

Chapter 10

Calibration and Validation of Agent-Based Models of Land Cover Change

The An Ngo and Linda See

Abstract This chapter considers two important issues in the development of agent-based models, i.e. calibration and validation. These terms are defined and framed into a step-by-step process. Each step is then explained in further detail and illustrated using an agent-based model of shifting cultivation developed by Ngo (2009) as part of his PhD research project. Although the process of model validation presented here is applicable to agent-based models in general, some of the finer details are more relevant to agent-based models of land use and land cover change.

10.1 Introduction

Model validation is a process for determining if a model is able to produce valid and robust results such that they can serve as the basis for decision makers (Berger et al. 2001). The validation process provides the information needed to assess how well the model approximates the real world system and meets the original objectives of the model development. Before the outputs of a model are validated, there is a calibration process whereby the model parameters are determined using real world data. Together both calibration and validation represent one of the seven challenges of agent-based modeling (Crooks et al. 2007). One of the main reason for this

The An Ngo (✉)
Hanoi University of Agriculture, Trau Quy, Gia Lam, Hanoi, Vietnam
e-mail: ntan@hua.edu.vn

L. See
IIASA, Laxenburg, A-2361 Austria
e-mail: see@iiasa.ac.at

challenge is that the concepts related to validation are still being debated, and conflicts remain in the way that validation terminologies are used (Carley 1996; Crooks et al. 2007; Troitzsch 2004). Moreover, the different techniques for validation are quite varied, which has led to a confusing situation for modellers. Therefore, it is important to have a systematic approach to the overall validation process, and one that is integrated throughout the development phase of an agent-based model (ABM). This chapter attempts to provide such an approach to ABM validation.

Numerous publications have been devoted to reviewing different validation methods for ABMs (Berger et al. 2001; Carley 1996; Klügl 2008; Parker et al. 2002; Troitzsch 2004; Windrum et al. 2007). Among these, several types of validation are mentioned, e.g. empirical validation, statistical validation, conceptual validation, internal validation, operational validation, external validation, structural validation and process validation. However, Zeigler (1976) provides a good characterization of these methods into three main types:

- **Replicative validation:** where model outputs are compared to data acquired from the real world;
- **Predictive validation:** where the model is able to predict behaviour that it has not seen before, e.g. that which might come from theories or which might occur in the future; and
- **Structural validation:** where the model not only reproduces the observed system behaviour, but truly reflects the way in which the real system operates to produce this behaviour.

In this chapter, the focus is on structural validation, which in broad terms consists of the following four processes as defined below (Carley 1996; Klügl 2008):

- **Face Validation:** is often applied at the early phase of a simulation study under the umbrella of conceptual validation. This technique consists of at least three methodological elements:
 - **Animation assessment:** involves observations of the animation of the overall simulated system or individual agents and follows their particular behaviours.
 - **Immersive assessment:** monitors the dynamics of a particular agent during the model run.
 - **Output assessment:** establishes that the outputs fall within an acceptable range of real values and that the trends are consistent across the different simulations.
- **Sensitivity Analysis:** assesses the effect of the different parameters and their values on particular behaviours or overall model outputs.
- **Calibration:** is the process of identifying the range of values for the parameters and tuning the model to fit real data. This is conducted by treating the overall model as a black box and using efficient optimisation methods for finding the optimal parameter settings.
- **Output Validation:** involves graphically and statistically matching the model's predictions against a set of real data.

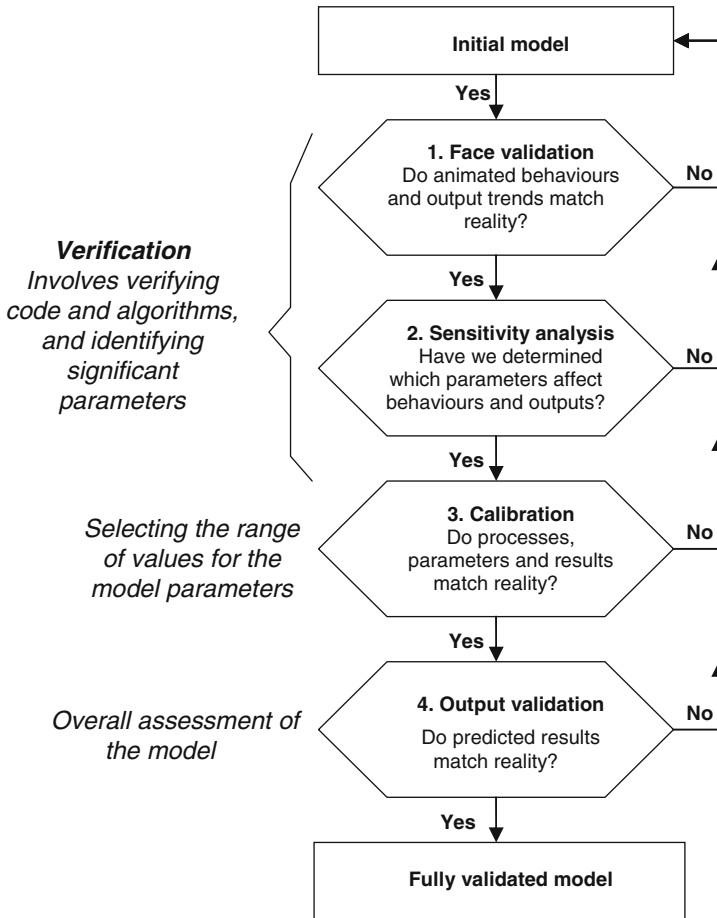


Fig. 10.1 General validation process of an ABM

Face validation and sensitivity analysis are sometimes collectively referred to as verification (Parker et al. 2002). The different processes above are often carried out iteratively in a step-by-step process as illustrated in Fig. 10.1.

A model is able to generate reliable and valid results within its experimental frame only if these validation processes are wholly implemented. However, there are very few examples of where comprehensive system validation has been applied to ABMs. For land use and land cover change modeling in particular, many studies have only concentrated on *output validation* (e.g. Castella and Verburg 2007; Jepsen et al. 2006; Le 2005; Wada et al. 2007) whereas the other steps mentioned above have not been treated explicitly. Therefore, the results may not truly reflect the way the system operates as per the definition of structural validation provided earlier.

The rest of the chapter discusses each of the stages in the validation process (Fig. 10.1) in more detail, providing examples from an ABM of shifting cultivation (SCM) as described in Ngo (2009) and Ngo et al. (2012).

10.2 Verification of ABMs

Verification is the process whereby the logic of the model is tested for acceptability and validity. Basically the model is checked to see if it behaves as it should. Crooks et al. (2007, p. 10) refer to this as testing the “inner validity” of the model. Verification often involves examining processes within the model and then comparing the model outputs graphically or statistically against the real data. However, the level of detail needed for verification is less than that required for calibration (Carley 1996).

As defined previously in Fig. 10.1, model verification consists of face validation together with the sensitivity analysis. Face validation is conducted to ensure that the processes and initial outcomes of the model are reasonable and plausible within the basic theoretical framework of the real system. Sensitivity analysis, on the other hand, is applied to examine the effect of the model parameters on the outcome of the model. Parameters with no significant effect are then removed from the model to make it more coherent and easier to operate. The sensitivity analysis is, therefore, necessary in the pilot phase of complicated simulation studies as the parameters that are identified as being important are those that will require calibration or identification using optimisation or some other means.

10.2.1 Face Validation

Face validation should be applied to several aspects of the model in its early development phase. The dynamic attributes of the agents can be analysed visually across many iterations of the model. All behaviours such as those used for identifying the relationships between agents, and the automatic updating of related parameters are checked for consistency and accuracy. These processes are essentially the *animation* and *immersive* assessments referred to in Sect. 10.1, which can be undertaken in a visual and qualitative way.

A simple example of visual validation is demonstrated in Fig. 10.2, which has been conducted for the SCM of Ngo (2009). Figure 10.2 shows the results of the dynamic monitoring of a random household agent with their relatives over time. As time increases (on an annual time step), the household characteristics of the agent are updated gradually from a state when the household was young to when the first partitioning occurs and the first son marries, forming a new household. Replacement by the second son then takes place when the head of household agent dies to form a new household. Visual analyses like these were used to determine whether the SCM (Ngo 2009) was able to produce acceptable results when simulating real human relationships in a shifting cultivation system.

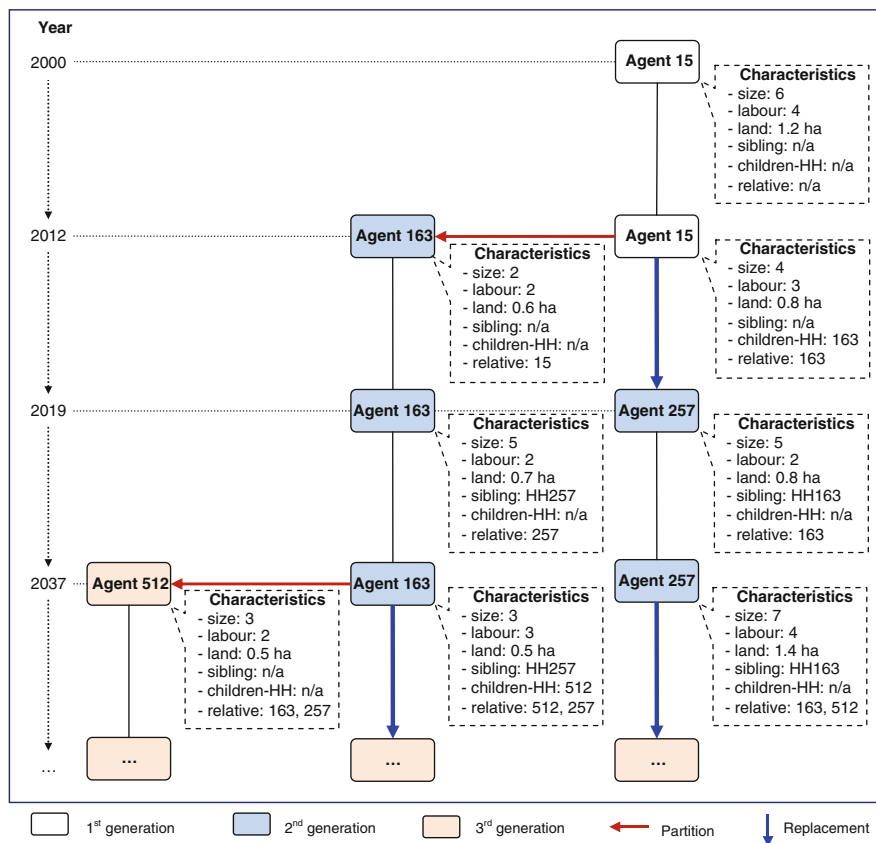


Fig. 10.2 Dynamic monitoring of selected household agents over time in the shifting cultivation model (Ngo 2009)

The second part of the face validation process relates to output assessment in order to ensure that the simulated results fall within an acceptable range of real values across the simulations. The simulated results might include the important parameter values which were used to describe an agent’s characteristics. The analyses can be conducted as follows. Firstly, the model is run several times (where all inputs are held constant) in order to generate the initial outputs related to the characteristics of the agents. The number of runs should be sufficiently large as to be statistically significant (e.g. 30). These data were then analysed visually to ensure that they fell within the range that corresponds to the real world (based on a comparison with survey data obtained from fieldwork).

A statistical comparison between the data from the simulated runs and the real data is shown in Table 10.1. In terms of the statistical distribution, it is important to check for Standard Errors (SE) and compare the mean values of the simulated results with real world values to ensure that the model can provide consistent results. Once

Table 10.1 Household data from the model simulation and the survey data collected in 2007 (Ngo 2009)

	Model outputs in 2007			Survey data in 2007		
	Mean	SE	SE (%)	Lower bound	Upper bound	Mean
Age of household heads (year)	40	0.149	0.37	39	45	42
Household size (# of people)	6.27	0.054	0.89	6.17	7.54	6.86
Household labour (# of labourers)	3.70	0.047	1.27	3.31	4.34	3.83
Land per capita (ha)	0.18	0.000	0.00	0.18	0.21	0.19

the simulated results appear to be consistent (e.g. $SE < 5\%$), their mean values can be then compared with the ranges of the real data, which is often indicated by the lower and upper bounds in statistical terms.

The simulated data in Table 10.1 shows that the model output results have SEs of less than 5% compared to the mean values, indicating that the results are consistent and can therefore be compared with the findings from the survey. The mean values of the model outputs fall within the upper and lower bounds of the survey data, which confirms that the SCM can produce household characteristics that are similar to the survey data.

Another assessment of the output within the face validation framework is to check how consistently the model can produce the same or similar outcomes between the different model runs. There are several ways to do this but the *Test for Homogeneity of Variances* (Winer 1971) is one possible approach. In practice, we might measure the variances of the simulated results for several time steps (i.e. $t, t+1, t+2, t+3, t+n$) with several replications. If the hypothesis is accepted, i.e. the variations between model runs are similar, then the model would pass this test.

Regarding ABMs related to land use and land cover change analysis, it is also important to compare output values from model runs produced at different scales. Since the level of detail is reduced at lower resolutions, there will most likely be some difference between the model outputs run at varying scales. However, if this difference is not statistically significant, then the model could be run at the coarser scale to reduce the running time of the model. This reduction in computational time could be very significant if the model is applied to a large area.

10.2.2 Sensitivity Analysis

In an ABM context, sensitivity analysis is often carried out to examine the effect of input parameters and their values on model behaviours and model outputs. This analysis is essential for selecting significant parameters for the simulation before the model is calibrated or used in scenario analysis. A common approach is to modify only one parameter at a time, leaving the other parameter values constant (Happe 2005). However, this approach is not as easily applicable to agent-based systems (Manson 2002) and sensitivity analysis has often been undertaken in an

unstructured way (Kleijnen et al. 2003). In order to avoid oversimplification of the underlying model due to leaving out possible interactions between input parameters, Kleijnen et al. (2003) and Happe (2005) have suggested that the sensitivity analysis should be conducted systematically by applying the statistical techniques of Design of Experiments (DOE) and metamodelling (Box et al. 1978; Kleijnen and Van Groenendaal 1992).

The suitability of DOE techniques in the context of ABMs has been recognised previously as it can help to determine the importance of the input parameters and also provide information about model behaviour and the logic employed in the programme (Happe 2005; Kleijnen et al. 2003). In DOE terminology, model input parameters are called factors, and model output measures are referred as to responses. A full factorial design consists of i factors, with an assumption that each factor takes j levels and therefore involves $n = i^j$ factor setting combinations. This means that n simulations are required to determine the effect of i factors. However, this procedure can only be applied to a small number of factors because the computation time increases exponentially with each additional factor and each additional factor level (or categories in which each factor is divided). It is obvious that alternative methods are therefore necessary to undertake a sensitivity analysis of the model if the number of factors is large.

To deal with the computational problem due to the large number of factors, Bettonvil and Kleijnen (1997) proposed the Sequential Bifurcation (SB) technique which is essentially a method to determine the most important factors among those that affect the performance of the system.

SB operates with three assumptions: (i) the importance of factors to the model performance can be approximated as a first-order polynomial; (ii) the sign of each factor effect is known; and (iii) errors in the simulation model are assumed to be zero. The overall procedure can be described as follows. Firstly, the analysed parameters are converted to binary variables with values of 0 or 1, which correspond to low and high simulation outputs, respectively. The simplest approximation of the simulation model output y is a first-order polynomial of the standardised variables $(x_1, \dots, x_j, \dots, x_K)$, which has main effects β_j and overall mean β_0 , and can be expressed as:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_j x_j + \dots + \beta_K x_K \quad (10.1)$$

The manner of the variable standardisation mentioned above implies that all the main effects in (10.1) are non-negative: $\beta_j \geq 0$. In terms of DOE, the standardised variables also indicate that the combination of experimental factors relates to the switch-on (1) and switch-off (0) of the equation's elements. To deal with the interaction between factors, i.e. the dependence of a specific factor on the levels of other factors, (10.1) can be approximated as:

$$y = \beta_0 + \sum_{j=1}^K \beta_j x_j + \sum_{j=1}^{K-1} \sum_{j'=j+1}^K \beta_{j,j'} x_j x_{j'} \quad (10.2)$$

where $\beta_{j,j'}$ is the two factor interaction effect between factor j and j' .

Secondly, SB is operated in an iterative procedure where the next factor is selected based on the outputs of previous factor combinations that have already been simulated. The procedure might contain several stages, depending on the lower limit of the effect level defined by the users. The first stage always estimates the simulated results from the two extreme factor combinations, namely y_o (all factors low) and y_K (all factors high). If $y_o < y_K$, then the sum of all the individual main effects is important and the second stage of SB is entered. SB then splits the factors into two subsets of equal size and continues the estimation process for each subgroup, which is the same as that described in the first stage, and the procedure continues in an iterative manner. SB terminates when the effect level (i.e. $y_j - y_o$) reaches the lower effect limit defined by the user.

More detailed instructions on how to apply the SB technique can be found in Bettonvil and Kleijnen (1997) and Ngo (2009). In general, the effective level of the factor found by the SB indicates its sensitivity. The factors that are identified by the SB as having little importance or were less effective should be eliminated from the model. The remaining factors or model parameters will then need to be calibrated if unknown *a priori*. In the SCM of Ngo (2009), sensitivity analysis was used to eliminate a number of variables from the model, leaving a subset for calibration.

10.3 Model Calibration

Once the sensitivity analysis is completed, the next stage in validation (Fig. 10.1) is calibration of the model. The calibration process is conducted to identify suitable values for the model parameters in order to obtain the best fit with the real world. This process, therefore, involves the optimisation of the parameters. There are many different optimisation methods available (Fletcher 2000) but a genetic algorithm (GA) is particularly well suited for implementing this task. A GA has novel properties such as being able to undertake a parallel search through a large solution space (Holland 1992). GAs have also been used to calibrate other ABMs in the past (Heppenstall et al. 2007; Rogers and Tessin 2004).

10.3.1 The Principle of Parameter Optimisation Using GAs

A GA applies the principle of “survival of the fittest” from the field of genetics to a population of competing individuals or solutions within a given environment called the search space (Soman et al. 2008). The procedures involved in a GA are similar to the process that occurs in genetics where the parameters in the GA play the role of chromosomes; the range of data is the genotype; while the results of the model runs are the phenotype. The general steps in a GA are illustrated in Fig. 10.3.

The GA starts with a randomly generated number of solution samples which is collectively called the population, which is the first generation of the species.

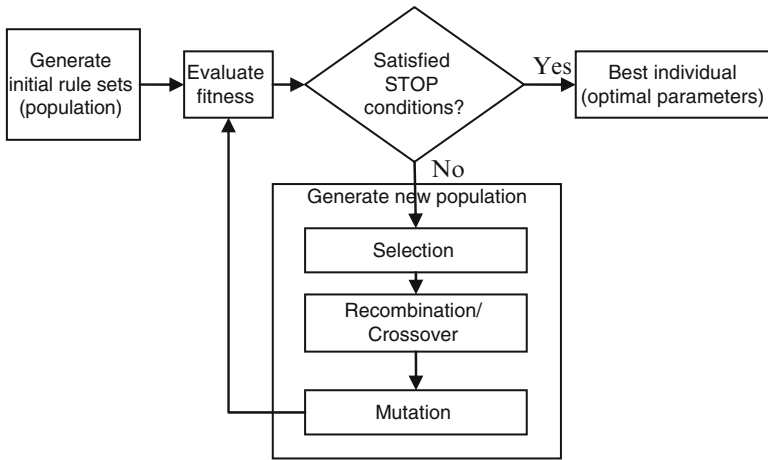


Fig. 10.3 The general steps in a genetic algorithm

A single solution or individual in the population is the combination of parameters with particular values. The solution is therefore equivalent to a natural chromosome with a specific genotype. The next step is the evaluation of fitness using the objective function specified by the user. If there is any individual with a fitness value that satisfies the threshold condition, the programme is terminated and the best individual will be the best solution. Otherwise the GA will operate in a loop creating new generations or populations. Within the loop, individuals (i.e. chromosomes) with higher fitness values are given a higher probability of mating with each other, so as to produce offspring that may better fit the environment.

Several selection methods for selecting the best fit individuals are available such as roulette wheel, tournament, rank and elitism (Mitchell 1996). The most popular method is tournament selection, which is not only suitable for a medium and small population size but also provides marginally better accuracy compared to the roulette wheel selection (Al-Ahmadi et al. 2009). The tournament selection chooses the best fit individuals from several random groups iteratively. For example, if a total of 35 best fit individuals must be selected out of a population of 50 members, the tournament will firstly select a random group (e.g. a group of three random members); within this group, a best fit individual will be the first selected member. These selection processes continue with the next random group to choose the second member until the 35th member is reached. All selected individuals are then entered into the recombination or crossover step which replaces the old chromosomes with the new ones. In the crossover phase, two selected individuals from two random tournament groups perform crossover with a certain number of gene exchanges.

The process of selection and recombination do not inject new genes, so the solution can converge to a local optimum (Soman et al. 2008). The process of mutation, which prevents GAs from premature convergence to a local optimum, is performed to achieve local perturbation by randomly replacing the parameter values with new ones. The frequency of the replacement and the level of perturbation (i.e. the number

of parameter values that are replaced) is defined by the mutation rate. Selection, recombination and mutation are then applied to each generation iteratively until an optimal solution is reached. The condition to be satisfied could be reached after a maximum number of generations or if there is observed stability in statistics such as the mean and/or variance of the population fitness values from a generation (Soman et al. 2008). The optimisation programme ends when the terminated conditions are matched and the optimal solution is reported.

10.3.2 Measurement of the Fitness of a GA

There are several techniques for measuring fitness and errors in the simulation model such as the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), Median Absolute Percentage Error (MdAPE), the Relative Operating Characteristic (ROC), a Confusion Matrix (CM), the Kappa Index of Agreement (KIA), Fractal Analysis (Mandelbrot 1983) and Multiple Resolution Goodness-of-fit (MRG). These techniques and goodness-of-fit statistics measure different aspects of the model performance, and may therefore be suited to different objectives. The selection of which evaluation measures to use depends upon the purpose of the validation and the characteristics of the measures, i.e. what the different measures are intended to show.

With respect to the GA, the RMSE is the most commonly used fitness or error measure (Chatfield 1992) because it indicates the magnitude of error rather than relative error percentages (Armstrong and Collopy 1992). This statistic measures the squared differences between the simulated or predicted values and the observed or reference values:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_{1,i} - x_{2,i})^2}{n}} \quad (10.3)$$

where $x_{1,i} - x_{2,i}$ is the difference between variable i from data source 1 (i.e. the simulated result) and data source 2 (i.e. the reference or observed data); and n is the total number of variables. The RMSE provides a global measure of performance that aggregates all individual differences into a single measure of predictive power.

Other measures of evaluation such as the ROC and the MRG are more suited to evaluation of the model outputs once the model is calibrated so are described in more detail in Sect. 10.4.

10.3.3 Interpreting Calibration Results from the GA

In practice, a GA does not produce a single unique set of parameters but a range of solutions that sit on a Pareto front (Madsen et al. 2002; Yapo et al. 1998). This means that the GA operations will produce a range of different parameter combinations

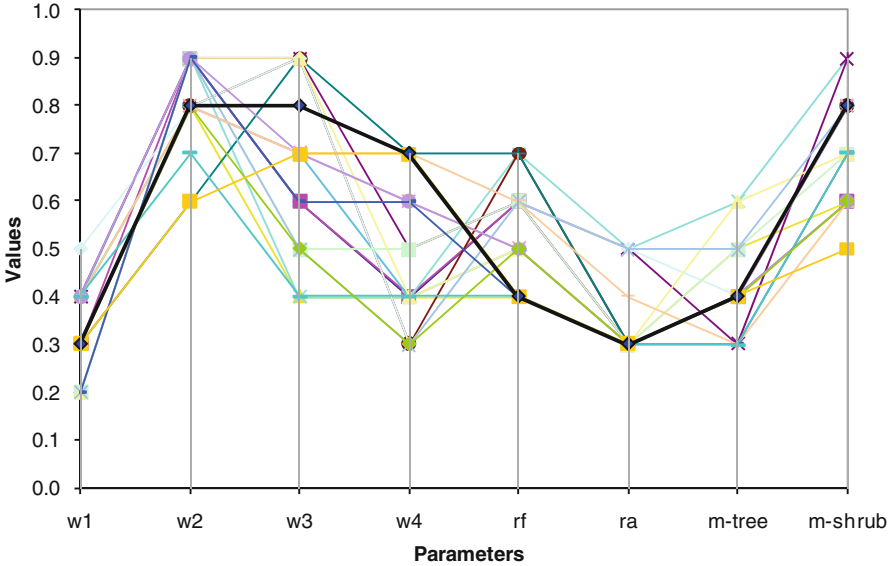


Fig. 10.4 The calibrated parameters provided by 30 GAs (Ngo 2009)

that can give acceptable solutions, rather than generating a single solution. An example of the optimised parameter set from the SCM of Ngo (2009) is shown in Fig. 10.4.

Each line in Fig. 10.4 represents a solution that consists of values for eight calibrated parameters. For each parameter there is a range of possible solutions, indicating the error in the values found by the GAs produced in several runs. Therefore, the additional step that needs to be done is to check for the standard errors for each parameter from all runs. If the errors are not high and the relationships between the values of the parameters and real conditions are reasonable, the solution will be potentially accepted.

As explained above, all parameter combination sets provided by the GA are potentially acceptable solutions. In addition, each parameter clusters around a central value, suggesting that there is a global optimum for the multiple objectives. However, later analyses using the ABM such as validation of model outputs and scenario analyses will require a consistent set of parameters. The way of selecting a set of parameters for further analysis depends strongly on the purpose of the modeller. An acceptable way could be to run the model several times with different parameter sets provided by the GA and then compare the output(s) that is considered as important or significant by the modeller. The parameter set that yielded the highest average fitness value compared to the real data is one method for selection. For example, the bold line in Fig. 10.4 is a parameter set that provided highest fitness values for land cover and was therefore selected as the best solution for the SCM (Ngo 2009).

10.4 Validation of Model Outputs

The final stage in the validation process (Fig. 10.1) is validation of the ABM outputs. This is the most important process in model development because it ensures that the model has the right behaviour for the right reasons (Klügl 2008; Qudrat-Ullah 2005; Troitzsch 2004). Validation of the model outputs is concerned with how well they represent real world behaviour and they are, therefore, compared with actual observations (Parker et al. 2002).

The measurement techniques that determine how the model outputs match the real data are varied. However, the Relative Operating Characteristic (ROC) and the Multiple Resolution Goodness-of-fit (MRG) are two good measures for validating ABM model outputs. These two measures are explained in more detail below.

10.4.1 *Relative Operating Characteristic (ROC)*

The ROC is used to evaluate the performance of a classification or prediction scheme by identifying where instances fall in a certain class or group (Beck and Shultz 1986). The classification is based on the value of a particular variable in which the boundary between classes must be determined by a threshold or cut-off value. An example would be the prediction of illegal cultivation measured by the SCM (Ngo 2009), where the threshold value used to predict whether or not a household would cultivate illegally in the protected forest is a value between 0 and 1. The result is therefore a two-class prediction, labelled either as positive (illegal) (p) or negative (not illegal) (n). There are four possible outcomes from a binary predictor: *true positive*, *false positive*, *true negative* and *false negative*. A *true positive* occurs when both the prediction and the actual value are p ; *false positive* when the prediction is p but the actual value is n ; *true negative* when the predicted value is n and the actual value is also n ; and *false negative* when the predicted value is n while the actual value is p . The four outcomes can be formulated in a two by two confusion matrix or contingency table as shown in Fig. 10.5 (Fawcett 2003). Definitions of precision, accuracy and specificity are also provided.

The ROC evaluation is based on the ROC curve, which is a graphical representation of the relationship between the sensitivity or tp-rate and the specificity or $1 - \text{fp-rate}$ of a test over all possible thresholds (Beck and Shultz 1986). A ROC curve involves plotting the sensitivity on the y-axis and $1 - \text{specificity}$ on the x-axis as shown in Fig. 10.6.

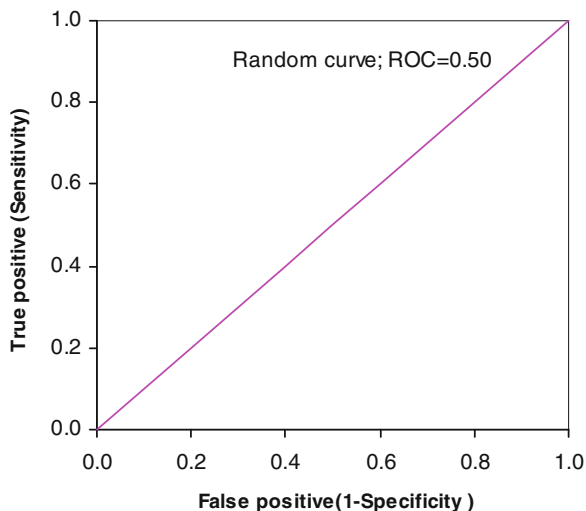
This graphical ROC approach makes it relatively easy to grasp the inter-relationships between the sensitivity and the specificity of a particular measurement. In addition, the area under the ROC curve provides a measure of the ability to correctly classify or predict those households with and without illegal cultivation. The

		True class	
		<i>P</i>	<i>N</i>
Predicted outcome	<i>P</i>	True Positives (<i>TP</i>)	False Positives (<i>FP</i>)
	<i>N</i>	False Negatives (<i>FN</i>)	True Negatives (<i>TN</i>)
Total		<i>P</i>	<i>N</i>

$tp\text{-rate} = TP/N = \text{sensitivity};$
 $fp\text{-rate} = FP/N;$
 $\text{precision} = TP/(TP + FP);$
 $\text{accuracy} = TP+TN/(P+N);$
 $\text{specificity} = TN/(FP+TN)$
 $= 1 - fp\text{-rate}$

Fig. 10.5 The confusion matrix to calculate the ROC (Adapted from Fawcett 2003)

Fig. 10.6 A basic ROC curve (Adapted from Fawcett 2003)



ROC area under the curve (AUC) would reach a value of 1.0 for a perfect test, while the AUC would reduce to 0.5 if a test is no better than random (Fawcett 2003).

The ROC has been proposed as a method for land cover change validation (Pontius and Schneider 2001). However, it is less useful in terms of capturing the spatial arrangement of the model outputs in relation to the real world results (Pontius and Schneider 2001). Thus, in the case of the SCM (Ngo 2009), the ROC is more useful for validating the number of illegal cultivators than the area of illegal cultivation predicted by the SCM (Ngo 2009).

10.4.2 Multiple Resolution Goodness-of-Fit (MRG)

Multiple resolution goodness-of-fit (MRG) has been proposed for measuring the spatial patterns of the model output at several resolutions. This measurement is especially relevant when validating the spatial outputs of ABMs that model land cover and land use change (Turner et al. 1989).

The MRG procedure is expressed in (10.4), which measures the fit at a particular sampling window size (F_w), which is then aggregated for all samples (Costanza 1989):

$$F_w = \frac{\sum_{s=1}^{t_w} \left[1 - \frac{\sum_{i=1}^p |a_{1i} - a_{2i}|}{2w^2} \right]_s}{t_w} \tag{10.4}$$

where F_w is the fit for the sampling window size w , a_{ki} is the number of cells of category i in the image k within the sampling window, p is the number of different categories in the sampling window, s is the sampling window of dimension w by w which moves across the image one cell at a time, and t_w is the total number of sampling windows in the image of window size w .

The fit for each sampling window is calculated as 1 minus the number of cells that would need to change in order that each category has the same number of cells in the sampling window irrespective of where they appear in the image.

The weighted average of all the fits, F_t , over all window sizes is then calculated to determine the overall degree of fit between the two maps as follows:

$$F_t = \frac{\sum_{w=1}^n F_w e^{-k(w-1)}}{\sum_{w=1}^n e^{-k(w-1)}} \tag{10.5}$$

where F_w is defined above in (10.4) and k is a constant. When $k=0$, all window sizes have the same weight while for $k=1$, only the smaller windows sizes are important. For the purpose of matching the spatial pattern of land use, a value of k of 0.1 gives an ‘adequate’ amount of weight to the larger window sizes (Costanza 1989).

The MRG is a much more suitable way of assessing the fitness of the spatial outputs compared to the more conventional methods used in ABM model output validation such as a confusion matrix or kappa statistic calculated at a single resolution only. The Kappa test, for example, can be used to measure the fit between two land cover maps based on a pixel-by-pixel comparison, but it ignores the relationships between one measured pixel and its neighbours. Hence, it will only tell us whether the total number of pixels in each land cover category is significantly different between the two maps, and not say anything about the accuracy of their spatial arrangement (Costanza 1989). The MRG, however, captures the details of the spatial and temporal patterns in the data. More details on the application of MRG can be found in Costanza (1989). The use of the MRG in validating the model outputs of the SCM can be found in Ngo (2009).

10.5 Summary

Calibration and validation are crucial stages in the development of ABMs yet remain a key challenge (Crooks et al. 2007). This chapter has defined these terms and presented the process as a series of steps that should be followed when building a model. Although the process is generic to ABMs in general, particular attention was given to ABMs of land use and land cover change, especially in terms of the measures for evaluating the output of the model. More specifically, examples from the calibration and validation of the SCM of Ngo (2009) were provided to illustrate the process. It should be noted that this represents only one view of the calibration and validation process based on experience gained through building an ABM of shifting cultivation. There are clearly a range of methods available that could be used in or adapted to any part of the calibration and validation process, e.g. different methods of parameter optimization, different measures of evaluating performance, etc. Until more guidance is provided in the literature, calibration and validation will remain a key challenge.

References

- Al-Ahmadi, K., Heppenstall, A., Hogg, J., & See, L. (2009). A Fuzzy Cellular Automata Urban Growth Model (FCAUGM) for the city of Riyadh, Saudi Arabia. Part 1: Model structure and validation. *Applied Spatial Analysis and Policy*, 2, 65–83.
- Armstrong, J. S., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, 8, 69–80.
- Beck, J. R., & Shultz, E. (1986). The use of relative operating characteristic (ROC) curves in test performance evaluation. *Archives of Pathology & Laboratory Medicine*, 110, 13–20.
- Berger, T., Goodchild, M., Janssen, M.A., Manson, S. M., Najlis, R., & Parker, D. C. (2001). Methodological considerations for agent-based modelling of land-use and land-cover change. In D. C. Parker, T. Berger, & S. M. Manson (Eds.), *Agent-based models of land-use and land-cover change. Report and Review of an International Workshop*. Irvine.
- Bettonvil, B., & Kleijnen, J. P. C. (1997). Searching for important factors in simulation models with many factors: Sequential bifurcation. *European Journal of Operational Research*, 96, 180–194.
- Box, G. E. P., Hunter, J. S., & Hunter, W. G. (1978). *Statistics for experimenters: An introduction to design, data analysis, and model building*. New York: Wiley.
- Carley, K. M. (1996). *Validating computational models* (Working Paper). Pittsburgh: Carnegie Mellon University.
- Castella, J.-C., & Verburg, P. H. (2007). Combination of process-oriented and pattern-oriented models of land-use change in a mountain area of Vietnam. *Ecological Modelling*, 202, 410–420.
- Chatfield, C. (1992). A commentary on error measures. *International Journal of Forecasting*, 8, 100–102.
- Costanza, R. (1989). Model goodness of fit: A multiple resolution procedure. *Ecological Modelling*, 47, 199–215.
- Crooks, A. T., Castle, C. J. E., & Batty, M. (2007). *Key challenges in agent-based modelling for geo-spatial simulation* (CASA Working Paper 121). URL: <http://www.casa.ucl.ac.uk/publications/workingPaperDetail.asp?ID=121>

- Fawcett, T. (2003). *ROC graphs: Notes and practical considerations for researchers* (HP Labs Technical Report HPL-2003-4).
- Fletcher, R. (2000). *Practical methods of optimization* (2nd ed.). Chichester: Wiley.
- Happe, K. (2005). *Agent-based modelling and sensitivity analysis by experimental design and metamodeling: An application to modelling regional structural change*. European Association of Agricultural Economists, Copenhagen.
- Heppenstall, A. J., Evans, A. J., & Birkin, M. H. (2007). Genetic algorithm optimisation of an agent-based model for simulating a retail market. *Environment and Planning B: Planning and Design*, 34, 1051–1070.
- Holland, J. H. (1992). *Adaptation in natural and artificial systems*. Cambridge, MA: MIT Press.
- Jepsen, M. R., Leisz, S., Rasmussen, K., Jakobsen, J., Moller-Jensen, L., & Christiansen, L. (2006). Agent-based modelling of shifting cultivation field patterns, Vietnam. *International Journal of Geographical Information Science*, 20, 1067–1085.
- Kleijnen, J. P. C., & Van Groenendaal, W. (1992). *Simulation: A statistical perspective*. Chichester: Wiley.
- Kleijnen, J. P. C., Sanchez, S. M., Lucas, T. W., & Cioppa, T. M. (2003). *A user's guide to the brave new world of designing simulation experiments*. Tilburg: Tilburg University, Center for Economic Research.
- Klügl, F. (2008). A validation methodology for agent-based simulations. In *Proceedings of the 2008 ACM Symposium on Applied Computing – Advances in Computer Simulation*. ACM. New York, NY, USA, 39–43.
- Le, Q. B. (2005). *Multi-agent system for simulation of land-use and land cover change: A theoretical framework and its first implementation for an upland watershed in the central coast of Vietnam*. Göttingen: ZEF.
- Madsen, H., Wilson, G., & Ammentorp, H. C. (2002). Comparison of different automated strategies for calibration of rainfall-runoff models. *Journal of Hydrology*, 261, 48–59.
- Mandelbrot, B. B. (1983). *The fractal geometry of nature*. Oxford: WH Freeman and Co.
- Manson, S. M. (2002). Validation and verification of multi-agent systems. In M. A. Janssen (Ed.), *Complexity and ecosystem management – The theory and practice of multi-agent systems*. Cheltenham: Elgar.
- Mitchell, M. (1996). *An introduction to genetic algorithms*. Cambridge, MA: MIT Press.
- Ngo, T. A. (2009). *Simulating spatial patterns of shifting cultivation – A village case study from the uplands of Vietnam*. Unpublished PhD thesis, School of Geography, University of Leeds, Leeds.
- Ngo, T. A., Drake, F., & See, L. (2012). An agent-based modelling application of shifting cultivation. In A. J. Heppenstall, A. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models for geographical systems* (pp. 611–627). Dordrecht: Springer.
- Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J., & Deadman, P. (2002). Multi-agent systems for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers*, 93, 314–337.
- Pontius, R. G., & Schneider, L. C. (2001). Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems and Environment*, 85, 239–248.
- Qudrat-Ullah, H. (2005). Structure validation of system dynamics and agent-based simulation models. In Y. Merkuryev, R. Zobel, & E. Kerckhoffs (Eds.), *Proceedings of the 19th European Conference on Modelling and Simulation* (pp. 481–485). Riga: ECMS.
- Rogers, A., & von Tessin, P. (2004, May). Multi-objective calibration for agent-based models. In: *Agent-Based Simulation 5*, Lisbon.
- Soman, S., Misgna, G., Kraft, S., Lant, C., & Beaulieu, J. (2008). An agent-based model of multi-functional agricultural landscape using genetic algorithms. In *American Agricultural Economics Association Annual Meeting*, Orlando.
- Troitzsch, K. G. (2004). Validating simulation models. In *Proceedings of the 18th European Simulation Multi-conference*. SCS Europe, Magdeburg.

- Turner, M. G., Costanza, R., & Sklar, F. H. (1989). Methods to evaluate the performance of spatial simulation models. *Ecological Modelling*, 48, 1–18.
- Wada, Y., Rajan, K. S., & Shibasaki, R. (2007). Modelling the spatial distribution of shifting cultivation in Luangprabang, Lao PDR. *Environment and Planning B: Planning and Design*, 34, 261–278.
- Windrum, P. F., Fagiolo, G., & Moneta, A. (2007). Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies and Social Simulation*, 10(2), 8. URL: <http://jasss.soc.surrey.ac.uk/10/2/8.html>
- Winer, B. J. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.
- Yapo, P. O., Gupta, H. V., & Sorooshian, S. (1998). Multi-objective global optimization for hydrologic models. *Journal of Hydrology*, 204, 83–97.
- Zeigler, B. P. (1976). *Theory of modelling and simulation*. New York: Wiley.