UDC 631.52:001.8:58.08
CHERNUSKYI V.V., CSc
Institute for Agriculture of Polissia NAAS
e-mail: isgpkor@rambler.ru

METHODOLOGICAL APPROACH TO CREATING COMPUTER-AIDED SYSTEMS FOR DATA COLLECTION AND ANALYSIS IN SELECTION PROCESS IN BREEDING AGRICULTURAL CROPS

Obtaining the maximum data array is one of the most important methodological objectives in an experiment. Found was a need to establish and analyze not only additive, but multiplicative variation in an experimental system. In particular, having recorded the full range of phenotypic variance in the gradient vector field of ontogenetic development component and integrated macrosign, we significantly increase the accuracy of genotype identification and therefore the efficiency of selection. Owing to the nonlinear synergetic selection principle introduced into breeding practice, created were crops cultivars adapted to the local environment, such as blue lupine, field pea, and winter rye.
Keywords: *additive; multiplicative variance; selection; macrosign synergy; cultivars*

Introduction. There is a difference of principle in the experiment scheme and data analysis within linear and nonlinear system. In particular, according to B.A. Dospiehov, within a univariate complex system of variance there are following components:

total variation $C_Y = \sum X_i$,

variation options $C_V = \frac{\sum(\sum V}{n}$ } - additive component

varying repetitions $C_P = \frac{\sum(\sum P}{l}$ } - multiplicative component

random variation $C_Z = C_Y - (C_V$ }- noise component

where C -is a corrective factor.

Most of the traditional systems of experiment statement primarily use the additive part as information-valuable platform, with multiplicative and noise components, which are also significant for the parametric information content, being neglected by experimenters. A similar principle of selection of valuable source material exists in breeding also.

P.F. Rokytskyi and D. Folkoner offered a classic formula for phenotypic and genotypic variability:

$$\sigma_P^2 = \sigma_G^2 + \sigma_E^2 + 2r_{GE}\sigma_G\sigma_E$$

where phenotypic variance is represented as the sum of genotypic and paratypic variability and "genotype-environment" interaction.

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2 + \sigma_{AAA}^2 + \sigma_{AAD}^2 + \cdots$$

Genotypic variability in turn consists of additive, dominant and recessive epigenetic interactions.

Of the greatest value for breeders is additive component of genotypic variation, which is fully transmitted to progeny being the least distorted genotypic information. In addition, according to current epigenetic theory of microevolution, which comprises the principles of reproduction and ontogenesis in plants, of great importance are the multiplicative parts of a genotype [1]. In particular, it was established for winter rye that productivity sighs are multiplicative. This is the reason of multiple values of the same signs in the population, as well as ambiguity and variability of phenotypic correlation coefficients between the individual elements of productivity and between them and the yield. The processes associated with the implementation of the population productivity are largely dependent not only on the composition of the genotype, but also on the environment, that is they are always mediated and dynamic [2].

A. V. Smyriaiev [3] established that, when improving biometric optimization sampling method for comparison experimental hybrid populations, a parameter D (genetic diversity of population) should be used not to estimate D itself, but its transformed value $Y = ln(D)$, which align D error variance. It was established that transformation $Y = l (D)$ is sensitive to differences in excess areas and asymmetries of initial sign distribution.

Therefore, one of the major methodological breeding problems remains a problem of identification of genotype for its phenotype and the associated complexity of selection pedigree plants. As opposed to pure controlled genetic experiments, a breeder has to do also with the so-called effect of "genotype environment" which significantly distorts the genotype index estimation in the total phenotype system. In particular, based on a study on genotype-environmental effects contribution to the formation of quantitative traits both in inbred and outbred plants [4], it was found that "genotype–environment" interaction is epigenetic by its nature, with this interaction being recorded as epistatic effect while assessing genetic variant in uneven progeny.

In this regard, there is a need to introduce an additional environmental (specifying) index to existing parameters of the phenotype development, i.e. phenotype equals to "genotype + environmental effects", as well as "genotype ×environment".

To study this interaction effect it is necessary to accumulate a large data array specifically about the impact of biogenic, abiogenic and, in particular, anthropogenic factors including by years of research. Field registries for experimental results including in terms of score visual evaluation are useful to this purpose to some extent, but the amount of data obtained in this way cannot be significant, and scoring in this case is the summation of visual images of virtual parametric scales formed in the researcher's mind, it is but subjective. This is why a significant amount of data is left out of information field of an experiment.

Besides, when carrying out experimental studies, breeders have to deal with large amount of information, which typically undergoes averaging, compression and archiving, i.e. obtaining mean values. However, during these operations, a quality-analytical component of the received information is being lost. The modern analytical basis of computer information processing allows operating directly with the entire (not shortened) numerical array, with processing actually taking a short period of time due to the high processing speed of the information systems. To obtain the same information in analogue form and translate it into a numeric array takes much more time. Experts from the Ukrainian Institute for Examination of Plant Varieties did the relevant work on unification and standardization of varieties results. In particular, they developed an optimized passport system for entering and storing the descriptions of varieties under examination.

To contribute these ideas we have worked out a hyper-matrix for entry and storage of digital data i.e. multi-level complex matrices with vectors – gradients and elementary matrices being the elements. With this form of multi-level hierarchical branched way of storing information it is possible to direct its processing and obtain functional relations and interactions of many experimental factors in form of multiply connected regions, states models with optimized parameter management, variation calculation for determining extrema of functions and other methods of searching for the optimal parameters of the component features in the formation of complex productivity characteristics of a variety.

Information models become more common in plant breeding. In particular, offered are the methodological aspects of forming crops database, classification of their characteristic, principles of creating a descriptive model of a variety as well as provided software and hardware environment of the information product [5] (Fig. 1). To analyze the relationship within "genotype-phenotype-environment" system in wheat proposed was a computer system WHEATPGE [6]. Also, proposed were ways of optimizing selection process for hybrid progeny in apple within frames of modern breeding programs using computer neural network programs for analysis [7].

Therefore, the main goal of our research was to develop a methodology and principles of selection experiment based on entire phase-parameter space of complex hyper matrices including the effects of temporal relations in "phenotype-genotype-environment" to identify areas of optimized vector-gradient directions of selection.

Materials and methods. The object was a process of formation and evolution of the parametric system for component and integrated agronomic traits in traditional crops of Polissia, such as yellow and blue lupine, field pea, winter rye, potato, in nurseries at different stages of selection process when selecting on temporal fields (2003 – 2013) under varying intensity influence of environmental factors. According to the “scheme of structural parametric pool distribution on time series from different years of selection”, a synthesis on a single surface of all series variation in phenotypes over years (environment is being taken into account), over signs and generations is carried out. The surface analysis is conducted according to the established pattern of an analytic geometry method, i.e. vector-gradient-spline on curved polynomial surfaces. The main advantage of this method is the ability to predict behaviour of phase-parametric systems. This is due to the possibility of time series analysis of the system function.

Research results. At present stage of computer technologies, the capabilities of multifactorial n-dimensional nonlinear analysis have been improved greatly. In particular, we have used the principles of optimizing relationships of component signs in the formation of complex sign maximum by simulating biological oscillators (Fig. 2) in the real parametric fields of signs in the samples dated of different years of research under the action of different vector-gradient environmental factors. We obtained trajectories of optimal control (selection) of component signs in the formation of complex one by way of example of signs of field pea samples “number of seeds in the bean” – “weight of seeds in a bean.”

A founder of gene geography, A.S. Serebrovskyi proposed a vector system for detecting graphic differences between chicken populations on the ground of quantitative signs. We contribute to his theory having proposed a gradient vector system for assessing parametric characteristics of quantitative traits in plants on the basis of modern computer technologies. The main point of the method is that the quantitative parameters of signs are reduced to a single integral average estimation, regardless of the absolute figures order, and within the macro command system of “Statistics” software possible is temporal (by year of experiment) scrolling traits vectors for gradient principle.

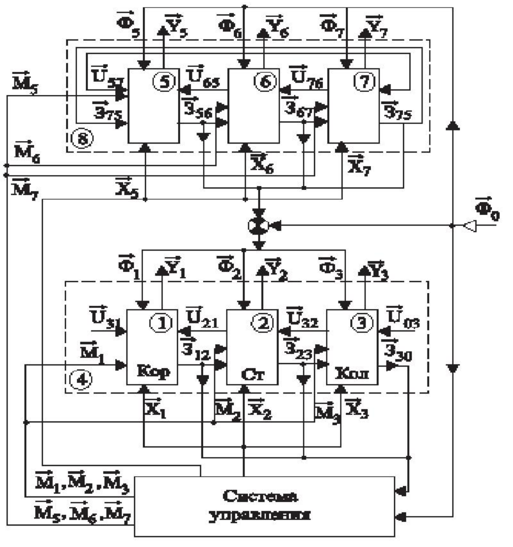


Fig. 1. Information model of cereals development.
Flows of information: of the external factors; of the objective function; of the mutual influence of the model components; of state of the model components; 1- root system; 2- stem; 3- ear; 4- plant; 5- disease; 6 - weeds; 7 - pests.

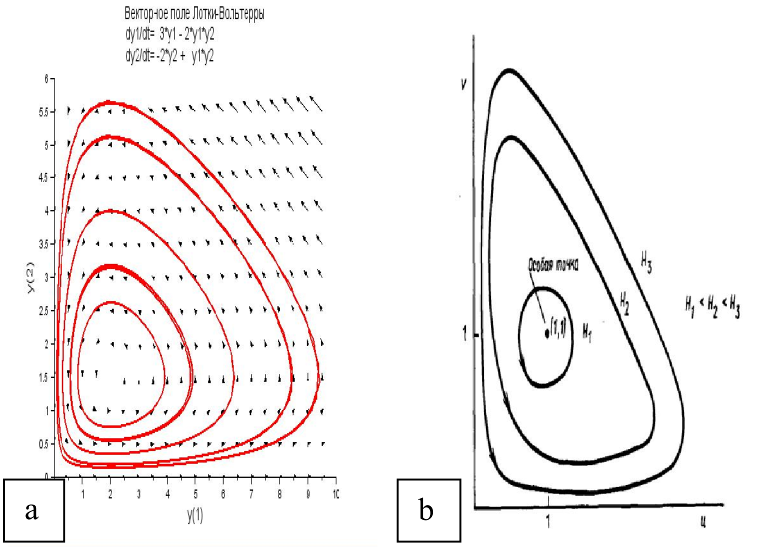


Fig. 2. a) Oscillator of Lotka-Volterra equations in the shape of geometric regions for nonlinear analysis of signs relationship (in our case KBN- MNB) in field pea samples (based on program Scilab), 2012
b) Theoretical field of closed trajectories of biological oscillator of the factors nonlinear interaction, including specific points by (J. Murray, 1983). ***$au+v-ln(u^av)=H=const$***

As a result of these approximations, a stable parametric core of normal genotype is distinguishing. The samples icons feature the most stable and balanced in size vectors cores have the largest breeding value (Fig. 3, 4).

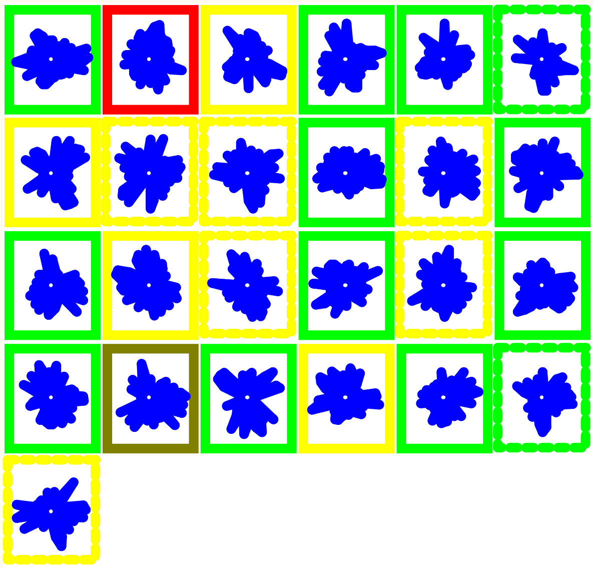


Fig. 3. Parametric characteristics of quantitative traits in winter rye samples

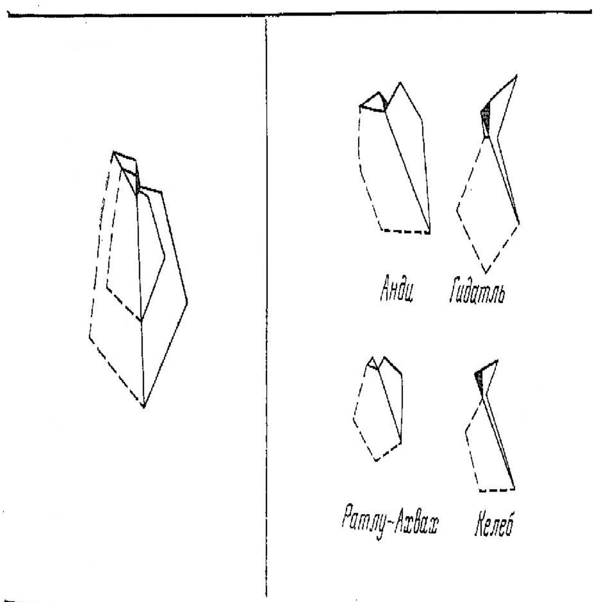


Fig. 4. Characteristics of chicken populations of different gene geographic origin (Andean and Avar) as reflected in vector-gradient system according to A.S. Serebrovskyi.

The need to take into account all the components and interactions between phenotypic variances exists due to the fact that, according to many researchers, biological systems are not "rough", i.e. the slightest influence of external factors, especially on the multiplicative component, causes significant changes in the phase-parameter space of the system. Based on example of a graphical model of the complete variation set (Fig. 5), it was found that even small coordinate-parametric changes in the component system (X, Y) lead to significant coordinate changes in the complex trait (Z). V.A. Drahavtsev et al. have offered a new approach to express estimation of genotypic and genetic (additive) variations in the plant productive capacity by discriminant principle of some plants distribution (Fig. 6) in wheat population that splits in two-dimensional coordinate system of traits [8].

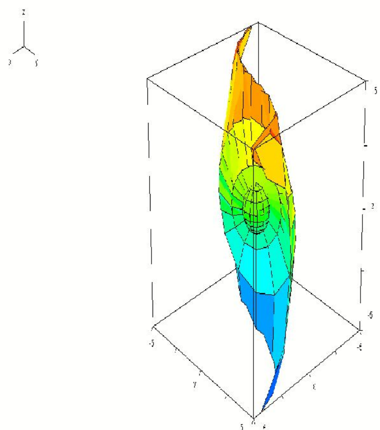


Fig. 5. Graphical model of full range of variance in general. Centre (core) is an additive part, peripherals are multiplicative complex fluctuations.

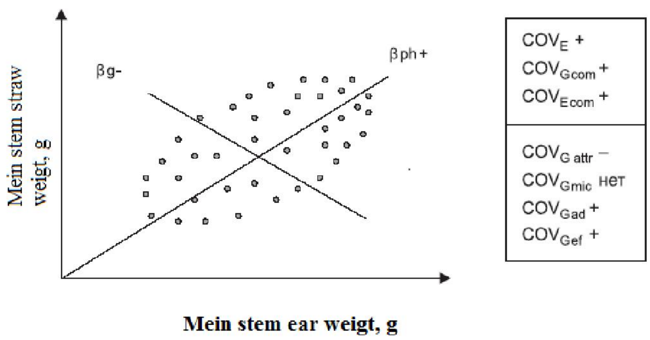


Fig. 6. Distribution of some plant in wheat population, which split in two-dimensional sighs coordinate system.

Resulted from experiments carried out over many years in breeding nurseries of various level for traditional in Polissia crops, such as yellow and blue lupine, field pea, winter rye, potato, accumulated is a collection of subcomponent interaction surfaces of component features, component features in the formation of complex one, traits and environment factors. Form of the surfaces varies significantly and is differentiated according to the number of components involved in its formation, especially with time series. Important is the methodological basis for the surface formation principle according to the chosen model, in particular linear-nonlinear, logistic, exponential, phase-parametric including branches in a case of one-dimensional one-parameter bifurcation, dimensional bifurcation with a large number of parameters, bifurcation of “collection” type, bifurcation of “swallow tail” type (Fig. 7).

However, it is possible to combine received collection of real surfaces in canonical figures of analytic geometry that could be described by certain canonical equations (e.g. 7a, 7d through 8a, 7b, 7e through 8b, 7b, 7f through 8c). Using coefficients λ (lambda) it is possible to build diagonal matrices, which will describe abstract surfaces.

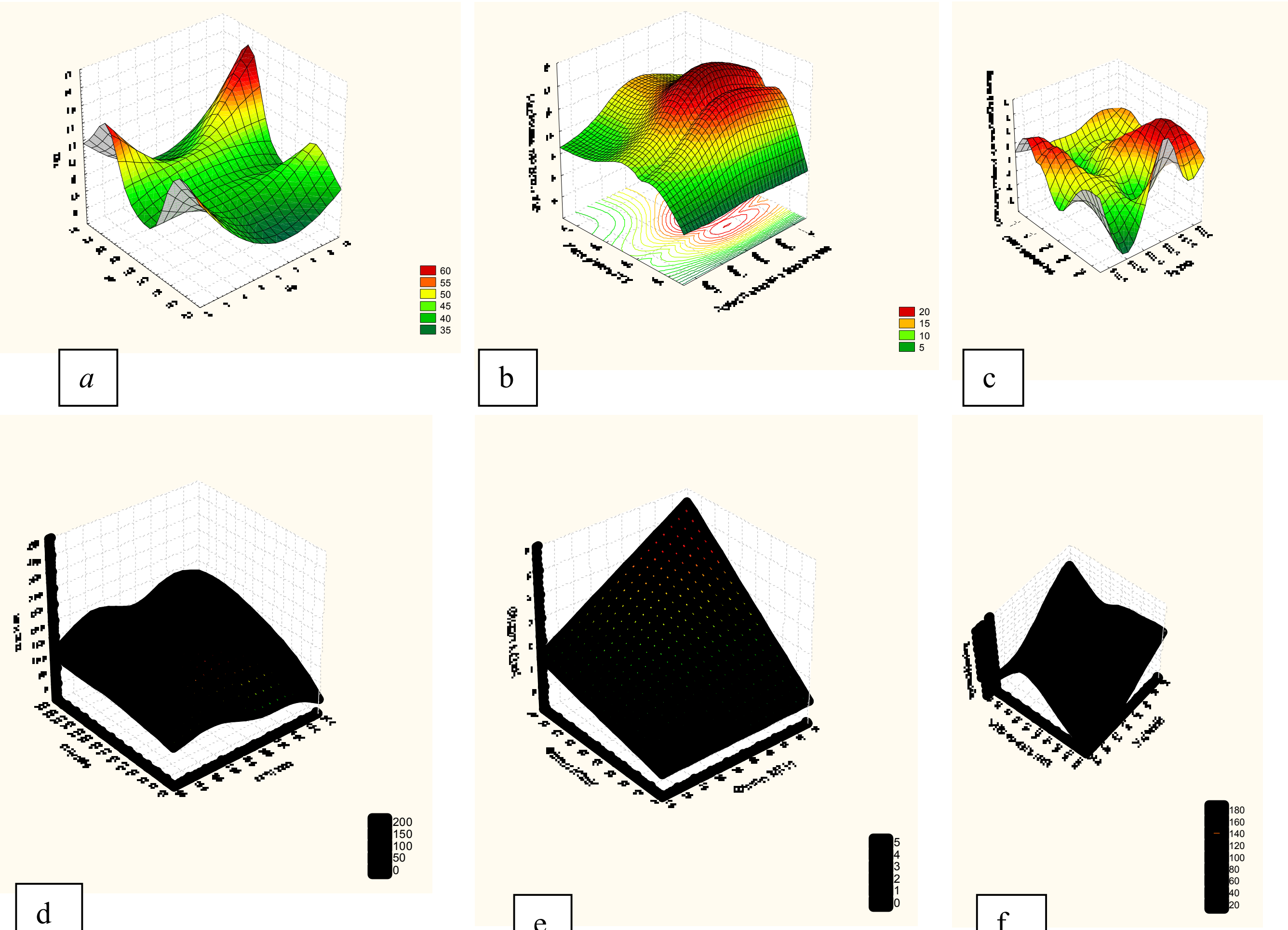
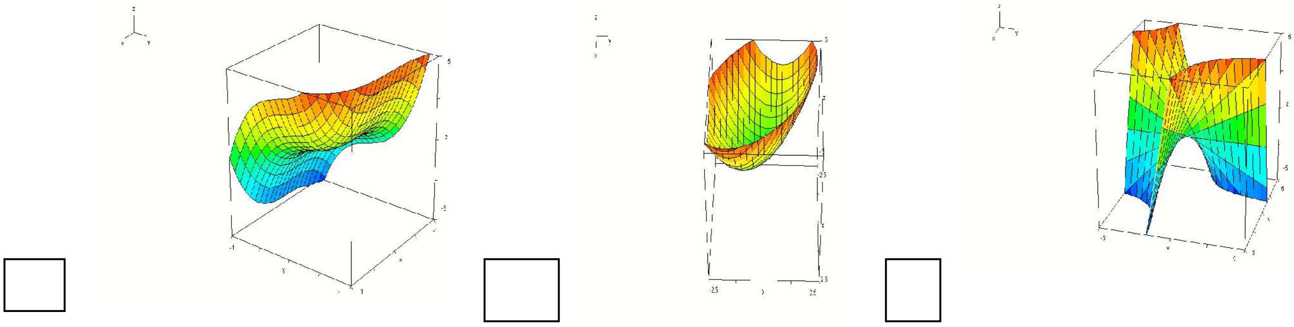


Fig. 7. Collection of surface shapes including: a - hyperbolic paraboloid; b - paraboloid; c - COM surface; d - fold; d - simple linear dependence; e - differentiated maxima.

On their base it is possible to build predictive models to identify the optimal combinations of component traits in the formation of the complex one, including under the influence of endogenous and exogenous environmental factors (Fig. 8).



Rice. 8. Abstract of surface for the decision of polynomials of different types (polynomial of $z=x^2+y^2+x+y+xy$)

Conclusions

1. To improve the efficiency of selection, a rational way could be to using a complete version of the analysis of phenotypic variation in samples taking into account the genotype-environmental effects.

2. One of the promising methods for finding optimized combinations of component traits while forming complex one is the method of vector-gradient splicing on curved surfaces.

3. Introduction of modern IT technologies including computer-aided data collection and processing to breeding practice provides a significant resource-saving effect.

References

1. Літун П.П. Теорія і практика селекції на макроознаки. Методологічні проблеми / П.П. Літун, В.В. Кириченко, В.П. Петренкова [та ін.]. – Харків : [б. и.], 2004. – 130 с.

2. Тороп Е. А. Морфогенетические закономерности формирования продуктивности озимой ржи (*Secale cereale L.*) : автореф. дис. на соискание учёной степени доктора биологических наук: спец. 06.01.05 «Селекция и семеноводство сельскохозяйственных растений» / Е. А.Тороп. – Рамонь, 2011. – 38 с.

3. Смиряев А.В. Совершенствование биометрического метода оптимизации выборки растений при экспериментальном сравнении гибридных популяций / А.В. Смиряев, Т.И. Хупацария, С.С. Баженова [и др.] // Известия ТСХА. – 2012. – Выпуск 2. – С. 71-86.

4. Сюков В.В. Вклад генотип-средовых эффектов в формирование количественных признаков у инбредных и аутбредных растений / В.В. Сюков, Е.В. Мадякин, Д.В. Кочетков // Весник ВОГиС. – 2010. – Т. 14, № 1. – С. 141-147.

5. Альт В.В. Методология формирования баз данных по сортам пшеницы и ячменя / В.В. Альт, П.Л. Гончаров, Н.А. Сурин // Вестник ВОГиС. – 2008. – Т. 12, № 4. – С. 717-725.

6. Генаев М.А. Компютерная система WheatPGE для анализа взаимосвязи фенотип-генотип-окружающая среда у пшеницы / М.А. Генаев, А.В. Дорошков, Е.В. Морозова [и др.] // Вавиловский журнал генетики и селекции. – 2011. – Т. 15, № 4. – С.784-793.

7. Дубравина И.В. Актуализация использования математических аппаратов в селекции яблони на заданные признаки / И.В. Дубравина, С.В. Пермякова // Научный журнал Кубанского ГАУ. – 2013. – № 90 (06). – С. 1-15.

8. Драгавцев В.А. Новые подходы к экспрессной оценке генотипической и генетической (аддитивной) дисперсий свойств продуктивности растений / В.А. Драгавцев, Г.А. Макарова, А.А. Кочетов [и др.] // Вавиловский журнал генетики и селекции. – 2012. – Т. 16, № 2. – С. 427-436.

Анотація

Чернуський В.В.

Методологічні підходи до створення системних автоматизованих комплексів збору та аналізу даних у процесі добору у зв'язку з селекцією сільськогосподарських культур

Отримання максимального масиву інформації – одне з найважливіших методологічних завдань експерименту. Виявлена необхідність встановлення і аналізу не тільки адитивної, але і мультиплікативної частини дисперсії експериментальної системи. Зокрема за рахунок обліку повного спектру фенотипової мінливості на векторно-градієнтному полі онтогенетичного розвитку компонентних і комплексних макроознак значно підвищилась точність ідентифікації генотипів і відповідно ефективність добору. За рахунок впровадження в селекційну практику системи нелінійно-синергетичного принципу добору створено ряд сортів люпину вузьколистого, гороху польового, жита озимого різних напрямів господарського використання адаптованих до умов Полісся.

Ключові слова: *адитивна, мультиплікативна дисперсія, добір, синергетика макроознак, сорти*

Аннотация

Чернусский В.В.

Методологические подходы к созданию системных автоматизированных комплексов сбора и анализа данных в процессе отбора в связи с селекцией сельскохозяйственных культур

Получение максимального массива информации – одно из важнейших методологических заданий эксперимента. Обнаружена необходимость установления и анализа не только аддитивной, но и мультипликативной, части дисперсии экспериментальной системы. В частности за счет учета полного спектра фенотипической изменчивости на векторно-градиентном поле онтогенетического развития компонентных и комплексных макропризнаков значительно повысилась точность идентификации генотипов и соответственно эффективность отбора. За счет внедрения в селекционную практику системы нелинейно синергетического принципа отбора создан ряд сортов люпина узколистного, гороха полевого, ржи озимой разных направлений хозяйственного использования адаптированных к условиям Полесья.

Ключевые слова: *адитивная, мультипликативная дисперсия, отбор, синергетика макропризнаков, сорта*