19th International Congress on Modelling and Simulation, Perth, Australia, 12–16 December 2011
http://mssanz.org.au/modsim2011

# Testing and evaluating large-scale agricultural simulation models

**Ian R Johnson**

*Melbourne School of Land and Environment, University of Melbourne, Vic 3010*

*Email: ian@imj.com.au*

**Abstract:** Large-scale simulation modelling of agricultural and other biological systems is a young science that has grown with increasing computer power over the last twenty or thirty years. Recently, we have become accustomed to using models as an authoritative source of information for research and policy making in important areas such as climate change or carbon dynamics. In complex systems, models may offer the only way for us to examine the challenges and concerns we have about the future in a time of pressure on resources and uncertainty. The complexity of biological systems makes it unrealistic for either models or experimental observation to address the full scope of underlying processes. By working harmoniously with models and experimental data, there is great potential to increase our understanding of these complex systems and the predictive power of the models. To ensure this is possible, it is the responsibility of modellers to be transparent about their methodology, to present information in a manner that is accessible to review, and for the models to be reproducible.

An essential part of model development is testing and evaluation. This is usually done by comparing the model with experimental data. Agricultural systems are complex and involve the integration of many processes such as: plant photosynthesis, growth and development; soil water and nutrient dynamics; and, for pasture systems, animal intake, metabolism and growth. Often, model testing is restricted to whole-system behaviour. If the model and data agree, then some level of 'validation' is said to have been achieved but if they disagree, then the usual conclusion is that the model is at fault and either the structure requires revision or parameters need adjusting. One difficulty with this process, particularly with complex biophysical models with a large number of parameters, is that similar results can be obtained for different combinations of parameter values, raising the question as to which parameter combinations are appropriate. Indeed, with validation defined this way, models will inevitably be proven to be invalid. Validation as generally applied is more appropriately termed 'verification' and it is well established that no scientific hypothesis, or model, can ever be verified. It is suggested that we focus on the notion of testing and evaluating models, recognising that they are always, at best, an approximation of the real world, just as the factors influencing observational data cannot be completely known. This opens the way to work closely with models and data together to gain greater understanding of the underlying system.

It is virtually impossible to provide a complete mathematical description of large models for assessment in normal journal articles and so peer-review is generally restricted to whole systems behaviour and is unlikely to identify errors and limitations in the underlying model structure. However, peer-review is important and it is suggested that complete and clear descriptions of the models be made available, and that models are, in principle, completely reproducible from the available documentation.

Some examples taken from models of canopy photosynthesis and whole-pasture simulation will be presented to illustrate these points.

***Keywords***: *Model testing and evaluation, biophysical simulation model, pasture model, complex systems.*

## 1. INTRODUCTION

Large-scale simulation modelling of agricultural and other biological systems is a relatively young science that has grown with increasing computer power over the last twenty or thirty years. Recently, we have become accustomed to using models as an authoritative source of information for research and policy making in important areas such as climate change or carbon dynamics. In complex systems, models may offer the only way for us to examine the challenges and concerns we have about the future in a time of pressure on resources and uncertainty. Early agricultural simulation models were used primarily as research tools to examine the interactions of different aspects of the system but, as the models have become more developed, and computing power has allowed greater inclusion of biophysical detail and management strategies, they are being increasingly used in advisory work, policy decision making, and the assessment of the impact of climate variability and potential climate change. The scope of the model applications in this conference is clear evidence of this. The role of models as research tools obviously continues but, as these research models improve, they then can be used with increasing confidence to address more practical problems.

With this increase in use comes the responsibility of careful testing and evaluation of the models. This is often termed 'model validation' and, at least in the agricultural sciences, the usual method is to construct the model, 'calibrate' it on a particular data set, and then compare the model performance with an independent data set (eg Boote *et al*, 1996). As it is generally used, attempts to validate a model are more accurately described as verification and, as has been emphasised by many authors (eg Oreskes *et al.*, 1994; Thornley and Johnson, 2000), this is never possible with complex biophysical systems. I shall address this issue here and suggest that a more flexible approach to model testing and evaluation can be a better strategy for model development.

## 2. COMPARING MODELS WITH DATA

When models are compared with experimental data it is often the case that model parameters are adjusted to improve the 'fit' of the model to the data. However, due to interactions between the structure of the model and its parameters, this has the potential to mislead. In philosophy, the idea of 'validating' a theory is represented using simple concepts, such as Popper's classic example of the statement that 'all swans are white'. This statement seemed irrefutable until 1697, when the black swan was discovered in Australia by Willem de Vlamingh (giving the Swan River its name). This type of simple, real life example demonstrates general principles of logic in philosophy but I am going to consider a practical example from a published model in an attempt to apply this in the field of agricultural modeling.

As an example, consider the simulation of canopy net photosynthesis which lies at the heart of many mechanistic biophysical crop and pasture simulation models. The example here is taken from the canopy photosynthesis model of Johnson *et al.* (2010), and all model simulations were generated using PlantMod (Johnson, 2010). The daily rate of canopy net photosynthesis, $P_{n,day}$, is the difference between the daily rate of canopy gross photosynthesis, $P_{g,day}$, and the daily canopy respiration rate, $R_{day}$, so that

$$P_{n,day} = P_{g,day} - R_{day} \tag{1}$$

where each term has units mol $CO_2$ ($m^{-2}$ ground) $d^{-1}$. Gross photosynthesis is the assimilation of atmospheric $CO_2$ to form sugars through the light driven reactions of photosynthesis, while respiration is the efflux of $CO_2$ associated with the utilisation of sugars to generate energy in the form of ATP (adenosine triphosphate) for plant metabolic processes. The rate of net canopy photosynthesis defines the net carbon assimilation, and therefore growth, of the plant. The instantaneous rate of canopy gross photosynthesis, $P_g$ mol $CO_2$ ($m^{-2}$ ground) $s^{-1}$, is generally calculated by incorporating a description of light interception and attenuation through the canopy and of single leaf gross photosynthesis in response to irradiance, or photosynthetic photon flux (PPF). These are then combined to give the rate of gross photosynthesis which, in turn, is integrated through the day to give $P_{g,day}$.

The basic approach for light attenuation through the canopy defines the PPF, or irradiance, as declining exponentially through the canopy. However, when the sun is not obscured by clouds, there will be direct and diffuse irradiance components within the canopy. These components can be derived from standard theory, which is described in detail in Johnson (2010). Figure 1 shows the PPF as a function of cumulative leaf area index (LAI) through the canopy, where LAI has units ($m^2$ leaf) ($m^{-2}$ ground), along with the direct and

diffuse components. It should be noted that, within the canopy, some leaves will intercept the diffuse irradiance only, while those in direct sunlight will intercept both the diffuse and direct components of irradiance. The decline in irradiance through the canopy is directly related to the canopy extinction coefficient and the value used here (0.5) is typical of cereal or grass canopies. Also, in Fig. 1a, 70% of the irradiance incident on the canopy comes from the direct solar beam, which is typical of clear-sky conditions.

The most widely used equation for the rate of leaf gross photosynthesis is the non-rectangular hyperbola, and this is illustrated in Fig. 1 with parameters that are typical of temperate cereals or grasses growing under well watered conditions with no nutrient stress. It can be seen that the rate of leaf gross photosynthesis increases fairly linearly as irradiance increases and then approaches an asymptote. The parameters in this equation relate to the initial slope of the curve, the asymptote, and the sharpness of the curve. Of these, the initial slope and asymptote are related to temperature and $CO_2$.

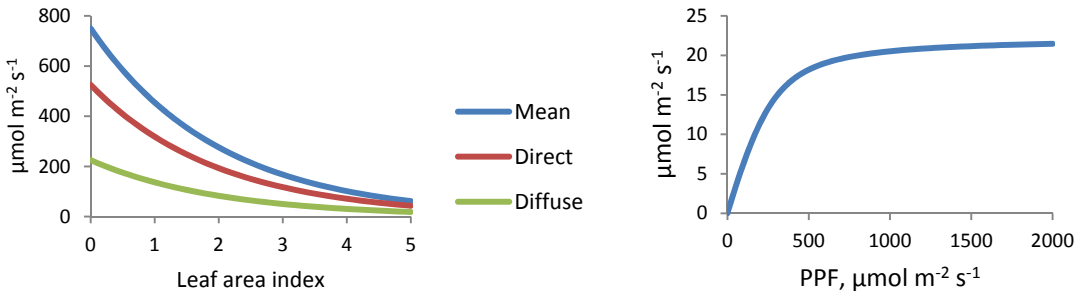


**Figure 1.** Left: irradiance, or photosynthetic photon flux (PPF) as a function of leaf area index through the depth of the canopy: the direct and diffuse components are shown along with the mean value. Right: the rate of leaf gross photosynthesis as a function of PPF.

The leaf photosynthetic potential depends on the concentration of photosynthetic enzymes which is, in turn, related to the protein, or nitrogen, concentration in the leaf. (Note that protein and nitrogen are often used interchangeably in these discussions, with nitrogen being a direct indicator of photosynthetic enzymes.) A widely applied assumption (see Johnson *et al.*, 2010 for references) is that the leaf protein concentration within the canopy is directly proportional to the irradiance so that the protein concentration also declines exponentially through the canopy. This assumption leads to convenient simplifications in the theory. However, it is not consistent with experimental data, as discussed by Kull (2002). When these model components are all combined, the influence of the underlying assumptions can be explored, in particular, the effect of considering direct and diffuse components of irradiance rather than just the mean irradiance, and using the common assumption of exponential decline of photosynthetic enzymes through the canopy or a distribution that is more consistent with observation.

The description of the protein concentration distribution through the canopy has a strong influence on the overall behaviour of this model. Although protein concentration has a direct effect on leaf gross photosynthesis, it also affects maintenance respiration through the respiratory costs of resynthesising degraded proteins. Thus, increasing protein concentration will increase both the flux of $CO_2$ into the canopy, gross photosynthesis, and the maintenance respiratory losses. To explore possible protein distributions, we (Johnson *et al.*, 2010) used a more general distribution function, of which the exponential distribution was a special case, and then derived the parameters that resulted in the maximum rate of shoot growth rate for a prescribed set of environmental conditions.

The light response curves for daily canopy net photosynthesis, $P_{n,day}$, are shown in Fig. 2 with the direct and diffuse components of irradiance included (D+), not included (D-) so that irradiance is assumed to be homogeneous in the horizontal plane, and with the protein distribution defined as an exponential decline (E+) or the more general function (E-). In each case, the optimum protein concentration was derived that maximised the shoot growth rate at mean daily irradiance of 750 µmol $m^{-2}$ $s^{-1}$, day and night temperatures of 22 and 12 °C, ambient $CO_2$ (380 µmol $mol^{-1}$) and a canopy with leaf area index of 3. The corresponding protein distributions are also shown in Fig. 2.

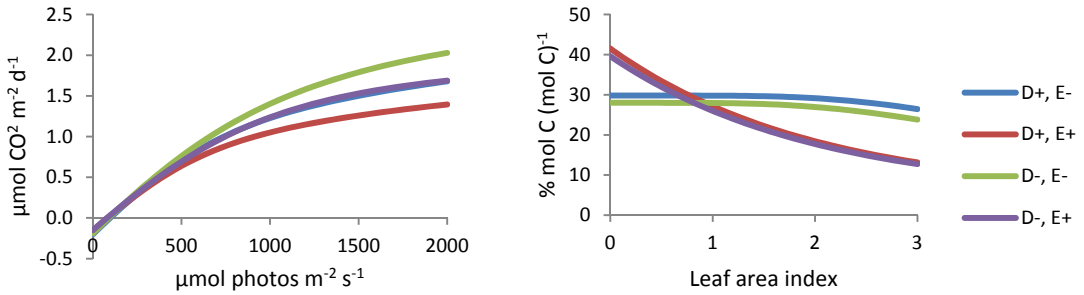


**Figure 2.** Left, daily canopy net photosynthetic rate as a function of PPF. Right, corresponding protein distribution through the canopy. D+ refers to incorporating direct and diffuse components of PPF; D- treats the PPF as homogeneous. E+ uses an exponential protein concentration through the canopy; E- has a more general protein distribution.

These simulations serve to illustrate important points regarding the concept of model validation. It should be noted that none of the parameter values have been selected specifically for these illustrations – they are all the generic defaults used in Johnson *et al.* (2010). First, let us assume that having a general protein distribution (E-) rather than exponential and incorporating direct and diffuse components of irradiance (D+), are more realistic representations of the real system, which is clearly consistent with observation – the sun is often visible in the sky and measured protein, or nitrogen, distributions do not generally follow an exponentially declining distribution through canopies. This means that in Fig. 2 the (D+, E-) curves are the most realistic. However, it can be seen that this curve and the (D-, E+) curve for $P_{n,day}$ are almost identical, so that having an incorrect treatment of the light in the canopy combined with an incorrect protein distribution results in the same response as having these prescribed correctly. If we also had data for canopy photosynthesis, and compared the model output with the data then, based on the principle that a good fit to data validates the model, both sets of assumptions would have resulted in a 'validated' model. Conversely, if we only had measurements of leaf protein concentration distribution through the canopy, and compared the model simulations, and assuming the data were similar to the E- illustrations, it would be likely that either the D+ or D- (include or ignore direct and diffuse irradiance components respectively) would have given a satisfactory fit to the data. (Note that although the E- simulations have a fairly uniform protein distribution, it does tend to decline significantly for greater LAI.) Clearly, getting good agreement between a model and experimental observation cannot be seen to prove that the model has been validated or is correct. Indeed, this raises the very question as to what 'validation' or 'correct' mean in this context.

For more complex large-scale biophysical models, it becomes more difficult to understand the interactions between the different components. A practical mechanistic pasture simulation model, for example, as illustrated in Fig. 3, includes plant growth in response to environmental conditions, soil water dynamics and use by the plants, soil organic and inorganic dynamics, and animal intake and metabolism. These components are interrelated and are also affected by management. For example over-grazing will reduce plant ground cover which in turn affects the transfer of organic matter to the soil and also animal intake. Numerous similar examples can be presented.

How do we gain confidence in models of this type? The relatively simple example with canopy photosynthesis and protein distribution illustrates that models with different assumptions have the potential to give good agreement with observation, depending on what aspects of the system are measured. Obviously, by measuring details of the system, we can hope to improve the reliability of the models. However, since it is inconceivable that we can measure every aspect of the system, and equally inconceivable that we can incorporate every detail in a model with accurately defined parameters, the models can only be an approximation of reality. In this sense, we are working with *open* systems (Oreskes *et al.*, 1994). In that paper they addressed the notions of model verification and validation in a clear and accessible manner, and convincingly argued that the concept of model validation is flawed.

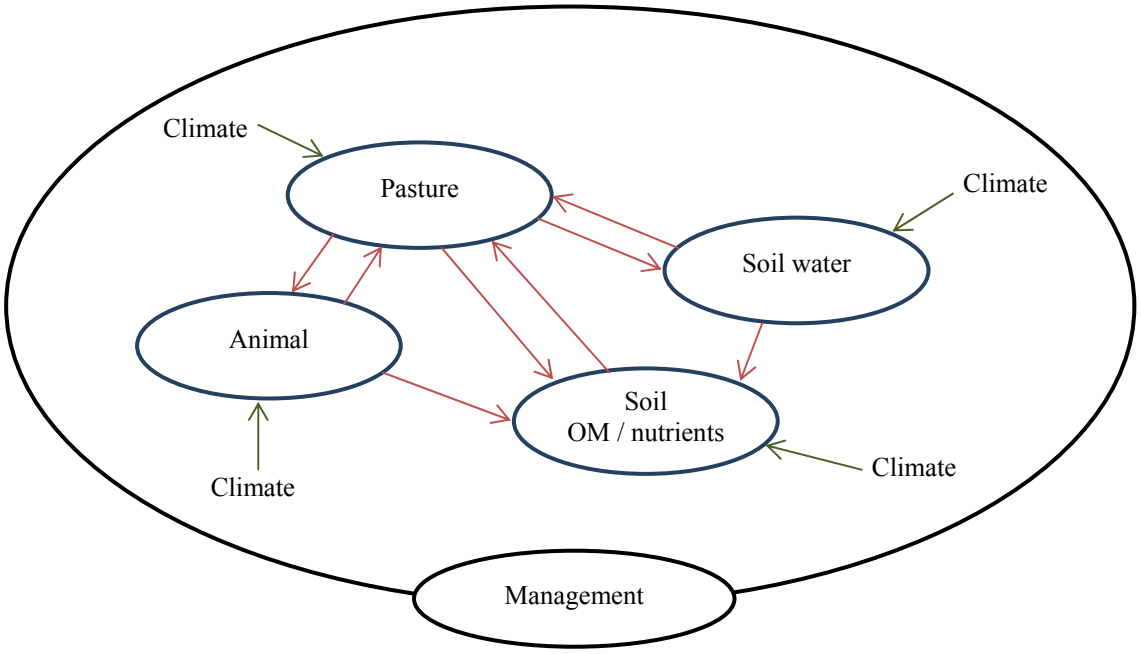


**Figure 3**. Overview of pasture simulation models. The red arrows indicate the complex interactions between components of the system.

## 3. WHAT DOES MODEL VALIDATION MEAN?

The notion that no matter how many times a response is observed there can be no guarantee that it will occur in the future has been referred to as Hume's problem of induction after the 18th century philosopher David Hume. Repeated observations increase our confidence in the likely outcome. For example, we can be confident that pure water at atmospheric pressure will boil at 100°C since it has been demonstrated to be the case many times. While this type of question was occupying philosophers, practical scientists were happily applying scientific principles and theories to great effect. In 1959 in his landmark text *The Logic of Scientific Discovery*, Karl Popper turned the question of validation around completely by saying that a theory could only be considered 'scientific' if attempts could be made to falsify it. So, if a theory can be tested against experimental data then the more those tests demonstrate agreement between theory and observation, the greater the confidence we have in the theory. However, returning to the question of the temperature at which water boils, while repeated measurements will show that it boils at 100°C at atmospheric pressure, it takes only one measurement at 2000 m altitude to show that it does not always boil at 100°C. Based with this extra information we can either abandon the theory as falsified or modify it. Modifying it to include the effect of altitude then leads to greater understanding of the natural world.

When theories and experiment disagree the temptation is to assume that the theory is at fault. However, it is important not to abandon theories lightly as there may be flaws in the data, or the underlying assumptions of the factors influencing the data. One of the most dramatic examples of this concerns the discovery of the planet Neptune. Observations of the orbit of Uranus did not fit precisely with Newton's theory of gravitational motion. Rather than conclude Newtonian theory was wrong, it was recognised that something was missing from the calculations, and that there must be another planet that was influencing the orbit of Uranus, which led to the discovery of Neptune in 1846. This was a fine example of theory and observation working closely together to improve our understanding of our solar system.

Newtonian mechanics worked so well that it was difficult to imagine that they were not a completely accurate representation of the physical world. Then came quantum mechanics and Newtonian theory that had stood unchallenged for over 200 years, broke down. The theory was not abandoned of course, but its limitations and scope of applicability were recognised.

The idea that we can, for all intents and purposes, accept a theory and use it with confidence to broaden our knowledge was discussed by Kuhn (1962) in his seminal text *The Structure of Scientific Revolutions*. Theories that are soundly based on past achievements and that provide a basis for future research were

termed *paradigms* – Newtonian mechanics are one such paradigm. Normal science, as defined by Kuhn, was the application of these theories to address new problems. Eventually, if the underlying theories proved to be inadequate or inaccurate, they would be replaced by a new paradigm.

### 3.1. Model verification

First consider the notion of model verification – to say that a model has been verified is to say that it has been proven to be true. However, this is only possible if we know that we have accounted for all possible aspects of the system being modelled, that is, we can only verify models of closed systems. For example, for a crop model, the system is never closed because we cannot guarantee that we have incorporated all of the influencing factors such as nutrient availability, or pest and disease problems. Furthermore, soil physical properties vary around the paddock and so the parameters required to run the model are incomplete. In the example in the previous section we saw that virtually identical results for the light response of canopy photosynthesis could be generated with different model assumptions – this means that the two model constructions are empirically equivalent. We can only treat the agreement with the data as a guide and continue to delve deeper into the workings of the system to try to identify which set of assumptions is the most appropriate. Even then, we have not reached the 'truth' or have verified the model, but we have gained some understanding and insight into how the observations of the system and our theory of how we think the system works are related.

The one aspect of model verification that can, arguably, be achieved is in the solution of mathematical equations. Once a set of equations has been written down then, if they can be solved analytically involving no numerical approximation techniques, the solution can be said to be verified. However, just because the solution has been verified does not mean that the underlying model has been verified. As an example consider throwing a ball. Applying Newton's laws of motion, and neglecting wind resistance, it can be shown (using high school mathematics and physics) that the trajectory of the ball follows a parabolic path given by

$$y = x\tan(\theta) - \frac{gx^2}{2v^2\cos^2(\theta)} \tag{2}$$

where $(x, y)$ are the horizontal and vertical coordinates, $(v, \theta)$ are the speed and angle the ball is thrown, and $g$ is the acceleration due to gravity. This solution can be verified as being the solution to the underlying equations based on Newton's laws, but it may fail when compared with experimental data since, while the equations are a closed system, the model being used ignores factors such as wind speed and wind resistance.

Although eqn (2) is a verifiable solution to a set of equations, in complex simulation models we cannot solve the equations analytically and must resort to numerical techniques. This means that once the underlying mathematical equations have been written we must be sure that we have solved the equations accurately, that is, the answers we get are actually answers to the equations we have written down. In numerical analysis, this is termed consistency. In this case, solving complex systems of equations using numerical techniques involves approximations, and we cannot be sure that the numerical solution really is the solution to the equations. To address this, we must conduct a series of checks. In some cases, it may be possible to find a special case where the equations can be solved analytically so that we can compare the numerical and analytical solutions. This is often not possible and so we have to rely on other strategies such as checking carbon, water and nutrient mass balances.

In summary, we cannot expect to verify our models of complex systems as being 'true' representations of the system since we cannot include all factors that influence the behaviour of the system and, furthermore, the model will always be characterised by having incomplete input data.

### 3.2. Model validation

Model validation is, in my experience, the most widely used term for testing a model in agricultural simulation modelling. Validation is different to verification in that it denotes legitimacy rather than truth. To quote Oreskes *et al.* (1994) "*... a model that does not contain known or detectable flaws and is internally consistent can be said to be valid.*" Thus, for example, if a set of mathematical equations are written down to describe a complex mathematical system, and are then translated to computer code, it is possible to validate

the code as being an accurate representation of the equations. The output, or results, of solving these equations depends on the validity of the code (as defined here), and also on the quality of the inputs and accuracy of the underlying assumptions.

In mathematics, the term 'valid solution' to equations, or sets of equations, may also be used to impose realistic constraints. For example, with the ball trajectory given by eqn (2), we can also impose conditions of validity on this solution, such as that the values of $x$ are constrained to ensure $y$ must be positive if the ground is horizontal, or $x$ must be positive if the ball is being thrown away from a wall.

The term 'validation' is generally used incorrectly. First, it is often used in the context of comparing model results with observational data, which is actually an attempt at model verification as discussed above and which, for open systems, is not possible. Secondly, to claim a model has been validated is to suggest it is an accurate representation of all of the processes in the system being modelled: as Oreskes *et al.* (1994) say, this is not even a theoretical possibility.

### 3.3. Model calibration

Model calibration is a term that is quite widely used in agricultural simulation modelling. This is the process of adjusting model parameters until the model and data are in good agreement. This process should be treated with caution. Returning to the canopy photosynthesis model above it was seen that different choices of model parameters could give virtually identical results. While this is a relatively simple example, with large scale biophysical models it is often difficult to be sure about possible interactions between model parameters. Any model that has been calibrated in this way and regarded as validated will inevitably fail some future test, either due to inadequate input parameters, limitations in the model, lack of information regarding the conditions under which the data were collected, or possible errors in the actual data being used.

Adjusting model parameters to improve agreement between a model and data set can, however, be a useful process. Suppose, for example, a hydrological model is being applied in a particular situation where measurements of soil water content are available but not the soil physical properties. By adjusting the soil physical parameters to make the model solution for soil water content as close as possible to the data, it may be possible to infer what the soil parameters may be. However, this process must be applied with caution since other aspects of the system may be distorting the results.

## 4. TESTING AND EVALUATING MODELS

If model verification or validation is a fundamentally flawed concept, how do we test and evaluate models? It seems clear that models must be tested in the context within which they have been designed to apply. For example, a site-specific model for irrigation scheduling based on calibrated measurements of soil water content, may be of enormous practical value to a particular farmer, but of no value to a neighbouring farmer growing a different crop on a different soil type. Conversely, if a model is designed, for example, as a mechanistic biophysical wheat simulation model then details of the biophysics should be examined as well as the main final yield – examining the components helps understand possible interactions between the components and the parameters that may be giving spurious results.

My interest lies in deterministic mechanistic pasture simulation models and so I shall take examples from this sphere. In my view these types of models have considerable potential to be applied in regions where specific data may be limited, or not available, although there is almost always some valuable local information. The examples taken here are from the SGS Pasture Model (Johnson *et al*., 2003), which has the same underlying biophysical core as DairyMod (Johnson *et al.*, 2008) with the modules being represented in Fig. 3. The models have been applied to a range of research questions addressing issues such as climate variability, drought, business risk, and the impacts of climate change. In the following examples, I shall not present data comparisons, although some can be found in Johnson *et al.* (2008). Full documentation is given in Johnson (2008).

The following discussion about model structure, testing and evaluating, is largely my own view based on my experience of developing complex mathematical biophysical simulation models in a variety of research programs and in collaboration with scientists from many different fields.

Basic criteria for models are that they should be:

- Modules at similar levels of complexity. When modules are combined, as in Fig. 3, the interacting modules should be consistent with each other.
- Interpretable parameters. Where possible, the model parameters should have a biophysical interpretation so that model users can adjust parameters based on their understanding of the underlying science.
- Robust. They will respond to variable conditions such as grazing intensity or irrigation management.
- Transportable. If a model has a mechanistic response of growth to environmental conditions, then it should be studied with a range of climate inputs applicable to different locations.
- Comparison with data. Model testing obviously involves comparison with available experimental data, but all of the caveats discussed above must apply. These can be summarized by saying that if the model and data agree, then this provides confidence in the model, although getting the 'right' answers for the 'wrong' reasons must always be considered a possibility, whereas if the model and data do not agree, then this provides an opportunity to explore why by scrutinising the model *and* the data.

To illustrate some of these ideas, long-term simulations have been run using the SGS Pasture Model with daily climate data from the SILO database (Jeffrey *et al.*, 2001) for the period 1901 to 2008 inclusive. Locations in northern NSW with native pastures and WA with improved pastures are used as examples.

First consider Glen Innes and Moree in Northern NSW which have average annual rainfalls of 830 and 567 mm $yr^{-1}$ respectively. These two locations, which have winter dominant rainfall that is subject to considerable variability, are at about the same latitude and are 200 km apart. However, as well as having lower rainfall, Moree also is considerably hotter. The model has been run for a generic native $C_3$ / $C_4$ pasture, such as wallaby grass and redgrass, typical of this region (Lodge *et al.*, 2003), and set-stocked with wethers with a normal mature body weight of 60 kg. In practice there will be a mixture of species, and this is discussed later. Soil physical parameters are for a clay-loam. For the simulations, if the animals' weight falls below 35 kg they are supplementary fed until their weight returns to 45 kg. Although, in reality, pastures should be de-stocked if substantial supplementary feed is required, one objective of these simulations is to assess variation in pasture production through supplementary feed requirement. The simulations were run at a range of stocking densities.

The long-term mean values for pasture intake and supplementary feed are shown in Fig. 4. It can be seen that, as stock density increases, average pasture intake increases to a maximum. However, before the maximum is reached the average supplementary feed required starts to increase. Thus, aiming for the highest average pasture intake in the long-term inevitably requires supplementary feed. It should be noted that there is considerable variation in the results throughout the simulation period, and Fig. 4 only shows the long-term averages. Also, note that the regional average stocking densities are around 7 and 2.5 wethers $ha^{-1}$ for Glen Innes and Moree respectively and these are the values where supplementary feed begins to be required. Furthermore, pasture systems with other species such as a winter legume will be more productive and be able to carry greater stock numbers.

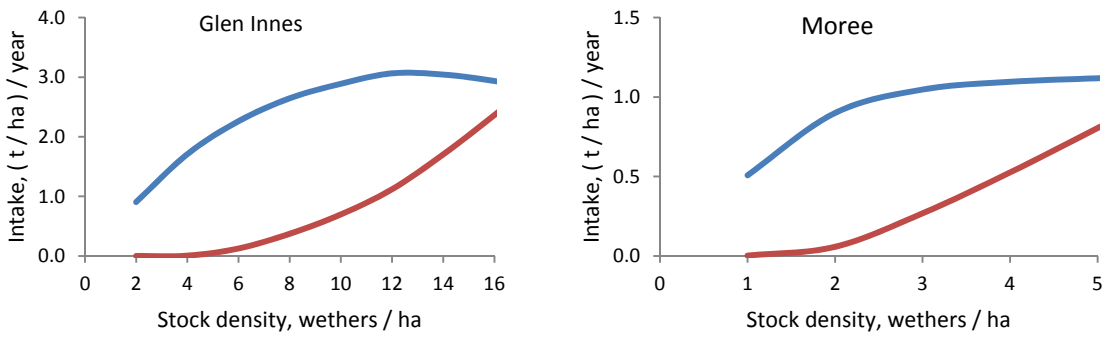


**Figure 4.** Mean annual pasture intake (blue) and supplementary feed (red) for Glen Innes and Moree for a range of stocking densities. Note the different scales on the axes.

I suggest that these simulations, as means of testing and evaluating the model, are complementary to comparison with data. They show that the model is robust since, using the same model parameters, the results reflect the known trends and patterns for two locations which, while fairly close together, have quite different rainfall and temperature characteristics. This gives confidence to use the model to explore

speculative 'what-if' questions. In the simulations, generic $C_3$ and $C_4$ species are used. In practice, there will be a range of species present and no attempt has been made to simulate these individual species. This is an example where it is clearly recognised that the model is an approximation of the real system. Nevertheless, the results are informative.

Now consider a different set of questions for two locations in WA, Albany and Geraldton, which are about 820 km apart. Unlike the previous locations, these are both characterised by strongly winter dominant rainfall, with Albany being cooler and wetter. The average rainfall for Albany (from the SILO climate files for 1901 to 2008) is 900 mm $yr^{-1}$, compared with 460 mm $yr^{-1}$ for Geraldton. For these simulations we shall just look at cutting trials and focus on pasture production. The simulations are for a well fertilized kikuyu pasture on a freely draining sandy-loam soil, with the pasture cut on the last day of each month and the growth rate calculated as the harvested yield divided by the number of days. Obviously there are different ways to calculate growth rate, such as the daily shoot growth minus any senescence, but the cut simulations are closely aligned with what can be readily duplicated in the field. The long-term monthly growth rates are shown in Fig. 5 and it can be seen that they follow a similar pattern with reliable growth in spring. However, growth generally starts and finishes earlier in Geraldton owing to the higher temperature and earlier decline in rainfall. One important characteristic of these simulations is that spring growth is quite reliable due to the winter dominant rainfall that generally ensures that the soil water content is high during spring when temperatures are ideal for growth.

Again, these results are quite plausible and are consistent with expectations in this region. However, no attempt has been made to simulate species persistence and it is likely that kikuyu is less persistent at Geraldton than Albany, and a more persistent species for this location is Rhodes grass. However, apart from persistence, their growth characteristics will be fairly similar and there is value in exploring simulations with the same parameter set at the different locations.

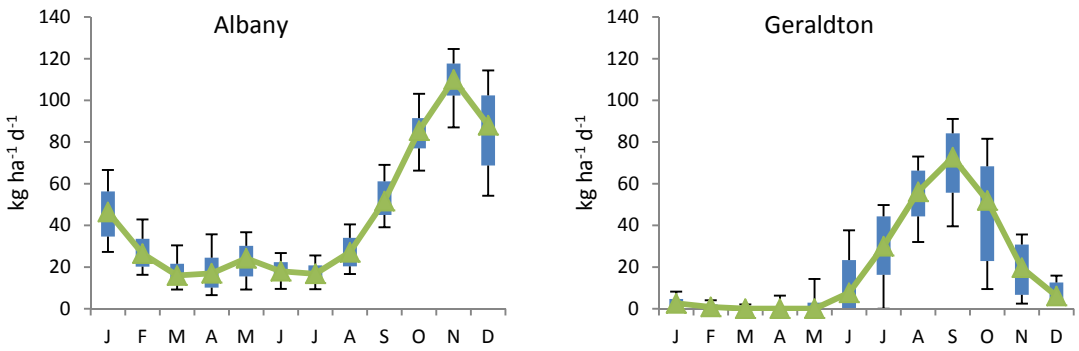


**Figure 5**. Monthly median growth rates and percentiles for Albany and Geraldton. The graphs show the monthly median (green), 25-75 percentiles (boxes), 10-90 percentiles (whiskers).

Having confidence with the model, we can consider some basic trends. To look at the impact of variable rainfall on the system, Fig. 6 shows both the annual yields from the cutting simulations and the deep drainage (2m) are plotted as functions of annual rainfall. For Albany, the results show that there is little correlation between annual yield and rainfall, whereas there is a very close linear relationship between through drainage and rainfall. This suggests that reductions in rainfall will primarily affect deep drainage, which may have important consequences for groundwater flows, but that there is little overall impact on yields. Closer examination of these results (not shown here) indicates that the length of the wet period is an important factor in annual yield. The situation is quite different at Geraldton where a close relation between yield and rainfall can be seen, while there were only 4 years with noticeable through drainage. The contrasting responses for these two sites to variation in rainfall are consistent with our understanding of pasture systems in these regions and again give confidence in the robustness of the model.

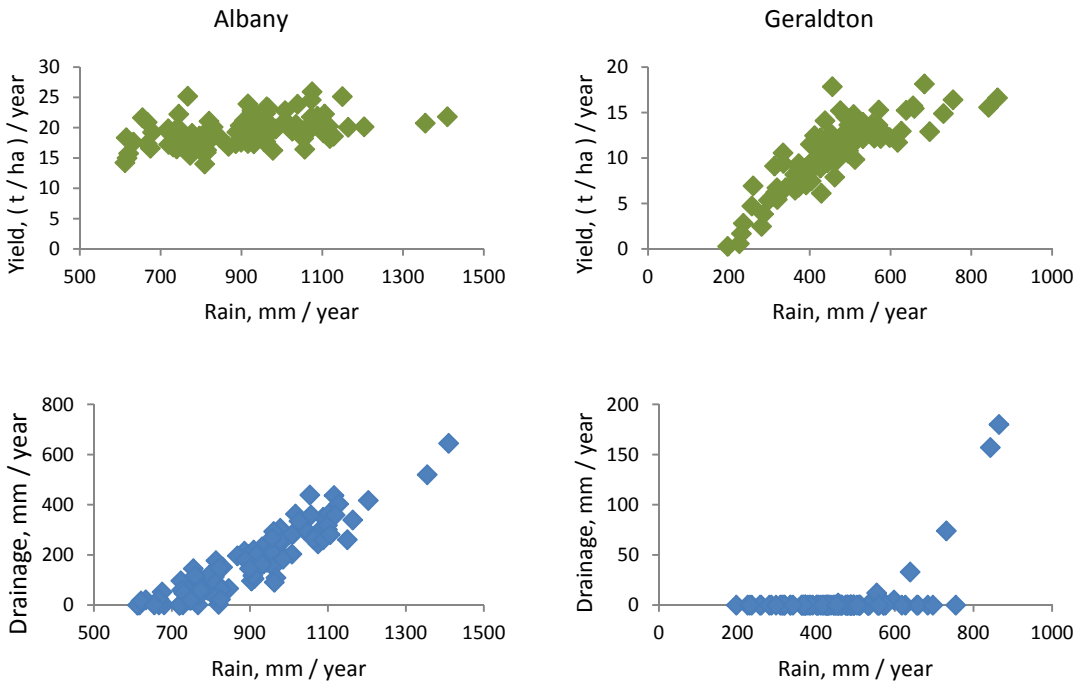


**Figure 6.** Annual yield (top) and through drainage (bottom) for Albany (left) and Geraldton (right) for the cutting trial simulations from 1908 to 2008.

The exploratory simulations show that the same model structure, with different soil characteristics and pasture species parameters, applied at very different locations, gives results that are consistent with expectations. Collecting data like this is impossible since these are 100 year simulations. In these simulations I have referred to generic native $C_3$ and $C_4$ pasture species for northern NSW whereas there will generally be a variety of pasture species. However, by considering a generic $C_3$ / $C_4$ mixture, this captures much of the general pasture species dynamics, while fully recognizing that it is a simplified representation of the actual system. Similarly, for WA a kikuyu sward has been considered. It is likely that Rhodes grass will be more persistent in Geraldton than kikuyu but, apart from persistence, their growth characteristics are quite similar. Again, by keeping the simulations as simple as possible, and acknowledging that the model is not attempting to simulate species persistence, results for the two sites can be easily compared. The overall responses, being consistent with our general understanding of pasture system dynamics at these locations gives us confidence to use the model in more complex analyses.

## 5. MODEL REQUIREMENTS

Full model documentation is usually too large for publication in conventional journal articles. Consequently, model publication is generally restricted to looking at the model behaviour in some context, often by way of comparison with experimental data to 'validate' the model. As discussed earlier, the use of the term validation in this context is more closely verification and, as is the case with all scientific hypotheses and theories, verification is never possible. The examples presented here are methods to test and evaluate models in addition to data comparison. If the model is mechanistic and is aimed to be applicable at a range of locations then it should be thoroughly tested at different locations even if there is an absence of actual field data. Alternatively, if the model requires parameters that must be defined locally, then this should be made clear. For example, growth models that use an empirical growth curve rather than a mechanistic growth model may need location parameters beyond those relating to factors such as soil physical properties. As part of the discussion about how models are published, I suggest the following guidelines.

### 5.1. Documentation and reproducibility

It is unlikely that the normal peer-review process will identify errors and limitations in the underlying model structure but, since peer-review is important, a model that is being presented or used in a research paper

should be required to have a link to the full model documentation, as distinct to a general model overview paper. In essence, this is equivalent to the materials and methods sections of experimental papers that are aimed at making it possible to replicate field experiments, apart from climate and general variability factors. Although the model structure may be documented to a degree in several papers, there is value in having a single source of documentation available. Clear documentation is particularly important for people who are using a complex model but have limited understanding of the detailed model structure.

### 5.2. Transparency and accessibility

The underlying structure of the model should be as transparent to the user as possible and model parameters should have some easily interpreted meaning. For example, when decay rates for soil organic matter pools are presented in units of time, such as $d^{-1}$ or $yr^{-1}$, it is helpful to convert these to half-lives. Thus, a soil organic matter pool decaying at 0.4% $d^{-1}$ has a half-life of 173 days, or just under 6 months, whereas a decay rate of 0.005% $d^{-1}$ has a half-life of 38 years. Difficulties can arise when a model is constructed from a series of regression curves but, in those cases, it should be made clear that that is the case and as much information as possible should be presented relating to the curve structure.

Complex biophysical simulation models comprise several sub-models, as illustrated in Fig. 3. As well as ensuring that users have access to the model parameters, there is value in allowing them to explore the behaviour of the individual sub-models. For example, the screen shot in Fig. 7 is for the growth component of the animal model in the SGS Pasture Model and DairyMod – in this case cattle. Users can see all the model parameters and can explore simple simulations looking at growing or mature animals, weight gain (or loss), body composition and energy dynamics. For example, variations in animal activity costs, due to factors such as hilly or flat terrain, are readily explored. Consequently, even if the users do not have a full working knowledge of the underlying model structure, they can gain insight and understanding regarding the model behaviour and the influence of individual parameters. Obviously there are various ways of giving access to model parameters and this is just one such example.

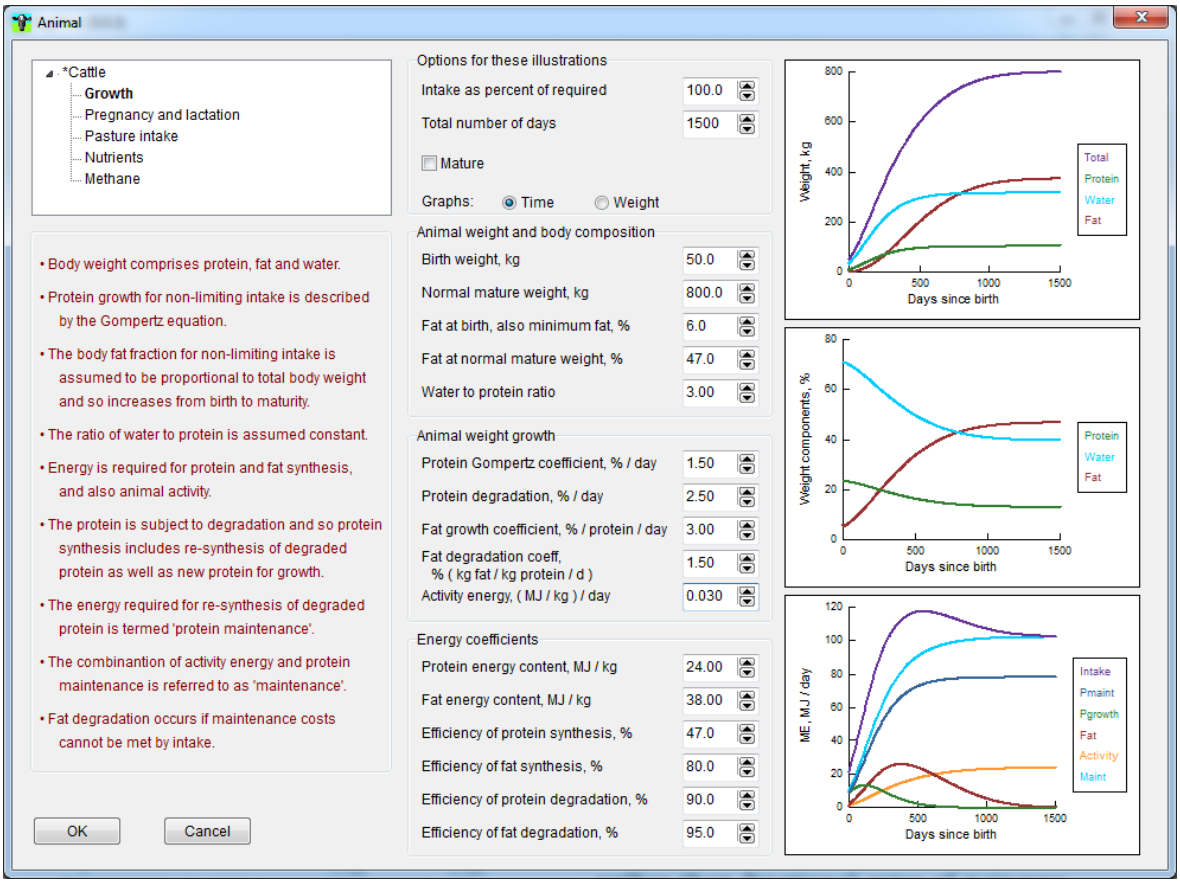


**Figure 7**. Screen shot of part of the animal module in the SGS Pasture Model and DairyMod.

### 5.3. Clarity of output

It is helpful to have as much access as possible to the model output, and not only summarised aspects of the simulations. Crop and pasture production models generally involve carbon and nutrient (eg nitrogen) dynamics. An important check for internal model consistency is to examine the full carbon and nutrient balances to ensure that all inputs, outputs and material retained are correctly accounted for. I have found this to be one of the most valuable debugging tools in developing complex models. By providing users with extensive detailed output they have the opportunity to explore the model behaviour and dynamics in ways that the model developers may not have considered.

## 6. DISCUSSION

I have addressed the notion of 'validating' models and, based on well-established literature in the philosophy of science, suggest that the process of validation that is generally applied is really an attempt to verify the model – that is, to prove it is 'true' (Oreskes *et al.*, 1994; Oreskes, 2003). It is impossible to prove a scientific theory true and, in the case of complex biophysical systems involving many interactions between model components and variation in physical conditions, there will always be factors that are not considered both in the structure of the model and the measurement of the data. Model testing and evaluation is about more than comparison with data, although data comparison has an obvious role to play. It is also about exploring the structure and behaviour of the model, its robustness in response to different factors such as climatic conditions, soil properties, grazing management, irrigation management and so on. This requires an openness to accept model limitations and to continually seek to improve the models, while at the same time acknowledging that the data are also subject to limitations. Suggesting that models have been accurately calibrated and validated may instill a false confidence and expectation that is misleading.

As models are being used increasingly for research and policy making in important areas such as climate change, carbon dynamics and drought assistance, both the performance and structure of the models must be open to scrutiny. Part of this involves clear and transparent testing and evaluation of models in a practical manner, recognising that no model will ever be proven to be a completely accurate representation of the real system. By working hand-in-hand with models and data there is great potential to improve our understanding of these complex systems and so improve our predictive ability.

Models are always an approximation of the real system, incorporating assumptions and simplifications that are part of the development process. By simplifying the system it actually opens the way for creative exploration of the behaviour, just as controlled experimental research provides valuable data for helping to understand factors such as leaf photosynthesis. Just as no model can ever be validated, as a corollary, there is no single right way to construct a biophysical model. Model development requires judgment about complexity and tractability. We should welcome models based on a range of underlying approaches and levels of complexity and use models to their strengths, recognising that one size does not fit all.

## ACKNOWLEDGEMENTS

The pasture simulation modelling examples presented here use the SGS Pasture Model (SGS: Sustainable Grazing Systems) and DairyMod. I am grateful to Meat and Livestock Australia and Dairy Australia for funding to develop these models, and to many colleagues with whom I have collaborated on projects using the models.

## REFERENCES

Boote, K.J., Jones, J.W. and Pickering NB (1996). Potential uses and limitations of crop models. *Agronomy Journal,* **88,** 704-716.

Jeffrey, S.G., Carter, J.O., Moodie, K.B. and Beswick, A.R. (2001) Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environmental Modelling & Software* **16,** 309–330.

Johnson, I.R. (2010). *PlantMod: exploring the physiology of plant canopies.* IMJ Software, Melbourne, Australia, www.imj.com.au/software/plantmod.

Johnson, I.R., Thornley, J.H.M., Frantz, J.M. and Bugbee B (2010). A model of canopy photosynthesis incorporating protein distribution through the canopy and its acclimation to light, temperature and CO2. *Annals of Botany*, **106,** 735-749.

Johnson, I.R. (2008). *Biophysical pasture model documentation: model documentation for DairyMod, EcoMod and the SGS Pasture Model.* IMJ Consultants, Melbourne, Australia. http://www.imj.com.au/consultancy/wfsat/wfsat.

Johnson, I.R., Chapman, D.F., Snow, V.O., Eckard, R.J., Parsons, A.J., Lambert, M.G. and Cullen, B.R. (2008). DairyMod and EcoMod: biophysical pastoral simulation models for Australia and New Zealand. *Australian Journal of Experimental Agriculture,* **48,** 621-631.

Johnson, I.R., Lodge, G.M. and White, R.E. (2003). The Sustainable Grazing Systems Pasture Model: description, philosophy and application to the SGS National Experiment. *Australian Journal of Experimental Agriculture,* **43,** pp 711-728.

Kuhn, T.S. (1962). *The structure of scientific revolutions.* University of Chicago Press, USA.

Kull, O. (2002). Acclimation of photosynthesis in canopies: models and limitations. *Oceologica,* **133,** 267-279.

Lodge, G.M., Murphy, S.R. and Harden S. (2003). Effects of grazing and management on herbage mass, persistence, animal production and soil water content of native pastures. 2. A mixed native pasture, Manilla, North West Slopes, New South Wales. *Australian Journal of Experimental Agriculture*, 43, 891-905.

Oreskes, N., Shrader-Frechette, K. and Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, **263,** 641-646.

Oreskes, N. (2003). *The role of quantitative models in science*. In 'Models in Ecosystem Science', eds Charles D. Canham, Jonathan J. Cole, and William K. Lauenroth (Princeton: Princeton University Press), 13-31.

Popper, K. (1959). *The logic of scientific discovery.* Hutchinson, London

Thornley, J.H.M. and Johnson, I.R. (2000). *Plant and crop modelling.* Blackburn Press, Caldwell, NJ.