Validation & Verification

How to validate ABMs

Debraj Roy

January 18, 2022

Computational Science Lab

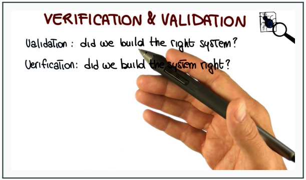
Table of contents

Introduction

Validation & Verification are terms using in Modelling and Simulation

Validation : The process of checking if your conceptual model is an accurate representation of the real system (did you build the right model)

Verification : The process of checking if your implementation (code) matches the conceptual model you have described.



Validation?

Validation

/val de (ə)n/

the action of ***checking*** *or* ***proving*** *the validity or accuracy of something.*

— Oxford English Dictionary

- What do we mean by accurate?
- When is a model valid (enough)?
- Can we ever prove a model valid?

ABM Validation & Verification?

In Agents?

At first glance it may seem that it's hard to distinguish ***validation*** and ***verification*** for ABM.

Your model is naturally algorithmic.

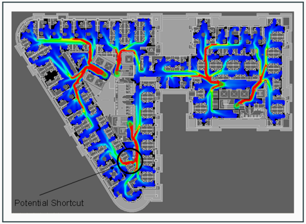
However, it's always a very good idea to write formal descriptions for your agent-based model: rules, constants, parameters, etc. see (Epstein 2002) for an example.

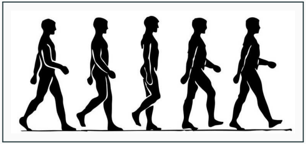
For what remains we will focus on Validation - techniques for verification in other modelling domains can be used in ABM also.

Validation?

In traditional modelling, validation can be thought about in a number of ways (Zeigler, Praehofer, and Kim 2000)

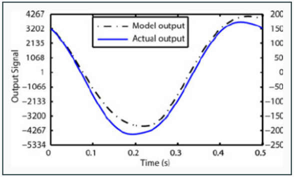
- Replicative Validity
- Predictive Validity
- Structural Validity





Replicative validity?

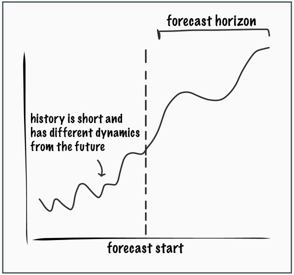
Replicative validity simply asks for the measured input and output parameters, under the conditions of our experimental frame, can we distinguish the model from the systems?



Predictive validity?

Predictive validity covers all of replicative validity plus the ability to predict as yet unseen input and output data ?

Note that the model may be invalid in this case for a number of reasons. Firstly, the model is wrongly calibrated or was conceptually incorrect. Secondly, the new observations may have been under and new experimental frame (conditions have changed). Think about the example of a financial system, under 'stable' conditions and under 'crisis' conditions.

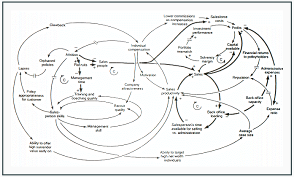


Structural validity?

Structural validity looks at the inner workings of the model. Does the model replicate the way in which the system transitions step-by-step or component-by component.

A model in this sense is supposed to somehow reflect the state transitions of the real system. This distinguishes 'real' models (mathematical models, ABMs, CAs, Networks, etc.) from simple predictive models (regression, Neural Networks etc.).

Predicting is not enough - it should operate in a similar way.



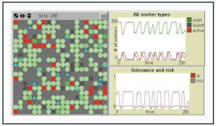
Challenges of ABM

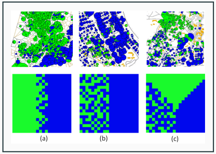
In this course we covered various Agent-based Models. We've also asked the question – *Why Model?*

Validation in the traditional sense requires data, it requires an experimental frame and observations.

A lot of the agent-based models (e.g., El-Farol, Civil Violence, etc.) don't claim to have predictive validity.

Also in many cases ABMs model systems that are hard to experiment with or generate experimental frame (e.g., models of infectious diseases)





Frameworks for ABM Validation

We'll now exam two papers that discuss issues, options and methodologies for validating ABM.

Windrum (Windrum, Fagiolo, and Moneta 2007) : Alternatives for validation in ABM.

Kulgl (Klügl 2008) : A general proposal and outline for ABM validation

Windrum

There is heterogeneity in the ABM community. Because of this there is variation in the validation methodology.

Stylised Facts : Looking at output traces, testing counter factuals. Models of Epstein etc.

Predictive Simulations : Empirical validation, looking at input/output variables and comparing to data.

Three Approaches

In (Windrum, Fagiolo, and Moneta 2007) the authors describe three methods/approaches for conducting validation of ABM. Their focus is on economic models, but some of these methods can be applied more broadly - especially within social science models.

We'll go through and then discuss.

- Indirect calibration
- Werker-Brenner approach
- History-friendly approach

Windrum

Indirect Calibration

Indirect Calibration

Four-step procedure that first performs validation and then indirectly calibrates the model by focusing on parameters that are consistent with output validation.

1. Identify stylised facts that are to be reproduced. Macro level quantities that should be reproduced (e.g., relation of unemployment vs. GDP, distribution of wealth - form)
2. Develop the model by trying to recreate the micro dynamics to best match the known/empirical evidence about behaviour (e.g., data on decision making, trader behaviour, mobility choices of individuals, etc.)
3. Use constraints on the stylised facts to restrict parameter space of micro dynamics/behaviours.
4. Reason about the causal mechanisms that drive the model to capture the stylised facts. Then verify that the model can generate another stylised fact not specifically designed for.

Let's consider an example of how this might work.

Consider a model related to crowd dynamics (e.g., social force).
See (Johansson et al. 2008).

1. We may identify a density vs flow relation as a stylised fact that has been measured.
2. Develop the model by thinking about how people avoid each other at a local scale (separating away from each other). Build in this decision making into the model
3. Adjust the model parameters relating to speed, force coefficients etc. such that you create the same shape density vs flow relation
4. Reason that this density flow relations should create other phenomena (or not) under certain conditions e.g., stop and go waves, oscillation of passing direction at bottlenecks.

Indirect Calibration - Limitations

There are some weaknesses to this approach:

1. No real empirical validation is done - the focus is not on creating values that exactly match the real world observations. The stylised facts should be very carefully chosen so as not to be too 'obvious' or 'trivial'. If the same stylised facts can be produced many ways - then the value of the model can be questioned.
2. The third step in the process, refining the parameter space raises questions about counter-factuals and how to vary/analyse the parameters within the space.

Windrum

Werker-Brenner Approach

Werker-Brenner Approach

Three-step procedure for empirical calibration. The idea assumes that you have a set of plausible models.

1. Use existing empirical knowledge to calibrate parameters of the models. Using wide-ranges for parameters which are unknown or for which there is little data.
2. Validate against other output measures for the calibrated parameters is step 1. Bayesian inference can be used - each model specification has a likelihood of being correct based on ratio of simulation output vs. real output. Throw away parameter settings that are inconsistent (unlikely).
3. Then perform a further round of calibration on remaining models and parameters sets. Also conduct an 'expert' evaluation of the model - try to identify common underlying aspects/structure.

Werker-Brenner Approach - Limitations

There are some weaknesses to this approach:

1. Eliminating the set of models assumes that a true model is initially identified. The likelihood based elimination would be comparative.
2. The method creates bias for parameters for which we have data - you need to be sure that you've measured the right parameters. Otherwise the model may favour parameters that are less important.
3. Requires sufficiently high quality data - to be sure that the statistical approach works.

Windrum

History-Friendly Approach

History-Friendly Approach

A quantitative approach that uses specific historical case studies.

1. Build the model using a wide-range of available data - from detailed studies to anecdotal evidence.
2. Calibrate and validate model to particular case - retro-dict historic events. Find parameters and model where Simulated history = Real history
3. Then use sensitivity analysis to understand when or how the model breaks down - when are we no longer able to retrodict.

History-Friendly Approach - limitations

1. Tailored to particular historical evolution - overfit.
2. Backward induction from traces to models and parameters may lead to one plausible model - but there may be many.
3. Testing of counter-factuals, how do we generate alternative histories - the model may simply not be capable of this.

Klugl (Klügl 2008)

Klugl Methodology

Klugl (Klügl 2008) presents another possible methodology. She also describes different forms of validity - some that are consistent with the ideas of Epstein and why we model.

Types of Validity

Kulgl identifies two dimensions of validity related to the:

a. Method used for validation
b. The element of the model validated

Dimension One - Method

Kulgl defines two methods for established validity.

Facial Validity : Human intelligence assessing the model output/design. E.g., structured walkthrough, expert assessment, animations, etc. Plausibility checking

Empirical Validity : Direct statistical comparison with the either (a) the real system or (b) another model

Dimension Two - Element

We can validate different aspects of the model

Behaviour validity : This considers the model as a black(grey)-box where we simply observe input-output behaviour of the system.

Structural validity : This looks at the internals of the model - do we capture the causal relations, physical components, right actors, etc.

Behaviour can be classified further to predictive or replicative (Zeigler, Praehofer, and Kim 2000).

Challenges of ABM

- Characteristic Output Descriptors: At macro scale and micro scale. Micro scale challenging for data, and how to generalise individual behaviour.
- Focus on Transient Dynamics: How to properly compare time-series analysis in ABM.
- Non-linearities & Brittleness: Chaotic behaviour, or impreciseness hidden (e.g., update sequence)
- Size of validation task: Computationally and Data hungry.
- Impossible Falsification: Over-parametrization leads to models that can fit any output data.

References

Epstein, Joshua M. (2002). "Modeling civil violence: An agent-based computational approach". In: *Proceedings of the National Academy of Sciences* 99.suppl 3, pp. 7243–7250. DOI: 10.1073/pnas.092080199. eprint: http://www.pnas.org/content/99/suppl_3/7243.full.pdf. URL: http://www.pnas.org/content/99/suppl_3/7243.abstract.

Johansson, Anders et al. (2008). "From crowd dynamics to crowd safety: a video-based analysis". In: *Advances in Complex Systems* 11.04, pp. 497–527.

Klügl, Franziska (2008). "A Validation Methodology for Agent-based Simulations". In: *Proceedings of the 2008 ACM Symposium on Applied Computing*. SAC '08. Fortaleza, Ceara, Brazil: ACM, pp. 39–43. ISBN: 978-1-59593-753-7. DOI: 10.1145/1363686.1363696. URL: http://doi.acm.org/10.1145/1363686.1363696.

Windrum, Paul, Giorgio Fagiolo, and Alessio Moneta (2007). "Empirical validation of agent-based models: Alternatives and prospects". In: *Journal of Artificial Societies and Social Simulation* 10.2, p. 8.

References II

Zeigler, Bernard P, Herbert Praehofer, and Tag Gon Kim (2000). *Theory of modeling and simulation: integrating discrete event and continuous complex dynamic systems*. Academic press.