KEYNOTE PRESENTATIONS

Chairman:

Dr L G Copping
L G Consultants, Saffron Walden

Session Organisers:

Dr L G Copping
L G Consultants, Saffron Walden

and

Dr H G Hewitt
The University of Reading

From Genomics to Agronomics - A road map for the new millennium

D J Nevill
Novartis Crop Protection, CH-4332 Stein, Switzerland

ABSTRACT

The crop protection scientist has a cornucopia of new technologies with which to address challenges in the marketplace. Prediction of field performance will increasingly rely on an understanding of functional genomics, xenobiotic interactions, and bioinformatics. We need to exploit new tools for the analysis of biological systems, for high throughput screening of chemicals and phenotypes, and in precision farming. Product types (agrochemicals, cultivar resistance and biological control) have several common requirements in establishing field performance. The efficient exploitation of technological opportunities across the discovery process will allow us to dedicate more of our limited capacity for field experiments to studies of integrated crop management systems.

INTRODUCTION

There has been massive change in crop protection technology and markets since the last conference in this series six years ago. In this introductory paper, I shall review these changes and their influence on assessing field performance. The change in approach to field evaluation is reflected in the title of this conference. In 1994 it was "Comparing glasshouse and field pesticide performance II"; now it is "Predicting field performance in crop protection". I would like to devote a few sentences to analysing this difference. "Comparing" has become "predicting". The implication is that we have knowledge in advance of field testing which can provide understanding of the effects that will be measured. A further requisite of a prediction is that models exist or can be developed to integrate characteristics measured in the laboratory, glasshouse or growth chamber in such a way that field-relevant data are generated. The other major change in the title of this conference is from "pesticide" to "crop protection". This changes our focus totally from an applied agent to concentrate on the crop and the constraints to yield which prevent the crop achieving its genetic potential. These can be both biotic factors (disease, pest, weed) as well as abiotic (temperature, light, moisture imbalance). Advances in molecular biology allow us to consider seriously the management of both these aspects of crop protection whereas previously synthetic chemicals had limited us to managing the biotic constraints. Within this more traditional area of disease, weed and pest control, the crop focus also encourages us to utilise a systems approach to integrate chemical control with resistant cultivars, biological control and cultural/agronomic measures.

These changes from "comparing" to "predicting" and from "pesticide" to "crop protection" also illustrate the fundamental changes that are occurring at an industry level. Can we still talk of a pesticide or even a crop protection industry? Or is it agribusiness? Our perception of products and markets plus the underlying technology platforms determines our framework for glasshouse and field testing.

TECHNOLOGY PLATFORMS

Synthesis chemistry and whole organism biology have been the fundamental technology platforms in the crop protection industry for more than 50 years. These areas of expertise are now being complemented by several areas of basic knowledge which are expanding rapidly. The most dynamic areas of technology change are in functional genomics, xenobiotic interactions, automation, and bioinformatics. The synergy between these combined areas provides us with powerful new tools to enhance our understanding of crop protection and bring novel products to the market.

Functional Genomics

We already have the complete genomic sequences for several organisms such as yeast, *Caenorhabditis*, *Arabidopsis* and *Drosophila*. The first crop plant, rice, will have been sequenced within a short time. This huge amount of basic information leads to both opportunities and challenges since knowing the order of Cs, As, Ts & Gs on a chromosome is not in itself highly useful. It is only the foundation from which we can begin in a comprehensive manner to understand the expression of genes, their function and interaction.

We are now developing the ability to uncover the functions of gene products predicted by DNA sequences of entire genomes. Functional genomics involves several levels of interaction that have been named the genome, transcriptome, proteome and metabolome (Table 1, Oliver 1998).

Table 1. Levels of gene function

Level of analysis	Definition	Status	Method of analysis
Genome	Complete set of genes of an organism or its organelles	Context independent	Systematic DNA sequencing
Transcriptome	Complete set of messenger RNA molecules present in a cell, tissue or organ	Context dependent (the complement of m-RNAs varies with changes in physiology, development or pathology)	Hybridisation arrays SAGE High throughput Northern analysis
Proteome	Complete set of protein molecules present in a cell, tissue or organ	Context dependent	2D gel electrophoresis, peptide mass fingerprinting Two hybrid analysis
Metabolome	Complete set of metabolites present in a cell, tissue or organ	Context dependent	IR-spectroscopy Mass spectroscopy NMR

All levels other than the genome itself are context dependent in the sense that they vary according to organism development, physiology and pathology. The transcriptome is the simplest context-dependent level, but its effects are indirect because messenger RNAs are transmitters of information and are not functional cellular entities. The proteins and metabolites in a cell are directly functional, but their analysis is technically demanding and their interactions are highly complex. Methods are now becoming available to study such effects comprehensively for simple organisms such as yeast (Uetz *et al*, 2000).

Xenobiotic interactions

Our understanding of the way active principles interact with a crop plant or a target organism advances less explosively than in the area of genomics. However, we have improved understanding of cuticular properties, translocation and degradation processes. Progress in adjuvant technology has led to improved performance of agrochemicals. Empirical knowledge on deposition, adhesion, retention and uptake of agrochemicals has helped us to select better molecules, improve their formulation and, therefore, their performance. Compilation of experimental data has been correlated with observed biological effects using simple statistical techniques. However, the emphasis of research is shifting to the production of more fundamental mechanistic information coupled with models of xenobiotic behaviour.

Table 2. The effects of biotic, abiotic and adjuvant factors on agrochemical efficacy

Efficacy factors	Plant and environment effects	Adjuvant effects
Deposition	Carrier volume Nozzle selection Crop architecture Leaf orientation	Flow rate Droplet spectrum
Retention	Droplet velocity Leaf surface character Environmental conditions	Droplet size Surface tension
Uptake	Leaf waxes Cuticle age and composition Environmental conditions	Surfactant concentration Surfactant character Wetting & drying Spreading Surface tension Cuticle vs. stomata
Translocation	Growth stage Plant physiology Environmental conditions	Uptake efficiency Contact phytotoxicity

Each factor controlling agrochemical efficacy is itself influenced by a complex of characteristics (Table 2, Zabkiewicz 2000). Spray deposition models begin to give us better understanding and will help us define improved application guidelines. Simulation models may eventually be linked to GIS and GPS systems to provide analyses of real time operations. Adhesion and retention are also dependent on well-understood physico-chemical parameters and are also amenable to modelling.

There has been gradual progress during the last decade in our understanding of xenobiotic uptake into the target organism. The use of model experimental systems has allowed us to analyse differences between plant species as well as the influence of environmental factors and additives on uptake (Bromilow and Chamberlain 1995, Kirkwood 1999). Similar investigations are underway for root uptake where our knowledge of uptake phenomena is even more limited. Many questions remain unanswered: for example, it is not yet determined whether there are separate leaf entry pathways for ionised and neutral molecules. Further research is needed to provide a holistic synthesis of the various processes involved in xenobiotic uptake. We are almost at a breakthrough point where we can begin to use predictive models in the selection of formulation components and in the pre-evaluation of new crop protection molecules.

Xenobiotic interactions with the physical environment have been intensively studied over the last 30 years, mainly because of our interest in potential adverse environmental impacts of agrochemicals. However the same environmental fate models can also provide important understanding of robustness of product performance and the link between glasshouse and field (Fig 1, Boesten, 2000).

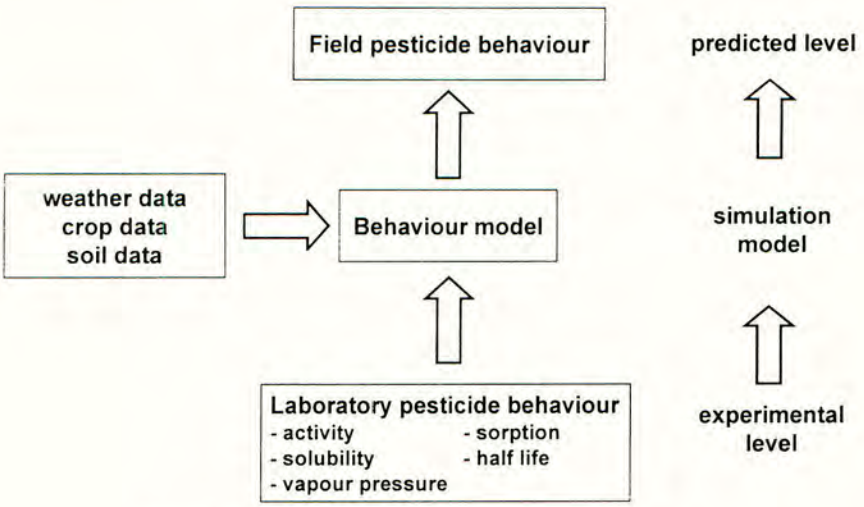


Figure 1. A schematic model linking laboratory to field.

The present environmental concerns over agrochemical residues in the atmosphere, water, soil and food will not go away. Improved active ingredients, formulations and application methods will be needed to minimise non-target effects. The difficulties in forecasting product behaviour and the cost of empirical field screening can only be reduced by a better mechanistic understanding of xenobiotic behaviour. This approach then leads itself to the development of empirical or process-driven models which may be incorporated into computer-based decision support systems (Zabkiewicz, 2000).

Automation

We now have the ability to carry out routine tasks with high throughput, high precision, and high reproducibility using robotics. This technology change has had fundamental impact on synthesis chemistry, gene sequencing, target finding as well as screening operations. Basic experimental functions such as preparation and measurement of input factors, manipulation of the test targets and assessment of the end result can all be partially or fully automated. Advantages are achieved when clearly defined experimental protocols need to be conducted in a highly uniform and repetitive manner. Robotics cannot replace inventive steps such as discovering and implementing new synthesis routes or performing individual biological experiments to test a novel hypothesis. Automation also begins to play a role in the field in such areas as precision application and direct sensing of treatment effects.

Bioinformatics

New technologies are capable of generating vast amounts of data which must be captured, stored and mined. The sequencing of a 130 Mb *Arabidopsis* genome is only just the beginning. Gene expression and metabolic profiling techniques will generate yet more data points as will phenotypic analysis in glasshouse and field. High throughput screening over several targets for 10^5-10^6 compounds per year demands significant computer support for the logistic, assessment and database management processes.

Research in computational biology is divided into two main areas: the analysis and interpretation of data; and the development of new algorithms and statistics leading to predictive models. An investigator needs to interrogate complex data sets in different ways to seek links with known phenomena – perhaps on different '-ome' levels (see section above). An obvious task may be to provide structures and functions of gene products of interest. Biochemical pathways to which a transcript belongs could be identified or interacting gene products could be found. Intelligent programmes are needed which may learn from previous experiments, searches and iterative analyses. Only then can we interpret interactions of tens of thousands of genes in terms of whole organism phenotypes whether in the field or in the glasshouse. Standardisation of data types and formats will be essential to facilitate meta-analyses across many studies to give robust and reliable results.

The power of bioinformatics lies in its potential to bring together disparate data from different organisms and different disciplines to unify biology (Thornton, 1998). The understanding and modelling of functionality is an essential component of the design and optimisation of new crop protection agents.

NEW TOOLS

Increased understanding of molecular biology, xenobiotic interactions, automation and bioinformatics has led to new methods for the discovery and development of active principles for crop protection and for their optimal use in the field. Such tools include: new bio-analytical methods; high throughput screens on *in vitro* and *in vivo* levels; marker-assisted breeding to integrate traits more rapidly and efficiently into elite cultivars; and, in the field, precision agriculture.

Analytical Methods

Rapid advances in functional genomics together with automation and bioinformatics begin to allow the analysis of transcriptomes, proteomes and metabolomes.

In principle, a whole genome represented by oligonucleotide probes can be deposited in a precise array on a few square centimetres of silicon wafer. Fluorescence-tagged c-DNA from test organisms can be washed over the gene chip which will illuminate where an oligo matches to the expressed gene. As our genome knowledge increases, so does the availability of such arrays for specified organisms e.g. yeast, *Arabidopsis*, rice. It will quickly become possible to understand the effects of pathogenesis or abiotic stress at the transcription level and look for remediation possibilities.

Such arrays have a wide variety of uses:

- Comparisons across environmental conditions will provide understanding of pathways regulating plant responses to environmental factors such as temperature, water, salt, nutrients and light intensity.
- Comparisons between healthy and diseased states will improve our understanding of the molecular basis of plant response to pathogen or pest attack.
- Differential expression patterns in various tissues will increase our understanding of organism development and our ability to manipulate it.
- Integration with other functional analyses will lead to discovery of new promoters and new gene functions.
- Comparison of organisms treated or untreated with xenobiotics may improve knowledge of mode of action
- Analysis of transgenics will give evidence on the specificity of a DNA insertion
- Mammalian expression arrays may provide early information on the toxicology of new molecules.

Most classical crop protection chemicals do not influence transcription directly, but have downstream effects on cell physiology. Therefore, a gene expression profile will often only give a secondary fingerprint of the effects of an applied chemical. New tools are becoming available which will allow us to analyse rapidly and to record changes in proteins and metabolites as a response to a novel stimulus. These methods will facilitate the determination of mode of action as well as allow more detailed comparison of performance under different environmental conditions.

Target-based high throughput screens

In vitro screens against validated targets have been taken up by the agrochemical industry because of technology and chemical diversity available for pharmaceutical discovery. They are amenable to high levels of miniaturisation and give fast read-outs of activity. They also utilise the information that is rapidly becoming available from functional genomics. The target-based screen allows the use of ultra-high chemical throughput (100,000 compounds per day have been achieved in the pharmaceutical industry) for compound discovery, if this is desirable (Dove, 1999). It also gives a direct read-out of activity at the target site which is not confounded by whole-organism effects such as uptake, translocation or metabolism. However, the *in vitro* screen is far distant from the field situation and also suffers from a number of unique disadvantages.

- There is no clear strategy to select screening targets out of the potential pool of thousands of lethal targets in an organism.
- Indirect or cascade effects may not be seen or the target may have a different conformation in the assay to *in vivo*.
- Assay development takes time and resources and may be a bottleneck in the screening process.
- A multidimensional structure-activity optimisation will be needed to discover a new agrochemical, however the *in vitro* screen may also support a conventional synthesis programme.

Active compounds from such screens need validation in whole organism tests. A major challenge for us is, therefore, to select those screening targets which are likely to be most

efficacious in the field and to develop assay systems which can cope with many 100s of new targets per year.

Whole organism-based high throughput screens

Our research advantage over the pharmaceutical industry is that we can work directly with our target organisms (Harrison, 1999). This allows us to establish a database with a constant foundation where the spectrum of organism assays stays more-or-less fixed and the numbers of compounds increase over time. Using miniaturised systems we can easily screen more than 100,000 compounds per year. The number is dependent on the amount of each compound which is available, the complexity of the application method, and the test substrate. Highly simplified assay systems allow us to screen in excess of half a million units per year. This may be particularly useful for natural product screens where there are many fermentation products. The approach may lead either to a novel synthetic chemical or to an active principle which could be produced by a transgenic plant.

The organism-based screen should allow us to rank the biological activity of compounds in a similar way to that which would be obtained in the field. A more detailed profiling of the compounds may be desirable, but is not essential since it can be carried out with lower throughput technology. In order to achieve a high chemical throughput, we may need to reach compromises between the practicability of an assay and its field relevance (Table 3). The field relevant assay may be more complex and sensitive to experimental conditions leading to a higher requirement for technical skills to achieve the necessary level of uniformity and reproducibility. The required throughput may also define the properties of the assay or the simpler assay may be used as a pre-screen for the more field-relevant test. The design of the assay alone cannot fully define field relevance; integration of known properties of test chemicals is important and field validation is essential.

A major challenge, which will decide the approach to high throughput screening, is the discussion of chemical diversity. At present it is still unclear how many compounds of what structural types need to be screened to provide sufficient active 'hits' for further chemical optimisation. Synthesis methods also improve to allow more follow-up projects. Using the techniques of high-speed synthesis, dependent on the structural features of the optimisation project, a chemist may synthesise ten times as many compounds now as were possible five years ago. In this dynamic interaction between biology and chemistry, pragmatic rather than strategic decisions define whether we try to screen 10^5 or 10^6 compounds per year.

Table 3 A comparison of a model vs. a more field-relevant assay for herbicide screening

Assay features	**Model system**	**Field-relevant system**
Test organism	Algal suspension	*Stellaria* whole plants
Growth conditions	Liquid culture, 96-well plate	Soil, 96-well seed tray
Application method	Pipette	Spray
Application timing	Not relevant	Post- or pre-emergent
Read-out	+ or -	symptoms
Chemical needs	ng per test	µg per test
Compounds / year	10^6	10^5
'Hit' rates	High	Moderate
Field correlation	Variable, depends on chemistry	Similar ranking
Technical skills	Low, routine maintenance	High, to obtain uniformity

Marker assisted breeding

Deliberate breeding for cultivar resistance to pest or disease resistance has a 100-year history and inadvertent selection for resistance goes back to the origins of agriculture. In the last 10 years, our interest in cultivar resistance has increased due to the availability of transgenic traits such as the δ-endotoxin from *Bacillus thuringiensis*. Ongoing research into novel insecticidal and fungicidal principles will bring more products to the field in the next few years. Additionally, a more comprehensive approach to the exploitation of the native germplasm will facilitate the discovery of more non-transgenic principles, including an improved exploitation of resistance to abiotic stresses. This non-transgenic approach may also be much better accepted by the public.

Seedling-expressed traits with monogenic inheritance can be best brought into elite germplasm using direct phenotypic selection. For traits with more complex genetic control (e.g. durable disease resistance) or for those which are difficult to measure precisely (e.g. heat or drought tolerance during flowering), we need to use new breeding techniques (Mackill, 1999). DNA-based markers facilitate reduced generation times and enable faster and more efficient breeding programmes, giving lead times to the market which are similar to agrochemicals. The use of marker-based approaches requires precise comparison of the genetic and phenotypic maps, which, especially for quantitative trait loci, implies careful phenotypic analysis and an understanding of potential genotype x environment interactions. On the simplest level, this begins with the two environments – glasshouse and field.

Precision Farming

Precision farming is a package of techniques for optimising field inputs and outputs. It is primarily aimed at enabling the farmer to maximise his profitability from arable production. However, many of the techniques of precision farming can also be used in field trials, giving us a new level of detail and accuracy for monitoring field conditions and explaining environmental variability.

Key technologies for precision farming are:

- Data collection systems e.g. soil sampling, aerial remote sensing, ground-based real time sensing, yield monitoring, hand-held input devices
- Navigation systems such as satellite-linked global positioning
- Data management systems which enable land mapping, modelling, decision support and expert systems
- Data communication systems via internet, satellite etc.
- Application systems allowing variable planting or variable application of fertiliser or agrochemicals.

Digitised maps are now frequently available for soil properties and local weather conditions. Global positioning systems may then be used to assist in the definition of field sites as well as in establishing the history of a data set. Improved models of environmental dissipation enable us to have a better understanding of product fate. For example, root-zone models relate weather and soil data to product dissipation in the soil and enable predictive statements of residuality and carry-over potential from glasshouse and laboratory studies. In general, predictive models,

which allow us to integrate large data sets from various sources, should also improve our understanding of the relationship between glasshouse and field results.

NEW PRODUCTS

Synthetic chemicals remain essential for cost-effective crop protection. The need for further innovation is driven by factors such as inadequate performance, pest resistance and environmental concerns. New chemicals need to be highly active, highly specific but broad enough to allow commercial success, toxicologically and environmentally benign, cheap and should also possess novel customer benefits (e.g. improved rainfastness, application flexibility, crop safety). There are opportunities to find new sources of chemical diversity and to integrate chemistry with cultivar resistance and biological control.

Sources of Diversity

The diversity of synthetic chemistry is almost infinite. However, most of these compounds are unsuitable as agrochemicals. They may be inactive, too reactive, too general as toxins or too expensive. The hurdles for a new crop protection active ingredient are very high – the state of the art is already high with many excellent products on the market. There are more new compounds available to be purchased than can be immediately handled by organism-based high throughput screening systems. Therefore, pre-screen decisions are needed to try to select out those compounds which are obviously undesirable. This may be on the basis of chemical-types which have been previously successful in the field or known highly reactive species. Diversity-based decisions may be essential at present, but small chemical differences may lead to big biological effects, therefore, care is needed. A challenge is to create a large enough chemical database of biological activity to be able to compare the use of diversity tools through to the field.

There is still a high level of interest in natural products as sources of diversity. Fermentation broths are increasingly available for screening from biodiversity projects and various third party groups. These are highly complex mixtures which are normally inactive and need very simple screens with very high throughput. They also need methods to separate active extracts and to establish novelty at an early stage of testing, since many active principles are commonly occurring antibiotics. Reproducibility of results is a key issue often influenced by small changes in fermentation conditions. Common agreement on methodology and rigorous quality control procedures are essential to ensure that positive results from early screens can be repeated in field experiments.

Cultivar Resistance

Plant breeders have intensively evaluated cultivar resistance in glasshouse and field for decades. Resistance was an important tool in an IPM programme, with little emotional impact until the advent of transgenic crops. The field evaluation of transgenic cultivar resistance (especially herbicide, insect or to a lesser extent disease resistance) has become a subject for intense international debate. A simple analysis of efficacy in glasshouse and field is inadequate, because a full risk assessment (as for synthetic crop protection) is required.

Potentially negative side effects of these crops are appropriately analysed alongside the evaluation of the beneficial main effect.

To be technically thorough in this risk analysis, we need to answer two questions for the transgenic cultivar:

- Does the inserted DNA only impact the expected metabolic pathway leading to pest, disease or herbicide resistance?
- Is there an adequate human and environmental safety assessment of the transgene's products?

The first question will be answered most efficiently with the tools of functional genomics described earlier. We should soon have the ability to look at the transcriptome, proteome and metabolome of the transgenic plant to ensure there are no unexpected effects of the inserted gene. The second question can be answered using the processes already available for evaluating the safety of synthetic crop protection chemicals. The essential difference between crop protection achieved through synthetic chemistry or cultivar resistance is the application method. This influences the exposure of organisms (including humans) to the active principle. In other areas, the parameters for establishing the toxicity to targets or non-targets can be similar for both product types and can comprise similar models and protocols.

Biological Control

The use of living organisms as pest control agents generally does not provide robust levels of economic pest management. In the disruptive, variable and often diverse environments of modern arable agriculture, the biological control organism finds it difficult to maintain its target below a population level that causes economic damage. However under certain situations there may be high and sustainable benefits:

- The control of exotic pests through the introduction of natural enemies from native habitats
- The suppression of large uniform populations of weeds or pests especially if the life cycle is long compared to that of the biological control organism
- The management of certain pests under controlled environment conditions e.g. in glasshouses
- The management of secondary pests where these have become highly damaging due to indiscriminate use of broad spectrum pesticides.

The assessment of biological control organisms in the glasshouse and the extrapolation of these results to the field may be difficult:

- The organisms may be environmentally sensitive or may rapidly lose viability if not preserved with appropriate formulations
- Manufacture by fermentation may lead to batch differences
- The organism may not be able to find its host or prey in the field due to spatial or temporal diversity.

Risk assessment is a particular challenge for biological control and is an area where careful comparison between glasshouse and field conditions is essential. Organisms are able to reproduce and multiply after introduction. This may sound obvious, but it is a key difference to the application of synthetic chemicals. For risk assessment, this means that although the hazard from the organism may be low, its exposure to the environment may be permanent.

Several classical release programmes have led to undesirable loss in non-target diversity. It is, therefore, essential that all potential risk factors (especially host range and interactions with non-targets) are thoroughly evaluated under controlled conditions before release (Thomas and Willis, 1998).

A major aim for us should be to ensure that we are fully utilising the biological control that is naturally available to us in integrated crop management programmes.

New chemical opportunities

Knowledge of gene function and the production of new products from plants leads to new opportunities for synthetic chemicals and also an additional new level of meaning to the words "crop protection".

Synthetic chemicals will be applied to plants as foliar sprays or seed treatments to regulate gene function in a specific manner. This approach is already commercialised in a small number of product types where plant resistance is induced, e.g. to pathogens by acibenzolar (Ruess *et al.*, 1996) or to certain herbicides using 'safeners' (Davies and Caseley, 1999).

Chemical switches will be needed to regulate the manufacture of novel products in plants. These may be protein-based products for pharmaceuticals (including vaccines or antibodies). The crop itself may need protection from side effects of synthesising a large quantity of foreign product until a certain stage in its life cycle. Or it may be advantageous to control the activation of a transgenic trait (e.g. an insecticidal principle) to certain time points during the season or when certain pest thresholds are met. The use of such chemical regulation may also help to alleviate concerns about the uncontrolled release of transgenic products into the environment, since the product would only be produced when the chemical is applied. Several of the available approaches are still speculative and will need validation at the biochemical, organism and field levels.

A NEW RELATIONSHIP BETWEEN FIELD AND LOBAROTRY BENCH

In this review we have seen the palette of new technologies available to crop protection scientists. This is the push compelling us to change. There is also a strong pull effect from new customer and consumer demands in the marketplace. To exploit the opportunities generated by this combination of factors, we are not only moving the interfaces between glasshouse and field, but also re-engineering the processes which bring us new products.

Integrating new knowledge

Crop protection discovery was conventionally thought of as a linear sequential process (Fig 2). A series of sequential screening steps selects out molecules for further evaluation. Selection criteria are based on primary biological activity and the information per molecule increases as the number of molecules per step declines. Such a process could now be extended backwards with target-based screens to increase the capacity of the first sieves. However, this misses an opportunity to use not only more information, but also different kinds of information earlier in the discovery process than ever before. Using our new technologies, we should be able to

establish key product characteristics pre-field (including essential toxicological and environmental parameters).

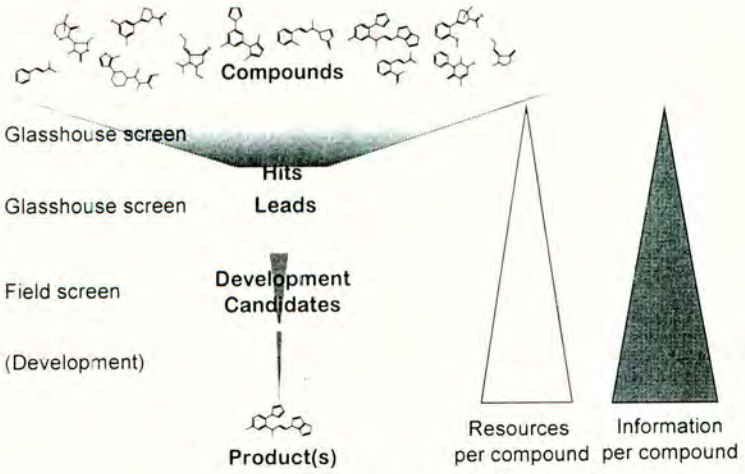


Figure 2. The conventional crop protection discovery process

New demands on field testing

Our crop protection products are components of integrated crop management systems. Their use, therefore, needs to be fine tuned in order to maximise the benefits they will bring within this framework:

- Integration of chemicals in mixtures or rotations
- Fit of application timing and dose to the pest situation
- Integration of chemical, genetic, biological and cultural pest management methods
- Adaptation of use recommendations to cultivar, environment, yield and financial expectations
- Link to decision support systems in precision farming.

The legislative framework for product registration has also evolved to greater complexity. The more stringent requirements in risk assessment for human and environmental safety as well as the need for efficacy testing in Europe also increase demands for field resources.

Major aims of our field experimentation will, therefore, be systems optimisation and product registration. In order to use our expensive and limited field resources effectively, we will depend heavily on the reliability of comprehensive laboratory and glasshouse evaluation for knowledge of the basic characteristics of the molecule.

A new process for crop protection discovery

The interaction between the availability of new knowledge and the demand of the marketplace can stimulate new dynamics in crop protection discovery. A linear process becomes

insufficiently responsive to change. The interaction between functional units becomes more flexible based on expertise from our key technology platforms (Fig 3). Important components of this iterative process are:

- A strong project orientation to maintain momentum towards a market product plus a system of portfolio management for the allocation of resources across multifunctional work areas
- Multidisciplinary project teams comprising representatives of the technology platforms to facilitate communication.

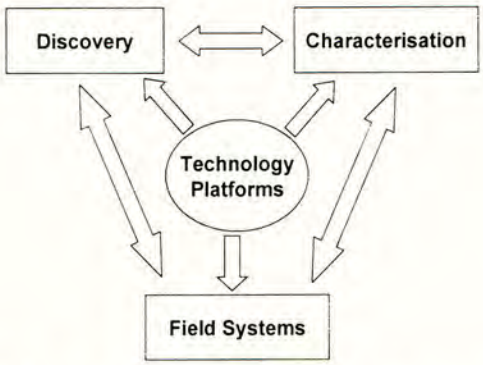


Figure 3. Interactions in the crop protection discovery process

CONCLUSIONS

Crop protection scientists have many fascinating years ahead of them. We are participating in fast technical and market change. This generates new opportunities for us, but also leaves important knowledge gaps which need urgent action:

- The connection between genomics and the phenotype in the glasshouse and in the fluctuating field environment
- The mechanism of interactions between xenobiotics and target organisms to allow prediction of effects e.g. uptake from knowledge of physico-chemical properties
- Bioinformatics platforms and predictive models to unify the components of biological systems.

The exploitation of these technological opportunities also requires that we have excellent skilful people working in interdisciplinary project teams towards realistic goals.

We are inundated with new chemical and biological information. With the sequencing and functional analysis of the major crop genomes over the next 2-3 years we shall have more biological data than in the previous 20 years. Our challenge is to utilise this information in a cost-effective manner to create product-relevant knowledge.

The primary aim of our business strategies is changing both its product and customer focus. We have been highly successful in marketing pesticides, but now, in an increasingly competitive world, we are looking to the establishment of integrated crop packages for processors and retailers. Our product offer increases in complexity and needs much more technical support through detailed field experimentation. This reduces capacity available for

simple product screening. We urgently need to match resource allocation across the discovery process with these new areas of supply and demand for knowledge. Robust and reliable testing in laboratory and glasshouse will ensure the field capacity is used where it is needed – for crop protection optimisation within a consumer-oriented production system.

Our future is sure to be fascinating, but perhaps every vision has a downside. If we are successful, then by our next meeting, crop protection biology will have become a branch of computer science.

ACKNOWLEDGEMENT

I am grateful to Dr Adrian Friedmann for his useful input on xenobiotic uptake.

REFERENCES

Boesten J J T I (2000). From Laboratory to field: uses and limitations of pesticide behaviour models for the soil/plant system. *Weed Research* **40(1)**, 123-38

Bromilow RH and Chamberlain K (1995). Principles governing uptake and transport of chemicals. In: *Plant Contamination: Modelling and Simulation*, S Trapp and JC Farlane eds, Boca Raton, Lewis Pub.; CRC Press, 37-68.

Davies J; Caseley JC (1999). Herbicide safeners: a review. *Pesticide Science* **55**, 1043-58.

Dove A (1999). Drug screening – beyond the bottleneck. *Nature Biotechnology* **17**, 859-63

Harrison W (1999). Agrochemical research and development in the 21st century: bloom or bust. *Journal of Biomolecular Screening* **4(2)**, 61-5.

Kirkwood R C (1999). Recent developments in our understanding of the plant cuticle as a barrier to the foliar uptake of pesticides. *Pesticide Science* **55**, 69-77.

Mackill D J (1999). *Genome analysis and breeding in Molecular Biology of Rice*, ed. K. Shimamoto, Springer, 17-42

Oliver S G (1998). Systematic functional analysis of the yeast genome. In: *Trends in Biotechnology* **16**, 373-8

Ruess W; Mueller K; Knauf-Beiter G; Staub T (1996). Plant activator CGA245704: an innovative approach for disease control in cereals and tobacco. *Brighton Crop Protection Conference – Pests and Diseases*, **1**, 53-60.

Thomas M B; Willis A J (1998). Biocontrol - risky, but necessary. *Trends in Ecology and Evolution* **13**, 325-9

Thornton J M (1998). The future of bioinformatics. In: *Trends Guide to Bioinformatics*, Elsevier Science, 30-2

Uetz P; Giot L; Cagney G; Mansfield T A; Judson R S; Knight J R; Lockshon D; Narajan V; Srinivasan M; Pochart P; Oureshi-Emili A; Li Y, Godwin B; Conover D; Kalbfleich T; Vijayadamodar G; Yang M; Johnston M; Fields S; Rothberg J M (2000). A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623-7

Zabkiewicz J A (2000). Adjuvants and herbicidal efficacy - present status and future prospects. *Weed Research* **40(1)**, 139-49

The evolution of the statistician : new trends for the 21st century

D J Pike
AB Statistics International & Department of Applied Statistics, The University of Reading, P.O.Box 240., Whiteknights Road, Reading RG6 6FN, UK

ABSTRACT

This paper will focus on the conjecture that ***"Statisticians are redundant today, and numerate chemists and biologists are much more cost-effective employees".*** It will acknowledge that this is a perfectly logical assertion, and will draw on 35 years experience to outline why it might be so. It will, however, illustrate that this *logical assertion* may be based on a misunderstanding of the potential of the statistician, which has existed since science began. The statistician, of course, has failed in more than equal measure to make the case for the Defence!

The positive side of things is that evolution and natural selection operate in the Statistical World as well as in the Natural World. The paper will develop and illustrate the importance to industry (much more widely than individuals may think) of a modern statistician, equipped with modern tools and training who can play a key team role in the advancement of today's scientific challenges.

INTRODUCTION

It is clear that there is a need for someone to grasp the nettle and ask "Does industry need statisticians any longer?" I do this because I believe with passion that the answer is "Yes"; but I am acutely aware of three things:

a. antipathy on the part of senior management to spend money in an area which has come to be regarded as a *support service,*
b. implicit belief on the part of today's new graduate scientists that they can do it all themselves – in *Excel,*
c. failure by the statistical profession to evolve in a way which retains the trust and confidence of the scientists it has served for so long.

All I contend are wrong, but it is probably not their fault. The problem is the misconception that statistics is only analysis of data - and here I blame the statisticians too. If it were so, I would confirm the view and retire! But not yet!!

Let me illustrate the case for the Prosecution with the data set shown in Table 1. This presents artificial data on crop yield in a weedy environment at two levels of application of a herbicide. The crop, the yield units and the herbicide are irrelevant. The key questions are :

1. What would you like to do with the data?
2. How would you do it?
3. What do the results mean?

Table 1 : Artificial data set on Crop Yield

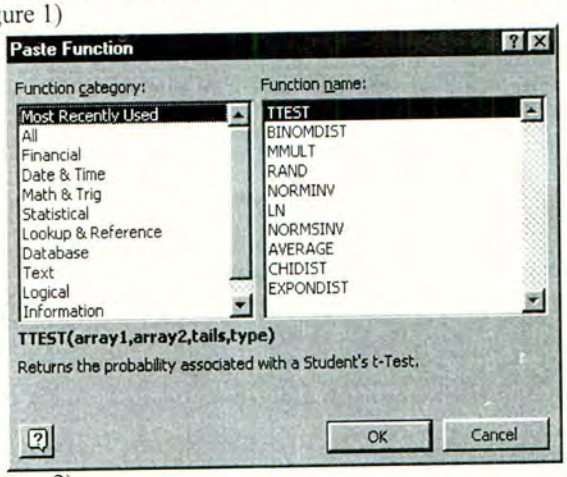
Replicate Number	1	2	3	4	5	6	7
Low Herbicide Level	4.6	5.3	4.2	6.1	3.9	4.7	4.9
High Herbicide Level	5.1	5.8	5.5	6.3	6.1	5.3	5.7

When I was a young statistician, 35 years ago, I was an indispensable part of this equation. The answers to the questions were :

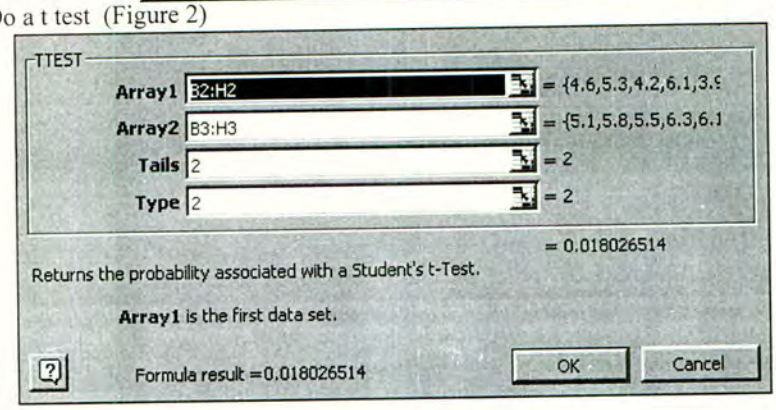
1. Is there a significant difference?
2. Do a 2-sample t test
3. Use a calculating machine
   Compute the two mean values *(you can probably remember the formula)*
   Compute the two sample variances *(can you remember the formula?)*
   Pool the variances to produce a combined estimate of the variance *(pardon?)*
   Calculate the t value using the appropriate formula *(!)*
   Look up the significance level in a set of tables *(!)*
4. Provide an interpretation in terms which the experimenter can understand

Do I have a role today? I am aware that today's discussion would go something like this.

1. Use Excel (Figure 1)



2. Do a t test (Figure 2)



3. If P is less than 0.05 it is significant - and here P=0.018.

In this example there are immediately two potential issues (or perhaps there are three) and they can also be related to the image and attitude of the senior manager to whom the experimental scientist reports. I offer two managerial images:

i. *Don't bother me with facts, my mind's made up".*
   A lost cause for all of us!
ii. *OK! Lots of information, but what does it all mean?*
   Remember that information does not equate to knowledge!

The issues are:

a. The result is as given: $P < 0.02$.
   Manager (i) may or may not be happy
   Manager (ii) wants an interpretation
      What did you test?
      One tailed or two?
      Were the variances equivalent?
      What is your conclusion?
         AND
      How does the statistical procedure you have carried out match up to your objectives?

b. The result is given as: $P < 0.12$.
   The consequences are unchanged except that Manager (ii) might ask For advice on future sample sizes!!

c. The result is given as: $P < 0.051$ (or 0.049)
   Do you have a view – or an answer?

I hope that this part of the thesis I am presenting is clear – even if it was not before! If it works here, then it also works for analysis of variance, for logistic regression, for nonlinear modelling, for hierarchical cluster analysis - and so on. The computer software (and hardware) revolution has led to both science and statistics becoming technique driven rather than objective driven.

The issue we need to address is whether we are missing anything critical on the way along, and how we might do better with the application of vision and understanding.

THE PRESENT SITUATION

Information today is easy to collect (by the megabyte) and the challenge has become to find an appropriate statistical technique to process the data. My nervousness is that:

- Information does not equate to knowledge
- Clever techniques can easily produce answers – to questions which have not been asked
- What we may be missing is the precise definition of objectives and
   i. Cost-effective study design to meet these objectives
   ii. Objective driven interpretation of results

So, to recap, analysis is becoming technique based, and the mathematicians and statisticians have begun to dig their own graves, encouraged by the available computing power. So now we have a new generation of buzz words – Neural Networks, Fuzzy Logic, Image Analysis, Partial Least Squares, Empirical Bayes, Markov Chain Monte Carlo, and so on. As a consequence, computer literate scientists, bewildered by the new jargon of the born-again statistician, can legitimately be forgiven for going it alone; but perhaps cannot be forgiven for creating their own jargon which is in turn unintelligible to the practical and interested statistician!!

My plea is two-fold:

To statisticians I say, heed the wisdom of David Finney (2000)

> *"I suspect that professional statisticians may be losing all control of – perhaps even losing all concern for – what is done in the name of our discipline? ..... Any road back will be long and tortuous, but unless we find it we fail to keep faith with the lead that giants gave us 75 years ago."*

To scientists and managers I say, remember that there are three phases to all studies:

- study design and model definition,
- data collection and analysis,
- interpretation of results,

and that there are statisticians out there who are keen and able to fulfil key roles in helping to provide effective advice on all three aspects.

A FUTURE SCENARIO

Strangely enough, what I advocate is to step back in order to move forward, and to remember that in today's commercial world cost-effective research is an important consideration. I offer four examples.

1. Factorial Experimentation.

The Society of Chemical Industry and the British Crop Protection Council have for more than 15 years now been promoting conferences on the prediction of Field Performance in Crop Protection. To date many questions seem to have been raised, but relatively few answers have emerged. Why might this be, when my logic at least feels it should be otherwise? I can propose two scenarios to initiate the debate:

- there **is** no predictive ability and we are deluding ourselves to search,
- we are missing one or more key factors in our studies.

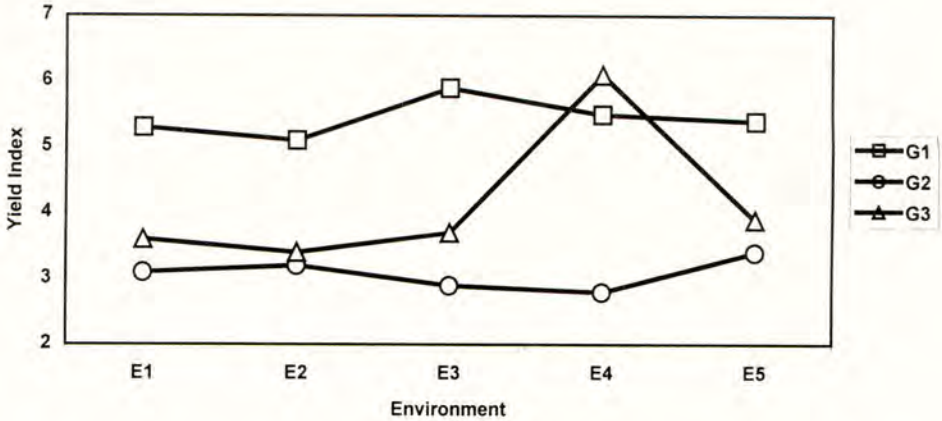
Common sense tells me to reject the first.

Plant breeders have long had a similar problem and seem to have made some progress through the study of Genotype*Environment interaction. Figure 3 shows the performance of three genotypes over 5 environments, where G1 is stable & high yielding, G2 is stable & low yielding and G3 is a niche genotype only to be recommended for environment 4. Logic suggests there should be a parallel for crop protection scientists from identifying potential

environmental factors of importance, experimenting with all factors together, and isolating:

- factors of consistent importance
- factors of no importance
- factors which interact

Figure 3. Yield Performance of 3 Genotypes at 5 Environments



From this study there could well be a way forward using a study based on factorial designs first proposed by Yates (1935) for investigating joint effects of fertiliser combinations. Yates realised, and showed, that experimentation with one factor at a time was a disaster if there were any factor interactions; but surprisingly the experimental precision of factorial experiments was greatest when there were no such interactions. This was perhaps the first such clear example of a *win-win* scenario.

2. New Product Discovery.

In some ways scientific advance and computer speed seem to have removed the demand for thought in scientific investigation – all we need now is the sledgehammer!

Consider the problem of screening and discovery based on combinatorial chemistry. The former approach of developing a careful design strategy to hit the target quicker (and perhaps to understand why) can now be replaced by a philosophy of *"to find the needle, put the haystack through the screen."*

This works well if the objective is simply to find the winner, and cost is not a problem. Frequently, however, production of new prototypes can be expensive. In such a situation the objective of experimentation is to be able to select a subset of the haystack to screen and, from the analysis, to learn what might drive success and to decide where to look next. Industrial statisticians have been able to advise successfully for many years on strategies for this type of study (Addelman 1963,1969; John, 1962, 1966, 1969; Pike, 1982). I make the point that recent development of sophisticated software procedures does not replace the building blocks on which sound science is based. Thought and the nature of the foundations still have a role to fulfil.

The strategy involves the careful design of appropriate fractions of factorial designs, and then augmenting them in the light of analysis of the initial results. Pike (1982) shows how this can be used to study combinations of up to 15 factors using less than 80 products, starting with a set of 32. Pike (2000) gives a more recent, and highly topical, example involving 8 factors.

3. Product Formulation.

I have long been concerned about the design and analysis of formulation experiments, where the challenge is to get the blend of Active, Oil and Wetter correct, so as to achieve the maximum synergy between the three with the minimum antagonism. This area is another where the problems are well understood and some solutions well established, both in the petroleum industry (Scheffé, 1958) and by subsistence farmers in Africa and India who achieve their synergy through intercropping. The mixtures designs pioneered by Scheffé have immediate application in formulation research. Actually the problem in practice is a little more complex, since it is required to optimise the relative volumes of Active, Oil and Wetter, and also the concentration of the active. Thus, even for a single product we are looking for a combination of a mixture design and a design to optimise the level of a quantitative factor. You do not need to stretch the imagination too far to see the potential minefield when you consider tank mixing!

I would suggest that to attempt to optimise formulation without some focused statistical advice is a bit like trying to cross that minefield without a metal detector.

4. The Use of Biological Control Agents.

I am also keen to see interest develop in the use of biological control agents, which I believe have the potential for synergy when mixed with reduced concentrations of synthetic chemicals. It sounds a bit like a *win-win* commercial scenario – the chance to charge a premium for a product with less synthetic and an environmentally friendly tag!

However, to reach Nirvana leaves the need to jump through a few hoops – even after the selection of target pests and material to mix. How about Cleaver control in Oilseed Rape using fluroxypyr and *Phoma spp*, or reduced application of OrganoPhosphates to control pests in stored grain products?

The traditional story about synthetic chemicals (an this is an idealised example) is that they control pests (Figure 4), but that efficacy drops with time (Figure 5). The solution is either to advise spraying at a higher rate or to use repeat spraying. Biological control agents are generally less active than synthetics, but efficacy is maintained over a longer period of time (Figure 5). The challenge is to look for potential synergy in the mixture, with the synthetic at a lower rate (Figure 6).

From the statistical perspective there is a need for flexible mathematical models and objective oriented mixture designs related to commercial targets. The knowledge is sprinkled through almost 50 years of the published statistical literature; but this application is an exciting new challenge for the 21st Century statistician – if the industry is interested! I would also like to see postgraduate students encouraged to think, rather than follow their noses through yet another study of Markov Chain Monte Carlo methods!!

Figure 4. % Pest Control vs Rate

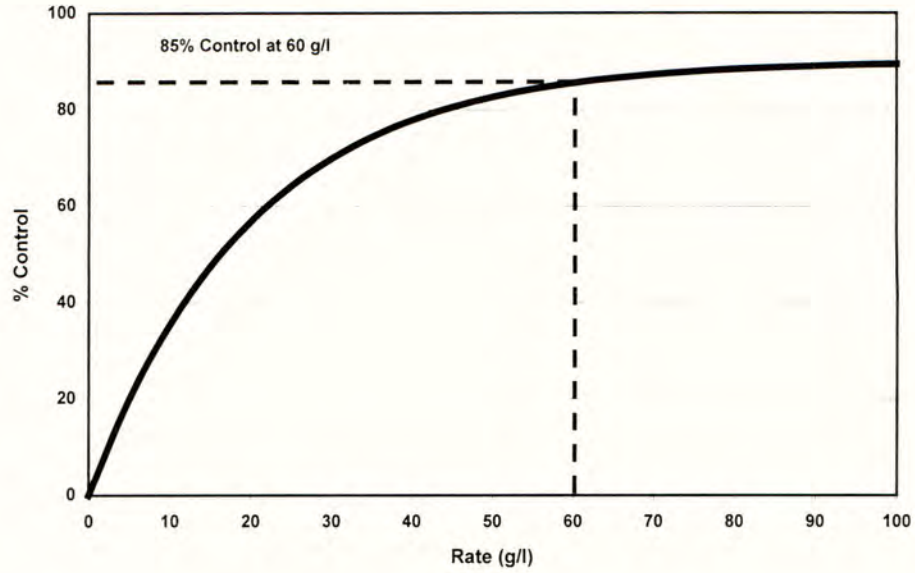


Figure 5. Control as a function of time for Synthetic and Biological.

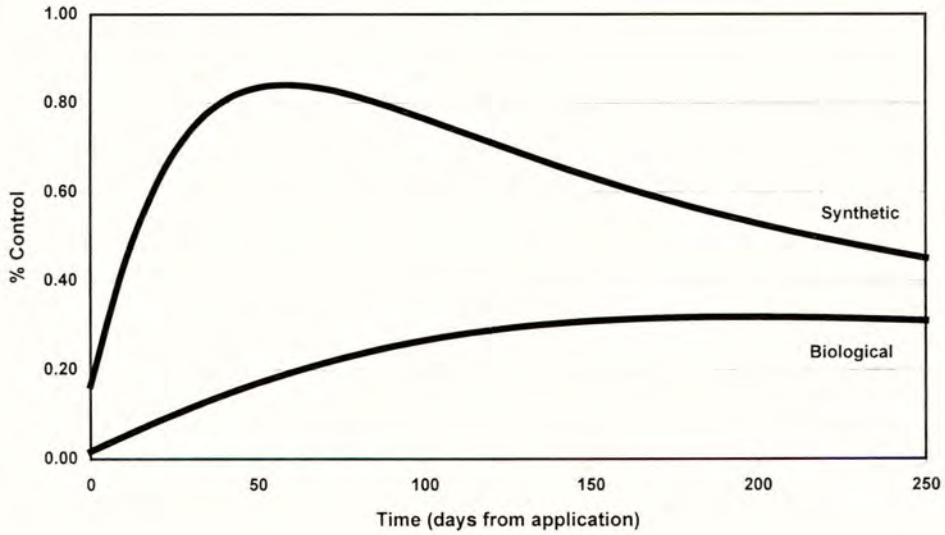
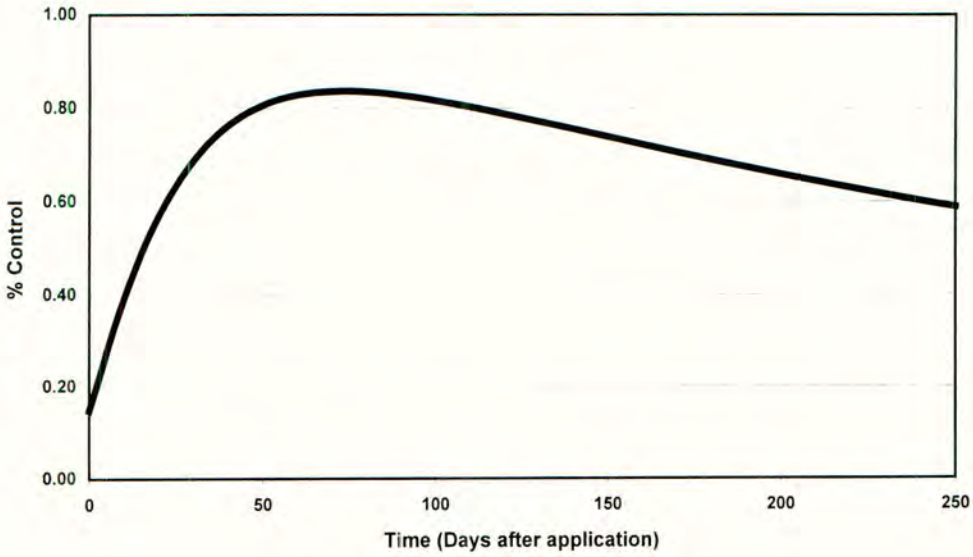


Figure 6. Potential Control from a mixture of Biological and Synthetic



5. The Impact of Biotechnology.

If we build on the philosophy of the last four sections, and consider potential biotechnological advance, it would seem logical to try to produce chemical mixtures to target combinations of genes which create appropriate phenotypic behaviour in plants, be it disease resistance, herbicide resistance or whatever.

Think about it. Analysis of Gene-Expression Microarray data (even if Newton *et al* (1999) is the best way to proceed) can potentially isolate chemicals which enhance or suppress genes which code for specific traits. How is such knowledge best utilised? The minute the word *"combination"* is mentioned the basis of the study is that of mixtures, of suitable relative concentrations and of potential interactions – not to mention environmental effects. To provide an efficient and cost-effective approach to the production and evaluation of such compounds, I would contend that a statistician is a key member of the team. And I would introduce the need for a further concept into the study – that of modelling.

Whenever a study involves quantitative factors - and we can consider here such aspects as concentration, time or temperature – effective product design and experimental design requires model definition as a first stage in the process. The *"suck it and see"* approach is still an option, but with the risk that the taste will be unpalatable.

A suitable design strategy here involves:

- model the likely effect of individual chemicals, possibly as a function of concentration, and other environmental factors,

- model the potential effect when such chemicals are combined, to take into account interaction in the form of synergy or antagonism,
- a model contains parameters to be estimated, some of which will be of critical importance in the light of the objectives of the study,
- design multifactor mixtures experiments to provide precise estimates of the most important parameters of the study,
- interpret the results of the study to postulate the form of the *optimal* combination, and predict its likely performance.

WHERE DO WE GO FROM HERE?

Both statisticians and scientists need to remember that Confucius is reputed to have said,

"He who would perfect his work must first sharpen his tools".

For the statistician this means remembering that most of the major statistical advances of the last century were made by mathematicians operating as biologists, geneticists, and industrial statisticians driven by a keen and practical interest in the problems not the theory. The theory evolved to solve the problems. There is a need to get back to basics.

For the scientist it means recognising that the most fruitful scientific advances are generally made through team collaborations of appropriate experts. There needs to be space in your team for the statistician.

ACKNOWLEDGEMENTS

I am very grateful to Geoff Hewitt for what seems like half a lifetime of fruitful discussion about topics like this. To him it must seem longer, having had to cope with my agrochemical naivety!

REFERENCES

Addelman S (1963) Techniques for constructing fractional replicate plans. *J.Amer.Statist.Assoc.* **58,** 45-71

Addelman S (1969) Sequences of two-level fractional factorial plans. *Technometrics* **11,** 477-509

Finney D J (2000) Whither ... goes now the practice of statistical science? *RSS News* **27/7,** 1-2

John P W M (1962) Three quarter replicates of 2^n designs. *Biometrics* **18,** 172-184

John P W M (1966) Augmenting 2^{n-1} designs *Techgnometrics* **8,** 469-480

John P W M (1969) Some non-orthogonal fractions of 2^n designs *J.Roy.Statist.Soc.* **B31,** 270-275

Newton M A; Kendziorski C M; Richmond C S; Blattner F R; Tsui K W (1999) On differential variability of expression ratios: Improving statistical inference about gene expression changes from microarray data *Technical Report #139, Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison*

Pike D J; Smith J R (1982) Priority and compromise in fractional factorial designs. *Proc.XI^th Int. Biometric Conf.* 51-57

Pike D J (2000) Experimental Design for new product development *Paper presented at RSS 2000 Conference, Reading 13-15 September 2000*

Scheffé H (1958) Experiments with mixtures *J.Roy.Statist.Soc.* **B20,** 344-360

Yates F (1935) Complex experiments (with discussion) *J.Roy.Statist.Soc.* **B2,** 181-247