

УДК 528.8.04:004.021

## ПРОГНОЗИРОВАНИЕ УРОЖАЙНОСТИ ОЗИМОЙ ПШЕНИЦЫ ПО РАЗЛИЧНЫМ СПУТНИКОВЫМ ДАННЫМ

А.В. Колотий

*Институт космических исследований НАНУ-ГКАУ,  
Национальный университет биоресурсов и природопользования,  
Национальный технический университет Украины «КПИ»*

*andrew.k.911@gmail.com*

В статті проводиться порівняльний аналіз використання різноманітних супутникових даних для прогнозування врожайності озимої пшениці в Україні на рівні окремих областей. В якості предикторів використовуються нормалізований різницевий вегетаційний індекс NDVI, індекс здоров'я рослинності VHI та продукт FAPAR, що характеризується часткою фотосинтетично активної сонячної радіації, поглинутої рослинністю.

*Ключові слова: прогноз врожайності, регресійна модель, інформаційна технологія, MODIS, NDVI, FAPAR, VHI.*

This article aims at the estimation of the results of usage various satellite data to predict winter wheat yield in Ukraine at the level of administrative regions (oblast). As the predictors normalized difference vegetation index (NDVI), vegetation health index (VHI) and fraction of absorbed photosynthetically active radiation index (FAPAR) are used.

*Keywords: yield, the regression model, information technology, MODIS, NDVI.*

В статье проводится сравнительный анализ использования различных спутниковых данных для прогнозирования урожайности озимой пшеницы в Украине на уровне отдельных областей. В качестве предикторов используются нормализованный разностный вегетационный индекс NDVI, индекс здоровья растительности VHI и продукт FAPAR, характеризующий долю фотосинтетически активной солнечной радиации, поглощенной растительностью.

*Ключевые слова: прогноз урожайности, регрессионная модель, информационная технология, MODIS, NDVI, FAPAR, VHI.*

### Введение

В свете проблем продовольственной безопасности задача прогнозирования урожайности является актуальной не только для Украины, как одного из крупнейших производителей зерна, но и для международного сообщества [1, 2, 3]. Современное состояние развития систем дистанционного зондирования Земли (ДЗЗ) из космоса позволяет рассматривать спутниковые данные в качестве источника оперативной и объективной информации о состоянии окружающей среды в целом и на регулярной основе получать информацию о значительных по площади территории.

Анализируя публикации, связанные с прогнозированием урожайности озимой пшеницы в Украине, следует констатировать факт, что большинство работ близкой тематики направлены на прогнозирование урожайности для всей территории страны [3] или для отдельных районов [4]. Проблеме прогнозирования урожайности на уровне отдельных областей, которые

являются основными единицами административного деления, уделяется меньше внимания [5, 6, 7]. В работах [5, 6, 7] авторами предложен метод прогнозирования урожайности озимой пшеницы с использованием вегетационного индекса NDVI на уровне областей Украины, проведено сравнительное исследование биофизического и регрессионного подхода для решения задачи прогнозирования урожайности озимой пшеницы. Построены региональные регрессионные модели и разработаны автоматизированные информационные технологии для прогнозирования урожайности на всей территории Украины, а также для каждой области отдельно. Полученные результаты сопоставлены с данными официальной статистики по данным Государственной службы статистики Украины.

## **1. Постановка задачи**

Для оценки достоверности прогнозных моделей, построенных на основе временных рядов индекса NDVI, полученного из продукта MOD13Q1.005 [9], в данной статье предлагаются альтернативные регрессионные модели, использующие в качестве предикторов временные ряды индекса здоровья растительности VHI для территории Украины, полученные с помощью прибора AVHRR [10], а также продуктов FAPAR, полученных с использованием данных SPOT Vegetation [11]. Кроме того, хотя индекс NDVI и связан со степенью развития биомассы, однако его физическая интерпретация не является достаточно прозрачной, поэтому актуальной является задача оценки эффективности использования в задаче прогнозирования урожайности вегетационных индексов и продуктов, обладающим более выраженным физическим смыслом (таких как VHI и FAPAR).

## **2. Используемые данные**

Ниже рассмотрим более подробно вышеупомянутые вегетационные индексы, которые будут использованы для проверки полученных ранее результатов [5, 6, 7].

NDVI (полученный из продукта MOD13) – нормализованный разностный индекс растительности, один из наиболее распространенных и используемых индексов для решения задач, использующих количественные оценки растительного покрова.

Рассчитывается по формуле:

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

где NIR - отражение в ближней инфракрасной области спектра, RED - отражение в красной области спектра. Согласно (1) плотность растительности (NDVI) в заданной точке изображения равна разности интенсивностей отраженного света в красном и инфракрасном диапазоне, деленной на сумму их интенсивностей.

VHI (полученный с помощью AVHRR – Advance Very High Resolution Radiometer) – вегетационный индекс, основанный на отражении видимого света растительным покровом, характеризующем здоровье культур. Данный индекс базируется на сочетании индекса VCI (Vegetation Condition Index), который характеризует угнетённость растительного покрова и индекса температурного режима TCI (Temperature Condition Index), которые были предложены в 1995 году Ф. Коганом:

$$VHI = 0.5 * VCI + 0.5 * TCI, \quad (2)$$

$$VCI = 100 * (NDVI - NDVI_{\min}) / (NDVI_{\max} - NDVI_{\min}) \quad (3)$$

$$TCI = 100 * (BT_{\max} - BT) / (BT_{\max} - BT_{\min}) \quad (4)$$

где  $BT$ ,  $BT_{\max}$ ,  $BT_{\min}$  – усредненные сезонные значения яркостной температуры, ее абсолютный максимум и минимум соответственно,  $NDVI$ ,  $NDVI_{\max}$ ,  $NDVI_{\min}$  – усредненные значения индекса NDVI [12].

Более подробно применение вегетационного индекса NDVI и индекса здоровья VHI для прогнозирования урожайности описано в работах [2, 3]

Продукт FAPAR (получен с помощью Spot Vegetation) – он содержит результаты измерения поглощенной фотосинтетической радиации и фракции абсорбирования, и выступает в качестве интегрального показателя состояния и здоровья растительного покрова. Данный продукт играет важную роль при определении первичной продуктивности фитосферы, и может быть использован для определения количественных показателей растительности [13].

Нередко измерения FAPAR происходят во время ясной или малооблачной погоды, что усложняет получение усредненных данных по указанному продукту, рекомендуется использовать измерения, сделанные в пасмурную погоду. FAPAR определяется долей, присущей фотосинтезу активной радиации (ФАР), поглощаемой листовым покровом, а иногда и почвой и определяется следующим уравнением:

$$FAPAR = \frac{(PAR_{\downarrow AC} - PAR_{\uparrow AC})(PAR_{\downarrow BC} - PAR_{\uparrow BC})}{PAR_{\downarrow AC}} \quad (5),$$

где  $PAR_{\downarrow AC}$  и  $PAR_{\uparrow AC}$  - поступающий (сверху вниз) и отраженный (снизу вверх) ФАР через лиственный покров, и  $PAR_{\downarrow BC}$  и  $PAR_{\uparrow BC}$  - соответствующие условия под лиственным покровом. Из формулы (5) можно получить соотношение (6) между отраженным ФАР под поверхностью листвы ( $\rho_{BC}$ ) и над нею ( $\rho_{AC}$ ):

$$FAPAR = [1 - \rho_{AC}(t)] - [1 - \rho_{BC}(t)](PAR_{\downarrow BC} / PAR_{\downarrow AC}) \quad (6),$$

где соотношение  $PAR_{\downarrow BC} / PAR_{\downarrow AC} = 1 - FIPAR$ ,  $FIPAR$  - часть ФАР, которая была задержана лиственным покровом. Во многих измерениях  $FIPAR$  используют вместо FAPAR, так как этот продукт легче измеряется и он является полным аналогом продукта FAPAR.

### 3. Построение регрессионной модели

Были построены линейные регрессионные модели для каждой области отдельно. В качестве предиктора использовались значения индекса NDVI, усредненные на уровне областей по маске посевных территорий для каждого 16-дневного композита, полученные из продукта MOD13; еженедельные значения индекса VHI; значения индекса FAPAR, усредненные на уровне области для каждого еженедельного (VHI) или 10-дневного (FAPAR) композита по маске посевных территорий.

Отметим, что за последнее десятилетие для урожайности озимой пшеницы наблюдается положительный линейный тренд для всех областей в связи с улучшением сельскохозяйственных технологий [2]. Урожайность будем аппроксимировать следующим образом:

$$Y_i = T_i + dY_i, \quad (7)$$

где  $Y_i$  – урожайность в год  $i$ ,  $T_i$  – трендовый компонент, который связан с сельскохозяйственными технологиями,  $dY_i$  – случайный компонент, который обусловлен метеорологическими условиями и состоянием посевов в текущем году  $i$ .

Трендовый компонент аппроксимировался с помощью линейной регрессии [14]:

$$T_i = a_0 + a_1 * i, \quad (8)$$

где  $i$  – год.

Таким образом, регрессионная модель построена для отклонения урожайности от тренда со значением вегетационного индекса NDVI, FAPAR и VHI:

$$\begin{aligned} dY_i^{NDVI} &= Y_i - T_i = f(NDVI_i) = b_0 + b_1 * NDVI_i \\ dY_i^{VHI} &= Y_i - T_i = g(VHI_i) = k_0 + k_1 * VHI_i \\ dY_i^{FAPAR} &= Y_i - T_i = h(FAPAR_i) = h_0 + h_1 * FAPAR_i \end{aligned} \quad (9)$$

где  $NDVI_i$  – 16-дневный композит NDVI прибора MODIS за некоторый период года  $i$ ,  $VHI_i$  – еженедельные данные с AVHRR,  $FAPAR_i$  – 10-дневные композиты, построенные по данным SPOT Vegetation.

Для обеспечения качества регрессионных моделей (минимизации ошибки прогноза) будет использована процедура кросс-валидации (leave-one-out cross-validation - *LOOCV*) [14] – в виду небольшого объема выборки, используемой для обучения регрессионных моделей. Для каждой области строится модель на основе данных за все года, кроме одного. На основе построенной модели выполняется прогноз на текущий год, который сравнивается с фактическим значением урожайности по данным официальной статистики.

В рамках этой процедуры, в качестве предиктора в модели по временным рядам VHI и FAPAR, будем использовать те значения продуктов, которые обеспечивают минимальное значение среднеквадратической ошибки *RMSE* [15]:

$$RMSE = \sqrt{\frac{1}{n} \sum_i (P_i - O_i)^2} \quad (10)$$

где  $P_i$  и  $O_i$  – прогнозируемое и наблюдаемое (по данным статистики) значения урожайности озимой пшеницы соответственно,  $n$  — число лет, данные которых используются для построения модели (например, если для идентификации параметров модели используются данные за 2000-2012 годы, то  $n=13$ ).

В таблице 1 представлены результаты поиска дней года для VHI, FAPAR и NDVI, которые обеспечивают минимальное значение среднеквадратической ошибки *RMSE* при обучении на данных 2000-2012 гг., приведенные к временной шкале данных NDVI.

Таблица 1.

Дни, обеспечивающие минимальные ошибки прогноза для VHI, FAPAR и NDVI при обучении на данных 2000-2012 гг.

<i>Область</i>	<i>FAPAR</i>	<i>NDVI</i>	<i>VHI</i>
<i>Винницкая</i>	129	97	145
<i>Волынская</i>	129	145	113
<i>Днепропетровская</i>	145	145	161
<i>Донецкая</i>	129	129	113
<i>Житомирская</i>	129	97	145
<i>Закарпатская</i>	97	113	129
<i>Запорожская</i>	145	129	145
<i>Ивано-Франковская</i>	97	145	145
<i>Киевская</i>	97	97	113
<i>Кировоградская</i>	161	129	161
<i>Луганская</i>	129	129	113
<i>Львовская</i>	129	113	145
<i>Николаевская</i>	145	129	145
<i>Одесская</i>	145	129	145
<i>Полтавская</i>	161	129	161
<i>Ровенская</i>	113	97	145
<i>Сумская</i>	161	129	113
<i>Тернопольская</i>	129	113	113
<i>Харьковская</i>	129	129	113
<i>Херсонская</i>	145	129	145
<i>Хмельницкая</i>	113	113	145
<i>Черкасская</i>	161	129	113
<i>Черновицкая</i>	113	113	145
<i>Черниговская</i>	97	97	97
<i>Республика Крым</i>	145	129	129

Как показали результаты проведенного исследования, при использовании различных источников информации время для построения прогноза выпадает на конец апреля – начало мая (113-й и 129-й день года для рядов FAPAR, преимущественно 113-й для рядов VHI, 97-й, 113-й или 129-й дни для рядов NDVI).



Следует отметить, что довольно высокий коэффициент детерминации моделей, как и низкий уровень ошибок прогноза, наблюдается в основном у областей степной и частично лесостепной агроклиматических зон Украины. Эти области являются основными поставщиками озимой пшеницы и среди их посевных площадей данная культура занимает наибольшую часть (в сравнении с прочими областями).

В таблице 2 приведены ошибки моделей, построенных на данных NDVI, VHI и FAPAR, оценивается среднеквадратическое отклонение по областям Украины за 2010-2012 гг, а также приводится средняя величина недооценки или переоценки.

Исследования показали, что величина ошибок (RMSE и средней) моделей, построенных по рядам NDVI, VHI, FAPAR не позволяет однозначно выделить наилучший источник данных. Для ряда лет наименьшие ошибки прогноза дает VHI, для ряда – FAPAR. Однако, учитывая небольшое объем данных, нельзя утверждать, что NDVI всегда обеспечивает менее точный прогноз.

Таблица 2.

Ошибки моделей, обученных на различных данных

		<b>2010</b>	<b>2011</b>	<b>2012</b>
<b>NDVI</b>	<b>RMSE</b>	8,2	6,2	6,8
	<b>average</b>	6,8	-3,7	-3,4
<b>VHI</b>	<b>RMSE</b>	<b>6,3</b>	<b>5,1</b>	7,0
	<b>average</b>	5,5	-3,8	<b>-3,6</b>
<b>FAPAR</b>	<b>RMSE</b>	8,88	5,18	<b>5,64</b>
	<b>average</b>	7,56	<b>-2,14</b>	<b>-0,49</b>

## 5. Выводы

В работе построены линейные регрессионные модели для прогнозирования урожайности озимой пшеницы в Украине на основе разнородных спутниковых продуктов. Разработанные алгоритмы могут быть реализованы с помощью высокопроизводительных систем [16, 17, 18]. В качестве предикторов использованы вегетационные индексы с разной степенью «физичности», вычисленные по данным различных спутников. А именно, регрессорами предложенных моделей являются индекс NDVI, вычисляемый по данным прибора MODIS, индекс «здоровья растительности» VHI, построенный на основе данных прибора AVHRR, и продукт FAPAR, имеющий наибольший физический смысл и получаемый на основе данных SPOT VEGETATION. Проведено сравнение ошибок прогноза при использовании различных предикторов: NDVI, VHI, FAPAR. Анализ результатов показал, что

использование индекса NDVI дает большую ошибку прогноза, чем FAPAR и VHI, но ограниченный объем данных не позволяет утверждать, что NDVI всегда будет обеспечивать менее точный прогноз. Поэтому авторам кажется целесообразным не ограничиваться одной прогнозной моделью, а использовать ансамблевый подход на основе предикторов различной природы.

## Литература

1. Галлего Х. Анализ эффективности различных подходов для классификации посевов на основе спутниковой и наземной информации / Галлего Х., Кравченко А.Н., Куссуль Н.Н., Скакун С.В., Шелестов А.Ю., Грипич Ю.В. // Проблемы управления и информатики. – 2012. – №3. – С. 123-134.
2. Kogan F. Forecasting crop production using satellite-based vegetation health indices in Kansas, USA / Kogan F., Salazar L., Roytman L. // International Journal of Remote Sensing. — 2012. — 33, N 9. — P. 2798–2814.
3. Becker-Reshef I. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data / I. Becker-Reshef, E. Vermote, M. Lindeman, C. Justice // Remote Sensing of Environment. — 2010. — 114, N 6. — P. 1312–1323.
4. Куссуль Н.Н. Регрессионные модели оценки урожайности сельскохозяйственных культур по данным MODIS / Н.Н. Куссуль, А.Н. Кравченко, С.В. Скакун, Т.И. Адаменко, А.Ю. Шелестов, А.В. Колотий, Ю.А. Грипич // Сборник научных статей "Современные проблемы дистанционного зондирования Земли из космоса". — 2012. — Том 9, №1. — С. 95–107.
5. Kogan F. Winter wheat yield forecasting in Ukraine based on Earth observation, meteorological data and biophysical models / F. Kogan, N. Kussul, T. Adamenko, S. Skakun, O. Kravchenko, O. Kryvobok, A. Shelestov, A. Kolotii, O. Kussul, A. Lavrenyuk // International Journal of Applied Earth Observation and Geoinformation — 2013. — vol. 23 — PP. 192-203.
6. Kogan, F. Winter wheat yield forecasting: A comparative analysis of results of regression and biophysical models / Kogan, F., Kussul, N.N., Adamenko, T.I., Skakun, S.V, Kravchenko, A.N., Krivobok, A.A., Shelestov, A.Yu., Kolotii, A.V., Kussul, O.M., Lavrenyuk, A.N. // (2013) Journal of Automation and Information Sciences 45 (6) PP. 68 – 81
7. Коган Ф. Прогнозирование урожайности озимой пшеницы: сравнительный анализ результатов регрессионных и биофизических моделей / Коган Ф., Куссуль Н.Н., Адаменко Т.И., Скакун С.В., Кравченко А.Н., Кривобок А.А., Шелестов А.Ю., Колотий А.В., Куссуль О.М., Лавренюк А.Н. //

- Международный научно-технический журнал «Проблемы управления и информатики», 2013, №3, С. – 138-150.
8. Колотий А.В. Выбор информационных признаков для построения прогнозов урожайности озимой пшеницы для территории Украины // Наукові праці ДонНТУ Серія "Інформатика, кібернетика та обчислювальна техніка". - 2013. - Випуск 2(18), С. 99-105.
  9. Сайт для отримання супутникових продуктів MODIS. – [Електронний ресурс] – Режим доступу: <http://e4ftl01.cr.usgs.gov/MOLT/MOD13Q1.005/>
  10. Сайт NOAA для отримання супутникових продуктів. – [Електронний ресурс] – Режим доступу: <http://www.star.nesdis.noaa.gov/smcd/emb/vci/VH/>
  11. Сайт консорціуму Geoland . – [Електронний ресурс] – Режим доступу: <http://www.geoland2.eu/portal/service/ShowServiceInfo.do?serviceId=BB808F80>
  12. Bhuiyan C. Various Drought Indices For Monitoring Drought Condition In Aravalli Terrain Of India // International Journal of Applied Earth Observation and Geoinformation 8 — 2006. — PP. 289-302.
  13. Gower S. T. Direct and indirect estimation of leaf area index, FAPAR, and net primary production of terrestrial ecosystems / S. T. Gower, C. J. Kucharik, J. M. Norman // Remote Sensing of Environment. – 1999. – Vol. 70. – №. 1. – P. 29-51.
  14. Jansen, M.J.W. Validation of CGMS // Workshop for Central and Eastern Europe on agrometeorological models: theory and applications in the MARS project Ispra, Italy, 1994 (Eds.) J.F. Dallemard, P. Vossen, Luxembourg: Office for Off. Publ. of the EU, 1994. — P. 159-170.
  15. Бідюк П.І. Аналіз часових рядів. / П.І. Бідюк, В.Д. Романенко, О.Л. Тимошук – Київ: «Політехніка», НТУУ “КПІ”, 2013. – 600 с.
  16. Kussul N. Grid technologies for satellite data processing and management within international disaster monitoring projects / Kussul N., Shelestov A., Skakun S.// Grid and Cloud Database Management, 2011. P. 279-305.
  17. Kussul O. Assessing security threat scenarios for utility-based reputation model in Grids / Kussul O., Kussul N., Skakun S.// Computers & Security. – 2013. – vol. 34. – P. 1-15.
  18. Куссуль Н. Н. Grid-системы для задач исследования Земли.: Архитектура, модели и технологии. / Куссуль Н.Н., Шелестов А.Ю. – К.: “Наукова думка”, 2008. – 452 с.