Proceedings of the Thirty-eighth Meeting of the Agricultural Research Modellers' Group

EDITED BY

L. A. CROMPTON AND T. R. WHEELER

School of Agriculture, Policy and Development, The University of Reading, Whiteknights,
PO Box 237, Reading RG6 6AR, UK

(*First published online 4 September 2006*)

This group, which is concerned with the applications of mathematics to agricultural science, was formed in 1970 and has since met at approximately yearly intervals in London for one-day meetings. The thirty-eighth meeting of the group, chaired by Professor A. Kassam of The University of Reading, was held in the Kohn Centre at the Royal Society, 6 Carlton House Terrace, London on Friday, 31 March 2006 when the following papers were read.

PLENARY LECTURE
Agricultural modelling: a possible road map

J. H. M. THORNLEY*

Centre for Ecology & Hydrology, Penicuik, EH26 0QB UK

SUMMARY

In this contribution a view is given of the current state of agricultural modelling in the UK. I begin with the ontogeny of this group. A brief description of modelling is presented to define the conceptual framework. The importance of modelling objectives is stressed. The possible significance of the avalanche of data now coming from the high-through-put techniques in the '-omics' research areas (genomics, proteomics and metabolomics) is discussed.

The 103 contributions given at these meetings since 1998 are categorized in terms of topic, type of model, and type of research. The results are considered. At the topic (submodel) level, there are few biological themes where improved submodels would not be appreciated, although some submodels do seem to be 'good enough'. In animals as well as plants, better treatments for development and allocation might be singled out for attention. Although most work is now mechanistic, there is little basic work at submodel level; applied work may suffer from this imbalance. Few contributions on forest models, at any level, have been presented in this forum. Crop modelling especially might benefit from greater interaction with forest modelling, as might forest modelling from crop modelling. There have been no mechanistic contributions at the ecosystem level. It seems likely that, taking a 20-year view, mechanistic ecosystem models of engineering strength will be much needed. The lack of competence in this area, in the UK and worldwide, is due, in part, to the level of maturity of this area of science, but this is possibly accentuated by fragmentation of effort and the current vogue for short-termism.

There is now a broad consensus that a model is *de rigeur* for any agricultural or ecological research programme which aims to take a firm grasp of the responses of such systems, which can be complex. Mechanistic models are required to provide the understanding needed for intelligent and flexible management, given the near certainty of changing environmental and economic conditions. The -omics data avalanche increases the importance of connecting to the molecular level, but does not change matters qualitatively. Large mechanistic models, of animals, crops and ecosystems, present two challenges, neither of which is being adequately addressed: first the sheer scale of the model and the scope of the science which must be represented poses organizational problems; second, making the connections from the higher levels down to the molecular level of the protein and gene is hampered by lack of work at intermediate (often physiological) levels.

Key words: agriculture, ecosystem, model

INTRODUCTION

When asked to contribute to this meeting, it was suggested that I might take a forward look, and attempt to define some of the challenges for the modelling community in agriculture and ecology in the near future. Crystal-ball gazing is always a dubious exercise, but it is as well to consider first where we have come from and where we are now, before speculating about future directions.

This group held its first meeting in the spring of 1970 at the headquarters offices of the (then) Agricultural Research Council in Great Portland Street. We dubbed ourselves the 'Crop Science Model-Builders' Working Party', using the initials CSMP, after the eponymous IBM software, the Continuous

* Correspondence to: Dr J. H. M. Thornley, 6 Makins Road, Henley-on-Thames, Oxfordshire RG9 1PP, UK. Email: johnthornley@care4free.net

System Modelling Program, successor of DYNAMO and precursor of ACSL (Advanced Continuous Simulation Language, ACSL 2000). Since this first meeting, which I initiated with the help of colleagues at the Glasshouse Crops Research Institute (GCRI), the group has met usually annually. We have now arrived at our thirty-eighth meeting, just a few months after I arrived at a significant anniversary.

The format of the meetings has changed little since our beginnings. However, in 1980 I moved from GCRI to the Grassland Research Institute (GRI), where my research interests now included ruminants and I acquired new colleagues, one of whom was Jim France. The scope of the meetings was broadened to include animals and other parts of the agricultural scene. In the early eighties, Kenneth Blaxter and Jim France arranged for the abstracts of our papers to be published in the *Journal of Agricultural Science, Cambridge*, which led to a notable increase in quality and quantity of contributions to the forum. The organization of the meeting moved to the University of Reading with Jim France when he moved there from GRI (subsequently renamed IGAP, then IGER), where it has remained since Jim France's departure for Canada. Numbers attending have increased by some 50–100%, not a great deal over 36 years. There is, thank heaven, no comparison with the dotcom boom. Clearly, however, the meetings have a modest but continuing appeal. Those of us who believe that, in science as well as in other areas of human affairs, 'slow' is better than 'whizz-bang', may be quietly pleased with this ontogeny.

Over the years, I have, from the beginning, given quite a few talks at our meetings. Until today, I have already talked about something specific: with a diagram, a few equations, and some simulations; an easy formula. Therefore today, talking about generalities, I risk being out of my depth, banal and trivial. Since I have worked mostly at home over the past ten years, I may also be out-of-touch and out-of-date.

Forty years ago I very definitely had my own science agenda. I followed it, often rather pigheadedly, but that worked out well for me. If I had attended a meeting at which an old-timer had spoken about the current state of things, lamentable or otherwise, and then made statements about what we need to get on with, I hope my reaction would have been one of kindly tolerance, listening, but thinking quietly to myself things like: predicting the future is always wrong, past his sell-by date, old scientists do more harm than good, etc.

Thus, I feel somewhat diffident about what I am going to say to you. No doubt most of my thoughts on the topic will be wrong. I hope very much that you all have your own science agendas, to which you will adhere stubbornly, as that is the only way to get worthwhile work done. But I also believe that a personal science agenda, essential for each of us, has the status of a model, a working hypothesis, to motivate and allow us to get on with things. Thus it needs continual re-assessment against any inputs which come along.

MODELLING

Here I discuss briefly some of the general principles of modelling in biology and agriculture in order to define our conceptual framework. The 'modelling project' in our corner of science may be summarized as theoretical research aiming to predict quantitatively the properties and dynamics of biological systems relevant to our interests, and at the same time provide a useful degree of understanding (defined by the modeller's objectives). These systems include animals, crops, ecosystems, and components (subsystems) of these systems. The methods and ideas include analysis (reductionism), synthesis (holism) and 'emergent' properties. Many of the concepts can be traced back to De Wit (1970) who largely pioneered crop simulation. To simplify and permit later categorization, we make binary classifications (e.g. empirical or mechanistic; basic research or development), in what is in reality a continuum.

Objectives or why build models?

What do we hope to achieve by building a model? Kant (1724–1804) said 'There is nothing so practical as a good theory'. There are several possible reasons for building models.

1. To describe/summarize data.
2. To integrate (subsystem) knowledge leading to an understanding of a complex system.
3. To make predictions.

Answering the 'Why build models?' question defines objectives. Objectives determine the sort of model constructed and its (biological) scope.

At the outset of any modelling project, the most important initial exercise is to develop a clearly thought-through and consistent set of objectives. People build models for many different reasons, and with many different goals in mind. Much of the controversy which sometimes emerges in modelling discussions is rooted in differing objectives. The size and scope of models can also vary widely: modelling the mammary gland, or photosynthesis is a rather different exercise from modelling a whole animal or a plant ecosystem. The former could be a short-term project and is hardly multidisciplinary; the latter is essentially a long-term commitment of arguably at least ten years and requires detailed knowledge of large areas of biology, needing team-work and

collaboration. An agreed and coherent set of objectives provides motivation and direction to researchers.

Types of model

There are two principal types of model, empirical and mechanistic, and depending on objectives (above), we can choose to build one or the other.

1. To describe/summarize data. — An empirical model.
2. To understand a system. — A mechanistic model.
3. To make predictions. — Either sort of model.

An empirical model simply re-describes the data to which it is applied. Usually, an empirical model has no scientific content, reflecting neither conservation of matter nor the laws of thermodynamics. Examples of empirical models are: feeding response of an animal with weight gain regressed on feed input; crop response to fertilizer with yield regressed on fertilizer application.

A mechanistic model is integrative, based on analysis of the system into components. It is ordinary reductionist science. 'Understanding' means grasping the relationship between empirical data on one level and empirical data on lower levels. The components of a mechanistic model are empirical, although they may have some theoretical basis. Examples of mechanistic approaches are: feeding response of an animal in terms of rumen function and processes of nutrition and metabolism, driven by feed input; crop response to fertilizer in terms of plant-soil processes, nutrient uptake, soil water and photosynthesis.

For making predictions, any sort of model that works will do. It isn't necessary to understand how something works in order to be able to predict behaviour. In many areas it has been common for man to be able to make predictions before an understanding is achieved, for example, planetary movements before Newton and the day length response of plants. However, prediction with understanding – that is using a mechanistic model – always offers more options than empirical prediction alone. And the deeper the understanding extends towards the molecular level, the more options become available.

Although some agricultural scientists and ecologists are wary of modelling, it has to be emphasized that modelling and mathematics are the servants of science. The hypotheses expressed in mathematics and computer programs are derived from biological concepts. It is essential to make suitable biological assumptions and then use appropriate but often simple mathematics. The model and computer program provide a powerful framework for representing, exploring and applying ideas about how it is thought that the system works.

Empirical and mechanistic models can be further categorized as: deterministic or stochastic, and dynamic or static, with obvious meaning.

Complexity in biological models

It is sometimes said, particularly of models of animals, crops and ecosystems, that the model is 'too complex'. However, even large models are essentially simple – their 'complexity' is usually due to there being many simple elements. In dynamic deterministic mechanistic models (the type of most relevance to biology), each element of the model consists of a state variable, Y, generally denoting the quantity of a substance. The differential equation defines the rate of change of quantity as being the inputs (the positive fluxes arising from chemical synthesis or transport) less than the outputs (the negative fluxes arising from degradation or transport). That is: $dY/dt = \text{inputs} - \text{outputs}$. Large models consist of many such elements, where fluxes from one substance (or pool) become inputs to other substances (or pools).

A further simplification is that chemical conversion and transport are the only significant processes occurring in biology. The mathematics of chemical (or biochemical) conversions and transport (diffusion, convection) provide the tools for building a wide range of biological models.

All mechanistic models are wrong

Mechanistic models are reductionist. Reduction must stop somewhere. Therefore mechanistic models are always incomplete. Mechanistic models are always in some sense wrong but, of course, some are more wrong than others. They provide a limited but valuable view of reality. Hopefully, the limitations do not lead to serious errors, but that can never be guaranteed. Because all models are wrong, they are easily criticized. As human beings, we tend to defend our modelling creations, sometimes unwisely. Most models do some things well, others poorly or not at all. A model which does a few things well, perhaps breaking new ground, can be of great value, and should not be discarded just because of the dubious bits which have to be tacked on to make it all work.

Basic research and development

Another distinction we make below is to categorize models as 'basic', or 'applied'. Basic research contributes to basic knowledge or what the economists may call intellectual capital. Applied work or development is the process of combining elements of basic knowledge to provide tools which can be used for other objectives, for instance, in agricultural production (Doyle & Thornley 1982). Allocation of

resources between basic and applied work can affect (at least theoretically) the growth rate of the knowledge-application complex as a whole.

Evaluation and validation of models

Evaluation is a process which includes all methods of critiquing a model. It is not a wholly objective process. It incorporates properties such as simplicity, plausibility, elegance, generality and applicability. It should always begin with the modeller's objectives in mind, before proceeding to wider issues. A mechanistic model can and should be evaluated at two levels: that of the whole system via its predictions; and that of the subsystems via the biological assumptions.

Validation is an issue that is sometimes misunderstood. To validate means to demonstrate, over a specified domain of application, that a model has acceptable predictive accuracy. First, validity is not a property of the model alone. It depends on environment and management for the application in mind. A model might be valid for use in south-east Britain, but quite useless in the north-west of the country. Validity is not an all-or-nothing concept, but rather continuous, between (say) zero and unity. The statement, 'This model has been validated' is almost meaningless.

THE '-OMICS' DATA AVALANCHE

The development of high-through-put measuring techniques in the '-omics' research areas (genomics, proteomics and metabolomics) is now producing large quantities of data at the molecular level. The data are very extensive and also very expensive. How to make best use of these data is becoming a pressing problem, not least for those producing the data. Suggestions cover a wide spectrum, from sober realism to unbridled hype, although happily the former appears to be gaining ascendancy.

In some quarters (O'Malley & Dupré 2005), there has been talk about 'the emerging postgenomic discipline of systems biology', 'going beyond reductionism' and 'The developing anti-reductionist consensus'. Possibly to most of us at these meetings, and to many others, this is amazing nonsense, as spelt out for instance by Bothwell (2006). However, the serious question which needs addressing is this. Does the massive amount of data now being produced by genomics, proteomics and metabolomics change fundamentally the 'modelling project' (defined above)?

The short answer to this question is no, but the emphasis and possibilities of some modelling work may be changed. Nothing has changed qualitatively. The biology is essentially the same. We still operate within the 'normal' science loop of hypothesis, analysis, synthesis, prediction, test against real-world data, and back to hypothesis (Popper 1958).

What has changed is that there is now a greatly increased opportunity for relating predictions to properties defined at the protein/genome level, with all the possibilities that such knowledge may provide. It still seems not to be viable to try to construct models going from ecosystem (or animal or plant) to molecule in one bound. However, surely it is more important than ever to join top and bottom, step-by-step, to all the intermediate levels, of organ, tissue, cell and sub-cellular. When these connections are firmly made, then both our understanding and our ability to intervene will be maximized.

CONTENT OF OUR RECENT MEETINGS

Having defined some terms, now let us look at the contributions at these meetings over the last 8 years, and see if they tell us anything. In Table 1, the 103 contributions given at these meetings since 1998 are placed in 16 topic categories (a number which fits comfortably across an A4 landscape page). Then each contribution is assigned two attributes. The first attribute for type of model has five flavours: (mechanistic + empirical) × (deterministic + stochastic) + unclassified. The second attribute for type of research has two flavours: basic + development.

Looking at the bottom right corner of Table 1, where the sums over years and topics are given, it is seen that, of the 103 papers, there are: 56 mechanistic deterministic, 30 empirical deterministic, 9 mechanistic stochastic, 1 empirical stochastic, and 7 unclassified. There have been 11 basic contributions.

Most work is now mechanistic. This is very different from and much to be preferred to the situation 30 years ago when what might be called the statistical paradigm dominated agricultural research. I remember an early talk I gave at a meeting (not of this group), suggesting a mechanistic approach to the problem of carbon allocation in plants. A fairly senior scientist stood up and said he couldn't see the point – a regression equation with fewer parameters would fit the data well enough. Such attitudes were a great impediment to modellers trying to publish modelling papers.

However, there is comparatively little basic work. Because applied research and development (ARD) uses basic building blocks, it seems possible that ARD efforts are suffering from this imbalance. One wonders if this is due to the short-term nature of much present funding, the competitive peer-review system which surely must level down rather than levelling up, or the type of person who is attracted into agricultural modelling preferring applied work.

At the topic or sub-model level, there are a few notable features. Some topics have scored two or

J. H. M. THORNLEY

Table 1. *Categorization of contributions to the 1998/2005 meetings. For example, 'os' denotes 'Osborne', the first author of the Abstract. First attribute, with five flavours, indicated by font:* (i) **mechanistic deterministic**; (ii) *empirical deterministic*; (iii) <u>mechanistic stochastic</u>; (iv) <u>*empirical stochastic*</u>; (v) *unclassified – normal font. Second attribute, indicated by case: upper-case (XY) – basic research; lower-case (xy) – applied work or development. Abstracts can be found in Crompton & Wheeler (2005), France & Crompton (1999, 2000, 2001, 2002, 2003, 2004), and France (1998). Abbreviations: env. denotes environment; biochem. denotes biochemistry; B denotes basic research.*

Year	Plant processes	Crops Grass-land	Forests	Roots/ soil	Weeds	Fruit; Post-harvest	Pol-lution	Env. Climate	Pests	Diseases	Animal processes	Animals	Gene/ biochem; genetics	Farm	Land use	General	
2005 (10)		**os**		**RO**	**to**, *po*	**Pe**		la				**pa**, **mi**	<u>TE</u>			co	**6**, *1*, <u>2</u>, 1, 2B
2004 (14)		**zh**, *we*	**ba**				**mc**, pr	*wi*, *sh*, **th**	<u>na</u>, <u>CH</u>	**PA**, ho		**gr**			*au*		**7**, *4*, 2, 1, 2B
2003 (14)	**la**, **be**	**ch**, <u>wa</u>			**me**, ph		**ke**, co	fi		<u>NA</u>	**sc**	**st**	<u>CL</u>			fe	**10**, *3*, 1, 2B
2002 (14)	**TH**	*Ch*		**CH**, ad		**Xa**	*Mi*	ri		**pa**, **BE**	**zu**, *ke*	<u>ye</u>		**sa**		hi	**8**, *4*, 1, 1, 3B
2001 (14)	**MA**	*st*, *po*, *hi*	**ba**					*Bi*		**pi**	**cr**, *to*, **zu**, *ke*	*ya*		ro		cr	**6**, *6*, 2, 1B
2000 (13)	*di*, *ma*	**mi**, *ca*		**RO**	*ki*		*Ma*	**ri**	**br**	**pa**	**zu**	**mi**				ro	**7**, *5*, 1, 1B
1999 (14)	*th*, *mu*	**ma**, *ya*, **gi**, *ke*		*fa*, **la**, mo					<u>sk</u>		**zu**, *fr*, **ma**	*al*					**7**, *6*, 1, 0B
1998 (10)		**be**, *ba*, li, *ya*		**mo**			co				**cr**	*to*				**ad**, <u>br</u>	**5**, *4*, 1, 0B
103	**5**, *3* (2B)	**8**, *10*, 1	**2**	**7**, *1* (3B)	**3**, *2*	**2**	**3**, *4*	**4**, *3*, 1	**1**, *3* (<u>1</u>B)	**6**, <u>1</u> (3B)	**8**, *4*	**5**, *3*, <u>1</u>	<u>2</u> (2B)	**1**, 1	*1*	**1**, <u>1</u>, 5	**56**, *30*, 9, <u>1</u>, 7
Total	8	19	2	8	5	2	7	8	4	7	12	9	2	2	1	7	11B/103

fewer contributions: forests (2); fruit, post-harvest (2); gene level (2); farm (2); land use (1); and ecosystem (0). I return to the zero score for ecosystems below. Although improved submodels are always appreciated, there are topics where our submodels appear to be good enough for the purposes of agriculture and ecology. For example, in crop modelling, this includes photosynthesis, respiration and water relations. In animals as well as plants, organism development and allocation might be singled out as worth attention. Pest models tend to be stochastic, which may or may not be necessary; I take the position that one always builds the simpler model, i.e. the deterministic one, and only moves towards the more complex stochastic model when the deterministic model fails. From an ecosystem viewpoint, because now many sub-models are 'good enough', the difficulties are more concerned with handling large programs and complex responses.

Possibly some of our animal and crop models are also 'good enough' for certain purposes, although the fact that they do not reach down to molecular levels suggests that much remains to be done.

Our meetings have attracted few contributions on forests, although forest modelling has a great deal in common with crop and grassland modelling. Possibly this is because forests are not perceived as very important in the context of the UK. However, my feeling is that there could be valuable interaction and interchange between crop and forest modellers. Perhaps the title of these meetings could be broadened and invitations specifically extended to UK forestry departments.

It may be serious that there has been so little work at the ecosystem level, either at these meetings or in actuality in the UK. The climate of the planet depends to a significant degree on how the principal ecosystems respond to change. Ecosystem models will surely be much needed. Such models present two challenges. The first is that posed by the sheer size of the model required and the breadth of science encompassed. This reflects the complexity of the real system, and, in a sense, the models required are big science. The second challenge is to build models which reach down to the molecular level, single proteins and genes. This requires rebuilding those intermediate levels of research, often physiology at the cell, tissue and organ levels, which were allowed to wither in the polarization of research towards the gene and towards very short-term applied research. Both challenges must be met in order to develop ecosystem models of engineering strength. This requires an appropriate research environment, which should be reasonably stable, multidisciplinary, well-connected to experimental programmes. It should also permit adequate support for the four essential legs of large models: research at the submodel level, development of the ecosystem model,

application to data and documentation. Some modelling researchers are dismayed by the wasteful fragmentation of many modelling research programmes. The UK Meteorological Office may provide perhaps an example of how 'big' models can be organized and implemented, although the basic science in this area is relatively well-established and agreed. Some of you might have a feeling of *déjà vu*, as some scientists have been quietly saying these things for many years.

CONCLUSIONS

1. It is now quite generally accepted in agriculture and ecology that a model is *de rigeur* for research programmes aiming to take a firm grasp of system responses.
2. Most work is now mechanistic. Mechanistic models are required to provide the understanding needed for intelligent and flexible management, given the near certainty of changing environmental and economic conditions.
3. Models can contribute quantitative prediction, understanding, integration of complexity and clarity.
4. Objectives determine the level of reduction applied and the type of understanding obtained.
5. The large quantity of data now coming from genomics, proteomics and metabolomics underlines the importance of extending models, step-by-step through the intermediate levels of organ, tissue and cell, to the molecular level. Models connected to the molecular level maximize understanding and possibilities for intervention. However, the '-omic' data avalanche has not brought anything qualitatively new to the table. 'Traditional' modelling techniques appear, at least in principle, to be up to the task.
6. Little basic work at the sub-model level now takes place. This could hamper applied research and development.
7. In animals and plants, organism development and allocation might merit attention.
8. Agricultural modelling, particularly of crops, could benefit from better contacts with the forest modelling community. Worldwide, forests are an important ecosystem, and are increasingly managed sustainably, for biodiversity, recreation and as a local resource.
9. There is virtually nothing mechanistic going on at the ecosystem level – a hiatus which could be important. This may be due to the short-term and fragmented nature of much modelling research, as well as the level of maturity of parts of biology. Perhaps a long-term major initiative is needed: to construct ecosystem-level models of engineering strength reaching down to molecular

levels (although perhaps in two or more steps). There are four essential legs of large-model work which need to be nurtured: (i) research at the sub-model level, (ii) model construction and development, (iii) application to data and (iv) documentation. Current techniques and concepts appear sufficient for this task; current research arrangements are not.

10. Suggestions and questions for the future of this group:

   (a) Widen the title, to include ecology and environment.

   (b) Perhaps the group is growing out of the annual one-day meeting format, and would be well served by a two- or three-day residential meeting, possibly every third year.

   (c) How are items 5 and 9 to be progressed? Can this group help? Is there a role for networking subgroups, specializing in particular areas, with agreed complimentary objectives? Or are such matters left to spontaneity? A strength of the group has been its openness, informality and diversity, as well as its friendliness and warmth; these require preserving and enhancing.

REFERENCES

ACSL (2000). Advanced Computer Simulation Language. Huntsville, USA: Aegis Research Corporation.

CROMPTON, L. A. & WHEELER, T. R. (2005). Abstracts of communications. Proceedings of the thirty-seventh meeting of the agricultural research modellers' group. *Journal of Agricultural Science, Cambridge* **143**, 319–324.

BOTHWELL, J. H. E. (2006). The long past of systems biology. *New Phytologist* **170**, 6–10.

DE WIT, C. T. (1970). Dynamic concepts in biology. In *Prediction and Measurement of Photosynthetic Productivity* (Ed. I. Setlik), pp. 17–23. Wageningen, The Netherlands: Pudoc.

DOYLE, C. J. & THORNLEY, J. H. M. (1982). An economic model of research and development expenditure. *Oxford Agrarian Studies* **11**, 173–187.

FRANCE, J. (1998). Abstracts of communications. Proceedings of the thirtieth meeting of the agricultural research modellers' group. *Journal of Agricultural Science, Cambridge* **131**, 237–243.

FRANCE, J. & CROMPTON, L. A. (1999). Abstracts of communications. Proceedings of the thirty-first meeting of the agricultural research modellers' group. *Journal of Agricultural Science, Cambridge* **133**, 341–350.

FRANCE, J. & CROMPTON, L. A. (2000). Abstracts of communications. Proceedings of the thirty-second meeting of the agricultural research modellers' group. *Journal of Agricultural Science, Cambridge* **135**, 321–329.

FRANCE, J. & CROMPTON, L. A. (2001). Abstracts of communications. Proceedings of the thirty-third meeting of the agricultural research modellers' group. *Journal of Agricultural Science, Cambridge* **137**, 113–122.

FRANCE, J. & CROMPTON, L. A. (2002). Abstracts of communications. Proceedings of the thirty-fourth meeting of the agricultural research Modellers' group. *Journal of Agricultural Science, Cambridge* **139**, 103–111.

FRANCE, J. & CROMPTON, L. A. (2003). Abstracts of communications. Proceedings of the thirty-fifth meeting of the agricultural research modellers' group. *Journal of Agricultural Science, Cambridge* **140**, 479–487.

FRANCE, J. & CROMPTON, L. A. (2004). Abstracts of communications. Proceedings of the thirty-sixth meeting of the agricultural research modellers' group. *Journal of Agricultural Science, Cambridge* **142**, 243–250.

O'MALLEY, M. A. & DUPRÉ, J. (2005). Fundamental issues in systems biology. *BioEssays* **27**, 1270–1276.

POPPER, K. R. (1958). *The Logic of Scientific Discovery*. London: Hutchinson.

ABSTRACTS OF COMMUNICATIONS

Empirical models of standing biomass for short rotation coppice. M. E. BALDWIN, G. M. MORGAN AND E. J. POOLE. *Biometrics Division, Forest Research, Alice Holt Research Station, Wrecclesham, Surrey GU10 4LH, UK*

Increasing concerns over the role of fossil fuel use in climate change has driven research into 'carbon neutral' alternative energy sources. Early research identified closely spaced poplar and willow coppice as a suitable system for producing wood fuels. This system became known as 'Short Rotation Coppice' (SRC) due to the short (2–5 years) interval between harvests.

No tools were available for predicting SRC yield in the UK. In order to provide data for developing SRC yield models, 49 trial sites containing poplar and willow varieties were established on agricultural land throughout the UK according to a commercial planting design and managed for two 3-year rotations.

Models were fitted using GenStat (VSN International Ltd., Hertfordshire, UK). In particular, linear regression models, generalized linear models and mixed models using the REML functionality were used. Multivariate methods such as cluster analysis and principal co-ordinate analysis were also used to explore the structure of the data. Soil texture (grouped as clay, loam or sand), extractable phosphorus, extractable potassium, extractable nitrate, frost days, growing degree days, annual rainfall, seasonal rainfall, D100 (shoot diameter in mm, at 1 m perpendicular to the ground), soil pH, latitude, longitude, crop age and variety were identified as having a significant impact on the biomass. Using these variables, two models were created. The first model estimates the amount of standing biomass within an existing SRC plantation using observed D100 measurements and the second model predicts the potential amount of biomass produced by combinations of site and SRC variety. The second model also warns the user if their chosen combination of site location and SRC variety is likely to suffer from *Melampsora* spp. rust fungus, a pathogen that can cause serious damage to some SRC varieties and significantly reduce yield. The two models have been converted into software which was written using C++.

Each model produces yield estimates for 16 willow varieties and 13 poplar varieties grown in England, Scotland, Wales and Northern Ireland (Tubby *et al.* 2006). Validation of models through the comparison with data from commercial growers indicates these models are predicting biomass (oven-dried tonnes per hectare per year) within an acceptable range.

TUBBY, I., MORGAN, G., MATTHEWS, R. W., EVANS, S. P., HENSHALL, P., BALDWIN, M. E., TAYLOR, P. & POOLE, J. (2006). Yield models for short rotation coppice of poplar and willow. In *Forest Research Annual Report and Accounts, 2004–2005*, pp. 46–53. Edinburgh: The Stationery Office.

Establishing the risk of introgression of insecticide resistance for *Bemisia tabaci*. I. DEMON[1,2], F. VAN DEN BOSCH[1] AND M. JEGER[2]. [1]*Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK*, [2]*Imperial College London, Wye Campus, Ashford, Kent TN25 5AH, UK*

Introgression is defined as the introduction of a new gene or genes into a population by crossings between two populations. Sometimes introgression is planned, for instance in agriculture it is used as a method of improving the quality of cultivated crops (Ahoton *et al.* 2003). In cases where conservation of a species is sought after, introgression is most undesirable. The focus of the present paper is on the introgression of insecticide resistance between resistant and susceptible biotypes of *Bemisia tabaci*. *Bemisia tabaci* is a species complex consisting of a range of biotypes, known to have a high degree of inter-biotype reproductive isolation. This whitefly has the tendency to develop insecticide resistance and this is well documented in literature (Cahill *et al.* 1996). Less is known about the distribution of such resistance and how it is influenced by differences in life histories between biotypes. Crosses between fully compatible biotypes are more successful than crosses between partially compatible biotypes, but even so some hybridization occurs (Byrne *et al.* 1994). In areas where insecticide resistant and sensitive biotypes of *Bemisia tabaci* coexist, introgression of the resistance gene will have considerable consequences for whitefly management. To determine the risk of introgression of insecticide resistance into a population of susceptible biotypes, given that a hybridization event has occurred and to assay the relative importance of different life history parameters on the probability of introgression we use a stochastic branching process model. Based on the model explicit extinction equations are derived from which the probability of introgression is calculated. It is shown that a fitness cost expressed through the average number of eggs laid, has the largest effect on the introgression probability as compared to when the fitness cost is

expressed through other life history characteristics. The results change when a reproductive isolation mechanism is considered, for which it is shown that the fitness cost expressed through the male survival and mating probability have the highest effect on the probability of introgression.

Ahoton, L., Lacape, J. M., Baudion, J. P. & Mergeai, G. (2003). Introduction of Australian diploid cotton genetic variety into upland cotton. *Crop Science* **43**, 1997–2005.

Byrne, F. J., Cahill, M., Denholm, I. & Devonshire, A. L. (1994). Biochemical identification of interbreeding between B-type and non B-type strains of the tobacco whitefly *Bemisia tabaci. Biochemical Genetics* **33**, 13–23.

Cahill, M., Jarvis, W., Gorman, K. & Denholm, I. (1996). Resolution of baseline responses and documentation of resistance to buprofezin in *Bemisia tabaci* (*Homoptera: Aleyrodidae*). *Bulletin of Entomological Research* **86**, 117–122.

Quantifying uncertainty in the simulation of crop-climate processes. A. CHALLINOR[1] AND T. WHEELER[2]. *Department of Meteorology*[1] *and Department of Agriculture*[2] *University of Reading, Whiteknights, Reading, Berkshire RG6 6BB, UK*

Recent advances in the integration of process-based crop and climate simulation models have enabled quantitative estimates of the impacts of climate on crop productivity across seasonal to multi-decadal timescales (Challinor *et al.* 2006). Fully-coupled crop-climate modelling that simulates crop growth and surface fluxes of heat and moisture, together with land use information, can create a consistent representation of the crop-climate system (Osborne 2005). With the exception of some studies that have used probabilistic seasonal forecasts, most studies of this kind have simulated bio-physical processes deterministically. However, estimates of the impact of climate variability and change on crops are subject to uncertainty in both climate and crop simulations. Climate modelling studies have begun to quantify this uncertainty, by varying the value of parameters within the climate model within a range informed by expert judgement (Murphy *et al.* 2004). The present authors have extended this technique in order to quantify the uncertainty associated with simulations of groundnut (*Arachis hypogaea* L.) yield across India in both the current climate and under climate change (doubled CO_2).

The impacts of physical (climate model) and biological (crop model) uncertainty on mean yields were generally comparable in magnitude. Uncertainty in the temporal variability of yield, however, was dominated by physical uncertainty. The impact of biological uncertainty was more systematic across space than that of physical uncertainty. As a result, the relative importance of the two sources of uncertainty varied spatially. Physical uncertainty increased under climate change. Biological uncertainty increased only for the crop model parameters that control the response of the crop to CO_2 and to the mean and variability of temperature.

The methods used in the present study can improve model estimates of the impacts of climate variability and change by quantifying uncertainty more objectively, and by identifying key processes and parameters which can be further constrained by observations.

Challinor, A. J., Wheeler, T. R., Osborne, T. M. & Slingo, J. M. (2006). Assessing the vulnerability of crop productivity to climate change thresholds using an integrated crop-climate model. In *Avoiding Dangerous Climate Change* (Eds J. W. Schellnhuber, N. Cramer, G. Nakicenovic, T. Yohe & T. M. L. Wigley), pp. 187–194. Cambridge: Cambridge University Press.

Osborne, T. M. (2005). Towards an integrated approach to simulating crop-climate interactions. Ph.D. thesis, The University of Reading.

Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M. & Stainforth, D. A. (2004). Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature* **430**, 768–772.

Optimizing shrimp performance testing with a mechanistic growth model. A. B. DOESCHL-WILSON[1,2], D. JIANG[2], B. KINGHORN[2,3] AND H. VAN DER STEEN[4]. [1]*Animal Health and Nutrition Department, Scottish Agricultural College, Bush Estates, Penicuik, EH26 0PH, UK,* [2]*SyAqua, 3033 Nashville Road, PO Box 348, Franklin, KY 42134, USA,* [3]*School of Rural Science and Agriculture, University of New England, Armidale, New South Wales 2351, Australia,* [4]*StoneBridge Genetics, 1284 Sydney Court, Bowling Green, KY 42103, USA*

Compared with the sophisticated breeding programmes of established livestock species, the current breeding programmes for shrimp are under-developed, lacking amongst other components well-designed performance test strategies that maximize response to selection.

The present work demonstrates how mechanistic growth models coupled with optimization methods and other mathematical tools used in the breeding industry can assist in the design of performance test regimes. A shrimp model was developed to simulate growth and survival for various shrimp genotypes under diverse environmental conditions (Jiang 1999). The model was integrated into a computational framework, which calibrates the model parameters based on real data, calculates expected accuracies for the predicted genetic merits, both for these parameters and the resulting trait predictions, and applies optimization routines to determine the performance test regimes corresponding to maximal prediction accuracies.

In the case study presented here, the impact of various performance test periods and methods of sampling body weight on achieving the breeding objective of harvest weight were assessed. It was found that the accuracy of predicting genetic merit increases non-linearly as the body weight measured approaches the targeted harvest weight, with moderate increase after body weight reaches approximately 0·80 of the harvest weight. This suggests that the performance test period could be shortened without compromising prediction. Additional sampling points, when spaced sufficiently apart, generally improve the accuracies. Further, repeated measurements in a shorter performance test period provide similar or higher accuracies than single measurements taken close to harvesting, indicating that the loss of prediction power associated with selection at an earlier growth stage could be compensated by more frequent measurements. Finally, individual measurements, which are at present difficult to obtain in shrimp, can be added to family means in a selection index. However, this results in only about 10% more genetic gain compared to selection on family means alone.

JIANG, D. (1999). *Observed and modelled effect of salinity, temperature and dietary protein/energy on growth of juvenile Litopenaeus vannamei and nitrogen dynamics in static aquaculture systems*. Ph.D. thesis, Texas A&M University, USA.

Crop rotations and weed population dynamics: effects of cyclic permutations and initial conditions. S. K. MERTENS. *Biomathematics and Bioinformatics Division, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK*

Periodic matrix models use separate projection matrices to project a class-structured population through a particular environment (Caswell & Trevisan 1994). These are very useful for examining the effects of crop rotations on weed population dynamics as the dynamics within each crop can be represented by a separate projection matrix and the seed population can be classed by depth in the soil. Analysis of matrix population models commonly focuses on asymptotic behaviour, which is achieved when populations are in their stable structure. In non-periodic matrix models, though, it has been shown that long-term results may be misleading as conditions rarely remain constant long enough for the stable structure to be reached (Fox & Gurevitch 2000; Yearsley 2004).

Analytic methods are therefore developed to calculate the effects of perturbations on population size at any time, for periodic matrix models. This involves differentiating population size with respect to an element of the matrix projecting the population through a particular environment. As the derivative is of population size, all eigenvalues and eigenvectors are involved, as well as time. For full details of the method see Mertens *et al.* (in press). In particular, methods of calculating the effects of perturbations on population size both within and at the end of environmental cycles are shown. This is an advantage over asymptotic perturbation analysis where the effect is calculated over an entire cycle of environments. Being able to examine effects within the cycle is useful when one wants to compare predictions with the effects of actual perturbations to an experimental system.

Population projections and new methods are used to explore how cyclic permutations of a 2-year crop rotation affect weed population size and its sensitivity to perturbations (Mertens *et al.* in press). For example, for a rotation made of carrots and minimum-tilled wheat, where both are managed without herbicides, when most seeds are concentrated in the upper layer, starting with a wheat crop will lead to a population size that is nearly an order of magnitude greater than if the rotation had been started with a carrot rotation. The long-term population growth rate, though, would have been the same regardless of the starting crop (Mertens *et al.* 2002). In the long-term, perturbations in the carrot crop of the transition governing the contribution of seeds in the second soil layer to the population in the top layer (c_{12}) will have the greatest effect on population size. However, when the population is concentrated in the top layer, in the first four cycles of a carrot-wheat rotation, perturbation of c_{11} (contribution of top layer seeds to the next top layer population) will have the greatest effect, but in a wheat-carrot permutation with the same initial conditions, the greatest effect will be achieved by perturbing the transition c_{44} (contribution of seeds in deepest layer back to the deepest layer). A different set of initial conditions will result in a different set of patterns. The different patterns can be explained by examining how the starting environment interacts with the initial conditions.

By examining the transient dynamics, it has been shown that cyclic permutations of crops, and not just non-cyclic permutations (Mertens *et al.* 2002), can also play an important role in the management of weed populations. This is particularly important when a grower is considering a change to a new crop rotation, but is wondering which crop to start with, or once a new rotation has been established, where interventions to lower population size would be most effective. Finally, the methods are not limited to weeds and crop rotations, but can be usefully applied to other populations subject to periodically recurring environments, such as periodic fires.

CASWELL, H. & TREVISAN, M. C. (1994). Sensitivity analysis of periodic matrix models. *Ecology* **75**, 1299–1303.

FOX, G. A. & GUREVITCH, J. (2000). Population numbers count: Tools for near-term demographic analysis. *American Naturalist* **156**, 242–256.

MERTENS, S. K., VAN DEN BOSCH, F. & HEESTERBEEK, J. A. P. (2002). Weed populations and crop rotations: exploring dynamics of a structured periodic system. *Ecological Applications* **12**, 1125–1141.

MERTENS, S. K., YEARSLEY, J. M., VAN DEN BOSCH, F. & GILLIGAN, C. A. G. (in press). Transient population dynamics in periodic matrix models: Methodology and effect of cyclic permutations. *Ecology*.

YEARSLEY, J. M. (2004). Transient population dynamics and short-term sensitivity analysis of matrix population models. *Ecological Modelling* **177**, 245–258.

The genetic population structure of plant pathogen fungi due to different types of dispersal. L. U. WINGEN[1], J. K. M. BROWN[1] AND M. W. SHAW[2]. [1]*Department of Disease and Stress Biology*, *John Innes Centre*, *Norwich NR4 7UH*, *UK*, [2]*School of Biological Sciences*, *Plant Science Laboratories*, *University of Reading*, *Whiteknights*, *Reading*, *Berkshire RG7 6AS*, *UK*

Several plant pathogenic fungi, e.g. *Puccinia striiformis* (wheat yellow rust), can be dispersed by wind over long distances and infect host plants at the new location hundreds or thousands of kilometres away (Brown & Hovmøller 2002). Long-distance dispersal can play a major role in the onset, range and duration of epidemics. This is especially true as modern crop plants often cover huge areas with only limited numbers of different cultivars. Many population genetic models simulate dispersal by using exponentially bounded dispersal functions. In contrast, long-distance dispersal is better described by negative power-law functions (Kot *et al*. 1996). This may be the consequence of atmospheric turbulence, which are major forces of the dispersal by wind, following power-laws as well.

With an individual based computer simulation developed from that of Shaw (1995), the present authors have investigated the influence of the type of dispersal on the population genetic structure. Individuals are haploid asexual organisms with simple genomes (32 bit) which get mutated to produce genetic diversity.

Populations of nearly stable population size generated by power-law distribution functions are compared to those that were generated by using exponential bounded distributions. The virtual populations are analysed statistically with common population genetic measurements like G_{ST} and with a measure capturing self-similarity, named 'Conditional Incidence' (CI) both used at different spatial scales.

It is shown that the population structure which arises from fat-tailed distribution (power-law) is distinctly different from that one generated by exponential bounded dispersal functions. The G_{ST}, measurement is poorly suited to differentiate these population genetic structures especially at lower scales. The CI self-similarity analysis discriminates the population structures better as differences are noticeable at large and at small scales. For negative power-law functions with small exponents (f(x) proportional to x^{-2} and smaller) the self-similar structure of the generated populations is evident at these scales.

It is planned to be able to improve and define the new statistics further, so that they can be applied to data from real population genetic studies.

BROWN, J. K. M & HOVMØLLER, M. S. (2002). Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science* **297**, 537–541.

KOT, M., LEWIS, M. A. & VAN DEN DRIESSCHE, P. (1996). Dispersal data and the spread of invading organisms. *Ecology* **77**, 2027–2042.

SHAW, M. W. (1995). Simulation of population expansion and spatial pattern when individual dispersal distributions do not decline exponentially with distance. *Proceedings of the Royal Society London B* **259**, 243–248.

The roles of orthogonal and diagonal interactions in grid-based spatial simulations. C. P. D. BIRCH. *Centre for Epidemiology and Risk Analysis*, *Veterinary Laboratories Agency Weybridge*, *New Haw*, *Addlestone*, *Surrey KT15 3NB*, *UK*

The rectangular grid is established as the usual framework for applied spatial modelling, either through cellular automata or more complicated models. One potential advantage of grids is the computational efficiency of nearest neighbourhoods. However, the nearest neighbourhood in a rectangular grid is ambiguous, because each cell has four orthogonal neighbours with which it shares an edge and four diagonal neighbours with which it shares nodes. There has been little discussion of the relative weighting of diagonal and orthogonal interactions. Most simulations using nearest neighbourhoods either weight diagonal interactions equal to orthogonal interactions, or set them at zero.

A very simple abstract simulation of biological spread, including a trade off between orthogonal and diagonal spread, demonstrated quantitative and qualitative differences between equivalent simulations, differing only in the relative strength of orthogonal and diagonal spread (Birch 2006). Simulations with strong diagonal interactions had the property that clusters of colonized cells could generate continued spread in conditions which prevented the spread of isolated cells, increasing by up to

50 % the range of values of a resistance parameter at which spread could continue. In contrast, when there were no diagonal interactions, isolated cells could spread whenever any cluster of cells could spread: in some examples, simulations focused on only orthogonal spread achieved complete occupation in less than half the time required for equivalent simulations with spread equally distributed among orthogonal and diagonal neighbours. The explanation for these results is that diagonal interactions (and more distant interactions) link the neighbourhood interactions of adjacent cells, so that spatial dynamics become sensitive to clustering. This result is the underlying explanation for related dynamics seen in some more elaborate simulations of biological invasion (Birch *et al.* 2000).

The relative strength of diagonal versus orthogonal interactions can be calculated by simple and long-established methods, from basic information about spatial dynamics (Birch 2006). Predominantly orthogonal interactions would be typical of large numbers of short range or contact events, such as in many physical processes and populations of some soil flora and fauna. Diagonal interactions indicate processes with significant range, including many biological processes, ranging from translocation along rhizomes of bracken (Birch *et al.* 2000) to distant seed dispersal, or animal migration. Critics may argue that models in continuous space avoid the distinction between diagonal and orthogonal interactions, but their argument is only valid if the models are also in continuous time. Moreover, once understood the differences between diagonal and orthogonal interactions can be exploited to simulate thresholds and clustering characteristic of the systems being modelled.

BIRCH, C. P. D. (2006). Diagonal and orthogonal neighbours in grid-based simulations: Buffon's stick after 200 years. *Ecological Modelling* **192**, 637–644.

BIRCH, C. P. D., VUICHARD, N. & WERKMAN, B. R. (2000). Modelling the effects of patch size on vegetation dynamics: bracken (*Pteridium aquilinum* (L.) Kuhn) under grazing. *Annals of Botany* **85** (Supplement B), 63–76.

Modelling the structural diversity and interactions of agriculture within a life cycle assessment framework. D. L. SANDARS, A. G. WILLIAMS AND E. AUDSLEY. *Institute of Water and Environment, Cranfield University, Silsoe, Bedford MK45 4DT, UK*

Environmental Life Cycle Assessment (LCA) calculates the environmental burdens associated with a product, service or process. This was applied to 10 agricultural commodities to produce an inventory (LCI), illustrated below by lamb meat.

LCI models typically deal with production line systems where something is manufactured in linear steps. Agriculture is more complex and outputs arise at multiple stages and from co-existent systems. Lamb comes from finished lamb systems and as a by-product from producing breeding stock. Several organic and non-organic systems co-exist. These are best modelled as networks.

The system was modelled as a set of animal production activities linked by flows of animals. The solution is the amount, X, of each activity (e.g. upland cross breeding flocks), i, that produces the desired mass of output Z (e.g. prime lamb meat),

$$Z = \sum_{i=1}^{n} z_i X_i \qquad (1)$$

where z_i is the output of activity i, and also satisfies the set of flows between activities:

$$\sum_{i=1}^{n} c_{ij} X_i = 0, \quad j = 1 \ldots p \qquad (2)$$

where c_{ij} is constraint j on activity i. Demands are negative and supplies are positive and supply must equal demand (equation 2). For example, purebred lowland flocks produce rams, which are, in turn, demanded as terminal sires by lowland finishing flocks.

The total amount of material k (e.g. ammonium nitrate) flowing into the system is

$$M_k = \sum_{i=1}^{n} m_{ik} X_i, \quad k = 1 \ldots q \qquad (3)$$

where m_{ik} is the flow of material k into activity i.

The LCI for the system is the total of each burden l (e.g. nitrate leaching to water)

$$B_l = \sum_{k=1}^{p} M_k b_{kl}, \quad l = 1 \ldots r \qquad (4)$$

where b_{kl} is the amount of burden l produced by the use or disposal of material k and M_k is the total amount of material.

The LCI identifies the contribution of each material

$$B_{kl} = M_k b_{kl} \qquad (5)$$

or activity

$$B_{il} = X_i \sum_{k=1}^{q} m_{ik} b_{kl} \qquad (6)$$

and thus identifies particular hotspots. The model is solved by Gaussian elimination.

At the commodity level, 1 t of lamb meat required 44 GJ of non-renewable energy and involved 6·7, 20 and 18 non-organic ewes in various hill, upland and lowland systems respectively, as well as 2·6 organic ewes in both the uplands and lowlands. The system also produced 156 kg of wool and 411 kg of mutton.

Defra funded the project (IS0205) at Silsoe Research Institute.

Solubilization and uptake of strongly-sorbed solutes by plant roots. A. POTOTSKY[1], T. ROOSE[1], G. J. D. KIRK[2], D. L. JONES[3] AND D. TOMOS[4]. [1]*Mathematical Institute*, *University of Oxford*, *24-29 St Giles'*, *Oxford OX1 3LB*, *UK*, [2]*National Soil Resources Institute*, *Cranfield University*, *Silsoe*, *Bedford MK45 4DT*, *UK*, [3]*School of Agricultural & Forest Sciences*, *University of Wales Bangor*, *Gwynedd LL57 2UW*, *UK*, [4]*Department of Biological Sciences*, *University of Wales Bangor*, *Gwynedd LL57 2UW*, *UK*

Plants are known to enhance their uptake of strongly sorbed solutes from soils by excreting solubilizing agents from their roots. An example is the mobilization of soil phosphate by excretion of particular organic anions. Preliminary work is reported to extend a model of this phenomenon developed by Kirk (1999) for the scale of the individual root, to the scale of complex branching root systems and the whole plant, so as to allow for interactions within and between root systems and regulation of uptake by the whole plant. Roose & Fowler (2004) have shown how analytical solutions of equations for uptake by individual roots can be used to calculate uptake by a branching root system. As a first step toward deriving analytical solutions of the coupled equations for phosphate solubilization by organic anions for this purpose, Kirk's model has been analysed to find out which processes are important at particular time scales and how sensitive the model is to changes in different parameter values. An asymptotic solution of the equation has been derived for organic anion transport, reaction and decomposition, in the limit in which the root radius is small compared to the transport length, using the 'point-source' approximation of the corresponding Green's function. The solution is valid for times greater than 1 hr. A solution of the organic anion equation has also been derived for smaller times using the method of the Laplace transform. Further, the small-time asymptotic solutions of the organic anion equation have been used to solve the phosphate equation allowing for solubilization in the corresponding time limit. Using this analytic solution it has been found that the 'zero-sink' model of phosphate uptake breaks down at small values of the initial phosphate concentration. The analytical results have been compared against numerical solutions of the full set of model equations and conducted a rigorous sensitivity analysis of the whole model using the analytical solutions. The sensitivity analysis shows that the model results are most sensitive to the root radius, diffusion coefficient, water content, and flux of organic anion across the root surface. In continuing research, this information is being used to derive an analytical solution to the coupled phosphate-organic anion equations for long times.

KIRK, G. J. D. (1999). A model of phosphate solubilization by organic anion excretion from plant roots. *European Journal of Soil Science* **50**, 369–378.

ROOSE, T. & FOWLER, A. C. (2004). A mathematical model for water and nutrient uptake by roots. *Journal of Theoretical Biology* **228**, 173–184.

Applying Bayesian Model Averaging to mechanistic models. J. M. GIBBONS[1], G. M. COX[1], A. T. A. WOOD[2], J. CRAIGON[1], S. J. RAMSDEN[1] AND N. M. J. CROUT[1]. [1]*Division of Agricultural and Environmental Sciences*, *School of Biosciences*, *University of Nottingham*, *Sutton Bonington*, *Loughborough*, *Leicestershire LE12 5RD*, *UK*, [2]*Division of Statistics*, *School of Mathematical Sciences*, *University of Nottingham*, *University Park*, *Nottingham NG7 2RD*, *UK*

Model averaging and Bayesian Model Averaging (BMA) in a mechanistic model context are investigated. Model averaging is a group of methods for combining predictions from several models into a single set of predictions. The methods are especially useful when there is a set of similarly performing models, which differ in their predictions. The strength of the approach is that it considers model in addition to parameter uncertainty. Predictions are combined by weighting with factors related to model performance, resulting in ensemble predictions. BMA is model averaging in a Bayesian framework where the model weights are Posterior Model Probabilities (Hoeting *et al.* 1999). BMA was applied to a model that predicts the plant uptake of radio-caesium from contaminated soils (Absalom *et al.* 2001). Using five model selection criteria (AIC, BIC, Residual Sum of Squares (RSS), MDL and ICOMP), ten models, including the full model, were selected for averaging. These models were based on those of Cox *et al.* (in press). The model predictions and averaged predictions were compared using a calibration data set and an independent data set. Several approximation methods for calculating posterior model probabilities were compared with a full Bayesian approach implemented using a Markov Chain Monte Carlo (MCMC) method and a Metropolis-Hastings algorithm. A simplified BMA approach requiring only the maximum likelihood parameter estimates and Laplace approximation of the integrated likelihoods is also described. The PMPs estimated using the MCMC approach and the Laplace approximation strongly weighted models with fewer parameters. The BIC-based PMP estimates ranked the models in the same order as the Laplace approximation, but gave

more weight to the models with more parameters. The AIC-based estimates of the PMPs differed considerably from the other methods. However, in terms of RSS all the methods produced similarly performing predictions (range 17·3 to 19·4) for the independent data set. Predictions of individual points differed among models (e.g. the median prediction of one point ranged from $-2·0$ to $-1·8$ across models). The model averaged predictions captured this uncertainty. In terms of RSS and individual point variation the simplified BMA approach performed as well as the full approach. It is concluded that BMA is a valuable approach in mechanistic model development and suggest when it is appropriate to apply simplified BMA.

ABSALOM, J. P., YOUNG, S. D., CROUT, N. M. J., SANCHEZ, A., WRIGHT, S. M., SMOLDERS, E., NISBET, A. F. & GILLETT, A. G. (2001). Predicting the transfer of radio-caesium to plants using soil characteristics. *Journal of Environmental Radioactivity* **52**, 31–43.

COX, G. M., GIBBONS, J. M., WOOD, A. T. A., CRAIGON, J. & CROUT, N. M. J. (in press). Towards the systematic simplification of mechanistic models. *Ecological Modelling*.

HOETING, J. A., MADIGAN, D., RAFTERY, A. E. & VOLINSKY, C. T. (1999). Bayesian model averaging: A tutorial (with Discussion). *Statistical Science* **14**, 382–401.

Formulating an early-warning model for *Septoria tritici*.

D. E. TE BEEST[1], S. PIETRAVALLE[2], M. W. SHAW[3] AND F. VAN DEN BOSCH[1]. [1]*Rothamsted Research, Harpenden, Hertfordshire AL5 2 JQ, UK*, [2]*Central Science Laboratory, Sand Hutton, York YO41 1LZ, UK*, [3]*Department of Agricultural Botany, School of Plant Sciences, University of Reading, Reading RG6 6AS, UK*

Weather is known to be of great importance in the development of pathogens such as *Septoria tritici* (*Mycosphaerella graminicola*), causing Septoria leaf blotch on wheat (Shaw & Royle 1993). By quantifying the weather–disease relationships an early-warning model to predict *Septoria tritici* epidemics can be formulated.

An adapted version of window pane, an algorithm that iteratively searches through a predefined time period for a selection of weather factors, is used. It identifies which weather variables in which window are best able to predict a disease occurrence (Coakley *et al.* 1985; Pietravalle *et al.* 2003). The effect of cultivar resistance and the pre-sowing period are included, and the use of replicate observations allowed. The aim is to develop an early warning system and therefore only the period preceding growth stage thirty-one is searched.

The predictive model combining rain roughly in the period March and April and wind in January and February has the best prediction accuracy (95%); however, validation on unbiased data is required. A model combining rain in March and April with minimum temperature in January and February is the second best model (89%). The final predictive models predicts poorly for susceptible cultivars, suggesting an infection on a susceptible cultivar is less dependent on the variables incorporated in the models, most importantly less dependent on rain.

With the early-warning predictive model an early indication to disease risk can be given which informs growers about disease risk, which can have economic and environmental benefits. With the predictive model and the identified weather variables the effect of climate change on future disease risk can also be assessed.

COAKLEY, S. M., MCDANIEL, L. R. & SHANER, G. (1985). Model for predicting severity of *Septoria tritici* blotch on winter wheat. *Phytopathology* **75**, 1245–1251.

PIETRAVALLE, S., SHAW, M. W., PARKER, S. R. & VAN DEN BOSCH, F. (2003). Modelling of relationships between weather and *Septoria tritici* epidemics on winter wheat: A critical approach. *Phytopathology* **93**, 1329–1339.

SHAW, M. W. & ROYLE, D. J. (1993). Factors determining the severity of epidemics of *Mycosphaerella Graminicola* (*Septoria tritici*) on winter wheat in the UK. *Plant Pathology* **42**, 882–899.

A new method to estimate boundary models for biological datasets.

A. E. MILNE AND R. M. LARK. *Biomathematics and Bioinformatics Division, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK*

Webb (1972) first proposed the boundary line as a model for a biological dataset where one variable is a biological response (e.g. crop yield) to an independent variable (e.g. available water content of the soil). The boundary line is some upper (or lower limit) on the value of the response variable at any given value of the independent variable.

The boundary line concept has been used to model diverse biological responses, including crop yield responses to soil factors (nutrients, water and pollutants) and trace gas emissions from soil. However, it has been subject to criticism (e.g. by Schmidt *et al.* 2000). A particular weakness is that the methods that have been used to estimate the boundary are somewhat *ad hoc*, with no statistical basis on which the strength of evidence for the boundary line model can be quantified.

In this paper a novel statistical procedure to estimate the boundary line is presented. A censored bivariate probability distribution is fitted to the data

by maximum likelihood. The boundary line is the censoring function. Consider a random variable $X(x, y)$ which is bounded (or censored) above by $y = \hat{\mu}(x)$. If it is assumed that the standard deviation of the measurement error is $\hat{\sigma}$, then the following variable is observed:

$$Z(x, y) = \min_y(X(x, y), Y(x, y)) \qquad (1)$$

where $X(x, y) \sim f(x, y)$ and $Y(x, y) \sim g(y) = N(\hat{\mu}(x), \hat{\sigma})$, and $\min_y(\cdot)$ is the minimum value of the y component.

This approach to estimation of the boundary line has various advantages. First, it entails no arbitrary procedures such as binning to define observations on the boundary. The assumptions that are made are formal ones about the statistical model. Second, the statistical model is conceptually homologous with the biological model that lies behind boundary line analysis. Biologically the boundary is thought of as an upper limit on possible values of y, given x, and in circumstances where x is often limiting, many observations are expected to be found near the boundary. A further advantage is that confidence intervals for the estimated parameters can be calculated using the Fisher information matrix. Finally, because the model is fitted to all of the data it can be compared to a bivariate normal model using Akaike's information criterion (AIC) (Akaike 1973), which allows comparison of model performance based on a compromise of parsimony and close fit.

The method was demonstrated using simulated and real datasets. The results from the simulated data sets showed that theoretically the method worked well. The results from real data sets were also encouraging. For example, data from an experiment looking at the relationship between the integral over time of normalized difference vegetation index (NDVI, a remotely sensed index of canopy density) and yield of a wheat crop was considered. A linear boundary line model with $\hat{\mu}(x) = ax + b$ was fitted to the data. The AIC values showed that the boundary line model was more suitable than a regression model.

AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proceedings of the Second International Symposium on Information Theory*, (Eds B. N. Petrov & F. Csaki), pp. 267–281. Budapest: Akademiai Kiado.

SCHMIDT, U., THÖNI, H. & KAUPENJOHANN, M. (2000). Using a boundary line approach to analyze N_2O flux data from agricultural soils. *Nutrient Cycling in Agroecosystems* **57**, 119–129.

WEBB, R. A. (1972). Use of the boundary line in analysis of biological data. *Journal of Horticultural Science* **47**, 309–319.

A stochastic model for the spread of *Mycobacterium avium* subsp. *paratuberculosis* in rabbit populations.
R. S. DAVIDSON[1,2], G. MARION[2] AND M. R. HUTCHINGS[1]. [1]*Scottish Agricultural College, West Mains Road, Edinburgh EH9 3JG, UK*, [2]*Biomathematics and Statistics Scotland, James Clerk Maxwell Building, King's Buildings, Mayfield Road, Edinburgh EH9 3JZ, UK*

Rabbits have recently been identified as the key wildlife source of *Mycobacterium avium* subsp. *paratuberculosis* (M.a.p.), the bacterium causing Johne's disease in cattle (Daniels *et al.* 2003; Judge *et al.* 2006). The main route of transmission is via ingestion of contaminated faeces. The combination of high disease prevalence and levels of infection in rabbits, high levels of faecal contamination of pasture and the lack of avoidance of rabbit faeces by grazing cattle result in cattle ingesting potentially infective doses of M.a.p. every day (Judge *et al.* 2005).

Two stochastic models of the process of infection among rabbits have been constructed in order to assess the disease dynamics and likely persistence of infection in rabbit populations. The first is a time independent model for the disease prevalence as a function of age in a group of rabbits in which the population structure is assumed to be in equilibrium, so that the model includes horizontal and vertical infection, but no demographic processes. Fitting this relatively simple model to field data enables the estimation of the overall infection rate and its separation into horizontal and vertical components, with the result that the infectious contact rate β is estimated to be 0·037 while the vertical transmission probability is estimated at 0·14. Joint confidence regions over the space of these disease transmission parameters are obtained. The corresponding rates have subsequently been passed into a more detailed numerical time dependent model which takes account of non-equilibrium demographic processes such as birth, death, spatial inhomogeneity, migration and age dependence, in addition to horizontal and vertical infection, in order to explore the stability of the infection in response to various ranges of the disease transmission parameters. The modelling has shown that M.a.p. is likely to persist in rabbit populations for extended periods and thus any disease control strategy should include both the domestic and wildlife hosts.

JUDGE, J., GRIEG, A., KYRIAZAKIS, I. & HUTCHINGS, M. R. (2005). Ingestion of faeces by grazing herbivores – risk of inter-species disease transmission. *Agriculture, Ecosystems and Environment* **107**, 267–274.

JUDGE, J., KYRIAZAKIS, I., GRIEG, A., DAVIDSON, R. S. & HUTCHINGS, M. R. (2006). Routes of intraspecies transmission of *Mycobacterium avium* subsp. *paratuberculosis*

in rabbits (*Oryctolagus cuniculus*): a field study. *Applied and Environmental Microbiology* **72**, 398–403.

DANIELS, M. J., HUTCHINGS, M. R., BEARD, P. M., HENDERSON, D., GRIEG, A., STEVENSON, K. & SHARP, J. M. (2003). Do non-ruminant wildlife pose a risk of para-tuberculosis to domestic livestock and vice versa in Scotland? *Journal of Wildlife Diseases* **39**, 10–15.

Modelling protein metabolism by the hindlimb tissues in lambs. L. A. CROMPTON[1], J. FRANCE[2] AND M. A. LOMAX[3]. [1]*Animal Science Research Group, School of Agriculture, Policy and Development, University of Reading, Whiteknights, PO Box 237, Reading, Berkshire RG6 6AR, UK*, [2]*Centre for Nutrition Modelling, Department of Animal and Poultry Science, University of Guelph, Guelph, Ontario N1G 2W, Canada*, [3]*Division of Biomedical Sciences, Imperial College London, Wye Campus, Ashford, Kent TN25 5AH, UK*

Amino acid isotope dilution kinetics have been combined with arterio-venous difference preparations to simultaneously estimate protein synthesis and degradation by the tissue beds of the fore and hind limbs in humans and animals. The estimates of tissue protein turnover are assumed to represent muscle, the predominant tissue, but have always included a contribution from the non-muscular tissues (skin, bone and adipose tissue). Isotope dilution kinetics have been applied to construct and solve a mathematical model of tyrosine (TYR) metabolism across the hindlimb tissues of lambs in the steady state, which allows the estimation of amino acid inflows and efflows from hindlimb tissues. The kinetic model has been used to resolve *in vivo* isotopic data and to estimate the fractional rates of constitutive protein synthesis, degradation and accretion. Curve fitting models have been developed to describe the changes in fractional synthesis rate (FSR) and fractional accretion rate (FAR) with plasma amino acid concentration.

The kinetic model contains one intracellular pool representing free TYR and two extracellular pools representing the arterial and venous TYR pools supplying and draining the tissue bed. Conservation of mass principles were applied to each pool to generate two sets of ordinary differential equations (one set for total TYR and one set for labelled TYR), which describe the dynamic behaviour of the system and were subsequently solved for the unknown steady state flows. Inputs required for model solution are plasma flow rate across the hindlimb, TYR concentrations and plateau specific radioactivity (SRA) in the arterial and venous pools and plateau TYR SRA in the intracellular pool. Using these data and a limited number of assumptions, the model can be solved to calculate the steady state flow rates for TYR inflow, efflow and bypass and TYR flows representing the synthesis and degradation of constitutive protein. Fractional synthesis and degradation rates can be estimated for hindlimb tissue from model solutions and the TYR content of hindlimb muscle (32·811 mmol/kg wet wt).

The model has been used to resolve *in vivo* radio-isotope data obtained from experiments with multi-catheterised lambs during a constant jugular vein infusion of [2,3-³H]TYR tracer. TYR kinetics were measured across the hindlimb of ten lambs (36 kg) fed a range of feed intakes as described previously (Crompton & Lomax 1993). The nitrogen intake of the lambs ranged from 3–38 g/day and plasma TYR concentration varied from 29–89 nmol/ml. The kinetic model outputs demonstrated that the flow of TYR into constitutive protein only accounted for on average 0·694 of the TYR inflow into the hindlimb, mean (SEM) values were 1·44 (0·14) and 2·08 (0·25) nmol/min/g respectively. Comparison of hindlimb FSRs with FSRs measured directly in muscle, showed that the contribution of non-muscular tissues to hindlimb tissue was 0·482 (0·060).

Curve fitting models were developed to describe the relationship between the plasma TYR concentration (x axis, nmol/ml) and the fraction rates of hindlimb constitutive protein synthesis and accretion (y axis, %/day). Both models were sigmoidal in shape. For synthetic rate the model used was $FSR = y_{max}/[1 + (k/x)^c]$, and parameter estimates (SE) were $y_{max} = 8·11$ (0·46)%/d, $k = 30·5$ (1·37) nmol/ml, $c = 5·21$ (1·49). Residual sum of squares = 3·67, $R^2 = 0·890$ and Durbin–Watson statistic = 2·36. For accretion rate the model used was $FAR = a + \{b/[1 + (k/x)^c]\}$, with parameter estimates, $a = -1·07$ (0·35)%/d, $b = 3·22$ (0·55)%/d, $k = 45·4$ (3·10) nmol/ml and $c = 21·0$ (28·2). Residual sum of squares = 2·62, $R^2 = 0·893$ and Durbin-Watson statistic = 1·05. The parameters of the two equations lend themselves to direct physiological interpretation. Both fractional rate models fitted the data satisfactorily and can be related to our current knowledge on the nutritional control of protein turnover. The models have the advantage of providing a simple biological description of changes in tissue protein synthesis and accretion with changing peripheral amino acid concentration.

CROMPTON, L. A. & LOMAX, M. A. (1993). Hindlimb protein turnover and muscle protein synthesis in lambs: a comparison of techniques. *British Journal of Nutrition* **69**, 345–358.