

DOI: <https://doi.org/10.37162/2618-9631-2022-4-114-129>

УДК 631.559:633.11+519.22

**Использование главных компонент  
для построения регрессионных моделей оценки  
средней районной урожайности озимой пшеницы  
для отдельных областей  
Центрально-Черноземного УГМС**

*А.Д. Клещенко, О.В. Савицкая*

*Всероссийский научно-исследовательский институт  
сельскохозяйственной метеорологии, г. Обнинск, Россия  
cxm-dir@obninsk.ru*

Показаны результаты разработки прогностических моделей на основе использования главных компонент для расчета ожидаемой урожайности озимой пшеницы для районов Центрально-Черноземного УГМС. В качестве входной информации использованы спутниковые, метеорологические и статистические данные. Спутниковые данные: индексы NDVI, VCI; метеорологические данные: средняя температура воздуха за одну и три декады, сумма осадков за одну и три декады, средний дефицит влажности воздуха за одну и три декады, гидротермический коэффициент Селянинова. Показаны результаты расчета ожидаемой урожайности озимой пшеницы для выделенных групп районов отдельных областей на территории деятельности Центрально-Черноземного УГМС. Расчеты проведены на основе интеграции спутниковой и наземной информации за период с 2012 по 2021 год по разработанным прогностическим моделям с использованием главных компонент.

*Ключевые слова:* спутниковая информация, метод главных компонент, метод проекции на латентные структуры, озимая пшеница, средняя районная урожайность

**Using the principal component analysis  
to construct regression models for estimating the average  
regional winter wheat yield for separate regions  
of the Central Chernozem Administration  
for Hydrometeorology and Environmental Monitoring**

*A.D. Kleshchenko, O.V. Savitskaya*

*All-Russian Research Institute of Agricultural Meteorology, Obninsk, Russia  
cxm-dir@obninsk.ru*

The results of the forecast model development based on the principal component analysis for calculating the expected winter wheat yield for the regions of the Central Chernozem Administration for Hydrometeorology and Environmental Monitoring (AHM) were shown. Satellite, meteorological and statistical data were used as input information. Satellite data included the NDVI and VCI indices, and meteorological data included average air temperature for one and three 10-day periods, total precipitation per one and three 10-day

periods, average air humidity deficit for one and three 10-day periods, Selyaninov hydro-thermal coefficient. The results of calculating the average winter wheat yield for the selected groups of separate regions on the territory of the Central Chernozem ANEM were presented. The calculations were based on the integration of satellite and ground-based information for the period from 2012 to 2021 and on the developed forecast models using the principal component analysis.

*Key words:* satellite data, principal component analysis, orthogonal projections to latent structures, winter wheat, average regional yield

## Введение

Пшеница является ведущей зерновой культурой во многих странах мира. Повышение ее урожайности имеет важное значение для продовольственной безопасности населения планеты в условиях увеличения его численности и сокращения ресурсов пахотных земель. Поэтому вопрос оценки ожидаемой урожайности пшеницы представляет большой научный и практический интерес, ему посвящено большое количество работ [8, 15, 10, 11]. В указанных работах представлена лишь небольшая часть имеющейся информации по указанной проблеме. Информация об ожидаемом урожае зерновых культур, и в первую очередь пшеницы, имеет большое значение для работников сельского хозяйства страны и позволяет заблаговременно определять ожидаемые ресурсы сельскохозяйственной продукции, потребность в технике для уборки, транспортировки урожая, планы импорта и экспорта. Большинство методов оценки и прогноза урожайности озимой пшеницы основано на использовании метеорологической и агрометеорологической информации, полученной на сети станций наблюдений Росгидромета. Однако с улучшением пространственного и временного разрешения данных дистанционного зондирования Земли, наблюдающемся в последние годы, стало возможным решение самых разнообразных задач в области сельского хозяйства, в том числе оценки состояния посевов и прогнозирования урожайности с применением спутниковой информации [6, 7, 9, 13]. Использование спутниковой информации позволяет перейти от точечных агрометеорологических обследований к площадному детализированному мониторингу территории.

На основе интеграции спутниковых измерений и наземных метеорологических данных в ФГБУ «ВНИИСХМ» созданы методы оценки ожидаемой урожайности по субъектам и административным районам юга России [4–6]. Результаты испытаний этих методов показали достаточно хорошую согласованность рассчитанных и фактических урожайностей озимой пшеницы. Для территории ФГБУ «Северо-Кавказское УГМС» метод оценки средней районной ожидаемой урожайности озимой пшеницы в настоящее время успешно прошел производственные испытания, которые длились два года с 2020 по 2021 год. Разработка подобных методов требуется и для других районов Европейской территории России, в частности для территории ФГБУ «Центрально-Черноземное УГМС». В данной статье

представлены результаты исследований по оценке средней районной урожайности озимой пшеницы проекционными математическими методами для указанной территории.

### Краткая характеристика агрометеорологических условий рассматриваемой территории

На юге России сосредоточены достаточно большие площади посевов озимой пшеницы (табл. 1), что создает достаточно однородный тип растительности, и это позволяет получать более надежные значения отражательных характеристик посевов озимых культур. Поэтому наблюдается достаточно высокая связь между спутниковыми индексами и урожайностью озимой пшеницы. Для этой территории характерна также высокая корреляция между урожайностью и метеорологическими параметрами [3, 6], обусловленная недостаточной влагообеспеченностью и частой засушливостью.

**Таблица 1.** Посевные площади озимой пшеницы для субъектов Северо-Кавказского УГМС и Центрально-Черноземного УГМС за 2021 год

**Table 1.** Winter wheat sown areas for the subjects of the North Caucasian UGMS and the Central Chernozem UGMS for 2021

Субъект	Посевная площадь, тыс. га	Субъект	Посевная площадь, тыс. га
<b>Северо-Кавказское УГМС</b>		<b>Центрально-Черноземное УГМС</b>	
Краснодарский	1682	Белгородская	328
Ростовская	2900	Липецкая	231
Волгоградская	1076	Курская	427
Ставропольский	1817	Воронежская	504

Центрально-Черноземная зона Российской Федерации также является важным районом интенсивного земледелия. Распаханность территории очень высокая и составляет более 80 % общей площади сельскохозяйственных угодий. Однако площади, занятые посевами озимой пшеницы, существенно меньше, чем в субъектах Северо-Кавказского УГМС (табл. 1). Справедливости ради следует заметить, что и общая площадь субъектов Северо-Кавказского УГМС существенно больше.

На территории Центрально-Черноземного УГМС в период активной вегетации озимых культур периодически также, хотя и реже чем на более южных территориях, наблюдаются засушливые условия, вызывающие снижение урожайности. Засушливые явления причиняют большой ущерб сельскому хозяйству в сельскохозяйственных районах Российской Федерации. С целью ранней диагностики засух и оценки их влияния на состояние

и продуктивность посевов сельскохозяйственных культур в ФГБУ «ВНИИСХМ» в 2002 году был создан Центр мониторинга засух (ЦМЗ). В ЦМЗ ведется разработка методов оценки атмосферной и почвенной засух (интенсивность, ареал распространения и их продолжительность). Для примера в табл. 2 представлены расчеты, полученные в ЦМЗ минимального и максимального числа декад с сильной и средней атмосферной засухой по станциям в пределах субъектов Северо-Кавказского и Центрально-Черноземного УГМС за последние 5 лет с 2018 по 2022 год.

**Таблица 2.** Число декад с сильной и средней атмосферной засухой в период активной вегетации озимых культур (с мая по вторую декаду июня)

**Table 2.** The number of decades with severe and medium atmospheric drought during the period of active vegetation of winter crops (from May to the second decade of June)

Год	2018		2019		2020		2021		2022	
Интенсивность засухи	Сильн.	Средн.	Сильн.	Средн.	Сильн.	Средн.	Сильн.	Средн.	Сильн.	Средн.
<b>Северо-Кавказское УГМС</b>										
Краснодарский	0-6*	0-4	0-4	0-3	0-2	0-3	0-1	0-2	0-2	0-2
Ставропольский	0-6	0-4	0-4	1-4	0-3	0-6	0-1	0-4	0-2	0-2
Ростовская	0-6	0-3	0-3	0-4	0-3	0-3	0-1	0-2	0-3	0-3
Волгоградская	0-4	0-3	1-6	0-3	0-2	0-3	0	0-4	0-2	0-4
<b>Центрально-Черноземное УГМС</b>										
Белгородская	0-2	0-1	0-3	0-2	0-1	0-1	0-2	0-2	0-2	0-2
Липецкая	0-2	0-1	0-1	0-5	0	0	0	0-3	0	0-1
Курская	0-1	0-1	0-1	0-3	0-1	0-1	0	0-1	0	0-1
Воронежская	0-4	0-2	1-5	0-3	0-2	0-3	0	0-2	0-1	0-2

*Примечание.* \* Диапазону значений соответствует минимальное и максимальное число декад с сильной и средней атмосферной засухой по станциям в пределах субъекта.

Как следует из таблицы, территория Центрально-Черноземного УГМС менее засушлива, чем территория Северо-Кавказского УГМС, в период активной вегетации озимых культур. Практически для всех субъектов Северо-Кавказского УГМС интенсивность, площадь охвата и продолжительность атмосферной засухи выше. Для этой территории влагообеспеченность является лимитирующим фактором и неперенным условием получения высокого урожая. Это проявляется также в более высоких коэффициентах корреляции между урожайностью озимой пшеницы и метеорологическими параметрами (в первую очередь осадками) для субъектов Северо-Кавказского УГМС по сравнению с субъектами Центрально-Черноземного УГМС, хотя и в этом случае наблюдается достаточно высокая корреляция. Следует отметить, что наибольшая урожайность

в субъектах, расположенных на территории Центрально-Черноземного УГМС, наблюдается в благоприятные годы по условиям увлажнения, а засушливые условия вызывают существенное снижение урожайности. Для примера в табл. 3 представлено снижение урожайности озимой пшеницы в беспрецедентную засуху 2010 года.

**Таблица 3.** Снижение урожайности озимой пшеницы от засухи в 2010 г.  
**Table 3.** Winter wheat yield reducing due to drought in 2010

Область	Урожайность, ц/га		Снижение урожайности, %
	Средняя за 2005–2009 гг.	2010 г.	
Белгородская	33,0	20,7	37
Липецкая	35,4	20,5	42
Курская	30,2	21,7	28
Воронежская	27,4	15,4	44

Спутниковый мониторинг является эффективным средством для оценки засух и состояния посевов сельскохозяйственных культур [2, 16]. Наиболее используемым спутниковым индексом для этих целей является так называемая нормализованная разность, представляющая собой отношение разности между значениями отражательных характеристик в ближнем инфракрасном и красном диапазонах спектра к их сумме (NDVI). Более подробно о спутниковых индексах будет сказано ниже. Значения NDVI позволяют достаточно надежно судить о складывающихся агрометеорологических условиях вегетации посевов. Например, в условиях засухи 2010 года максимальное значение NDVI для озимых культур для областей Центрально-Черноземного УГМС составило: Белгородская – 0,76, Липецкая – 0,78, Курская – 0,76, Воронежская – 0,73. В 2020 году, благоприятном по условиям увлажнения, максимальное значение NDVI для озимых культур существенно превышало значения 2010 года и соответствовало для Белгородской обл. – 0,87, Липецкой – 0,88, Курской – 0,87, Воронежской – 0,85. Однако более надежные оценки могут быть получены при комплексировании спутниковых и метеорологических данных.

### Исходные данные

В ФГБУ «ВНИИСХМ» разработан метод оценки ожидаемой среднерайонной урожайности озимой пшеницы на основе совместного использования наземной метеорологической информации и спутниковых данных для субъектов Северо-Кавказского УГМС. В качестве спутниковой информации используются данные со спектрального радиометра Modis с пространственным разрешением 250 м (сервис ВЕГА-PRO [12]). Информация

Modis получила широкое применение за счет высокого временного и достаточно высокого пространственного разрешения, непрерывной бортовой калибровки и наличия доступных многолетних архивных материалов. Это позволило использовать данные Modis для определения многих характеристик подстилающей поверхности. Для оценки состояния растительности применяются вегетационные индексы, которые представляют собой линейные или дробно-линейные комбинации значений сигналов в двух диапазонах спектра: красном и ближнем инфракрасном. В работе были использованы следующие спутниковые индексы: упомянутый выше спутниковый индекс «нормализованная разность» – NDVI (Normalized Difference Vegetation Index) и индекс условий роста растительности – VCI (Vegetation Condition Index). Эти индексы рассчитываются по следующим формулам:

$$\text{NDVI} = \frac{IK - K}{IK + K}, \quad (1)$$

где  $IK$  – отражение в ближней инфракрасной области спектра;  $K$  – отражение в красной области спектра.

$$\text{VCI}_j = \frac{\text{NDVI}_j - \text{NDVI}_{\min}}{\text{NDVI}_{\max} - \text{NDVI}_{\min}} \cdot 100\%, \quad (2)$$

где  $\text{NDVI}_j$  – значение NDVI для даты  $j$ ;  $\text{NDVI}_{\max}$  – значения максимальных NDVI внутри всего набора данных;  $\text{NDVI}_{\min}$  – значение минимальных NDVI внутри всего набора данных;  $j$  – дата.

В качестве наземной метеорологической информации использовались следующие параметры, получаемые по данным наблюдений на гидрометеорологических станциях Росгидромета: средняя декадная температура воздуха ( $T$ ), средняя температура воздуха за 3 декады ( $T3$ ), сумма осадков за декаду ( $P$ ), сумма осадков за 3 декады ( $P3$ ), среднедекадный дефицит влажности воздуха ( $D$ ), средний дефицит влажности воздуха за 3 декады ( $D3$ ). Кроме того, использовался гидротермический коэффициент Селянинова (ГТК), характеризующий соотношение тепла и влаги:

$$\text{ГТК} = \frac{\sum P}{0,1 \sum T}, \quad (3)$$

где  $\sum P$  – сумма осадков за период со среднесуточными температурами воздуха выше  $10^\circ\text{C}$ , мм;  $\sum T$  – сумма температур за тот же период.

Следует отметить, что станции метеорологической сети Росгидромета расположены неравномерно и с недостаточной плотностью, поэтому в пределах территории района часто отсутствует станция, характеризующая условия погоды этого района. Для таких случаев использовался метод обратных взвешенных квадратов расстояний [14], позволяющий получать средние районные значения метеорологических параметров для тех районов, в которых не было станций.

Статистические данные представлены значениями фактической средней районной урожайности, полученные с сайта Федеральной службы государственной статистики в базе данных показателей муниципальных образований начиная с 2012 г. [1]. Исследование проводилось за период с 2012 по 2021 год. Для увеличения объема выборки районы объединялись в группы с учетом агроклиматических условий рассматриваемых субъектов на основе карты агроклиматического районирования, разработанной Д.И. Шашко. Агроклиматическое районирование было разработано в 1967 году, поэтому для уточнения сходства и различия агроклиматических условий проводилось сопоставление средних значений урожайностей за период с 2012 по 2021 год для районов, попавших в определенную агроклиматическую зону. Таким образом в результате анализа для Белгородской, Липецкой, Курской областей все районы были объединены в пределах каждой области в одну группу. Для Воронежской области было выделено две группы (рис. 1).



**Рис. 1.** Агроклиматические группы и значения средней районной урожайности (ц/га) для Белгородской, Курской, Липецкой и Воронежской областей.

**Fig. 1.** Agroclimatic groups and average values of district yields (dt/ha) for the Belgorod, Kursk, Lipetsk and Voronezh regions.

## Результаты исследования

Метод оценки средней районной урожайности озимой пшеницы, разработанный для юга России, основан на применении проекционного математического метода (метода главных компонент), который позволяет

выделить в больших массивах данных скрытые переменные и проанализировать связи, которые существуют в изучаемой системе. Метод главных компонент (МГК) был использован для увеличения числа переменных в прогностической модели и устранения сложностей, которые вызывает мультиколлинеарность. Результаты авторских проверок метода также показали его хорошую оправдываемость [5].

Цель настоящей работы – представить результаты исследований по разработке метода оценки средней районной урожайности озимой пшеницы применительно к субъектам территории Центрально-Черноземного УГМС на основе интеграции наземных и спутниковых данных. Как показано выше, ввиду метеорологических особенностей рассматриваемых территорий прямой перенос метода, разработанного для Северо-Кавказского УГМС, не совсем корректен для территории субъектов Центрально-Черноземного УГМС.

Как было сказано ранее, для районов Северо-Кавказского УГМС наблюдаются высокие коэффициенты корреляции урожайности озимой пшеницы как со спутниковой, так и с метеорологической информацией. Для районов Центрально-Черноземного УГМС корреляция с метеорологическими параметрами в большинстве случаев несколько ниже, чем со спутниковыми данными. При использовании метода главных компонент исходные данные (метеорологические, спутниковые) проецируются в новую систему координат, образованную ортогональными осями, проведенными в направлениях наибольшей дисперсии в исходных данных. Наиболее значимыми переменными в модели становятся те показатели, которые имеют наибольшую дисперсию. Но не обязательно, что эти переменные будут иметь наибольшую связь со значениями урожайности озимой пшеницы. Для районов Северо-Кавказского УГМС это условие не столь значимо, поскольку все входные переменные имеют достаточно высокие коэффициенты корреляции с урожайностью озимой пшеницы, для районов Центрально-Черноземного УГМС у метеорологических параметров коэффициенты корреляции ниже, чем у спутниковых данных, но при этом дисперсия у них не ниже дисперсии спутниковых параметров. Поэтому в прогностической модели наибольший вес должны иметь те переменные, которые максимально коррелированы с урожайностью.

Эта задача может быть решена при использовании метода проекции на латентные структуры (ПЛС). В последние годы в различных странах появился ряд исследований, посвященных применению метода ПЛС для решения задач, связанных с оценкой продуктивности посевов [17, 18, 20]. В Китае в ряде работ описано использование метода ПЛС для оценки влияния независимых факторов (внесение удобрений, климатические факторы, питательные вещества почвы) на урожайность озимой пшеницы и определения из них наиболее важных [20]. В 2020 г. в Китае проведено интересное исследование об использовании алгоритма ПЛС для оценки урожайности озимой пшеницы на основе спутниковой информации [18].



В результате исследования были выбраны четыре спутниковых индекса (зеленый нормализованный разностный индекс растительности – GNDVI, оптимизированный индекс растительности с поправкой на почву – OSAVI, нормализованный разностный вегетационный индекс – NDVI, индекс отражения старения растений – PSRI), наиболее чувствительные для оценки ожидаемой урожайности. На основе этих индексов с использованием алгоритма ПЛС была построена многомерная модель для оценки ожидаемой урожайности. Точность оценок этой модели была выше на 20 % по сравнению с алгоритмом линейной регрессии и на 13 % по сравнению с методом главных компонент.

При построении регрессионной ПЛС-модели проводится схожее с МГК разложение исходных данных, однако в отличие от МГК-моделирования одновременно раскладываются две матрицы: исходные данные  $X$  (метеорологические, спутниковые) и данные  $Y$  (урожайность озимой пшеницы). При этом расчет новых координат (главных ПЛС-компонент) ведется таким образом, чтобы получить наибольшую корреляцию их с дисперсией в матрице  $Y$ . Максимальный вес в ПЛС-модели получают те исходные переменные, которые коррелированы с изменением значений в данных  $Y$ . Следует отметить, что перед построением моделей все переменные матрицы  $X$  центрируются на среднее и нормируются. В процессе центрирования из каждой переменной вычитается ее среднее значение, нормирование происходит путем деления на среднеквадратическое отклонение. Эти процедуры необходимы для того, чтобы все переменные имели одинаковый статистический вес в процессе построения ПЛС-модели. Подробное описание ПЛС можно найти в [19].

В настоящее время существует большое количество различных пакетов программного обеспечения, в которых реализованы алгоритмы многомерной обработки данных. В данной работе используются наиболее популярные специальные библиотеки для MATLAB, которые обладают большой гибкостью в настройках и позволяют вносить различные изменения в алгоритмы расчетов, однако требуют определенных навыков программирования.

Для примера в табл. 4 показаны коэффициенты корреляции между метеорологическими, спутниковыми параметрами и урожайностью озимой пшеницы и коэффициенты параметров в прогностической модели, полученные с использованием метода МГК и метода ПЛС для Липецкой области для третьей декады мая. Оценки получены с использованием первой компоненты.

Как следует из таблицы, наибольшие коэффициенты корреляции наблюдаются со спутниковыми индексами, в ПЛС-модели вес спутниковой информации выше, чем вес метеорологических данных. Эта многофакторная модель построена для одной главной компоненты. Выбор оптимальной размерности модели (числа ГК) является важным этапом. Метод ПЛС

является итерационной процедурой, в которой новые компоненты добавляются последовательно, одна за другой. Важно определить, в какой момент остановить этот процесс, поскольку при малом числе компонент описание данных может быть неполным, а избыточное число главных компонент может привести к некорректным оценкам. Перекрестная проверка (кросс-валидация по  $k$  блокам) является надежным статистическим методом для выбора числа главных компонент. В процессе перекрестной проверки выполняется обучение модели на подмножестве доступных данных и их оценка на другом дополнительном подмножестве.

**Таблица 4.** Коэффициенты корреляции между метеорологическими, спутниковыми параметрами и урожайностью озимой пшеницы и коэффициенты параметров в прогностической модели для Липецкой области для третьей декады мая

**Table 4.** Correlation coefficients between meteorological, satellite parameters and winter wheat yields and coefficients of parameters in the prognostic model for the Lipetsk region for the third decade of May

	NDVI	VCI	T	D	T3	D3	ГТК
Коэффициент корреляции	0,84	0,83	-0,54	-0,58	-0,53	-0,62	0,55
Коэффициенты (МГК)	0,16	0,16	-0,13	-0,15	-0,13	-0,17	0,14
Коэффициенты (ПЛС)	0,20	0,19	-0,12	-0,13	-0,12	-0,15	0,13

*Примечание.* NDVI – среднее за декаду значение вегетационного индекса; VCI – среднее за декаду значение индекса условий роста растительности; T – средняя декадная температура воздуха, °C; D – средний за декаду дефицит влажности воздуха, гПа; T3 – средняя температура воздуха за 3 декады, °C; D3 – средний дефицит влажности воздуха за 3 декады; ГТК – значение за месяц.

Оценка ошибки определяется на основе разницы в рассчитанных и фактических данных путем расчета среднеквадратической ошибки прогноза ( $\sigma$ ):

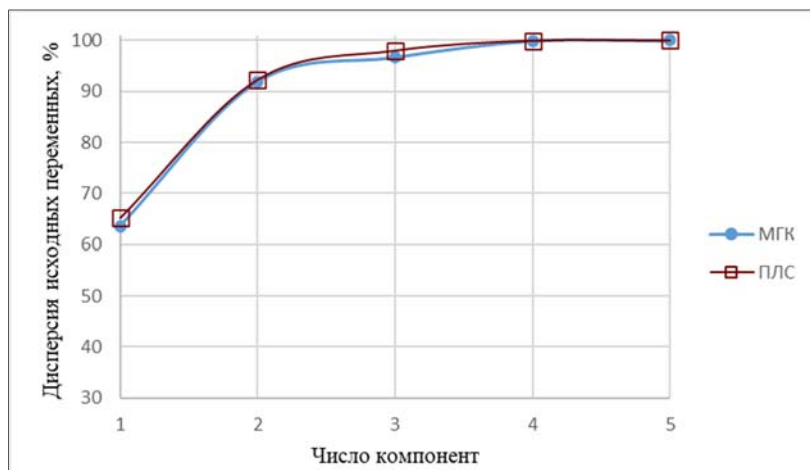
$$\sigma = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{n}}, \quad (4)$$

где  $y_i$  – фактическое значение урожайности;  $\bar{y}_i$  – рассчитанное значение урожайности;  $n$  – число наблюдений.

При этом предоставляется достаточное количество данных как для обучения, так и для проверки модели. Если в выборке выделить часть наблюдений для обучения и оставшуюся часть использовать для проверки, то уменьшение объема обучающих данных создает проблему недообучения и риска потери важных закономерностей и тенденций. Использование

метода кросс-валидация по  $k$  блокам позволяет решить эту проблему. При проверке данные делятся на  $k$  подмножеств. Каждое из  $k$  подмножеств используется в качестве проверочного набора, а другие  $k-1$  подмножеств объединяются, чтобы сформировать обучающий набор. Для получения обобщенного результата ошибка усредняется по всем  $k$  испытаниям. Каждое наблюдение попадает один раз в набор для проверки и  $k-1$  раз в обучающий набор. Таким образом увеличивается число наблюдений, используемых для обучения и проверки. Перестановка тренировочного и тестового наборов также повышает эффективность метода. В расчетах использовалась перекрестная проверка при параметре  $k$ , равном 10.

Для примера рассмотрим расчет среднеквадратической ошибки с использованием перекрестной проверки для Белгородской области для второй декады мая. Для расчетов использовались следующие параметры, коррелирующие с урожайностью озимой пшеницы: NDVI, VCI, средняя за декаду температура воздуха, средний за декаду дефицит влажности воздуха и средняя температура воздуха за 3 декады, показавшие наиболее высокую корреляцию с урожайностью. Были построены пять компонент. Среднеквадратическая ошибка существенно сокращается при использовании двух компонент. При добавлении третьей и четвертой компоненты ошибка сильно не меняется. Таким образом, из сказанного следует, что двух компонент достаточно для использования в модели. Причем две компоненты покрывают более 90 % общей дисперсии исходных данных (рис. 2).



**Рис. 2.** Изменение процента общей дисперсии исходных переменных от числа главных компонент для районов Белгородской области для второй декады мая.

**Fig. 2.** The total variance of the initial variables percentage change of the from the number of principal components for the districts of the Belgorod region for the second decade of May.

Относительная ошибка рассчитанной средней районной урожайности озимой пшеницы с 2012 по 2021 г. с использованием одной компоненты по методу МГК составляет 9,2 %, по методу ПЛС – 8,1 %; с использованием двух компонент по методу МГК – 7,6 %, по методу ПЛС – 7,4 %. В табл. 5 для примера представлена относительная ошибка рассчитанной урожайности озимой пшеницы с 2012 по 2021 г. для районов Белгородской области.

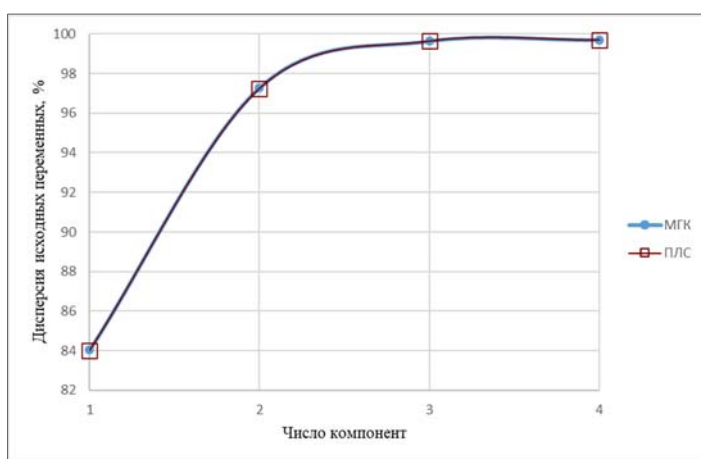
**Таблица 5.** Относительная ошибка рассчитанных значений урожайности озимой пшеницы за период с 2012 по 2021 г. для районов Белгородской области

**Table 5.** Relative error of calculated values of winter wheat yield for period from 2012 to 2021 for areas of the Belgorod region

Район Белгородской области	1 декада мая		2 декада мая		3 декада мая		1 декада июня	
	МГК	ПЛС	МГК	ПЛС	МГК	ПЛС	МГК	ПЛС
Алексеевский	8.0	7.7	6.7	6.4	8.6	7.4	10.2	10.1
Борисовский	7.8	7.1	9.9	9.3	13.6	13.3	16.3	16.1
Валуйский	6.7	5.1	6.1	5.9	8.4	7.4	9.2	9.2
Вейделевский	10.8	9.0	6.3	6.2	7.8	7.8	9.6	9.5
Волоконовский	5.4	4.2	6.4	6.0	8.5	7.8	8.6	8.4
Грайворонский	7.7	7.2	8.7	8.3	8.9	8.0	9.1	8.9
Ивнянский	7.6	7.4	8.3	8.3	6.8	6.5	8.5	8.6
Корочанский	7.4	7.8	5.5	5.0	8.3	7.6	9.3	9.1
Красненский	8.9	8.1	10.1	9.8	11.1	10.2	11.2	11.1
Красногвардейский	8.1	7.4	7.3	7.1	8.6	8.1	9.6	9.4
Краснояружский	6.0	5.5	6.2	6.1	5.0	4.8	4.5	4.5
Новооскольский	7.1	6.2	6.8	6.6	6.9	7.0	6.7	6.8
Прохоровский	8.2	8.2	8.5	8.7	8.9	8.5	8.6	8.7
Ракитянский	6.7	6.0	7.2	7.0	5.9	5.6	6.2	6.0
Чернянский	7.1	6.9	7.1	7.0	9.2	8.5	10.4	10.2
Шебекинский	6.7	6.8	9.1	8.9	10.4	9.5	8.7	8.4
Яковлевский	7.1	7.5	9.2	9.7	9.1	8.7	7.3	7.2

Как следует из табл. 5, практически во всех случаях ошибка расчетов при использовании метода проекции на латентные структуры ниже по сравнению с методом главных компонент. Аналогичные результаты получены при расчетах для всех районов рассматриваемых субъектов.

Следует отметить, что достаточно редко для отдельных групп районов и в определенную декаду коэффициенты корреляции со спутниковыми индексами не превышали коэффициентов корреляции с отдельными метеорологическими параметрами. В таких случаях точность расчетов по МГК и ПЛС приблизительно одинакова. Например, для второй группы районов Воронежской области (15 районов) в 3 декаде мая коэффициенты корреляции с NDVI, VCI составляли 0,66, со средней температурой воздуха за 3 декады – -0,66, со средним дефицитом влажности воздуха за 3 декады – 0,69. В результате расчетов по МГК и ПЛС были получены четыре компоненты. На рис. 3 показан график изменения процента общей дисперсии исходных переменных от числа главных компонент.



**Рис. 3.** Изменение процента общей дисперсии исходных переменных от числа главных компонент для районов Воронежской области для 3 декады мая.

**Fig. 3.** The total variance of the original variables percentage change of the from the number of principal variables component for districts of the Voronezh region for the 3 decade of May.

Из рисунка видно, что компоненты, построенные по двум методам, объясняют практически одинаковый процент дисперсии исходных данных. Первая компонента объясняет 84 % дисперсии, вторая компонента – 13 % дисперсии, что в сумме с первой компонентой составляет 98 %. Для выбора числа компонент использовался метод перекрестной проверки. В результате расчета было получено, что с увеличением числа компонент средне-квадратическая ошибка не уменьшается, следовательно, для этого набора параметров для расчета ожидаемой урожайности достаточно использовать одну компоненту. Для рассматриваемого случая относительная ошибка рассчитанной урожайности озимой пшеницы с 2012 по 2021 год по методу МГК и методу ПЛС совпадает и составляет 11,5 %.

## Заключение

В результате проведенного исследования были получены прогностические модели на основе метода проекции на латентные структуры для расчета ожидаемой урожайности озимой пшеницы для районов Центрально-Черноземного УГМС. Проверка полученных моделей показала достаточно надежную их работу. Проведена оценка близости расчетов модели к фактическим данным. Относительная ошибка расчетов между фактической и рассчитанной урожайностью озимой пшеницы составила 7–14,9 %. Практически во всех случаях ошибка расчетов при использовании метода ПЛС ниже по сравнению с корреляционно-регрессионным методом и методом главных компонент. Предложенный метод может иметь практическое применение в оперативной работе прогнозистов-агрометеорологов Центрально-Черноземного УГМС после проведения производственных испытаний.

## Список литературы

1. База данных показателей муниципальных образований. Федеральная служба государственной статистики. <https://www.gks.ru/dbscripts/munst>.
2. Клещенко А.Д., Асмус В.В., Страшная А.И., Кровотынцев В.А., Вирченко О.В., Савицкая О.В., Береза О.В., Василенко Е.В., Сухарева В.В., Моргунов Ю.А., Косякин С.А. Мониторинг засух по данным наземной и спутниковой информации // Метеорология и гидрология. 2019. № 11. С. 95-108.
3. Клещенко А.Д., Вирченко О.В., Савицкая О.В. Спутниковый мониторинг состояния и продуктивности посевов зерновых культур // Труды ВНИИСХМ. 2013. Вып. 38. С. 54-70.
4. Клещенко А.Д., Савицкая О.В. Оценка пространственно-временного распределения урожайности зерновых культур и стандартизованного индекса осадков (SPI) по спутниковой и наземной информации // Труды ГГО. 2014. Вып. 571. С. 147-161.
5. Клещенко А.Д., Савицкая О.В. Оценка урожайности озимой пшеницы с использованием метода главных компонент на основе комплексирования спутниковой и наземной информации // Метеорология и гидрология. 2021. № 12. С. 127-136.
6. Клещенко А.Д., Савицкая О.В., Косякин С.А. Оценка средней районной урожайности озимой пшеницы по спутниковой и наземной метеорологической информации // Гидрометеорологические исследования и прогнозы. 2020. № 3 (377). С. 103-121.
7. Кононенко С. М., Старостина Т. В. Спутниковые данные в прогнозе урожайности пшеницы для административных районов Сибири // Интерэкспо Гео-Сибирь. 2018. Т. 1, № 4. С. 94-101.
8. Лебедева В.М., Страшная А.И. Основы сельскохозяйственной метеорологии, Т. 2. Методы расчетов и прогнозов в агрометеорологии. Обнинск: ФГБУ «ВНИИГМИ-МЦД», 2012. 216 с.
9. Лузян Е.А, Барталев С.А., Крашенинникова Ю.С., Плотников Д.Е., Толпин В.А., Уваров И.А. Анализ развития озимых культур в южных регионах европейской части России весной 2018 года на основе данных дистанционного мониторинга // Современные проблемы дистанционного зондирования Земли из космоса. 2018. Т. 15, № 2. С. 272-276.
10. Моисейчик В.А., Богомолова Н.А. Возможности физико-математических методов долгосрочных агрометеорологических прогнозов состояния озимых зерновых культур к весне по субъектам Российской Федерации и в целом по Российской Федерации в условиях глобального изменения климата // Труды ВНИИСХМ. 2007. Вып. 36. С. 133-149.
11. Полевой А.Н. Прикладное моделирование и прогнозирование продуктивности посевов. Л.: Гидрометеиздат, 1988. 320 с.
12. Спутниковый сервис анализа вегетации. – URL: <http://pro-vega.ru>. (дата обращения: 2022-09-23).

13. Страшина А.И., Тарасова Л.Л., Богомолова Н.А., Максименкова Т.А., Береза О.В. Прогнозирование урожайности зерновых и зернобобовых культур в центральных черноземных областях на основе комплексирования наземных и спутниковых данных // Труды Гидрометцентра России. 2015. Вып. 353. С. 128-153.

14. Ткачева Ю.В. Методика интерполяции кусочно-линейных данных об эмиссиях автотранспорта на регулярную модельную сетку // Труды Гидрометцентра России. 2018. Вып. 368. С. 170-180.

15. Уланова Е.С. Агрометеорологические условия и урожайность озимой пшеницы. Л: Гидрометеиздат, 1975. 302 с.

16. Щербенко Е.В. Дистанционные методы выявления сельскохозяйственной засухи // Современные проблемы дистанционного зондирования Земли из космоса. 2007. Т. 2. С. 409-419.

17. Elfatih M. Abdel-Rahman, Onesimo Mutanga, John Odindi, Elhadi Adam, Alfred Odindo, Riyad Ismail. A comparison of partial least squares (PLS) and sparse PLS regressions for predicting yield of Swiss chard grown under different irrigation water sources using hyperspectral data // Computers and Electronics in Agriculture. 2014. Vol. 106. P. 11-19.

18. Peng-PengZhang<sup>1</sup>, Xin-XingZhou<sup>1</sup>, Zhi-XiangWang, Wei Mao, Wen-Xi Li, FeiYun, Wen-ShanGuo, Chang-WeiTan Using HJ-CCD image and PLS algorithm to estimate the yield of field-grown winter wheat // Scientific Reports. 2020. Vol. 10, № 5173. P. 1-10.

19. Wold S., Sjostrom M., Eriksson L. PLS-regression: a basic tool of chemometrics // Chemom. Intell. Lab. Syst. 2001. Vol. 58. P. 109-130.

20. Yutong Hu, Xiaorong Wei, Mingde Hao, Wei Fu. Partial Least Squares Regression for Determining Factors Controlling Winter Wheat Yield // Agronomy Journal. 2017. Vol. 110. P. 281-292.

## References

1. Baza dannykh pokazatelei munitsipal'nykh obrazovaniy. Federal'noi sluzhby gosudarstvennoi statistiki, Available at: <https://www.gks.ru/dbscripts/munst> [in Russ.].

2. Kleshchenko A.D., Virchenko O.V., Savitskaya O.V., Sukhareva V.V., Morgunov Y.A., Kosyakin S.A., Asmus V.V., Krovotyntsev V.A., Vasilenko E.V., Strashnaya A.I., Bereza O.V. Drought monitoring based on ground and satellite data. *Russ. Meteorol. Hydrol.*, 2019, vol. 44, no. 11, pp. 772-781. DOI: 10.3103/S1068373919110074.

3. Kleshchenko A.D., Virchenko O.V., Savitskaya O.V. Satellite monitoring of the condition and productivity of grain crops. *Trudy FGBU «VNIISXM» [Proceedings of All-Russian Research Institute of Agricultural Meteorology]*, Obninsk, 2013, pp. 54-70 [in Russ.].

4. Kleshchenko A.D. Savitskaya O.V. Estimation of space-time distribution yield of grains and standardized precipitation index on satellite and land information. *Trudy GGO [Proceedings of Voeikov Main Geophysical Observatory]*, 2014, vol. 571, pp. 147-161 [in Russ.].

5. Kleshchenko A.D., Savitskaya O.V. Estimation of Winter Wheat Yield Using the Principal Component Analysis Based on the Integration of Satellite and Ground Information. *Russ. Meteorol. Hydrol.*, 2021, vol. 46, no. 12, pp. 881-887. DOI: 10.3103/S1068373921120104

6. Kleshchenko A.D., Savitskaya O.V., Kosyakin S.A. Estimation of the average regional yield of winter wheat according to satellite and ground meteorological information. *Gidrometeorologicheskie issledovaniya i prognozy [Hydrometeorological research and forecasts]*, 2020, no. 3, pp. 103-121 [in Russ.].

7. Kononenko S. M., Starostina T. V. Satellite data in the forecast of wheat yield for the administrative regions of Siberia. *Interexpo Geo-Sibir' [Interexpo Geo-Siberia]*, 2018, vol. 1, no 4, pp. 94-101 [in Russ.].

8. Lebedeva V.M., Strashnaya A.I. Osnovy sel'skokhozyaystvennoy meteorologii, Metody raschetov i prognozy v agrometeorologii. [Fundamentals of agricultural meteorology, Methods of calculations and forecasts in agrometeorology]. Obninsk, FGBU «VNIIGMI-MTsD» Publ., 2012, vol. 2, 216 p. [in Russ.].

9. Lupyay E.A., Bartalev S. A., Krashennnikova Yu. S., Plotnikov D. E., Tolpin V. A., Uvarov I. A. Analysis of winter crops development in the southern regions of the European part of Russia in spring of 2018 with use of remote monitoring. *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa [Current problems in remote sensing of the Earth from space]*, 2018, vol. 15, no 2, pp. 272-276 [in Russ.].

10. Moiseychik V.A., Bogomolova N.A. Possibilities of physical and mathematical methods of long-term agrometeorological forecasts of the state of winter crops by spring in the constituent entities of the Russian Federation and in the Russian Federation as a whole in the context of global climate change. *Trudy GU «VNIISXM» [Proceedings of All-Russian Research Institute of Agricultural Meteorology]*, 2007, no. 36, pp. 133-149 [in Russ.].

11. Polevoy A.N. Prikladnoe modelirovanie i prognozirovanie produktivnosti posevov [Applied modeling and forecasting of crop productivity]. Leningrad, Gidrometeoizdat Publ., 1988, 320 p. [in Russ.].

12. Sputnikovyi servis analiza vegetatsii. Available at: <http://pro-vega.ru> [in Russ.].

13. Strashnaya A.I., Tarasova L.L., Bogomolova N.A., Maksimenkova T.A., Bereza O.V. Forecasting of the yield of grain and leguminous crops in the Central Chernozem Regions based on integration of land and satellite data. *Trudy Gidromettsentra Rossii [Proceedings of Hydrometcenter of Russia]*, 2015, no 353, pp. 128-153 [in Russ.].

14. Tkacheva Yu.V. Methodology of interpolating piecewise linear data on vehicle emissions on a regular model grid. *Trudy Gidromettsentra Rossii [Proceedings of Hydrometcenter of Russia]*, 2018, no 368, pp.170-180 [in Russ.].

15. Ulanova E.S. Agrometeorologicheskie usloviya i urozhaynost' ozimoy pshenitsy [Agrometeorological conditions and productivity of winter wheat]. Leningrad, Gidrometeoizdat Publ., 1975, 302 p. [in Russ.].

16. Shcherbenko E.V. Remote methods for detecting agricultural drought. *Sovremennyye problemy distantsionnogo zondirovaniya Zemli iz kosmosa [Current problems in remote sensing of the Earth from space]*, 2007, vol. 2, pp. 409-419 [in Russ.].

17. Elfatih M. Abdel-Rahman, Onesimo Mutanga, John Odindi, Elhadi Adam, Alfred Odindo, Riyad Ismail. A comparison of partial least squares (PLS) and sparse PLS regressions for predicting yield of Swiss chard grown under different irrigation water sources using hyperspectral data. *Computers and Electronics in Agriculture*, 2014, vol. 106, pp. 11-19.

18. Peng-PengZhangl, Xin-XingZhoul, Zhi-XiangWang, Wei Mao, Wen-Xi Li, FeiYun, Wen-ShanGuo, Chang-WeiTan Using HJ-CCD image and PLS algorithm to estimate the yield of field-grown winter wheat. *Scientific Reports*, 2020, vol. 10, no. 5173, pp. 1-10.

19. Wold S., Sjostrom M., Eriksson L. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.*, 2001, vol. 58, pp. 109-130.

20. Yutong Hu, Xiaorong Wei, Mingde Hao, Wei Fu. Partial Least Squares Regression for Determining Factors Controlling Winter Wheat Yield. *Agronomy Journal*, 2017, vol. 110, pp. 281-292.

*Поступила 06.10.2022; одобрена после рецензирования 16.11.2022;  
принята в печать 29.11.2022.*

*Submitted 06.10.2022; approved after reviewing 16.11.2022;  
accepted for publication 29.11.2022.*