling practice in the context of mechanistic effect models for risk assessment of plant protection products, EFSA Journal
Schmolke, A., Thorbek, P., DeAngelis, D.L., Grimm, V. (2010). Ecological models supporting environmental decision making: a strategy for the future. Trends in Ecology Evolution, 25(8), 479-486