

Регрессионные модели оценки урожайности сельскохозяйственных культур по данным MODIS

Н.Н. Куссуль¹, А.Н. Кравченко¹, С.В. Скакун¹, Т.И. Адаменко³, А.Ю. Шелестов²,
А.В. Колотий¹, Ю.А. Грипич¹

¹Институт космических исследований НАНУ-НКАУ,
03187 Киев, просп. Глушкова, 40, корп. 4/1

E-mail: inform@ikd.kiev.ua

²Национальный университет биоресурсов и природопользования Украины,
03187 Киев, ул. Героев Обороны, 15

E-mail: andrii.shelestov@gmail.com

³Украинский гидрометеорологический центр,
01034 Киев, ул. Золотоворотская, 6-В

E-mail: adamenko@meteo.gov.ua

В статье предложена информационная технология построения модели прогнозирования урожайности сельскохозяйственных культур по спутниковым данным на основе индукционного подхода. В качестве модели выбрана регрессионная модель первого порядка, а в качестве предиктора – 16-дневный композит индекса NDVI за фиксированный период, соответствующий максимальному значению NDVI, вычисленный по маске соответствующей культуры. Маска определяется автоматически методом кластеризации временных рядов NDVI (k-средних или карт Кохонена). Показано, что точность регрессионной модели прогнозирования на основе данных по области в целом несколько выше, чем точность прогноза, построенного путем усреднения прогнозных значений по отдельным районам. Важным преимуществом предлагаемого подхода является его применимость не только для автоматизированного прогнозирования урожайности озимой пшеницы, но и других видов сельскохозяйственных культур.

Ключевые слова: прогноз урожайности, регрессионная модель, маска культуры, информационная технология, MODIS, NDVI, кластеризация, карты Кохонена.

Введение

Современный этап развития космических технологий наблюдения Земли характеризуется значительным увеличением объемов и снижением стоимости спутниковой информации. Эта тенденция, наряду с бурным развитием вычислительной техники и повышением требований к системам поддержки принятия решений, ставит на повестку дня разработку технологий автоматизированной обработки информации для решения важных народнохозяйственных задач. Первоочередной задачей сегодняшнего этапа развития информационного общества является решение задач продовольственной безопасности. Задача построения автоматизированных систем поддержки принятия решений в аграрном секторе на основе спутниковой информации, в том числе систем прогнозирования урожайности, является чрезвычайно актуальной и для Украины.

На сегодня существует множество методов и моделей прогнозирования урожайности, в том числе и для Украины. Однако, как правило, они построены для отдельного региона с использованием ограниченного объема данных и обеспечивают прогноз для отдельной культуры (чаще всего, озимой пшеницы). Поэтому актуальной задачей является анализ регулярности и качества спутниковых данных, а также точности и рабочести моделей с целью их реализации в автоматизированных системах поддержки принятия решений. Поэтому в данной статье предпринимается попытка проанализировать данные прибора MODIS, который на сегодня является наиболее регулярным источником бесплатной спутниковой информации до-

статочного разрешения, и построить унифицированную технологию их автоматизированной обработки, позволяющую прогнозировать урожайность не только озимой пшеницы, как это делают большинство авторов научных публикаций, но и групп яровых и поздних яровых культур.

В статье будут проанализированы существующие на сегодня методы прогнозирования урожайности сельскохозяйственных культур и предложена информационная технология обработки спутниковых данных, обеспечивающая построение регрессионной модели прогнозирования урожайности в автоматизированном режиме.

Существующие методы прогнозирования урожайности

Мониторинг сельскохозяйственных культур из космоса, в частности заблаговременная оценка урожайности до наступления жатвы, представляет научный интерес, начиная с 1970-х годов (Wall et al., 2007). Спутниковые данные дистанционного зондирования Земли являются источником своевременной, объективной и относительно недорогой информации. Их применимость к прогнозированию урожайности ввиду их повторяемости, значительного покрытия и быстрого доступа была продемонстрирована для разных масштабов и разных географических регионов (Manjunath et al., 2002; Prasad et al., 2006; Becker-Reshef et al., 2010; Куссуль и др., 2008; Савин и др., 2010). Для прогнозирования урожайности применяются различные методы, в том числе визуальная оценка состояния полей, статистическая обработка, поиск по году-аналогу, модели роста и регрессионные подходы (Chipanshi et al., 1999; Doraiswamy, Cook, 1995; Maselli, Rembold, 2001; Pinter et al., 1981; Wall et al., 2007). В частности, вегетационный индекс NDVI активно используется для мониторинга сельскохозяйственных культур и прогнозирования урожайности с начала 1980-х годов (Tucker et al., 1980).

Для оценки урожайности сельскохозяйственных культур используются различные подходы, в том числе биофизические модели, в которых в качестве входных параметров применяются данные дистанционного зондирования с целью калибровки модели и ее настройки. Основным недостатком таких моделей является необходимость задания многочисленных специфических параметров для каждой культуры: характеристик почвы, методов обработки культур, агрометеорологических данных, сроков посева. Среди таких моделей следует отметить CERES, WOFOST, CROPSYST и STICS. Данные дистанционного зондирования наиболее часто применяются для построения регрессионных моделей (Wall et al., 2007), описывающих соотношение между различными вегетационными индексами статистическими данными по урожайности. Эти методы, как правило, просты, и их реализация не требует задания многочисленных параметров. Основным недостатком эмпирических регрессионных методов является то, что соотношения между урожайностью и вегетационными индексами, как правило, строятся для отдельных районов или областей, причем зачастую для ограниченных выборок данных, поэтому такие модели не являются рабочими и не распространяются на другие области (Doraiswamy et al., 2003; Moriondo et al., 2007). Тем не менее, индукционный подход и основанные на нем регрессионные модели являются наиболее распространенными в силу ограниченных требований к задаче и простоты в применении. Именно этот подход будет использован в данной статье. Отличительной особенностью предлагаемого подхода является попытка построить унифицированную технологию автоматизированной обработки спутниковых данных, позволяющую прогнозировать урожайность не только озимой пшеницы, как это делают большинство авторов научных публикаций, но и групп яровых и поздних яровых культур.

Алгоритм обработки данных для построения регрессионной модели прогнозирования урожайности

В качестве обучающей выборки для настройки параметров регрессионной модели используются значения индекса NDVI, вычисляемые по данным прибора MODIS, которые в автоматическом режиме загружаются с сервера USGS и статистические данные по урожайности для отдельных районов Винницкой области и области в целом. Обобщенный алгоритм обработки спутниковых данных представлен на рис. 1, где проиллюстрированы 2 подхода к построению модели: с учетом масок посевов по основным группам культур (озимые, яровые и поздние яровые) и без такого учета.

В простейшем случае для построения прогнозной модели урожайности сельскохозяйственных культур по временному ряду NDVI данные об урожайности напрямую сопоставляются со средним NDVI по заданной территории (району). Процедура построения прогнозной регрессионной модели представлена левой ветвью на рис. 1. При этом возникает очевидная проблема смешения сигналов NDVI от различных типов поверхностей. Основной вклад в суммарный сигнал вносят посевы сельскохозяйственных культур (озимые, яровые, поздние яровые), а также естественная растительность (леса, луга, водно-болотные объекты).

Поскольку указанные типы растительного покрова имеют пересекающуюся во времени динамику развития, использовать средний NDVI в качестве предиктора урожайности имеет смысл только в самом начале вегетационного сезона для прогнозирования урожайности озимых культур либо в конце сезона для прогнозирования урожайности поздних яровых, что ограничивает его применимость.

Для преодоления недостатков использования временных рядов NDVI с усреднением их по площади в данной работе предпринята попытка автоматизированного построения масок посевов сельскохозяйственных культур на основе кластеризации временных рядов NDVI. Каждому полученному кластеру экспертным путем сопоставлялся один из классов растительного покрова: земли не сельскохозяйственного назначения, озимые, яровые и поздние яровые культуры. Процедура автоматизированной обработки данных для этого случая представлена правой ветвью алгоритма на рис. 1.



Рис. 1. Алгоритм обработки данных для построения регрессионной модели прогнозирования урожайности по спутниковым данным

Построение маски озимых культур методом k-средних

Сигнал NDVI для озимых культур и лесов достаточно хорошо отделяется от сигналов других классов, поэтому для построения маски озимых оказалось достаточно использования простейших методов кластеризации, например метода k-средних.

Объектом исследования в задаче оценки урожайности сельскохозяйственных культур по данным MODIS была выбрана Винницкая область Украины ввиду ее значительной территориальной протяженности и нахождения сразу в нескольких климатических зонах. Число кластеров в задаче кластеризации задавалось равным 4, при меньшем числе кластеров в ряде случаев не удавалось выделить класс озимых, а увеличение количества кластеров приводило только к распределению рядов NDVI других классов (не озимых) по дополнительным кластерам. Кластеризация и построение маски выполнялось для периода 2000-2009 гг.

На рис. 2 показаны центры кластеров (в данном случае это «усредненный» временной ряд NDVI для различных типов покрова) для четырех лет: 2001 (типичный год, средняя урожайность), 2003 (экстремально низкая урожайность в результате массового вымерзания озимых), 2007 (средняя урожайность, засуха в летний период) и 2008 (экстремально высокая урожайность). В большинстве случаев динамика NDVI для озимых культур имеет типичный вид, однако в 2003 г. временной ряд NDVI озимой пшеницы не имеет ярко выраженного максимума в весенний период, что может быть связано с малой площадью озимых в этом году и слиянием класса озимых с классами естественной растительности.

На рис. 3 показаны графики зависимости от времени коэффициента линейной корреляции урожайности озимой пшеницы со средним NDVI и с NDVI, усредненным по маске озимых культур. Видно, что усреднение NDVI с использованием маски озимых приводит к увеличению коэффициента корреляции для большинства периодов съемки. Для среднего NDVI максимальное значение коэффициента корреляции наблюдается в конце марта, что соответствует возобновлению вегетации озимых культур, и в дальнейшем корреляция уменьшается. Это связано с развитием биомассы других типов растительности, что вносит шум в сигнал NDVI озимых культур. Для NDVI, усредненного по маске озимых культур, максимум корреляции наблюдается в более позднее время (середина мая), что приблизительно соответствует периоду максимума зеленой биомассы озимой пшеницы. Тем самым маскирование озимых культур позволило снизить помехи в NDVI, привносимые остальными типами растительного покрова.

В результате использования масок вместо простого усреднения удалось повысить лучшее значение коэффициента корреляции на 15% (с 0,58 до 0,65).

Анализ данных и выбор предикторов регрессионной модели

По данным MODIS по маске озимой пшеницы для каждого из 27-ми районов Винницкой области были рассчитаны средние значения NDVI на промежутке времени с 2000 по 2009 год с шагом в 16 дней (1, 17, ..., 353). Гидрометцентром Украины были предоставлены данные по урожайности для каждого из районов за 10 лет.

Для регрессионного анализа этих временных рядов и построения математических моделей необходимо определить момент времени, когда корреляция между NDVI и урожайностью достигает наибольшего значения. Как показали исследования, графическое представление которых можно увидеть на рис. 4, наиболее высокая корреляция между NDVI и урожайностью наблюдается на 145 день года ± 16 дней (129, 145 и 161 дни соответственно). Поскольку максимальная корреляция между значениями NDVI и урожайности в зависимости от метеорологических условий приходится на разные дни года (рис. 4), проанализируем возможность использования в качестве предикторов усредненное значение NDVI за фиксированный день года и максимальное значение NDVI. Это будет сделано в следующем разделе.

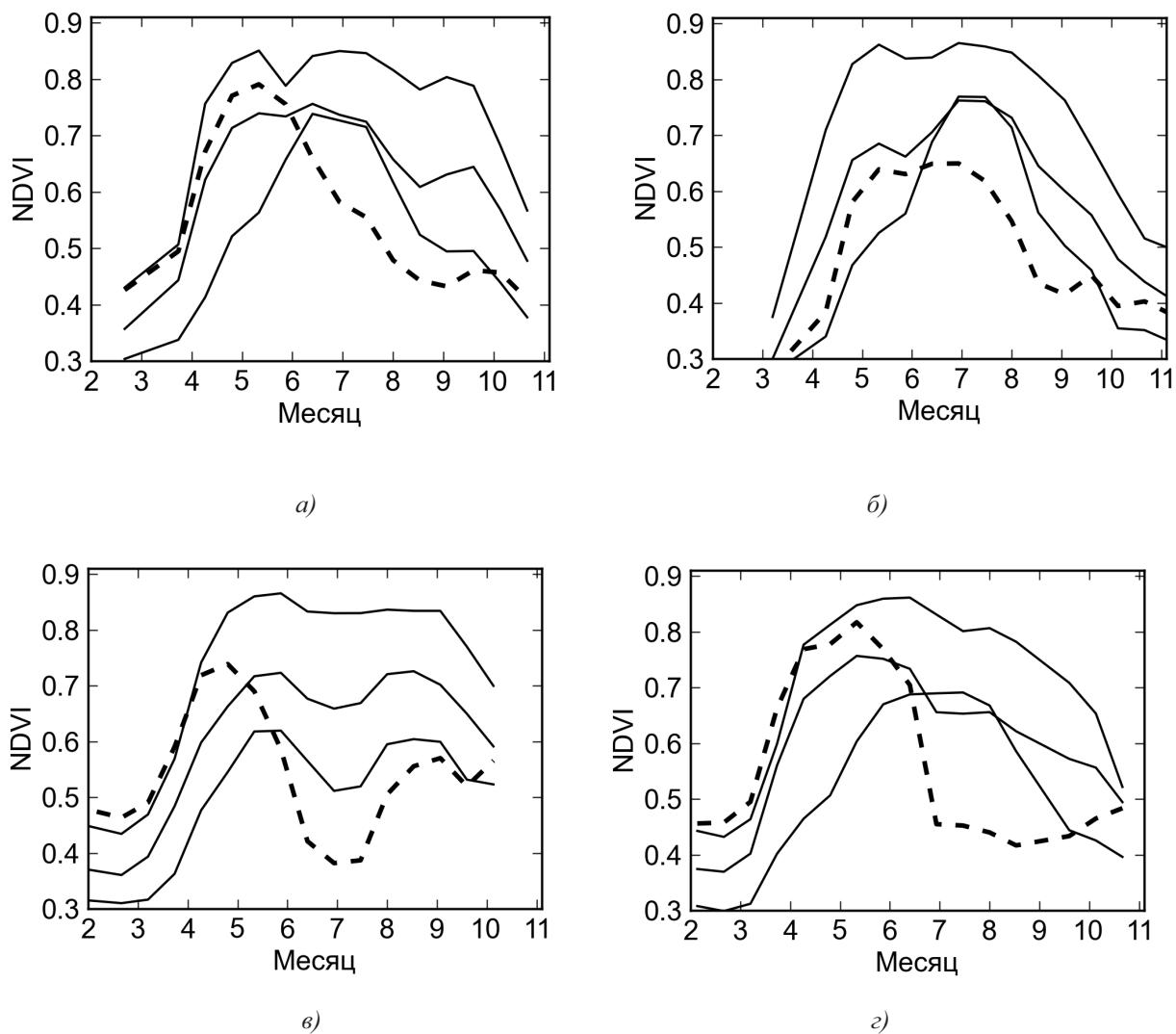


Рис. 2. Годовая динамика NDVI в Винницкой обл. за различные годы:
а – 2001 г., б – 2003 г., в – 2006 г., г – 2008 г.

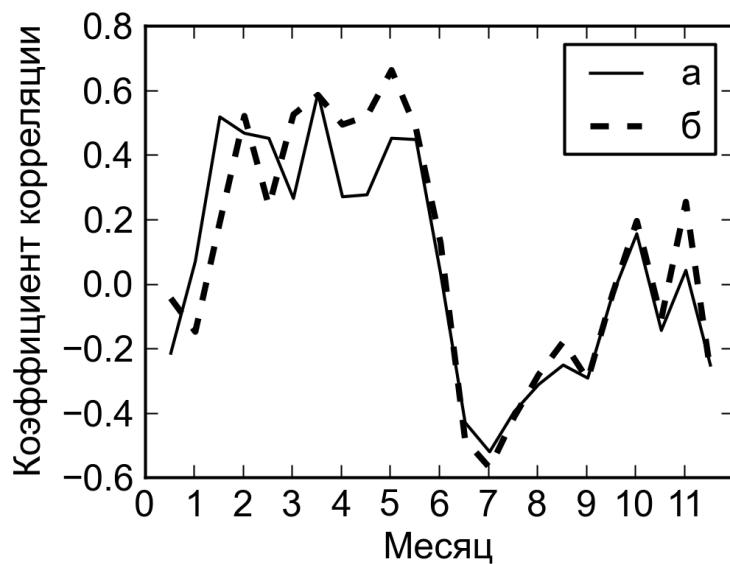


Рис. 3. График зависимости корреляции NDVI за отдельный 16-дневный период:
а – средний NDVI, б – средний NDVI по маске озимых культур

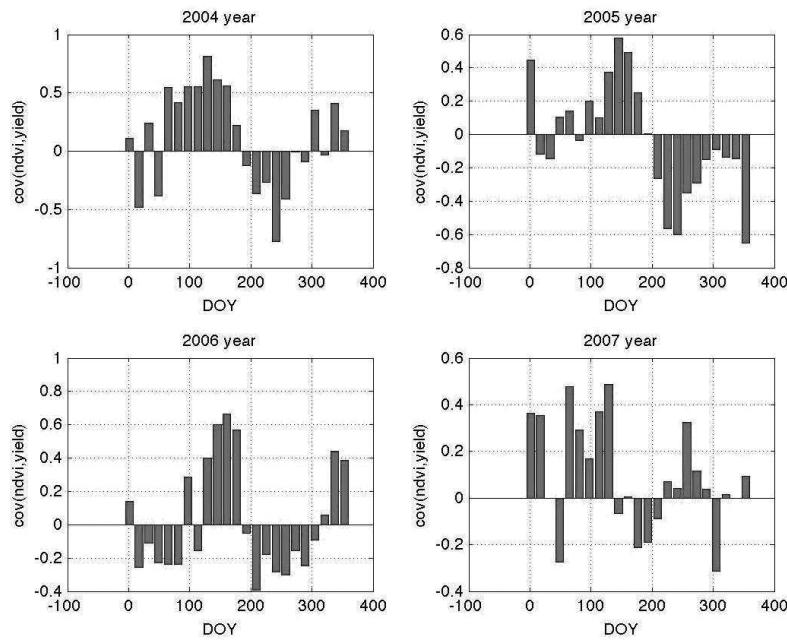


Рис. 4. Корреляция NDVI и урожайности за 2004-2007 годы

Построение и анализ регрессионных моделей зависимости урожайности озимой пшеницы от NDVI

С целью выявления закономерностей и зависимостей между NDVI и урожайностью были построены регрессионные модели для каждого из 27-ми районов Винницкой области Украины. Рассмотрим некоторые из них.

На рис. 5 представлена регрессионная модель зависимости урожайности озимой пшеницы от значения NDVI за 145 день года для 16-го района Винницкой области. Как видно из рис. 5, данные по урожайности для этого района являются достаточно согласованными. Модель характеризуется следующими параметрами:

$$\text{Yield} = 12.4328 * \text{NDVI} - 6.5625$$

$$R^2 = 0.73.$$

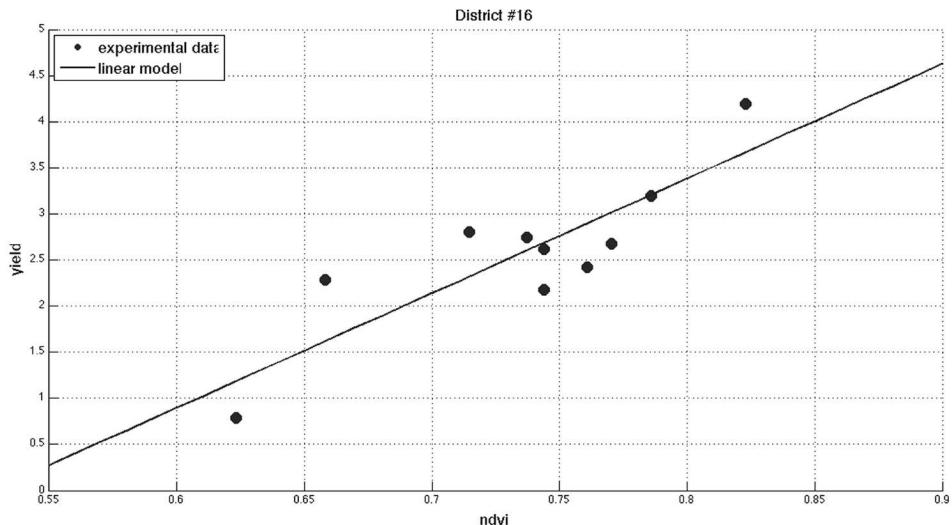


Рис. 5. Модель урожайности для 16-го района Винницкой области

Отметим, что не для всех районов регрессионные модели обладают столь хорошими статистическими характеристиками. Например, для 8-го района данные более противоречивы: для одного и того же значения NDVI в разные годы урожайность отличается почти вдвое (рис. 6). Это отражается и на статистических параметрах модели:

$$\text{Yield} = 9.5249 * \text{NDVI} - 4.1004$$

$$R^2 = 0.23$$

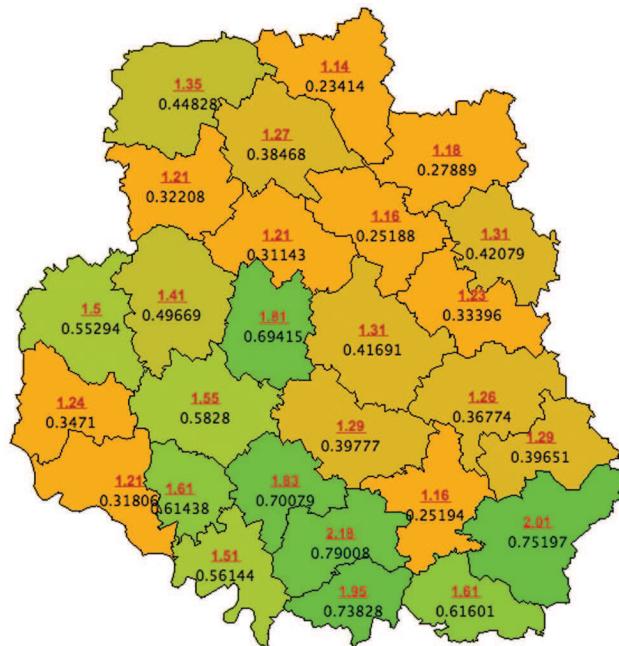


Рис. 6. Пространственное распределение точности моделирования урожайности по районам Винницкой области (предиктор — значение NDVI для фиксированного дня года)



Рис. 7. Пространственное распределение точности моделирования урожайности по районам Винницкой области (предиктор — максимальное значение NDVI за вегетационный период)

Аналогичные модели были построены для всех районов Винницкой области. Поскольку в зависимости от метеорологических условий от года к году максимальные значения NDVI приходятся на разные дни, была проанализирована возможность построения регрессионной модели урожайности, опираясь на максимальные значения NDVI. На рис. 6–7 показано пространственное распределение полученных регрессионных моделей по районам Винницкой области. Черным показаны значения статистик R^2 по всем районам, красным – эффективность построенной модели, мерой которой является отношение среднеквадратического отклонения для временных рядов урожайности к среднеквадратическому отклонению результатов регрессионной модели. Как видно из рис. 6–7, лучшие статистические характеристики моделей получены для районов, сосредоточенных в южной части области. Сравнивая статистические показатели моделей для разных предикторов (значение NDVI для фиксированного дня года на рис. 6 и максимальное значение NDVI за вегетационный период на рис. 7), отметим, что во втором случае (рис. 7) модели обладают худшими статистическими характеристиками (более высокое среднеквадратическое отклонение прогнозируемой урожайности, более низкое отношение среднеквадратического отклонения для временных рядов урожайности к среднеквадратическому отклонению результатов модели). Следовательно, в качестве предиктора регрессионной модели целесообразно использовать значение индекса NDVI за фиксированную дату (145 день), когда NDVI находится в окрестности максимума.

Анализ точности моделей урожайности для отдельных районов и Винницкой области в целом

В статистике важную роль играет прогнозирование урожайности сельскохозяйственных культур не для отдельного района, а для области в целом. Для оценки урожайности по области можно использовать два подхода:

1) Оценить значение урожайности культуры по области, используя в качестве обучающей выборки для настройки коэффициентов регрессии значения NDVI и урожайности для каждого района;

2) Построить регрессионную модель на основе усредненных значений NDVI и статистических данных урожайности по области.

Проанализируем точность прогнозирования урожайности для каждого из подходов. В **первом** случае для идентификации параметров линейной регрессионной модели воспользуемся в качестве обучающей выборки значениями NDVI и данными по урожайности по всем 27-ти районам Винницкой области за 10 лет (270 точек).

Во втором случае регрессию построим по 10-ти усредненным по области значениям NDVI и усредненным статистическим данным урожайности для области. Графическое представление данной модели приведено на рис. 9.

$$Yield = 11.5509 * NDVI - 5.5803$$

$$R^2 = 0.43$$

Ее графическое представление приведено на рис. 8.

Во втором случае регрессию построим по 10-ти усредненным по области значениям NDVI и усредненным статистическим данным урожайности для области. Графическое представление данной модели приведено на рис. 9.

Модель, построенная на основе усредненных обучающих данных по области, обладает несколько лучшими статистическими характеристиками. Таким образом, ввиду меньшего объема необходимых данных и незначительного уменьшения разброса, более предпочтительным является использование второго подхода к построению модели – настройка параметров регрессии на основе усредненных по области обучающих данных.

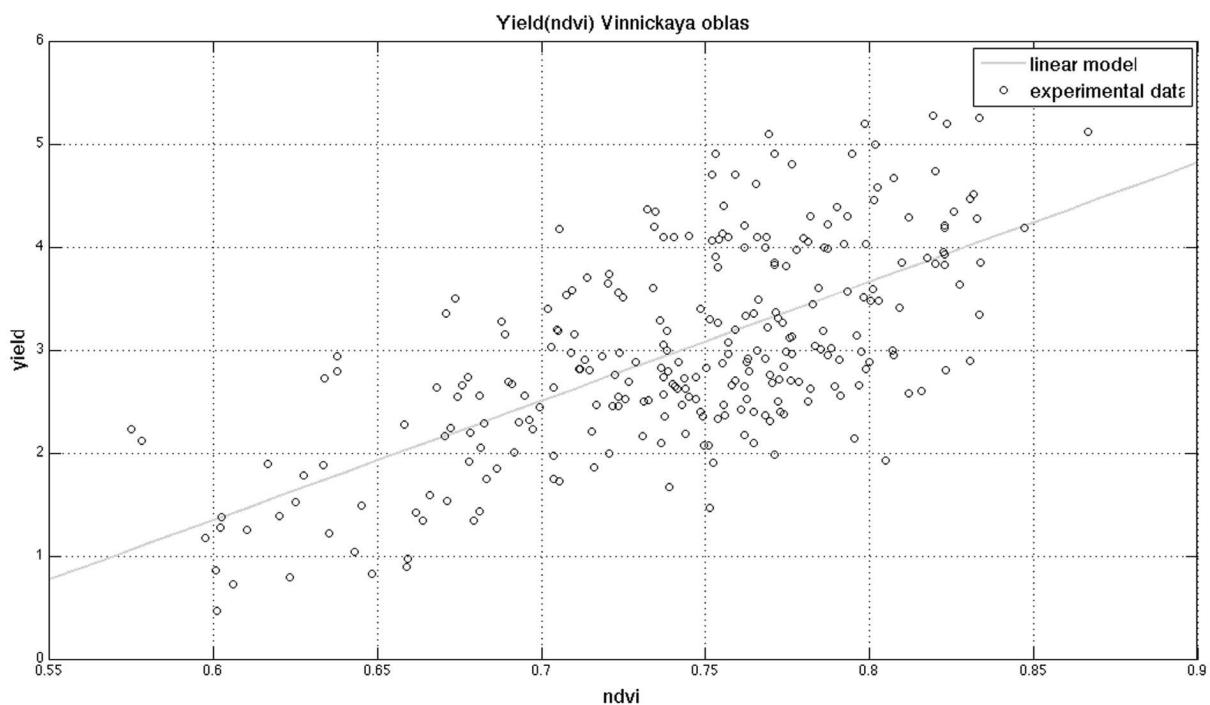


Рис. 8. Регрессионная модель прогнозирования урожайности по области на основе обучающих данных по отдельным районам

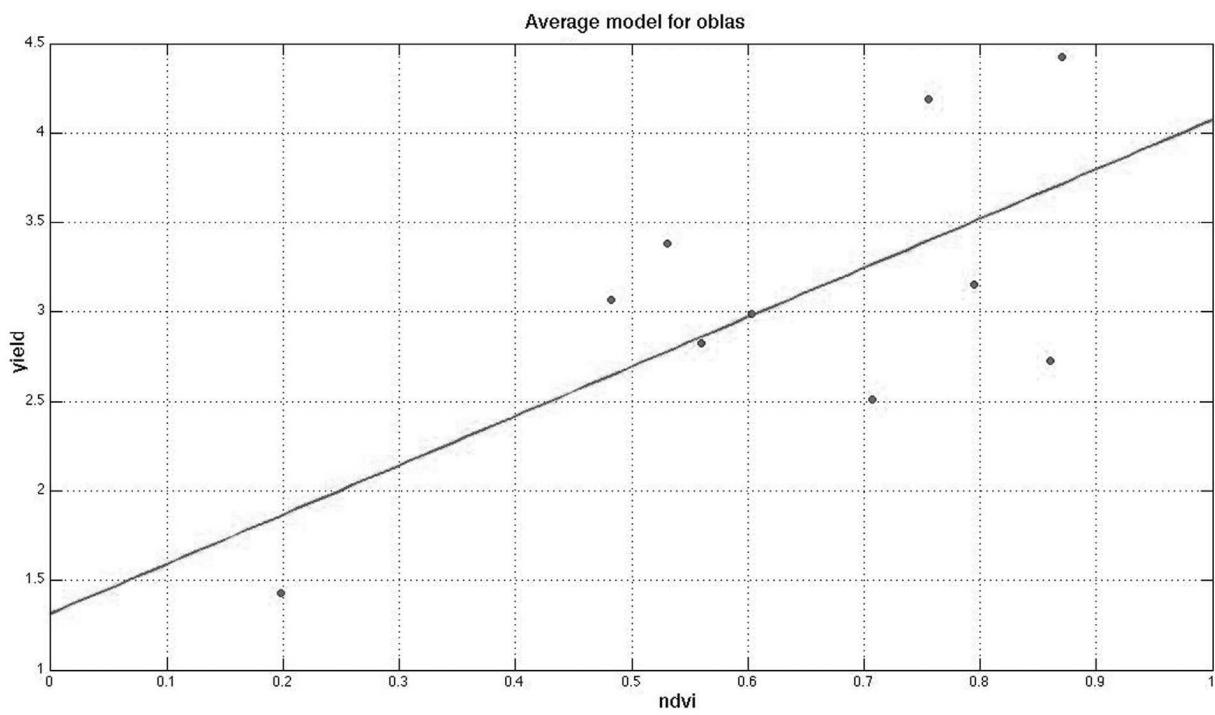


Рис. 9. Общая модель урожайности по области, построенная на основе усредненных значений NDVI и данных урожайности по области

Построение регрессионной модели прогнозирования урожайности яровых культур

Описанная выше технология построения регрессионной модели для прогнозирования урожайности озимой пшеницы может быть распространена на другие виды сельскохозяйственных культур, а именно, яровые и поздние яровые культуры.

Рассмотрим задачу построения регрессионной модели зависимости урожайности яровых культур от значений NDVI по районам Винницкой области за 2009 год. Аналогично описанному выше подходу прогнозирования урожайности озимой пшеницы, для построения прогнозной модели будем использовать 16-дневные композиты вегетационного индекса NDVI, вычисленные на основе данных прибора MODIS. Сначала строятся маски яровых культур для каждого района. Для анализа временных рядов NDVI и идентификации яровых воспользуемся методом кластеризации на основе нейронных сетей (НС) Кохонена (Haykin, 1999; Kussul et al., 2008). Самоорганизующиеся карты Кохонена – это специальный класс искусственных НС, работа которых основана на конкурентном принципе обучения. НС данного типа представляют собой одну из реализаций интеллектуального подхода к анализу данных и позволяют автоматически выявить статистически значимые свойства входных данных. После автоматической кластеризации данных выполним постобработку (калибровка или классификация) результатов с использованием наземных данных и визуального инспектирования НС – соответствующие кластеры отнесем к классу яровые.

Для оценки точности классификации выполним сравнение площадей, вычисленных по спутниковым данным, и статистических данных о посевных площадях ярового ячменя (который имеет наибольший удельный вес среди яровых в Украине) для каждого из районов. По результатам сравнения смещение составляет 26%. Погрешность определения посевных площадей для яровых культур по всей Винницкой области составила 10%. Результаты сравнения показаны на рис. 10.

По полученной маске яровых для каждого района вычислим среднее значение NDVI. В качестве предиктора модели выберем максимальное значение NDVI за вегетационный период. Результаты построения модели представлены на рис. 11. Как видно из рис. 11, полученная регрессионная модель для прогнозирования урожайности ярового ячменя обладает достаточно хорошими статистическими характеристиками и может быть использована в автоматизированных системах поддержки принятия решений. Средняя абсолютная ошибка и стандартное отклонение прогнозирования урожайности по районам Винницкой области составили, соответственно, 0,29 т/га и 0,23 т/га. Средняя относительная ошибка по районам составила 13,8%. Если оценить урожайность для всей области (на основе площадей, полученных по спутниковым данным, и урожайности по районам, и последующей агрегации валового сбора для всей области), абсолютная ошибка составила 0,05 т/га, или 1,9%. Можно видеть, что при переходе от районов к области ошибка прогнозирования урожайности существенно уменьшается (в данном случае от 13,8% до 1,9%).

Выводы

Таким образом, в работе предпринята попытка построить информационную технологию прогнозирования урожайности на основе индукционного подхода. В качестве модели выбрана регрессионная модель первого порядка, а в качестве регрессора – 16-дневный композит индекса NDVI за фиксированный период, соответствующий максимальному значению NDVI, вычисленный по маске соответствующей культуры. Маска определяется автоматически методом кластеризации (k -средних или карт Кохонена) временных рядов NDVI.

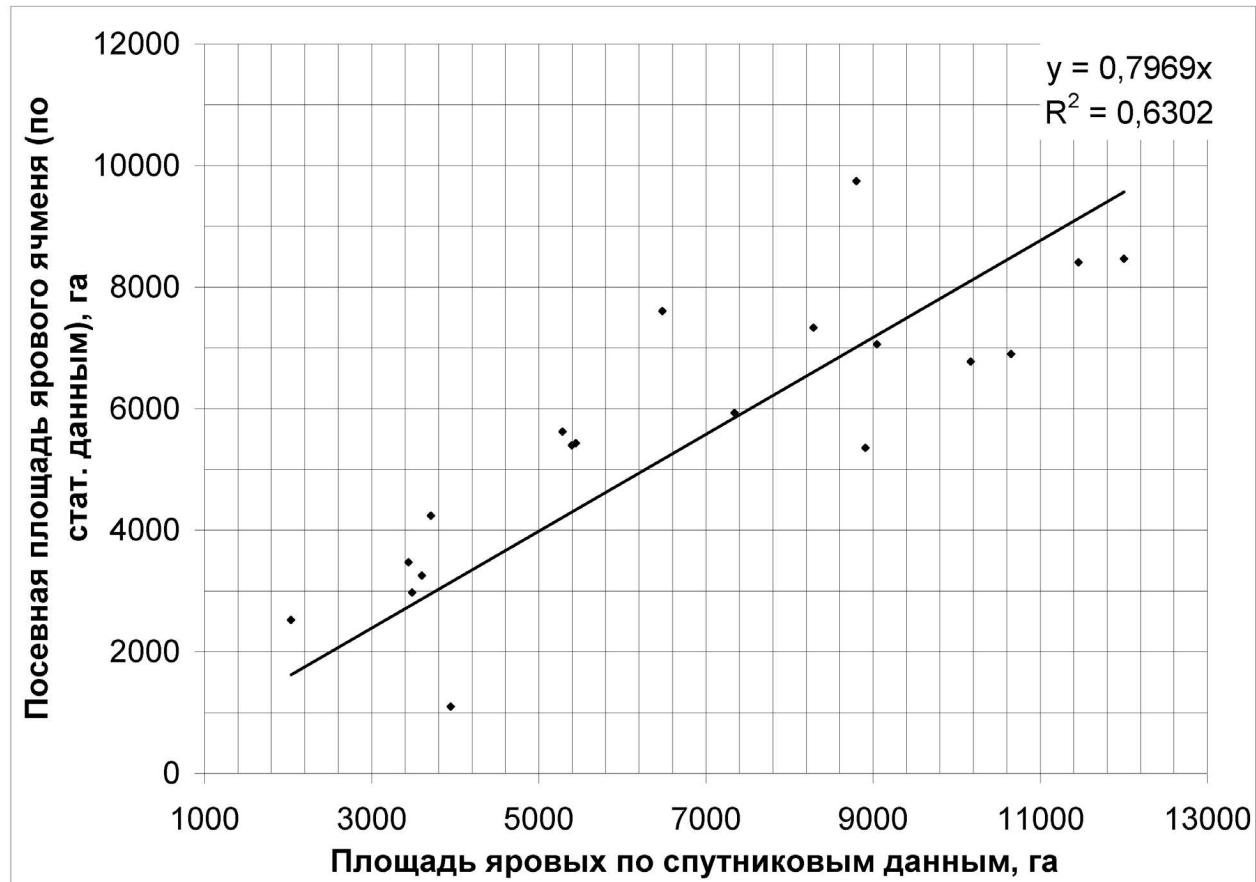


Рис. 10. Сравнение посевных площадей яровых культур по районам, определенных на основе спутниковой информации и данных статистики

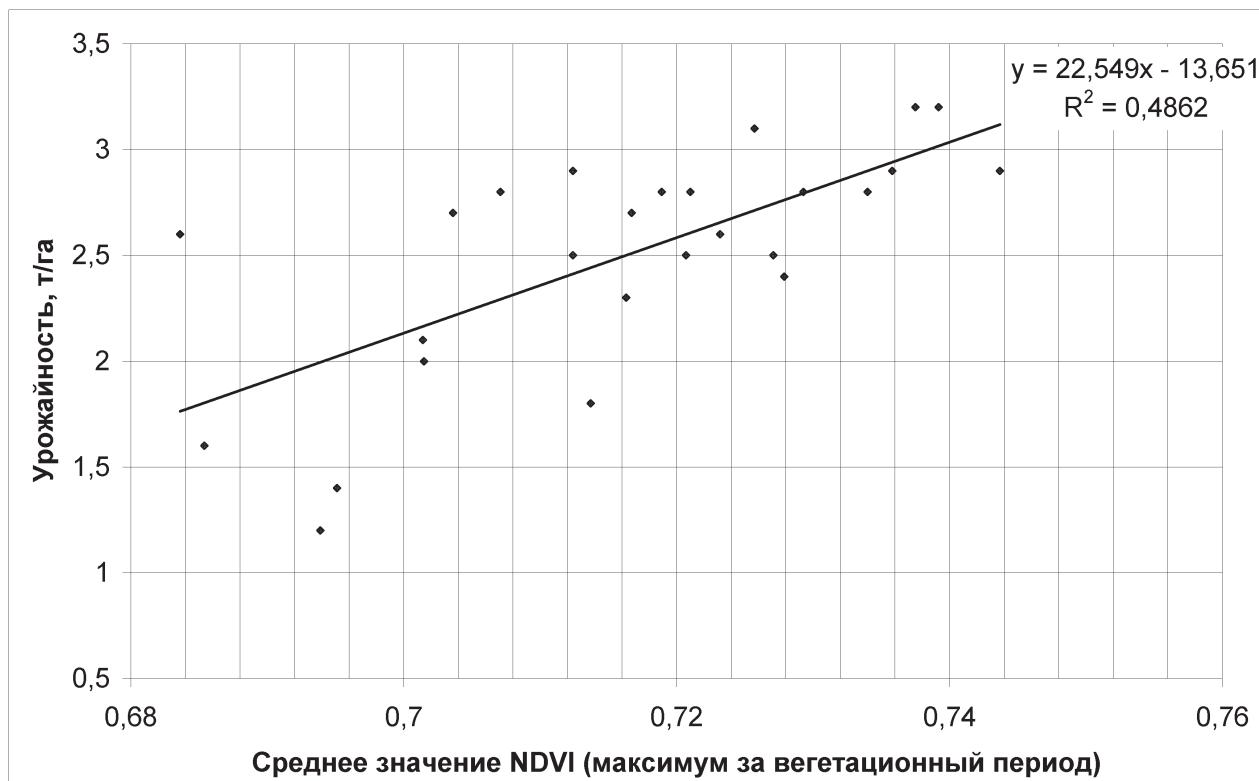


Рис. 11. Регрессионная модель прогнозирования урожайности по максимуму NDVI

Показано, что точность регрессионной модели прогнозирования на основе данных по области в целом несколько выше, чем точность прогноза, построенного путем усреднения прогнозных значений по отдельным районам. Это можно объяснить неточностью статистических данных для отдельных районов, которая нивелируется при усреднении данных по области. Для повышения точности прогнозирования на уровне отдельных районов необходимо учитывать «выбросы» (противоречивость) данных, вносящие шум в достаточно малочисленную обучающую выборку (10 обучающих векторов для каждого района), и отфильтровывать такие данные. Анализ исходных статистических данных свидетельствует о том, что такая противоречивость статистических данных характерна для некоторых районов и никак не связана с площадью посевов. Она может быть связана с влиянием на урожайность не учтенных в модели факторов (например, метеорологических условий в период созревания урожая), неполным соответствием индекса NDVI объему биомассы и «человеческим фактором», т.е. ошибками в статистических показателях (что не удивительно, поскольку процесс сбора статистики в Украине до сих пор не автоматизирован). Вычислительный эксперимент показывает, что фильтрация таких шумов («выбросов» в данных) существенно повышает статистические характеристики и рабочесть модели. Для учета дополнительных факторов, влияющих на урожайность, или «нивелирования» недостатков выбранного в качестве предиктора модели индекса NDVI в дальнейшем предполагается использовать ансамблевый метод и при построении прогноза урожайности учитывать результаты нескольких разных моделей.

Точность прогноза урожайности по построенной модели согласуется с результатами (Becker-Reshef et al., 2010). Важным преимуществом предлагаемого подхода является его применимость не только для автоматизированного прогнозирования урожайности озимой пшеницы, но и других видов сельскохозяйственных культур. Разработанную технологию предполагается в дальнейшем использовать в геоинформационных системах поддержки принятия решений как на уровне министерства, так и на уровне отдельного хозяйства (Куссуль и др., 2010; Шелестов и др., 2011).

Работа выполнена при поддержке гранта CRDF “Analysis of Climate Change & Food Security based on Remote Sensing & In Situ Data Sets” (№ UKB2-2972-KV-09).

Литература

1. 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.
2. Chipanshi A.C., Ripley E.A., Lawford R.G. Large-scale simulation of wheat yields in a semi-arid environment using a crop-growth model // *Agricultural Systems*, 1999. 59. P. 57–66.
3. Doraiswamy P.C., Cook P.W. Spring wheat yield assessment using NOAA AVHRR data // *Canadian Journal of Remote Sensing*, 1995. 21. P. 43–51.
4. Doraiswamy P.C., Moulin S., Cook P.W., Stern A. Crop yield assessment from remote sensing // *Photogrammetric Engineering and Remote Sensing*, 2003. 69. P. 665–674.
5. Haykin S. *Neural Networks: A Comprehensive Foundation* // Upper Saddle River, New Jersey: Prentice Hall, 1999.
6. Kussul N., Shelestov A., Skakun S. Intelligent Computations for Flood Monitoring // International Book Series «Advanced Research in Artificial Intelligence» (ed. Markov K., Ivanova K., Mitov I.), 2008. N 2. P. 48–54.
7. Maselli F., Rembold F. Analysis of GAC NDVI data for cropland identification and yield forecasting in Mediterranean African countries // *Photogrammetric Engineering and Remote Sensing*, 2001. 67. P. 593–602.
8. Manjunath K.R., Potdar M.B., Purohit N.L. Large area operational wheat yield model development and validation based on spectral and meteorological data // *International Journal of Remote Sensing*. 2002. 23. P. 3023–3038.

9. Moriondo M., Maselli F., Bindi M. A simple model of regional wheat yield based on NDVI data // European Journal of Agronomy, 2007. 26. P. 266–274.
10. Pinter P.J., Jackson R.D., Idso S.B., Reginato R.J. Multidate spectral reflectances as predictors of yield in water stressed wheat and barley // International Journal of Remote Sensing, 1981. 2. P. 43–48.
11. Prasad A.K., Chai L., Singh R.P., Kafatos M. Crop yield estimation model for Iowa using remote sensing and surface parameters // International Journal of Applied Earth Observation and Geoinformation, 2006. 8. P. 26–33.
12. Tucker C.J., Holben B.N., Elgin J.H., McMurtrey J.E. Relationships of spectral data to grain yield variation // Photogrammetric Engineering and Remote Sensing, 1980. 46. P.657-666.
13. Wall L., Larocque D., Leger P.M. The early explanatory power of NDVI in crop yield modelling // International Journal of Remote Sensing, 2007. 29. P. 2211–2225.
14. Куссуль Н.Н., Соколов Б.В., Зельк Я.И., Зеленцов В.А., Скаакун С.В., Шелестов А.Ю. Оценка рисков стихийных бедствий на основе разнородной геопространственной информации // Проблемы управления и информатики, 2010. N 6. С. 97–110.
15. Куссуль Н.Н., Ильин Н.И., Скаакун 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» (ed. Markov K., Ivanova K., Mitov I.), 2008. N 3. P. 103–109.
16. Савин И.Ю., Барталев С.А., Лупян Е.А., Толчин В.А., Хвостиков С.А. Прогнозирование урожайности сельскохозяйственных культур на основе спутниковых данных: возможности и перспективы // Современные проблемы дистанционного зондирования Земли из космоса, 2010. Т.7. № 3. С. 275–285.
17. Шелестов А.Ю., Кравченко А.Н., Волошин С.В., и др. Web-портал системы агромониторинга // Наука и инновации, 2011. № 3.

Crop Yield Forecasting Regression Models based on MODIS Data

N.N. Kussul¹, A.N. Kravchenko¹, S.V. Skakun¹, T.I. Adamenko³, A.Yu. Shelestov²,
A.V. Kolotii¹, Yu.A. Gripich¹

¹ Space Research Institute NASU-NSAU
03680 Kyiv, 40 Glushkov Prospekt, build. 4/1
E-mails: inform@ikd.kiev.ua

² National University of Life and Environment Sciences of Ukraine
03187 Kyiv, 15 Geroev Oborony str.
E-mail: andrii.shelestov@gmail.com

³ Ukrainian Hydrometeorological Center
01034, Kyiv, 6-V Zolotovoritska str.
E-mail: adamenko@meteo.gov.ua

Information technology for crop yield forecasting model creation is proposed on the base of satellite data and inductive approach. A first order regression model is used with 16-days composite of maximum NDVI for crop mask as a predictor. A crop mask is built automatically using clusterization of an NDVI time series (using k-means or Kohonen maps). It is shown that accuracy of the forecasting model for the entire region is better than for separate districts. The proposed approach can be used not only for winter wheat crop yield forecasting, but also for other crops, in particular spring crops

Keywords: Crop yield forecasting, regression model, crop mask, information technology, MODIS, NDVI, clusterization, Kohonen map.