

К вопросу автоматизации построения вариограмм в задачах точного земледелия*

В. П. Якушев¹, В. М. Буре^{1,2}, О. А. Митрофанова^{1,2}, Е. П. Митрофанов^{1,2}

¹ Агрофизический научно-исследовательский институт, Российская Федерация, 195220, Санкт-Петербург, Гражданский пр., 14

² Санкт-Петербургский государственный университет, Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9

Для цитирования: *Якушев В. П., Буре В. М., Митрофанова О. А., Митрофанов Е. П.* К вопросу автоматизации построения вариограмм в задачах точного земледелия // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2020. Т. 16. Вып. 2. С. 177–185. <https://doi.org/10.21638/11701/spbu10.2020.209>

Большинство задач точного земледелия (ТЗ) основаны на оценке изменчивости агроэкологических параметров (дифференцированное внесение удобрений, выделение неоднородных зон сельскохозяйственного поля и др.). В связи с этим актуальным и перспективным направлением является изучение пространственной структуры таких данных с помощью геостатистических методов. Для выполнения валидации модели вариограммы, определения ее параметров, анализа анизотропии и осуществления других этапов эксперимента требуется значительное количество численных расчетов. Вручную их все провести крайне сложно, потому необходима автоматизация таких процессов. Большинство существующих программ исключает определенные действия, которые могут иметь важное значение при решении задач ТЗ (нет возможности исследования анизотропии или пространственного тренда, ограничено число теоретических моделей вариограммы и т. п.), а использование языков программирования (R, Python и др.) требует глубоких экспертных знаний. Поэтому возникла необходимость автоматизировать решение определенного круга задач геостатистическими методами. Для реализации рассматриваемого модуля оптимальным представляется применение языка программирования R, обладающего рядом существенных преимуществ: открытый исходный код и бесплатный доступ, значительное количество поддерживаемых и регулярно пополняемых пакетов, широкие графические возможности, кросс-платформенность и др. Представлены общие предложения по автоматизации построения вариограмм и дальнейшего использования в решении определенного круга задач ТЗ.

Ключевые слова: вариограммный анализ, точное земледелие, геостатистика, автоматизация, язык программирования R.

Введение. Технологии точного земледелия (ТЗ) основываются на исследованиях пространственной вариабельности агроэкологических параметров. Прецизионное применение сельскохозяйственных ресурсов позволяет не только экономить финансовые и трудозатраты, но и повышать урожайность, а также снижать вредное воздействие на окружающую среду [1]. В последние годы активно развиваются различные системы поддержки принятия решений для предоставления специализированных программных инструментов и баз знаний сельскохозяйственным производителям [2]. Такие системы предназначены для оперативного сбора, хранения и обработки множе-

* Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 19-29-05184 мк).

© Санкт-Петербургский государственный университет, 2020

ства типов агроэкологических данных (данные дистанционного зондирования (ДДЗ), агрохимические показатели почвы и т. п.).

Геостатистика включает в себя ряд методов, позволяющих осуществлять анализ вариабельности пространственно распределенных данных. Этот подход активно используется не только в геологии [3], но и в других областях, оперирующих пространственно структурированными параметрами, например в метеорологии [4], геохимии [5], экологии [6] и др. В последние десятилетия он становится также актуальным и востребованным для решения задач ТЗ [7]. Основные этапы геостатистического анализа состоят из построения экспериментальной вариограммы, подбора ее теоретической модели, а также при необходимости пространственной интерполяции методом кригинга.

Для выполнения валидации модели вариограммы, определения ее параметров, анализа анизотропии и осуществления других этапов эксперимента требуется большое количество численных расчетов. Вручную их провести крайне сложно, в связи с чем необходима автоматизация таких процессов. В таблице представлены наиболее используемые бесплатные программы, включающие в себя ряд этапов геостатистического анализа. Однако следует отметить, что большинство из них исключают определенные действия, которые могут иметь важное значение при решении задач ТЗ (нет возможности исследования анизотропии или пространственного тренда, ограничено число теоретических моделей вариограммы и т. п.). Целесообразно отдельно выделить язык программирования R, который ориентирован на решение множества задач с помощью статистического анализа. Главным его недостатком является необходимость глубоких экспертных знаний, для пользователей в области решения практических и научно-исследовательских задач ТЗ он слишком сложный. Поэтому следует автоматизировать решение определенного круга задач геостатистическими методами.

Таблица. Наиболее используемые бесплатные программы, включающие в себя методы геостатистики

Характеристики	Программные инструменты				
	GeoDa	GMT	GSLIB	R (GStat, geoR,...)	SAGA GIS
Вариограммный анализ	-	-	+	+	+
Кригинг	-	-	+	+	+
Кокригинг	-	-	+	+	-
Альтернативные методы пространственной интерполяции	+	+	-	+	+
ГИС-функции	+	-	-	-	+
ОС	Windows, MacOS, Linux	Windows, MacOS, Linux	Windows, Linux, Dos	Windows, MacOS, Linux	Windows, Linux
Последняя версия	1.14.0 (2019)	6.0.0 (2019)	1.5 (2009)	4.0.0 (2020)	7.6.2 (2020)
Страна	США	США	США	Международный проект	Германия

Объекты и методы. К основным задачам ТЗ, требующим применения методов геостатистики, относятся:

— оценка пространственного распределения агроэкологических данных методом кригинга, кокригинга и т. п. (например, для дифференцированного внесения удобрений [7]);

- изучение внутриполевой неоднородности с использованием вариограммного анализа [8];
- мониторинг мелиоративных и прецизионное применение ирригационных систем [9, 10];
- оценка эффективности перехода к дифференцированным технологиям на конкретной сельскохозяйственной территории [11].

В большинстве задач статистический анализ осуществляется на основе набора агроэкологических данных $Z(x_i)$, это могут быть значения вегетационного индекса NDVI (Normalized Difference Vegetation Index), полученные с помощью аэрофото- и космоснимков в ближнем инфракрасном и видимом диапазонах; агрохимические и физические показатели почвы (влажность, электропроводность, содержание питательных веществ и т. п.) и др. В первую очередь необходимо автоматизировать вариограммный анализ, поскольку именно на этом этапе геостатистики существенные недостатки выявлены в доступных программах для решения определенного круга задач ТЗ.

Вариограмма представляется графически как функция расстояний между наблюдениями, оценивается она на основе экспериментальной вариограммы [12]:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [Z(x_i) - Z(x_i + \mathbf{h})]^2,$$

где \mathbf{h} — вектор расстояния между двумя точками; $N(\mathbf{h})$ — число пар экспериментальных точек, разделенных вектором \mathbf{h} . Для ее построения целесообразно проводить несколько этапов предварительного анализа: визуализация исходных данных на карте, расчет основных статистических показателей, проверка ограничений и предположений геостатистики (выбросы в данных, наличие пространственного тренда), построение гистограммы частот, выбор шага и количества лагов, анализ анизотропии.

На заключительном этапе подбирается теоретическая модель вариограммы (сферическая, экспоненциальная, гауссова, модель с эффектом дырок и др.). Основные компоненты вариограммы: «самородок» (c_0) — значение начальной изменчивости; «ранг» (a) — расстояние, на котором данные становятся полностью независимыми; «порог» (c) — предельное значение вариограммы, достигаемое на расстоянии ранга.

В качестве проверки построений предлагается использовать классический подход, основанный на вычислении стандартного отклонения значений экспериментальной вариограммы и подобранной теоретической модели. Однако целесообразно также протестировать применение более новых подходов, например предложенный в работе [13] механизм подбора модели вариограммы, который одновременно учитывает пространственную структуру изучаемого параметра и сопутствующую ей точность интерполяции.

Результаты и их обсуждение. Геостатистический модуль, автоматизирующий вариограммный анализ для задач ТЗ, предлагается внедрить в программный веб-сервис, разрабатываемый в Агрофизическом НИИ. На рис. 1 схематично представлена архитектура этого проекта. Платформа предназначена для оперативного сбора, хранения и всесторонней обработки разнородных данных, используемых для широкого круга востребованных задач в области производства растениеводческой продукции.

Для реализации рассматриваемого модуля оптимальным представляется использование языка программирования R, обладающего некоторыми существенными преимуществами: открытый исходный код и бесплатный доступ, большое количество поддерживаемых и регулярно пополняемых пакетов, широкие графические возмож-

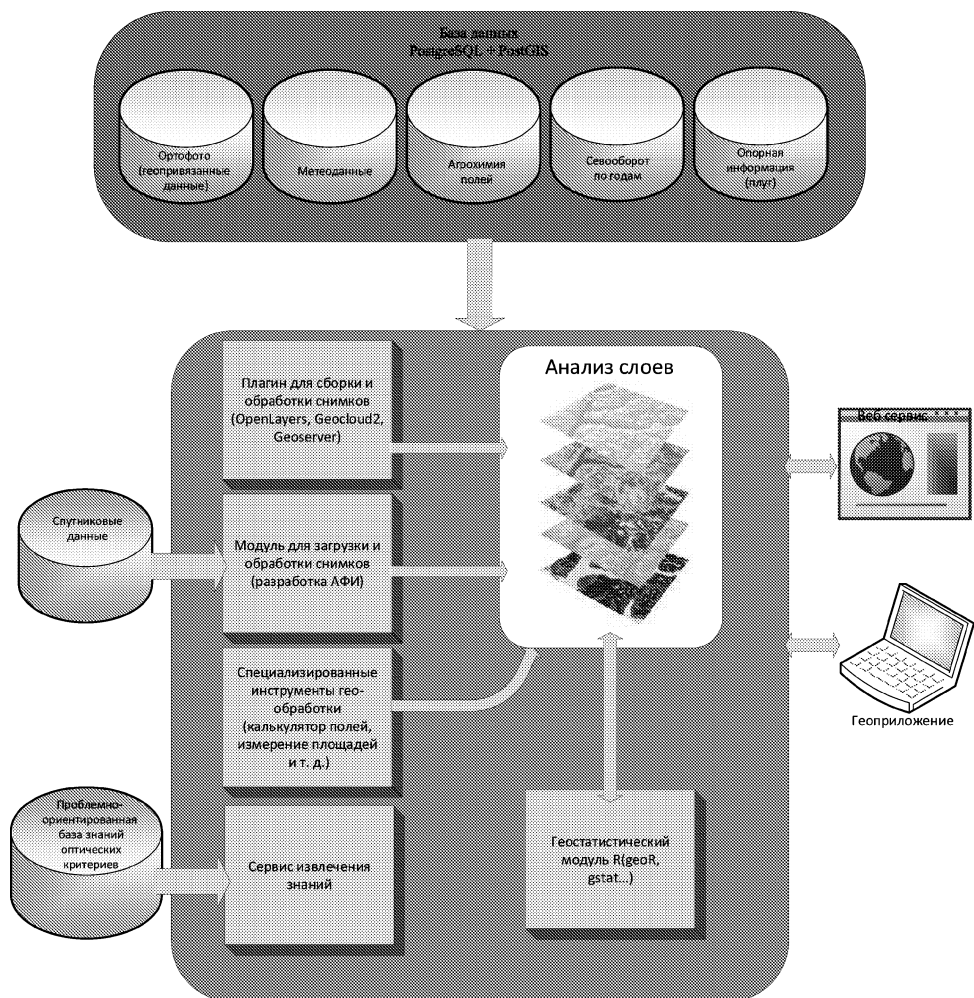


Рис. 1. Структурная схема веб-сервиса, разрабатываемого в Агрофизическом НИИ

ности, высокая совместимость с другими языками программирования (например, Python, который предполагается применять для реализации веб-сервиса и геоприложения), кросс-платформенность, значительный объем инструментария статистического анализа и машинного обучения. На рис. 2 схематично представлен общий алгоритм реализации вариограммного анализа.

При реализации модуля предполагается, что пользователю будут доступны дополнительные функции анализа (не по умолчанию, а в качестве отдельных опций), например для анализа выбросов в экспериментальных данных. Основной алгоритм можно также схематично представить следующим образом:

Data: $Z(x, y) = (Z_1, \dots, Z_n)$ – набор агроэкологических геoprивязанных данных, $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n)$ – координаты экспериментальных точек, n – количество наблюдений.

Result: График вариограммы и основные результаты построения (параметры модели, нагетт-дисперсия и т. п.).

```

//
Визуализация исходных данных (plot);
Output: карта распределения исследуемого параметра;
// Блок предварительного анализа:
trend ← оценка присутствия пространственного тренда (линейная корреляция пере-
менной с координатами, cor.test);
if trend == TRUE then
{
    Сообщение пользователю об обнаружении дрефта;
    Устранение тренда построением его модели (строятся несколько основных, срав-
ниваются коэффициенты детерминации, выбирается наилучшая) и элиминацией
(вычитаются модельные значения из данных, дальнейший анализ проводится на
остатках);
}
Построение гистограммы расстояний (hist);
Расчет оптимальных значений шага и числа лагов с помощью оценки разброса рас-
стояний, их запись;
Output: Рекомендуемые параметры экспериментальной вариограммы;
Input: Пользователь либо принимает рекомендуемые, либо вводит свои параметры
экспериментальной вариограммы;
// Блок построения экспериментальной вариограммы:
anis ← оценка присутствия анизотропии (строятся экспериментальные вариограммы
в разных направлениях и рассчитываются отклонения полученных построений друг
от друга, в случае превышения допустимого значения запоминаются направления
анизотропии и возвращается TRUE, variogram);
Расчет и запись основных параметров теоретической вариограммы (определяются
начальное значение и предельное, а также расстояние, где оно достигается);
if anis == TRUE then
{
    В параметры теоретической вариограммы записывается информация об
анизотропии;
    Output: Графики экспериментальной вариограммы по направлениям, сообщение
пользователю, рекомендуемые параметры теоретической вариограммы;
}
else
    Output: График экспериментальной вариограммы, рекомендуемые параметры
теоретической вариограммы;
Input: Пользователь либо принимает рекомендуемые, либо вводит свои параметры
теоретической вариограммы, из выпадающего списка выбирает модель;
// Блок построения теоретической вариограммы:
Построение теоретической модели (vgm);
Output: Наложение на график экспериментальной вариограммы теоретической мо-
дели;
Подгонка теоретической модели (валидация, fit.variogram);
Output: Скорректированный график, результаты построений.
//

```

У существующих программных инструментов, реализующих геостатистические



Рис. 2. Общий алгоритм автоматизации вариограммного анализа

подходы, выделяются две основные проблемы: либо расчеты чрезмерно упрощены для пользователя, соответственно упускается часть важных статистических проверок и результатов, либо применение средств автоматизации требует от пользователя глубоких экспертных знаний. Данный модуль от существующих аналогов отличает ряд особенностей: на каждом этапе построений рассчитываются оптимальные параметры для дальнейшего анализа, пользователь может согласиться применять предложенные программой либо ввести свои показатели; полученные результаты могут быть напрямую автоматически применены для решения определенной задачи ТЗ (например, выдать заключение о целесообразности перехода к дифференцированным технологиям на конкретном сельскохозяйственном поле на основе вычисленной нагдет-дисперсии).

Заключение. В работе рассмотрен вопрос автоматизации построения вариограмм для решения определенного круга задач ТЗ, связанных с исследованием пространственной вариабельности агроэкологических данных. Предложены общий алгоритм реализации, а также структура веб-платформы с составным геостатистическим модулем, отвечающей за оперативный сбор, хранение и обработку геопривязанных параметров.

Данный проект позволит автоматизировать большой объем вычислительных процессов, а также внедрить информационные технологии в процесс производства растениеводческой продукции.

Литература

1. Якушев В. В. Точное земледелие: теория и практика. СПб.: Агрофиз. науч.-исслед. ин-т, 2016. 364 с.
2. Chen N., Zhang X., Wang C. Integrated open geospatial web service enabled cyber-physical information infrastructure for precision agriculture monitoring // Comput. Electron. Agric. 2015. Vol. 111. P. 78–91.
3. Hohn M. E. Geostatistics and petroleum geology. 2nd ed. Dordrecht, The Netherlands: Springer, Science+Business Media, 1999. 235 p.

4. Gomez J. L., Pastoriza F. T., Alvarez E. G., Oller P. E. Comparison between geostatistical interpolation and numerical weather model predictions for meteorological conditions mapping // *Infrastructures*. 2020. Vol. 5. N 15. DOI:10.3390/infrastructures5020015
5. Paz-Ferreiro J., Vazquez E. V., Vieira S. R. Geostatistical analysis of a geochemical dataset // *Bragantia*. 2009. Vol. 69. P. 121–129.
6. Olthoff A. E., Gomez C., Alday J. G., Martinez-Ruiz C. Mapping forest vegetation patterns in an Atlantic–Mediterranean transitional area by integration of ordination and geostatistical techniques // *Journal of Plant Ecology*. 2018. Vol. 11. Iss. 1. P. 114–122.
7. Якушев В. П., Жукковский Е. Е., Петрушин А. Ф., Якушев В. В. Вариограммный анализ пространственной неоднородности сельскохозяйственных полей для целей точного земледелия: метод. пособие. СПб.: Агрофиз. науч.-исслед. ин-т, 2010. 52 с.
8. Cambardella C. A., Moorman T. B., Novak J. M., Parkin T. B., Karlen D. L., Turko R. F. et al. Field-scale variability of soil properties in central Iowa soils // *Soil Science Society of America Journal*. 1994. Vol. 58. P. 1501–1511.
9. Jiang Q. X., Fu Q., Wang Z. L. Research on precision irrigation in Western Semiarid Area of Heilongjiang province in China based on GIS // *Computer and Computing Technologies in Agriculture*. 2008. Vol. 1. P. 359–370.
10. Moustafa M. M., Yomota A. Use of a covariance variogram to investigate influence of subsurface drainage on spatial variability of soil-water properties // *Agricultural Water Management*. 1998. Vol. 37. P. 1–19.
11. Якушев В. П., Буре В. М., Митрофанова О. А., Митрофанов Е. П. Применение методов геостатистики для анализа целесообразности перехода к технологиям дифференцированного внесения агрохимикатов // *Вестн. С.-Петерб. ун-та. Прикладная математика. Информатика. Процессы управления*. 2020. Т. 16. Вып. 1. С. 31–40.
12. Демьянов В. В., Савельева Е. А. Геостатистика: теория и практика. М.: Ин-т проблем безопасности развития атомной энергетики РАН; Наука, 2010. 327 с.
13. Li Z., Zhang X., Clarke K. C., Liu G., Zhu R. An automatic variogram modeling method with high reliability fitness and estimates // *Computers and Geosciences*. 2018. Vol. 120. P. 48–59.

Статья поступила в редакцию 3 марта 2020 г.

Статья принята к печати 28 мая 2020 г.

Контактная информация:

Якушев Виктор Петрович — акад. РАН, д-р с.-х. наук; vyakushev@agrophys.com

Буре Владимир Мансурович — д-р техн. наук, проф.; vlb310154@gmail.com

Митрофанова Ольга Александровна — мл. науч. сотр.; omitrofa@gmail.com

Митрофанов Евгений Павлович — мл. науч. сотр.; mjeka@bk.ru

On the issue of semivariograms constructing automation for precision agriculture problems*

V. P. Iakushev¹, V. M. Bure^{1,2}, O. A. Mitrofanova^{1,2}, E. P. Mitrofanov^{1,2}

¹ Agrophysical Research Institute, 14, Grazhdanskiy pr., St. Petersburg, 195220, Russian Federation

² St. Petersburg State University, 7–9, Universitetskaya nab., St. Petersburg, 199034, Russian Federation

For citation: Iakushev V. P., Bure V. M., Mitrofanova O. A., Mitrofanov E. P. On the issue of semivariograms constructing automation for precision agriculture problems. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, 2020, vol. 16, iss. 2, pp. 177–185. <https://doi.org/10.21638/11701/spbu10.2020.209> (In Russian)

* This work was supported by the Russian Foundation for Basic Research (grant N 19-29-05184 mk).

Most precision agriculture (PA) problems are based on assessing the variability of agroecological parameters (differentiated fertilizer application, allocation of heterogeneous zones of the agricultural field, etc.). In this connection, the study of the spatial structure of such data using geostatistical methods seems to be a relevant and promising direction. To validate the semivariogram model, determine its parameters, analyze anisotropy, and carry out other stages of the experiment, a large number of numerical calculations are required. Manually all these calculations are extremely difficult to perform, therefore, automation of these processes is necessary. Most existing programs exclude certain actions that may be important in solving PA problems (there is no possibility of studying anisotropy or spatial trend, the number of theoretical variogram models, etc., is limited), and the use of programming languages (R, Python, etc.) requires deep expertise. Therefore, there was a need to automate the solution of a certain range of problems by geostatistical methods. For the implementation of the module under consideration, it seems optimal to use the R programming language, which has a number of significant advantages: open source code and free access, a large number of supported and regularly updated packages, wide graphical capabilities, cross-platform, etc. General suggestions for automating the construction of semivariograms are presented and further use in solving a certain range of PA tasks.

Keywords: variogram analysis, precision agriculture, geostatistics, automation, programming language R.

References

1. Iakushev V. V. *Tochnoe zemledelie: teoriia i praktika [Precision agriculture: theory and practice]*. Saint Petersburg, Agrophysical Research Institute Publ., 2016, 364 p. (In Russian)
2. Chen N., Zhang X., Wang C. Integrated open geospatial web service enabled cyber-physical information infrastructure for precision agriculture monitoring. *Comput. Electron. Agric.*, 2015, vol. 111, pp. 78–91.
3. Hohn M. E. *Geostatistics and petroleum geology*. 2nd ed. Dordrecht, The Netherlands, Springer, Science+Business Media Press, 1999, 235 p.
4. Gomez J. L., Pastoriza F. T., Alvarez E. G., Oller P. E. Comparison between geostatistical interpolation and numerical weather model predictions for meteorological conditions mapping. *Infrastructures*, 2020, vol. 5, N 15. DOI:10.3390/infrastructures5020015
5. Paz-Ferreiro J., Vazquez E. V., Vieira S. R. Geostatistical analysis of a geochemical dataset. *Bragantia*, 2009, vol. 69, pp. 121–129.
6. Olthoff A. E., Gomez C., Alday J. G., Martinez-Ruiz C. Mapping forest vegetation patterns in an Atlantic–Mediterranean transitional area by integration of ordination and geostatistical techniques. *Journal of Plant Ecology*, 2018, vol. 11, iss. 1, pp. 114–122.
7. Iakushev V. P., Zhukovskii E. E., Petrushin A. F., Iakushev V. V. *Variogramnyi analiz prostranstvennoi neodnorodnosti sel'skokhoziaistvennykh polei dlia tselei tochnogo zemledeliia*. Metodicheskoe posobie [A variogram analysis of the spatial heterogeneity of agricultural fields for precision agriculture. Toolkit]. Saint Petersburg, Agrophysical Research Institute Publ., 2010, 52 p. (In Russian)
8. Cambardella C. A., Moorman T. B., Novak J. M., Parkin T. B., Karlen D. L., Turko R. F. et al. Field-scale variability of soil properties in central Iowa soils. *Soil Science Society of America Journal*, 1994, vol. 58, pp. 1501–1511.
9. Jiang Q. X., Fu Q., Wang Z. L. Research on precision irrigation in Western Semiarid Area of Heilongjiang province in China based on GIS. *Computer and Computing Technologies in Agriculture*, 2008, vol. 1, pp. 359–370.
10. Moustafa M. M., Yomota A. Use of a covariance variogram to investigate influence of subsurface drainage on spatial variability of soil-water properties. *Agricultural Water Management*, 1998, vol. 37, pp. 1–19.
11. Iakushev V. P., Bure V. M., Mitrofanova O. A., Mitrofanov E. P. Primenenie metodov geostatistiki dlia analiza tselesoobraznosti perekhoda k tekhnologiiim differentsirovannogo vneseniia agrokhimikatov [The use of geostatistical methods to analyze the transition feasibility to the differential application of agrochemicals technologies]. *Vestnik of Saint Peterburg University. Applied Mathematics. Computer Science. Control Processes*, 2020, vol. 16, iss. 1, pp. 31–40. (In Russian)
12. Dem'ianov V. V., Savel'eva E. A. *Geostatistika: teoriia i praktika [Geostatistics: theory and*

practice]. Moscow, Nuclear safety institute of the Russian Academy of Sciences, Nauka Publ., 2010, 327 p. (In Russian)

13. Li Z., Zhang X., Clarke K. C., Liu G., Zhu R. An automatic variogram modeling method with high reliability fitness and estimates. *Computers and Geosciences*, 2018, vol. 120, pp. 48–59.

Received: March 03, 2020.

Accepted: May 28, 2020.

Authors' information:

Viktor P. Iakushev — RAS Academician, Dr. Sci. in Agriculture; vyakushev@agrophys.com

Vladimir M. Bure — Dr. Sci. in Technics, Professor; vlb310154@gmail.com

Olga A. Mitrofanova — Junior Researcher; omitrofa@gmail.com

Evgenii P. Mitrofanov — Junior Researcher; mjeka@bk.ru