Investigation of a complex simulation model of plant growth. I.

Sensitivity analysis reveals simple meteorological trends

Padmaja Ramankutty\textsuperscript{1,2,3,*}, Megan Ryan\textsuperscript{1,2}, Roger Lawes\textsuperscript{4}, Jane Speijers\textsuperscript{3}, Michael Renton\textsuperscript{1,2,4}

\textsuperscript{1}School of Plant Biology and Institute of Agriculture, The University of Western Australia, 35 Stirling Hwy, Crawley WA 6009, Australia
\textsuperscript{2}Future Farm Industries CRC, The University of Western Australia, 35 Stirling Hwy, Crawley WA 6009, Australia
\textsuperscript{3}Department of Agriculture and Food WA, 3 Baron-Hay Court, South Perth WA 6151, Australia
\textsuperscript{4}CSIRO Sustainable Ecosystems, Private Bag 5, Wembley WA 6913, Australia

ABSTRACT

This paper demonstrates methodology for a form of sensitivity analysis used to assess a complex computer simulation model of plant growth in order to determine the relationship between a particular response and its principal drivers. This is the first step in the development of simple statistical models that accurately emulate this outcome of the computer simulation model using only these associated principal drivers as inputs. It also has value in its own right, in that it reveals essential features of how relationships between the response of interest and its principal drivers are modelled in the simulation, which leads to a better conceptual understanding of the simulation model and may also reveal its limitations.

To illustrate this methodology, the Agricultural Production Systems sIMulator (APSIM) was used and relationships between meteorological and soil type input data and biomass production for the perennial pasture species lucerne \textit{(Medicago sativa)} were examined. It was hypothesised that, despite the underlying complexity of the APSIM model, this examination would reveal clear and relatively simple relationships between the principal drivers and the resulting production, which would be suitable for constructing emulators. The examination

\*Corresponding author: Tel: +61-8-6488-2206; fax: +61-8-6488-1108; email: ramanp01@student.uwa.edu.au
found that the relationships between rainfall and biomass produced and between radiation and
biomass produced are sigmoidal. However, the relationships between temperature and
biomass are more complex and are probably best approximated by models based on splines.
Thus the relationships between input data and output data are simple enough to develop
simple statistical models or 'emulators' capable of predicting APSIM-generated aboveground
lucerne biomass production using just meteorological input data. The development and
statistical validation of these emulators are then described in a companion paper.

Keywords: Lucerne, Sensitivity analysis, Perennial pasture, Simulation, Model simplification

1. Introduction

This paper demonstrates methodology for a form of sensitivity analysis used to assess a
complex computer simulation model of plant growth in order to determine the relationship
between a particular response and its principal environmental drivers. This is the first step in
the development of simple statistical models that accurately emulate this outcome of the
computer simulation model using only these associated principal drivers as inputs. The actual
development of this emulator is discussed in part 2 of this series of 2 papers (Ramankutty et
al. in review). The methodology also has value in its own right, in that it reveals the essential
features of how the relationships between the response of interest and its principal drivers are
modelled in the simulation, which leads to a better conceptual understanding of the simulation
model and may also reveal its limitations.

Computer simulation models (also known as mechanistic or complex models) which
represent plant growth have been developed to assist in understanding and managing bio-
physical processes in nature. These models cover a large diversity of agricultural and
ecological activity. There are models for animal production such as DairyMod (Johnson et al.
2008), generic crop models such as SUCROS (van Ittersum et al. 2003), models for growth of parasitic plants (Hautier et al. 2010), models for vegetation growth (Brolsma et al. 2010), and models for predicting the effect of cropping on field margin flora (Fargue-Lelièvre et al. 2011). Such mechanistic models use computer simulation to represent many of the physical and biological processes involved in plant growth, thereby increasing the clarity and understanding of how these processes and their underlying mechanisms are involved in predicting plant development. This is the strength of complex plant growth models: by taking a mechanistic approach, these physical and biological processes are represented at a relatively high level of detail and realism.

However, there are disadvantages to using such a mechanistic and complex approach. Their detail can make them inaccessible to people who have limited understanding of biological processes. Yield prediction using these types of models requires extensive inputs such as detailed quantitative information, first on plant physiology and phenology for each new species, and then on soil characteristics and meteorological data for each new location. Furthermore, such models may consist of several connected modules which are often developed by different people with particular areas of expertise. As such, the models can lack transparency, which can make it difficult for even experienced agronomists to modify them for a new species or a new environment without specialised training. Modification can be especially challenging in situations where not all of the requisite information is available or where significant structural changes are required because of a change in understanding of the driving processes (Holzworth et al. 2008).

Simpler models can circumvent the difficulties associated with detailed mechanistic growth models. For example, French & Schultz (1984) developed a very simple model that predicts potential wheat yield based solely on rainfall. This model accounts for about 70% of total variability in yield. The Pycal model (Tennant 1996) is basically the French & Schultz
model modified to include stored water at the start of the season. The Oliver model (Oliver et al. 2009) is a modification to the French & Schultz model that includes an estimate of the amount of water lost from the soil through evapotranspiration and also tracks stored soil water levels throughout the season. The STIN model (Stephens et al. 1989) uses a simple regression on a water stress index to predict actual plant production. The success and transparency of these types of models is due in part to the fact that they use only 1 or 2 factors to predict the outcome. However, this then limits their accuracy of prediction, since there are likely to be several additional factors significantly influencing plant production, and these factors are also likely to have significant interactions.

Sensitivity analyses involving complex simulation models are not new to the world of agriculture or ecology (Cariboni et al. 2007; Jiménez-Martínez et al. 2010; Adam et al. 2011). Often such analyses are used in the development of the simulation model to identify which parameters are most important in affecting the model’s predictions (Saltelli et al. 2008). However, here we use a form of detailed sensitivity analysis for a different reason: to clearly identify which driving environmental factors would be the most important to include in a simplified growth model, what functional forms could be used to model these factors, and how important it would be to include interactions between these factors in such a model. Conducting such a sensitivity analysis would provide direct clarification and increased transparency of the *modus operandi* of the complex plant growth model, thus helping a wider audience with less specialised knowledge to better understand the overall outcomes of the ways in which the underlying processes have been modelled. It would also facilitate the development of the previously-mentioned simplified growth model, which would then be based only on the principal drivers and their interactions as identified by the sensitivity analysis (Brooks et al. 2001).
This approach has many other advantages. It allows for examination of the internal workings of the complex plant growth model without having to acquire in-depth knowledge of the development, testing and application of this model. Thus the simulation model is treated as a ‘black box’ for the purposes of this analysis, and the focus is on determining the relationships between the principal drivers and the response of interest. Time is not spent revisiting the computer code which is assumed to have already been developed and extensively validated by experts. Another advantage of this approach is that it is very simple and straightforward and can be easily generalised to apply to any computer simulation model to assess the relationships between any of its outcomes and the associated principal drivers.

To demonstrate this methodology, we chose from the many available options to assess perennial pasture production using the Agricultural Production Systems sIMulator, APSIM, a mechanistic model of plant growth particularly suited for Australian climatic and environmental conditions (McCown et al. 1996; Keating et al. 2003). APSIM is a complex model consisting of many interlinked modules (Keating et al. 2003). It is very flexible since it contains generic modules for processes such as plant yield, which can be extended to represent new species or even individual varieties/cultivars. Several modules can be linked together to create a model of a single paddock or a whole farm and these models can be modified by including or excluding modules as required (McCown et al. 1996). Therefore, APSIM is a valuable model in that it combines a vast amount of existing knowledge of biological processes into a single model which can be used to generate simulations of various scenarios, thus limiting the need for extensive and expensive experimentation. This is the first time that such a sensitivity analysis has been reported for APSIM.

For the current study, to simplify our sensitivity analysis of APSIM and also provide greater clarity of the methods used in the analysis, we chose to examine a single plant species. While APSIM was primarily developed as a cropping systems model (Wang et al. 2002), its
developers have recently included modules for simulation of the perennial pasture species lucerne (*Medicago sativa*) (Robertson et al. 2002), bambatsi panic (*Panicum coloratum*) ([http://www.apsim.info](http://www.apsim.info)) and Rhodes grass (*Chloris gayana*) (Lawes & Robertson 2008). Perennial pastures have increased in prominence in the cropping regions of Australia over the last 10 years in a partial attempt to deal with dryland salinity and climate variability (Cocks 2001; Dear & Ewing 2008). The APSIM lucerne module (Probert et al. 1998; Robertson et al. 2002) is probably the best studied, best represented and most validated of all the perennial pasture modules in APSIM and was therefore chosen as the example for this study.

Note that while many papers have focused on validating the APSIM model using field or experimental data (Asseng et al. 2000; Verburg & Bond 2003; Yunusa et al. 2004), this study instead focused on using sensitivity analysis to analyse APSIM itself to elucidate relationships that exist between meteorological and soil input data and the resulting lucerne biomass production predicted by APSIM, based on the accumulated wisdom of experts that has been coded into the APSIM model. We hypothesised that, despite the intricacies of the internal structure of APSIM and the numerous and varied input factors generally required to produce accurate simulations from this model, our exploration would reveal relatively clear and simple relationships between input environmental data and the resulting simulated biomass production, thus clearly illustrating which drivers are most important, and how these principal drivers affect lucerne biomass production in APSIM. In the second paper of this series of 2 papers, we developed a simple statistical model, known as an emulator, of APSIM-generated lucerne biomass production which is based only on these principal drivers.

### 2. Methods

Data for this study were generated exclusively from APSIM by running simulations using the lucerne module. We assumed that the main factors that affect lucerne growth in Western
Australia are plant specific attributes, soil types and characteristics, pasture management such as grazing frequency, amount and distribution of rainfall, air temperature extremes, as well as radiation levels and day length. Previous extensive studies (Leach 1970; Brownlee 1973) indicate that the optimal grazing frequency for lucerne is between 6 and 8 weeks and so we assumed lucerne to be harvested every 2 months and grazing frequency was not further examined in this study.

The weather file for Badgingarra, Western Australia (30.34° S, 115.54° E) was used as the base file for meteorological data and this file was modified as needed to examine the effects of meteorological factors on aboveground biomass production. Badgingarra was chosen as a location “where climatic conditions do not greatly constrain lucerne growth” (Robertson 2006). This meteorological data was obtained from the Australian Bureau of Meteorology SILO database (www.bom.gov.au/silo) and is described in Jeffrey et al. (2001). The file contains historical daily information including rainfall, radiation, and maximum and minimum temperatures for approximately 100 years (1889–2006).

To thoroughly research the full range of effects of rainfall, radiation, temperature and soil type on APSIM-generated lucerne biomass production, simulations were run over a wide range of meteorological values, many of which represented extreme, idealised or even unrealistic conditions. The main purposes of investigating this extended range of environmental conditions were to ensure that an emulator developed as a consequence of this study would be broad enough to cope with any environment found in Western Australia, and also to explore the limitations in APSIM’s range of validity. However, in each case, a comparison was made with actual weather data from Badgingarra. The purpose of this comparison was to show the trends and limits of actual biomass production that occurs with the real weather of this site as opposed to what can be achieved in our ‘manufactured environments’ of extreme or idealised conditions.
Using APSIM simulation, lucerne was planted on 15 May 1889 (early autumn), the first year in which readings are available from the Badgingarra weather file. Plants were allowed to develop under actual climatic conditions present at Badgingarra for the first 11 years until monitoring started on 1 December 1900. This was to ensure that plants started out at the same level of growth at the beginning of the monitoring period for each simulation run and that only production of well-developed lucerne would be monitored. Starting the monitoring in 1900, ensured that ample years of simulation data were available for detailed trend analysis.

Once monitoring started, plants were cut to a height of 70 mm and harvested every 2 months. Lucerne biomass production monitoring was initiated on 1 December 1900 and continued at two-monthly intervals (to coincide with cut times) until 30 November 2006. Note that APSIM assumes that, once planted, lucerne will continue to grow indefinitely unless deliberately killed.

Soils in APSIM are defined by several parameters. New soils can be created by altering some or all of these parameter values. While the standard practice is to assign the soil type that is most consistent with actual soils found at a given site, for the purposes of this study on the effects of soil type on biomass production in APSIM, 3 disparate soil types were considered and their parameters were derived from previously surveyed soils. Important parameter values that are applicable to lucerne growth and development in each of these soil types are given in Table 1. In this version of APSIM, pH is assumed to have no effect on plant growth and P is assumed to be non-limiting, as is N because lucerne is a legume.
Table 1. Soil and plant information relevant to lucerne growth and development in the 3 soil types assessed in this study as coded in APSIM.

<table>
<thead>
<tr>
<th>Soil type</th>
<th>Maximum potential root depth(^a) (m)</th>
<th>PAWC(^b) (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor sand</td>
<td>2.1</td>
<td>54.9</td>
</tr>
<tr>
<td>Gravelly, loamy sand</td>
<td>2.5</td>
<td>111.6</td>
</tr>
<tr>
<td>Red/gray clay</td>
<td>1.8</td>
<td>134.7</td>
</tr>
</tbody>
</table>

\(^{a}\) Lucerne roots extend to a maximum depth of 2 m in APSIM.

\(^{b}\) PAWC = plant available water capacity

2.1 Changing rainfall

Effects of rainfall on biomass were examined first, since rainfall has been shown to be the main driver of pasture biomass production in Western Australia (Hill 1996). Several simulations were run using a single soil type to eliminate variability in biomass due to differences in soil types. The original Badgingarra weather file was altered from 1 December 1900 by fixing daily solar radiation to 12, 20 or 28 mj m\(^{-2}\) and average daily air temperature to 15, 23 or 30\(^{\circ}\)C. The levels of radiation and temperature span a range of growing radiation and temperature conditions for lucerne in Western Australia. Since APSIM requires maximum and minimum temperatures as inputs, daily average temperatures were converted into equivalent maximum and minimum values with a difference (Temp Diff) of 10\(^{\circ}\)C between them. Thus 9 new weather files were created, each with daily rainfall at its actual values as recorded in the original weather file. With temperature and radiation fixed, any variation in biomass is solely attributed to the variation in rainfall amount and distribution.

To separate the effects of rainfall amount and frequency on biomass, these 9 weather files were further modified by changing the actual rainfall, thus creating 3 additional weather files for each of the 9 previously created files. Rain was systematically set to fall every 7, 14 or 28
days with the amount that fell on each occasion decreasing by a certain percentage every 2
months from an initial value at 1 December 1900. Rainfall was set to decrease (as opposed to
increase) across time in order to prevent drying out of the soil profile during early stages of
the simulation runs, which might have stunted plant growth. For rain falling every 7 days, the
initial amount was set at 100 mm per rainfall event and the decrease was 1.47%. For rain
falling every 14 days, the initial amount was 200 mm and the decrease was 2.94%. For rain
falling every 28 days, the initial amount was 400 mm and the decrease was 5.88%. Thus the
total amount of rain in a given 2-month period was the same regardless of rainfall frequency.
These percentages were chosen to ensure that rainfall reached 0 mm several months before
the end of the monitoring period in November 2006.

The first set of simulations was run using poor sand and these 36 modified weather files
as inputs. Total rainfall for each 2-month period was plotted against total lucerne biomass
production for that period.

2.2 Changing soil type

The method described in the section above was repeated using gravelly loamy sand, and
red/gray clay, soils of markedly different water-holding capacities (Table 1). Consequently
the effects of changes in soil type and also any interaction effects of soil type and rainfall on
lucerne biomass production were able to be observed. All subsequent simulations were run
using poor sand as on the whole the trends and patterns observed in the results were the same
across all 3 soil types, despite the notable differences in their water-holding capacities.

2.3 Changing radiation

The effects of solar radiation on biomass production in APSIM were investigated. In this case,
rainfall amount and frequency were fixed at 10, 20 or 30 mm to fall every 7 days. Temp levels
and Temp Diff were fixed as per section 2.1 (Changing rainfall). As per section 2.1, 9 weather files were created using actual radiation data from Badgingarra. Then an additional file was created for each of these, where radiation was set to decrease by 0.054 mj m\(^{-2}\) every 2 months starting from 35 mj m\(^{-2}\), thereby reducing to approximately 0 mj m\(^{-2}\) by the end of the monitoring period. Simulation results from these 18 weather files were used to plot average radiation for each 2-month period against total lucerne biomass production for that period.

2.4 Changing temperature

The effects of air temperature were considered as per those described above. Rainfall amount and frequency were fixed at the same 3 levels as per section 2.3, radiation was fixed at the same 3 levels as in the section 2.1, and 9 weather files were created using actual temperature data from Badgingarra. Three additional weather files were created for each of these 9 files. Temp Diff values of 5, 10 and 15\(^\circ\)C were analysed using initial maximum temperatures of 37.5, 40 and 42.5\(^\circ\)C, respectively, to ensure that daily average temperature in a given 2 month period was the same regardless of the Temp Diff value. Daily maximum and minimum temperatures were decreased by 0.0545\(^\circ\)C every 2 months so that average daily temperatures reduced to approximately 0\(^\circ\)C by the end of the monitoring period. Results from simulations using these 36 weather files were used to plot daily average temperature for each 2-month period against total lucerne biomass production for that period.

3. Results

The primary interest of this study was in the pattern of trends observed when values of a factor or factors were varied, not in the values themselves observed for either factors or the resulting biomass produced. Therefore, actual levels of biomass produced and rainfall,
radiation or temperature values at which maximum production occurred are rarely mentioned in this section.

It should be noted that the biomass levels reached in some of these simulations were, on occasion, much higher than those obtained using the actual data at Badgingarra. This is due to the process of the sensitivity analysis leading to some simulations under extreme conditions that are unlikely to be observed in reality. In particular, the process of the sensitivity analysis that seeks to explore the full range of values of driving factors breaks the natural correlations between driving factors. High rainfall, high radiation and high temperatures are unlikely to occur together in Badgingarra, for example. However, these simulation runs at extreme values were necessary to reveal in full detail the trends resulting from varying each factor.

3.1 Changing rainfall

In general, increasing rainfall resulted in increased biomass from zero to an upper limit, which depended on both temperature and radiation.

The graphs relating to actual rainfall showed high variability in APSIM-generated lucerne biomass which was similar for all temperature and radiation combinations examined (Fig. 1). This variability resulted from differing rainfall frequencies that occurred with actual rainfall. Examination of the results from the modified weather files for a fixed temperature (maximum=20°C, minimum=10°C), fixed radiation (28 mj m\(^{-2}\)) and fixed 2-monthly rainfall (approximately 150 mm) showed that decreasing rainfall frequency resulted in increased 2-monthly biomass production. Decreasing rainfall frequency also tended to increase variability in the biomass response to rainfall at high radiation levels. However there was little difference between the trends shown across these frequencies and therefore rainfall frequency was held constant at 7 days in future investigations.
The results from the modified weather files showed more clearly other impacts of changing rainfall on biomass. Once a lower rainfall threshold had been reached, biomass increased linearly from zero with increasing rainfall, to an upper limit at a higher rainfall threshold. The rainfall thresholds and the upper biomass limits were affected by radiation, air temperature and frequency of rainfall.

Increasing radiation from 12 to 28 mj m\(^{-2}\) increased the maximum level of biomass and the rainfall at which maximum biomass was reached but the rate of change in biomass remained constant. In contrast, increasing average temperature from 15 to 30°C reduced the maximum level of biomass, increased rainfall at which maximum biomass was reached, and decreased the rate of change in biomass. There was also an interaction between radiation and average temperature with the effect of radiation on maximum biomass decreasing as temperature increased. At an average temperature of 15°C, maximum biomass increased by approximately 5 t ha\(^{-1}\) when radiation increased from 12 to 28 mj m\(^{-2}\), whereas at 30°C maximum biomass increased by only 2 t ha\(^{-1}\) for the same change in radiation.

It is noticeable that for all of the response lines, for each radiation level, there was some variation in biomass as rainfall increased above that required to produce maximum biomass. This is an artifact that arises because the number of days in a 2-month period varies from 59 to 62. This artifact is also present in Fig. 4 in relation to radiation and in Fig. 5 in relation to temperature. There were several noticeable gaps in the data from just under 300 mm to over 400 mm where there were no data points, indicating that rainfall levels when averaged over any 2-month period did not occur in this range at Badgingarra. This arises simply as a consequence of the way in which the periods have been defined and would not be present if cutting and monitoring had taken place more frequently. For similar reasons, there were gaps in the data in Figs. 4 & 5 relating to radiation and temperature, respectively.
3.2 Changing soil type

There was a slight decrease in maximum biomass reached on gravelly, loamy sand compared with the other 2 soils but this was only seen at high radiation in combination with low rainfall frequency (Figs. 2 & 3). There seemed to be more variability in predicted biomass in the panel for actual rainfall for red, brown sandy loam over clay than for either poor sand or for gravelly, loamy sand. The same levels of maximum biomass were reached in both poor sand and red, brown sandy loam over clay for each temperature–radiation combination. But in red, brown sandy loam over clay there was less variability in maximum biomass reached as well as the rate of biomass increase, compared with poor sand. Nevertheless, the trends and patterns observed were the same across all 3 soil types on the whole, which was why further analyses were conducted only for poor sand.

3.3 Changing radiation

The panel of graphs relating to biomass versus actual radiation showed clusters of points which correspond to the different 2-month periods in each year (Fig. 4). For rain of 20 mm and 30 mm falling every 7 days, biomass clearly increased with increasing radiation for all temperatures.

The panel of graphs for 2-monthly decreasing radiation showed that biomass increased from zero to a maximum once a radiation threshold was reached. The maximum biomass was dependent on both the amount of rainfall available and temperature. At low temperatures, greater radiation was required to initiate biomass production than required at medium or high temperatures but this seemed to be independent of rainfall level. Similarly at low temperatures, greater radiation was required to reach maximum biomass production. As rainfall increased, maximum biomass production was attained at higher radiation levels. Once
maximum biomass production was reached, biomass declined slightly or remained constant as radiation increased, depending on temperature and rainfall.

3.4 Changing temperature
Biomass declined with increasing actual temperature regardless of radiation or rainfall levels (Fig. 5). As in the earlier graphs showing actual radiation, clusters of points correspond to the different 2-month periods in each year. The change in the rate of decrease of biomass production for temperature values at about 25°C for some graphs (mainly those at medium and high radiation and rainfall levels and low Temp Diff levels) reflected the effects of exceeding optimal lucerne growing temperatures (Dolling 2006).

The panels with decreasing temperatures showed a more complete biomass response to temperature as the range of temperatures had been extended beyond the 2-month averages at Badgingarra. Very low temperatures resulted in no biomass. Biomass then increased rapidly to a maximum, with both this maximum and the temperature at which it was reached being dependent on rainfall, radiation and the level of Temp Diff. Further increases in temperature then resulted in a gradual decrease in biomass to low levels.

At low radiation levels, there appeared to be little effect of rainfall or Temp Diff on the response of biomass to changes in temperature. At high radiation levels and medium or high levels of rainfall there was clearly more biomass produced for a given average daily temperature than for corresponding low levels of rainfall.

There appeared to be an interaction effect of rainfall with Temp Diff and radiation. At low rainfall, biomass production reached greater maximum levels with decreasing Temp Diff. However at medium and high levels of rainfall, this occurred only for high radiation; for both medium and low radiation there was no change in maximum biomass produced with changes in either Temp Diff or rainfall.
Fig. 1. Effect of changes in rainfall amount and frequency on APSIM-generated lucerne biomass production in a poor sand soil. The black dashed vertical and horizontal lines on the extreme left panels show predicted biomass for a rainfall of 150 mm and thus help to elucidate how an increase in rainfall frequency for a fixed amount of rainfall results in less harvested biomass.
Fig. 2. Effect of changes in rainfall amount and frequency on APSIM-generated lucerne biomass production in a gravelly, loamy sand soil.
Fig. 3. Effect of changes in rainfall amount and frequency on APSIM-generated lucerne biomass production in a red/grey clay soil.
Fig. 4. Effect of changes in radiation on APSIM-generated lucerne biomass production in a poor sand soil.
Fig. 5. Effect of changes in temperature on APSIM-generated lucerne biomass production in a poor sand soil.
4. Discussion

This study reveals that there are clear and relatively simple relationships between lucerne biomass predicted by APSIM and the meteorological input factors that we considered. Surprisingly, there was no visible effect of soil type on APSIM-generated lucerne biomass production with simulated regular rainfall, in spite of the large differences in water holding capacity of the 3 soil types used (Table 1). This failure of soil type to demonstrate any effect on APSIM-generated lucerne biomass production probably reflects the regularity of simulated rainfall which means there were no very large or very small rainfall events. This resulted in plants being able to take up all soil water at each rainfall event, except that which evaporated, as the amount of rainfall never exceeded soil water holding capacities of any of the soils. The fact that soil type made some difference with real rainfall reflects the fact that real rainfall was much more variable, with occasional large rainfall events and more frequent but irregular small rainfall events. Large rainfall events potentially exceeded soil water holding capacity, especially if occurring in close succession, and thus the amount of water from these events taken up by plants depended on soil capacity. Similarly, the amount of water from very small rainfall events not evaporating and thus available to plants would depend on soil water holding capacity. Thus we expect that if we had tested lower or higher rates of rainfall than we did, and if these occurred in less regular patterns, then soil type would have had more of an effect in APSIM.

The relationships between rainfall and biomass produced and between solar radiation and biomass produced are essentially sigmoidal, with a close to linear relationship between a minimum and maximum threshold. Sigmoidal functions are not uncommon in the field of biology and are often used to model growth of ecological populations and other natural processes (Berger 1981; Streibig 1988; Barrow & Mendoza 1990). Negligible differences in maximum biomass production due to rainfall frequency are due to the regularity of simulated
rainfall patterns as explained above; at more extreme frequencies, or with less regular patterns we would eventually expect to see more of an effect.

The relationships between temperature and biomass are indicated by unimodal functions that increase to a single maximum and then steadily decrease thereafter. This probably reflects the fact that some processes within the APSIM simulation depend on temperature, with maximum efficiency at an intermediate optimal temperature and reduced efficiency at more extreme higher and lower temperatures. While these effects could be modelled with piece-wise linear functions, or quadratics, the intricacies shown in these graphs suggest that they are probably best approximated by splines rather than by curves with a relatively simple functional form. The advantages of splines are that they are simple to construct and can flexibly approximate complex functional forms with a high degree of accuracy. Splines have been used for modelling a variety of natural phenomena, such as calculating bacterial swimming speed (Stock 1976), air pollution (Leitenstorfer & Tutz 2007), breast cancer survival analysis (Gray 1992) and soil temperatures (Yang 2004).

Notwithstanding these main effects, it is the effect of APSIM-generated lucerne biomass production on the interaction between the amount and frequency of rainfall, solar radiation and air temperature that precisely defines the shape of the overall response surface. Our study suggests that noticeable high-order interaction effects exist between rainfall amount, solar radiation, daily average temperature and the difference between maximum and minimum daily temperatures. For example, in Fig. 3, the minimum threshold for rainfall amount where biomass starts to increase is higher when rain falls every 14 days than every 7 days, and higher still when it falls every 28 days. This is probably due to the fact that more rainfall will be lost to evaporation when rainfall occurs in smaller more frequent amounts, and so more rainfall overall is required for the same effect. Other interactions give similar insights into the
overall effect of the interaction between the collective processes underlying the APSIM simulation, and will also need to be included in emulators.

Little knowledge of the internal formulation and constructs involved in the APSIM model were required in obtaining these trends and patterns. However, we have demonstrated not only characteristics of the relationships modelled in APSIM between primary drivers of lucerne biomass production and resulting production output but also that these relationships are fundamentally relatively basic. The implication here is that this methodology can be used to examine any complex mechanistic model to efficiently elicit the effects of the main drivers on a specific outcome; detailed knowledge of the structure and composition of the mechanistic model is not required to accomplish this task.

These findings advocate the development of a statistical emulator of APSIM-generated lucerne biomass production that includes sigmoidal functions to represent the effects of rainfall amount and solar radiation, a spline function to represent the effect of temperature and some (possibly linear) higher-order interaction effects of these variables. This statistical emulator and other plausible emulator models of APSIM-generated lucerne biomass production are constructed, statistically validated and compared in part 2 of this series (Ramankutty et al. in review).

5. Conclusion

This study revealed simple relationships between meteorological data and soil type data that drives the complex growth model APSIM and the lucerne biomass predictions it outputs. These results increase the general transparency of the APSIM model and, by extension, other similar complex growth simulators. They also highlight the viable potential of building emulators for APSIM and other growth simulators, and give a good indication of what functional forms may be used for such emulators.
Acknowledgements

The authors would like to thank Michael Robertson, Jeremy Whish and Alison Laing for their invaluable comments and advice. They would also like to thank Art Diggle, Perry Dolling, Bill Bowden, Geoff Moore, Daniel Real and Krys Haq for their suggestions and guidance, and the Future Farm Industries CRC for their financial support. Formatting and other edits by Chris Davies are greatly appreciated.


Tennant D (1996) Potential yield calculator, user's guide. Department of Agriculture and Food WA, South Perth


Verburg K, Bond WJ (2003) Use of APSIM to simulate water balances of dryland farming systems in south eastern Australia, CSIRO Land and Water, Canberra
