Colloquium Biometricum 40
2010, 63–71

# UNREPLICATED EXPERIMENTS IN EARLY STAGE BREEDING PROGRAMS

**Katarzyna Marczyńska[1], Stanisław Mejza[1], Wojciech Mikulski[2]**

[1]Department of Mathematical and Statistical Methods
Poznań University of Life Sciences
Wojska Polskiego 28, 60-637 Poznań, Poland
e-mails: kasiapaw@au.poznan.pl; smejza@au.poznan.pl
[2]Pland Breeding Station Szelejewo
Szelejewo Drugie 1, 63-820 Piaski

## Summary

In plant breeding trials, during the early stages of the improvement process, it is not possible to use an experimental design that satisfies the requirement of replicating all the treatments because of the large number of genotypes involved, the small amount of seed and the low availability of resources. Hence, the unreplicated designs are used for early generation testing when hundreds or even thousands new genotypes need evaluation in the same trial using a limited amount of seed that is enough for one replicate only. To control the real or potential heterogeneity of experimental units, control (check) plots are arranged in the trial.

There are many methods of using information resulting from check plots. In the paper the main tool of exploring this information will be based on a response surface methodology (RSM). At the beginning we will try to identify response surface characterizing experimental environments. The obtained response surface will be then used to adjust the observations for genotypes. Finally, so adjusted data will be used for inference concerning the next steps of breeding program. The theoretical considerations will be illustrated with the example dealing with spring barley.

**Key words and phrases:** breeding program, experimental design, unreplicated experiments, check plots , response surface

**Classification AMS 2010:** 62K99, 62F99,62P10

## 1. Introduction

The paper deals with the selection problems in early stage of a breeding program. We restrict our attention to the breeding program in which the selection decision of lines and strains of plants for further breeding is taken on the basis of the performed experiment.

Two problems decide on the design of experiment, i.e. very large number of lines and limited amount of seeds. Because of those two facts the experiment are usually created in one replication of one of several objects using some standard or standards repeated several times. It means that final decision concerning selection of about 20% to 30% of the hybrids for further breeding program is based on unreplicated field experiment. Hence, the procedure of selection at this stage should be extremely cautious and supported by proper statistical techniques.

## 2. Overview of methods of estimation of genotype effects

In this section we give an short overview of the some selection methods usually used for inference (selection) in unreplicated breeding trials. The methods for unreplicated breeding trials are proposed by Cullis and Gleesson (1989) and Kempton and Fox (1997). The variability of units with their geometrical structure in the experiment is used by the proposed methods to adjust the average values of observed characteristics on hybrids.

The important methods of selecting genotypes from unreplicated experiments are proposed by Kempton and Fox (1997). Usually in the field experiment besides of hybrids the standard (control crop) is applied. The estimate of the line effect is calculated as the deviation of the observation for this line from the (weighted) mean value of the two standards the closest to the estimated line.

Linear variance model approach proposed by Williams, (1986) not only eliminates the effect of blocks, but the linear trend of soil variation within a block.

Many statistical techniques utilize check plots in unreplicated experiments (cf. Utz, 1997). In this method some check plots occurring around the plots with genotype (line) are used to adjust the estimate of that line effect.

For more details connected with inference from unreplicated experiments the reader is referred to, for example, Ambroży *et al.* 2008a, 2008b, Bakinowska *et al.* 2009.

All of the usually applied adjusting methods for unreplicated experiments are proper to some specific structure of soil fertility (cf. Dobek and Kala, 1995, Kristensen and Ersboll, 1992, 1995). The disadvantage of them is the fact that before and also after performing the experiment we do not know which kind of soil structure occurs in our experiment. Hence, we can not say which of existing methods is proper to given experimental situation. The method of inference proposed below is free of disadvantage mentioned above. It is always proper because the trend of soil is identified estimated and later used for adjustments.

## 3. Response surface methodology

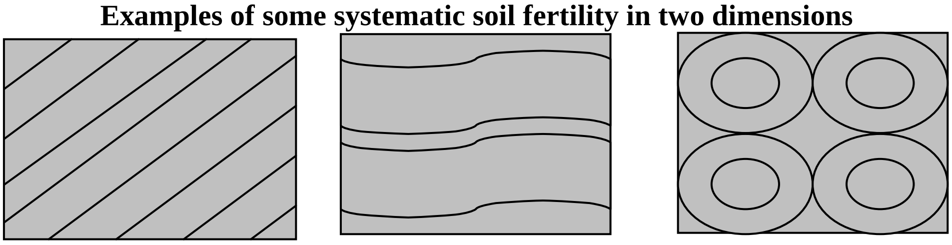
In this approach the main role is played by the density and geometrical structure of check plot treatment (standards) in the experiment. In the paper the main tool of exploring information resulting from check plots will be based on a response surface methodology (RSM).

Response surface methodology (RSM) is a set of techniques that comprises:

1. Setting up a series of experiments (designing a set of experiments) that will yield adequate and reliable measurements of the response on interest,
2. Determining a mathematical model that best fits the data collected from the design chosen in (1), by conducting appropriate tests of hypotheses concerning the model's parameters,
3. Determining the optimal setting of the experimental factors that produce the maximum (or minimum) value of the response,
4. Using estimated response surface to forecast the observation in a given places of the experimental field.

If discovering the best value, or values, of the response is beyond the available resources of the experiment, then response surface methods are aimed at obtaining at least a better understanding of the overall system. In any system in which variable quantities change, the interest might be in assessing the effects of the factors on the behavior of some measurable quantity (the response). Such an assessment is possible through regression analysis. Using data collected from a set of experimental trials, regression helps to establish empirically the type of relationship that is present between the response variable and its influencing factors. There are many types of the structure of the soil fertility (environment). Some of them are illustrated below.

**Examples of some systematic soil fertility in two dimensions**



Simple systematic design — Semi – balanced systematic design — Completely balanced systematic design

## 4. Research material and methods

Material for study were the results of grain yield obtained in unreplicated breeding trials with standards, including 410 lines of spring barley performed at “Modzurów” plant breeding station. The observed trait in the experiment was the grain yield in kg per plot. Area of one plot was $10m^2$ (Tab.1). Experiment was conducted in 2005/2006 season. In the statistical analysis the yield of 333 experimental plots (Fig.2) was taken into account. The measured (observed) area of the plots was 1m x 10m i.e. $10m^2$ (Fig. 1).

**Table 1.** Basic parameters of the experiment with spring barley

| | |
|---|---|
| Species | Spring barley |
| Number of studied lines | 410 |
| Number of standard objects | 1 |
| Number of tested plots between standards | 5 |
| Numbers of rows | 37 |
| Number of columns | 9 |
| Plot size | 10 $m^2$ |

**Table 2.** Regression models for standards objects

| Estimate of the regression model $\mathbf{z = a + b_1x + b_2x^2 + b_3x^3 + c_1y + c_2y^2 + c_3y^3 + dxy}$ | Equation coefficients |
|---|---|
| **a** | 4.573666 |
| $\mathbf{b_1}$ | 0.160087 |
| $\mathbf{b_2}$ | -0,006222 |
| $\mathbf{b_3}$ | 0.000046 |
| $\mathbf{c_1}$ | 0.226874 |
| $\mathbf{c_2}$ | -0.015809 |
| $\mathbf{c_3}$ | 0.000277 |
| **d** | 0.001037 |
| $\mathbf{R^2}$ | 0.73715636 |

The statistical analysis was performed at two stages. At the beginning we estimate the two dimensional surface equation for standards (marked bold) to characterize the variability of the soil (Tab.2) in the experimental field. The estimate of response surface of the yield (significant at alpha = 0.05) is in the Table 2. The coefficient of determination is equal to 74%. The shape of the response surface characterizing the soil fertility of experimental field is presented in the Figure 3.

In the next step we calculate the forecast for all hybrids. The coordinates of the hybrids were the central point of the experimental field. It means that the distances in one direction were: ( y ) 0.5; 1.5; 2.5; … 36.5[m] while second one were: ( x ) 5, 10, 15 … 85[m].

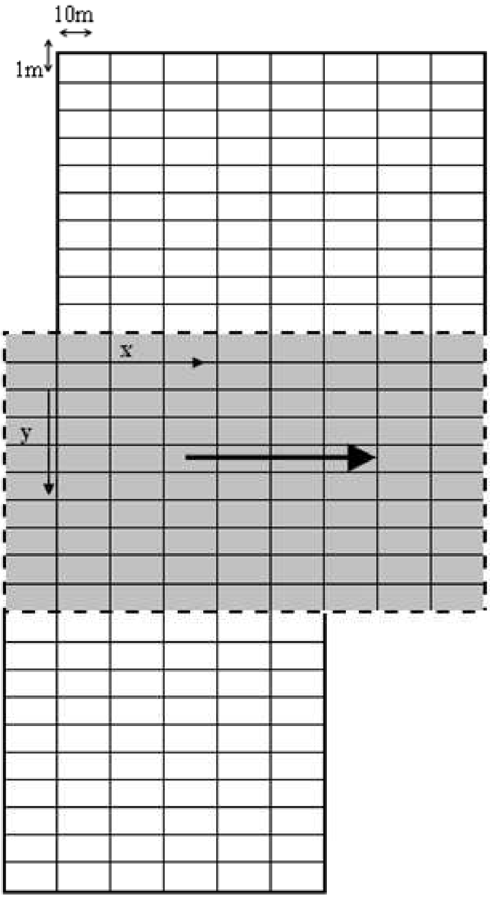



Fig. 1. The general scheme of experimental field

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **4813** | **4948** | **5069** | **5204** | **5325** | **5460** | **5581** | **5654** | **5775** |
| 4814 | 4947 | 5070 | 5203 | 5326 | 5459 | 5582 | 5653 | 5776 |
| 4815 | 4946 | 5071 | 5202 | 5327 | 5458 | 5583 | 5652 | 5777 |
| 4816 | 4945 | 5072 | 5201 | 5328 | 5457 | 5584 | 5651 | 5778 |
| 4817 | 4944 | 5073 | 5200 | 5329 | 5456 | 5585 | 5650 | 5779 |
| 4818 | 4943 | 5074 | 5199 | 5330 | 5455 | 5586 | 5649 | 5780 |
| **4819** | **4942** | **5075** | **5198** | **5331** | **5454** | **5587** | **5648** | **5781** |
| 4820 | 4941 | 5076 | 5197 | 5332 | 5453 | 5588 | 5647 | 5782 |
| 4821 | 4940 | 5077 | 5196 | 5333 | 5452 | 5589 | 5646 | 5783 |
| 4822 | 4939 | 5078 | 5195 | 5334 | 5451 | 5590 | 5645 | 5784 |
| 4823 | 4938 | 5079 | 5194 | 5335 | 5450 | 5591 | 5644 | 5785 |
| 4824 | 4937 | 5080 | 5193 | 5336 | 5449 | 5592 | 5643 | 5786 |
| **4825** | **4936** | **5081** | **5192** | **5337** | **5448** | **5593** | **5642** | **5787** |
| 4826 | 4935 | 5082 | 5191 | 5338 | 5447 | 5594 | 5641 | 5788 |
| 4827 | 4934 | 5083 | 5190 | 5339 | 5446 | 5595 | 5640 | 5789 |
| 4828 | 4933 | 5084 | 5189 | 5340 | 5445 | 5596 | 5639 | 5790 |
| 4829 | 4932 | 5085 | 5188 | 5341 | 5444 | 5597 | 5638 | 5791 |
| 4830 | 4931 | 5086 | 5187 | 5342 | 5443 | 5598 | 5637 | 5792 |
| **4831** | **4930** | **5087** | **5186** | **5343** | **5442** | **5599** | **5636** | **5793** |
| 4832 | 4929 | 5088 | 5185 | 5344 | 5441 | 5600 | 5635 | 5794 |
| 4833 | 4928 | 5089 | 5184 | 5345 | 5440 | 5601 | 5634 | 5795 |
| 4834 | 4927 | 5090 | 5183 | 5346 | 5439 | 5602 | 5633 | 5796 |
| 4835 | 4926 | 5091 | 5182 | 5347 | 5438 | 5603 | 5632 | 5797 |
| 4836 | 4925 | 5092 | 5181 | 5348 | 5437 | 5604 | 5631 | 5798 |
| **4837** | **4924** | **5093** | **5180** | **5349** | **5436** | **5605** | **5630** | **5799** |
| 4838 | 4923 | 5094 | 5179 | 5350 | 5435 | 5606 | 5629 | 5800 |
| 4839 | 4922 | 5095 | 5178 | 5351 | 5434 | 5607 | 5628 | 5801 |
| 4840 | 4921 | 5096 | 5177 | 5352 | 5433 | 5608 | 5627 | 5802 |
| 4841 | 4920 | 5097 | 5176 | 5353 | 5432 | 5609 | 5626 | 5803 |
| 4842 | 4919 | 5098 | 5175 | 5354 | 5431 | 5610 | 5625 | 5804 |
| **4843** | **4918** | **5099** | **5174** | **5355** | **5430** | **5611** | **5624** | **5805** |
| 4844 | 4917 | 5100 | 5173 | 5356 | 5429 | 5612 | 5623 | 5806 |
| 4845 | 4916 | 5101 | 5172 | 5357 | 5428 | 5613 | 5622 | 5807 |
| 4846 | 4915 | 5102 | 5171 | 5358 | 5427 | 5614 | 5621 | 5808 |
| 4847 | 4914 | 5103 | 5170 | 5359 | 5426 | 5615 | 5620 | 5809 |
| 4848 | 4913 | 5104 | 5169 | 5360 | 5425 | 5616 | 5619 | 5810 |
| **4849** | **4912** | **5105** | **5168** | **5361** | **5424** | **5617** | **5618** | **5811** |

Legend:
**4813** - Number of standard objects
4814 - Number of tested plots

Fig. 2. Location of the field plots

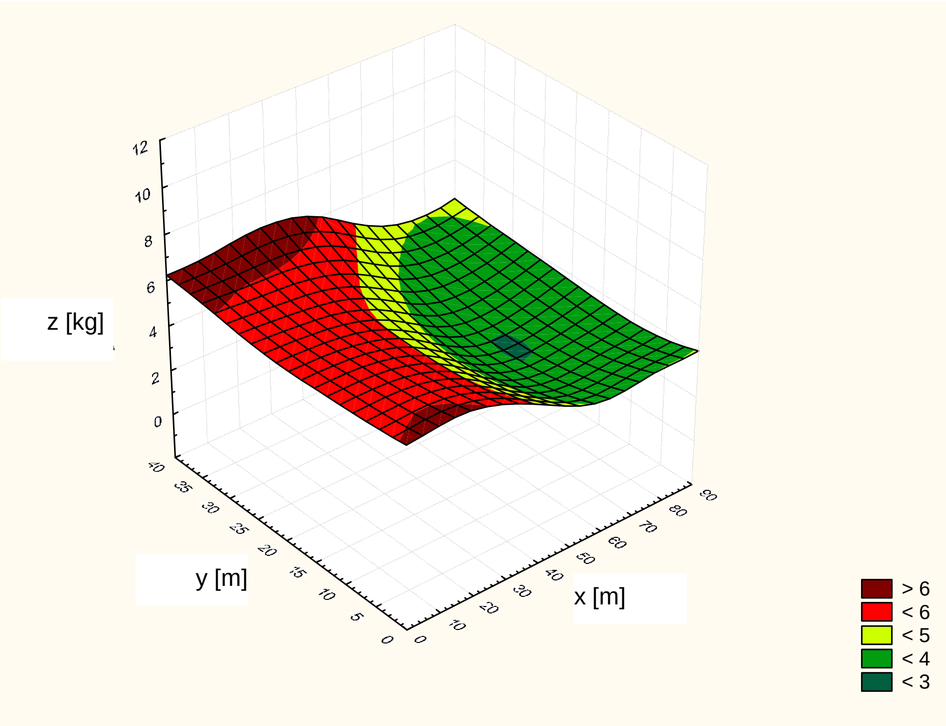



**Fig. 3.** The response surface characterizing the soil fertility of experimental field

The forecast obtained estimate the yield resulting directly from the soil fertility. We assume that the observed yield results directly from two components; one due to the soil fertility and second due to the hybrid effect. Then, the difference between observed yield and forecast can be treated as the estimate of genotype (hybrid) effect. And these differences will play the main role in the selection procedure. The ranking of genotypes with respect to above estimates is presented in Table 3. The statistical selection is based on Tukey's test for comparing all pairs.

The critical value for Tukey's test is equal to 5.1 (at significant level 0.05). All calculations were performer by statistical package STATISTICA. It means that two hybrids we recognize as different when the difference between their estimates is larger than 5.1.

**Table 3.** The ranking of genotypes

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **4844** | -4.194 | **4946** | -0.548 | **5612** | -0.105 | **5190** | 0.254 | **5608** | 0.662 | **5440** | 1.003 |
| **5653** | -2.85 | **5589** | -0.547 | **5457** | -0.088 | **5801** | 0.268 | **5194** | 0.668 | **5445** | 1.034 |
| **5652** | -2.635 | **5645** | -0.537 | **5354** | -0.077 | **4926** | 0.287 | **5451** | 0.668 | **4931** | 1.056 |
| **5356** | -2.572 | **5179** | -0.519 | **5796** | -0.075 | **4834** | 0.289 | **5100** | 0.67 | **5450** | 1.069 |
| **5651** | -2.448 | **4846** | -0.489 | **4832** | -0.068 | **5637** | 0.302 | **5347** | 0.676 | **4933** | 1.071 |
| **5650** | -2.076 | **5628** | -0.486 | **5073** | -0.041 | **5426** | 0.323 | **5084** | 0.678 | **5176** | 1.099 |
| **5778** | -1.822 | **5200** | -0.484 | **5798** | -0.036 | **5594** | 0.327 | **5351** | 0.682 | **5089** | 1.111 |
| **5776** | -1.805 | **5585** | -0.477 | **5427** | -0.007 | **5789** | 0.335 | **5804** | 0.683 | **4938** | 1.113 |
| **4845** | -1.803 | **4814** | -0.472 | **4824** | -0.006 | **5102** | 0.337 | **5441** | 0.699 | **4941** | 1.116 |
| **5582** | -1.799 | **5622** | -0.471 | **5429** | -0.005 | **5358** | 0.346 | **5078** | 0.708 | **5443** | 1.132 |
| **5326** | -1.505 | **5786** | -0.47 | **5788** | 0.01 | **5091** | 0.346 | **5332** | 0.709 | **5074** | 1.134 |
| **5649** | -1.489 | **4841** | -0.451 | **5076** | 0.029 | **4945** | 0.347 | **5619** | 0.722 | **5444** | 1.165 |
| **5583** | -1.485 | **5780** | -0.443 | **5792** | 0.033 | **5095** | 0.378 | **5103** | 0.725 | **5341** | 1.183 |
| **5584** | -1.368 | **4847** | -0.422 | **5344** | 0.036 | **4927** | 0.388 | **5616** | 0.73 | **5182** | 1.19 |
| **5172** | -1.359 | **5783** | -0.408 | **5625** | 0.038 | **5345** | 0.42 | **5171** | 0.745 | **5435** | 1.233 |
| **5459** | -1.329 | **5592** | -0.401 | **5614** | 0.055 | **5610** | 0.439 | **5187** | 0.746 | **5434** | 1.247 |
| **5357** | -1.31 | **5627** | -0.389 | **5600** | 0.055 | **5330** | 0.452 | **5085** | 0.747 | **4919** | 1.255 |
| **5644** | -1.286 | **5794** | -0.368 | **5621** | 0.056 | **5452** | 0.471 | **5346** | 0.756 | **5339** | 1.257 |
| **4838** | -1.285 | **4915** | -0.346 | **5640** | 0.056 | **5425** | 0.478 | **5596** | 0.759 | **5177** | 1.269 |
| **5779** | -1.17 | **5094** | -0.345 | **4827** | 0.07 | **5352** | 0.497 | **5438** | 0.77 | **5079** | 1.278 |
| **5101** | -1.158 | **5072** | -0.341 | **5348** | 0.102 | **5803** | 0.506 | **5620** | 0.786 | **5188** | 1.29 |
| **5646** | -1.155 | **5634** | -0.334 | **4823** | 0.109 | **5169** | 0.509 | **5597** | 0.79 | **5088** | 1.318 |
| **5777** | -1.13 | **5639** | -0.317 | **5791** | 0.115 | **5184** | 0.515 | **4822** | 0.82 | **5098** | 1.357 |
| **4916** | -1.07 | **5590** | -0.309 | **5797** | 0.117 | **5338** | 0.523 | **5595** | 0.842 | **5439** | 1.359 |
| **5638** | -1.015 | **5633** | -0.307 | **4815** | 0.13 | **4937** | 0.529 | **5096** | 0.843 | **5808** | 1.382 |
| **5202** | -0.97 | **5191** | -0.3 | **5170** | 0.134 | **5071** | 0.533 | **5080** | 0.844 | **5433** | 1.413 |
| **5173** | -0.962 | **5603** | -0.293 | **5455** | 0.14 | **5432** | 0.543 | **5329** | 0.846 | **5077** | 1.423 |
| **5183** | -0.96 | **5428** | -0.272 | **5199** | 0.142 | **4842** | 0.544 | **5195** | 0.848 | **5810** | 1.429 |
| **5784** | -0.94 | **5598** | -0.263 | **4816** | 0.145 | **5615** | 0.545 | **5335** | 0.85 | **4840** | 1.451 |
| **5083** | -0.938 | **4828** | -0.235 | **4821** | 0.155 | **4830** | 0.548 | **5453** | 0.857 | **5809** | 1.513 |
| **5795** | -0.933 | **5360** | -0.23 | **5328** | 0.155 | **5606** | 0.561 | **5446** | 0.857 | **4929** | 1.52 |
| **4947** | -0.921 | **5607** | -0.224 | **5800** | 0.173 | **5609** | 0.563 | **4914** | 0.862 | **5447** | 1.553 |
| **5327** | -0.901 | **5629** | -0.221 | **4829** | 0.174 | **5193** | 0.564 | **4917** | 0.868 | **5086** | 1.563 |
| **5591** | -0.888 | **4922** | -0.212 | **4817** | 0.174 | **5802** | 0.565 | **4943** | 0.871 | **5175** | 1.614 |
| **5647** | -0.86 | **5643** | -0.208 | **5333** | 0.183 | **5353** | 0.587 | **4934** | 0.876 | **4944** | 1.626 |
| **5586** | -0.809 | **5602** | -0.204 | **5782** | 0.187 | **4932** | 0.591 | **5197** | 0.878 | **5092** | 1.631 |
| **5631** | -0.789 | **5350** | -0.182 | **4833** | 0.194 | **4836** | 0.602 | **5342** | 0.88 | **5097** | 1.652 |
| **5604** | -0.776 | **5613** | -0.161 | **5359** | 0.195 | **5456** | 0.603 | **4818** | 0.908 | **4940** | 1.669 |
| **5588** | -0.691 | **5201** | -0.144 | **5807** | 0.196 | **5203** | 0.607 | **5340** | 0.953 | **5082** | 1.769 |
| **4839** | -0.633 | **5189** | -0.14 | **4935** | 0.203 | **5090** | 0.616 | **5449** | 0.956 | **4923** | 1.995 |
| **5641** | -0.619 | **5178** | -0.136 | **5181** | 0.206 | **4826** | 0.617 | **5334** | 0.969 | **4913** | 2.236 |
| **5632** | -0.596 | **5437** | -0.124 | **5601** | 0.219 | **4939** | 0.624 | **5196** | 0.972 | **4848** | 2.282 |
| **5070** | -0.59 | **5635** | -0.119 | **5790** | 0.223 | **5623** | 0.636 | **4928** | 0.973 | **4921** | 2.492 |
| **5806** | -0.579 | **4925** | -0.118 | **4835** | 0.238 | **5185** | 0.641 | **5431** | 0.979 | **5785** | 3.362 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **5458** | -0.555 | **4820** | -0.118 | **5626** | 0.252 | **4920** | 0.651 | **5104** | 0.98 | **5336** | 4.846 |

## 5. Conclusions

The observed yield results directly from two components; one due to soil fertility and second due to hybrids effect. The difference between observed yield and forecast can be treated as the estimate of genotype effect The calculations showed that the response surface methodology seemed thus to be a good approach for predicting the yield of plots with the line in such unreplicated trials.

## Acknowledgments

The paper was supported by the grant from the Polish Ministry of Science and Higher Education No. NN310 185337.

## References

Ambroży K., Bakinowska E., Bocianowski J., Budka A., Pilarczyk W., Zawieja B. (2008). Statystyczne wspomaganie decyzji selekcyjnych na wczesnych etapach hodowli zbóż. I. Metody oceny efektów obiektowych. *Biuletyn IHAR* 250, 21–28.

Ambroży K., Bakinowska E., Bocianowski J., Budka A., Pilarczyk W., Zawieja B.(2008). Statystyczne wspomaganie decyzji selekcyjnych na wczesnych etapach hodowli zbóż. II. Empiryczne porównanie metod oceny efektów obiektowych. *Biuletyn IHAR,* 250, 29–39.

Bakinowska E., Bocianowski J., Budka A., Pilarczyk W., Zawieja B., Ambroży K. (2009). Estymacja wariancji błędu w hodowlanych doświadczeniach jednopowtórzeniowych z replikowanymi obiektami wzorcowymi. *Biuletyn Instytutu Hodowli i Aklimatyzacji Roślin,* Nr 251, 5–14.

Cullis B.R., Gleeson A.C. (1989). The efficiency of neighbour analysis for replicated variety trials in Australia. *Journal of Agricultural Science,* 113, 223–239.

Dobek A., Kala R. (1995). On the analysis of experiments with unreplicated varieties. *Biuletyn Oceny Odmian.* Zeszyt 26–27, 73–82.

Elandt-Johnson R.C. (1966). Multi-dimensional orthogonal polynomials for cartain models in multivariate analysis. *The Indian Journal of Statistics*, Series B, Vol.28, Parts 3&4.

Kempton R.A., Fox P.N. 1997. *Statistical methods for plant variety evaluation*. Chapman & Hall.

Khuri A.I., Cornell J.A. (1987). Response Surfaces. Designs and Analyses. *Library of Congress Cataloging in Publication Data*. New York, Milwaukee.

Kristensen K., Ersboll A.K. (1992). The use of geostatistical methods in planning variet trials. *Biuletyn Oceny Odmian*. Zeszyt 24-25, 139–157.

Kristensen K., Ersboll A.K. (1995). The use of geostatistical methods in variety trials, where some varieties are unreplicated. *Biuletyn Oceny Odmian.* Zeszyt 26–27, 113–122.

Utz H. F. (1997). PLABSTAT. A computer program for statistical analysis of plant breeding experiments. Version 2N. Institute *of Plant Breeding, Seed Science and Population Genetics,* University of Hohenheim, Germany.

# DOŚWIADCZENIA JEDNOPOWTÓRZENIOWE WE WCZESNYCH STADIACH HODOWLANYCH

## Streszczenie

Istnieje wiele metod oceny efektów obiektowych w doświadczeniach jednopowtórzeniowych z wzorcami. W pracy głównym narzędziem oceny tych efektów będzie metoda powierzchni reakcji (RSM). Na początku postaramy się zidentyfikować i oszacować powierzchnię reakcji charakteryzując środowisko doświadczenia na podstawie poletek z obiektem wzorcowym. Uzyskana powierzchnia reakcji wykorzystana zostanie do estymacji efektów występujących w doświadczeniu genotypów. Jako ocenę efektu genotypu przyjmujemy w pracy różnice pomiędzy obserwacją z poletka a prognozą uzyskaną z oszacowanej powierzchni reakcji, charakteryzującą zmienność glebową. W końcu, tak oszacowane dane zostaną wykorzystane do selekcji genotypów do następnego programu hodowlanego. Rozważania teoretyczne są ilustrowane przykładem doświadczenia nad jęczmieniem jarym.

**Słowa kluczowe:** Doświadczenia hodowlane, doświadczenia jednopowtórzeniowe, poletka kontrolne, powierzchnia reakcji

**Klasyfikacja AMS 2010:** 62K99, 62F99, 62P10