

УДК 519.254

Наук. співроб., к.т.н. Кравченко О. М., інженер Колотій А. В.,  
к.т.н., доцент Олефір О.С., студ. Дмитренко О. А.

Інститут космічних досліджень НАН та НКА України,  
Національний технічний університет України «Київський  
політехнічний інститут»

## РЕГРЕСІЙНА МОДЕЛЬ ВРОЖАЙНОСТІ ОЗИМИХ КУЛЬТУР НА ОСНОВІ ВЕТАЦІЙНОГО ІНДЕКСУ

### Abstract

*Oleksii M. Kravchenko, scientific researcher, PhD; Andrii V. Kolotii, engineer,  
Oleksandr S. Olefir, assoc. prof., PhD; Olexandra Dmytrenko, student  
Evaluating plant's condition with the help of satellite data*

*It is always good to know what to expect from future because in this case it's possible to make a good plan to allocate existing resources, especially when the subject is food. The purpose of this paper is to show the possibility to predict the amount of harvest and explain how it may be done. One of the crops with which Ukraine is famous all over the Europe is wheat; this crop will be the basis for the paper.*

### Вступ

Україна є одним з найбільших постачальників пшениці в Європі. На сьогоднішній день наша країна входить до п'ятірки найбільших експортерів зерна, має певні договірні зобов'язання перед іншими країнами [1]. Таким чином, важливим є найскоріше зробити висновки стосовно майбутнього врожаю: передбачити кількість зерна, що буде отримано.

Визначивши, який чинник (чинники) впливають на врожайність пшениці та побудувавши модель залежності врожайності від стану розвитку культури за певний період, можна передбачити, який буде врожай.

На сьогоднішній день є можливість використовувати сучасні супутники для спостереження за станом рослинності. Зокрема, за допомогою приладу MODIS супутника TERRA отримуються високочастотні знімки безхмарного покриття України раз в 16 днів з просторовим розрізненням 250м, якого достатньо для моніторингу стану полів України. Дані для досліджень регулярно обчислюються Інстит. космічних досліджень НАН та НКАУ [2].

## Постановка задачі

На основі статистичних даних про врожайність в певному регіоні та відповідних часових рядів вегетаційних індексів для цього ж регіону необхідно побудувати регресійну модель врожайності.

## Термінологія

*Вегетаційний індекс* (VI) – основний параметр, що показує стан рослинності, який обчислюється за супутниковими даними. Значення VI, здебільшого, варіюються в залежності від кількості зеленої біомаси.

*NDVI* – найвідоміший індекс оцінки стану рослинності. NDVI простий для обчислення, має найширший динамічний діапазон з поширених VI і найкращу чутливість до змін у рослинному покриві. Він помірно чутливий до змін ґрунтового та атмосферного фону, окрім випадків з бідною рослинністю (рослинний покрив присутній на більш ніж 30% поля). NDVI вираховується за формулою

$$NDVI = \frac{NIR - VIS}{NIR + VIS},$$

де *NIR* – коефіцієнт відбиття у близькому інфрачервоному діапазоні (0,8 мкм), а *VIS* – у червоному діапазоні (0,6 мкм) [3].

*Кореляція (кореляційна залежність)* – статистична залежність двох або декількох випадкових величин (або величин, які можна з деяким допустимим ступенем точності вважати такими). При цьому, зміна значень однієї або декількох з цих величин призводять до систематичної зміни значень іншої або інших величин. Математичною мірою кореляції двох випадкових величин є коефіцієнт кореляції, який може змінюватись від -1 до 1.

Нехай *X*, *Y* – дві випадкові величини, визначені на одному імовірнісному просторі. Тоді їх коефіцієнт кореляції задається формулою:

$$R_{X,Y} = \frac{M[X,Y] - M[X] \cdot M[Y]}{\sqrt{M[X^2] - (M[X])^2} \cdot \sqrt{M[Y^2] - (M[Y])^2}},$$

де *M* – математичне очікування [4].

## Алгоритм

Для побудови моделі залежності врожайності від вегетаційного індексу використовується лінійна регресія. Перед побудовою регресійної моделі вираховуються предиктори за допомогою кореляції. З'ясовується день, в який спостерігається найбільша залежність врожаю за минулі роки

від NDVI (цей вегетаційний індекс використовується в даній моделі). Також необхідно знайти степінь довіри моделі – коефіцієнт детермінації, що визначається за наступною формулою:

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_{i1} - (p_{11}X_{i1} + p_{12}))^2}{(n-1)D[Y]},$$

де  $Y$  – вектор стовпець розмірності  $n \times 1$  – врожайність за вибраний період,  $X$  – вектор стовпець розмірності  $n \times 1$  – NDVI за той самий період,  $p_{11}$  та  $p_{12}$  – коефіцієнти лінійної регресії,  $D$  – дисперсія.

На Рис. 1 є дані по кореляції – зв'язку між врожайністю та NDVI по Вінницькій області за 2000-2005 роки. Горизонтальна вісь – дні року; вертикальна – значення кореляції. Від'ємна кореляція означає, що залежність обернена, а не пряма. Значення кореляції на малюнку по 2000 року сильно скачуть та менші за 0.5. Таким чином, дані за цей рік не відображають сильну залежність. Погодні умови в цьому році були нестабільні. Була засуха. Дані по 2002 та 2004 роках можна назвати стабільними; максимальне значення кореляції перевищує 0.7 на 113 та 129 днях років відповідно. 0,7 - достатньо великий показник: врожайність сильно залежить від індексу NDVI. Таким чином, можна використовувати відомості за ці дні для побудови регресійної моделі.

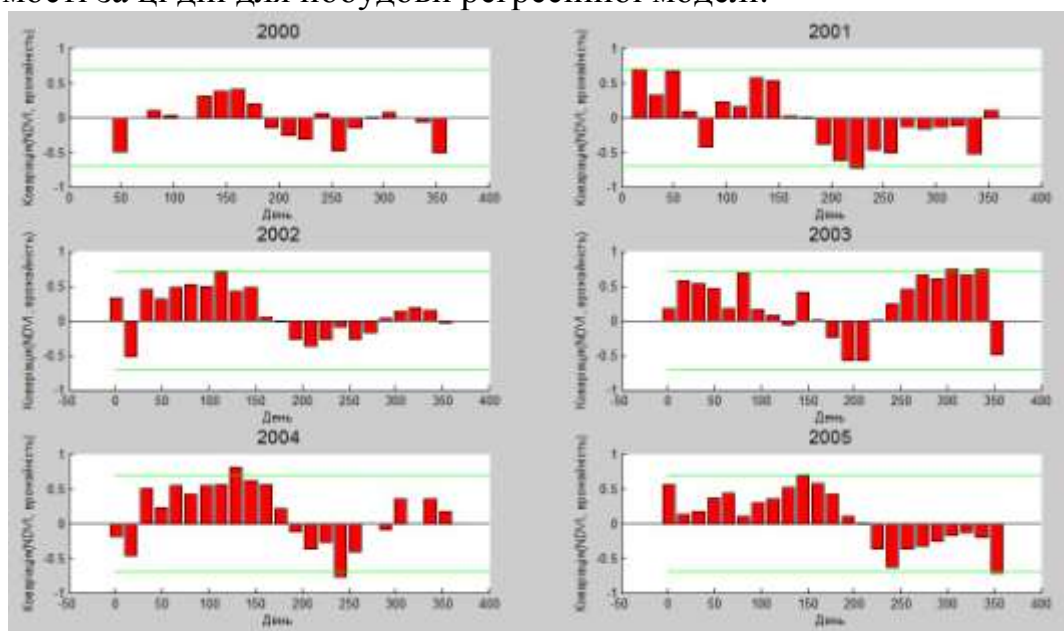


Рис. 1. Кореляція від NDVI та врожайності за 2000-2005 роки

На Рис. 2 зображено графік лінійної регресії по Вінницькій області та табличні значення (27 крапок, що відповідають 27-ми районам Вінницької області). На вертикальній осі відкладено врожайність за 2009 рік; на

горизонтальній – значення NDVI на 145-му дні року. Лінійна регресія підходить для відображення залежності врожайності від NDVI, судячи з графіку;  $R^2 = 0,53$ . Стандартне відхилення за отриманою моделлю становить 45 центнерів. Значення кореляції для цих даних дорівнює 0,78. Для побудови більш точної моделі можна досліджувати або іншу залежність, або взяти до розгляду додаткові дані.

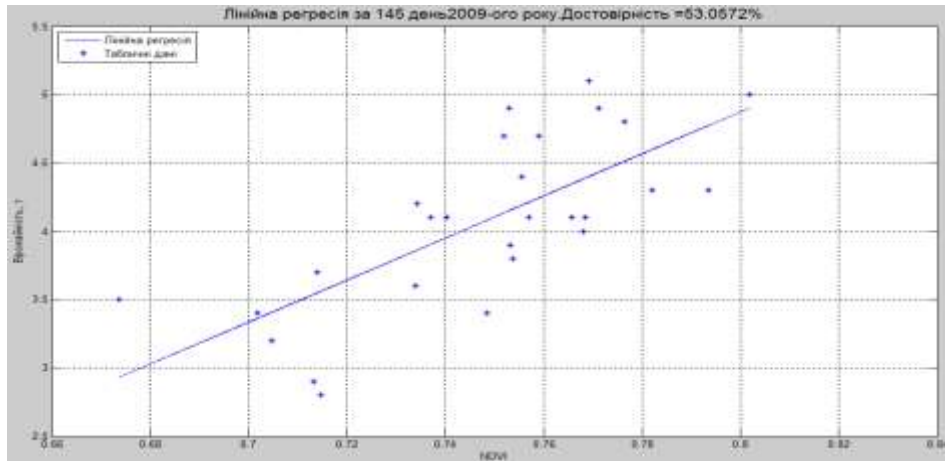


Рис. 2. Графік лінійної регресії озимої пшениці за 145 день 2009-го року по 27-ми районах Вінницької області

## Висновки

В результаті дослідження побудовано модель залежності врожаю від NDVI. Отримана модель гарно наближає реальні дані. Маючи велику базу оброблених даних можна користуватись принципом років-аналогів, тобто, бачачи, що погода та інші умови, що впливають на ріст культури в один із попередніх років близькі до поточних, можна спрогнозувати врожай року-аналога [5]. В подальшому планується долучити інші дані для роботи з регресійною моделлю: метеорологічні дані та дані про проективне покриття.

## Література

- 1) Світові запаси пшениці зменшаться через зниження виробництва в Україні [Електронний ресурс] / <http://vkurse.ua/ua/economics/iz-zasnizheniya-proizvodstva-v-ukraine.html>, свободный. – Загол. з екрану.
- 2) Куссиль Н.Н., Шелестов А.Ю., Скакун С.В., Кравченко А.Н. Интеллектуальные вычисления в задачах обработки данных наблюдения Земли, монографія – К: Наукова думка, 2007. - 196 с.

- 3) Дубинин М. Вегетационные индексы [Электронный ресурс] <http://gis-lab.info/qa/vi.html>, свободный. – Загол. з экрану.
- 4) Гмурман В.Е. Теория вероятностей и математическая статистика. Учеб. пособие для вузов. Изд. 6-е, стер. — М: Высш. шк., 1998. — 479 с. ил.
- 5) Савин И. Ю., Барталеев С. А., Лупян Е. А., Толпин В. А., Хвостиков С. А. Прогнозирование урожайности сельскохозйственных культур на основе спутниковых данных: возможности и перспективы // Современные проблемы дистанционного зондирования Земли из космоса. – 2010. – №3. – 275-285.