Centre for Policy Modelling, MMUBS, Manchester

Stockholm Environmental Institute, Oxford Office

Universität Kassel

Politechnika Wroclawska

IIASA, Vienna

Macaulay Institute, Aberdeen

Uniwersytet Wroclawski

Working Paper on Case Study Structure, Stakeholder/Agents and Validation Data

Deliverable No. 8 of Project 012816: CAVES –
Complexity, Agents, Volatility, Evidence and Scale

Duration: 2005-2008

Funded under the EU 6FP NEST programme.

CAVES validation protocol and application

Lead author: Takeshi Takama and Anton Cartwright

Executive Summary

- Models are constructed because researchers are restricted in their observation of the phenomena of interest. When the model is being used for inference and decision making the burden is on the model's designers to demonstrate the degree of correspondence between the model and the real world phenomena it seeks to represent, and to establish the confidence that can be placed in model inference.
- Validation of agent based models has the ability to enhance the credibility of their findings, facilitate comparisons between models and inform the confidence that can be placed in inference from the model.
- Because natural systems are open, it is not possible to establish benchmarks of absolute truth against which to check model veracity. For ABM, which are not so much concerned with forecasting reality as they are generating insights, this need not be a problem provided modellers do not make undue claims about the level of confidence that can be placed in their work.
- Knowing what the model is required for, and the level of certainty that is required, can inform the model design and application.
- Modellers in the CAVES study should reflect on the general principles that inform validation in traditional models. These include effective model design which is typically a prerequisite for validation. Model design and knowing why the model was constructed will suggest the best form of validation.
- Validating a component of the model may be legitimate to validate stakeholders, inputs, assumptions or outputs depending on the aims of the model.
- Validation of ABM depends on agent interviews should emphasise to stakeholders that ABMs are aimed at insights and not forecasts.
- The use of non-quantitative information assessment techniques such as Role Playing Game and Companion Modelling can be useful in gauging the pedigree of information used in ABMs.
- This paper proposes a guidelines of "validation for purpose" with sensitivity matrix. The matrix report high priority validation activities to the project team

Table of Contents

1 INTRODUCTION

The benefits of simulation models cannot be fully realised without an understanding of how accurately the model represents the dynamics that it seeks to capture. Models are constructed because researchers are restricted in their observation capacity of the phenomena of interest by either time[1], space or budget[2] (Oreskes, et al. 1994 p. 644). When the model is being used for inference and decision making the burden is on the model's designers to demonstrate the degree of correspondence between the model and the real world phenomena it seeks to represent and to establish the confidence that can be placed in model inference.

It will be suggested that validation refers to the various activities that provide evidence that models serve (or do not serve) the purpose for which they were formulated – validation denotes legitimacy. Verification and validation are often confused even amongst modellers. Both validation and verification are important in the process of computer simulation modelling (Balci. 1994). Model validation deals with building the right model – an appropriate characterisation of the real world. In contrast, model verification deals with building the model right, which relates to the transformation of a problem formulation into a model presentation. Unlike validation, verification does not necessarily denote an accurate depiction of the real world or "truth", but instead describes model legitimacy – no known flaws or logic inconsistencies (Oreskes, et al. 1994). Unlike verification, standardisation and guidelines for validation activities has not formally developed[3].

The CAVES project involves the application of ABM to land use management issues in three case studies. The CAVES models do not aim to predict land and water use under future scenarios, but seek to provide insight into complex social and environmental systems that might prove useful in understanding how these systems respond to shocks. There is a particular focus on how social networks influence the response to shocks. Agent based models are well suited to this purpose, but in line with the observations of Janssen and Ostrom (2006) a key question for the case studies and the broader CAVES project, involves how to ensure the robustness of model validation findings. How do

---

1 e.g. the researchers do not have enough time observe reality and/or that the events referred to are in the past or future and so these are not completely unobservable.

2 Also, it is constructed to reduce complexity of study problmes.

3 These requirements have been adopted from the software engineering environment that defines principles for the programming of agent-based models (Wooldridge 2002). Computer programmers have developed standard tools and approaches to help the verification process of software development. For example, Unit Test checks if each functions or method of a program code works as a programmer expects (Beck 1994, Beck 1999) and regression testing (automated testing) provides an easy means of checking whether a whole program works as expected whenever a program code has been modified (Brooks 1995 p.122). Revision control systems such as CVS (http://www.nongnu.org/cvs/) and SubVersion (http://subversion.tigris.org/) are often used to record versions of program codes, so that programmers can check different results from different versions and can safely modify the program codes (Singh et al. 2004). These standardised verification tools are available to agent based modellers as well (GilbertandTroitzsch 1999 p.22). In contrast, validation techniques are not well developed in the software engineering domain, especially relative to the other sciences. For example, roughly 20% of software engineering papers have no validation component, and a further third make only weak reference to validation (Zelkowitz and D. Wallace. 1997). Critically, the validation of empirical ABM as they are applied to the social sciences mainly on the in-field data collection of human and environmental behaviours. In this sense the application of software engineering validation is insufficient and the onus on AB modellers is to devise appropriate – and ideally standardised - means of model validation (See more in Gilbert 2004).

researchers ensure that model outcomes reflect persistent (locally generic) aspects of the system under study, rather than a modeller's choice of parameter settings, initial conditions, or software/hardware platform preferences? Perhaps even more importantly, is it possible to develop standardised validation procedures that will make the approach adopted in CAVES replicable?

This working paper does not provide definitive answers to these questions. On the contrary, the experience of the CAVES research teams to date lends credence to the notion that a standardised approach to validating social systems, and agent based models in particular, remains elusive if not impossible . The paper does, however, suggest procedures that may be followed in order to ensure a greater level of confidence and credibility in the models. The challenge is to show the validity of agent based models and their findings without forfeiting their ability to capture complex social realities. Successful validation will enhance the credibility of agent based models and allow for comparisons across the three case studies (e.g. Axtell, et al. 1997). The remainder of this validation paper is structured as:

- Section 2 reflects on the problems of validating ABMs
- Section 3 describes approaches that can be used to address the problems of ABM model validation: appropriate design, component validation, and the validation for purpose
- Section 4 concludes the paper.
- The appendix proposes a sensitivity matrix to prioritise validation activities through a life-cycle of modelling. These actions can be taken to aid validation in the CAVES project.
- The complimentary paper about the case validation introduces validation actions planned in the CAVES project such as methodology, approaches and protocols.

2 PROBLEMS WITH ABM VALIDATION

The initial attraction of ABMs was the ability to relax the assumptions applied in conventional social science models and introduce heterogeneous agents that interact in multiple different ways. Successful agent based models yield general insights but remain applicable in specific cases. They also allow for the scaling up of the processes of interactions between a few agents to interactions among many agents. Collectively these features of ABM allow for a more accurate depiction of the "real world" and allow for the explanation of macro-level phenomena such as spatial patterns and levels of cooperation that were often overlooked in conventional models. It is the same features (and especially the heterogeneity amongst models), however, that have frustrated attempts to develop a standard validation approach, especially as they have moved from the extensions of convention of formal models to more empirical work and validation has consequently become an important extension by which to judge a model. As Janssen & Ostrom (2006) point out, very few applications of ABM have been empirically tested.

Whilst the use of agent based (AB) models has become increasingly popular in the social sciences the validation of such models remains difficult. This is principally due to the fact that the systems that

ABMs seek to capture are infinitely complex – they are not "closed systems" for which a set of definitive rules can be generated (Oreskes, et al. 1994). The inability, or failure, of agent based modellers to validate their work, results in many of their findings being considered anecdotal (Fagiolo, Windrum et al. 2006). The lack of a widely recognised validation approach in ABM tends to undermine the credibility of model findings and may be one of the reasons why research involving agent based modelling has struggled to gain its due recognition in the mainstream literature (Leombruni and Matteo Richiardi. 2005, Richiardi, et al. 2006).

In an information technology age it is increasingly the case that it is not the availability of data that constrains effective decision making but the reliability of the data. Placing a measure of confidence in modelled results is a prerequisite for credible insights. Validation seeks to determine the level of confidence that can be placed in a model and is frequently used to upgrade the veracity of a research process. Central to the issue of validation is an acknowledgement of the difference between precision and accuracy. Accuracy relates to the degree of conformity of a modelled result with its actual or true result. Precision relates to the degree to which further models or measurements will produce the same or similar results. Technically a result is considered valid if it is both precise and accurate.

In the social sciences and particularly in agent-based models, both accuracy and precision are rare. Inaccuracy and a lack of precision can enter agent based models at all stages. Inputs can be either inaccurate or imprecise, or both; model assumptions can be wrong or poor proxies of the systems they seek to represent; and the modelled outputs can reflect the product of both data and model impediments. The literature (Fagiolo, Windrum et al. 2006, Schreiber 2002, See, for example, Carley. 1995) acknowledges at least 4 distinct sources of inaccuracy: technical (inexactness), methodological (unreliability), epistemological (recognized ignorance) and societal (social robustness). All of these contribute to uncertainty and reduced confidence. Indeed imprecision and a lack of veracity can cascade throughout the research and modelling process compounding inaccuracy in the final conclusions. This does not necessarily pose a problem. In the quest for insights, as opposed to projections, researchers in the social sciences have devised many ways of coping with a lack of, or poor, data.

3 PROCEDURES FOR ADDRESSING THE DIFFICULTY OF ABM VALIDATION

As mentioned above, validation in empirical ABM is difficult, however, there are a few validation procieures ABM modellers can tackle to improve model accuracy and precision. If the aim in social simulation is a "tentatively adequate" (Hendry and J. F. Richard. 1983) approximation of complex and unobservable systems in which agents interact with each other and with their environment, then external validation aims to ensure that this approximation remains credible. Unlike more conventional statistical models for which a range of formal tests for validity and representativeness exist, validating Agent Based models with an emphasis on process rather than prediction, relies on more pragmatic approaches. Whilst it might not be possible to develop a definitive approach to validating models that simulate open systems, there are number of procedures that modellers can follow to establish the level of confidence that can be placed in the model findings, and where possible to improve that level of confidence. Therefore, this section introduces 1) model design approach, 2) component validation approach such as macro- and micro-validation, and 3) validation for purpose approach.

3.1 The importance of model design

To a large extent the difficulties experienced in validating ABMs can be mitigated through effective model design. Well constructed models, by definition, lend themselves to better validation. This section begins with a summary of modelling approaches and principles, many of which include in-built validation. This section introduces the Keep It Simple, Stupid (KISS) principle, the Keep It Descriptive, Stupid (KIDS) principle, and the Take A Previous Model and Add Something (TAPAS) approach to model design.

KISS

According to KISS, parsimony - capturing the most salient features of a real world situation with as few parameters as possible – should guide all modelling efforts. Fewer parameters can make for easier insights and, in statistical analysis, reduce the degrees of freedom. But it is equally important to ensure that the simplifications or stylisations do not detract from the model's credibility. The insights that are facilitated by simple approaches are what guide the Keep It Simple, Stupid (KISS) principle to model design (Axelrod 1997). However, the real world is almost always "wondrously complicated", and accordingly we should be aware of the "inherent weaknesses of the beautiful human mind" (Hoffmann, et al. 1997). Researchers often accept and apply the merit of simplicity in their work. Provided that the KISS approach, with its emphasis on parsimony, is not used as an excuse for expedience it provides a powerful concept in model design. As Gilbert (2004) mentions, over parameterised and overly complex models run the risk of concealing underlying principles and useful insights, and render their application difficult. This is particularly true of Multi-Agent Simulations, in which the objective is not to implement a detailed decision-making process involving numerous data and complex calculation, but rather to see how simple behaviours lead to complex phenomena Where a model effectively captures and represents a real world situation with few parameters it is easy for stakeholders and outside experts to understand the model and so to validate its workings and its outputs.

KIDS

The Keep It Descriptive, Stupid (KIDS) principle aims to make the simulation as descriptive as the information and resources will permit (Edmonds, Moss 2005) and then adapts this simulation as becomes necessary. Unlike the KISS approach, KIDS principles do not aim for generalisations, but for context specific, and at times intricate, detail. There is an express acknowledgement in KIDS models that the modelled phenomena are frequently complex and that anecdotes from interviews or in species identification[4] for example, provide useful insights. The advantage of KIDS – relative to KISS – models is in their accessibility. Good KIDS models, because of their descriptive content, are more amenable to scrutiny and criticism by stakeholders and experts than KISS models, which too easily become enigmatic and obscure. This attribute of KIDS models also makes them easier to validate and less capable of "garbage in garbage out" errors. Validation of KIDS models is typically via stakeholder and expert review. The downside of this approach is that it does not say how to start

[4] Small details will be important to distinguish closely-related species. Therefore, detailed description is important for species identification, but not simplicity.

modelling a complex phenomenon, which has too many descriptive characteristics to start with, i.e. which characteristics should researchers have to describe first?

TAPAS

Take A Previous Model and Add Something (TAPAS) is a heuristic approach that draws on credible existing models to provide cost and time effective modelling solutions (Frenken. 2006). The TAPAS approach is potentially compatible with KISS and KIDS. Inherent in this approach is the belief that incremental additions and adaptations to previously validated models provide more capability in explaining complex systems than efforts that begin from scratch. The principle behind the TAPAS approach is captured in the frequently cited words of Newton, "If I have seen further it is by standing on the shoulders of giants" (Newton 1676).

The TAPAS approach lends itself to the validation technique of "docking". Docking involves the conceptual alignment of models and can be used to check if an agent-based model is a special case of another model (Axtell, et al. 1997). Docking probably has the most potential to provide ABM with a standardised validation technique, although it does not replace the need for validation with the external or "real world" environment. By using agent-based modelling platforms such as Swarm, RePast[5], and Mason[6], which include a standardised random number generator, social network module, etc., it is possible to reduce the validation task. The danger with TAPAS models and their facilitation of docking validation techniques is that they reproduce and compound modelling errors. Perhaps even more seriously TAPAS approaches have the tendency to assimilate and replicate existing biases and entrenched but misplaced intuitions – two things that many ABMs seek to challenge. It is incumbent on researchers to ensure that the use of an existing model in the formulation of further models or in the cross-referencing of a model is appropriate and credible according to evidence and fieldwork-based prior knowledge of the study and that it does not compromise the potential for ABMs to provide novel insights. It also seems to exclude participant involvement in the co-construction and conceptual development of models, e.g. as in the companion modelling technique discussed later.

These approaches to model design are not mutually exclusive and can be applied to the same model where appropriate. For example, KISS may be a good place to start thinking about the problem while TAPAS may provide a good means to start building the model. KIDS might be most appropriate for the inclusion of field-data and other empirical information. These modelling principles are useful when conceptualising a theoretical approach, but tell us little about how to proceed. Within the field of agent-based modelling, it would be useful to have practical model design guidelines, which encourage modellers to think carefully about each stage from design through to validation. Understanding the flow between micro level input and macro level output within agent based modelling will facilitate the process of empirical validation.

3.2 Micro-macro validation approaches

5 http://repast.sourceforge.net/

6 http://cs.gmu.edu/~eclab/projects/mason/

Once models have been appropriately designed, it may be possible to validate components of the model that are observable. Validating a component of the model that is well understood adds credibility and reliability even when the entire model can not be definitively validated. Carley (1995) identifies model components that could be singled out for validation. The components of ABM validation are largely divided into two: 1) Macro-validation and 2) Micro validation, due to the key characteristics of the modelling technique, i.e. ABM is developed based on micro-level input to study emergent phenomena at the macro level. The participatory research tools, which are useful for the component validation, are explained at the end of this section.

Macro-level (output) validation

Macro-validation refers to the extent to which modelled results concur with the 'real world'. Observed real world data are generated by unknown processes, real world data-generation processes, (rwDGP), which lead to various stylised facts or statistical properties (Windrum, et al. 2007 S.2.2-3). Modellers are looking for the mechanisms of rwDGP, which are unknown (Figure 1). Modellers try to estimate the rwDGP within a model, which is a process of generating simplified data and abstracting an area formulating an artificial system, the modelled data-generation process (mDGP). Modellers are expecting to see the explanation of stylised facts or statistical properties through the mDGP. For this purpose, modellers have to attempt to match mDGP and rwDGP. In these cases an empirical validation is a backward induction (or abduction) by comparing real world observed data with output data from the model (bottom of Figure 1). This "backward inductive validation" has a pitfall since the duplication of data from mDGP to that of rwDGP is not necessary to prove that the mDGP has the properties of rwDGP. A researcher may get the same output from a model which matches reality by chance though with completely different processes. "Correlation is not proof of causality" needs to be foremost in the minds of researchers. Simply because the observed data fits the construction of a model does not necessarily mean that the model approximates the reality correctly. For example, around the 1929's Wall Street Crash, some observers noted a close correlation between shareprices of New York and London with the levels of solar radiation. They believed that solar radiation or sun sport affected to the stock markets because solar radiation affected agricultural business and the business affected to the global economy(Garcia-Mata and F. I. Shaffner. 1934). However, this correlation is now considered purely by chance (Cass and K. Shell. 1983, Mirowski. 1984).

Moreover, if an agent-based model has a stochastic component[7], obviously the macro-level validation based on the comparison of multiple sets of outputs will be more difficult. Every run of the model, using the same parameters, produces a slightly different output. Even without a stochastic component, macro-level validation can be difficult. Macro-level (or system level) behaviours in ABM reflect iterations of interaction between environmental states and agents, i.e. environmental systems influence the actions of agents, which feedback to the state of the environment (Wooldridge 2002). For this reason, the same model parameters can give rise to a range of macro-level outcomes.

[7] As a black box for unknown, but necessary parameters.

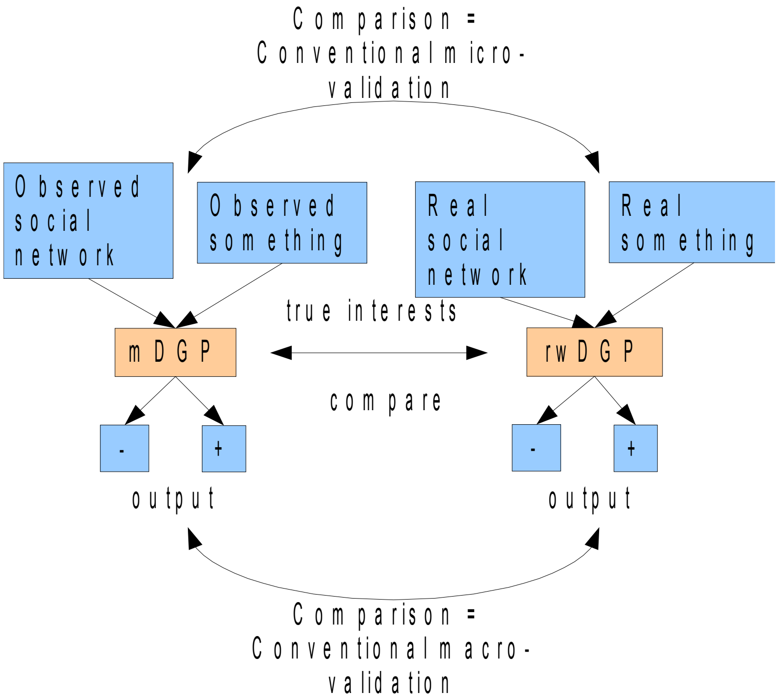

Figure 1:Micro- and Macro- validation. It is difficult to compare processes, i.e. mDGP and rwDGP, but it is possible to compare the inputs and outputs of the two processes. These comparisons are input (micro) validation and output (macro) validation, respectively.

Computer simulation enables the running of models with large parameter spaces and non-linear interaction. However, the testing of these interactions is very difficult. Certainly, checking every parameter of every time-step exhaustively is not feasible in complex ABM. Therefore, some non-linear tests[8] have been developed. These include (Miller. 1998): 1) Multivariate sensitivity analysis, 2) Model breaking and validation, 3) Extreme case scenario discovery, and 4) Policy discovery. In quantitative modelling, calibration is a part of the validation processes as it tries to match an estimate to real world evidence. Also, from a broader viewpoint, calibration and estimation by simulation are not very different since the distinction between the two concepts is vague for the parameters of real world phenomena (Hansen and J. J. Heckman. 1996 p.91). Calibration has a long history and has various tools which enable it to be carried out(e.g. Train 2003).

Output validation, once again, is only as good as the reliability of the system data. The most famous

[8] These tests are not exhaustive, so they are sample-based.

output validation methodology is the “Turing Test” named after mathematician Alan Turing. The test is whether a group of experts are able to tell the difference between data generated by the model and reality. Turing’s original application was to use this as a test of artificial intelligence, i.e. he suggested having a machine and a man both pretending to be a woman - the machine would pass the test if it fooled observers as often as the man. Similarly, if the output of our computational modelling is indistinguishable from real events, a substantial level of validation has been achieved. However, given that most ABM are less concerned with predicting or approximating reality than with generalising about it, the Turing test may have limited applicability in the context of ABM (Edmonds in press). To the extent that it is applicable, the test would question whether the simulation results generated are plausible.

A related form of output validation is what Schreiber (2002) calls "Face Validity" testing, where the model results are presented to persons who are knowledgeable about the source problem, and asking whether this model is reasonably compatible with their knowledge and experience (Sargent. 1987). This is one of the approaches being adopted in the Odra (See the complimentary case validation paper). Certainly, a first step in many models involves establishing that the model results fit with the sensibilities of the substantive expert in the modelling team. Presenting the model at conferences and in publications is another way of getting the kind of feedback needed to appraise the facial validity of the model. Face validity can be applied to input, assumptions or output but is most commonly used in establishing the validity of output. The broad knowledge and experience of the substantive experts serves as the standard against which the model is scrutinised.

“Model – phenomena tests” involve comparing modelled results with real world data. Model-phenomena tests can be used to validate various components of the model but are most commonly used for output validation. For instance, “Historical Data Validity” tests can be used to compare a model's results with the results of data[9] and “Predictive Data Validity” tests compare modelled forecasts with actual outcomes (Sargent. 1987). Another related concept that mixes historical data and prediction is the “Out of Sample Forecast” test which use historical data to tune the model and another portion of data to test the predictive outcomes of the model.

Narratives explaining why the model is appropriate and reasonable and identifying the model’s limitations provide a further powerful way of ensuring that it remains grounded in reality. This is particularly the case for descriptive models in which detail and anecdotes can be checked for resonance (and possibly accuracy) with stakeholders.

“Extreme-Bounds Analysis” or “Extreme Condition Testing” involves the use of very high or very low values for the inputs and/or parameters of the model to test whether the model continues to make sense at the margins (Leamer. 1985). For instance, we should be surprised if a socio-environmental model eliminates all the agents, but trade continues. While a model that generates absurd results for extreme values may not need to be rejected purely on those grounds, researchers should at least bracket any results they claim with a warning about the model failures.

[9] The data need not necessarily be collected beforehand, but doing so ensures that data collection is not biased by the results of a target model.

Micro-level validation

Similar to the macro-validation is the other pair of a backward inductive validation to justify the soundness of mDGP against rwDGP (top of Figure 1). In other words, micro validation is the comparison of real world observed data with output data from the model. This section divides the micro validation into input validation and assumption validation, but these two are highly related.

1) Input validation

With agent based modelling, the researcher explicitly describes the decision processes of simulated stakeholders at the micro level. Structures emerge at the macro level as a result of the attributes given to agents (actions) and their interactions with other agents and their environments. Developing such models requires information about how agents make their decisions, how they forecast future developments, and how they remember the past. What do they perceive, believe and/or ignore? How do agents exchange information? What is the structure of agent interactions (trade, kin, organisation, proximity)?

Input validation requires that the structural conditions, institutional arrangements, and behavioural dispositions incorporated into the model capture the salient aspects of the actual system. It further requires that the data used in models are accurate, and as a result requires credible historical information. The "iterative participatory"[10] modelling approach represents a common form of input validation and involves researchers joining with stakeholders in a repeated looping, for example, through of a four-stage modelling process: field study and data analysis; role-playing games; agent-based model development and implementation; and computational experiments. At the end of this process researchers should be confident that their model assumptions and data sources will yield robust results.

Model inputs can, alternatively, be validated using “degenerative techniques” that involve interrupting or removing certain components of the model and ensuring that the impact on results is consistent with the understanding that informed the model. This approach is similar to “traces testing” that looks at individual agents as they work through the modelling environment (Sargent. 1987).

2) Assumption validation

In some instances validating assumptions can provide an early and powerful means of knowing whether a modelling process is likely to deliver reliable insights. In social science, for example, an assumption that “risk aversion increases with wealth” may assist in validating a model. The breaches of these generally accepted norms are considered grounds for additional scrutiny whereas compliance with them provides a measure of confidence in the model. Scrutinising assumptions can reveal whether the rules that constitute the research model reflect the norms and realities of stakeholder experiences.

The assumption validation is particularly important for ABM. Conventional modelling approaches, e.g. neo-classical economics, tend to place emphasis on analytical tractability over descriptive

[10] The iterative participatory approach has already proven its value in the CAVES study.

accuracy (Friedman. 1953). Therefore, these conventional approaches try to avoid including many variables and parameters. In contrast, the paradigm of ABM is biased towards empirical reality and so it focuses more on descriptive accuracy (Windrum, et al. 2007 S2.3).

Reality of Assumptions (ROAS)[11] is a principle, which asserts that the soundness of model assumptions are as important, if not more important than the forecasting power of a model. ROAS has to be weighted more significantly in ABM as the reason mentioned above. Validation issues at a micro-level challenge our ability to obtain real world evidence to support the assumptions. For example, an agent is assumed to have a concept of goal or desire, etc. However, how realistically can we measure the goals of village farmers in a real village? It is not appropriate to ask stakeholders their goal directly since when a goal is psychological information, it is unconsciously embedded in actors behaviour (Bryman 1988 p.112). Furthermore, if researchers need to formulate realistic assumptions, behavioural data of actors are required. Again, behavioural data are not easily available in environmental science.

In more “formal” sectors, these issues can be set aside. In sectors like stock markets or transportation, the goals of agents can be easily to be assumed with some degree of reliability, i.e. making more money through buying and selling stocks or to get to a destination as quickly and as cheaply as possible, respectively. Moreover, the massive quantity of data on financial stock markets, transportation, and traffic is available to researchers. In contrast, some different approaches are required to validate ABM at the micro-level in more “informal” sectors as the motivation and goals of rural farmers may diverse and complex and there is no such a large database exist. This viewpoint was supported by other ABM researchers who participated in the two workshops held, i.e. Modelling Social Vulnerability in Montpellier on 3-7 April 2006 and CASG Seminar in Oxford on 6 June 2006.

For ABM of “informal” sectors, researchers interview stakeholders and modellers analyse the information to integrate it into the computer model. An obvious means of validation is the feedback processes with the stakeholders, i.e. to give the opportunity to review either the decision rules or the model itself by interview participants or key informants. However, it is likely that the model will be too complex to make interview participant review practical, i.e. 'reactivity' problem. In any form of participant observation, there is interaction between subject and investigator. The research situation changes the behaviour of both subjects and investigator and the outcomes of observation. (Cook 1994 pp. 89-90). Therefore, participatory tools and practical methods are required to ease the reactivity problem.

Besides the participatory tools, which are discussed next, “Model – Model” tests provide a further means of validating assumptions. These tests draw on existing validated models that explore the same issues to evaluate the credibility of newly developed models. “Docking” is a widely applied form of model-model testing and involves ensuring that the model assumptions are comparable with those applied in formally accepted literature (Axtell, et al. 1997).

[11] ROAS is conventionally used in validation, but “possibility” instead of “reality” may be a better term as modelled “reality” is difficult to achieve in the CAVES project, i.e. POAS instead of ROAS. (See Section 3.3 for more on this).

Tools to validate models with community knowledge

Although validation techniques mentioned above workwell, they cover only a half of the validation issues. These techniques are useful only if field research can collect good empirical evidence, i.e. the inputs and outputs of rwDGP. This requires researchers to check the quality of the macro- and micro-level data. Also, the questions the CAVES project are seeking to explore are very site-specific, so that it is most likely that we have to collect the empirical information of a particular research area at the current time, so previously collected data may irrelevant. The techniques and tools to obtain these information are discussed in this section. A further way of overcoming the difficulties of validating model assumptions involves participatory research processes including role playing or scenario testing. These activities involve presenting specific situations to participants, and collecting responses for macro- and micro- validation.

Participatory research has different degrees of engagement with stakeholders, i.e. the scale ranges from eliciting knowledge from stakeholders to the co-management of natural resource issues with co-learning taking place at all the point in between (Lynam, et al. 2007)[12]. In some situations, clear questions can be asked of stakeholders, but in others, it is not possible where questions or information are sensitive or confidential. For example, the Polish and Grampian models, in particular, involve spatially explicit data, which would pose issues for confidentiality and could potentially cause anxiety for participants when the future scenarios based onthe model outputs are presented. Therefore, it is important to choose tools appropriate to the situation. Lynam (2007) provides a good review on selected participatory tools to incorporate community knowledge, preferences, and values into decisions. The selected tools include Bayesian Belief Networks (BBN), Discourse-based valuation, 4Rs framework, Participatory mapping, Pebble Distribution Method , Vision/pathway scenario, Alternative scenario analysis, Spidergram, Venn diagram, and Who Counts Matrix.

Alternatively, participants could be asked to make choices given hypothetical scenarios while playing a 'game' which can help to identify their decision-making processes and heuristics. This process is used when applying Knowledge Elicitation Tools (KnETs) (Bharwani. 2006), to a case, as it filters the responses from such a game through a pattern recognition algorithm to formulate a decision tree, where the order of decisions or factors for consideration in a decision are identified. The South Africa team has some experience working with this tool which also allows for verification and validation of the decision rules with stakeholders in an iterative process. A prototype 'game' using KnETs was also experimented with by the Polish team who have received training on the method. For another type of formalisation and classification of knowledge, the Grampian team has been using an Ontology approach extensively(Gruber. 1993) to contribute to their modelling effort. The South African team also used the Ontology approach during the initial stages of the project to explore attributes of the case study (Polhill, Ziervogel 2006). This method classifies attributes within a domain and then serves as a grammar checker to verify the consistency of the conceptualisaiton that is represented.

The validation of model inputs and output through non-interviewee methods is also possible. For quantitative data, it is fairly easy to validate data through cross referencing with existing research. Utilising available data from other regions to test the model may be possible if input and output factors are similar between the regions. Validating qualitative research can be more problematic. Standard qualitative research validation techniques include the utilisation of multiple data sources (in this case examples include interviewees, key informants, data from other research projects, statistics, and published literature) (Marsland, Wilson et al. 2000). Interviews involving participatory research

[12] Adopting stakeholder problems and addressing these issues during the entire research process.

techniques are site specific. As a result, it is hard to use an interview result as a cross-reference of other interview results (Gonsalves 2005 p.38). It might be agreed that utilising best practice is an academic standard for data validation, and should also be identified as part of the validation technique.

It is clear that communication between modellers, field researchers, and stakeholders is an important part of the validation process. There are several tools for incorporating community knowledge, preferences, and values into decision making. This section further discuss role playing games and companion modelling techniques as they match well with the CAVES project.

1) Role playing games

Role playing games (RPG) are often used in participatory research to elicit local knowledge or to co-learn on certain issues with stakeholders. Therefore, RPG provide a useful approach for validation for the specific purpose of agent-based modelling. A RPG usually involves relationships between the three: the conceptual model, the controlled experiment, and the observed reality in participatory research (Barreteau. 2003a S4.2). For example, experimental economics implements a conceptual model in a controlled experiment in order to understand features of an observed reality. In contrast, policy exercises use background conceptual models in a controlled experiment with stakeholders of an observed reality. These controlled environments are, in a way, pseudo-closed systems, so that the input variables of the RPG will be selected.

When a game is played with stakeholders after modelling, the game will validate the model as stakeholders compare their knowledge and thoughts with the model results. Moreover, during this iterative process, a model design and a game design will be co-evolutionary and improved as the outcome of each activity is fed back into the design of the other. The interaction is not only between the model and the observed reality, but also between modellers, field workers, and stakeholders. The co-evolutionary interaction allows them to learn about the models and the stakeholders' reality. Likewise, the stakeholder may gain an improved clarity and understanding of processes. Some agent-based modelling researchers have formalised this process and refer to it as 'companion modelling' (Barreteau, et al. 2001, Barreteau. 2003b, Barreteau. 2003a).

2) Companion modelling

Companion modelling is an agent based simulation approach that makes opportunistic use of various forms of social simulation including computational simulation and role playing games. In the environmental context, companion modelling has proven particularly successful in generating shared understanding and in strengthening the collective decision making ability of stakeholders (a community of "constructed knowers" as opposed to "silent knowers") sharing a common resource, (See Gurung, et al. 2006) such as in the context of water management in Bhutan. The ability for stakeholders in companion modelling role playing exercises to provide feedback on the game itself provides an in-built validation protocol.

Models are used in a cyclic process composed of three repeated stages (Figure 2): (i) Field studies and bibliography, which supply information and hypotheses for modelling and raise questions to be resolved using the model; (ii) Converting current knowledge into a formal tool (model) to be used as a simulator; (iii) Simulations, conducted according to an experimental protocol (computer model or role playing game), challenge the former understanding of the system and raise new questions for a new batch of field studies. The principle is to integrate various stakeholders' points of view and to develop platforms for collective learning (Gurung, et al. 2006).

The goal is a shared vision (examples of use of the technique are mainly in the area of resource management) that leads to new indicators, shared monitoring procedures, information systems and concrete alternatives for action.

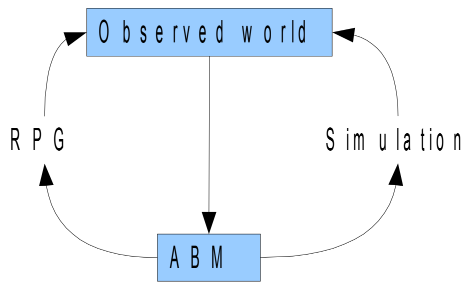

Figure 2: Companion modelling

In the application of a Companion Model in the Bhutan water-sharing simulation the researchers went to some lengths to ensure that the role-playing games and the generated results were credible (Gurung, et al. 2006). One limitation noted by the researchers in this regard was the need to make the game (in which stakeholders deployed their land parcels under various crops in response to water and climate scenarios) "playable". This imposed limitations on the number of players and the number of time-steps that could be introduced. In spite of this, the ability for stakeholders to alter the rules of the game before proceeding to future games ensured that the model resonated with local realities and enhanced the level of confidence that stakeholders and researchers were able to place in the generated results.

3.3 Modelling for a purpose

Contemplating some fundamental questions allows researchers to establish the validation challenge and begin a heuristic process that can serve validation well:

- Does the model tell us what we want to know?
- Is the model plausible given our understanding of the processes?
- Can we understand why the model is doing so well?
- Did we derive a better understanding of our empirical observations?

- Does the behaviour of the model coincide with the understanding of the relevant stakeholders about the system?

Conventionally, validation means "*you built the right thing (model)*"(Chrissis et al. 2003 p563). However, what does "right" mean? In physical science, "right" means to build a model, which matches reality. That is, the validation of physical science is feasible as it may be done in experiments, i.e. in a closed system. In social science, this approach can be problematic. The reality of social science is not in a laboratory so that it is very difficult to observe reality especially for complex studies like the CAVES project. Thus, what we will do in the CAVES project are to: 1) confirm our studies, and 2) validate the models against the 'purpose' of model.

Confirmation

The concept of confirmation was introduced by a paper in Science (Oreskes, et al. 1994). Oreskes writes that the concept of conventional validation against reality is misleading. A model is often misleadingly considered as an accurate representation of physical reality in Earth science. In natural science including policy related science, validation is not possible as its system is never closed and the results of a model can never be the same.

A laboratory experiment controls all input variables, so that this is a closed system. In this case, a unique result can be obtained from a model and so the result is always unique. Therefore, these propositions are only possible to test in a closed system. In an open system, a complete set of input variables are not known. The underlining concept is that the goal of science theories outside of lab based physical science should be adequacy, but not "truth". In this way, researchers need a concept other than verification and validation, namely "confirmation". If a theory or law of science matches with observations, these are *confirmed*. Confirmation is a matter of degree so that these scientific results are more confirmed or less confirmed by multiple results[13]. Therefore, although the validation and verification of an open-system should proceed with caution, the models can be still useful as they can confirm incorrect intuitions and biases.

Confirmation is a potentially good working definition for considering stakeholder validation (Polhill, Small 2007). With stakeholder validation, the 'legitimacy' of the model is established with respect to the expectations of a particular community, somewhat akin to the peer review process in the scientific community. Since the question of internal inconsistency is unlikely to be one that stakeholders can evaluate (in fact, it is possible that we could establish this separately through ontologies), we would essentially be trying to ascertain whether the stakeholders could detect any flaws in the model (Ibid.).

Confrontation between field and modelling processes has to be continuous due to the openness and uncertainty of features of real social systems. For example, companion modelling creates this permanent confrontation which is what makes it a good validation or *confirmation* technique. The cyclical approach between field and modelling processes in companion modelling constantly discusses the assumptions of the model and feeds them back to the field process (Barreteau. 2003b).

Like open systems, ABM also has many input variables because ABM is biased towards empirical reality (See Section 3.2 about "Assumption validation"). As a result of this the model may have many variables and parameters in trying to describe real world. However, this also means that all the

---

[13] This concept of "adequacy" in confirmation is similar to the Grounded Theory, which neither rejects nor accepts a theory. Instead, it says a theory is more true or less true (Strauss 1987).

'possible worlds' also expand hugely and consequently, this makes validation very difficult in ABM. One way to reduce this parameter space is reduce the attributes of the target model to its most salient features and this can be better achieved when the specific *purpose* of the model is identified.

Fit for purpose

In the CAVES meeting, we decided to validate the model against the *purpose of model*, but not 'reality', *per se*. In a way, this is related to the idea of confirmation as it is impossible to validate open systems, such as the case studies in the CAVES project. This approach makes validation exercises in the project feasible. Moreover, this "validation for a specific purpose" is documented in previous studies (e.g. Guideline. 1996, Sargent. 2000, Sargent. 2004) [14].

The principle of "validation for purpose" provides a further means of guiding model design and has elements of validation built into the approach. By identifying the aim and beneficiaries of a model *ex ante* it guides modellers as to when they should cease trying to enhance a model's predictive power or credibility as explained in the previous section on confirmation. If we accept that models of natural systems can never represent definitive "truth", then this approach of modelling to a standard that is adequate for the modellers purpose, is the only approach to validation.

Models for a specific purpose ask "what do we want to use the model for" which sets and limits the validation task against the established benchmark. When modelling complex social systems which can never be definitively understood, this approach is the only intellectually honest validation approach. This approach denotes a reference to an accepted standard whose absolute value can never be fully known. Establishing this benchmark or acceptable standard will depend on what the purpose of the modelling exercise is. Understanding who, and for what purpose a model will be used requires intensive communication between stakeholders and modellers.

This is different from the conventional understanding of validation, which checks if a model is a good representative of the reality; the approach described here gives rise to a new form of validation. For example, through the process, modellers and field workers are forced to understand stakeholders' problems, and stakeholders gain a better understanding of modelling problems, their constraints and potential solutions.

Approaches to 'validation for purpose'

This section explains how this validation for purpose will be used in practice. For example, if the purpose of a model is to understand decision-making processes of stakeholders, a validation for the purpose needs to replace the "real world Data Generation Process" (rwDGP) by the "decision maker's Data Generation Process" (dmDGP) (See page 10). As explained in the section 3.2, usually the true interest of modelling research is about a "process" such as "data generation process" (DGP). A process is a flow, which generates output from some input variables, i.e. input -> DGP -> output. Comparing the outputs of rwDGP and mDGP is the conventional macro-level validation (Figure 3). Comparing observed reality/assumptions and the "real" reality is the conventional micro-level

[14] Also the definition of validation in Wikipedia is: "Validation checks that the product design satisfies or fits the intended usage".

validation. Therefore, validating mDGP is the backward induction (or abduction) from these micro- and macro- validation.

If it is possible to handle these micro-and macro validations for rwDGP, the validation against the reality is feasible. However, as discussed this is likely to be unachievable due to the openness of the reality. In this case, conventional validation against reality will be impossible and the results will not be useful.

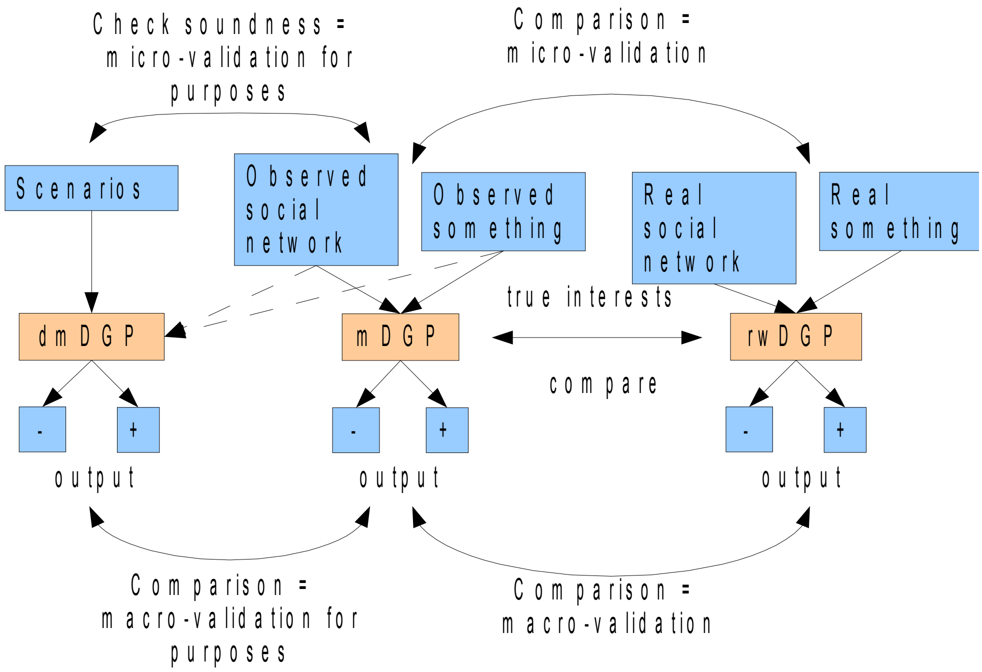

Figure 3: Difference between conventional validation and validation for purpose

Therefore, replacing rwDGP with dmDGP will make validation more achievable and the results of the process useful[15]. The macro level validation becomes the comparison between the output of the model and the answer of decision makers. The output of dmDGP can be observed by a workshop including an expert panel, for example. Instead of estimating real input variables, researchers can show potential and possible input factors as scenarios to stakeholders[16]. Some observed reality can be used for the scenarios, but it is still important to check the soundness of observed realities and the scenarios. The input variables will be used in the dmDGP.

In this example, the purpose of this model is not to display the processes of the real world. This model can be used as a decision-support tool in a different sense. Instead of forecasting future outcomes, the model shows elements and logics of decision-making process, e.g. logics in programming codes, some equations, distributions, etc. Decision making cannot happen with perfect information. Every decision maker makes some assumptions and hypotheses. In this example, the validation process forces decision makers to state the assumptions they make. This necessity of being explicit about assumptions as well as checking them is a benefit as it ensures transparency and robustness in their decision making process.

---

15 In effect, this is the iterative, stakeholder-driven process of knowledge elicitation, verification and validation that the KnETs process involves (Bharwani. 2006).

16 In this case, the scenarios are used as 'what-if' questions to isolate the exercise from the reality. This can be problematic, but reduce the number of possible worlds the model needs to look at.

Sensitivity matrix for validation for purpose

It will be still too costly and time consuming to carry out absolute model validation over the domain for its intended purpose (Sargent. 2004). A validation test needs sufficient accuracy, which can be interpreted as "a model is good enough for a certain purpose." Of course, determining that a model has sufficient accuracy will not guarantee the absolute validity of a model in all possible aspects. However, this "good enough" approach is necessary as the cost of model validation is significant especially when validation requires a high confidence and accuracy (Figure 4). Therefore, this section projects a selection and prioritisation process for validation activities.

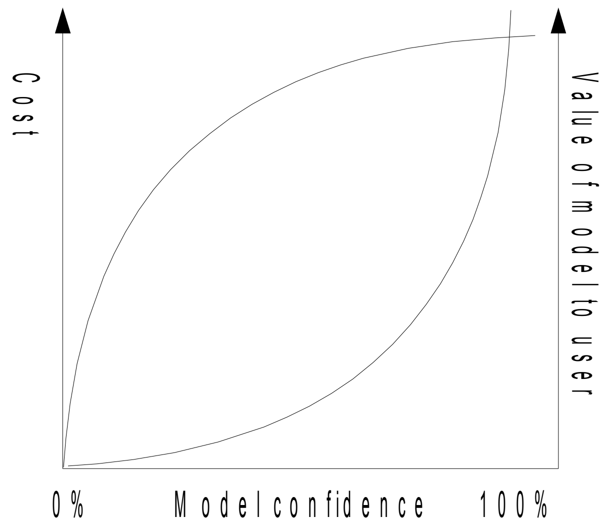

Figure 4: Cost, value, and confidence of model (Source Sargent 2000)

During the CAVES meeting at IIASA in March 2007, guidelines for validation were discussed. We identified four categories of validation activities overall listing all important validation activities to achieve the purpose of modelling: 1) Design (scoping), 2) Process, 3) Techniques, 4) Documentation (output). Table 2 in the Annex shows lists of the validation activities discussed during the IIASA meeting:

- The design category is about model design issues. As explained above, the model design is important to make validation processes easy and to help understand the purpose of the validation. This category has to be defined before models are created and field work carried out. This makes the purpose of modelling clear and, eventually, the scope of validation activities can be narrowed down.

- The process category shows what kinds of validation activities researchers should do once standardised modelling and data collection start. This category is something we need to check during validation to make our activities acceptable by stakeholders.

- The techniques category is like a toolbox of validation activities. So, researchers can find an appropriate tool for a specific activity of validation by looking though the list. If these tools are used appropriately, the validation processes will be made easier and the process will be more transparent.

- The documentation category is a set of documentation issues related to validation. Developing communication protocols can help to build a shared understanding between modellers, fieldworkers, and stakeholders. These four categories are important to get the validation of the CAVES project fully acceptable by stakeholders and modelling team.

For the auditing of the validation processes, this paper proposes a sensitivity matrix, which is adapted from Downing (2004). The sensitivity matrix assesses the importance of validation activities across the range of validation categories. The sensitivity matrix has the potential validation activities for a given model as its rows. The activities are not only strictly conventional validation activities such as “Check if stakeholders' perception are translated into the model correctly”, but also include larger aspects of modelling activities such as “Decide what we are validating for the target audience” and “Make model outputs available to field workers as they need to ask new questions in the field” (See Table 2 in Annex).

The columns of the matrix are the categories of validation. In the CAVES project, the columns are 1) Design (scoping), 2) Process, 3) Techniques, 4) Documentation (output), but each project team has to decide which categories of validation activities or processes are relevant to them. It is likely that some iteration and refinement will be warranted in the matrix. Nonetheless, the categories of columns will overlap somewhat. There are no hard and fast rules for separating certain validation activities. For example, “Reverse design gaming for validation” can be considered as a design issue, process issue or technical issue.

The sensitivity of each validation activity to each category will be identified in this process. The team has to fill the matrix by scoring each cell. A rapid, scoping exercise might use high, medium, or low though if time permits a five-point scale is sufficient for most analyses. Table 1 shows an example of the sensitivity matrix and this will be explained in more detail later.

Three technical issues need to be understood:

- The rating of sensitivities depends on the importance of an activity within a category. In other words, how much a category or process of validation is affected to get adequate confidence in a model if a particular validation activity is not carried out, i.e. Is a model “good enough”? For instance, if a modelling team does not check the quality and amount of data are adequate, this will hugely affect the success of the validation process, and also have some impact on the technique category, but not as much in the design category. Therefore, it is necessary to specify what the consequences or outcomes of omitting the identified validation activities are.

- It is possible to aggregate the ratings, across the rows, down the columns and for the overall matrix. This provides an overall score that may be useful, but should be done only with caution. The results are likely to be sensitive to individual ratings and rankings of validation activities across projects. There are several ways to aggregate ratings. The example below shows a simple normalised sum. However, stakeholders may be concerned primarily with 'hot spots' of significant validation activities. In this case, counting the number of high scores (e.g., those with a 4 or 5) is a better approach than summing all of the values.

- The categories of validation have to be determined before this sensitivity matrix exercise. This can be achieved by listing all possible validation activities first and clustering them as was done at the CAVES Vienna meeting. Alternatively, if there are well known categories in the domain, pre-defined categories can be used, e.g. the TAPAS approach.

The matrix shows which validation activities are most important in order to produce satisfactory results and therefore which categories of validation a project should focus on. In other words, this process identifies high priority validation activities that modellers and field researchers should focus on while also allowing them to re-think any missing validation activities. This will help to achieve a wide-range of validation activities. The result is a scoping of the validation process and this may be necessary as one is unlikely to accomplish all validation activities in a complex project such as CAVES.

Example: Sensitivity matrix for the CAVES project.

This example is a tentative version of a sensitivity matrix for the CAVES project and based on the list and categories of validation activities agreed in the CAVES Vienna meeting. The sensitivity matrix has to be agreed with the project members, later. This example is intended to show the technique rather than the results of this matrix. As an example, this matrix uses only ten validation activities, but it is important to list and carry out the exercise with all potential activities.

The example uses the five-point scale scoring system and shows a simple approach to aggregation. The importance index (right end of Table 1) relates to the overall importance of each validation category to the project. For instance, "List all assumptions in the model" is more important to certify a proper design for validation (with a score of 5) than any other category. The score is calculated as the sum of the columns for each row divided by the total possible score (20), given in percentage. For this example, this is (5+2+3+3)/20*100.

Table 1: An example of sensitivity matrix example for the CAVES project

	Validation categories				Importance	
	Design (scoping)	**Process**	**Techniques**	**Documentation**	Σ	%
Decide on the purpose of the model	5	1	1	1	8	40

List all assumptions in the model	5	2	2	3	12	60
Check if stakeholders' perceptions are translated into the model correctly	2	5	1	1	9	45
Discuss the validation issue between modellers and field workers in person every6 months	1	5	1	3	10	50
RPG as a data gathering and validation tool	1	1	4	1	7	35
Cross-actor confirmation to checkquality of information	1	3	4	1	9	45
Ontology, scripts, and journals for documentation	1	1	3	3	8	40
Reverse design gaming for validation	2	2	2	1	7	35
Document all specifications of scope, procedure, and criterion of validation before validation begins	4	2	1	5	12	60
Make model outputs available to field workers as they need to find out more questions in the field	1	2	2	5	10	50
... others						
Focused category index (Σ)	23	24	21	24		
Focused category index (%)	46	48	42	48		

In this example, "List all assumptions in the model" and "Document all specifications of scope, procedure, and criterion of validation before validation begins" have the highest score. That is, these validation activities have to be high priority in the validation plan. This does not mean that the rest of activities are not important. Although this matrix is just a guideline, researchers are able to select some key activities in the context of a particular project through this exercise.

Similarly, the focused category index (bottom lines of Table 1) is the aggregate score for a specific category (a column) across the validation activities (the rows). All four categories received similar scores; therefore, this shows that the considered validation activities are well distributed across all kind of validation issues. However, the actual distribution of activities is not known unless researchers decide what activities they are going to carry out. Thus, researchers should keep referencing the list to revise the focused category index and to check that a category is not overlooked.

Lastly, researchers using this sensitivity matrix approach have to bear in mind that this subjective scoring system may have some pitfalls (Sargent. 2004). This is a very subjective way to prioritise validation activities. Therefore, it will be advantageous to follow through this exercise with multiple people to check one's justifications and to note the reasons forone's decisions. This clarification of subjectivities will be more useful than hidden the subjectiveness of the method.

In conclusion, a project team needs to construct a first-cut of a sensitivity matrix that can then be used as a guideline of a validation activity check-list, so that researchers, carrying out *validation for purpose*, can aim to carry out as many as possible. Moreover, these lists and matrix have to be reviewed regularly with experts and stakeholders. For example, although the design category will be checked at the beginning of any research activity, this category has to be revised when the modelling purposes are changed. The high priority validation activities have to be carried out to achieve acceptability by stakeholders as a representation of their decision making and by modellers as *fit for the purpose* they intended in the design phase.

4 CONCLUSION

This paper first discusses the difficulties of validation in empirical ABM projects. Then, three possible procedures capable of easing validation difficulties are discussed. Validation is complicated in ABM by the fact that the systems that are generally studied are not closed, as in the case in a laboratory experiment for example. Open systems contain an infinite number of input variables and can never be fully known. It is not possible, even with unlimited resources, to capture all the input variables of an open system in a simulation model.

Therefore, this paper proposed three procedures to address the difficulties. First, good model designs such as KISS, KIDS, and TAPAS make validation activities easier by providing entry and focus points of validation activities. Second, validating a component of the model using macro- and micro-validation approaches, can provide increased confidence in the findings of the ABM as a whole. For these validation activities, some participatory research tools namely role playing games and companion modelling are useful. Third, this paper introduces a new validation idea, i.e. "validation for purpose". This subjective validation approach is based on the idea of confirmation. Validation is considered adequate and useful if the objectives of stakeholders, including modellers, are accomplished by an empirical ABM or if the modelling exercise is useful to the stakeholders. Sensitivity matrices can be useful in identifying priorities and in indicating to modellers the point at which they should be satisfied with the extent of the validation process.

In so doing, this concept paper introduces a shift from conventional validation with its focus on making the model as realistic as possible to a new type of validation which focuses on making the model useful to stakeholders and modellers, where the *use* is pre-defined and the guiding principle for when the model is *complete enough*. The complimentary paper on the practical validation of the case studies in the CAVES project provides further insight into this validation approach. Full implementation of this validation approach may not be achieved in all case studies in the CAVES project, though it will be trialled during the follow-up round of fieldwork in South Africa. Therefore, the next step from this research is to suggest more practical guidelines indicating how to implement the proposed validation approach in future empirical research and attempting to apply it to a real case situation will provide some experience and lessons on how to do this.

5 ANNEX

5.1 The list of validation activities check

These are the list of validation activities identified during the CAVES meeting at IIASA in March 2007. The lists are divided into four categories. However, in reality, it is difficult to categorise these lists as their components are not mutually exclusive. Therefore, it is more appropriate to categorise the in the form of the sensitivity matrix explained in Section 5.2.

Table 2: List of validation activities for purposes

Design (scoping)	Process	Techniques	Documentation (Output)
• Decide what is the purpose of a model • Decide whom we are validating for • Decide what we are validation for the target people • Decide what data a model needs, i.e. about structures, timeliness, rules, properties • List all assumptions in the model • Decide typology of validation, such as, micro/macro,	• At the beginning, modellers and others discuss purpose of models, assumptions, process, etc. • Start with reference model validation • Check if stakeholders' perception adequate to reality if it is possible, i.e. classical validation • Check if stakeholders' perception are translated into a model correctly	• Ontology, scripts, and journals for documentation • Cross-actor confirmation to check quality of information • Extreme condition test for macro (output) validation • RPG as a data gathering and validation tools • Sensitivity tests for macro (output) validation • Reverse design	• Document all specification of scope, procedure, and criterion of validation before validation beguines • Document any subsequent changes • Make model outputs available to field workers as they need to find out more questions in the field • Format model outputs to be readable to non-modellers, e.g. story and visual

socio-economics/biophysical, source of validation materials (Windrum, et al. 2007 S.3.7)	• Check if the quality and amount of data are adequate • Check if model output consistent with theory and other available data • Check if calibration and validation are done by the different people • Discuss the validation issue between modellers and field workers in person every 6 month • Check alternative assumption/hypothesis of model behaviours from stakeholder to find falsification criteria	gaming? for macro (output) validation • Interviews with actors • Laboratory experiments (create closed environment) • Turing test to compare the model results and the your actual behaviours • Mixture of models – cross checking?	formats • Format non-model in clear sentences as a story • Document advantages, disadvantages, and limitation of models • Preserve semantics in the model

5.2 Alternative Protocols in ABM validation

In addition to the methodological approaches to agent based modelvalidation discussed above, a number of more general validation protocols have been developed for use in the social-environmental science nexus. These approaches are based on the understanding that policies for sustainability cannot wait until all the facts are known and an acknowledgement that we must "plan and implement radical changes in technology and lifestyle, in spite of irreducible uncertainty, ignorance, and value-conflicts" (van der Sluijs, J., Risbey et al. 2002)[17].

The protocols have in common an inclusive approach (similar to bottom-up) on the premise that sustainability is a moral issue and so requires a commitment from civil society. Accordingly the relevant knowledge base must be robust in relation to the constraints and demands of this new context of use. It should be designed to fit for its various functions in the discursive, inclusive policy

[17] http://www.nusap.net/

processes on complex issues that are essential for consensus. All stakeholders (including those who produce, use and are affected by policy-relevant knowledge) should be equipped with tools for a critical self-awareness oftheir engagement with that knowledge as a means of creating the “robust knowledge” necessary for sustainability.

NUSAP

This protocol provides one means of conducting multidimensional uncertainty assessments and is particularly apt for application to agent-based models. In its simplest form NUSAP is a quantitative model validation protocol. As an approach, however, NUSAP engenders important principles for qualitative research and allows researchers to move beyond traditional biases of validation particularly quantification biases. NUSAP has been used to include and address the influence of "societal dimensions" such as value-laden assumptions (van der Sluijs, J.P., et al. 2005) that exert systematic biases on modelled processes. NUSAP assumes a heuristic approach capable of fostering a more informed public discourse and negotiated management of complex environmental problems.

The designers of NUSAP, Italian-based Silvio Funtwicz and Jerome Ravetz, present the protocol in response to a generally observed "physics-envy" – the certainty with which physical scientists (as opposed to social scientists) are able to quantify and validate their results. In its application, the NUSAP approach enables the different sorts of uncertainty in quantitative information to be displayed in a standardized and self-explanatory way and it enables the providers and users of such information to be clear about its uncertainties. As such NUSAP also fosters an enhanced appreciation of the issue of quality in information and enables a more effective scrutiny of both quantitative and qualitative information.

NUSAP is an acronym for five categories that describes different aspect of information:

- Numeral - This will usually be an ordinary number; but when appropriate it can be a more general quantity, such as the expression "a million" (which by implication is not the same as the number lying between 999,999 and 1,000,001).

- Unit - This may be of the conventional sort, but may also contain extra information, as the date at which the unit is evaluated (most commonly with money).

- Spread - Which is generalized from the "random error" or variance of experiments or modelled results in statistics. Spread is usually conveyed by a number (either , % or "factor of") it is not an ordinary quantity, for its own inexactness is not of the same sort as that of measurements. The last two parameters in the NUSAP approach are more qualitative.

- Assessment - Provides a place for a concise expression of the salient qualitative judgements about the information. In conventional statistical models this might involve an assessment of the “significance-level”. In the case of numerical estimates for policy purposes, it might be the qualifier "optimistic" or "pessimistic". In some experimental fields, information is given with two terms, of which the first is the spread, or random error, and the second is the "systematic error" which is estimated on the basis of the history of the measurement.

- Pedigree - Pedigree provides an evaluative description of the mode of production (and where relevant, of anticipated use) of the information. Each special sort of information has its own pedigree and it is incumbent on researchers to formulate the distinctions around which a special pedigree is constructed. In the process researchers gain, and disseminate, greater clarity about the characteristic uncertainties of their own field. In NUSAP pedigree is usually expressed by way of a matrix in which the columns represent the various phases of production or use of the information, and the rows represent normatively ranked descriptions.

6 REFERENCES

AXELROD, R., 1997. The Complexity of Cooperation. Princeton University Press. Princeton, NJ: .

Axtell, R., Axelrod, R., Epstein, J. and Cohen, M., 1997. Replication of Agent-Based Models, Aligning Simulation Models: A Case Study and Results. *The Complexity of Cooperation,* , pp. 183-205.

Balci, O., 1994. Validation, verification, and testing techniques throughout the life cycle of a simulation study. *Annals of Operations Research,* **53**(1), pp. 121-173.

Barreteau, O., 2003a. The joint use of role-playing games and models regarding negotiation processes: characterization of associations. *Journal of Artificial Societies and Social Simulation,* **6**(2), pp. 6-2.

Barreteau, O., 2003b. Our companion modelling approach. *Journal of Artificial Societies and Social Simulation,* **6**(2),.

Barreteau, O., Bousquet, F. and Attonaty, J.M., 2001. Role-playing games for opening the black box of multi-agent systems: method and lessons of its application to Senegal River Valley irrigated systems. *Journal of Artificial Societies and Social Simulation,* **4**(2), pp. 5.

BECK, K., 1999. eXtreme Programming eXplained, Embrace Change. Addison Wesley.

BECK, K., 1994. Simple Smalltalk Testing: With Patterns.

Bharwani, S., 2006. Understanding Complex Behavior and Decision Making Using Ethnographic Knowledge Elicitation Tools (KnETs). *Social Science Computer Review,* **24**(1), pp. 78.

BROOKS, F.P., 1995. The mythical man-month. Addison-Wesley.

BRYMAN, A., 1988. Quantity and Quality in Social Research. Routledge.

Carley, K.M., 1995. Computational and mathematical organization theory: Perspective and directions. *Computational & Mathematical Organization Theory,* **1**(1), pp. 39-56.

Cass, D. and Shell, K., 1983. Do Sunspots Matter? *The Journal of Political Economy,* **91**(2), pp. 193-227.

CHRISSIS, M.B., Konrad, M. and Shrum, S., 2003. Cmmi: Guidelines for Process Integration and Product Improvement. Addison-Wesley Professional.

COOK, T.E., 1994. Criteria of Social Scientific Knowledge: Interpretation, Prediction, Praxis. Rowman & Littlefield.

DOWNING, T.E., 2004. Vulnerability in Napa Assessments: Guidance, Examples and Team Exercises For Developing Rapid, Participatory Vulnerability Assessments In National Adaptation Programmes of Action. For the UN Environment Programme UN Institute for Training, Assessment and Research edn. Oxford: .

EDMONDS, B., in press. The Social Embedding of Intelligence-Towards producing a machine that could pass the Turing Test. In: G. PETERS and R. EPSTEIN, eds, *The Turing Test Sourcebook: Philosophical and Methodological Issues in the Quest for the Thinking Computer.* 1st edn. Dordrecht Dordrecht, The Netherlands: Kluwer, .

EDMONDS, B. and MOSS, S.J., 2005. From KISS to KIDS: An ``anti-simplistic'' Modeling Approach. Manchester Metropolitan University Business School.

FAGIOLO, G., WINDRUM, P. and MONETA, A., 2006. Empirical Validation of Agent-Based Models: A Critical Survey.

Frenken,K., 2006. Technological innovation and complexity theory. *Economics of Innovation and New Technology,* **15**(2), pp. 137-155.

Friedman,M., 1953. The Methodology of Positive Economics. *Essays in Positive Economics,* , pp. 3-43.

Garcia-Mata,C. and Shaffner, F.I., 1934. Solar and Economic Relationships: A Preliminary Report. *The Quarterly Journal of Economics,* **49**(1), pp. 1-51.

GILBERT, N., 2004. Open problems in using agent-based models in industrial and labor dynamics. In: R. LEOMBRUNI and M. RICHIARDI, eds, *Industry and Labor Dynamics: the agent-based computational approach.* Sigapore: World Scientific, pp. 401-405.

GILBERT, N. and Troitzsch, K.G., 1999. Simulation for the Social Scientist. Open University Press. Buckingham: .

GONSALVES, J.F., 2005. Participatory research and development for sustainable agriculture and natural resource management: a sourcebook. International Development Research Centre.

Gruber,T.R., 1993. A translation approach to portable ontology specifications. *Knowledge Acquisition,* **5**(2), pp. 199-220.

Guideline,I.C.H.H.T., 1996. Validation of Analytical Procedures: Methodology. *International Conference on Harmonization,* .

Gurung,T.R., Bousquet, F. and Trébuil, G., 2006. Companion Modeling, Conflict Resolution, and Institution Building: Sharing Irrigation Water in the Lingmuteychu Watershed, Bhutan. *Ecology and Society,* **11**(2), pp. 36.

Hansen,L.P. and Heckman, J.J., 1996. The Empirical Foundations of Calibration. *The Journal of Economic Perspectives,* **10**(1), pp. 87-104.

Hendry,D.F. and Richard, J.F., 1983. The Econometric Analysis of Economic Time Series. *International Statistical Review,* **51**(2), pp. 111-148.

Hoffmann,R., Minkin, V.I. and Carpenter, B.K., 1997. Ockham's Razor and Chemistry. *HYLE International Journal for Philosophy of Chemistry,* **3**(3), pp. 3-28.

Janssen,M.A. and Ostrom, E., 2006. Empirically Based, Agent-based models. *Ecology and Society,* **11**(2), pp. 37-49.

Leamer,E.E., 1985. Sensitivity Analyses Would Help. *The American Economic Review,* **75**(3), pp. 308-313.

Leombruni,R. and Richiardi, M., 2005. Why are economists sceptical about agent-based simulations? *Physica A: Statistical Mechanics and its Applications,* **355**(1), pp. 103-109.

Lynam,T., de Jong, W., Sheil, D., Kusumanto, T. and Evans, K., 2007. A Review of Tools for Incorporating Community Knowledge, Preferences, and Values into Decision Making in Natural Resources Management. *Ecology and Society,* **12**(1), pp. 5-19.

MARSLAND, N., WILSON, I., ABEYASEKERA, S. and KLEIH, U., 2000. *A Methodological Framework for Combining Quantitative and Qualitative Survey Methods.*

Miller,J.H., 1998. Active Nonlinear Tests (ANTs) of Complex Simulation Models. *Management Science,* **44**(6), pp. 820-830.

Mirowski,P., 1984. Macroeconomic Instability and the" Natural" Processes in Early Neoclassical Economics. *The Journal of Economic History,* **44**(2), pp. 345-354.

NEWTON, I., 1676. Letter to Robert Hooke, 5 Feb 1676. *Bloomsbury Biographical Dictionary of Quotations (1997).* Bloomsbury, .

Oreskes,N., Shrader-Frechette, K. and Belitz, K., 1994. Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences *Science,* **263**(5147), pp. 641.

POLHILL, J.G. and ZIERVOGEL, G., 2006. Using ontologies with case studies: an end-user perspective on OWL,*National Centre for e-Social Science Annual Conference,* 2006, .

POLHILL, G. and SMALL, L., 2007. Stakeholder validation of FEARLUS. Working paper edn.

Richiardi,M., Leombruni, R., Sonnessa, M. and Saam, N., 2006. A Common Protocol for Agent-Based Social Simulation.*Journal of Artificial Societies and Social Simulation,* **9**(1),.

Sargent,R.G., 2004. Validation and verification of simulation models. *Proceedings of the 36th conference on Winter simulation,* , pp. 17-28.

Sargent,R.G., 2000. Verification, validation and accreditation of simulation models. *Simulation Conference Proceedings, 2000.Winter,* **1**.

Sargent,R.G., 1987. An overview of verification and validation of simulation models. *Proceedings of the 19th conference on Winter simulation,* , pp. 33-39.

SCHREIBER, D., 2002. Validating Agent–Based Models: From Metaphysics to Applications, Prepared for the Midwestern Political Science Association's Annual Conference in Chicago. Working paper edn. UCLA: .

SINGH, L., Drucker, L. and Khan, N., 2004. Advanced Verification Techniques: A Systemc Based Approach for Successful Tapeout. Springer.

STRAUSS, A.L., 1987. Qualitative Analysis for Social Scientists. Cambridge University Press.

TRAIN, K., 2003. Discrete Choice Methods with Simulation. Cambridge University Press. Cambridge: .

VAN DER SLUIJS, J., RISBEY, J., QUINTANA, S.C. and RAVETZ, J., 2002. Uncertainty management in complex models: the NUSAP method, A. JAKEMAN and A. RIZZOLI, eds. In: 24-27 June, 2002 2002, pp13-18.

van der Sluijs, J.P., Craye, M., Funtowicz, S., Kloprogge, P., Ravetz, J. and Risbey, J., 2005. Combining Quantitative and Qualitative Measures of Uncertainty in Model-Based Environmental Assessment: The NUSAP System. *Risk Analysis,* **25**(2), pp. 481-492.

Windrum,P., Fagiolo, G. and Moneta, A., 2007. Empirical Validation of Agent-Based Models: Alternatives and Prospects. *Journal of Artificial Societies and Social Simulation,* **10**(2), pp. 8.

WOOLDRIDGE, M., 2002. An introduction to multiagent systems. John Wiley and Sones. Chichester: .

Zelkowitz,M.V. and Wallace, D., 1997. Experimental validation in software engineering. *Information* *and Software Technology,* **39**(11), pp. 735-743.