

Сравнительный анализ результатов регрессионных и биофизических моделей в задаче прогнозирования урожайности озимой пшеницы

Ф. Коган¹, Н.Н. Куссуль², Т.И. Адаменко⁴, С.В. Скаакун², А.Н. Кравченко²,
А.А. Кривобок⁶, А.Ю. Шелестов³, А.В. Колотий², О.М. Куссуль⁵, А.Н. Лавренюк⁵

¹ Национальное управление океанических и атмосферных исследований (NOAA)
5200 Auth Rd, Camp Springs MD 20746, USA
E-mail: Felix.Kogan@noaa.gov

² Институт космических исследований НАНУ-НКА
03187 Киев, просп. Глушкова, 40, корп. 4/1
E-mail: inform@ikd.kiev.ua

³ Национальный университет биоресурсов и природопользования Украины
03187 Киев, ул. Героев Обороны, 15
E-mail: andrii.shelestov@gmail.com

⁴ Украинский гидрометеорологический центр
01034 Киев, ул. Золотоворотская, 6-В
E-mail: adamenko@meteo.gov.ua

⁵ Национальный технический университет Украины «КПИ»
03056, Киев, пр. Победы, 37
E-mail: olgakussul@gmail.com

⁶ Украинский научно-исследовательский гидрометеорологический институт
03650, Киев, просп. Науки, 37
kryvobok@uhmi.org.ua

В данной статье ставится задача оценить относительную эффективность использования спутниковых данных для прогнозирования урожайности озимой пшеницы в Украине на уровне отдельных областей, а также сравнить эффективность прогнозирования на основе эмпирических и биофизических моделей. В качестве эмпирических моделей урожайности рассматриваются линейные регрессионные модели зависимости урожайности от 16-дневного композита индекса NDVI на основе данных MODIS с разрешением 250 м, а также нелинейная регрессионная модель, в качестве предикторов которой используются метеорологические параметры. Эмпирический подход к прогнозированию сравнивается с биофизическим подходом, реализованным в адаптированной для Украины системе CGMS (Crop Growth Monitoring System), в основе которой лежит модель роста растительности WOFOST. Показано, что наиболее высокую точность прогноза урожайности озимой пшеницы обеспечивает подход на основе биофизических моделей роста. Недостатком этого подхода является сложность настройки модели (большой объем входных данных) и недостаточная заблаговременность. Регрессионные модели (линейная модель на основе спутниковых данных и нелинейная на основе метеорологических) показывают примерно одинаковую точность, которая с учетом достаточной заблаговременности позволяет использовать эти прогнозы на практике.

Ключевые слова: прогноз урожайности, регрессионная модель, биофизическая модель, маска культуры, информационная технология, MODIS, NDVI, кластеризация, CGMS, WOFOST.

1. Введение

В связи с увеличением населения Земли и влиянием глобальных изменений климата на производительность сельского хозяйства на первый план выходит проблема продовольственной безопасности. Решению этой проблемы посвящена инициатива стран Большой двадцатки по созданию Глобальной системы сельскохозяйственного мониторинга

(GLAM, 2010). Задача проекта состоит в объединении данных спутниковых и наземных наблюдений для мониторинга агрометеорологических параметров, роста сельскохозяйственных культур и динамики изменения водных ресурсов. Понятие агромониторинга включает два основных компонента: оценку посевных площадей и прогнозирование урожайности. Точность прогнозирования урожайности с заблаговременностью несколько месяцев до начала сбора урожая играет важную роль при решении задач глобального, национального и регионального уровней.

Украина является одной из наиболее развитых сельскохозяйственных стран и одним из крупнейших производителей сельскохозяйственных культур в мире. Согласно статистическим данным зарубежной сельскохозяйственной службы (FAS – Foreign Agricultural Service) Министерства сельского хозяйства США (USDA – U.S. Department of Agriculture) за 2011 г., Украина занимает 8-е место среди крупнейших экспортеров и 10-е – среди крупнейших производителей пшеницы в мире. Поэтому своевременный и точный прогноз урожайности в Украине является ключевым элементом для поддержки принятия решений в политике продовольственной безопасности в мире.

В настоящее время для прогнозирования урожайности используются два наиболее распространенных подхода: на основе эмпирических регрессионных моделей и биофизических моделей роста растительности. Эмпирические модели связывают урожайность с некоторыми предикторами (например, характеристиками биомассы, получаемыми со спутников, метеорологическими параметрами) и, как правило, не требуют большого числа входных параметров. Такие модели являются достаточно простыми в реализации, однако они недостаточно робастны. Эффективность таких моделей в значительной степени зависит от наличия и качества данных (Becker-Reshef et al., 2010).

Модели роста сельскохозяйственных культур позволяют получить биофизические параметры культур, такие как урожайность, биомассу, содержание воды и т.д. К этому классу относятся модели World Food Studies (WOFOST) (Boogaard et al., 1998), реализованная в рамках Европейской системы мониторинга сельскохозяйственных культур Crop Growth Monitoring System (CGMS), а также модели EPIC (Erosion Productivity Impact Calculator) (Williams et al., 1984) и CERES (Crop Environment Resource Synthesis) (Ritchie et al., 1985). Основная трудность применения таких моделей состоит в том, что они требуют многочисленных входных параметров для запуска модели, в частности информации о типе почвы, сортах культур, метеорологических данных, агротехнологии (сроках посева, всходов, цветения, фазы молочной зрелости и т.д.). Хотя эти модели достаточно надежны и робастны, их применение требует соответствующей адаптации и калибровки для учета сельскохозяйственной специфики конкретного региона.

Спутниковые данные играют важную роль в прогнозировании урожайности, так как они являются источником своевременной и объективной информации для больших территорий. Спутниковые продукты, такие как лиственный индекс LAI (Leaf Area Index), вегетационные индексы NDVI (Normalized Difference Vegetation Index) и EVI (Enhanced Vegetation Index), метеорологические параметры, получаемые со спутников Meteosat и NOAA-AVHRR, можно ассимилировать в модели роста растительности для повышения

их эффективности (de Wit and van Diepen, 2008; Fang et al., 2011; Curnela et al., 2011). Их также можно использовать в качестве предикторов в эмпирических регрессионных моделях. В частности, в работах (Becker-Reshef et al., 2010; Mkhabela et al., 2011; Kogan et al., 2012) описано применение вегетационного индекса NDVI и индекса здоровья растительности VHI (Vegetation Health Index) для прогнозирования урожайности.

Что касается Украины, то известные работы были в основном направлены на исследование методов прогнозирования на примере района или даже тестового полигона либо на прогнозирование урожайности в масштабе страны (Kogan et al., 2011). Прогнозированию урожайности на уровне отдельных областей уделялось меньше внимания (Куссуль и др., 2008, Куссуль и др., 2011; Куссуль, Скаун и др., 2011). Работы, посвященные сравнению различных подходов к моделированию на основе эмпирических моделей и моделей роста растительности, авторам неизвестны.

Поэтому в данной статье ставится задача оценить относительную эффективность использования спутниковых данных для прогнозирования урожайности озимой пшеницы в Украине на уровне отдельных областей, а также сравнить эффективность прогнозирования на основе эмпирических и биофизических моделей.

В качестве эмпирических моделей урожайности будут рассмотрены линейные регрессионные модели зависимости урожайности от 16-дневного композита индекса NDVI на основе данных MODIS с разрешением 250 м, а также нелинейная регрессионная модель, в качестве предикторов которой используются метеорологические параметры. Эмпирический подход к прогнозированию будет сравниваться с биофизическим подходом, реализованным в адаптированной для Украины системе CGMS (Crop Growth Monitoring System), в основе которой лежит модель роста растительности WOFOST (Кривобок и др., 2010; Адаменко и др., 2011).

2. Данные и методология

В состав Украины входят 24 области и Автономная Республика Крым, площадь которых составляет от 8097 до 33 310 кв. км (средняя площадь области составляет около 24 000 кв. км). В целом Украину можно разделить на следующие агроклиматические зоны: Полесье (зона смешанных лесов в северо-западной части, занимающая около 20% всей территории Украины), лесостепная (более 33%) и степная зона в южной части Украины (порядка 40% территории).

Пшеница является одной из самых важных сельскохозяйственных культур в Украине. За последние пять лет пшеница составила в среднем почти 45% производства всех зерновых культур. Озимая пшеница составляет в среднем 95% урожая пшеницы.

Для идентификации параметров моделей урожайности будут использованы официальные статистические данные по урожайности озимой пшеницы на уровне областей за период 2000–2009 гг. Валидация моделей будет выполнена на данных 2010 г.

Предложенный подход состоит в построении эмпирических регрессионных моделей для каждой области Украины (Kogan et al., 2012). В качестве обучающей выборки

для настройки параметров регрессионной модели используются значения индекса NDVI, вычисленные по данным прибора MODIS, которые в автоматическом режиме загружаются с сервера USGS, и статистические данные урожайности по областям за 2000–2009 гг. Значения индекса NDVI усредняются для каждой области по маске посевных территорий, построенной по карте классификации земного покрова ESA GLOBCOVER.

2.1. Линейная регрессионная модель прогнозирования урожайности озимой пшеницы по спутниковым данным

Следует отметить, что для урожайности озимой пшеницы наблюдается положительный линейный тренд для всех областей в связи с улучшением сельскохозяйственных технологий (Kogan et al., 2012). Таким образом, регрессия строится для отклонения урожайности от тренда в зависимости от вегетационного индекса NDVI:

$$dY_i = Y_i - T_i = f(NDVI_i) = b_0 + b_1 * NDVI_i, \quad (1)$$

где $NDVI_i$ – значение индекса NDVI, выбранное для определенного дня года i .

Для оценки значений b_0 и b_1 для уменьшения влияния выбросов в регрессионной модели используется робастная линейная регрессия (Street et al., 1988).

Для определения информационных признаков и проверки робастности модели была использована процедура кросс-валидации (leave-one-out cross-validation – LOOCV) (Jansen, 1995). В рамках этой процедуры модель строится на данных за все годы, кроме одного, который используется для валидации. В качестве предиктора модели (1) выбирается значение NDVI за тот промежуток времени, который обеспечивает минимальное значение среднеквадратической ошибки RMSE:

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

где P и O – прогнозируемое и наблюдаемое значения урожайности.

Оказалось, что наиболее информативным предиктором для степной зоны и большинства других областей является значение NDVI за 129 дней (начало мая).

Относительная эффективность использования спутниковой информации для прогнозирования урожайности вычисляется следующим образом:

$$Eff = \frac{\frac{1}{n} \sum_i (dY_i)^2}{RMSE^2}. \quad (3)$$

Относительная эффективность показывает, во сколько раз использование данных спутниковых наблюдений позволяет уменьшить ошибку прогнозирования урожайности по сравнению с моделью тренда. Конкретные значения относительной эффективности и коэффициента детерминации представлены в табл. 1. Распределение относительной эффективности по областям представлено на рис. 1.

Таблица 1. Относительная эффективность и коэффициенты детерминации регрессионных моделей усредненные по агроклиматическим зонам

<i>Агроклиматическая зона</i>	<i>Относительная эффективность</i>	<i>R-квадрат</i>
Полесье	1,182	0,479
Лесостепь	1,576	0,667
Степь	1,883	0,804

2.2. Нелинейная регрессионная модель прогнозирования урожайности озимой пшеницы по метеорологическим данным

Вторым подходом, исследуемым в данной работе, является регрессионное моделирование урожайности на основе метеорологических факторов. Для построения моделей в качестве предикторов использовались ежемесячные метеорологические данные, усредненные по областям. Для усреднения использовались ежедекадные данные всех действующих метеостанций для каждой области. Для построения прогноза урожайности рассматривался временной ряд метеопараметров за 12 месяцев до сбора урожая озимой пшеницы (период июль прошлого года – июнь текущего года). Использовались следующие параметры: максимальная, минимальная и средняя температура воздуха (36 значений), совокупный уровень осадков (12 значений), запас продуктивной влаги в слое 0–20 см за сентябрь–октябрь и апрель–июнь (5 значений).

Поскольку для построения робастной модели с большим количеством предикторов (53 в данной работе) необходим большой объем обучающей выборки, то вместо построения модели для каждой области ставится задача построения модели для каждой агроклиматической зоны. В качестве прогнозируемой величины используется нормированное значение отклонения урожайности от тренда:

$$ndY_i = \frac{dY_i}{\sqrt{\frac{1}{n} \sum_i dY_i^2}}. \quad (4)$$

При такой нормировке средний квадрат прогнозируемой величины для каждой области одинаков и равен 1.

Линейный регрессионный анализ предикторов и урожайности не выявил хороших моделей, что связано с наличием существенных нелинейных взаимосвязей. Примером такой взаимосвязи является зависимость уровня урожайности пшеницы от средней температуры в апреле. Поэтому для создания регрессии необходимо использовать нелинейные модели. Одним из современных методов создания нелинейных регрессионных моделей в ситуации большого количества предикторов являются ядерные методы (Bishop, 2006). В данной работе использовались регрессионные модели на основе гауссовских процессов (Rasmussen, 2006).

При использовании регрессии на основе гауссовских процессов оценка прогнозируемой величины $\bar{f}(x_*)$ при значении вектора предикторов x^* представляется в виде:

$$\bar{f}(x_*) = \sum_1^n \alpha_i k(x_i, x_*), \quad (5)$$

где $\alpha_i = [K(X, X)]^{-1} Y$ – коэффициенты линейной комбинации ядерных функций $k(x_i, x_*)$, вычисленные для тестового вектора x^* и для каждого вектора обучающей выборки x_i ; $Y = (y_i, i = \overline{1, n})^T$ – вектор значений прогнозируемой переменной; $K(X, X)$ – матрица значений ковариационной функции для всех парных комбинаций векторов предикторов x_i . В данной работе используется ковариационная функция ARD, которая позволяет выделить релевантные предикторы для прогноза (Rasmussen, 2006).

Конкретные значения относительной эффективности и коэффициента детерминации для регрессионных моделей, построенных с использованием метеорологических данных представлены в табл. 2.

Таблица 2. Относительная эффективность и коэффициенты детерминации регрессионных моделей, усредненные по агроклиматическим зонам
(жирным выделены отобранные модели)

Агроклиматическая зона	Момент создания прогноза	RMSE	Относительная эффективность
Полесье	Апрель	0,9805	1,0199
Полесье	Май	0,8397	1,1909
Полесье	Июнь	0,8280	1,2077
Лесостепь	Апрель	0,8208	1,2183
Лесостепь	Май	0,7011	1,4263
Лесостепь	Июнь	0,7968	1,2550
Степь	Апрель	0,5729	1,7455
Степь	Май	0,7634	1,3576
Степь	Июнь	0,7366	1,3576

2.3. Прогнозирование урожайности озимой пшеницы с использованием системы CGMS

Основной задачей CGMS является оценка влияния погодных условий на развитие и урожайность сельскохозяйственных культур в региональном масштабе. В системе метеорологические данные, агрометеорологическая информация, данные о типе почв используются в моделях роста и развития сельскохозяйственных культур и многолетних

трав WOFOST (Boogaard et al., 1998) и LINGRA (Schapendonk et al., 1998), соответственно, для расчета параметров биопродуктивности. Эти модели позволяют оценивать биомассу и урожайность сельскохозяйственных культур с дневным шагом по времени для различных типов культур. Результаты моделирования непосредственно используются не только для мониторинга посевов сельскохозяйственных культур, но и как входная информация для прогноза урожайности.

Затем результаты моделирования агрегируются к административным единицам, в нашем случае – районам, областям и в общем по стране, и используются в прогностическом блоке для количественной оценки прогноза урожайности. В качестве предикторов в CGMS можно задавать как метеорологические, так и параметры продуктивности сельскохозяйственных посевов. К метеорологическим предикторам относятся минимальная, максимальная и средняя температура, сумма осадков, суммарная солнечная радиация, потенциальная эвапотранспирация и так называемый климатический водный баланс.

В CGMS для прогноза урожайности используется линейное уравнение, в котором учитывается временной тренд и моделируемые параметры биопродуктивности (Vossen, 1992)

$$Y_T = b_0 + b_1 T + b_2 S_T, \quad (6)$$

где Y_T – прогнозируемая урожайность (т/га); $S_T = f(MET \text{ или } SIM)$ – функция, описывающая урожайность, обусловленную влиянием метеорологических условий, отличных от среднестатистических (одним из предикторов, т/га); T – год; b_0 – средняя официальная урожайность (т/га); b_1 – годовой линейный тренд официальной урожайности (т/га); b_2 – константа при S_T , которая находится в диапазоне от 0 до 1.

3. Полученные результаты

Проанализируем результаты, полученные каждой из моделей, и оценим эффективность прогнозирования на независимых данных 2010 г.

Таблица 3 представляет сравнение официальной статистики по урожайности озимой пшеницы с результатами прогнозирования с использованием линейной регрессионной модели, основанной на индексе NDVI по данным MODIS, нелинейной регрессионной модели, основанной на метеорологических данных и системы CGMS (прогноз на 20 июня), основанной на модели роста растительности WOFOST и метеорологических данных.

Таблица 3. Фактическая и спрогнозированная урожайность озимой пшеницы за 2010 г. для областей Украины с использованием подходов: 1) линейная регрессионная модель; 2) нелинейная регрессионная модель; 3) система CGMS

Область	Факт., ц/га.	1) MODIS NDVI		2) Meteo.		3) CGMS	
		Прогн., ц/га.	Прогн-факт, ц/га	Прогн., ц/га.	Прогн-факт, ц/га	Прогн., ц/га.	Прогн-факт, ц/га
Винницкая	32,9	41,6	8,7	38,3866	5,4866	38,1	5,2
Волынская	26,0	28,7	2,7	33,6198	7,6198	27,8	1,8
Днепропетровская	28,5	37,4	8,9	26,3064	-2,1936	29,6	1,1
Донецкая	29,8	36,8	7	26,733	-3,067	29,3	-0,5
Житомирская	25,2	30,4	5,2	32,7905	7,5905	22	-3,2
Закарпатская	21,2	32	10,8	33,5906	12,3906	-	-
Запорожская	26,1	34,2	8,1	28,5798	2,4798	27,4	1,3
Ивано-Франковская	24,8	27,9	3,1	32,028	7,228	20,7	-4,1
Киевская	25,1	36	10,9	38,2196	13,1196	27,7	2,6
Кировоградская	30,1	26,1	-4	27,302	-2,798	29,7	-0,4
Луганская	23,5	33,3	9,8	24,0395	0,5395	25,8	2,3
Львовская	25,3	30,2	4,9	32,8299	7,5299	29	3,7
Николаевская	28,5	25,7	-2,8	26,4399	-2,0601	26,4	-2,1
Одесская	28,3	32,4	4,1	24,6934	-3,6066	26,3	-2
Полтавская	26,1	34,2	8,1	37,6892	11,5892	30,1	4
Ровненская	29,0	30,3	1,3	32,6335	3,6335	28,1	-0,9
Сумская	22,0	36,6	14,6	33,6676	11,6676	23,7	1,7
Тернопольская	25,1	37,4	12,3	35,9743	10,8743	30	4,9
Харьковская	20,8	34,4	13,6	35,3365	14,5365	23,1	2,3
Херсонская	24,3	22,6	-1,7	25,3542	1,0542	21,7	-2,6
Хмельницкая	28,8	27,7	-1,1	32,6602	3,8602	33,9	5,1
Черкасская	35,0	41,2	6,2	42,6373	7,6373	39,8	4,8
Черновицкая	26,6	36,9	10,3	34,6116	8,0116	30,1	3,5
Черниговская	21,7	31,6	9,9	32,8652	11,1652	22	0,3
АР Крым	21,2	23	1,8	24,9001	3,7001	24,4	3,2
		СКО, ц/га. =	7,94	СКО, ц/га. =	7,79	СКО, ц/га. =	3,04

Как видно из рис. 2, наилучшие результаты с точки зрения величины среднеквадратического отклонения (СКО) (3,0 ц/га), были получены с помощью системы CGMS. Но эти прогнозы были получены путем запуска модели на 20 июня 2010 г., только за несколько недель до сбора урожая. Регрессионные модели на основе NDVI и метеорологических данных переоценивают урожайность в сравнении с данными официальной статистики в большинстве областей с СКО 7,9 ц/га и 7,8 ц/га, соответственно. Но эти прогнозы были получены в апреле–мае, то есть за 2–3 месяца до сбора урожая.

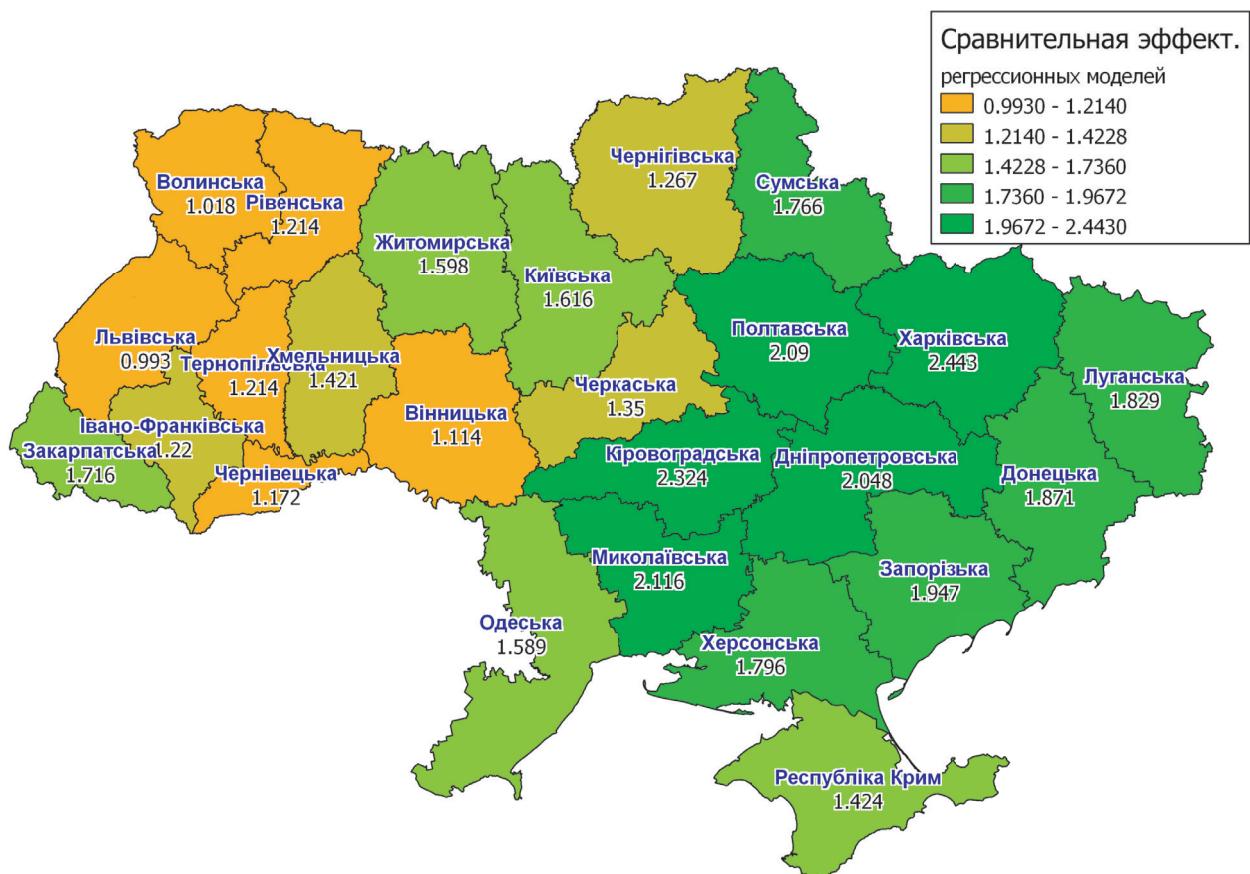


Рис. 1. Сравнительная эффективность регрессионных моделей для всех областей Украины

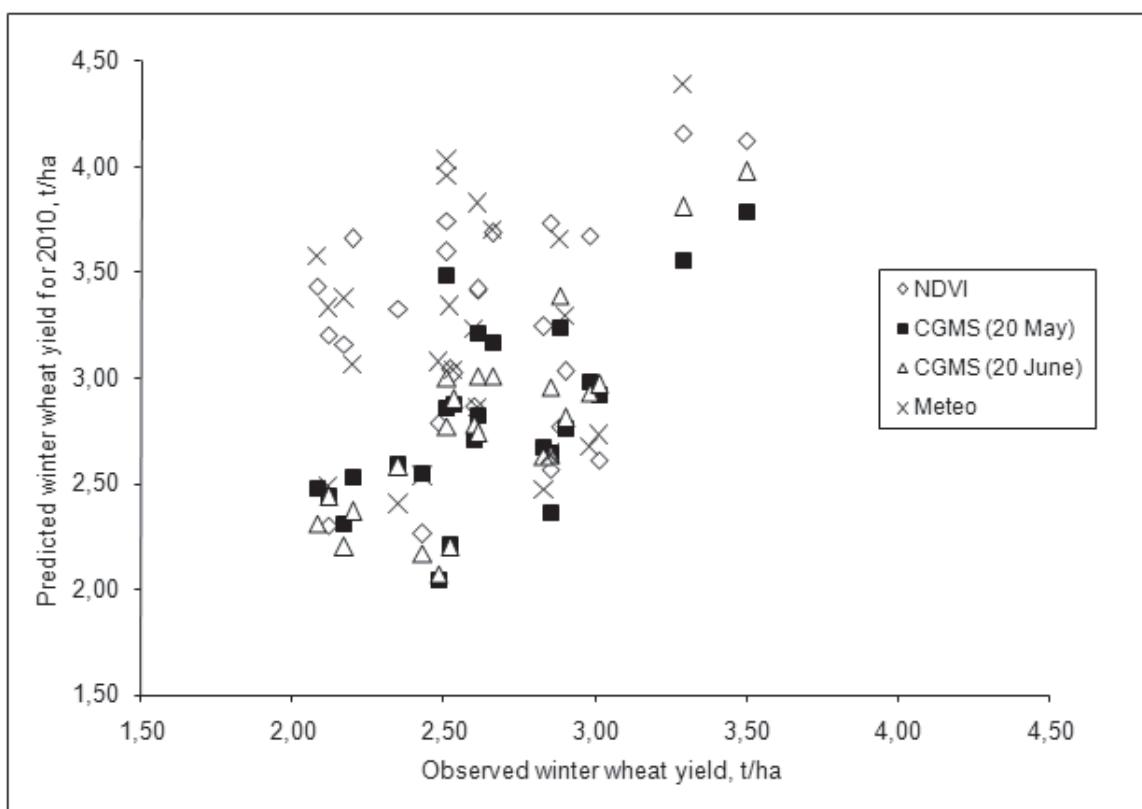


Рис. 2. Точечная диаграмма сравнения результатов прогнозирования урожайности озимой пшеницы с использованием различных подходов с данными официальной статистики

4. Выводы и направления дальнейших исследований

Таким образом, на основе выполненных исследований можно утверждать, что наиболее высокую точность прогноза урожайности озимой пшеницы обеспечивает подход на основе биофизических моделей роста. Недостатком этого подхода является сложность настройки модели (большой объем входных данных) и недостаточная заблаговременность. Регрессионные модели (линейная модель на основе спутниковых данных и нелинейная на основе метеорологических) показали примерно одинаковую точность, которая с учетом достаточной заблаговременности позволяет использовать эти прогнозы на практике. Проблемой построения регрессионных моделей, которая, на наш взгляд, не обеспечивает достаточной точности прогнозирования, является недостаточное количество данных (короткий временной ряд) для построения тренда и настройки параметров модели. Этот фактор обуславливает необходимость проверки робастности моделей не только на данных 2010 г., но и последующих лет.

Именно с этим будут связаны дальнейшие исследования авторов. Кроме того, планируется уделить внимание эффективной реализации предложенных моделей в распределенных системах (Куссуль и др., 2006; Kussul et al., 2009, 2010, 2011).

Благодарности

Эта работа поддерживается Фондом гражданских исследований и развития США (CRDF). Грант «Analysis of climate change & food security based on remote sensing & in situ data sets» (# UKB2-2972-KV-09).

Литература

1. Адаменко Т.И., Кривобок А.А., Кривошеїн А.О. Особенности адаптации системы CGMS для оперативной оценки состояния и прогноза урожайности озимой пшеницы в Украине // Труды УкрНИГМИ. 2011. № 261. С. 118–129.
2. Кривобок О.А., Кривошеїн О.О., Круківська А.В. Адаптація характеристик ґрунтового покриву України для використання в системі моніторингу умов росту і розвитку сільсько-господарських культур (CGMS) на основі картографічного методу // Національне картофрування: стан, проблеми та перспективи розвитку: зб. наук. пр. 30.09.10. С. 122–126.
3. Куссуль Н.Н., Скакун С.В., Лавренюк А.М., Колотий А.В. Индукционный подход к прогнозированию урожайности озимой пшеницы по спутниковым данным в Украине // Пятый Белорусский космический конгресс (25–27 октября 2011 года, Минск): Материалы конгресса. Т. 2. С. 116–120.
4. Куссуль Н.Н., Ильин Н.И., Скакун С.В., Лавренюк А.Н. Оценка состояния растительности и прогнозирование урожайности озимых культур Украины по спутниковым данным // International Book Series «Decision Making and Business Intelligence, Strategies and Techniques» (eds. Markov K., Ivanova K., Mitov I.). 2008. № 3. Р. 103–109.

5. Шелестов А.Ю., Куссуль Н.Н., Скаун С.В. Grid-технологии в системах мониторинга на основе спутниковых данных // Проблемы управления и информатики. 2006. № 1–2. С. 259–270.
6. Becker-Reshef I., Vermote E., Lindeman M., Justice C. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data // Remote Sensing of Environment. 2010. 114 (6). P. 1312–1323.
7. Bishop C. Pattern Recognition and Machine Learning. Springer, New York, USA. 2006. 740 p.
8. Boogaard H.L., Diepen C.A., Rötter R.P., Cabrera J.M.C.A., Laar H.H. van. User's guide for the WOFOST 7.1 crop growth simulation model and WOFOST control center 1.5. Technical document 52, Winand Staring Centre, Wageningen, the Netherlands, 1998. 144 p.
9. Curnela Y., de Wit A.J.W., Duveiller G., Defourny P. Potential performances of remotely sensed LAI assimilation in WOFOST model based on an OSS Experiment // Agricultural and Forest Meteorology. 2011. Vol. 151. No. 12. P. 1843–1855.
10. de Wit A.J.W., van Diepen C.A. Crop growth modelling and crop yield forecasting using satellite-derived meteorological inputs // International Journal of Applied Earth Observation and Geoinformation. 2008. Vol. 10. No. 4. P. 414–425.
11. Fang H., Liang S., Hoogenboom G. Integration of MODIS LAI and vegetation index products with the CSM-CERES-Maize model for corn yield estimation // International Journal of Remote Sensing. 2011. Vol. 32. No. 4. P. 1039–1065.
12. Hluchy L., Kussul N., Shelestov A., Skakun S., Kravchenko O., Gripich Y., Kopp P., Lupian E. The Data Fusion Grid Infrastructure: Project Objectives and Achievements // Computing and Informatics. 2010. Vol. 29. No. 2. P. 319–334.
13. Justice C., Becker-Reshef I., Parihar J.S. Global agriculture monitoring // Global Agricultural Monitoring Community of Practice (GEO Task: AG-07-03a). (Eds) Luxembourg: Publications Office of the European Union, 2010. 32 p.
14. Jansen M.J.W. Validation of CGMS. In: J.F. Dallemand, P. Vossen (eds). Workshop for Central and Eastern Europe on agrometeorological models: theory and applications in the MARS project. Ispra, Italy. EUR 16008 EN, Office for Off. Publ. of the EU, Luxembourg. 21–25 November 1994. P. 159–170.
15. Kogan F., Salazar L., Roytman L. Forecasting crop production using satellite-based vegetation health indices in Kansas, USA // International Journal of Remote Sensing. 2012. Vol. 33. No. 9. P. 2798–2814.
16. Kogan F., Menzhulin G., Shamshurina N., Pavlovsky A. New regression models for prediction of grain yield anomalies from satellite-based vegetation health indices // NATO Science for Peace and Security Series – C: Environmental Security Use of Satellite and In-situ Data to Improve Sustainability. Springer, 2011. P. 105–112.
17. Kussul N., Shelestov A., Skakun S. Technologies for Satellite Data Processing and Management Within International Disaster Monitoring Projects // In Fiore, S.; Aloisio, G. (Eds.). Grid and Cloud Database Management Grid, Springer, 2011. P. 279–306.

18. *Kussul N., Shelestov A., Skakun S.* Flood Monitoring on the Basis of SAR Data // In F. Kogan, A. Powell, O. Fedorov (Eds.). Use of Satellite and In-Situ Data to Improve Sustainability. NATO Science for Peace and Security Series C: Environmental Security, Springer, 2011. P. 19–29.
19. *Kussul N., Shelestov A., Skakun S.* Grid and sensor web technologies for environmental monitoring // Earth Science Informatics. 2009. Vol. 2. Nos. 1–2. P. 37–51.
20. *Mkhabela M.S., Bullock P., Raj S., Wang S., Yang Y.* Crop yield forecasting on the Canadian Prairies using MODIS NDVI data // Agricultural and Forest Meteorology. 2011. Vol. 151. No. 3. P. 385–393.
21. *Ritchie J.T., Otter S.* Description and performance of CERES-Wheat: A user oriented wheat yield model. ARS wheat yield project. Springfield, Virginia: USDA ARS, 1985. P. 159–175.
22. *Street J.O., Carroll R.J., Ruppert D.* A Note on Computing Robust Regression Estimates via Iteratively Reweighted Least Squares // The American Statistician. 1988. Vol. 42. No. 2. P. 152–154.
23. *Rasmussen C.E., Williams C.K.I.* Gaussian Processes for Machine Learning. The MIT Press, 2006. ISBN: 0-262-18253-X. 266 p.
24. *Schapendonk A.H.C.M., Stol W., Kraalingen D.W.G. van, Bouman B.A.M.*, 1998. LINGRA, a sink/source model to simulate grassland productivity in Europe // European J. of Agronomy. 1998. Vol. 9. P. 87–100.
25. *Vossen P.* Forecasting national crop yields of E.C. countries: the approach developed by the agriculture project. In: F. Toselli, J. Meyer-Roux (eds). // Proc. Conference on the application of remote sensing to agricultural statistics, Belgirate, Italy. EUR 14262 EN, Office for Official Publications of the EU, Luxembourg, 26–27 November 1991. P. 159–176.
26. *Williams J.R., Jones C.A., Dyke P.T.* A modeling approach to determining the relationship between erosion and soil productivity // Transactions of the ASAE. 1984. Vol. 27. P. 129–144.

Comparative analysis of regression and biophysical models for the problem of forecasting the yield of winter wheat

**F. Kogan¹, N.N. Kussul², T.I. Adamenko⁴, S.V. Skakun², A.N. Kravchenko²,
A.A. Krivobok⁶, A.Yu. Shelestov³, A.V. Kolotii², O.M. Kussul⁵, A.N. Lavrenyuk⁵**

¹ *National Oceanic and Atmospheric Administration
5200 Auth Rd, Camp Springs MD 20746, USA
E-mail: Felix.Kogan@noaa.gov*

² *Space Research Institute NASU-NSAU
Kyiv -03187, Glushkov Ave, 40, 4/1
E-mail: inform@ikd.kiev.ua*

³ *National University of Life and Environmental Sciences of Ukraine
Kyiv -03187, Heroyiv Oborony st., 15
E-mail: andrii.shelestov@gmail.com*

⁴ *Ukrainian Hydrometeorological Center
Kyiv -01034, Zolotovoritskaya st., 6-B
E-mail: adamenko@meteo.gov.ua*

⁵ *National Technical University of Ukraine “KPI”
Kyiv -03056, Peremogu ave, 37
E-mail: olgakussul@gmail.com*

⁶ *Ukrainian Hydrometeorological Institute
03650, Kyiv, Nayki Prospekt, 37
E-mail: kryvobok@uhmi.org.ua*

This article aims at estimation of the relative efficiency of the use of satellite data to predict winter wheat yield in Ukraine at the level of administrative regions (oblast), as well as to compare the effectiveness of prediction based on empirical and biophysical models. As an empirical model we use a linear regression model that uses as predictor 16-day NDVI composite derived from MODIS data at 250 m resolution. Non-linear regression model is considered as well as a predictor that uses meteorological parameters. These two empirical approaches are compared with the biophysical approach, implemented in CGMS (Crop Growth Monitoring System) system, based on WOFOST crop growth model and adapted for Ukraine. It is shown that the highest accuracy of the forecast yield of winter wheat provides an approach based on biophysical models of growth. The disadvantage of this approach is the difficulty of setting up a model (a large amount of input data) and the lack of lead time. Regression models (linear model based on satellite data and nonlinear, based on weather) show roughly the same accuracy, which is given sufficient lead time allows the use of these predictions in practice.

Keywords: forecast yield, the regression model, biophysical model, the mask of culture, information technology, MODIS, NDVI, clustering, CGMS, WOFOST.