

Прогнозирование урожайности сельскохозяйственных культур на основе данных дистанционного зондирования Земли (на примере сои)*

А. С. СТЕПАНОВ

Дальневосточный научно-исследовательский институт сельского хозяйства, с. Восточное, Хабаровский край, Россия

Контактный e-mail: stepanxx@mail.ru

Описан подход к прогнозированию урожайности сельскохозяйственных культур с использованием данных дистанционного зондирования Земли. В качестве основного параметра прогностической регрессионной модели использовались значения вегетационного индекса NDVI. В статье приведена оценка возможности раннего прогнозирования до достижения индексом NDVI максимальных значений с применением гауссианы в качестве аппроксимирующей функции, соответствующей еженедельным композитам NDVI. Для пахотных земель Тамбовского р-на Амурской области рассчитана ошибка определения максимума NDVI в зависимости от календарной недели прогнозирования. Построенная модель использована для предварительной оценки урожайности сои в регионе в 2018 г.

Ключевые слова: вегетационный индекс, функция Гаусса, пахотные земли, прогнозирование, урожайность, соя, Python.

Библиографическая ссылка: Степанов А.С. Прогнозирование урожайности сельскохозяйственных культур на основе данных дистанционного зондирования Земли (на примере сои) // Вычислительные технологии. 2019. Т. 24, № 6. С. 125–133. DOI: 10.25743/ICT.2019.24.6.015.

Введение

Технологии дистанционного зондирования Земли активно применяются в сельском хозяйстве, в том числе и для прогнозирования урожайности сельскохозяйственных культур [1]. В качестве основного показателя для построения прогнозной модели используется вегетационный индекс NDVI, имеющий высокую корреляцию с урожайностью сельхозкультур [2]. В регрессионную модель чаще всего включаются максимальное значение динамического ряда NDVI и интегрированные показатели, характеризующие климатические условия территории [3, 4]. Однако применение в модели фактического максимума NDVI снижает возможность раннего прогнозирования и, соответственно, практическую значимость прогноза для сельского хозяйства. Связано это не только с тем, что максимальное значение NDVI в большинстве регионов РФ достигается не ранее середины лета, но и с вариативностью даты прохождения максимума в пределах нескольких недель в разные годы [5]. Таким образом, проблема моделирования

*Title translation and abstract in English can be found on page 133.

© ИВТ СО РАН, 2019.

динамики вегетационного индекса для разных культур, оценка определения времени достижения максимума и раннее прогнозирование максимума являются актуальными задачами [6].

В качестве другого аспекта проблемы прогнозирования урожайности сельскохозяйственных культур можно выделить необходимость поиска оптимальной регрессионной модели, где в качестве одной или нескольких независимых переменных выступают вегетационный индекс и интегральные показатели, характеризующие метеорологические и климатические особенности региона. В качестве критерия выбора модели рассматриваются точность модели и средняя абсолютная ошибка прогноза по смоделированным данным. Немаловажным фактором, непосредственно влияющим на повышение точности модели, является необходимость предварительной обработки данных по посевным площадям и валовому сбору сельхозкультур.

В целом разработка методологического подхода к прогнозированию урожайности сельхозкультур на региональном уровне должна заключаться в поэтапном решении обозначенных выше задач в процессе поиска оптимальной модели, а также тестировании разработанной модели на примере конкретного региона или муниципалитета.

1. Объекты и методы исследования

Построение и оценка точности прогнозной модели осуществлялись на примере Тамбовского района Амурской области, являющегося одним из основных сельскохозяйственных районов юга Дальнего Востока и при этом лидирующей территорией по производству сои в РФ. Рассматривались семидневные (еженедельные) значения индекса NDVI (безоблачные композиты MODIS) за календарный год, вычисленные по маске пахотных земель. В исследовании использовались результаты обработки спутниковых данных, полученные посредством информационной системы Vega-Science [7], нормализованные значения NDVI за календарную неделю рассчитывались как средние значения за 5 лет наблюдений. Всего было сформировано 10 динамических рядов нормализованных индексов, соответствующих 2009–2018 гг. Максимальное значение NDVI достигалось в период 29–32-й календарной недели, что соответствует концу июля — началу августа. Определение максимума NDVI при раннем прогнозировании урожайности осуществлялось путем поиска параметров аппроксимирующей нелинейной функции, соответствующей распределению нормализованных значений NDVI предыдущих лет. Расчеты выполнены в среде программирования Python с использованием дополнительных научных пакетов [8, 9].

Данные по урожайности сои в период с 2007 по 2018 г. в Тамбовском районе Амурской области получены с применением базы данных показателей муниципальных образований Росстата. Для расчетов урожайности использовались ежегодные значения валового сбора сои и убранных посевных площадей культуры. В случае неоднородности динамических рядов предварительно осуществлялась фильтрация исходных данных статистическими методами.

2. Результаты исследования

Проведенный анализ динамических рядов индекса NDVI по маске пахотных земель показал, что изменение значений показателя в течение года соответствует нормальному распределению, и, соответственно, для аппроксимации рядов может быть использована

функция Гаусса

$$F(i) = \text{NDVI}_{\max} \exp \frac{-(i-b)^2}{2c^2}, \quad (1)$$

где i — номер недели, а b и c — неизвестные параметры гауссианы. Решение такой задачи обычно проводится нелинейным методом наименьших квадратов, в частности, в последнее время для решения подобных задач применяется алгоритм Левенберга — Марквардта [10].

На рис. 1 представлены фактические и смоделированные значения NDVI пахотных земель тестового региона в период 2016–2018 гг. По результатам выполнения алгоритма Левенберга — Марквардта:

в 2016 г. $b = 31.01 \pm 0.11$, $c = 8.07 \pm 0.10$,

в 2017 г. $b = 30.92 \pm 0.13$, $c = 8.04 \pm 0.12$,

в 2018 г. $b = 31.02 \pm 0.11$, $c = 8.06 \pm 0.10$.

Как видно из рис. 1, форма кривой, построенной по фактическим данным, вполне соответствует графику аппроксимирующей функции. Для прогнозирования максимума NDVI использовались два способа: по фактическому значению NDVI_i , где i соответствует номеру календарной недели, и по сглаженному значению, когда NDVI_i вычисляется как среднее значение вегетационных индексов ($i, i-1, \dots, i-3$) календарных недель. Максимальное значение NDVI определялось по формуле, вытекающей из (1):

$$\text{NDVI}_{\max} = \frac{\text{NDVI}_i}{\exp \frac{-(i-b)^2}{2c^2}}. \quad (2)$$

Как следует из рис. 2, прогнозирование максимума NDVI имеет смысл, начиная с 19–20-х календарных недель, что соответствует середине мая, при этом при раннем прогнозировании целесообразнее подход с использованием в (2) значения NDVI последней недели наблюдения. Начиная с 26-й календарной недели ошибка сглаженного прогноза в 2016–2018 гг. не превышала 5%.

Для дальнейшей оценки возможностей метода была рассчитана средняя абсолютная ошибка прогнозирования (в процентах), а также $\Delta \bar{x}$ максимума NDVI по маске пахотных земель в разные недели наблюдения в 2009–2018 гг. (см. таблицу). Как видно, для более точного определения максимума с 22-й недели целесообразно использовать сглаженное значение показателя NDVI. Средняя абсолютная ошибка прогнозирования

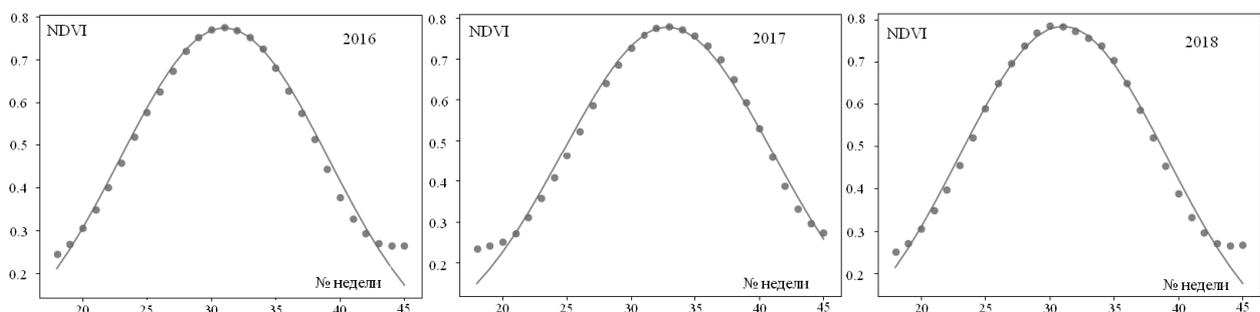


Рис. 1. Фактические (кривая) и аппроксимированные функцией Гаусса значения (точки) еженедельных композитов индекса NDVI (19–45-е недели календарного года) для пахотных земель Тамбовского р-на Амурской области в период 2016–2018 гг.

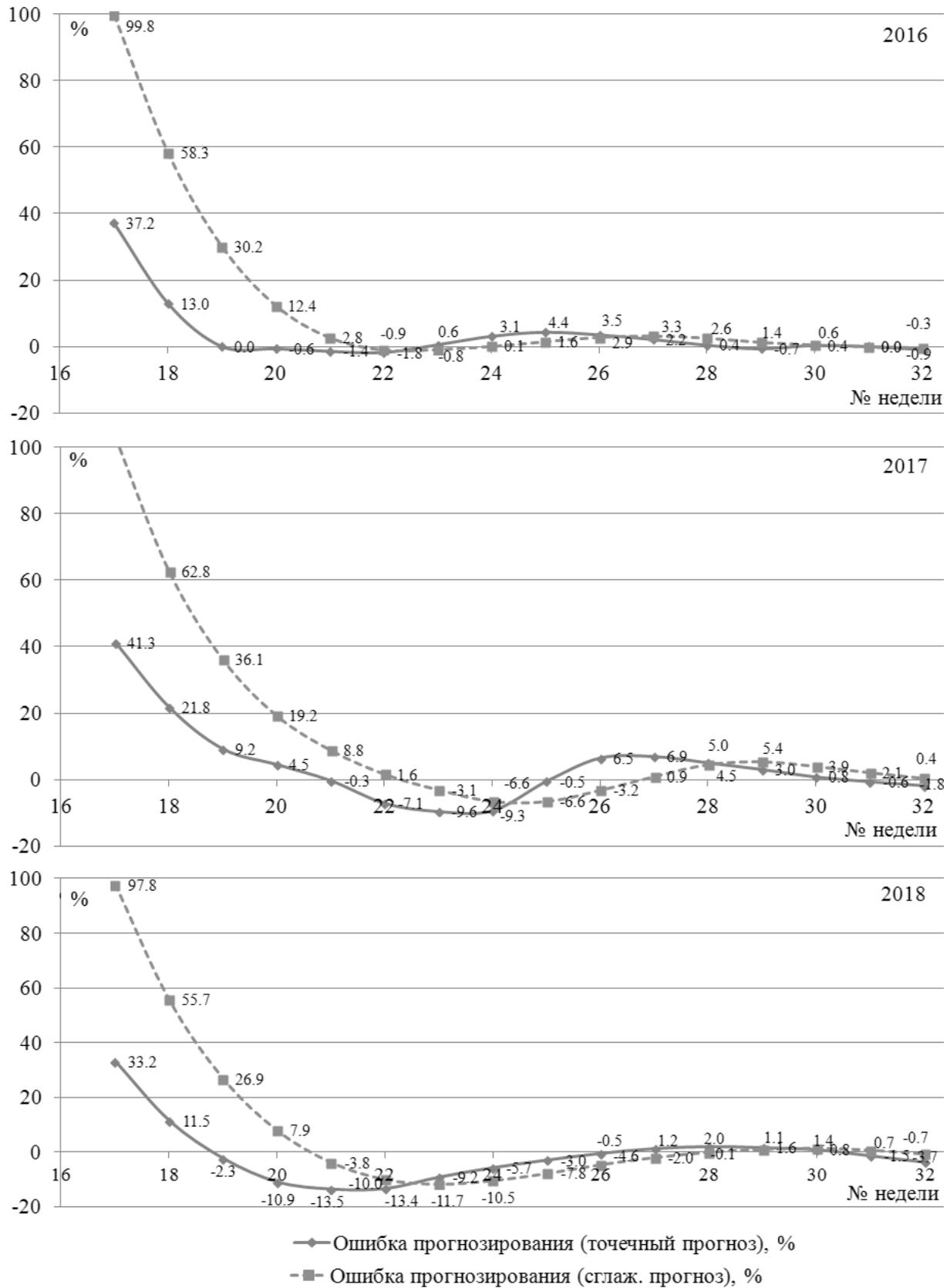


Рис. 2. Отклонение еженедельного прогнозируемого максимального значения NDVI от фактического максимума, % (2016–2018 гг., Тамбовский р-н Амурской области)

Средняя абсолютная ошибка прогнозирования максимума NDVI по маске пахотных земель в разные недели наблюдения в 2009–2018 гг. (Тамбовский р-н, Амурская обл.), %

№ недели	Сглаженный NDVI	7-дневный NDVI	№ недели	Сглаженный NDVI	7-дневный NDVI
19	32.5 ± 4.2	5.8 ± 2.6	26	5.1 ± 2.4	3.9 ± 2.3
20	15.2 ± 3.3	6.4 ± 2.9	27	3.7 ± 2.3	3.7 ± 2.4
21	5.4 ± 3.2	8.6 ± 4.3	28	3.2 ± 2.1	3.1 ± 1.5
22	6.1 ± 2.7	8.1 ± 4.3	29	3.0 ± 1.7	2.1 ± 0.7
23	6.7 ± 3.0	7.3 ± 3.5	30	2.4 ± 1.1	0.7 ± 0.2
24	6.2 ± 2.6	6.3 ± 2.4	31	1.2 ± 0.6	0.8 ± 0.9
25	6.0 ± 2.4	4.3 ± 2.3	32	0.6 ± 0.4	2.0 ± 1.5

за 10 лет в 29–32-е недели находилась в диапазоне 0.6–3.0%, в 27-ю и 28-ю неделю — 3.2–3.7%, а в 21–26-е недели — 5.1–6.7%. Ошибка точечного прогноз по данным 19-й и 20-й недель составила соответственно 5.8 ± 2.6 и 6.4 ± 2.9 %.

Для построения прогнозной модели получены данные по посевным площадям и валовому сбору сои в период с 2009 по 2018 г. в Амурской области. На следующем этапе построены две регрессионные модели, где в качестве зависимой переменной рассматривалась средняя урожайность сои (ц/га), а в качестве независимых предикторов первой модели принимались максимум NDVI и интегральные метеорологические показатели, гидротермический коэффициент и биологическая эффективность климата [11]. Во второй модели в качестве независимой переменной рассматривался только максимум NDVI. Далее для регрессионной модели с одной независимой переменной, построенной на основе данных 2009–2017 гг., оценивалась ошибка прогнозирования. В среде программирования Python были рассчитаны прогнозы урожайности сои с использованием индекса NDVI разных календарных недель 2018 г. Абсолютная ошибка прогнози-

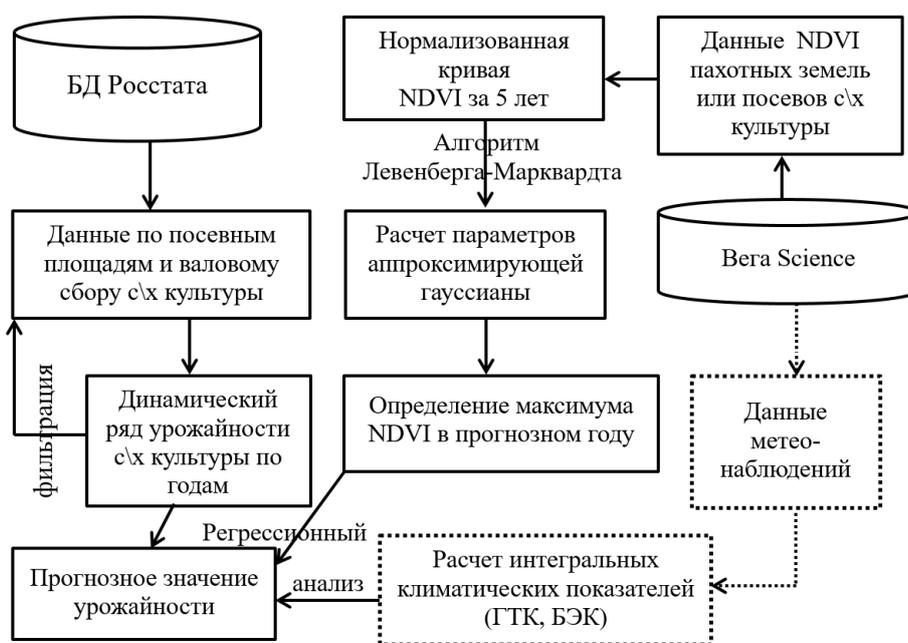


Рис. 3. Общая схема прогнозирования урожайности сельскохозяйственных культур на уровне региона

рования урожайности сои в период 29–32-е календарные недели составила 2.0–5.1 %, в 26–28-е недели — от 3.6 до 6.6 %, в 22–25-е недели — 4.9–9.1 %.

В общем виде разработанный подход к прогнозированию урожайности сельхозкультуры можно представить схемой (рис. 3). На первом этапе формируется динамический ряд урожайности культуры в исследуемом регионе за предшествующий многолетний период. Для расчетов значений динамического ряда используются данные по посевным площадям и валовому сбору культуры в регионе. В качестве источников данных могут быть использованы базы данных федерального и региональных подразделений Росстата, департаментов сельского хозяйства разных уровней, а также специально разработанные программные продукты. В случае неоднородности сформированного ряда дополнительно проводятся верификация и фильтрация данных.

На втором этапе для выбранных территорий с использованием системы Vega-Science формируются динамические ряды значений NDVI по маске пахотных полей или по маске соответствующей сельскохозяйственной культуры. По ежегодным значениям семидневных композитов NDVI рассчитываются нормализованные значения индекса за пятилетний период. Для нормализованных значений кривой с помощью алгоритма Левенберга — Марквардта проводится расчет параметров аппроксимирующей функции Гаусса. С использованием гауссианы по известным значениям NDVI с разной степенью точности определяется максимум NDVI текущего года. Окончательно для прогнозирования урожайности строится регрессионная модель. В случае моделирования с интегральными климатическими показателями исходные метеоданные могут быть также получены с использованием системы Vega-Science.

Заключение

Предложенный подход к определению урожайности сельскохозяйственной культуры продемонстрировал достаточно высокую точность, при этом метод обеспечивает возможность раннего прогнозирования. Аппроксимация динамических кривых, соответствующих значениям семидневных композитов индексов NDVI, может быть осуществлена с использованием функции Гаусса. Применение аппроксимирующей функции для прогнозирования годового максимума NDVI по маске пахотных земель в Тамбовском районе Амурской области показало хорошие результаты: средняя абсолютная ошибка прогноза в зависимости от недели прогнозирования находилась в диапазоне от 0.6 до 6.7 % в моделируемом периоде. Использование данных дистанционного зондирования Земли и разработанных программных модулей Python способствует оперативному формированию прогноза и соответственно возможности корректировки планов сельскохозяйственной отрасли. Дальнейшее развитие исследований в этом направлении с анализом NDVI по маске разных сельскохозяйственных культур позволит осуществлять прогнозирование урожайности на более раннем этапе и даст возможность не только корректировки планов, но и выполнения ряда мероприятий по повышению урожайности, таких как внесение удобрений и т. п. На следующем этапе исследований необходимо разработать интерфейс для интеграции разработанных программных модулей Python с системой Vega-Science, а также для автоматизации и обеспечения возможности использования программного продукта непосредственно специалистами в области сельского хозяйства. Результаты исследования могут быть востребованы как отдельными фермерскими организациями, так и департаментами сельского хозяйства разного уровня.

В исследовании использовались ресурсы Центра коллективного пользования “ИКИ-Мониторинг” [12] и ЦКП “Центр данных ДВО РАН” [13].

Список литературы / References

- [1] **Якушев В.П., Дубенок Н.Н., Лупян Е.А.** Опыт применения и перспективы развития технологий дистанционного зондирования Земли для сельского хозяйства // Совр. пробл. дистанц. зондирования Земли из космоса. 2019. Т. 16, № 3. С. 11–23.
Yakushev, V.P., Dubenok, N.N., Loupian, E.A. Earth remote sensing technologies for agriculture: application experience and development prospects // Current Problems in Remote Sensing of the Earth from Space. 2019. Vol. 16, No. 3. P. 11–23. (In Russ.)
- [2] **Панеш А.Х., Цалов Г.В.** Прогнозирование урожайности озимой пшеницы на основе сервисов геоинформационных систем // Вестн. АГУ. 2017. № 4. С. 175–180.
Panesh, A.Kh., Tsalov, G.V. Prediction of winter wheat productivity on the basis of geographic information system services // Vestnik AGU 2017. No. 4. P. 175–180. (In Russ.)
- [3] **Куссуль Н.Н., Кравченко А.Н., Скакун С.В. и др.** Регрессионные модели оценки урожайности сельскохозяйственных культур по данным MODIS // Совр. пробл. дистанц. зондирования Земли из космоса. 2012. Т. 9, № 1. С. 95–107.
Kussul, N.N., Kravchenko, A.N., Skakun, S.V. et al. Crop yield forecasting regression models based on MODIS data // Current Problems in Remote Sensing of the Earth from Space. 2019. Vol. 9, No. 1. P. 95–107. (In Russ.)
- [4] **Страшная А.И., Тарасова Л.Л., Богомолова Н.А. и др.** Прогнозирование урожайности зерновых и зернобобовых культур в центральных черноземных областях на основе комплексирования наземных и спутниковых данных // Тр. Гидрометеорол. науч.-исслед. центра РФ. 2015. Т. 353. С. 128–153.
Strashnaya, A.I., Tarasova, L.L., Bogomolova, N.A. et al. Forecasting of the yield for grain and leguminous crops in the Central Chernozem Regions based on integration of land and satellite data // Tr. Gidrometeorol. Nauch.-issled. Tsentra RF. 2015. Vol. 353. P. 128–153. (In Russ.)
- [5] **Савин И.Ю., Чендев Ю.Г.** Причины многолетней динамичности индекса NDVI (MODIS), осредненного для пахотных земель на уровне муниципалитетов Белгородской области // Совр. пробл. дистанц. зондирования Земли из космоса. 2018. Т. 15, № 2. С. 137–143.
Savin, I.Yu., Chendev, Yu.G. Reasons for long-term dynamics of NDVI (MODIS) averaged for arable lands of municipalities of Belgorod region // Current Problems in Remote Sensing of the Earth from Space. 2018. Vol. 15, No. 2. P. 137–143. (In Russ.)
- [6] **Буховец А.Г., Семин Е.А., Костенко Е.И., Яблоновская С.И.** Моделирование динамики вегетационного индекса NDVI озимой пшеницы в условиях ЦФО // Вестн. Воронеж. гос. аграрного ун-та. 2018. Т. 11, № 2. С. 186–199.
Bukhovets, E.G., Semin, E.A., Kostenko, E.I., Iablonovskaia, S.I. Simulation of the dynamics of the NDVI of winter wheat in the conditions of the Central Federal District // Vestnik of Voronezh State Agrarian Univ. 2018. Vol. 11, No. 2. P. 186–199. (In Russ.)
- [7] **Толпин В.А., Лупян Е.А., Барталев С.А. и др.** Возможности анализа состояния сельскохозяйственной растительности с использованием спутникового сервиса ВЕГА // Оптика атмосферы и океана. 2014. Т. 27, № 7. С. 581–586.
Tolpin, V.A., Loupian, E.A., Bartalev, S.A. et al. Possibilities of agricultural vegetation condition analysis with the VEGA satellite service // Atmospheric and Oceanic Optics. 2014. Vol. 27, No. 7. P. 581–586. (In Russ.)

- [8] **Сирант О.В.** Использование технологии Big Data для прогнозирования продаж // Технические науки — от теории к практике. 2016. № 12. С. 33–40.
Sirant, O.V. Use Big Data technology for sales forecasting // Tekhnicheskie Nauki — ot Teorii k Praktike. 2016. No. 12. P. 33–40. (In Russ.)
- [9] **Sheppard, K.** Introduction to Python for econometrics, statistics and Data Analysis. Oxford: Univ. of Oxford, 2018. 427 p.
- [10] **Gavin, H.P.** The Levenberg–Marquardt method for nonlinear least squares curve-fitting problems. Dep. Civ. Environ. Eng. Duke Univ. 2019. Available at: <http://people.duke.edu/hpgavin/ce281/lm.pdf> (accessed 23.09.2019).
- [11] **Обязов В.А., Носкова Е.В.** Многолетние изменения агроклиматических ресурсов Забайкалья // Вестн. ЗабГУ. 2015. № 8(123). С. 20–29.
Obyazov, V.A., Noskova, E.V. Long-term changes of agroclimatic resources of Transbaikalie // Transbaikal State Univ. J. 2015. No. 8(123). P. 20–29. (In Russ.)
- [12] **Лурия Е.А., Балашов И.В., Барталев С.А. и др.** Центр коллективного пользования системами архивации, обработки и анализа спутниковых данных ИКИ РАН для решения задач изучения и мониторинга окружающей среды // Совр. пробл. дистанц. зондирования Земли из космоса. 2015. Т. 12, № 5. С. 263–284.
Loupiian, E.A., Balashov, I.V., Bartalev, S.A. et al. IKI center for collective use of satellite data archiving, processing and analysis systems aimed at solving the problems of environmental study and monitoring // Current Problems in Remote Sensing of the Earth from Space. 2015. Vol. 12, No. 5. P. 263–284. (In Russ.)
- [13] **Sorokin, A.A., Makogonov, S.I., Korolev, S.P.** The information infrastructure for collective scientific work in the Far East of Russia // Sci. and Technical Inform. Proc. 2017. Vol. 4. P. 302–304.

Поступила в редакцию 3 октября 2019 г.

Forecasting of crop yields based on Earth remote sensing data (using soybeans as an example)

STEPANOV, ALEXEY S.

Far Eastern Agricultural Research Institute, Khabarovsk Territory, 680521, Russia

Corresponding author: Stepanov, Alexey S., e-mail: stepanxx@mail.ru

Purpose. Develop and describe a general approach to forecasting crop yields (using soybeans as an example).

Methodology. Crop yields were estimated using regression models. Values of the vegetative index (NDVI) were considered with Vega-Science system. The normalized NDVI values were approximated by the Gauss function using the Levenberg — Marquardt algorithm to enable early prediction with Python language.

Findings. Values of the normalized index were determined by the preceding five-years period. For normalized values, approximating Gaussians were constructed and the parameters of the Gaussian function were calculated. The maximum was predicted for the NDVI values at various calendar weeks of the simulated year. The maximum values of NDVI composites in 2009–2018 were accounted for 30–32 calendar weeks.

According to the simulation results, it was found that the average absolute error in predicting the maximum NDVI for 10 years at the weeks 29–32 did not exceed 3%, for weeks 27–28 — 4% and for the weeks 21–26 — 7%. At the next stage, a regression model was built to predict yield, where the calculated NDVI maximum was used as an independent variable, and soybean yield calculated according to the statistics of Rosstat on sown areas and gross soybean harvest in the region acted as an independent variable. Analysis of the error in predicting soybean yield for 2018 was obtained according to the simulation results of 2009–2017. It was shown that the absolute forecast error when using the data of 22–32 calendar weeks of 2018 did not exceed 9.1%.

Originality/Value. The proposed approach to determining crop yields demonstrate high accuracy, while the method provides the possibility of early forecasting. The use of Earth remote sensing data and developed software modules of Python contribute to the operational formation of the forecast and, accordingly, the possibility of adjusting the agricultural plans.

Keywords: vegetation index, Gauss function, arable land, forecasting, yield, soybeans, Python.

Cite: Stepanov, A.S. Forecasting of crop yields based on Earth remote sensing data (using soybeans as an example) // Computational Technologies. 2019. Vol. 24, No. 6. P. 125–133. (In Russ.) DOI: 10.25743/ICT.2019.24.6.015.

Received October 3, 2019