



PROJECT REPORT No. 263

**DEVELOPMENT OF WEB-BASED SOFTWARE FOR IDENTIFYING
ADAPTABILITY OF CEREAL VARIETIES TO LOCAL CONDITIONS**

OCTOBER 2001

Price £4.00

PROJECT REPORT No. 263

DEVELOPMENT OF WEB-BASED SOFTWARE FOR IDENTIFYING ADAPTABILITY OF CEREAL VARIETIES TO LOCAL CONDITIONS

by

M TALBOT

**Biomathematics & Statistics Scotland JCMB,
The King's Buildings, Edinburgh EH9 3JZ**

This is the final report of a two-year project that started in October 1998. The work was funded by a grant of £99,856 from HGCA (project no. 1966)

The Home-Grown Cereals Authority (HGCA) has provided funding for this project but has not conducted the research or written this report. While the authors have worked on the best information available to them, neither HGCA nor the authors shall in any event be liable for any loss, damage or injury howsoever suffered directly or indirectly in relation to the report or the research on which it is based.

Reference herein to trade names and proprietary products without stating that they are protected does not imply that they may be regarded as unprotected and thus free for general use. No endorsement of named products is intended nor is any criticism implied of other alternative, but unnamed products.

CONTENTS

	Page
Abstract	1
Summary	2
• Rationale	2
• Objectives	3
• Methodology - an overview	3
• The Varadapt System	4
• Conclusions and Implications	5
Technical Detail	7
• Introduction	7
• Data Analysis	8
• The Alogrithms	9
• Implementing the Algorithms	11
References	15
Appendix A1 : User documentation	16
Appendix A2 : Programmer Documentation	26
Appendix A3 : Data format	36

ABSTRACT

Data from variety trials provide national or regional estimates of variety performance and these averages represent information that is widely applicable. However the averages can conceal differences in variety responses, either to permanent features of individual fields, e.g. their location, soil type, PK status; or to management factors which may be under farmers' control, e.g. sowing date, manuring rate.

Estimation for local environments is not attempted for several reasons, but one of the main reasons is analytical complexity. Problems arise from:

- Trials data that are heavily influenced by seasonal and location effects, and which also involve the introduction of new varieties and the discarding of old varieties over time;
- The strong association that often occurs in trials between environmental factors, such as between soil N status and the rate of applied N, whose confounding effects are difficult to disentangle;
- The uncertainty about conditions in the local environment where variety recommendations are needed, e.g. the information about the soil N status of a field is likely to be imprecise;
- The absence of effective methods for putting all of the information together as useful decision tools for farmers.

Developments in statistics and computing e.g. Bayesian inference and Markov chain Monte Carlo computational methods, suggest that some of these problems can now be overcome. The aim of this project has been to apply these techniques to produce tools that will help growers interpret data from variety trials in order to identify varieties that may be particularly adapted to specific circumstances.

The algorithms have been implemented as Web-based programs using the Java language. A software package, named Varadapt, has been made available for access across the Internet at <http://www.bioss.ac.uk/varadapt>. Currently yield data from more than 300 winter wheat and spring barley trials for the years 1994-2000 are available for exploration via the website.

Varadapt provides users with the tools to explore interactively, from their own computers, all of the yield data collected in HGCA-sponsored variety trials. Specifically, it allows users to describe their conditions while the software seeks for evidence of adaptability, or otherwise, to those circumstances.

The small amount of yield information currently presented in published lists of recommended varieties represents both an extremely restricted summary of an extensive body of data as well as a limited picture of a variety's potential. This project produces methods for more fully exploiting the costly information collected and for helping farmers get the best performance from varieties.

SUMMARY

Rationale

Plant breeders aim to produce varieties that are adapted to a wide range of conditions and a measure of their success is the consistency with which varieties perform relative to each other in practice. Nevertheless, some varieties are better suited to one set of conditions than to others. An investigation showed that the size of differences in yields of barley and wheat varieties between 11 regions of the UK was approximately one-third the average difference between varieties nationally (Talbot, 2001). Since the regional boundaries in this study were chosen largely for administrative reasons it seems likely that the true scale of differences in variety performance between agronomically diverse regions may be even greater. Other investigations have shown differences in the response of winter wheat varieties to factors such as sowing date, and of barley varieties to nitrogen fertilizer.

Current practice recognizes these differences in variety performance by, for example, presenting in the NIAB Cereal Variety Handbook summaries of variety performance grouped by region, rotational position, soil type, time of sowing, and site fertility. However there are a number of difficulties with this form of presentation:

- The groupings are too broad to be of practical use to growers; for example, the North East region covers a distance of some 300 miles from north to south.
- It is difficult for growers to relate the results to their own specific circumstances; for example, a grower in the south-west region wishing to sow late will have difficulty in usefully combining the information from the relevant tables.

Estimation of a variety's potential for specific conditions has not been attempted in the past for a number of reasons:

- The structure of the trials' data is complex, being strongly influenced by seasonal and location effects, and with varieties not grown in all trials;
- Strong correlation between environmental factors often occur, such as between soil N status and the rate of applied N, whose confounding effects on varieties are difficult to disentangle;
- The uncertainty about conditions in the local environment where variety recommendations are needed, e.g. the soil type;
- The absence of effective methods for putting all of the information together as useful decision tools for farmers.

While seasonal difficulties remain, developments in statistical methods and computer technology suggest that problems with the handling of uncertainty and the combining of information from several sources can be overcome. Bayesian inference linked to Markov chain Monte Carlo computational methods can model the situation described above, and can also incorporate in a coherent way the grower's knowledge of local conditions as well as their likely effect on variety performance.

Developments in web technology also make it possible to provide users with interactive access to variety trials databases where they can pose questions on-line and interactively gain answers from exploratory analyses.

Objectives

The aim of this project is to bring together some of these developments and produce tools that help farmers interpret information from variety trials. The two main goals are:

- To produce algorithms for estimating from trials data the extent to which a variety's yields may differ from its average UK performance in specific circumstances;
- To implement these algorithms as practical tools for use by growers.

A broader strategic aim is to build a bridge between the grower's need for specific targeted information and the crop agronomist's concern that trial information may not always be robust enough to support such advice. The computational tools can provide independent measures of the strength of the case for adaptation of varieties to specific conditions.

Methodology - an overview

The statistical approach adopted is that of Bayesian multiple regression modelling with measurement errors. Further information on this approach is given in the TECHNICAL DETAIL section of the report, and also in a paper by Theobald, Talbot & Nabugoomu (2001).

The statistical methods have been incorporated as Java applets in the web-based Varadapt (Crop Variety Adaptation to Specific Conditions) program. The program uses information from the inquirer about his/her circumstances and links this to:

- yield results from variety trials,
- and details of the circumstances under which the trials were grown.

Varadapt then predict the extent to which varieties may be more or less adaptable to these specific conditions.

The heart of Varadapt is a Bayesian modelling approach which combines the data and the uncertainty associated with the data along with the user's information on local conditions.

The objective is to provide algorithms which, using the various data sources, dynamically identify useful relationships and predict possible adaptation of varieties to local circumstances. The tasks are to:

- Discover which are the most influential covariables and produce predictive model;
- Assess users' uncertainty about local circumstances;
- Predict expected response taking into account uncertainty about local conditions.

Covariables can include continuous measures such as soil organic matter, altitude, and may also include classificatory measures such as soil type.

There are many potential covariables of which only one or two may be relevant for a particular variety or location.

The Varadapt System

From the user's viewpoint Varadapt consists of a series of linked interactive WWW pages. The opening pages allow the user to choose a crop and select a variety for investigation. Succeeding pages provide facilities to:

- view a map showing the relative performance of the variety in trials throughout the UK over the past five years;
- specify the location of a farm, by using a clickable UK map, and see the yields of the varieties at the closest trial sites to that farm;
- explore the adaptation of a variety to various environmental and management factors including location, soil conditions, sowing date.

In addition to these exploratory tools Varadapt also allows users to specify combinations of factors that they think may be important in their circumstances and to search the database to see if there is evidence of adaptability to those conditions. It is at this stage that the Bayesian modelling procedures come into play.

Details of the input and output from all of the web pages are provided in Appendix A1. Programmer documentation is in Appendix A2. The data from variety trials used by Varadapt is organised in two database files and an example of the structure of these files are given in Appendix A3.

The software has been made available for access by users across the Internet. at <http://www.bioss.ac.uk/varadapt>. Currently yield data are available from winter wheat and spring barley variety trials sown in the UK over the period 1994-2000, some 600 trials in total. The software allows users to interactively explore variety trials databases to assess if a variety may be adapted to their individual circumstances.

Conclusions and Implications

Varadapt has been tested by crop specialists, plant breeders and a few growers. Tools for capturing feedback from users are incorporated in the software and these have produced supportive comments along with useful suggestions for improvements. Appropriate changes have been made in response to these suggestions.

There are a number of constraints that may limit the usefulness of the Varadapt approach. One is the restricted number of useful trial characteristics that are currently recorded as routine measures. For example, characters such as soil moisture availability, if recorded routinely, might be valuable in the longer term. Attention may need to be given in future to extending the range of crop management conditions recorded from variety trials.

A further important constraint is imposed by the extent to which varieties vary in their performance from year-to-year in response to seasonal factors. When we attempt to predict what a variety will achieve at a particular location in a future year it is difficult to build this seasonal uncertainty into our estimates of predictive precision. So, for example, if a variety yields 2% below its long-term average in a particular year because of seasonal effects, and Varadapt predicts that its yield under certain conditions will be 2% above average then this is a risk that cannot be avoided. The best that can be done is to identify permanent features of the environment that are affecting variety performance in the longer term. And, of course, this risk also applies to Recommended List relative yield figures.

The main benefits for the industry from the Varadapt approach are likely to come from:

- Ensuring that more useful information is extracted from the data being collected in the HGCA-funded variety evaluation programme;
- Helping the industry to identify at an early stage the conditions under which a variety is likely to perform at its best;
- The presentation of results in a manner which the users can translate to their own circumstances. In this sense Varadapt is a prototype of how results from the HGCA R&D programme may be delivered to levy-payers in future.

TECHNICAL DETAIL

Introduction

New crop varieties are evaluated by testing them alongside established varieties in trials distributed throughout the region in which they are to be eventually grown by farmers. Typically, the results from such trials are combined to produce measures of average variety performance. These averages are the basis on which varieties are recommended for growing in the whole region. However, the regions can be heterogeneous, e.g. with differing levels of soil fertility, or varying in earliness of season, and varieties will often respond differentially to such covariates. As a consequence, reliance on over-trials averages may not be most efficient when selecting varieties in all parts of the region. The question then arises: if farmers can identify, with some level of uncertainty, the position of their farms on the scales on which heterogeneity is measured, can we then use trials data to provide a *local-area* predictor of likely variety performance in their conditions which is better than the *regional* performance average?

The use of environmental covariates to analyse variety x environment data sets in order to estimate relative variety performance in target locations is considered by several workers, e.g. Freeman and Perkins (1971), Hardwick and Wood (1972), van Eeuwijk, Denis and Kang (1996) and Piepho, Denis and van Eeuwijk (1998). Their approaches largely concentrate on dealing with complete variety x environment data sets, with only a single error component, though Piepho *et al.* (1998) include random effects for locations and years, and also allow several quantitative and qualitative covariates.

This report describes a Bayesian approach to the more typical situation of a variety x location x year data set which is incomplete and includes one or more location x year covariates. This approach incorporates expert knowledge of the crop, and takes account of possible uncertainty in the values of the covariates at the location where the crop is to be grown.

The work of implementing the methods so that they can operate from web browsers has been a major task of this project. It was not practical to use a statistical package such as BUGS (Spiegelhalter, Thomas, Best and Gilks 1996, 1997) for routine computation of the required estimates. We had to develop the algorithms from scratch as Java programs. The main difficulty to be overcome has been to ensure that the computationally intensive algorithms deliver answers that are:

- tailored to the user's circumstances;
- providing a precision that is appropriate to the situation;
- within a timescale that is acceptable.

Meeting these standards required considerable experimentation before arriving at a solution that was robust, practical and, at the same time, rigorous.

Here are described details of the data analysis and of the algorithms as they have been implemented.

Data analysis

The variety trials data that are used to determine whether a variety may be adapted to specific conditions have two main components:

- the trial details including, for each trial, the location (i.e. National Grid reference), elevation of trial site, previous crop, soil texture, drainage, nutrient status and sowing date;
- a variety-by-trials table of dry matter grain yields.

Before presentation to the Java programs the data are analysed. A REML (Residual Maximum Likelihood) (Patterson 1997) analysis is carried out to estimate the mean yield of a trial after adjusting for differences in the varieties that were sown between trials. For each variety its yield in a trial is expressed as a percent of the mean yield of the trial. These percentage figures represent the amount by which varieties perform better or worse than average in that particular trial. It is these data which are used by the Java programs to estimate variety adaptability to specific conditions.

Appendix A1 contains an extract from part of a typical dataset as used by the Java programs.

The Algorithms

i) The case of no classificatory covariates

Suppose that we have available measurements of p continuous covariates, perhaps including some missing values, at each of n sites. One of these sites will be the location at which a prediction of the yield of a given variety is required; the remainder will be sites of trials in which the variety was grown. Denote the measured value of covariate j at site i by $x^*(i, j)$. The asterisk distinguishes the measurement from the unknown 'true' value $x(i, j)$. We adopt the classical measurement error model in which $x^*(i, j)$ is stochastically dependent on $x(i, j)$. Given $x(i, j)$, the conditional distribution of $x^*(i, j)$ is Normal with mean $x(i, j)$ and precision $\tau_x(j)$. In turn, the 'true' covariate value $x(i, j)$ is assumed independent of all others and Normally distributed with mean $\theta_x(j)$ and precision $\kappa_x(j)$. Let the yields of the variety in question be $\{y(i); i = 1, \dots, n\}$, all of which but one (by convention $y(n)$) will be known. *A priori*, they are independently Normally distributed with $y(i)$ having mean

$$\alpha + \sum_{j=1}^p \beta(j)x(i, j)$$

and precision τ_y . This is the standard linear regression model with independent Normal residuals. The regression intercept α , the slopes $\{\beta(j); j = 1, \dots, p\}$ and the residual precision τ_y are assigned priors as follows:

$$\tau_y \sim \Gamma(s_y, \lambda_y) \quad \beta(j) \sim N(\theta_\beta(j), \kappa_\beta(j)) \quad \alpha \sim N(\theta_\alpha, \kappa_\alpha)$$

We could have chosen to treat the measurement error precision $\{\tau_x(j); j = 1, \dots, p\}$ as random rather than fixed, but we did not consider it worthwhile. The data $\{x^*(i, j), y(i); i = 1, \dots, n; j = 1, \dots, p\}$ contain no information about this precision, and furthermore we have no other informative observations, such as repeated measurements of the same covariate values. Therefore, we would have to select an informative prior placing high probability on relatively high values of precision, since the whole exercise depends crucially on the worth of the measured covariates. The alternative, to assume that the measurement precision of the available covariates is known, seems reasonable.

ii) Incorporating classificatory covariates

We now consider a regression involving a single classificatory covariate having l levels. The interpretation of the variables $x^*, x, \tau_x, \theta_x, \kappa_x, y, \tau_y, s_y$ and λ_y is as before. When the classificatory covariate takes level k , the appropriate regression intercept and slopes are $\alpha(k)$ and $\{\beta(k, j); j = 1, \dots, p\}$ respectively. The regression coefficients for each of the l levels are independent of each other and assigned common Normal priors:

$$\alpha(k) \sim N(\theta_\alpha, \kappa_\alpha); k = 1, \dots, l$$

$$\beta(k, j) \sim N(\theta_\beta(j), \kappa_\beta(j)); k = 1, \dots, l; j = 1, \dots, p$$

These priors were assigned to the single intercept and single slope for each continuous covariate in the previous model. Whereas the parameters of these priors were fixed beforehand, however, now we treat them as random and assign them hyper-priors as follows:

$$\theta_\alpha \sim N(\mu_\alpha, \tau_\alpha)$$

$$\kappa_\alpha \sim N(s_\alpha, \lambda_\alpha)$$

$$\theta_\beta(j) \sim N(\mu_\beta, \tau_\beta); j = 1, \dots, p$$

$$\kappa_\beta(j) \sim N(s_\beta, \lambda_\beta); j = 1, \dots, p$$

By treating the hyper-parameters of the common distributions of the regression coefficients as random rather than fixed, we allow the posteriors of the coefficients corresponding to a given level of the classificatory covariate to 'borrow strength' from each other through their joint influence on the hyper-parameters. The ratio of the number of trial sites n to the number of factor levels l is likely to be modest in practice, perhaps of the order of 10. Therefore, we need all trial sites to be informative to some degree about the regression coefficients corresponding to all factor levels. Simply treating the classificatory covariate as a random rather than fixed effect in the manner we have done provides a learning mechanism, albeit a weak and unstructured one, by means of 'prior shrinkage'.

The factor level at site i , given by $z(i)$, is assumed known without error at all $n - 1$ trial sites but subject to uncertainty at the location for which we must make a yield prediction. We shall return to this point in the next section. The yields $\{y(i); i = 1, \dots, n - 1\}$ are independently Normally distributed with precision τ_y and the mean of $y(i)$ being

$$\alpha(z(i)) + \sum_{j=1}^p \beta(z(i), j)x(i, j).$$

Although we have formulated a regression model to include a single classificatory covariate, typical variety trials data sets may have half a dozen or more classificatory covariates of potential value to yield prediction. In such circumstances we propose simply to cross all of these factors to form a single 'dummy' factor and treat as above. The alternative, to have an extra index for all regression coefficients specifying the level of each additional factor, is unwieldy. It also gives no advantage over the dummy factor approach without imposing more prior structure on the set of coefficients corresponding to each combination of factor levels. Given the generality of the methods it seems appropriate to avoid over-structuring our model.

Implementing the algorithms

i) Eliciting information from the program user

The user is asked to supply covariate information relating to the site for which he requires a yield prediction. For each of the continuous covariates in the regression model, he is asked to provide a measurement and an estimate of the associated precision. Should he decline to enter a precision for measured covariate j , say, then by default the precision of his measurement is taken to be that of the trial site measurements, $\tau_x(j)$. Otherwise his entry becomes the 'local' precision $\tau_l(j)$ of measured covariate j . Allowing local and trial site measurement precisions to differ is a minor but practically important generalisation of our regression model. In particular, if the user is unable to supply measurements for some of the covariates he is asked about then the corresponding local precisions can be set arbitrarily close to zero.

The program also elicits from the user a prior distribution on the local level of the 'dummy' classificatory covariate. For each 'real' factor in the model, the user is asked to specify a subjective prior on its level at his site. Prior probabilities on the dummy factor levels are calculated by independently combining the priors for each constituent factor. Given the likely associations between real factors, this last step may seem dubious. The obvious alternative however, to interrogate the user for subjective probabilities of all combinations of real factor levels, is far too lengthy.

ii) Posterior sampling by Markov chain Monte Carlo

The factorisation of the joint distribution of all variables in the model is represented as

$$\begin{aligned}
p(\theta_\alpha, \kappa_\alpha, \theta_\beta, \kappa_\beta, \alpha, \beta, x, x^*, \tau_y, y) &= p(\theta_\alpha) p(\kappa_\alpha) \prod_{j=1}^p p(\theta_\beta(j)) p(\kappa_\beta(j)) \prod_{k=1}^l p(\alpha(k) | \theta_\alpha, \kappa_\alpha) \\
&\prod_{j=1}^p \prod_{k=1}^l p(\beta(k, j) | \theta_\beta(j), \kappa_\beta(j)) \prod_{j=1}^p \prod_{i=1}^n p(x(i, j) | \theta_x(j), \kappa_x(j)) p(x^*(i, j) | x(i, j), \tau_x(j)) \\
&p(\tau_y) \prod_{i=1}^n p(y(i) | \alpha, \beta, x, z, \tau_y).
\end{aligned}$$

Our shorthand notation is obvious: $\theta_\beta = (\theta_\beta(j); j = 1, \dots, p)$ and so on. The conditional distribution of any subset of variables given all others (the 'full conditionals') can then be trivially read off as proportional to the product of those terms on the right-hand side of the above equation containing those variables. For instance, the full conditional for the mean θ_α of the common distribution of the regression intercepts is given by

$$\begin{aligned}
p(\theta_\alpha | \dots) &\propto p(\theta_\alpha) \prod_{k=1}^l p(\alpha(k) | \theta_\alpha, \kappa_\alpha) \\
&\propto \exp\left\{-\frac{\tau_\alpha}{2}(\theta_\alpha - \mu_\alpha)^2 - \frac{\kappa_\alpha}{2} \sum_{k=1}^l (\alpha(k) - \theta_\alpha)^2\right\}
\end{aligned}$$

Here and later, ' \dots ' indicates conditioning on all other variables. Expanding the terms on the right-hand side and completing the square in θ_α , we find that

$$\theta_\alpha | \dots \sim N\left((\tau_\alpha \mu_\alpha + \kappa_\alpha \sum_{k=1}^l \alpha(k)) / (\tau_\alpha + l \kappa_\alpha), \tau_\alpha + l \kappa_\alpha\right)$$

Conjugacy in the prior for θ_α thus allows a simple Gibbs move to be used in updating this parameter in the MCMC sampler. The same applies to the precision of the regression intercepts κ_α :

$$\begin{aligned}
p(\kappa_\alpha | \dots) &\propto p(\kappa_\alpha) \prod_{k=1}^l p(\alpha(k) | \theta_\alpha, \kappa_\alpha) \\
&\propto \kappa_\alpha^{s_\alpha + l/2 - 1} \exp\left\{-\kappa_\alpha (\lambda_\alpha + \sum_{k=1}^l (\alpha(k) - \theta_\alpha)^2 / 2)\right\} \\
&\Rightarrow \kappa_\alpha | \dots \sim \Gamma(s_\alpha + l/2, \lambda_\alpha + \sum_{k=1}^l (\alpha(k) - \theta_\alpha)^2 / 2)
\end{aligned}$$

Indeed, conjugacy has motivated the choice of all prior distributions in the model, enabling us to sample most of the unknowns directly from their full conditionals. In the same manner as for θ_α and κ_α above, a little algebra yields the following results:

$$\begin{aligned}
\theta_\beta(j) | \dots &\sim N\left((\tau_\beta \mu_\beta + \kappa_\beta(j) \sum_{k=1}^l \beta(k, j)) / (\tau_\beta + l \kappa_\beta(j)), \tau_\beta + l \kappa_\beta(j)\right); j = 1, \dots, p \\
\kappa_\beta(j) | \dots &\sim \Gamma\left(s_\beta + l/2, \lambda_\beta + \sum_{k=1}^l (\beta(k, j) - \theta_\beta(j))^2 / 2\right); j = 1, \dots, p \\
\alpha(k) | \dots &\sim N\left((\kappa_\alpha \theta_\alpha + \tau_y \sum_{i:z(i)=k} (y(i) - \sum_{j=1}^p \beta(k, j) x(i, j))) / (\kappa_\alpha + n_k \tau_y), \kappa_\alpha + n_k \tau_y\right); k = 1, \dots, l \\
\beta(k, j) | \dots &\sim N\left(\frac{\kappa_\beta(j) \theta_\beta(j) + \tau_y \sum_{i:z(i)=k} x(i, j) \left(y(i) - \alpha(k) - \sum_{l \neq j} \beta(k, l) x(i, l)\right)}{\kappa_\beta(j) + \tau_y \sum_{i:z(i)=k} x(i, j)^2}, \kappa_\beta(j) + \tau_y \sum_{i:z(i)=k} x(i, j)^2\right); \\
&k = 1, \dots, l; j = 1, \dots, p \\
x(i, j) | \dots &\sim N\left(\frac{\kappa_x(j) \theta_x(j) + \tau_x(j) x^*(i, j) + \tau_y \beta(z(i), j) \left(y(i) - \alpha(z(i)) - \sum_{l \neq j} \beta(z(i), l) x(i, l)\right)}{\kappa_x(j) + \tau_x(j) + \tau_y \beta(z(i), j)^2}, \kappa_x(j) + \tau_x(j) + \tau_y \beta(z(i), j)^2\right); \\
&i = 1, \dots, n; j = 1, \dots, p \\
\tau_y | \dots &\sim \Gamma\left(s_y + n/2, \lambda_y + \sum_{i=1}^n \left(y(i) - \alpha(z(i)) - \sum_{j=1}^p \beta(z(i), j) x(i, j)\right)^2 / 2\right) \\
y(n) | \dots &\sim N\left(\alpha(z(n)) + \sum_{j=1}^p \beta(z(n), j) x(n, j), \tau_y\right)
\end{aligned}$$

Here, n_k denotes the number of sites (including that of the program user) at which the dummy factor level is k . Should the user specify a local precision $\tau_l(j)$ for the measurement $x^*(n, j)$ of covariate j at his location, then in the expression above for the full conditional distribution of $x(n, j)$ the trial site precision $\tau_x(j)$ is replaced by the local precision $\tau_l(j)$. In the event that the measurement of covariate j is missing at site i , the full conditional for $x^*(i, j)$ has a particularly simple form:

$$\begin{aligned}
x^*(i, j) | \dots &\sim N(x(i, j), \tau_x(j)), i \neq n \\
x^*(n, j) | \dots &\sim N(x(n, j), \tau_l(j))
\end{aligned}$$

Should $x^*(n, j)$ be missing, $\tau_l(j)$ will be set to some small positive value by default.

It is only for the unknown dummy factor level at the location of interest, $z(n)$, that we resort to a Metropolis updating method. The full conditional for $z(n)$ is given by:

$$p(z(n) | \dots) \propto p(z(n)) \exp \left(-\frac{\tau_y}{2} \left(y(n) - \alpha(z(n)) - \sum_{j=1}^p \beta(z(n), j) x(n, j) \right)^2 \right)$$

Recall that the prior distribution of $z(n)$ is elicited from the user. In the updating step, the proposed new level z^* is drawn uniformly at random from the $l-1$ levels besides the present one. Since this is a symmetric proposal distribution, the Metropolis acceptance probability of this proposal simplifies to $\min(1, R)$, where

$$R = \frac{p(z^*)}{p(z(n))} \exp \left(-\frac{\tau_y}{2} \left(\left(y(n) - \alpha(z^*) - \sum_{j=1}^p \beta(z^*, j) x(n, j) \right)^2 - \left(y(n) - \alpha(z(n)) - \sum_{j=1}^p \beta(z(n), j) x(n, j) \right)^2 \right) \right)$$

A single iteration of the MCMC sampler consists of a systematic sweep through moves updating each unknown in turn. The order of updating is as follows:

- a) residual precision τ_y ,
 - b) parameters $(\theta_\alpha, \kappa_\alpha)$ of the distribution of the regression intercepts,
 - c) parameters $(\theta_\beta(j), \kappa_\beta(j); j = 1, \dots, p)$ of the distribution of the regression slopes,
 - d) regression intercepts $(\alpha(k); k = 1, \dots, l)$,
 - e) regression slopes $(\beta(k, j); k = 1, \dots, l; j = 1, \dots, p)$,
 - f) 'true' covariates $(x(i, j); i = 1, \dots, n; j = 1, \dots, p)$,
 - g) only in the case of missing values, the measured covariates $(x^*(i, j); i = 1, \dots, n; j = 1, \dots, p)$,
 - h) local dummy factor level $z(n)$,
- local yield $y(n)$.

All moves except h) are accomplished by Gibbs sampling

REFERENCES

- Freeman, G.H., and Perkins, J.M. (1971), "Environmental and Genotype-Environmental Components of Variability. VIII. Relations Between Genotypes Grown in Different Environments and Measures of these Environments", *Heredity*, 27, 15-23.
- Hardwick, R.C., and Wood, J.T. (1972), "Regression Methods for Studying Genotype-Environment Interactions", *Heredity*, 28, 209-222.
- Patterson, H.D. (1997), "Analysis of Series of Variety Trials", in *Statistical Methods for Plant Variety Evaluation*, eds. R.A. Kempton, and P.N. Fox, London: Chapman & Hall, pp. 139-161.
- Piepho, H.P., Denis, J.B., and van Eeuwijk, F.A. (1998), "Predicting Cultivar Differences Using Covariates", *Journal of Agricultural, Biological, and Environmental Statistics*, 3, 151-162.
- Spiegelhalter, D., Thomas, A., Best, N., and Gilks, W. (1996), BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.5, (version ii), Cambridge: MRC Biostatistics Unit.
- Spiegelhalter, D., Thomas, A., Best, N., and Gilks, W. (1997), BUGS: Bayesian Inference Using Gibbs Sampling, Addendum to Manual, Version 0.6, Cambridge: MRC Biostatistics Unit.
- Talbot, M. (1984), "Yield Variability of Crop Varieties in the U.K.", *Journal of Agricultural Science, Cambridge*, 102, 315-321.
- Talbot, M. (2001), "Regional patterns in the response of cereal varieties", Report to HGCA Crop Evaluation Limited, Statistics Working Party, March 2001
- Theobald, C.M., Talbot, M., Nabugoomu, F. (2001), "A Bayesian Approach to regional and Local-Area Prediction from Crop Variety Trials", *Journal of Agricultural, Biological and Environmental Statistics*, (in press).
- Van Eeuwijk, F.A., Denis, J.B., and Kang, M.S. (1996), "Incorporating Additional Information on Genotypes and Environments in Models for Two-Way Genotype by Environment Tables", in *Genotype by Environment Interactions*, eds. M.S. Kang, and H.G. Gauch Jr, Boca Raton: CRC Press, pp15-50.

APPENDIX A1 : USER DOCUMENTATION FOR THE VARADAPT SYSTEM

A description is given of the purpose and functions of the main web pages that comprise the Varadapt system

SELECTING THE CROP



HTML page : home.htm

This is the opening page for Varadapt and provides initial links to the main parts of the system.

- **About Varadapt:** links to a page that explains the background to the system.
- **Using Varadapt:** helps the first-time user to operate the system.
- **Technical stuff:** some of the technical detail behind Varadapt.
- **WWW links:** where to go for more information about varieties.
- **Select a variety:** the first step before selecting a variety is to choose a crop. At present only spring barley and winter wheat are handled but the system can readily be extended to deal with other crops.
- **Feedback:** the comments of users can be entered here.

CHOOSING A VARIETY TO INVESTIGATE

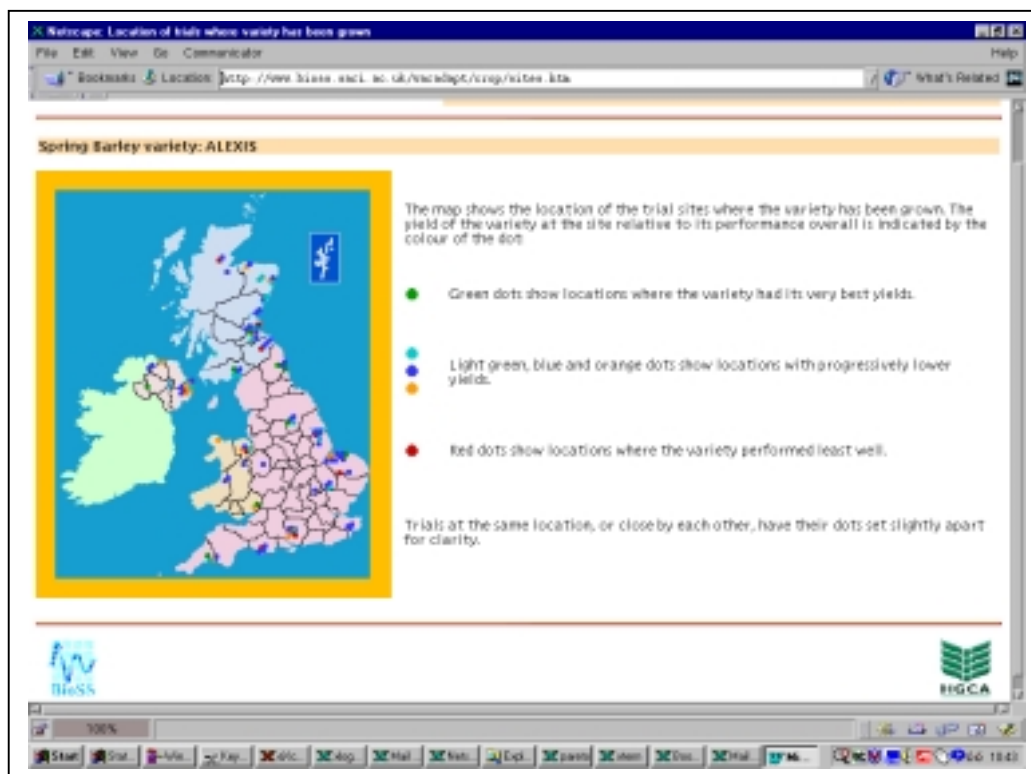


HTML Page: varsel.htm

Once the crop is selected the user is passed to this page that provides a list of varieties for which trials data are available. A variety is chosen by scrolling through the list of names and clicking on the appropriate one. Subsequently the user can move to an exploratory tool by clicking on one of the options listed in the bottom half of the page.

All Varadapt operations will focus on the variety chosen here until such time as the user returns to this page and selects a new variety.

VARIETY PERFORMANCE AT TRIAL SITES THROUGHOUT THE UK



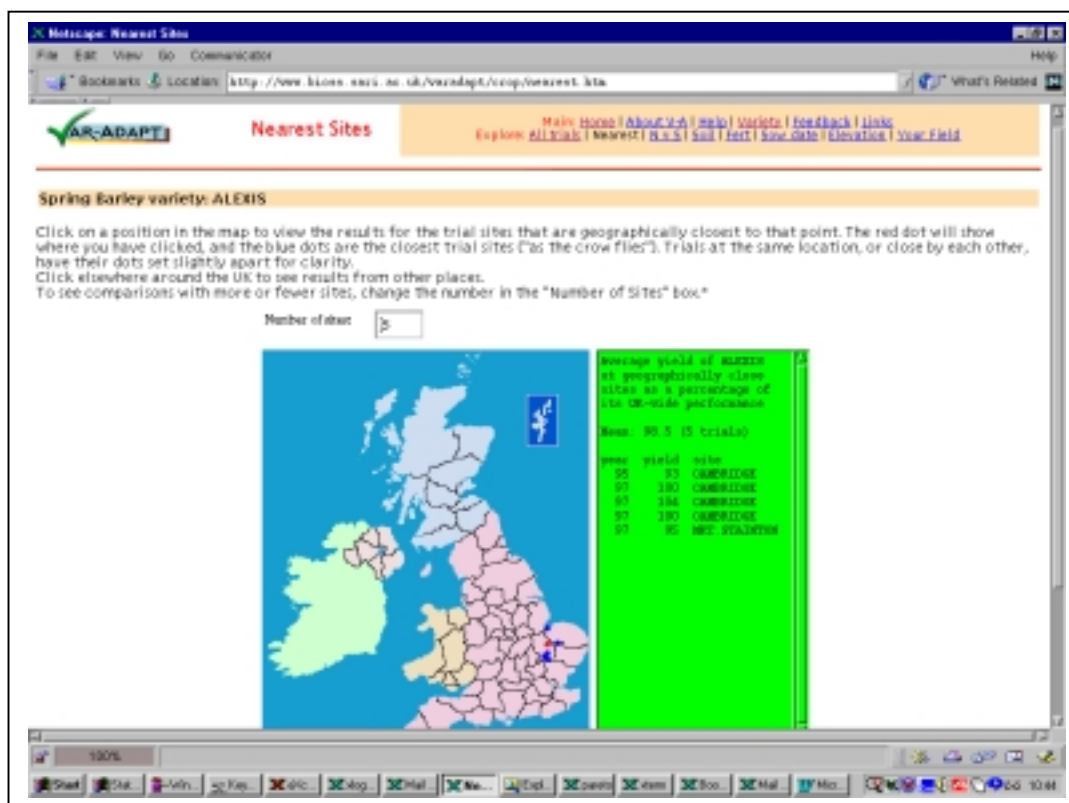
HTML page: [sites.htm](#)

This page is brought into operation by clicking on the line in the Select Variety page: “See the distribution of all sites where the variety has been grown”, or by clicking on the “All Sites” option at the top of the page.

A map of the UK is displayed with coloured dots representing the location of the trials where the variety identified at the top of the page has been sown. The colour of the dots indicates how well the variety has performed at that location relative to its overall performance, i.e. relative to its average yield in recent years.

- Dark green: the best performance of the variety was at these locations. When the trial locations are ranked in terms of performance of the variety then these were the places where the best 10% of performances were achieved.
- Light green: relatively good performance - in the top 66% to 90% of the variety’s trial yields.
- Blue: average performance – trials where the ranked variety performance is the range 33% to 66%.
- Orange: poorer than average performance - the bottom 33% to 10% of performances.
- Red: the worst 10% of yield results for the variety was at these locations.

PERFORMANCE OF VARIETY AT NEAREST TRIAL SITES



HTML page: nearest.htm

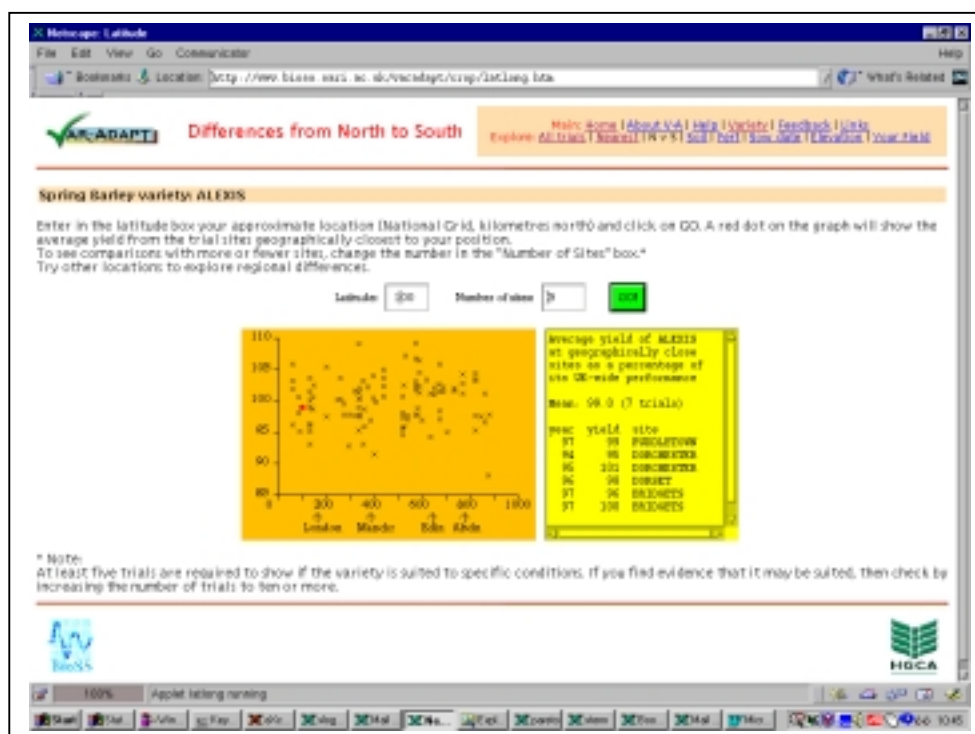
This page is brought into operation by clicking on the line in the Select Variety page: "View results for trial sites near your location.", or by clicking on the "Nearest" option at the top of the page.

A map of the UK is displayed. Point the mouse to a position on the map and click with the left mouse button. A red dot will appear at the location of the pointer and five blue dots will be displayed around the red dot. These blue dots identify the five closest trial sites in which the variety was sown. Closeness is defined as distance as 'the-crow-flies'. In the panel alongside the map appear details of the trials together with the yields of the variety expressed as a percentage of its average yield in all trials. A figure of 100 means that the variety yielded the same as its Recommended List figure in that trial. A figure of 105 indicates that the variety yielded 5% above that quoted in the Recommended List.

The panel also includes the average yield over the selected trials. If this is significantly different (statistically) from 100 then this is noted in the comments.

The user can move the pointer around the UK map successively clicking on locations to seek evidence of local adaptation. The number of close trials can be varied by changing the default number 5 in the box above the map.

NORTH / SOUTH DIFFERENCES IN VARIETY PERFORMANCE



HTML page: latlong.htm


This page is brought into operation by clicking on the line in the Select Variety page: "Evaluate effects associated with north/south location.", or by clicking on the "N v S" option at the top of the page.

A graph is displayed which shows the yield of the variety plotted against the distance from south to north of the trial locations in which the variety was grown. To see if the variety has performed better or worse than average at a particular position, enter in the box at the top the page a latitude figure and click on the Go button. A red dot should appear in the graph that indicates the average yield of the variety at the closest 5 trial sites to that latitude.

In the panel alongside appear the individual trial details together with the yields of the variety and their mean. If the mean is significantly different from 100 then this is noted in the comments.

More or fewer than 5 close locations can be chosen by changing the number in the box above the panel.

ADAPTABILITY TO SPECIFIC FIELD CONDITIONS


Soil Conditions



[Home](#) / [About Us](#) / [Help](#) / [Variety](#) / [Feedback](#) / [Links](#)
[English](#) / [All trials](#) / [Season 1](#) / [S.S.3](#) / [Soil](#) / [Test](#) / [Soil data](#) / [Location](#) / [Your field](#)

Spring Barley variety: ALEXIS

To learn how soil conditions may influence the variety's performance, please respond to as many of the following questions as you feel reasonably confident to answer. Then press the GO button below.

Previous crop outside?	<input type="text" value="yes"/>	108.2 (64)
Soil texture	<input type="text" value="sandy clay loam"/>	108.2 (36)
Soil drainage	<input type="text" value="free"/>	89.2 (70)
Soil pH	<input type="text" value="up to 6.5"/>	108.2 (39)
Soil P205	<input type="text" value="low"/>	90.4 (42)
Soil K20	<input type="text" value="low"/>	90.8 (20)
Soil Mg	<input type="text" value="medium"/>	90.8 (26)
Click	<input type="button" value="GO"/> when ready.	All data 108.2 (110)

The number in parentheses is the number of trials on which the mean is based.

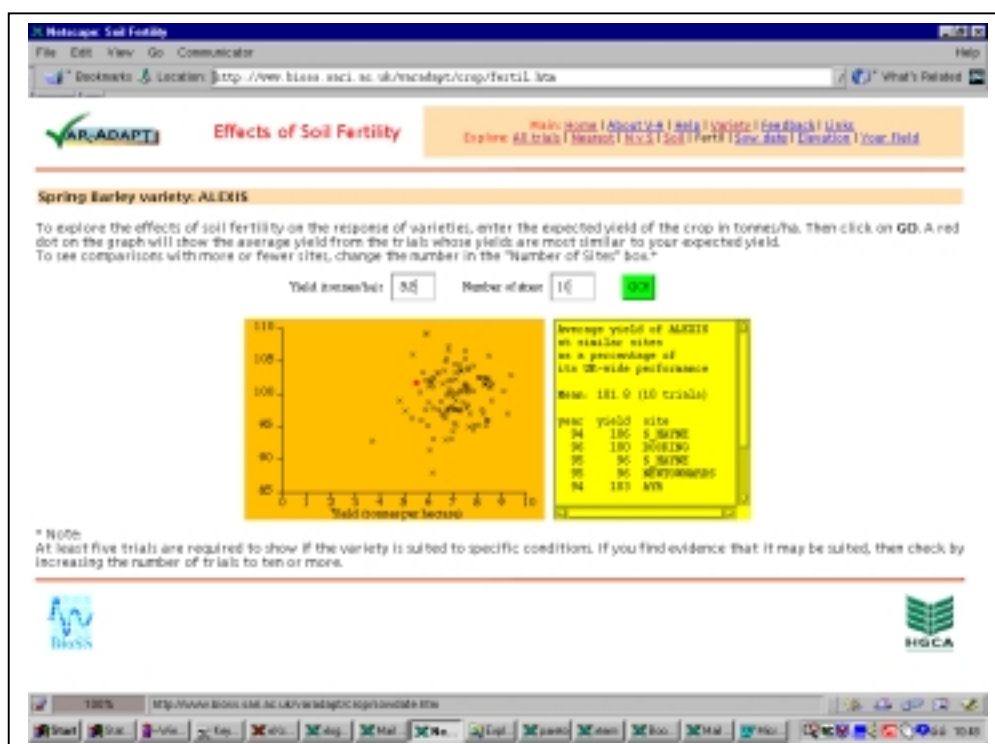



HTML page: soilcha.htm

This page is brought into operation by clicking on the line in the Select Variety page: “See how [soil conditions](#) affect the variety's yield. ”, or by clicking on the “Soil” option at the top of the page.

A list of soil characteristics is displayed and the user is asked to select the characteristics that best describes the circumstances of his/her field. Click on the Go button and the number of trials with each characteristic is displayed along with the mean yield of the variety in those trials. If the mean yield is significantly different (statistically) from 100 this fact is noted alongside the mean.

EFFECTS OF SOIL FERTILITY



HTML page: fertil.htm

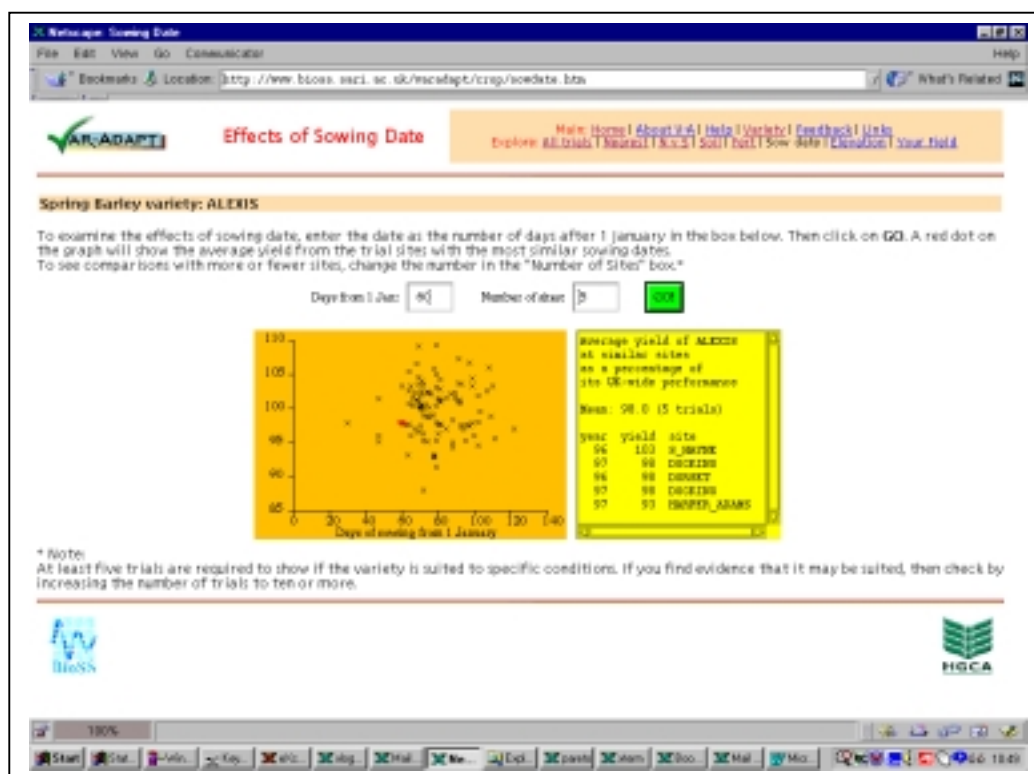
This page is brought into operation by clicking on the line in the Select Variety page: “Does the variety perform relatively better or worse in [fertile](#) soils? ”, or by clicking on the “Fert” option at the top of the page.

A graph is displayed which shows the yield of the variety plotted against the average yield of the trial in which it was sown. To see if the variety has performed better or worse than average at a particular trial yield level enter in the box at the top of the page a yield figure in tonnes/hectare and click on the Go button. A red dot should appear in the graph that indicates the average yield of the variety at the closest 5 trial sites to the specified yield.

In the panel alongside appear the individual trial details together with the yields of the variety and their mean. If the mean is significantly different from 100 then this is noted in the comments.

More or fewer than 5 close locations can be chosen by changing the number in the box above the panel.

EFFECTS OF SOWING DATE



HTML page: sowdate.htm

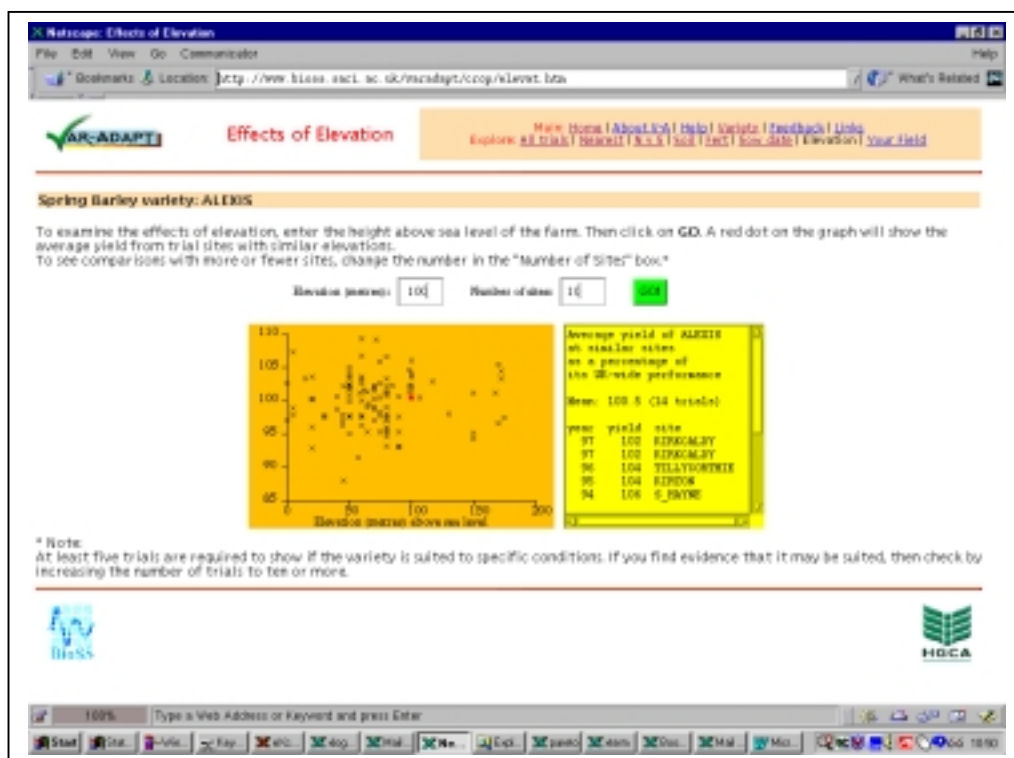
This page is brought into operation by clicking on the line in the Select Variety page: "What are the effects of [sowing](#) date?", or by clicking on the "Sow Date" option at the top of the page.

A graph is displayed which shows the yield of the variety plotted against days from 1 January to sowing of the trial in which the variety was grown. To see if the variety has performed better or worse than average at earlier or later sowings enter in the box at the top of the page a sowing days figure and click on the Go button. A red dot should appear in the graph that indicates the average yield of the variety at the closest 5 trial sites to that sowing date.

In the panel alongside appear the individual trial details together with the yields of the variety and their mean. If the mean is significantly different from 100 then this is noted in the comments.

More or fewer than 5 close locations can be chosen by changing the number in the box above the panel.

EFFECTS OF ELEVATION



HTML page: [elevat.htm](#)

This page is brought into operation by clicking on the line in the Select Variety page: "Check the performance in trial sites with a range of [elevations](#).", or by clicking on the "Elevation" option at the top of the page.

A graph is displayed which shows the yield of the variety plotted against the elevation of the trial location from sea level. To see if the variety has performed better or worse than average at an elevation enter in the box at the top the page an elevation figure and click on the Go button. A red dot should appear in the graph that indicates the average yield of the variety at the closest 5 trial sites to that elevation.

In the panel alongside appear the individual trial details together with the yields of the variety and their mean. If the mean is significantly different from 100 then this is noted in the comments.

More or fewer than 5 close locations can be chosen by changing the number in the box above the panel.

PREDICTING ADAPTABILITY FOR SPECIFIC CONDITIONS

HTML page: locate.htm

This page is brought into operation by clicking on the line in the Select Variety page: “to check if a variety might be adapted to your field, please click on Your Field”, or by clicking on the “Your Field” option at the top of the page.

A list of field characteristics is displayed and the user is asked to select the characteristics that he/she considers the most important factors and how the field in which the crop is to be grown is best described in terms of these characteristics. Click on the Go button and Varadapt will attempt to estimate the extent to which the variety in question may be adapted to the specific conditions. A report will be produced which indicates the expected yield of the variety under the specified conditions. The expected yield is expressed as a percent of the variety's UK Recommended List figure.

This part of the system involves computationally intensive calculations and the speed of response reduces as the number of characteristics selected increases. Equally, if the user specifies too many characteristics there will be too little information in trials to support any evidence for adaptation. So in selecting characteristics it is better to limit the number to no more than 2 or 3.

APPENDIX A2 : PROGRAMMER DOCUMENTATION FOR THE VARADAPT SYSTEM

The Java programs are in the following files:

Called directly from the applet and HTML page:

elevat.java	elevat.htm	Effects of Elevation
fertil.java	fertil.htm	Effects of Soil Fertility
latlong.java	latlong.htm	Differences from North to South
locate.java	locate.htm	Describing your Field
nearest.java	nearest.htm	Nearest sites
sites.java	sites.htm	Location of Trial Sites
soilcha.java	soilcha.htm	Soil conditions
sowdate.java	sowdate.htm	Effects of Sowing Date

Called from other java modules:

FFreader.java	Subroutines for free-format input
cropf.java	Reading the site information files ww_data and sb_data
mlr.java	Simulation software for the location applet
reposn.java	Moves points representing sites on a map to avoid overlap
varadapt.java	General subroutines called from several applets

Each file contains one Java class whose name is the same as that of the file (without the .java). If a routine or variable in one class is referred to by another class, its name has to be preceded by the class name and a dot, e.g. varadapt.probt.

STANDARD JAVA ROUTINES

All of the applet programs contain the following standard Java routines:

```
public void init ()  
    // called when the applet starts. There is one init for each applet.  
    // Most of them get the applet parameters from the HTML
```

```
public boolean action(Event evt, Object arg)  
    // handles clicking on the GO button or pressing return.  
    // There is one action method for each applet.
```

```
public void paint(Graphics g)  
    // draws the applet on the screen.  
    // There is one paint method for each applet. It is activated  
    // by a call to repaint().
```

COMMON JAVA ROUTINES

Most of the applets contain the following home-written routines. They do the same general tasks for each applet, but the action differs from one applet to another.

```
void makegr()          // creates the buttons, boxes etc on the applet
```

```
void prog()            // does the main calculations
```

```
void indat()            // reads the site information file  
                        // There is an indat in each class, but the  
                        // methods are subtly different for each class
```

```

public FFreader opencropfile (String filename)
    // reads the crop file name 'filename'.
    // This method is the same in all classes but it has to
    // be duplicated in each class; it won't
    // compile if it is put in the varadapt class.

```

Most applets have the same set of global variables:

```

final int maxtrials = 400, maxvars = 30;    // max. trials, varieties
final int maxmeas = 25;                    // max. measures
String [] getdat = new String [maxtrials];
    // data for the analysed attribute for each site (held as characters)
String [] sitename = new String [maxtrials]; // site names
String [][] gettrinfo = new String [maxtrials][maxmeas];
    // gettrinfo[i][j] holds value of attribute j in trial i
double[] xval = new double[maxtrials];
    // data for the analysed attribute for each site (held as a real number)
double[] yd = new double[maxtrials];
    // yields for current variety (held as a real number)
String[] year = new String[maxtrials];
    // year of each trial
int num=0; // total number of sites
int sub=0; // number of sites with this attribute known
int subop; // number of sites with known yield
int numdisp; // number of nearby sites for comparison
double eyld,entry; // entry = value that the user fills in the box
    // eyld = estimated yield for that value of entry
String cropname,variety;
int datacol; // column in sites file of relevant data
String graphtitle // Title printed at the top of the graph
String entrytitle // X-axis label on the graph

```

Applets *elevat*, *fertil* and *sowdate* contain only these standard routines. Other routines are mentioned below.

FFREADER CLASS

```

public class FFreader extends DataInputStream

// ----- Free Format Input Routines for Java

// ----- GLOBAL VARIABLES -----

public static double rmiss = -1e21; // Real missing value
public static int imiss = -32767; // Integer missing value
private static String line; // The current line being read
private static int linpos; // The current position on the line

```

```

public FFreader (InputStream is) // declares an FFreader object
public String fscans ()           // reads a string from the file
public String getLine ()         // reads a line from the file
String newLine ()               // reads a new line from the file
public void skipline()           // skips to the end of the current line
public int fscani ()             // reads an integer from the stream
public double fscanr ()         // reads a real number from the stream

```

CROPF CLASS

```
public class cropf
```

Methods findcropfile, readcropfile

```

* findcropfile finds the name of, and readcropfile reads the crop file for crop 'cropname'
* and updates the arrays
* getdat[t] - the data for the selected variety for trial t (as a string)
* gettrinfo[t][n] - attribute n for trial t (as a string)
* numsub[0] = number of varieties
* numsub[1] = number of trials

```

```

static final int maxtrials=400, maxvars=30, maxmeas=25;
// maximum number of trials, varieties and measures

```

```
public static String findcropfile (String cropname) {
```

```

public static int readcropfile (FFreader br, String getdat[],
String gettrinfo[][] , String variety, int numsub[]) {

```

LATLONG APPLET

```
public class latlong extends java.applet.Applet
```

Plots yield of a variety against the latitude of the sites where the variety was grown

```

static void drawgraph
(Graphics g, int xright, int ybot, String xtitle,
double[] xval, double[] yd, int subo, double entry, double eyld) {
// draws the graph of yd against xval
// with its lower right corner at position (xright,ybot)
// subo = number of trials

```

```

// entry = value of x specified by user
// eyld = the estimated yield at x=entry
// The Latitude applet uses a slightly different version of 'drawgraph'
// from the other applets (which use the one in varadapt)

```

```

void printcity (Graphics g, String city, int north)
    // prints the name of a city and an arrow below point 'north' on the
    // Y (latitude) axis

```

LOCATE APPLET

```

public class locate extends Applet

```

```

int rows = 9;
int maxcombins = 120; /* limit on combinations of levels */
int maxnodes = 5;    /* limit on number of nodes */

    // /* variables for storing postcodes and grid references */
    // redundant, but kept in case they are needed again

```

```

/** variables passed to the mlr module */

```

```

String cropname,variety;
static boolean stopped;

```

```

/** variables read from the mlr module */

```

```

int nodes;

```

```

/** setting up the applet */

```

```

public void stop() // this is called when the browser shows a different
    // page; it sets the variable 'stopped' to true.
    // The simulation software (mlr.java) checks 'stopped'
    // at every iteration and stops when 'stopped' is true.
    // In Netscape (and probably other browsers) an applet
    // does not stop running when the user switches to a
    // new Web page, so the applet has to know when to stop.

```

Routines dealing with post codes and grid references

```

    // now redundant, but kept in as comments should they be needed again

```

```

void outdiff (double simyield, double simsd, TextArea outbox)
    // shows the averages and differences in a text box
    // simyield = simulated average
    // simsd = simulated SD

```

```
static void progrout (int percent) // prints progress of calculation
```

MLR CLASS

```
public class mlr
```

```
    // Does the simulation  
    // This module, like the Locate module, used a lot of code for dealing with  
    // latitude and longitude. It was difficult to implement easily and has  
    // been commented out in case it is needed in future.
```

```
/***/ variates passed from the location applet */*/
```

```
static String cropname, variety;  
static int orignodes;      /* number of nodes in the applet */  
static int cond[];        /* original condition levels */
```

```
/***/ variables passed to the location applet */*/
```

```
public static int nodes;   /* number of nodes with conditions given */
```

```
public static int nsites;  /* the number of sites */
```

```
public static double  
    varyield, varsd,      /* variate yield and SD from data */  
    simyield, simsd;      /* estimated yield and SD from sim */  
/***/ variables local to this module */*/
```

```
private static String  
    gettrinfo[][] = new String[maxtrials][maxvars], /* farm data */  
    getdat[] = new String[maxtrials];               /* yield data for each trial */
```

```
private static boolean  
    xmiss[][] = new boolean[maxtrials][maxvars]; /* TRUE if xobs is missing */
```

```
private static double  
    xobs[][] = new double[maxtrials][maxvars], /* x observations */  
    yd[] = new double[maxtrials],             /* yields */  
    pclass[];                                  /* probability of each level */
```

```
private static int  
    nlev, /* product of levels in discrete classes */  
    npred, /* number of continuous nodes */  
    states[], /* no. of states in the simulation */  
    nrs, /* random number seed */  
    nodeno[], /* variate column for each node in simulation */
```

```

classe[] = new int[maxtrials],          /* the combined level of discrete factors */
faccol[] = {10,11,12,13,14,15,16,17,18,7};
                                           /* columns in data corresponding to rows in applet */

```

Reading the Data

```

static void getparam (String c, String v, int nod, int nodelev[])
    /* gets the major parameters from the locate applet */
    /* crop name, variety, number of nodes, levels of nodes */

static void adjpr(int[] condvalue)
    /* removes all nodes which aren't given conditions , i.e. condvalue<0 */

static void indat(FFreader br)    // reads the site information file.
    // It is different from the other indat's
    // because it also sets up a lot of
    // local arrays for simulation

private static double stod (String s)
    /* converts a string s to a real number */

private static void normalise()
    /** normalise the x observations (including the farmer's site) **/
    /** so they have mean 0 and variance 1 **/

```

Simulation software

```

static boolean dosim()          // calls normalise and simulate

private static void initrand(int fixed)
    /** initialises the random number generator.  If 'fixed' is 0 then it uses
    a random odd integer as the seed, otherwise the value of 'fixed' **/

private static double sdrand()
    // Returns a random real between 0 and 1.
    // The seed NRS is updated by the routine:
    // new seed = old seed * 1843829075 (mod 2**31)

private static double max(double a, double b) // returns larger of a and b
private static double min(double a, double b) // returns lesser of a and b
private static double square(double x)        // returns x squared

private static double rgamma(double s)
    /** * Returns one value from the gamma distribution GAMMA(s,1),
    * i.e. with pdf = y^(s-1) Math.exp(-y) / GAMMA(s) */

```

```
private static void gauss(double z[], int n)
    /* generates an n - vector of i.i.d.N(0, 1) r.v.'s z */

private static boolean simulate ()      // does the actual simulation
```

NEAREST APPLET

```
public class nearest extends java.applet.Applet

    * Plots on a UK map the sites growing the current variety that are
    * nearest to the selected site

int [] east = new int [maxtrials];      // east and north in km from grid origin
int [] nort = new int [maxtrials];
int [] xsit = new int [maxtrials];      // east and north in pixels on applet
int [] ysit = new int [maxtrials];

public void main (int inear,int innor)  // does most of the calculations

public boolean mouseDown (Event evt, int x, int y)
    // prints a dot on the map at the position (x,y) pointed to by the
    // mouse when it is clicked
```

SITES APPLET

```
public class sites extends java.applet.Applet

    * Displays a UK map with trial sites identified where variety was grown

int [] xvali= new int[maxtrials];      // east and north in pixels of the map
int [] yvali= new int[maxtrials];
double[] farmeast = new double[maxtrials]; // east and north on km
double[] farmnorth = new double[maxtrials];
```

SOILCHA APPLET

```
public class soilcha extends Applet
```

* Evaluates the effects of soil characters on the performance of a variety

```
void showstat (Label tex, int n, double low, double high) {  
    // calculates and displays the mean and count  
    // also calculates the difference and significance but they are not now displayed
```

REPOSN CLASS

```
public class reposn
```

* routines to re-position a dot on the map so it does not coincide with other dots

```
private static double square(double x)  
    // returns x*x
```

```
public static void repos (int[] x, int[] y, int nsites)  
    // moves points marking sites on the map so that they don't coincide
```

VARADAPT CLASS

```
public class varadapt
```

This contains general routines called by several applets

```
static String decpl(double x, int places)  
    // prints a real number to a specified number of decimal places.
```

```
static double getentry (TextField tex)  
    // returns the value of the Entry box: -1000 if blank or invalid
```

```
static int getdisp (TextField ch)  
    // returns the value of the No. of Sites box
```

```
static int appdat (String[] getdat, String[][] gettrinfo, int sub,  
    double[] xval, String[] sitename, String[] year, double[] yd, int column)  
    // copies the values in column 'column' of the site data  
    // (gettrinfo) into 'xval' and the corresponding yields from getdat[] into 'yd'  
    // also the trial name into 'sitename' and year into 'year' ignoring sites where  
    // the yield or attribute is unknown  
    // and returning the number of sites where they are known
```

```

static int sortdiff (double[] xval, String[] sitename, String[] year,
    double[] yd, double entry, int numdisp, int subo)
    // sorts the data (xval, yd) in order of absolute difference between
    // xval and entry.

static boolean checkentry (double entry, double[] xval, int subo)
    // checks entry level is not outside all points..

static double allmean (double[] yd, int subo)
    // calculates average of yield (yd) over all 'subo' sites

static void calcdiff (double[] yd, int numdisp, int subo, double output[])
    // calculates the mean and significance of the yield of the nearest sites
    // yd[] = yield for each site
    // numdisp = the number of nearby sites
    // subo = the number of sites altogether
    // output[0] = the significance: 2 = high, 1 = sig, 0 = not sig, -1 = unknown
    // output[1] = the mean from nearby sites

static double probtver (double tt, int idf)
    // calculates probability of t value tt with idf degrees of freedom

static double tabtver(double vv,int idf)
    // returns the T value of vv with idf degrees of freedom

static void limits (double[] x, int n, int[] results)
    // finds the extreme values of an array x, rounds them to round numbers
    // in results[0] and results[1] and calculates a suitable gap between
    // tick marks in results[2]

static void showdiff (int numdisp, double eyld,
    String variety, int stars, int wording, TextArea outbox)
    // shows the averages and differences in a text box
    // numdisp = number of similar sites
    // eyld = average of similar sites
    // wording = whether to put 'similar' (1) or 'geographically close' (2)
    // stars = significance value (0=not sig, 1=*, 2=**, -1=not calculable)

static void showsites (int numdisp, String[] year,
    double[] yd, String[] sitename, TextArea outbox)
    // prints the site name, year and yield of each site

static void drawgraph
    (Graphics g, int xright, int ybot, String xtitle,
    double[] xval, double[] yd, int subo, double entry, double eyld)
    // draws the graph of yd against xval
    // with its lower right cornet at position (xright,ybot)
    // subo = number of trials
    // entry = value of x specified by user
    // eyld = the estimated yield at x=entry

```

APPENDIX A3 : DATA FORMAT FOR THE VARADAPT SYSTEM

The variety trial data on which Varadapt operates are held in a text file that comprises several elements:

- Text title followed by a colon on a line on its own.
- A list of the variety names for which yield information is to follow. Each name must appear starting on a new line. A colon starting on a new line follows the last name. The list of names in the function 'listvars' in the *varsel.htm* page must contain the same varieties in the same order.
- Column headings for the table which follows. A colon starting on a new line follows this list.
- Details of the trials for which yield data are held. An asterisk must indicate an item that is missing or unknown. The details include:
 - **TNO** - Trial code number (1 upwards) which also appears with the yield data and provides a unique link between the corresponding entries in the two tables;
 - **CRO** - Crop code - SB spring barley, WW winter wheat
 - **YEA** - Year of harvest of the trial;
 - **REG** - Code letters identifying the region in which the trial was sown;
 - **SNO** - Trial number within the region;
 - **TYP** - Trial type code i.e. N - National List, R - Recommended List, F - Fungicide Treated, S - Secondary, B - plant breeders', O - Other;
 - **NAM** - Farm name or district in which trial was sown, NB the name should contain no spaces;
 - **NRE** - Cereal Variety Recommended List region codes : 1=SW, 2=SE, 3=CE, 4=NW, 5=NE;
 - **LON** - Longitude of trial location (kms east x 10);
 - **LAT** - Latitude of trial location (kms north x 10);
 - **WKS** - Weeks from 1 January to date of sowing of trial;
 - **NPR** - Previous crop code : 1=cereals, 2=not cereals;
 - **NST** - Soil type code; 1=sandy_clay, 2=sandy_clay_loam, 3=silty_clay, 4=silty_clay_loam, 5=clay_loam, 6=sandy_loam, 7=silt_loam, 8=peaty_loam, 9=sandy_silt_loam, 10=loamy_sand;
 - **NSD** - Soil drainage code : 1=free draining, 2=imperfect or poor drainage
 - **NPH** - Soil pH code: 1=6.5 and less, 2=6.5 to 7.0, 3=greater than 7.0;
 - **NSP** - Soil P : 1=2.0 and less on ADAS scale ; 2=2.0 to 3.0; 3=greater than 3.0;
 - **NSK** - Soil K : 1=1.5 and less on ADAS scale ; 2=1.5 to 2.5; 3=greater than 2.5;
 - **NSM** - Soil Mg : 1=1.5 and less on ADAS scale ; 2=1.5 to 2.5; 3=greater than 2.5;
 - **NAL** - Altitude of trial site above sea level : 1=40m and less; 2=40-60, 3=60-80,4=greater than 80.
 - **pH** - Soil pH;
 - **SP** -
 - **WN** -
 - **ALT** - Altitude of trial site above sea level in metres.

- **DYS** - Days from 1 January to date of sowing.
- The list of variety names repeated as above.
- Column headings for the trials-by-varieties table of yield below. A colon follows it on a new line.
- A trials-by-varieties table of yields. Each line has the following items - any items that are missing is indicated by an asterisk:
 - Trial code number - the same number as in the trials data table;
 - Year of harvest;
 - Region letter codes;
 - Trial number within region;
 - Yields for each variety in turn, in the order of the variety names listed at the top of the file, and expressed as a percent of the mean yield for the trial. See Technical Detail - Data Analysis section.

Example of Data File Format

UK Winter Wheat Variety Trials 1994-2000 - Fungicide Treated

:

MADRIGAL

EQUINOX

CONSORT

REAPER

RIALTO

RIBAND

HEREWARD

CHARGER

BUSTER

SOISSONS

SAVANNAH

:

WINTER WHEAT VARIETY TRIALS - UK 94-00 - SITE DATA

:

TNO	CRO	YEA	REG	SNO	TYP	NAM	NRE	LON	LAT	WKS	NPR	NST	NSD	NPH	NSP	NSK	NSM	NAL	PH	SP	WN
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----	----	----

ALT DYS

:

1	WW	0	S	10	R	WYE	1	6069	1493	6	2	4	1	*	*	*	*	2	9.63	3	3
50	279																				
2	WW	0	S	12	R	CIRENCESTER	1	4023	1976	7	2	4	1	3	1	2	2	4	10.20	3	3
110	288																				
3	WW	0	S	13	*	*	1	*	*	*	*	*	*	*	*	*	*	*	11.16	3	3
*	*																				
4	WW	0	S	91	O	WYE_EARLY	1	6069	1464	1	2	7	1	*	*	*	*	3	*	*	*
76	244																				
5	WW	0	SW	10	R	KINGSBRIDGE	1	2653	474	7	1	7	1	1	1	1	2	4	*	*	*
110	286																				
6	WW	0	W	10	R	USK	1	3479	2028	6	2	6	1	1	1	1	3	4	9.85	3	3
200	280																				
7	WW	0	EE	3	N	CAMBRIDGE	2	5434	2628	5	2	5	1	*	*	*	*	1	*	*	*
20	277																				
8	WW	0	EE	4	N	CAMBRIDGE	2	5434	2628	5	2	5	1	*	*	*	*	1	*	*	*
20	277																				
9	WW	0	EE	5	N	CPB_TWYFORD	2	6217	2667	6	2	2	1	3	2	2	2	3	*	*	*
61	283																				
.																					
.																					
.																					
.																					
364	WW	94	ES	204	R	DOWNIE	5	3525	7364	7	2	2	2	2	1	2	2	2	8.67	3	2
46	287																				
365	WW	94	NS	202	NR	ABERDEEN	5	3864	8101	9	2	6	2	1	3	3	2	4	10.73	3	3
140	304																				
366	WW	94	NS	203	R	LAURENCEKIRK	5	3697	7717	9	*	6	1	1	1	3	2	2	8.47	3	2
50	301																				

```

367 WW 94 NS 204 R INVERNESS 5 2535 8865 8 2 6 1 1 1 2 1 1 7.35 2 1
30 294
:
MADRIGAL
EQUINOX
CONSORT
REAPER
RIALTO
RIBAND
HEREWARD
CHARGER
BUSTER
SOISSONS
SAVANNAH
:
MADRI EQUIN CONSO REAPE RIALT RIBAN HEREW CHARG BUSTE SOISS SAVAN
:
1 0 S 10 95.8 101.8 93.6 100.2 99.3 97.2 104.3 105.9 104.6 103.7 94.8
2 0 S 12 97.8 98.3 95.6 * 101.6 * 101.3 * * * 101.8
3 0 S 13 96.8 101.8 101.8 99.1 98.6 98.4 99.7 102.9 103.0 103.1 99.0
6 0 W 10 99.1 101.5 101.8 99.3 100.2 102.4 102.3 101.4 97.9 93.2 102.0
11 0 EE 10 98.0 101.7 105.3 98.9 103.1 99.5 101.4 88.4 104.9 95.1 100.0
.
.
.
364 94 ES 204 * * 101.2 * 95.9 97.7 93.6 * * 92.7 *
365 94 NS 202 * * 100.1 102.3 96.2 101.8 96.6 97.7 101.3 95.5 *
366 94 NS 203 * * 98.4 * 98.4 99.9 102.1 * 103.3 95.9 *
367 94 NS 204 * * 94.2 * 93.7 106.5 105.4 * 104.2 100.6 *
:

```