Welham, S. J., Thompson, R. and Sylvester-Bradley, R. 2005. *Development of 'RL Plus': winter wheat variety performance in relation to site characteristics (HGCA Project Report No. 365).* Stoneleigh Home Grown Cereals Authority (HGCA).

Project Report No. 365

MARCH 2005

Price: £3.80

Development of 'RL *Plus*': winter wheat variety performance in relation to site characteristics

by

S. Welham[1], R Thompson[1] and R. Sylvester-Bradley[2]

[1]Rothamsted Research, Harpenden, Herts. AL5 2JQ
[2]ADAS Boxworth, Cambridge CB3 8NN

This is the final report of a 30 month project that commenced in June 2002. The work was funded by a contract for £120,898 from the Home-Grown Cereals Authority (Project 2698) and was co-funded by an in-kind contribution from Rothamsted Research (£14,220).

Contents

Acknowledgements

The project was undertaken by:

Roger Sylvester-Bradley[2], Mike Talbot[3], Alec Mann[3], Jane Ryall[4], Andy Horwell[4], Simon Kerr[4], Richard Fenwick[4], Robin Thompson[1], Sue Welham[1], David Cranstoun[5] and John Foulkes[6].

and was steered by:

Jim McVittie[7], Bill Handley[7] Paul Hickman[8] and Ethel White[9].

[1]Rothamsted Research, Harpenden, Herts. AL5 2JQ
[2]ADAS Boxworth, Cambridge CB3 8NN
[3]BioSS, JCMB, The King's Buildings, Edinburgh EH9 3JZ
[4]NIAB Huntingdon Road, Cambridge CB3 OLE
[5]SAC Bush Estate, Penicuik, Midlothian, EH26 0PH
[6]University of Nottingham, Sutton Bonington Campus, Loughborough LE12 5RD.
[7]HGCA, Caledonia House, 223 Pentonville Road, London N1 9HY
[8]Advanta Seeds UK Ltd., Station Road, Docking, King's Lynn PE31 8LS (for BSPB)
[9]DARD, Plant Testing Station, Crossnacreevy, Belfast BT6 9SH

Abstract

This project developed RL *Plus*, an augmented version of the Recommended Lists for Cereals and Oilseeds, published by the HGCA on CD and the internet (http://www.hgca.com) to provide the cereals and oilseeds industry with means of interrogating and analysing data from HGCA-funded variety trials. Given that RL *Plus* is fully documented and published in electronic form, it is not described further here. This report is confined to additional research analysing relative variety performance of winter wheat in terms of site characteristics.

A spreadsheet was constructed of treated yields of winter wheat from 43 varieties across 506 trials (Recommended List, National List or BSPB trials) in the UK from harvests during 1992-2002. This included site information collected from the trials, and supplemented with the site location (OS co-ordinates), soil types, meteorological data and drought index derived during the project. Complete, or near-complete, data existed for 249 trials from harvests during 1993-2003. These data were used to investigate factors associated with site variation in variety yields.

Data-mining techniques were used to identify site variables that explained variation in variety yields between sites. This information was used to build models to describe and predict patterns of variety variability due to site differences. Variation in variety yields could be modelled in terms of overall variety differences (43% variety variation accounted for), and variety interactions with large-scale trend due to geographic location (general climate, 16%), small-scale location trend specific to years (micro-climates, 14%), expected site yield (2%), late sown crops (crops sown on/after 30 October, 0.4%), sites with sandy or shallow soil (0.5%), sites with low soil K index (0.4%), differences between years (4.5%), differences between sites (unexplained by site variables, 2%), and other unexplained variation (18%). Further investigation suggested that other site variables, such as previous cropping, might also influence variety variability but that the relationship was local (differed between regions).

The results of the statistical analysis can be used to optimise use of the ‘Varieties on your Farm’ module of RL *Plus*. In general, geographic location appears to be the most important site variable influencing variation in variety yields across the UK. However, for particular varieties, the expected site yield, soil type or soil K index may be equally important.

Summary

Introduction

Variety trials show that there is variation in variety yields due to interactions between variety and the environment, often called genotype by environment interactions. These interactions mean that a predicted variety yield for a ‘typical’ site may not be appropriate to a specific site. The variability in trial environments can be broken down into differences between years, differences due to site characteristics, and interactions between site and year. Variation in variety yields due to year differences can be quantified, but cannot be used to predict variety yield at a specific site, due to the uncertainty in forecasting weather. Differences due to sites may be due to intrinsic site characteristics that do not change over time (eg. location, soil properties) or to crop management practices (eg. date of sowing) over which the farmer has control. If variety variability is related to these site variables, then the relationship can be used to give an improved prediction of variety yield at a specific site, although the prediction still cannot eliminate uncertainty due to unknown weather. The aim of the statistical analysis in this project was to identify variables that explained variation in variety yields between sites and build models to describe and predict patterns of variety variability due to site differences.

Materials and Methods

<u>Typing soils</u>

Soils data from RL trials included topsoil texture, drainage (free, imperfect, or poor), organic matter, pH, P, K, and Mg (the last 3 as indices) and available water capacity data (AWC) were obtained from the SSLRC for a subset of sites. These were used to estimate soil types as defined in the ‘Fertiliser Recommendations for Agricultural and Horticultural Crops’ MAFF Reference Book 209 (Anon, 2000). Five categories of mineral soil were differentiated according to texture and depth as follows:

Topsoil texture		*Drainage*	
	free	imperfect	poor
	RB209 soil type (for codes, see text)		
peaty loam	P	-	-
sandy loam	S	M	M
silt loam	Z	Z/M*	M
sandy silt loam	M	M	-
sandy clay loam	M/C*	C	C
silty clay loam	Z	Z/C*	C
clay loam	M	C	C

** judged according to locality*

codes being S for light sand soil, A for shallow soil, M for medium soils, C for deep clay soils, Z for deep fertile silty soils, O for organic soils and P for peaty soils. Shallow soils over rock were also identified at sites local to those (e.g. ‘Bridgets’ and ‘Cirencester’) known to have shallow soils, or where subsoil pH was low (<5.5). Organic and peaty soils were identified for the 31% of trials that had organic matter data. Soils types were distributed as follows: shallow (A) 11%, light sand (S) 9%, medium (M) 26%, deep clay (C) 29%, deep silty (Z) 14%, organic (O) 8%, and peaty (P) 3%.

Droughtedness

Monthly soil moisture deficits (SMDs) were calculated from available water capacities (AWCs), rainfall and potential evapo-transpiration (PE), and a summary ‘index’ of droughtedness was derived, based on recent research on UK drought effects on wheat from April through to harvest. AWC was derived from soil type (defined as above) as follows: shallow (A) 120mm, light sand (S) 110mm, medium (M) 180mm, deep clay (C) 200mm, deep silty (Z) 270mm, organic (O) 180mm, and peaty (P) 250mm. Monthly SMD was calculated assuming that rainfall and PE were evenly distributed through each month. The drought index ranged from 0.0-4.9, and had a median value of 1.0 for winter wheat.

The database

A spreadsheet was constructed of treated yields of winter wheat from 43 varieties across 506 trials (Recommended List, National List or BSPB trials) in the UK with harvest during 1992-2002. The data included site information collected from the trials, and supplemented with the site location (OS co-ordinates either provided or estimated from site names), RB209 soil types, monthly meteorological data and drought index derived as above. The trials selected were required to have good site information and contain at least 8 of the 43 varieties. Complete, or near-complete, data existed for 249 trials, harvested during 1993-2003. These data were used to investigate site factors associated with variation in variety yields.

Data-mining techniques were used to identify environmental variables that explained variation in variety yields between sites. This information was used to build models to describe and predict patterns of variety variability due to environmental differences.

Data mining

To explore the nature of variety variability as a response to different environments between sites, factor analysis models were fitted within a model for variety yields that accounted for all sources of variation. To avoid potential confusion between variability due to differences between years and variability due to differences between sites, factor analysis models were fitted for data from each year separately. The factor analysis model regards yields from different environments as different traits for each variety. The model then represents the large number of environments by a much smaller number of hypothetical factors.

For each year, the factor analysis model fits a regression model to the variety.site interaction effects where both the explanatory covariate (the 'factor' of factor analysis) and the regression coefficients are estimated. The estimated explanatory covariates represent combinations of environments that maximise variety variability. When considered in relation to site characteristics, these explanatory covariates may give insight into site characteristics that are related to variation in variety yield.

Modelling variety variability

An extended regression model was constructed to relate site characteristics to variations in variety yield. The model used standard linear regression to relate yield to the overall effect of each site variable, but allowed variation in the regression intercept and slope for each variety, fitted as random effects. Raw site yield data were adjusted to decrease correlation with other variables and improve interpretability. Results from the data mining analysis were used to build a preliminary model. The remaining site variables which had sufficient data present were converted to categorical variables so that both linear and non-linear responses could be detected; then these site variables were each in turn added to the model to check for an interaction with variety. A final model was constructed by including all the site variables found to be individually significant, then omitting terms to find the best subset.

To investigate whether there was local variation in the relationships, the same modelling process was undertaken with two subsets of the data: one encompassing Central, East and Southern England, an area of about 300 × 400 km; the other subset in Eastern Scotland, within ~50km of the coast from the Borders to Aberdeenshire, an area of 250 × 100 km. Both subsets form contiguous areas with reasonable geographic coverage of trial sites.

Results

Data mining

Factor analysis models were fitted to all data from individual years from 1998 to 2002. Models with four factors were fitted in 1999-2002, with only three factors required in 1998. The percentage of variety variability accounted for in total and by each of the estimated factors is shown in Table 1, with the site characteristics found to be related to the estimated factors. In each case, variation accounted for by the first factor includes the variety main effects.

Table 1. Summary of variety variability accounted for by estimated factors in factor-analysis models fitted to yield of winter wheat in treated trials with harvest during 1998-2002, with site characteristics related to the estimated factors.

Harvest Year	Number of factors	% Variance accounted for by factor				Total	Site characteristics related to factors
		1	2	3	4		
2002	4	53.8	10.2	7.6	5.9	77.5	Site yield, longitude, latitude, soil pH
2001	4	48.7	14.6	10.6	7.7	81.6	Site yield, longitude, latitude, soil pH, lodging, drought index, applied N, previous crop
2000	4	52.1	19.3	12.0	6.1	89.5	Latitude, soil pH
1999	4	71.0	10.5	7.5	5.4	94.4	-
1998	3	72.8	14.0	5.0	-	91.8	Site yield, longitude, soil Mg index

Table 2. Site variables tested in a random regression relationship for association with variety variability, in a model containing variety interactions with site yield and location.

Site variable	Category boundaries in definition of the variable	Evidence of interaction with variety?
Soil type	AS, MOP, CZ	Yes
Altitude	50m, 100m	No
Sowing date	7 Oct, 15 Oct, 30 Oct	Yes
Soil AWC	150, 225	No
Soil pH	6.25, 7.75	No
Soil P index	1.5, 3	No
Soil K index	1.5, 3	Yes
Drought index	0.1, 0.5, 1	No
Previous crop	Non-cereals, cereals	No
Soil Mg index	1.5, 3	No
N applied	200,225	No

Modelling variety variability

The data mining analysis strongly indicated site yield as a source of variation in variety yield. However, raw site average yields were confounded with site variables such as previous cropping, date of sowing and harvest year. The raw site yields were adjusted to predict expected site yields in a typical year for a first cereal crop and normal sow date. This allowed the effects of site variables to be more easily distinguished, and corresponded more closely with information available when predictions are made. Latitude and longitude were also strongly indicated as important variables from the data mining analysis. As the two variables together summarise average climatic conditions, the model was constructed in terms of a two-dimensional smooth surface in terms of latitude and longitude for each variety, fitted using a smoothing spline. This smooth surface represented large-scale trend in variety response to climate. It was expected that there might also be more local trend in variety yields, and a spatial model with exponential correlation was fitted to represent this small-scale trend.

Table 2 shows the definition of the additional site variables tested and whether evidence of an interaction with variety was found. Evidence of an interaction was found with soil type, date of sowing and soil K index, with the variety response associated with specific categories: soil type AS (shallow or sandy), sowing date before 7 October or after 30 October, and soil K index <1.5. On fitting a joint model containing variety interactions with location, expected site yield, and the additional variables shown to be associated with variety variability, it was found that not all of the additional variables were required and a final reduced model was constructed.

The final model contained terms to account for the overall effect of the following variables:

- expected site yield (in average year with previous non-cereal crop, normal sow date)
- previous crop (non-cereals, cereals)
- date of sowing (< 7 Oct, 7-15 Oct, 15-30 Oct, >30 Oct)
- geographic location (latitude, longitude)
- shallow or sandy soil
- low soil K index

with terms used to model variety variability:

- variety main effect (variety)
- variety $\times$ geographic location interaction, *ie*. two-dimensional spline, large-scale trend (variety.spl(location))
- variety interaction with expected site yield (variety.siteyld)
- variety interaction with late sow dates, *ie*. after 30 Oct (variety.latesown)
- variety interaction with shallow/sandy soils (variety.soilAS)
- variety interaction with low soil K index (variety.lowK)
- variety interaction with site, *ie*. residual variety.site variation unaccounted for by other environmental variables
- variety variation between years (variety.year)
- variety interaction with location within year, fit using spatial model, *ie*. within-year short-scale trend (variety.year.spatial(location))
- variety interaction with trial, *ie*. residual variety.site.year variation

and with additional terms in the model used to account for variation due to year, region, site and trial and their interactions.

Some of the variety interaction terms could be used to predict variety yields at given sites, but others could not be used for prediction. For terms associated with variety variability, Table 3 shows the variance component associated with each term, whether it could be used for prediction and an estimate of the average percentage of the total variety variation accounted for by the term. Variety main effects accounted for 43% of the variety variation in the model, with interactions with site accounting for 21% and interactions with year accounting for 36%. Although this model accounted for most of the variety × site variability, the year-to-year variation was still unpredictable. Predictions from the model, appropriate to a notional 'typical' year, may therefore tend to be inaccurate for any particular year. However, the model could predict which varieties tended to prefer which environments in a typical year.

Table 3. Variety interaction terms in the final model. Terms are classified according to whether they can be used to predict variety yield at a new site (given appropriate site information) and the average percentage of the total variety variation accounted for by each term.

Model term	Type of term	% variety variation accounted for
Variety	Predictive	43.3
Variety.spl(location)	Predictive	15.6
Variety.siteyld	Predictive	2.0
Variety.latesown	Predictive	0.4
Variety.soilAS	Predictive	0.5
Variety.lowK	Predictive	0.4
Variety.site	Non-predictive	1.9
Variety.year	Non-predictive	4.5
Variety.year.spatial(location)	Non-predictive	13.7
Variety.trial	Non-predictive	17.8

The three largest sources of variety variability (after the main effects) were the large-scale location trend, small-scale location trend and noise (variety.trial). The large-scale location trend could be used in prediction and may represent variety response to large-scale changes in climatic conditions across the UK. The following varieties showed strong evidence of large-scale location trend with other factors held fixed: Soissons, Hereward, Scorpion 25, Mercia, Hunter, Claire, Malacca, Tanker, Madrigal, Rialto and Spark.

For the small-scale location trend, the estimated spatial correlation parameter indicated strong correlation (0.89) between locations 20km apart, dropping to 0.75 for locations 50km apart and 0.56 for locations 100km apart. This suggests that micro-climates (~50km diameter) existed within which particular varieties performed better (or worse) than expected. However, this small-scale pattern appeared to be inconsistent across years, and so could not be used to predict relative variety yield at a given site. The variety.trial interaction term represented random variation in variety response across trials and could be regarded as random noise, in that it had no apparent pattern and would be unpredictable for new sites or future years.

Although the remaining site variables each accounted for only a small percentage of the variety variation, each variable showed evidence of a large effect on the yield of one or more varieties, and should not be ignored when predicting site yield for these varieties. Varieties Scorpion 25, Robigus, Xi 19 and Tanker had a positive interaction with expected site yield, *ie*. these varieties showed a larger increase in yield than average as overall site yield increased. In contrast, varieties Hereward, Mercia, Spark, Soissons, Malacca and Shamrock had a negative interaction with expected site yield, *ie*. these varieties showed a smaller increase in yield than average as overall site yield increased. The variety interaction effects with expected site yield showed a strong correlation with overall variety yield, indicating that high-yielding varieties tended to give a better response to high-yielding sites. Variety Riband had a negative interaction with late sowing, indicating a worse response then average to these conditions. Varieties Hereward and Rialto had a negative interaction with shallow or sandy soils, while Deben had a positive interaction. Finally, variety Hereward had a positive interaction with low soil K index.

Analysis of the two subsets (Central, East & South (CES) England or East Scotland) was used to examine whether effects were consistent across the whole UK. In both cases, there was no evidence of either large-scale or small-scale smooth location trend in variety response. The large-scale trend was not expected in these smaller subsets. The small-scale trend was expected to be detected, but there was possibly not enough information to estimate the spatial parameters within the subsets. Where variety interaction terms occurred in models for both the full data and subsets, the variance components were of similar size. For the CES England subset, variety yields were found to be related to position along a NW-SE axis within the area, expected site yield, late sown crops, sandy/shallow soils, previous crop and soil pH. For the East Scotland subset, variety yields were found to be related to latitude, expected site yield and late sown crops. However note that there were very few sandy/shallow soils or late sow dates within this subset. There was a high correlation between the variety main effects and variety.siteyield interactions obtained from the two subsets.

Discussion

The statistical analysis identified site variables associated with variation in relative variety yields between sites. These site variables were used to improve predictions of variety performance on a given site using information readily available to the grower. The subset analyses indicated that although the relationship of variety yield with expected site yield held across the UK, site variables such as previous crop and soil pH might have a more local interaction with variety. Further work and an extension of the methodology are required to detect local influence of site variables on variety yield.

The results of the statistical analysis can be used to optimise use of the 'Varieties on your Farm' module of RL *Plus*. In general, geographic location appears to be the most important site variable influencing variation in variety yields across the UK. However, for particular varieties, the expected site yield, soil type or soil K index may be equally important.

Technical Report: Modelling relative variety performance in terms of site characteristics

S J Welham & R Thompson, Rothamsted Research

1. Introduction

Within the UK, systems of testing over several years and many locations have been used to decide which varieties of arable crops should be adopted onto the National List and Recommended List, and summary results of the trials are published to guide growers in their choice of varieties. Traditionally, the summary results indicate average variety performance over several years of trials. However, Talbot (1984) demonstrated that substantial genotype by environment interaction existed in Recommended List trials. The genotype by environment interaction can be decomposed into variety interactions due to differences between years, differences between locations, and interactions with combinations of location and year. Genotype by year interactions can be quantified, but are not helpful in predicting future variety yield, due to the uncertainty in forecasting weather. Interactions due to trial locations may be due to intrinsic site characteristics that do not change over time (eg. geographic location, soil properties) or to crop management practices (eg. date of sowing) over which the grower has control. If variety by location interactions could be related to these site variables, then that relationship could be used to give an improved prediction of variety yield at a specific site, although the prediction still cannot eliminate uncertainty due to unknown weather.

More recently, internet-based tools such as RL *Plus* (http://www.hgca.com) have made it possible for growers to estimate variety performance from specified subsets of trials. Whilst this added flexibility may lead to more appropriate variety predictions, it is also possible that these results could be misleading if the subsets specified become too small for reliable inference or if there was substantial correlation between site characteristics. The aim of the work described in this paper was to use a large data set containing variety means from trials at different locations across the UK during 10 years to identify site characteristics that explained genotype by location interactions and to build models to describe and predict site-specific variety yields. The results of these analyses can be used to inform growers of important site characteristics and improve the design of tools such as RL *Plus*.

The problem of modelling variety by location interactions in terms of site characteristics has been previously examined, most recently by Theobald *et al.* (2002) and Denis *et al.* (1997, 1998). Denis *et al.* (1998) considered the use of multiple site covariates, but fit all variety.covariate interactions as fixed model terms. Denis *et al.* (1997) also used multiple site covariates with fixed variety.'site covariate' interactions, but also allowed genotype covariates, and fitted the variety.'genotype covariate' interactions as random to allow for correlation and heterogeneity between varieties. Theobald *et al.* (2002) used a Bayesian approach with site covariates so that variety effects and interactions could be considered as random. The approach in this paper extends the standard linear mixed model used for analysis of variety trial data (see for example Talbot, 1984) to fit variety effects and interactions with site covariates as random regression terms in the model. We also consider the problem of including two-dimensional site variables, such as geographic location, within the model. We accommodate variety interactions with location using thin-plate splines within the mixed model framework to account for large-scale trend, and allowing spatial correlation between locations to account for small-scale trend.

Section 2 describes briefly the data set used in the analysis. Section 3 describes the methods and results for the exploratory data analysis phase, which are used to develop the random regression models presented in section 4. Finally, in section 5 we discuss the merits and implications of our approach, and consider further work that could be done to improve the model and methods.

2. Data set

2.1 Classifying sites for soil type

The aim was to provide a categorisation of soils throughout the UK which (i) would be easily recognisable to the cereals industry, (ii) could be identified for most or all trials from the RL data (1992-2002), and (iii) had the best chance of relating to cereal crop performance, and hence may identify interactions with variety, or variety type. The data from RL trials had seven columns on soils: topsoil texture, drainage (free, imperfect, or poor), organic matter, pH, P, K, and Mg (the last 3 as indices). In addition, available water capacity data (AWC) were obtained from the SSLRC for about 60% of sites. The most widely recognised soil categorisation in current use by the industry is defined in the 'Fertiliser Recommendations for Agricultural and Horticultural Crops' MAFF Reference Book 209 (Anon, 2000). This provides five categories of mineral soil, differentiated according to texture and depth, plus organic and peat soils as follows: light sand soil (S), shallow soil (A), medium soil (M), deep clay soil (C), deep fertile silty soil (Z), organic soil (O) and peaty soil (P). This categorisation is to be used in future variety testing, and all new sites, from 2003, are being classed accordingly. Hence RB209 soil types were adopted and estimated from RL trials data, as shown in the following table.

Topsoil texture		*Drainage*	
	free	imperfect	poor
		RB209 soil type (for codes, see text)	
peaty loam	P	-	-
sandy loam	S	M	M
silt loam	Z	Z/M*	M
sandy silt loam	M	M	-
sandy clay loam	M/C*	C	C
silty clay loam	Z	Z/C*	C
clay loam	M	C	C

** judged according to locality*

It was possible to categorise all sites in this way, with the following modifications. Shallow soils over rock (coded A) were identified where sites (e.g. 'Bridgets' and 'Cirencester') or grid references local to these sites are known to have shallow soils. A few sites with AWCs much less than indicated by their topsoil texture (or with low pH values - see below) were also reclassified as shallow. Of the 31% of the treated winter wheat trials that had organic matter data, any soil with 9-16% organic matter, and any sandy loam with 5-9% organic matter, was classed as organic (O). Soils with 17% organic matter or more were classed as peaty (P). (Since 69% or soils were not open to this check, it is likely that some organic soils have not been recognised.) The three trials with topsoil pH<5.5 (all Scottish, in 1997) were also classed as shallow (A). It should be noted that 3 sites (one with 3 trials) with soil P index=0, and five sites with soil K index=0 were not excluded from the data. About 60 trials with no soil records were classified by association with sites of the same name. At a few sites where topsoils were not consistent, and are known to be variable, past cropping was used to adjudicate on soil type, it being unlikely that root crops would be grown on clay soils in the west of England! Soils types, according to RB209, were thus distributed as follows:

Soil type	*Sites* *%*
Shallow (A)	11
Light sand (S)	9
Medium (M)	26
Deep clay (C)	29
Deep silty (Z)	14
Organic (O)	8
Peaty (P)	3

2.2 Classifying sites for droughtedness

Various drought indices have been devised previously in both the USA (Palmer 1965) and the UK (Thomasson 1979). The method adopted here was based on the latter, but modified according to recent research, as follows (where AE is actual evapotranspiration and PE is potential evapotranspiration, as estimated by the Met. Office):

1. Soil moisture deficit (SMD) was calculated for each month, using monthly mean meteorological data, and assuming AE = 30% x PE before April, AE = 50% x PE during April and AE = PE (i.e. full crop cover) through May, June & July.
2. Working from the findings of Foulkes *et al.* (2001) it was assumed that drought effects would occur when
   - SMD in May = 0.5 x AWC
   - SMD in June & July = 0.65 x AWC
3. The extents (in mm) to which SMDs for each month exceeded these levels were added, and then divided by 100 to give a scale which, for almost all sites, ranged from 0-4 (although greater values were possible).

Note that only monthly met data were available, so it has been assumed that rainfall and evapo-transpiration were evenly distributed through each month. If most of the rain fell (for example) at the end of any month, the drought effect will inevitably have been underestimated, and *vice versa*.

AWCs used were not those obtained from SSLRC but were estimated from soil types as follows:

Soil type & code		Typical texture		Stoniness (%)		AWC # (%)		Depth (cm)		Total AWC (mm)
		topsoil	*subsoil*	*topsoil*	*subsoil*	*topsoil*	*subsoil*	*topsoil*	*subsoil*	
Shallow	A	ZCL	ZC	0	25	18	15	20	100	120
Light sand	S	LS	S	0	15	12	8	30	150	110
Medium	M	SCL	SC	0	0	17	15	25	120	180
Deep clay	C	CL	C	0	0	18	15	20	130	200
Deep silty	Z	ZL	ZL	0	0	22	19	30	140	270
Organic	O	OSCL	SC	0	0	18	15	25	120	180
Peaty	P	PL	C	0	0	30	18	30	120	250

modified from MAFF / ADAS Soil Texture leaflet 895.

Drought index values ranged from 0.0-4.9, and had a median value of 1.0 for winter wheat, 0.9 for oilseed rape, and 0.7 for spring barley.

2.3 Collation of site data

A spreadsheet was constructed from treated yields (t/ha) of winter wheat from 43 varieties across 506 trials (Recommended List, National List or British Society of Plant Breeders trials) in the UK with harvest during 1992-2002. Yield values were the predicted variety means using REML estimation for a mixed model with fixed variety effects and random terms to account for the block structure within the trial. Trials used 2, 3 or 4 replicates. The residual mean square error from each trial analysis was also stored. Site information collected from the trials included trial location (Ordnance Survey grid reference either provided for the trial or estimated from site names), altitude, date of sowing, soil pH, indication of soil type, index values for soil P, K and Mg status and amount of nitrogen applied. Missing values of site location were replaced where possible using knowledge of trial sites. These values were supplemented with soil type, redefined according to MAFF RB209 (Anon, 2000) as above, soil available water capacity (AWC) and drought index values derived as above. Trials included in the statistical analyses were required to contain at least 8 of the 43 varieties. Within this subset, complete or near-complete site data existed for 249 trials with harvest during 1993-2003.

3. Exploratory analysis: factor analysis models

3.1 Methods

Factor analysis models were fitted within a mixed model for variety yields, as described by Smith *et al.* (2001, 2002), to explore the nature of variety variation as a response to different environments between sites. Models were constructed for data from each year 1998-2002 separately in order to avoid potential confusion between variability due to differences between years and variability due to differences between trial sites. In this context, we regard yields from different trials as representing different traits for each variety, and model the covariance structure of the variety.trial interaction as separable, using a factor analysis model between trials within varieties, and assuming independence between varieties (see Smith 2001 for details). The factor analysis model represents the large number of environments by a much smaller number of hypothetical factors that we aim to interpret in terms of site characteristics. For each year, the factor analysis model effectively fits a linear regression model to the variety.trial interaction effects where both the explanatory covariate (the 'factor' of factor analysis) and the regression coefficients are estimated. The within-year model could be written as:

$$y_{ij} \quad = \quad c + t_i + u_{ij} + e_{ij} \tag{1}$$

where y_{ij} is the mean yield for variety i (i=1…m) from trial j (j=1…p) in the given year, c is a constant term, t_j represents the effect of trial j, u_{ij} represents the interaction of variety i and trial j, and e_{ij} represents error in the variety mean from the within trial analysis. Note that all effects are specific to the year under consideration. The constant and trial effects were fitted as fixed, and the variety.trial interaction and error were fitted as random. The variety.trial interaction was fitted assuming a separable structure, with independence between varieties, and a factor analysis model with up to 4 factors across trials. For k factors, the factor analysis model is defined by

$$u_{ij} \quad = \quad \sum_{r=1}^{k} \lambda_{jr} f_{ir} + \delta_{ij}$$

where λ_{jr} is the loading for trial j in factor r, and f_{ir} is the corresponding score for variety i in factor r. Using the analogy with linear regression, $\boldsymbol{\lambda}_r = (\lambda_{1r} \ldots \lambda_{pr})^{\mathrm{T}}$ is the estimated covariate, and f_{ir} is the corresponding (random) regression coefficient for variety i, with residual δ_{ij}. In addition, it is assumed $\boldsymbol{f}_r = (f_{1\mathrm{r}} \ldots f_{mr})^{\mathrm{T}} \sim$ N(0, $\boldsymbol{I}_m$) with $\boldsymbol{f}_r, \boldsymbol{f}_s$ independent for $r{\neq}s$ and $\boldsymbol{\delta}_j = (\delta_{1\mathrm{j}} \ldots \delta_{mj})^{\mathrm{T}} \sim \mathrm{N}(0, \psi_j I_m)$ with $\boldsymbol{\delta}_j$, $\boldsymbol{\delta}_l$ independent for $j{\neq}l$, and $\boldsymbol{f}_r$, $\boldsymbol{\delta}_j$ independent for all r, j. It then follows that for $\boldsymbol{u}_i = (u_{i1} \ldots u_{ip})^{\mathrm{T}}$,

$$\mathrm{var}(\boldsymbol{u}_i) \quad = \quad \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi}$$

where Λ is a $p \times k$ matrix with entries $[\Lambda]_{jr}=\lambda_{jr}$ and $\Psi=\mathrm{diag}(\psi_1 \ldots \psi_p)$, with $\boldsymbol{u}_i$, $\boldsymbol{u}_j$ independent for $i{\neq}j$. This covariance model allows for heterogeneity in the interaction effects between trials. Within each variety, the correlation between interaction effects from different trials reflects the overall similarity of variety responses between trials, and the variance of the effects depends on heterogeneity due to the estimated factors, $\boldsymbol{\lambda}_r$, and on the site specific variances, ψ_j. Because the data consisted of predicted variety means from each trial, plot error from the within-trial analysis and the variety.trial interaction become confounded. The plot error consists of noise that we wish to ignore, whereas we intend to model the variety.trial interaction. To separate the two terms, we approximate the plot error contribution for each trial by the trial residual mean square ($\hat{\sigma}^2_{p_j}$) divided by the variety replication used in the trial (n_j) and use a fixed diagonal variance matrix with

$$\mathrm{var}(e_{ij}) \quad = \quad \hat{\sigma}^2_{p_j} / n_j$$

Model (1) was fitted to data from each year by the REML method using the XFA option in ASREML (Gilmour *et al.*, 2002). Following analysis, the estimated factors were rotated to the principal components representation (see Smith *et al.*, 2001). The rotated factors then represented combinations of trials that maximised variety.trial interaction. When considered in relation to site characteristics, these explanatory covariates might give insight into site characteristics related to variation in variety yield. Rotated loadings were therefore plotted again the available site variables and simple linear regression was used to investigate these relationships. All post-processing was done using the Genstat statistical package (Payne, 2000).

3.2 Results

Factor analysis models were fitted to data from all trials with at least 8 varieties present within individual years from 1998 to 2002. Earlier years could not be used because the amount of data decreased rapidly. For 1997 and earlier, data was available for less than 20 varieties, compared with 38 varieties in 2000. Models with four factors were fitted in 1999-2002, with only three factors required in 1998. Following Smith *et al.* (2001), we calculated the percentage of genetic (variety) variation accounted for by rotated factor *r* as

$$100 \times \sum_{j=1}^{p} \tilde{\lambda}_{jr}^2 \Big/ \mathrm{trace}\left(\mathbf{\Lambda\Lambda}' + \mathbf{\Psi}\right)$$

where $\tilde{\lambda}$ represents the rotated loadings. The percentage of variety variation accounted for in total and by each of the estimated factors in each year is shown in Table 1. In each case, variation accounted for by the first factor includes the variety main effects. Site characteristics related to the estimated factors could be found in all years except 1999. Table 2 shows the site variables with adjusted R^2 value of >20% in a simple linear regression relating the rotated factor to the site variable.

Table 1. Summary of data sets with percentage of variety variation accounted for by estimated factors in factor-analysis models fitted to yield of winter wheat in treated trials with harvest during 1998-2002.

	Number of				% Variance accounted for by factor				
Harvest Year	varieties	trials	observations	factors	1	2	3	4	Total
2002	35	29	948	4	53.8	10.2	7.6	5.9	77.5
2001	36	34	713	4	48.7	14.6	10.6	7.7	81.6
2000	38	42	767	4	52.1	19.3	12.0	6.1	89.5
1999	29	37	587	4	71.0	10.5	7.5	5.4	94.4
1998	23	43	544	3	72.8	14.0	5.0	-	91.8

4. Random regression models

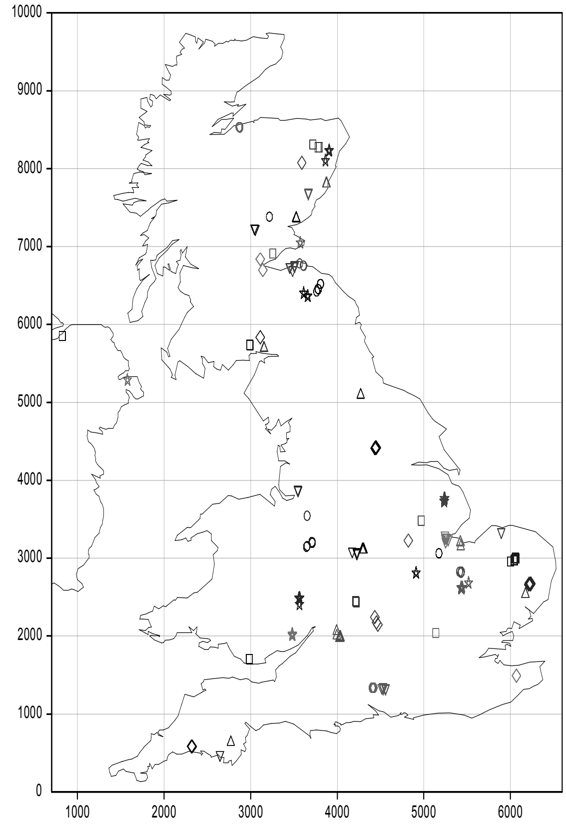
4.1 Methods

Within the dataset, although trials were sited in similar areas across years, these could be some distance apart. There was therefore a need to define locations to represent groups of trials in close proximity, which could be considered to experience the same climatic conditions. Locations were therefore defined using hierarchical cluster analysis on a similarity matrix constructed using Euclidean distances, using the complete linkage (furthest neighbour) clustering method, with the clustering threshold set so that sites within the same cluster (location) were <15km apart. This resulted in 56 location groups. Trial sites and their allocation to locations are shown in Figure 1.

Table 2. Site characteristics found to be correlated with the rotated factors estimated from model (1) for harvest years 1998-2002, with sign of correlation and adjusted R^2 value from simple linear regression of rotated factor on the site variable.

Harvest Year	Rotated factor	Site characteristics related to rotated factor	Sign of correlation	Adjusted R^2 (%)
2002	1	Site yield	+	31
	2	Site yield	+	25
	4	Longitude	-	41
	4	Latitude	+	30
	4	Soil pH	-	31
2001	1	Site yield	+	21
	1	Previous crop	-	20
	1	Soil pH	+	23
	2	Longitude	-	30
	2	Latitude	+	28
	2	Soil pH	-	38
	2	Drought index	-	28
	4	Applied N	+	21
	4	Lodging	-	20
2000	1	Latitude	+	22
	1	Soil pH	-	31
1998	1	Longitude	+	26
	1	Site yield	+	25
	2	Soil Mg index	-	23

Figure 1. Trial sites across the UK present in the statistical analysis, using OS grid coordinates at the scale 1 unit = 0.1km. Symbols and shades are used to group sites considered to be within the same location and locations have a diameter of <15km.



The factor analysis models strongly indicated site yield as a source of variation in variety yield. However, raw site average yields were strongly affected by site variables such as previous cropping, date of sowing and harvest year. The variety mean yields were used to predict expected site yields in a typical year for a first cereal crop and normal sow date. This decreased the correlation between site variables, and provided values that more closely corresponded with information available to the grower when predictions are required. The expected site yield values were obtained from a variance components model that accounted for type of previous cropping and date of sowing plus all sources of variation in the data, and could be written

$$y_{ijk} = c + p_{s(r)} + d_{t(r)} + v_i + Y_j + l_k + (Yl)_{jk} + T_{r(jk)} + (vY)_{ij} + (vl)_{ik} + (vYl)_{ijk} + (vT)_{ir(jk)} + e_{ir(jk)} \quad (2)$$

where y_{ijk} represents the trial mean yield from variety i (i=1…43) in harvest year j (j=1993…2002) at location k (k=1…56). Several trials might occur within the same location either within or across years, and trials are numbered as r=1…429 across years and locations, given by $r(jk)$. Then c is a constant term in the model, $s(r)$ indicates whether trial r is a first cereal crop ($s(r)$=1) or preceded by a cereal crop ($s(r)$=2) and $p_{s(r)}$ is the effect of previous cropping, $t(r)$ indicates the date of sowing for trial r as a categorical variable ($t(r)$=1: <7 Oct, 2: 7-15 Oct, 3: 15-30 Oct, 4: >30 Oct) and $d_{t(r)}$ is the effect of the sow date category. The constant, previous crop and sow date effects are all fitted as fixed terms in the model, with the remaining terms fitted as random. The main effects of variety i, year j, location k and trial r are represented by v_i, Y_j, l_k and T_r and are assumed mutually independent with associated variance components σ_i^2, σ_j^2, σ_k^2 and σ_r^2 respectively. Interactions of year.location, variety.year, variety.location, variety.year.location and variety.trial follow in the obvious notation for both effects and variance components, with all random effects assumed mutually independent. As for the factor analysis model, plot error from the within-trial analysis is represented by e_r with a fixed diagonal variance matrix with values

$$\mathrm{var}\left(e_{ir}\right) \;=\; \hat{\sigma}^2_{p_r} \Big/ n_r$$

The model was fitted by REML estimation using the mixed model facilities in Genstat (Welham & Thompson, 2000). Expected site yields for a first cereal crop sown during 7-15 October in an average year for an average variety were predicted for each trial from the fitted model using the methods of Welham *et al.* (2004). The expected site yield for trial r is denoted z_r.

Model (2) can be regarded as the baseline model for the variety mean yields. The aim of the statistical modelling in this section is to explain the variety.location and variety.trial interactions in terms of site variables and produce a model to produce improved predictions of relative variety yield at a specific site. For convenience, we abbreviate model (2) as

$$y_{ijk} = c + p_{s(r)} + d_{t(r)} + v_i + u_{ijk} + e_{ir} \qquad (3)$$

where u_{ijk} represents the composite term

$$u_{ijk} = Y_j + l_k + (Yl)_{jk} + T_{r(jk)} + (vY)_{ij} + (vl)_{ik} + (vYl)_{ijk} + (vT)_{ir(jk)}$$

with definitions as above.

Expected site yield was added to the baseline model (2) via an overall regression term representing the average yield response to changes in expected site yield, and random regression terms for the variety responses to expected site yield. The random regressions assume that the regression intercept and slope vary according to a bivariate normal distribution across the set of varieties, with variance and covariance parameters to be estimated. Use of a random rather than fixed interaction has the advantage that the predicted variety response to site yield then has minimum mean square of prediction. The model including expected site yield then takes the form

$$y_{ijk} = c + p_{s(r)} + d_{t(r)} + a_z z_r + v_i + w_{zi} z_r + u_{ijk} + e_{ir} \qquad (4)$$

where a_z is a fixed parameter representing the overall response in treated yield to a unit increase in expected site yield, and w_{zi} is a random parameter representing the deviation in response to a unit increase in expected site yield for variety i, with

$$\begin{pmatrix} \boldsymbol{v} \\ \boldsymbol{w}_z \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{I}_{43} \otimes \begin{bmatrix} \sigma^2_v & \sigma_{vz} \\ \sigma_{vz} & \sigma^2_z \end{bmatrix} \right) = \mathrm{N}\left(0, \boldsymbol{I}_{43} \otimes \boldsymbol{\Sigma}_2 \right)$$

for $\boldsymbol{w}_z = (w_1 \ldots w_{43})^{\mathrm{T}}$. The variety effects v_i then represent deviations in the intercept for variety i. The covariance parameter σ_{vz} allows correlation in variations between intercept and slope for each variety and makes the model invariant to changes of origin in the regression variable.

Latitude and longitude were also strongly indicated as important variables from the factor analysis models. As the two variables together summarise average climatic conditions, it was decided to build the model in terms of a two-dimensional smooth surface in terms of latitude and longitude for each variety, fitted using a smoothing spline. Let the OS co-ordinates of trial r be represented by (x_{1r}, x_{2r}) where x_{1r} is the latitude and x_{2r} is the longitude. The definition of the two-dimensional smoothing spline is given in Appendix

A, and requires definition of a set of knots. The knots were chosen to be representative of the trial locations and were determined as for the location grouping, but with the clustering threshold set to 70km. This gave 27 groups, and the 27 group average locations were used as the set of knot points. The two-dimensional spline term could then be added to model (4):

$$
\begin{aligned}
y_{ijk} = \; & c + p_{s(r)} + d_{t(r)} + a_z z_r + a_1 x_{1r} + a_2 x_{2r} + \\
& v_i + w_{zi} z_r + w_{1i} x_{1r} + w_{2i} x_{2r} + \sum_{j=1}^{24} E_{rj} (\delta_j + \varepsilon_{ij}) + u_{ijk} + e_{ir}
\end{aligned}
\tag{5}
$$

where E_{rj} represents the rjth element of matrix $\boldsymbol{E}_x$ defined in appendix A, $\boldsymbol{\delta} = (\delta_1 \dots \delta_{24})^{\mathrm{T}}$ are random effects associated with an overall smooth response to location, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1} \dots \varepsilon_{i24})^{\mathrm{T}}$ are random effects associated with the smooth response to location for variety i, with

$$
\boldsymbol{\delta} \sim \mathrm{N}(0, \sigma_{s1}^2 \boldsymbol{G}_{\mathrm{s}}), \quad \boldsymbol{\varepsilon}_i \sim \mathrm{N}(0, \sigma_{s2}^2 \boldsymbol{G}_{\mathrm{s}}),
$$

and $\boldsymbol{\delta}$, $\boldsymbol{\varepsilon}_i$ mutually independent for all i=1…43 and $\boldsymbol{\varepsilon}_i$, $\boldsymbol{\varepsilon}_j$ mutually independent for $i \neq j$. In addition, the random coefficients (v_i, w_{zi}, w_{1i}, w_{2i}) then have a joint distribution with zero mean and an unstructured 4×4 covariance matrix $\boldsymbol{\Sigma}_4$, with 10 unknown parameters, *ie*. for $\boldsymbol{w}_j = (w_{j1} \dots w_{j43})^{\mathrm{T}}$, j=1,2 then

$$
\begin{pmatrix} \boldsymbol{v} \\ \boldsymbol{w}_z \\ \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \end{pmatrix} \sim \mathrm{N}\left(0, \boldsymbol{I}_{43} \otimes \boldsymbol{\Sigma}_4 \right)
$$

The smooth spline surface was designed to represent large-scale trend in variety response to climate. It was expected that there might also be more local trend in variety yields, and a spatial model with exponential correlations was fitted to represent this small-scale trend. The exponential correlation model assumes correlations of the form $\varphi^{d(r,s)}$ between locations r and s, where $d(r,s)$ represents the Euclidean distance between the two locations and φ is a parameter to be estimated with $|\varphi|<1$. Two forms of the local spatial model were tried, the first model fitted spatial correlations across locations within varieties using the variety.location interaction, the second model fitted spatial correlations across locations within varieties and years using the variety.year.location interaction. The fit of the two models was compared using the residual log-likelihood of the fitted models, and the model with the highest log-likelihood was chosen.

There was little consistent evidence from the factor analysis models to link other site variables with variety by location interactions, and so each site variable was considered in turn to see if it could be used to improve the model. The remaining site variables were redefined as categorical variables, so that both linear and non-linear interactions with variety could be detected. Each categorical site variable in turn was added into model (5) as a fixed term, to represent the overall effect of each category, then as a random interaction with variety so that any interaction could be detected. The random interaction was fitted using either a common variance parameter or separate variance parameters for each level of the categorical variable. Fitting separate variance parameters indicated whether the interaction was associated with specific categories. The two random interaction models were compared to the model without the random interaction using -2 × difference in residual log-likelihood, and comparing this to a mixture of χ^2 distributions according to the results of Stram and Lee (1994). A final model was constructed by including all the site categorical variables found to be individually significant, then omitting terms to find the best subset, again using the change in residual log-likelihood to check for improvements in the model.

To investigate whether there was local variation in the relationships, the same modelling process was undertaken with two subsets of the data. The first subset was defined to lie within latitudes 1000-4000 and longitude 3000-7000, using the OS grid co-ordinates shown on Figure 1. This area encompassed Central, East and Southern (CES) England, an area of about 300 × 400 km and included 2320 data values. The other subset covered Eastern Scotland, within ~50km of the east coast between latitudes 6000-8500, an area of 250 × 100 km containing 995 data values. Both subsets formed contiguous areas with reasonable geographic coverage of trial sites.

All models were fitted by REML estimation of variance parameters using either the mixed model facilties in GenStat (Payne, 2000), or the program ASREML (Gilmour *et al.*, 2002).

4.2 Results

The estimated variance components from model (2) are shown in table 3, labelled according to the symbolic form of the model:

fixed ~ constant + prevcrop + fsowdate

random ~ variety + year + location + year.location + trial + variety.year + variety.location + variety.year.location + variety.trial + ploterror

var(ploterror) = diag[$\hat{\sigma}^2_{p_r} / n_r$; r=1…429]

where all variable names indicate factors, and the variance matrices for individual terms are identity matrices scaled by a variance component unless indicated otherwise. There were large components of variance due to years, locations, the year.location interaction and trials, reflecting patterns in variation of overall trial yield. The presence of a large location component indicated similarity between trials within the same geographical location, and justified the use and definition of this term. However, the large trial variance component indicated that there was also substantial variation between trials within locations.

Table 3. Estimated variance components from model (2) without site variables, model (6) with site variables for geographic location and expected site yield, and model (7), the final fitted model.

Model term	Estimated variance component from		
	Model (2)	Model (6)	Model (7)
year	.5375	.5010	.5027
location	.5420	0	0
year.location	.5785	.0599	.0631
trial	.6596	.0550	.0536
variety	.1629	.1501	.0015

variety.lin(long)	-	.0016	.0015
variety.lin(lat)	-	.0041	.0044
variety.lin(siteyield)	-	.0048	.0048
variety.spl(lat,long)	-	.2325	.2249
variety.latesown	-	-	.0090
variety.soilAS	-	-	.0073
variety.lowK	-	-	.0075
variety.year	.0255	.0150	.0152
variety.location	.0192	.0088	.0066
variety.year.location	.0408	.0457	.0466
variety.trial	.0603	.0630	.0606

After adding random regressions for expected site yield, latitude and longitude and the two-dimensional spline terms to fit model (5), it was discovered that the overall spline term had a zero variance component and so was dropped from the model. On investigating the need for local spatial models, it was found that the model did not converge when fitting spatial correlation within the variety.location term, but that the residual log-likelihood increased substantially when spatial correlation was fit across locations within the variety.year.location term. The symbolic form of the fitted model could then be written as:

fixed ~ constant + prevcrop + fsowdate + lin(siteyield) + lin(lat) + lin(long)

random ~ variety + variety.lin(long) + variety.lin(lat) + variety.lin(siteyield) + variety.spl(lat,long)
+ year + location + year.location + trial + variety.year + variety.location
+ variety.year.location + variety.trial + ploterror

var(ploterror) = diag[$\hat{\sigma}^2_{p_r} / n_r$; r=1…429]

var(variety, variety.lin(long), variety.lin(lat), variety.lin(siteyield)) = $\boldsymbol{I}_{43} \otimes \boldsymbol{\Sigma}_4$

var(variety.year.location) = $\boldsymbol{I}_{43} \otimes \boldsymbol{I}_{10} \otimes \exp(\varphi)$ (6)

where lin(x) indicates a linear function of the variable x after mean correction, and spl(x,z) indicates a two-dimensional spline in terms of variables x and z, as defined in Appendix A, exp(φ) represents an exponential correlation function with parameter φ, siteyield represents the expected site yield values, and lat, long represent the latitude and longitude in terms of OS grid co-ordinates for each trial. The estimated variance parameters for model (6) are shown in table 3. Including the expected site yield as a covariate reduced the location, year.location and trial variance components substantially. The year component was largely unaffected because the expected site yield values were adjusted to average over year effects. The variety.year component was reduced as a side-effect of introducing spatial correlation within the variety.year.location interaction. The variety.location component was reduced due to the interaction of variety interactions with expected site yield and the spine term representing large-scale trend. Within the small-scale spatial trend term, the spatial correlation parameter φ was estimated as 0.58, indicating a correlation of 0.58 between trials 100km apart. Model (6) was then the best model that could be found for variety interactions in terms of the expected siteyield and geographic location variables and was used as a baseline for testing whether other site variables could be used to improve the model.

Table 4 shows the definition of the additional site variables tested, the change in log-likelihood and whether there was evidence of an interaction with variety.

Table 4. Site variables tested in a random regression relationship for association with variety variability, in a model containing variety interactions with site yield and location.

Site variable	Category boundaries in definition of the variable	Evidence of interaction with variety?	Increase in log-likelihood on adding interaction with	
			common variance	separate variance for each category
Soil type	AS, MOP, CZ	Yes	0	2.98
Altitude	50m, 100m	No	0	0.18
Sowdate	7 Oct, 15 Oct, 30 Oct	Yes	0.64	4.68
Soil AWC	150, 225	No	0.63	0
Soil pH	6.25, 7.75	No	0	0
Soil P index	1.5, 3	No	0	0.18
Soil K index	1.5, 3	Yes	0.09	2.67
Drought index	0.1, 0.5, 1	No	0.39	0.74
Previous crop	Non-cereals, cereals	No	0.84	0.13
Soil Mg index	1.5, 3	No	0	0.01
N applied	200, 225	No	0.04	0.21

Evidence of an interaction was found with soil type, date of sowing and soil K index, with the variety response associated with specific categories: soil type AS (shallow or sandy), sowing date before 7 October or after 30 October, and soil K index <1.5. New variables were calculated to represent the single categories, taking value 1 for units within the category and zero otherwise. On fitting a joint model containing variety interactions with location, expected site yield, and the additional variables shown to be associated with variety variability, it was found that the early sowdate variable did not improve the model and so was omitted. The variables 'late sowdate', 'soil type AS' and 'low K index' were retained to give a final model, in symbolic form:

fixed ~ constant + prevcrop + fsowdate + lin(siteyield) + lin(lat) + lin(long) + soilAS + lowK

random ~ variety + variety.lin(long) + variety.lin(lat) + variety.lin(siteyield) + variety.spl(lat,long)
+ variety.latesown + variety.soilAS + variety.lowK
+ year + location + year.location + trial + variety.year + variety.location
+ variety.year.location + variety.trial + ploterror

$$\text{var(ploterror)} = \text{diag}[\ \hat{\sigma}^2_{p_r} / n_r\ ; r=1\ldots429\ ]$$

$$\text{var(variety, variety.lin(long), variety.lin(lat), variety.lin(siteyield))} = \boldsymbol{I}_{43} \otimes \boldsymbol{\Sigma}_4$$

$$\text{var(variety.year.location)} = \boldsymbol{I}_{43} \otimes \boldsymbol{I}_{10} \otimes \exp(\boldsymbol{\varphi}) \qquad (7)$$

where latesown is a 0/1 variable indicating trials sown after 30 October, soilAS is a 0/1 variable indicating trials on sites with sandy/shallow soil, and lowK is a 0/1 variable indicating trials with soil K index < 1.5. The estimated variance components from model (7) are shown in table 3. Within the small-scale spatial trend term, the spatial correlation parameter φ was estimated as 0.56. There were few changes in variance parameters from model (4), but there was a further reduction in the unexplained variety.location interaction.

The estimated correlations (in the variance matrix $\boldsymbol{\Sigma}_4$) between the variety intercepts and slopes for the random regressions on longitude, latitude and expected site yield are shown in table 5. There were high correlations between the variety intercepts and slopes for the regressions on longitude and expected site yield, indicating a connection between variety overall yield and response to both expected site yield and longitude.

Table 5. Estimated correlation parameters in variance matrix $\boldsymbol{\Sigma}_4$ from model (7) between the variety intercepts and slopes for the random regressions on longitude, latitude and expected site yield.

Variety intercepts	1	0.78	0.28	0.89
Variety.lin(long) slopes	0.78	1	0.11	0.78
Variety.lin(lat) slopes	0.28	0.11	1	-0.06
Variety.lin(siteyield) slopes	0.89	0.78	-0.06	1
	Variety intercepts	Variety.lin(long) slopes	Variety.lin(lat) slopes	Variety.lin(siteyield) slopes

Estimated fixed effects from model (7) are shown in table 6, and represent the effects of site variables on overall trial yield. These terms cannot be considered a sensible model for overall site yield because of the regression on expected site yield, although this term is essential in understanding variety interactions with site. The regression coefficient for expected site yield is greater than 1.0 due to shrinkage in the prediction of expected site yields, and because the other site conditions that impinge on overall yield all cause a reduction in yield. The estimated regression coefficients for longitude and latitude, and the effects of low soil K index and sandy/shallow soil are all close to zero, but are retained in the model so that variety interactions can be modelled as deviations from the overall value.

Table 6. Estimated fixed effects from model (7). Longitude and latitude are measured as OS grid co-ordinates. †Intercept term estimates trial yield for longitude=4189, latitude=4027 and expected site yield of 10.2 t/ha, for a trial with non-cereal previous crop sown before 7 October with non-sandy/shallow soil and soil K index > 1.5. Effects different from zero are indicated by *.

Model term	Estimated effect	Standard error
Intercept†	10.41*	0.24
Slope of regression on (longitude − 4189)	-0.61×10^{-5}	2.53×10^{-5}
Slope of regression on (latitude − 4027)	-0.40×10^{-5}	1.49×10^{-5}
Slope of regression on (expected site yield − 10.20)	1.370*	0.287
Cereal previous crop	-0.657*	0.056
Sowdate 7-15 October	-0.192*	0.062
Sowdate 15-30 October	-0.347*	0.071
Sowdate >30 Oct	-0.788*	0.075
Site with soil K index < 1.5	-0.066	0.071
Site with sandy/shallow soil	0.011	0.064

Some of the variety interaction terms in the model are suitable for use in prediction of relative variety yields at new sites, but other terms cannot be used for prediction. To assess the impact of the model on prediction of relative variety yields, the percentage of variety variation in the dataset associated with each variety interaction term was calculated as

$$100 \times \mathrm{trace}\left(\boldsymbol{Z}_i \boldsymbol{G}_i \boldsymbol{Z}_i^T\right) \Big/ \sum_i \mathrm{trace}\left(\boldsymbol{Z}_i \boldsymbol{G}_i \boldsymbol{Z}_i^T\right)$$

where $\boldsymbol{Z}_i$ is the design matrix with respect to the full dataset for the ith variety interaction term with variance matrix $\boldsymbol{G}_i$, where $\boldsymbol{G}_i$ includes any scalar parameters, and i sums over all variety (main effect and) interaction terms in the model. This calculation reflects both the amount of variation associated with the term and the importance of the term in the dataset. The contribution of longitude and latitude were calculated jointly, including both the correlated random regression terms and the two-dimensional spline term at the mean value of expected site yield (10.2 t/ha). The contribution for expected site yield was calculated at the mean value of longitude (=4189) and latitude (=4027). For variety interaction terms, Table 7 shows the estimated variance component associated with each term, whether it could be used for prediction and the calculated percentage of the total variety variation accounted for by the term. Variety main effects accounted for 43% of the variety variation in the model, with interactions with site accounting for 21% and interactions with year accounting for 36%. Although this model accounted for most of the variety × site variability, the year-to-year variation was still unpredictable. Predictions from the model, appropriate to a notional 'typical' year, may therefore tend to be inaccurate for any particular year. However, the model could predict which varieties tended to prefer which environments in a typical year.

Table 7. Variety interaction terms in the final model. Terms are classified according to whether they can be used to predict variety yield at a new site (given appropriate site information) and the percentage of the total variety variation in the dataset accounted for by each term.

Model term	Type of term	% variety variation accounted for
Variety	Predictive	43.3
Variety.spl(location)	Predictive	15.6
Variety.siteyld	Predictive	2.0
Variety.latesown	Predictive	0.4
Variety.soilAS	Predictive	0.5
Variety.lowK	Predictive	0.4
Variety.site	Non-predictive	1.9
Variety.year	Non-predictive	4.5
Variety.year.spatial(location)	Non-predictive	13.7
Variety.trial	Non-predictive	17.8

The three largest sources of variety variability (after the main effects) were the large-scale location trend, small-scale location trend and noise (variety.trial). The large-scale location trend could be used in prediction and may represent variety response to large-scale changes in climatic conditions across the UK. Figures 2-4 in appendix B show the predicted response to large-scale trends in climate across the UK for a selection of varieties as a deviation from variety mean yield. These surfaces were calculated assuming an expected site yield of 10.20 t/ha held constant across the UK, with a first wheat crop sown date 7-15 October on a site with soil K index > 1.5 and non-sandy/shallow soil. To assess the strength of the climatic trend, predictions were also made at the 27 knot points and compared to the mean across those 27 sites. Varieties Claire, Hereward, Hunter, Madrigal, Malacca, Mercia, Rialto, Scorpion 25, Soissons, Spark and Tanker showed large variation in yield across the 27 sites compared to the mean. These varieties are shown in figures 2-4, along with variety Consort that showed little response to climatic conditions. For the small-scale location trend, the estimated spatial correlation parameter indicated strong correlation (0.89) between locations 20km apart, dropping to 0.75 for locations 50km apart and 0.56 for locations 100km apart. This suggests that micro-climates (~50km diameter) existed within which particular varieties performed better (or worse) than expected. However, this small-scale pattern appeared to be inconsistent across years, and so could not be used to predict relative variety yield at a given site. The variety.trial interaction term represented random variation in variety response across trials and could be regarded as random noise, in that it had no apparent pattern and would be unpredictable for new sites or future years.

Although the remaining site variables each accounted for only a small percentage of the variety variation, each variable showed evidence of a large effect on the yield of one or more varieties, and should not be ignored when predicting site yield for these varieties. The BLUPs for variety main effects and interactions with expected site yield, late-sown crops, sandy/shallow soil and low soil K index are shown in table 10 in Appendix C. Varieties Robigus, Scorpion 25, Tanker and Xi 19 had a positive interaction with expected site yield, *ie*. these varieties showed a larger increase in yield than average as overall site yield increased. In contrast, varieties Hereward, Malacca, Mercia, Shamrock Soissons and Spark had a negative interaction with expected site yield, *ie*. these varieties showed a smaller increase in yield than average as overall site yield increased. The variety interaction effects with expected site yield showed a strong correlation with overall variety yield, indicating that high-yielding varieties tended to give a better response to high-yielding sites. Variety Riband had a negative interaction with late sowing, indicating a worse response then average to these conditions. Varieties Hereward and Rialto had a negative interaction with shallow or sandy soils, while Deben had a positive interaction. Finally, variety Hereward had a positive interaction with low soil K index.

Analysis of the two subsets (Central, East & South (CES) England or East Scotland) was used to examine whether effects were consistent across the whole UK. In both cases, there was no evidence of either large-scale or small-scale smooth non-linear location trend in variety response. The large-scale trend was not expected to be found at the scale of the subsets. The small-scale trend was expected to be found, but there was possibly not enough information to estimate the spatial parameters within the subsets. Where variety interaction terms occurred in models for both the full data and subsets, the variance components were of similar size (see table 8). For the CES England subset, variety yields were found to be related to position along a NW-SE axis within the area, expected site yield, late sown crops, sandy/shallow soils, previous crop and soil pH. For the East Scotland subset, variety yields were found to be related to latitude, expected site yield and late sown crops. However note that there were very few sandy/shallow soils or late sow dates within the East Scotland subset. Correlations between the estimated effects from the full analysis and the subsets are shown in Table 9.

Table 8. Estimated variance components from the final fitted model (7) for the full data set, and from models for the CES England and East Scotland subsets

Model term	Estimated variance component from		
	Model (7) full data set	CES England subset	East Scotland subset
year	.5027	.5521	.4314
location	0	0	.0345
year.location	.0631	.0670	.0074
trial	.0536	.0638	.0381
variety	.0015	.1780	.1853
variety.lin(long)	.0015	-	-
variety.lin(lat)	.0044	-	.0053
variety.lin(NW-SE axis)	-	.0021	-
variety.lin(siteyield)	.0048	.0064	.0075
variety.spl(lat,long)	.2249	-	-
variety.latesown	.0090	.0082	.0612
variety.soilAS	.0073	.0044	-
variety.lowK	.0075	-	-
variety.prevcrop	-	.0048	-
variety.pH	-	.0016	-
variety.year	.0152	.0294	.0368
variety.location	.0066	0	.0009
variety.year.location	.0466	.0277	.0194
variety.trial	.0606	.0532	.0613

Table 9. Correlation between BLUPs for variety main effects and interactions from the final fitted model (7) for the full data set, and from models for the CES England and East Scotland subsets

Model term	Correlation between BLUPs from		
	Full data vs CES England subset	Full data vs East Scotland subset	CES England vs East Scotland subsets
variety	0.912	0.861	0.811
variety.lin(lat)	-	0.636	-
variety.lin(siteyield)	0.961	0.715	0.752
variety.latesown	0.797	0.461	0.191
variety.soilAS	0.851	-	-

There was high correlation between the variety main effects and variety.siteyield interactions across all three models. There was also high correlation between the variety.latesown interaction effects from the full data set and the CES England subset, but much lower correlation with the East Scotland subset. This difference could be accounted for by the small number of late-sown crops in the East Scotland subset.

5. Discussion

The statistical analysis identified site variables associated with variation in relative variety yields between sites. Including these variables in the model increased the percentage of variety variation accounted for by terms useful in prediction from 53% in model (2), excluding site variables, to 62% in the final fitted model (7). The variety.location component of variance, that describes variety.site variation not accounted for by other site variables, was reduced to 34% of its original value. The final fitted model would therefore be expected to give an improved prediction for specific sites. However, there remains much variation associated with variety by year interactions that remained unexplained and would add much uncertainty to predictions. The results of the statistical analysis can be used to optimise use of the 'Varieties on your Farm' module of RL *Plus*. In general, geographic location appears to be the most important site variable influencing variation in variety yields across the UK. However, for particular varieties, the expected site yield, soil type or soil K index may be equally important. An improved version of RL *Plus* might use the results of these analyses as prior information in the model fitting process.

The subset analyses indicated that although the relationship of variety yield with expected site yield held across the UK, site variables such as previous crop and soil pH might have a more local interaction with variety. Preliminary work using varying coefficient models indicates that previous crop may have a spatially-varying interaction with variety. Further work and an extension of the methodology are required to determine the best methods to fit and detect local influence of site variables on variety yield.

Inclusion of expected site yield as a variable in the model gives rise to several potential problems. Within the model, expected site yield is treated as a known variable, whereas in fact it has been estimated with error from the response variable in the model. We have ignored these issues because of the clear influence of site yield on variety interactions. Further work is required to determine better statistical methods for inclusion of site yield in the model. Problems of interpretation also arise from the inclusion of expected site yield in the model, as expected site yield also varies spatially across the UK. The contour maps in Figures 2-4 were presented for a fixed value of expected site yield. An improved prediction of variety yield could be calculated taking into account both geographic position and site yield as a function of geographic position. In both cases, care is required in interpretation of the predicted surfaces.

Several statistical issues were raised during development and interpretation of the models. It had been hoped that use of the factor analysis models would lead to the detection of important interactions between site variables. Use of the factor analysis models within single years meant that there were too few sites within each year to reliably identify interactions. Using a combination of years within the factor-analysis models would have increased the number of sites and the chance of detecting interactions, but could have introduced year-to-year variation that could not be modelled by site variables. Development of an improved factor-analysis model that separates site and year variation is required to solve these problems. One advantage of the factor analysis models is their ability to model heterogeneity between environments, and the random regression models used here could possibly be improved by using a factor-analysis term to account for residual variation rather than simple variance components terms. Again, improvements in model fitting techniques are required to make this a realistic proposition. A curious feature of the scan for the inclusion of site variables into model (6) (see table 4) was that none of the homogenous variety interactions with site variables indicated that the interaction would improve the model, whereas an improvement was clear for several variables with the heterogeneous model. This might be explained by sub-optimal categorisation of the variables, although several other categorisations were also tried, but indicates a potential problem in detecting interactions with standard variance component models where the interaction occurs at only one factor level.

In conclusion, the statistical methods used in this project have proved useful in detecting and describing variety interactions with site variables. Improvements in the techniques could lead to further improvements in the modelling process, but this potential improvement is limited by the large proportion of variety variation associated with unpredictable weather factors. The same modelling process could be undertaken with summary weather variables, and could lead to an improvement in variety prediction if relatively reliable quantitative long-term weather forecasts became available.

6. References

Anon (2000) *Fertiliser Recommendations for Agricultural and Horticultural Crops,* MAFF Reference Book 209, 7th Edition.

Denis JB, Piepho HP & van Eeuwijk FA (1997) Modelling expectation and variance for genotype by environment data. *Heredity* **79**, 162-171.

Denis JB, Piepho HP & van Eeuwijk FA (1998) Predicting cultivar difference using covariates. *JABES* **3**, 151-162.

Foulkes, M.J., Scott, R.K. & Sylvester-Bradley, R. (2001). A comparison of the ability of wheat cultivars to withstand drought in UK conditions: resource capture. *Journal of Agricultural Science* **137**, 1-16.

Gilmour AR, Gogel BJ, Cullis BR, Welham SJ & Thompson R (2002) *ASREML User Guide Release 1.0*. VSN International Ltd, Hemel Hempstead UK, 290pp.

Green PJ & Silverman BW (1994) *Nonparametric regression and generalised linear models.* Chapman & Hall, London. 182pp.

Palmer WC. (1965). *Meteorological Drought.* Research Paper No. 45, 58pp. Department of Commerce, Washington DC. (cited by Dai, A. Trenberth, AE & Tarl, T. (1998). Global variations in droughts and wet spells: 1900-1995. *Geophysical Research Letters* **25**, 3367-3370.)

Payne RW (2000) *The Guide to GenStat*. VSN International, Hemel Hempstead, UK.

Smith AB, Cullis BR & Thompson R (2001) Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics* **57**, 1138-1147.

Smith A, Cullis B & Thompson, R (2002) Exploring variety-environment data using random effects AMMI models with adjustments for spatial field trend, Parts 1 & 2. In: *Quantitative Genetics, Genomics and Plant Breeding*. Ed: M S Kang. CABI, UK. 400pp.

Stram DA & Lee JW (1994) Variance components testing in the longitudinal mixed effects model. *Biometrics* **50**, 1171-1177.

Talbot M (1984) Yield variability of crop varieties in the UK. *Journal of Agricultural Science* **102**, 315-321.

Theobald C, Talbot M & Nabugoomu (2002) A Bayesian approach to regional and local-area prediction from crop variety trials. *JABES* **7**, 403-419.

Thomasson, A.J. (1979). Assessment of soil droughtiness. In: Soil survey applications (ed. M. G. Jarvis & D. Mackney) *Soil. Surv. TECH. Monogr.* **13**, 43-50.

Wahba G (1990) *Spline models for observational data.* SIAM: Philadelphia, 169pp.

Welham SJ, Cullis BR, Gogel BJ, Gilmour AR & Thompson R (2004) Prediction in linear mixed models. *Australian and New Zealand Journal of Statistics* **46**, 325-347.

Welham SJ & Thompson R (2000) Chapter 5: REML analysis of mixed models. In *The Guide to GenStat, Part 2: Statistics*, Ed. RW Payne. VSN International, Hemel Hempstead, UK.

Appendix A. Definition of two-dimensional smoothing spline

The two-dimensional smoothing spline used in the statistical analysis is derived as a thin-plate spline (Green & Silverman, 1994) with reduced knots, and fitted within a mixed model using REML estimation of the smoothing parameter. For simplicity, we consider a model with data $\boldsymbol{y} = (y_1 \dots y_n)$ measured at n locations with co-ordinates $(\boldsymbol{x}_1, \boldsymbol{x}_2)$, $\boldsymbol{x}_j = (x_{j1} \dots x_{jn})^{\mathrm{T}}$, j=1,2.

Following Green & Silverman (1994, p142), for a set of r knots $(\boldsymbol{t}_1, \boldsymbol{t}_2)$, $\boldsymbol{t}_j = (t_{j1} \dots t_{jr})^{\mathrm{T}}$, j=1,2, the bivariate function $g(u,v)$ is a natural thin-plate spline if and only if g is of the form

$$g(u,v) = a_1 + a_2 u + a_3 v + \sum_{j=1}^{r} c_j\, \eta\left(d\left(u,v;t_{1j},t_{2j}\right)\right)$$

for unknown parameters $\boldsymbol{a} = (a_1, a_2, a_3)^{\mathrm{T}}$ and $\boldsymbol{c} = (c_1 \dots c_r)^{\mathrm{T}}$ with

$$\begin{aligned} \eta(r) &= \frac{1}{16\pi} r^2 \log\left(r^2\right) \quad \text{for} \quad r > 0 \\ \eta(0) &= 0 \\ d(u_1,u_2;v_1,v_2) &= \left[(u_1 - v_1)^2 + (u_2 - v_2)^2\right]^{0.5} \\ \boldsymbol{T}^{\mathrm{T}}\boldsymbol{c} &= 0 \quad \text{for} \quad \boldsymbol{T} = \left(\boldsymbol{1}\ \boldsymbol{t}_1\ \boldsymbol{t}_2\right) \end{aligned}$$

The vector of spline values at the observed locations is then written

$$g(\boldsymbol{x}_1,\boldsymbol{x}_2) = \boldsymbol{X}\boldsymbol{a} + \boldsymbol{E}_x\boldsymbol{c}$$

where $\boldsymbol{X} = (\boldsymbol{1}\ \boldsymbol{x}_1\ \boldsymbol{x}_2)$, $[\boldsymbol{E}_x]_{ij} = \eta(\ d(\ x_{1i}, x_{2i}\ ;\ t_{1j}, t_{2j}\ )\ )$ and we explicitly impose the constraint $\boldsymbol{T}^{\mathrm{T}}\boldsymbol{c} = \boldsymbol{0}$ by finding a matrix $\boldsymbol{C}$ such that $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{T} = \boldsymbol{0}$, then $\boldsymbol{c} = \boldsymbol{C}\boldsymbol{\delta}$ for unknown parameters $\boldsymbol{\delta}$. The spline is fitted by minimising the penalised sum of squares

$$\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{a} - \boldsymbol{E}_x\boldsymbol{C}\boldsymbol{\delta}\right)^{\mathrm{T}}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{a} - \boldsymbol{E}_x\boldsymbol{C}\boldsymbol{\delta}\right) + \lambda\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{E}_t\boldsymbol{C}\boldsymbol{\delta}$$

in terms of $\boldsymbol{a}$ and $\boldsymbol{c}$, where $[\boldsymbol{E}_t]_{ij} = \eta(\ d(\ t_{1i}, t_{2i}\ ;\ t_{1j}, t_{2j}\ )\ )$ for a given smoothing parameter λ. Wahba (1990, p32) shows that the matrix $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{E}_t\boldsymbol{C}$ is positive definite. Following the arguments of Verbyla *et al.* (1999) in the univariate case, it can then be shown that, for fixed λ, the solution that minimises the penalised sum of squares is a best linear unbiased predictor (BLUP) from a mixed model with the constant, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ as fixed terms and random term with design matrix $\boldsymbol{E}_x$ and random effects $\boldsymbol{\delta} = (\delta_1 \dots \delta_{r-3})^{\mathrm{T}}$ with

$$\boldsymbol{\delta} \sim \mathrm{N}\left(0, \sigma_s^2\boldsymbol{G}_s\right)$$

where $\boldsymbol{G}_s = (\boldsymbol{C}^{\mathrm{T}}\boldsymbol{E}_t\boldsymbol{C})^{-1}$ and $\lambda = \sigma^2 / \sigma_s^2$ where σ^2 represents the residual variance. We then estimate the smoothing parameter using REML estimation in analysis of the mixed model via the variance component σ_s^2. The advantage of this formulation is that the two-dimensional smoothing spline term can be included alongside other terms in the mixed model.

Appendix B. Varietal response to large-scale climate trends modelled as a two-dimensional smoothing spline

Figure 2: Predicted response in treated yield (t/ha) to large-scale trends in climate for varieties Claire, Consort, Hereward and Hunter. Values shown are deviations from average yield for each variety, with expected site yield held constant at 10.2t/ha across the UK, assuming first wheat crop with sow date 7-15 October. Varieties Claire, Hereward and Hunter show strong evidence of response to location. Trial sites where each variety was grown are indicated by ×. Contour lines for yield deviations: 1=-0.3 2=-0.25, 3=-0.2, 4=-0.15, 5=-0.1, 6=0, 7=0.1, 8=0.15, 9=0.2, 10=0.25 11=0.3t/ha.

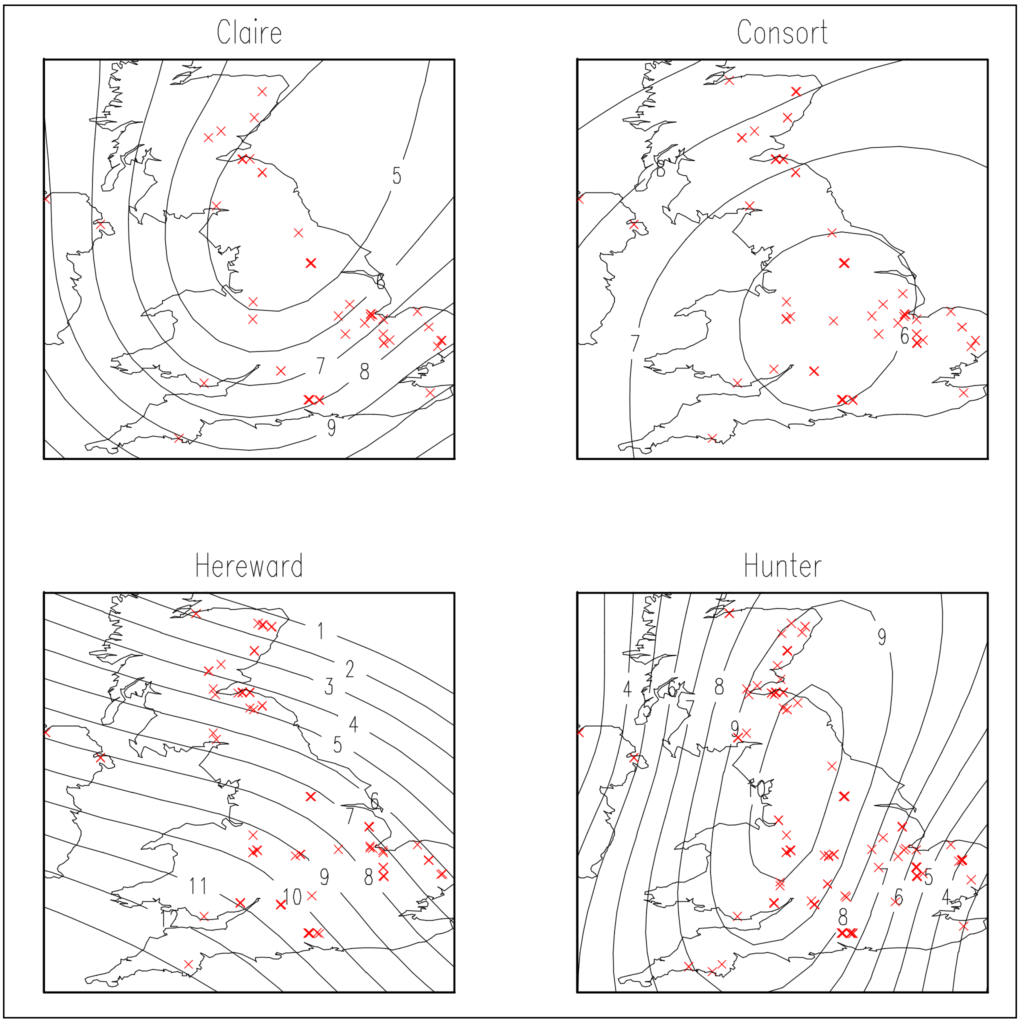


Figure 3: Predicted response in treated yield (t/ha) to large-scale trends in climate for varieties Madrigal, Malacca, Mercia and Rialto. Values shown are deviations from average yield for each variety, with expected site yield held constant at 10.2t/ha across the UK, assuming first wheat crop with sow date 7-15 October. All varieties show strong evidence of response to location. Trial sites where each variety was grown are indicated by ×. Contour lines for yield deviations: 1=-0.3 2=-0.25, 3=-0.2, 4=-0.15, 5=-0.1, 6=0, 7=0.1, 8=0.15, 9=0.2, 10=0.25 11=0.3t/ha.

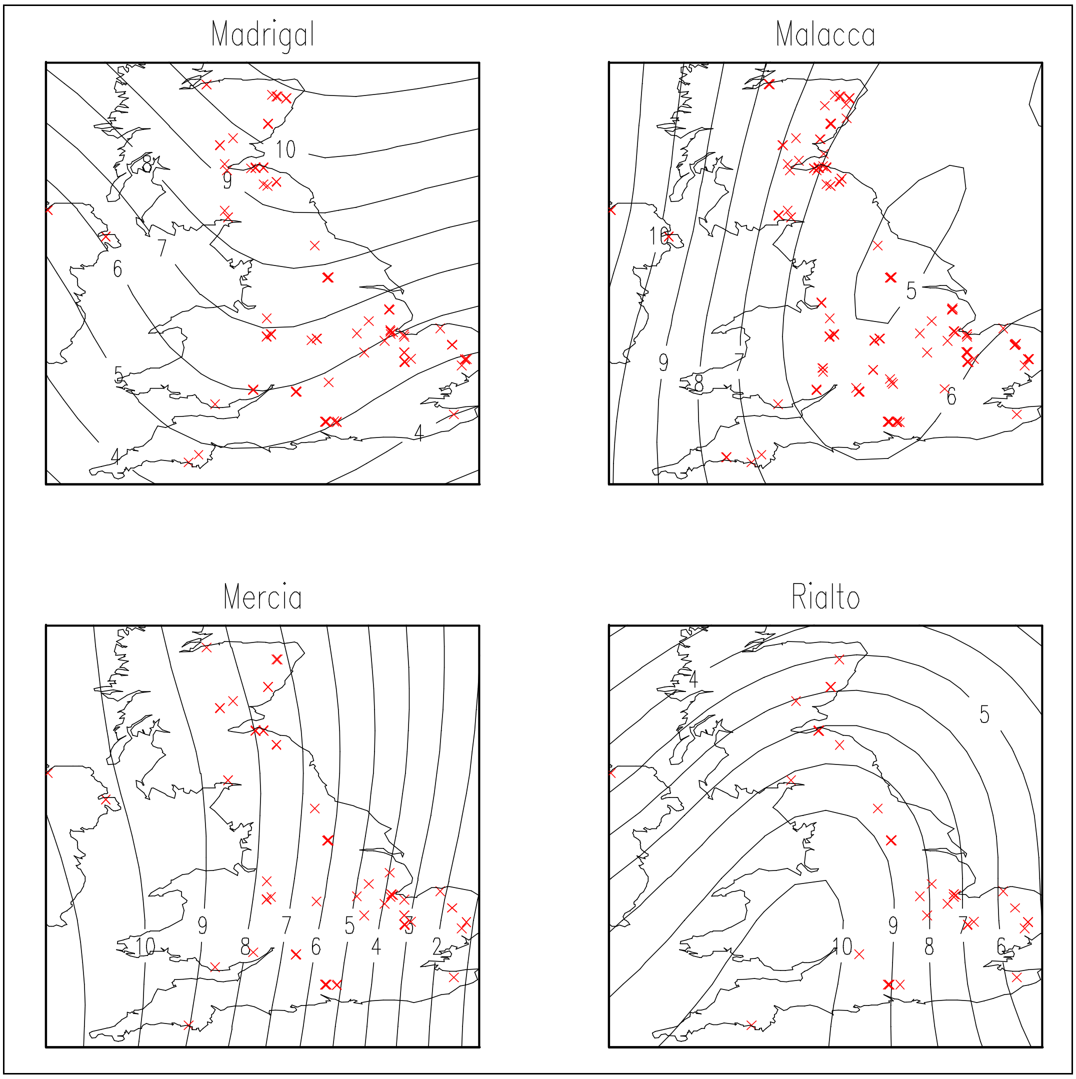
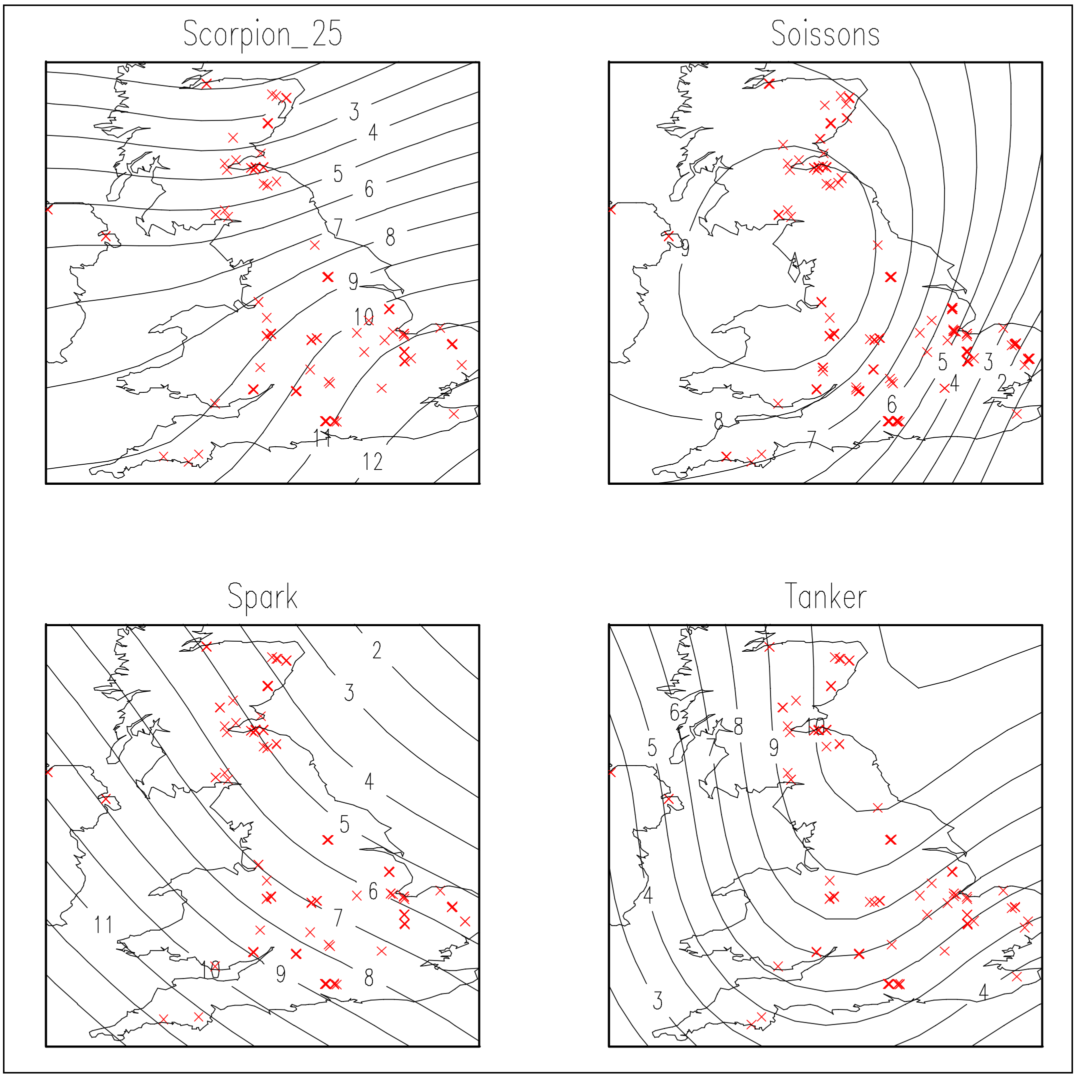


Figure 4: Predicted response in treated yield (t/ha) to large-scale trends in climate for varieties Scorpion 25, Soissons, Spark and Tanker. Values shown are deviations from average yield for each variety, with expected site yield held constant at 10.2t/ha across the UK, assuming first wheat crop with sow date 7-15 October. All varieties show strong evidence of response to location. Trial sites where each variety was grown are indicated by ×. Contour lines for yield deviations: 1=-0.3 2=-0.25, 3=-0.2, 4=-0.15, 5=-0.1, 6=0, 7=0.1, 8=0.15, 9=0.2, 10=0.25 11=0.3t/ha.



Appendix C. Variety effects (with prediction standard errors) in the final model for selected site variables

Table 10. Predicted effects (with prediction standard errors) in the final model for variety main effects and interactions with expected site yield, late sown crops (after 30 Oct), shallow or sandy soil or soils with low K index (<1.5). * indicates cases where predictor/(prediction standard error) > 2.

Variety	Variety main effects	Interaction effects for variety ×			
		siteyield	latesown	soilAS	lowK
Brigadier	0.091 (0.141)	0.033 (0.031)	-0.009 (0.069)	0.032 (0.061)	-0.074 (0.066)
Buster	-0.131 (0.141)	-0.007 (0.032)	0.043 (0.080)	-0.010 (0.062)	-0.035 (0.065)
Hereward	-0.807* (0.126)	-0.117* (0.027)	0.053 (0.060)	-0.107* (0.054)	0.141* (0.057)
Hunter	0.068 (0.147)	0.004 (0.032)	-0.035 (0.075)	0.046 (0.063)	-0.032 (0.070)
Hussar	-0.086 (0.136)	-0.038 (0.029)	0.015 (0.065)	-0.013 (0.058)	0.001 (0.063)
Mercia	-1.052* (0.150)	-0.231* (0.033)	0.010 (0.083)	-0.027 (0.064)	0.015 (0.071)
Rialto	0.084 (0.131)	0.039 (0.028)	-0.040 (0.060)	-0.112* (0.056)	0.093 (0.059)
Riband	-0.105 (0.130)	-0.019 (0.028)	-0.209* (0.062)	0.041 (0.056)	0.017 (0.059)
Spark	-0.923* (0.143)	-0.154* (0.032)	0.104 (0.068)	-0.052 (0.061)	0.047 (0.068)
Soissons	-0.635* (0.134)	-0.145* (0.029)	0.060 (0.065)	-0.027 (0.059)	0.009 (0.064)
Consort	-0.054 (0.129)	-0.018 (0.027)	-0.065 (0.069)	-0.006 (0.056)	0.022 (0.057)
Charger	-0.131 (0.136)	-0.058 (0.030)	-0.017 (0.067)	0.084 (0.059)	-0.108 (0.063)
Reaper	-0.030 (0.146)	0.009 (0.034)	0.083 (0.083)	0.054 (0.063)	0.057 (0.067)
Equinox	0.162 (0.139)	0.067* (0.031)	0.039 (0.074)	0.050 (0.061)	-0.065 (0.061)
Madrigal	0.200 (0.141)	0.008 (0.031)	0.008 (0.073)	0.035 (0.061)	0.000 (0.061)
Malacca	-0.456* (0.145)	-0.094* (0.032)	-0.049 (0.068)	-0.007 (0.063)	-0.016 (0.063)
Savannah	0.476* (0.140)	0.045 (0.030)	0.062 (0.069)	-0.011 (0.060)	-0.018 (0.060)
Buchan	0.030 (0.152)	-0.008 (0.034)	0.008 (0.074)	0.017 (0.065)	0.002 (0.065)
Claire	-0.133 (0.144)	-0.025 (0.032)	-0.022 (0.070)	-0.023 (0.063)	0.064 (0.063)
Shamrock	-0.428* (0.145)	-0.096* (0.032)	0.082 (0.069)	-0.087 (0.063)	0.013 (0.063)
Napier	0.318* (0.158)	0.045 (0.035)	-0.058 (0.072)	-0.016 (0.066)	0.015 (0.066)
Biscay	0.262 (0.159)	0.062 (0.035)	-0.011 (0.075)	-0.025 (0.069)	0.033 (0.070)
Deben	0.204 (0.159)	0.008 (0.035)	-0.037 (0.075)	0.153* (0.069)	0.101 (0.070)
Option	0.201 (0.159)	0.025 (0.035)	-0.049 (0.075)	0.004 (0.069)	0.064 (0.070)
Tanker	0.689* (0.159)	0.114* (0.035)	0.040 (0.075)	-0.026 (0.069)	-0.013 (0.070)
Access	0.353* (0.164)	0.065 (0.036)	-0.039 (0.080)	0.039 (0.073)	-0.031 (0.074)
Chatsworth	-0.221 (0.165)	-0.045 (0.037)	0.024 (0.083)	-0.019 (0.073)	0.012 (0.075)
Macro	0.206 (0.164)	0.037 (0.037)	-0.095 (0.083)	0.030 (0.073)	-0.089 (0.075)
Phlebas	0.165 (0.164)	0.066 (0.037)	-0.003 (0.083)	0.084 (0.073)	-0.001 (0.075)
Richmond	0.224 (0.164)	0.041 (0.037)	0.055 (0.083)	-0.023 (0.073)	0.016 (0.075)
Solstice	-0.117 (0.163)	-0.033 (0.036)	0.031 (0.078)	-0.087 (0.073)	0.088 (0.074)
Xi19	0.226 (0.163)	0.087* (0.036)	-0.039 (0.078)	0.065 (0.073)	-0.024 (0.074)
Brunel	0.069 (0.172)	0.020 (0.040)	-0.021 (0.084)	-0.001 (0.076)	-0.060 (0.075)
Carlton	-0.075 (0.173)	0.002 (0.040)	0.047 (0.085)	-0.028 (0.076)	-0.013 (0.076)
Chardonnay	0.030 (0.172)	-0.003 (0.040)	0.047 (0.084)	-0.026 (0.076)	-0.023 (0.075)
Einstein	0.281 (0.172)	0.043 (0.040)	0.029 (0.084)	0.015 (0.076)	0.042 (0.075)
Goodwood	0.144 (0.172)	-0.016 (0.040)	0.004 (0.084)	-0.048 (0.076)	-0.051 (0.075)
NSL WW39	0.013 (0.172)	0.002 (0.040)	0.048 (0.084)	-0.044 (0.076)	-0.066 (0.075)
Robigus	0.526* (0.172)	0.116* (0.040)	-0.046 (0.084)	0.033 (0.076)	-0.010 (0.075)
Scorpion 25	0.102 (0.172)	0.081* (0.040)	-0.032 (0.084)	0.026 (0.076)	0.013 (0.075)
Tellus	0.013 (0.172)	0.016 (0.040)	0.056 (0.084)	-0.062 (0.076)	-0.058 (0.075)
Warlock 24	0.172 (0.172)	0.065 (0.040)	-0.043 (0.084)	0.056 (0.076)	-0.008 (0.075)
Wizard	0.075 (0.173)	0.006 (0.040)	-0.028 (0.085)	0.027 (0.076)	-0.069 (0.076)