All models are wrong, some are useful… but how do you know?

<u>Andreas Nicolaidis Lindqvist</u>
Stockholm Resilience Centre, Stockholm University
RISE, Research Institutes of Sweden
andreas.nicolaidis@ri.se

David Collste, Celinda Palm, Sarah Cornell
Stockholm Resilience Centre, Stockholm University

Jeanette Melin
Swedish Defense Unive

Keywords: Fitness for purpose, Validation, Evaluation, Integrated Assessment Models, Model quality assurance

There is broad agreement that simulation models need to be fit-for-purpose to be of value (Beven and Young, 2013; Hamilton et al., 2021; Jakeman et al., 2006). However, despite broad scholarly agreement about the importance of ensuring models are fit-for-purpose, there is considerable ambiguity and even conflicting views of what this really means (Beven and Young, 2013; Hamilton et al., 2021). Even within individual modeling studies it is rare to find a clear definition (ex-ante) and subsequent assessment (ex-post) of model fitness. This is problematic for multiple reasons:

1. Lack off a clear definition of model fitness hinders the development of guidelines to how it is achieved (Hamilton et al., 2021).
2. Without well-defined theoretical framing of what constitutes fitness, tools and frameworks for robust and transparent assessment and evaluation of model quality cannot be developed (Hamilton et al., 2021).
3. Following from the preceding two assertions, ambiguity in the definition and methods for assessing fitness makes objective benchmarking and comparison of models impossible.
4. Without agree definitions and methods that facilitate robust assessment of fitness-for-purpose there is a risk that time and resources are spent on the development of models of little societal value.

Some would argue that following well-established guidelines for good modeling practices is enough to ensure that the resulting model is fit-for-purpose but we argue that when modeling to support policy, education or decision making, assessing fitness-for-purpose should stretch beyond the traditional narrow focus of what is "in" the model, and the checklist of best-practice steps conduced when building it, to include the broader context in which it was developed and used.

The aim of this paper is to address the challenge of assessing model fitness-for-purpose by presenting a structured framework for:

1) ***defining model fitness*** by first holistically framing the modeling context and then, given the contextual framing, assessing the relative importance of different fitness quality criteria;

2) ***developing metrologically useful quality criteria***, presenting conceptual descriptions of the selected quality criteria in the form of construct maps (Wilson, 2005) that form the basis for subsequent quantitative measurement of model fitness;

3) ***deriving robust quantitative measures of model fitness*** based on established metrological approaches grounded in Rasch measurement theory (Rasch, 1960).

We pilot the framework using the ongoing EU funded WorldTrans project (https://worldtranseu.wpcomstaging.com/about-us/) and the system dynamics-based global integrated assessment model (IAM) FRIDA v.1.0 as case study. We present preliminary results and discuss how the framework complement existing well-established methods for model evaluation (Barlas, 1996; Sterman, 2000). The proposed framework provides a structured and holistic approach to ensuring model utility, which we believe can enhance model uptake and policy relevance.

Another novelty with the presented framework is the adoption and integration of metrological theory and methods into the simulation modeling paradigm. To our knowledge, this is the first-time model quality criteria are assessed using metrologically rigorous standards and we see great scientific value in continuing exploring the untapped potential learnings and synergies between the metrological sciences and the system dynamics modeling domain. We invite for constructive discussions around methodological choices and further development of the presented research endeavor.

References

Barlas, Y., 1996. Formal aspects of model validity and validation in system dynamics. System Dynamics Review, 12(3): 183-210. DOI:https://doi.org/10.1002/(SICI)1099-1727(199623)12:3<183::AID-SDR103>3.0.CO;2-4

Beven, K., Young, P., 2013. A guide to good practice in modeling semantics for authors and referees. Water Resources Research, 49(8): 5092-5098. DOI:https://doi.org/10.1002/wrcr.20393

Hamilton, S., Pollino, C., Stratford, D., Fu, B., Jakeman, A.J., 2021. Fit-for-purpose environmental modeling: Targeting the intersection of usability, reliability and feasibility. Environmental Modelling & Software, 148: 105278. DOI:10.1016/j.envsoft.2021.105278

Jakeman, A.J., Letcher, R.A., Norton, J.P., 2006. Ten iterative steps in development and evaluation of environmental models. Environmental Modelling & Software, 21(5): 602-614. DOI:https://doi.org/10.1016/j.envsoft.2006.01.004

Rasch, G., 1960. Probabilistic Models for Some Intelligence and Attainment Tests. Danmarks Paedagogiske Institut.

Sterman, J., 2000. Business Dynamics: Systems Thinking and Modeling for a Complex World. McGraw-Hill.

Wilson, M., 2005. Constructing measures: An item response modeling approach. Constructing measures: An item response modeling approach. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, xix, 228-xix, 228 pp.