

Аппроксимация временных рядов NDVI в задаче раннего распознавания видов сельскохозяйственных культур по космическим снимкам

Н.С. Воробьева^а, А.В. Чернов^а

^а Самарский национальный исследовательский университет имени академика С.П. Королёва, 443086, ул. Луначева, 39Б, Самара, Россия

Аннотация

Работа посвящена выбору функции аппроксимации временных рядов NDVI в задаче раннего распознавания сельскохозяйственных культур по космическим снимкам. Анализируются аппроксимация кусочно-линейной, асимметричной гауссовой, двойной логистической функциями, рядом Фурье, полиномом и сплайном. Количество свободных параметров функций варьируется от 3 до 10. Показано, что наилучшей функцией аппроксимации является сплайн третьего порядка. В работе оценивается возможность его использования для построения модели временного ряда, и приводятся примеры генерации временных рядов.

Ключевые слова: временные ряды; NDVI; модель временного ряда; космические снимки; распознавание сельскохозяйственных культур; аппроксимация временного ряда

1. Введение

Распознавание видов сельскохозяйственных (далее – с/х) культур по космическим снимкам является важной задачей мониторинга земель сельскохозяйственного назначения. Результаты распознавания применяются для контроля использования субсидий, выделяемых с/х производителям, учета и контроля использования сельскохозяйственных земель, выявления неиспользуемых земель в течение ряда лет, а также контроля севооборота [1, 2].

В данной работе будем говорить о распознавании на территориях масштаба региона. Методы, предлагаемые в литературе для распознавания на территориях такой площади, основаны на использовании временных рядов [3,4,5], построенных для объектов исследования (полей). Временной ряд – это упорядоченный по дате получения космических снимков набор значений некоторого параметра (спектрального канала или вегетационного индекса), рассчитанный для объекта исследования (поля). Вегетационный индекс – вторичный (производный) продукт, получаемый в результате математических операций, выполняемых на каналах космического снимка [6]. Наиболее популярный из всех вегетационных индексов – NDVI (Normalized Difference Vegetation Index) – характеристика, рассчитываемая по значениям в красном ρ_{RED} и инфракрасном ρ_{NIR} каналах спектра:

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

Типичный набор исходных данных в методах распознавания на территориях большой площади включает границы полей, данные о посеянных культурах на небольшом количестве полей и набор космических снимков. Распознавание выполняется по набору данных, полученных за весь посевной сезон или его часть. Посевной сезон – временной промежуток, соответствующий периоду от начала посевных работ до окончания уборочных работ.

Наиболее важным является раннее распознавание (распознавание в начале посевного сезона) позволяющее быстро принять меры наземного контроля и оперативного управления. Раннее распознавание характеризуется малым количеством исходных данных, недостаточных для формирования обучающей выборки и невозможностью использования для распознавания методов [3,4,5], предлагаемых в литературе.

Для решения задачи раннего распознавания в работе [7] предложена информационная технология, позволяющая генерировать обучающую выборку по данным прошлых лет и далее использовать сгенерированные данные для распознавания культур текущего года. Технология состоит из двух этапов. На первом этапе по данным за прошедшие годы формируется набор моделей. На втором этапе производится работа с данными текущего года с обращением к набору моделей первого этапа. Модели используются для генерации обучающей выборки, состоящей из временных рядов, что позволяет провести раннее распознавание культур текущего сезона.

Одной из таких моделей, используемых на первом этапе, является *модель временного ряда*, представляющая собой некоторую функцию $g(c_1, \dots, c_m, x)$, заданную аналитически, зависящую от вектора детерминированных параметров $\mathbf{c} = (c_1, \dots, c_m)$ и описывающую временной ряд, рассчитанный для заданного поля в заданном году по набору снимков. Использование вместо значений временного ряда его модели позволяет избавиться от зашумленности, обусловленной такими явлениями как атмосферные явления, облачность, природно-климатические условия. На рисунке 1 приведен пример зашумленного (исходного) временного ряда и его модели. По оси абсцисс отложена относительная дата, рассчитанная от 1 марта, по оси ординат – значение NDVI.

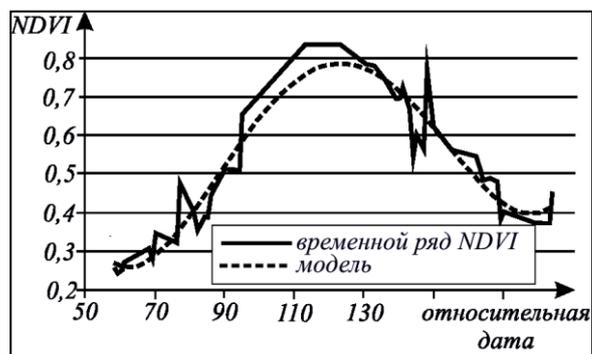


Рис. 1. Пример зашумленного и сглаженного временных рядов.

В данной работе решается задача поиска вида функции $g(c_1, \dots, c_m, x)$, наилучшим образом описывающую заданную совокупность временных рядов и позволяющую генерировать обучающую выборку для использования в методах распознавания видов с/х культур по космическим снимкам.

В литературе для моделирования и генерации временных рядов в задачах распознавания видов с/х культур, предлагаются следующие методы:

1. Синтезирование изображений, имитирующих величину отраженного излучения в различных спектральных диапазонах, расчет по ним временных рядов [8,9,10] и дальнейшее зашумление полученных временных рядов.

Недостатки метода: не учитываются особенности развития культур на конкретной территории в конкретном году, связь моделируемых временных рядов и реальных наземных измерений отсутствует. Использовать метод для задачи, решаемой в данной работе, невозможно.

2. Непосредственное моделирование кривой NDVI:

- a. Описание временного ряда двойной логистической функцией [10,11,12]. В предлагаемых работах поля с одинаковым видом культуры имеют схожие профили NDVI и описываются двойной логистической функцией с фиксированными значениями параметров.

Недостатки метода: напрямую предлагаемый метод невозможно применить, так как он не учитывает ежегодной изменчивости профилей NDVI, связанной с влиянием погодных условий.

- b. Формирование «идеальных» кривых по набору космических снимков за продолжительный промежуток времени [13] и их использование для распознавания. Перед распознаванием каждая кривая корректируется по реальным наземным данным текущего года – датам наступления фенологических фаз. Фенологическая фаза – это конкретный этап развития растения (всходы, кушение и т.д.).

Недостатки метода: для использования данного метода необходима дополнительная информация о датах наступления фенологических фаз на каждом поле региона. Такая информация отсутствует, поэтому предлагаемый метод не может быть использован.

- c. Сложные математические модели динамики NDVI, требующие дополнительных параметров (например, данных о температуре и влажности воздуха) [14].

Недостатки метода: такие модели невозможно применить для решения задачи, рассматриваемой в работе, так как дополнительные данные, требуемые для построения моделей, отсутствуют.

Как видно из представленного краткого обзора, ни один из предлагаемых в литературе методов не может быть использован в информационной технологии для моделирования и генерации временных рядов, что говорит об актуальности проводимых исследований.

2. Исходные данные

Для решения задачи, рассматриваемой в данной работе, использовались данные, полученные на территории Самарской области за 2011–2016 годы: наземные данные о полях и набор космических снимков.

Границы полей задавались векторной картой с указанием сведений о виде (классе) засеянной с/х культуры. Список классов следующий:

1. Многолетние травы и неиспользуемые земли.
2. Озимые.
3. Пар.
4. Ранние яровые.
5. Поздние яровые.

В первый класс вошли многолетние травы, неиспользуемые земли, залежи. Во второй – озимые пшеница, рапс, рожь, ячмень. В третий – земли, занятые парами. В четвертый – яровой ячмень, овес, рожь. В пятый – подсолнечник, кукуруза, нут, однолетние травы, соя, гречиха. Такое разбиение является довольно стандартным, вопрос о разбиении культур по таким классам выходит за рамки данной работы.

Для каждого поля проводился расчет временных рядов по набору данных Terra/MODIS, обработанных до уровня продукта MOD09GQ. Продукт MOD09GQ прошел радиометрическую, атмосферную коррекцию, имеет географическую привязку и представляет собой минимальный уровень комбинирования ежедневных данных разрешения 250 м. Для расчета временных рядов использовались снимки, соответствующие весенне-летнему посевному сезону с 1 апреля по 31 августа. В расчетах учитывалась маска облачности, облачные пиксели исключались из рассмотрения.

В результате за 2011-2016 годы был сформирован набор из 32524 временных рядов, соответствующих исходным полям. В таблице 1 приведена краткая статистика по количеству временных рядов в каждом классе.

Таблица 1. Статистика исходных данных – количество временных рядов в каждом классе

Многолетние травы и неиспользуемые земли	Озимые	Пар	Ранние яровые	Поздние яровые
5063	4753	7040	6436	9232

3. Аппроксимация временного ряда

Для поиска функции $g(c_1, \dots, c_m, x)$, рассмотрим структуру значений временного ряда.

Временной ряд можно представить как набор значений дискретной величины z , измеренной в моменты времени t_i , являющейся смесью полезного сигнала y и аддитивного шума v , имеющую пропуски в ежедневных наблюдениях по причине облачности:

$$z(t_i) = y(t_i) + v(t_i), \quad (2)$$

где $i = \overline{1, N}$, N – количество точек временного ряда.

Полезный сигнал y можно представить как непрерывную функцию, заданную на некотором временном промежутке, значения которой зашумляются аддитивным шумом. Полезный сигнал и есть та самая искомая функция $g(c_1, \dots, c_m, x)$. Будем рассматривать $g(c_1, \dots, c_m, x)$ на временном интервале $[x_s, x_f]$, где x_s, x_f – начало и конец временного интервала.

Возникает вопрос о виде функции, описывающей полезный сигнал. Наиболее популярные функции, используемые в методах распознавания видов с/х культур для предварительного сглаживания, фильтрации и аппроксимации временных рядов и отражающие динамику вегетационного индекса NDVI, следующие:

- двойная логистическая [10,11,12];
- асимметричная гауссова [15,16,17];
- ряд Фурье [18,19,20,21];
- полиномы [3,16, 22, 23].

Расширим этот список сплайном 3-го порядка и кусочно-линейной функцией, ранее не использовавшимися для аппроксимации временных рядов в методах распознавания видов с/х культур. Рассмотрим задачу поиска функции, наилучшим образом аппроксимирующей временные ряды, представленные в качестве исходных данных за 2011-2016 год.

Любой временной ряд из набора исходных данных представляет собой совокупность пар значений «дата» – «значение индекса». Для удобства аппроксимации интервал дат с 1 апреля по 31 августа, включающий в себя 153 точки наблюдения и соответствующий весенне-летнему посевному сезону, переведен в диапазон $[-1,1]$. Далее примеры всех временных рядов будут приводиться именно этого диапазона.

Рассмотрим подробно выбранные функции.

3.1. Двойная логистическая

Функция имеет 7 параметров и вычисляется по формуле:

$$g(x) = vb + \frac{k}{1 + e^{-c(x-p)}} - \frac{k + vb + ve}{1 + e^{d(x-q)}}, \quad (3)$$

где vb, k, c, p, ve, d, q – параметры.

3.2. Асимметричная гауссова

Функция имеет 7 параметров и вычисляется по формуле:

$$g(x) = c_1 + c_2 \cdot f(x; a_1, \dots, a_5),$$

$$f(x; a_1, \dots, a_5) = \begin{cases} e^{-\left(\frac{x-a_1}{a_2}\right)^{a_3}}, & x > a_1, \\ e^{-\left(\frac{a_1-x}{a_4}\right)^{a_5}}, & x \leq a_1 \end{cases}, \quad (4)$$

где $c_1, c_2, a_1, \dots, a_5$ – параметры.

3.3. Ряд Фурье

Функция вычисляется по формуле:

$$g(x) = a_0 + \sum_{k=1}^M (a_k \cdot \cos(k \cdot x \cdot w) + b_k \cdot \sin(k \cdot x \cdot w)), \quad (5)$$

где $w, a_0, a_k, b_k, k = \overline{1, M}$ – параметры.

Будем считать, что период первой гармоники равен длине рассматриваемого вегетационного сезона. Для аппроксимации рассматривается диапазон дат $[-1, 1]$, его длина равна 2, поэтому $w = \pi$ и аппроксимация временных рядов будет производиться функцией следующего вида с меньшим количеством параметров:

$$g(x) = a_0 + \sum_{k=1}^M (a_k \cdot \cos(k \cdot x \cdot \pi) + b_k \cdot \sin(k \cdot x \cdot \pi)). \quad (6)$$

Будем рассматривать функцию при $M = \overline{1, 4}$ (соответственно, количество параметров составляет 3, 5, 7, 9).

3.4. Полином

Функция аппроксимации вычисляется по формуле:

$$g(x) = \sum_{k=0}^M c_k x^k, \quad (7)$$

где $c_k, k = \overline{0, M}$ – параметры (коэффициенты разложения). Будем рассматривать аппроксимацию полиномами с числом параметров от 3 до 10.

3.5. Сплайн 3-го порядка

Предлагается аппроксимировать временной ряд сплайном 3-го порядка (далее для краткости просто «сплайн»), представляющим собой линейную комбинацию M базисных функций.

Пусть на интервале наблюдения $[x_s, x_f]$ задана равномерная сетка узлов $[x_1 = x_s, x_2, \dots, x_{M-1}, x_M = x_f]$ с шагом $h = (x_f - x_s) / (M - 1)$, $M \in N\{1, 2\}$. Базисная функция вычисляется по формуле:

$$u(x) = \begin{cases} \frac{3}{2}|x|^3 - \frac{5}{2}|x|^2 + 1, & 0 < |x| < 1 \\ -\frac{1}{2}|x|^3 + \frac{5}{2}|x|^2 - 4|x| + 2, & 1 < |x| < 2 \\ 0, & 2 < |x| \end{cases}. \quad (8)$$

Функция аппроксимации в виде сплайна 3-го порядка задается так:

$$g(x) = \sum_{k=1}^M c_k u\left(\frac{x - x_k}{h}\right), \quad (9)$$

где c_k – свободные параметры, коэффициенты разложения.

Функция (8) была предложена в [24] (функция интерполяции кубической свертки). На рисунке 2 приведен ее график.

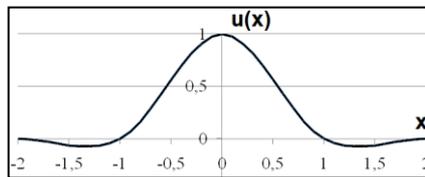


Рис. 2. Базисная функция для аппроксимации сплайном 3-го порядка.

Будем использовать функцию аппроксимации (9) с числом параметров от 3 до 10.

3.6. Кусочно-линейная функция

Включаем данный способ аппроксимации в рассмотрение, так как он наиболее прост и очевиден.

Кусочно-линейную аппроксимацию временного ряда, аналогично сплайну 3-го порядка, можно представить в виде разложения (9) по базисным функциям следующего вида:

$$u(x) = \begin{cases} 1 - |x|, & |x| < 1 \\ 0, & |x| \geq 1 \end{cases} \quad (10)$$

График базисной функции (10) приведен на рисунке 3.

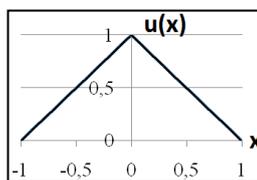


Рис. 3. Базисная функция для кусочно-линейной аппроксимации.

Будем использовать кусочно-линейную функцию аппроксимации с числом параметров от 3 до 10.

3.7. Сравнение аппроксимации временных рядов различными функциями

Проведем аппроксимацию временных рядов рассмотренными функциями и сравним результаты.

3.7.1. Критерии качества

Качество аппроксимации каждой из функций будет оцениваться по следующим критериям:

1. Статистика выполненных аппроксимаций.
2. Среднеквадратическая ошибка.
3. Максимальная ошибка.

Рассмотрим подробно приведенные критерии.

Вегетационный индекс NDVI принимает значения в диапазоне $[-1, 1]$, причем, для открытой почвы и растительности значения индекса всегда положительны. Поэтому будем считать аппроксимацию не выполненной, если значение аппроксимирующей функции вышло за диапазон $[0, 1]$. Критерий статистики выполненных аппроксимаций рассматривается для каждой конкретной функции аппроксимации без разделения на классы.

Максимальная и среднеквадратическая ошибки оцениваются для каждого класса, включающего определенное количество временных рядов (статистика приведена в таблице 1), и для всего набора временных рядов (32524 рядов).

Пусть $\varepsilon = \{\varepsilon_i, i = 1, K\}$ – невязки аппроксимации набора временных рядов (набор может включать все временные ряды или только ряды определенного класса). K – количество невязок, равное общему числу точек всех временных рядов из набора.

Среднеквадратическая ошибка аппроксимации $\varepsilon_{ско}$ вычисляется по формуле:

$$\varepsilon_{ско} = \sqrt{\frac{\sum_{i=1}^K \varepsilon_i^2}{K}} \quad (11)$$

Максимальная ошибка аппроксимации ε_{\max} набора временных рядов выбирается так, чтобы удовлетворять условию:

$$P(\varepsilon \leq \varepsilon_{\max}) = 0.99 \quad (12)$$

Среднеквадратическая ошибка показывает усредненное качество аппроксимации, максимальная ошибка позволяет отследить максимальные отклонения значений временного ряда от аппроксимирующей функции. Вероятность в (12) полагается равной 0.99, чтобы отбросить небольшое число максимальных невязок.

3.7.2. Результаты аппроксимации

На рисунке 4 приведен пример аппроксимации типичного временного ряда (красные точки) каждой из предлагаемых функций (черная линия). Все функции имеют по семь параметров.

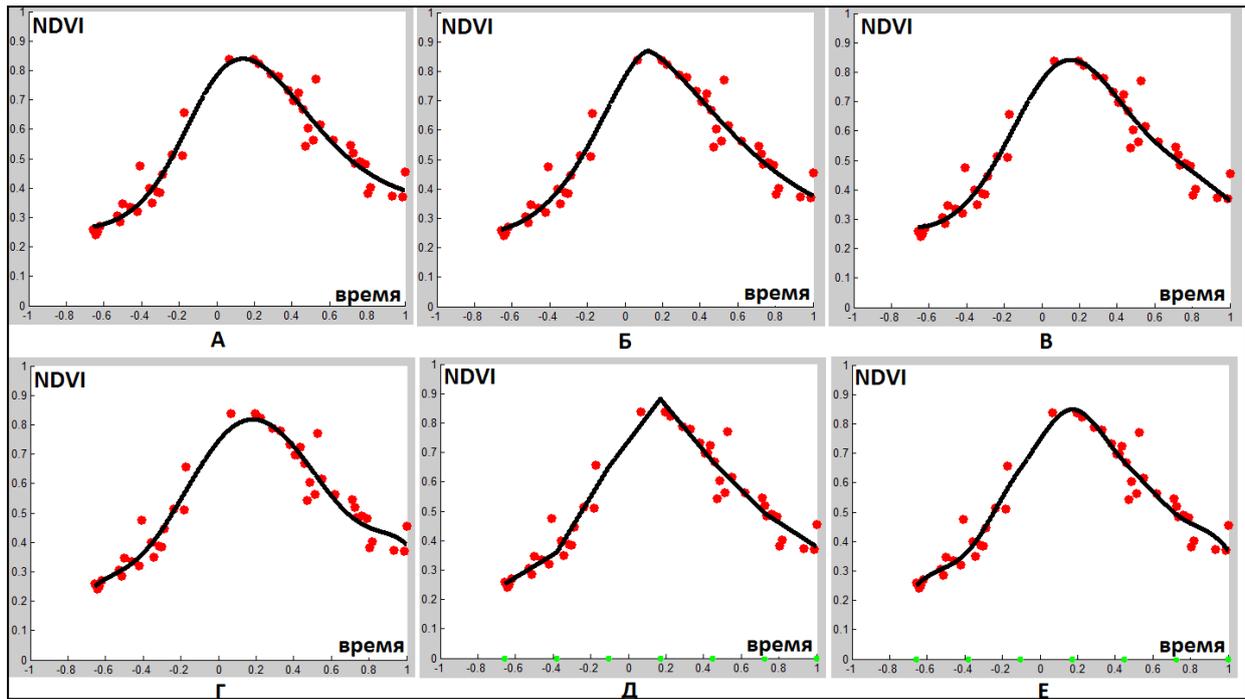


Рис. 4. Аппроксимация временного ряда предлагаемыми функциями (А – двойная логистическая, Б – асимметричная гауссова, В – ряд Фурье, Г – полиномы, Д – кусочно-линейная функция, Е – сплайн).

Для всех временных рядов из набора исходных данных была проведена аппроксимация в пакете MATLAB с помощью встроенной функции fit. Искомые значения коэффициентов функций определялись как решение задачи минимизации суммы квадратов невязок (метод наименьших квадратов). Для каждой функции аппроксимации были рассчитаны соответствующие критерии качества (для всего набора временных рядов и с разделением на классы).

Далее приведены сравнительные результаты аппроксимации.

На рисунке 5 приведены графики среднеквадратической и максимальной ошибок аппроксимации для различных функций в зависимости от числа параметров (от 5 до 10, для иллюстрации наиболее значимых отличий значений ошибок). Ошибки рассчитаны для аппроксимации всех временных рядов из набора. Аналогичные графики были построены отдельно для каждого класса, в статье они не приводятся, но общий вид закономерности сохраняется.

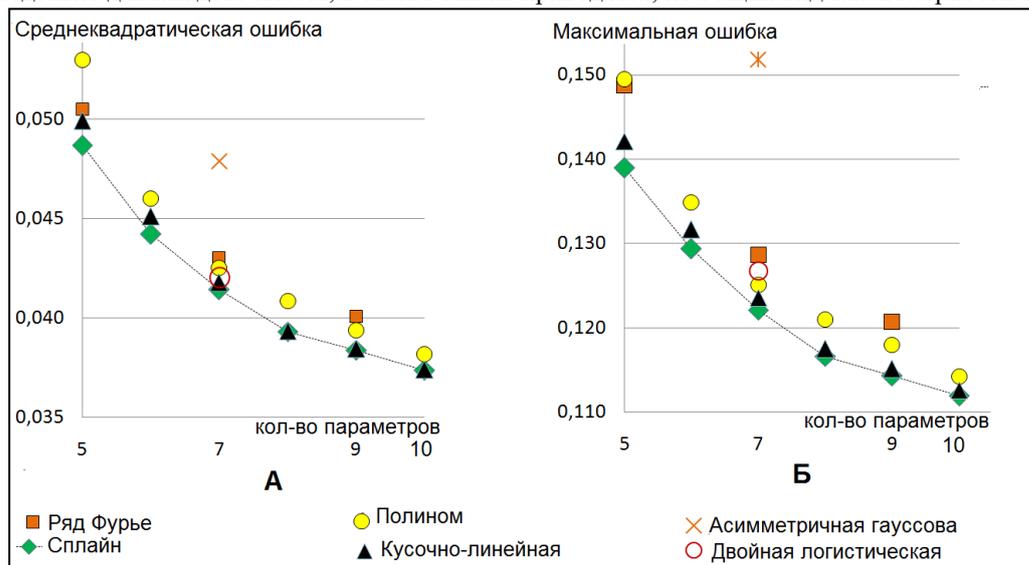


Рис. 5. Зависимость среднеквадратической (А) и максимальной (Б) ошибок от количества параметров различных функций аппроксимации.

В таблицах 2 и 3 приведены значения среднеквадратической и максимальной ошибок аппроксимации функциями с 7-ю параметрами с разделением по классам. Серым цветом в таблице отмечены минимумы по столбцам, соответствующие значениям минимальных ошибок аппроксимации для каждого класса, то есть выделены функции, наилучшим образом аппроксимирующие класс.

Таблица 2. Значения среднеквадратической ошибки для функций с 7-ю параметрами

Модель	многолетние травы и неиспользуемые земли	озимые	пар	поздние яровые	ранние яровые
Асимметричная гауссова	0,0464	0,0503	0,0525	0,0443	0,0471
Двойная логистическая	0,0385	0,0413	0,0437	0,0424	0,0427
Кусочно-линейная	0,0381	0,0411	0,0440	0,0408	0,0439
Полином	0,0386	0,0430	0,0457	0,0435	0,0448
Сплайн	0,0380	0,0405	0,0438	0,0405	0,0433
Фурье	0,0426	0,0409	0,0442	0,0427	0,0440

Таблица 3. Значения максимальной ошибки для функций с 7-ю параметрами

Модель	многолетние травы и неиспользуемые земли	озимые	пар	поздние яровые	ранние яровые
Асимметричная гауссова	0,1449	0,1628	0,1651	0,1411	0,1436
Двойная логистическая	0,1143	0,1202	0,1339	0,1308	0,1253
Кусочно-линейная	0,1101	0,1175	0,1337	0,1216	0,1259
Полином	0,1110	0,1182	0,1362	0,1236	0,1267
Сплайн	0,1089	0,1161	0,1325	0,1211	0,1248
Фурье	0,1330	0,1188	0,1322	0,1300	0,1273

На рисунке 6 приведена статистика неверных аппроксимаций: по оси абсцисс отложено количества параметров функции аппроксимации, по оси ординат процент неверных аппроксимации от общего числа временных рядов в наборе исходных данных.

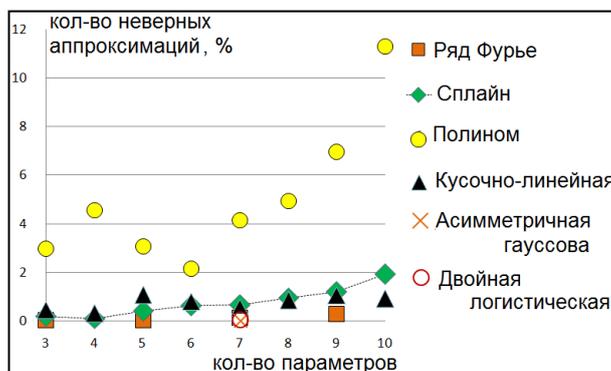


Рис. 6. Статистика неверных аппроксимаций.

Проведенные исследования по аппроксимации временных рядов, некоторые результаты которых приведены в таблицах 2,3 и на рисунках 5,6, позволяют сделать следующие выводы:

1. Худшее качество аппроксимации дает асимметричная гауссова функция.
2. Ряд Фурье плохо аппроксимирует временные ряды, построенные для классов «поздние яровые» и «многолетние травы и неиспользуемые земли».
3. Кусочно-линейная функция, сплайн, полином и двойная логистическая функция, начиная с четырех параметров, дают близкие ошибки аппроксимации. Сравним указанные функции:
 - 3.1. По количеству неверных аппроксимаций полином – наихудшая функция. Основная причина выходов значений полинома за диапазон [0;1] связана с особенностью полиномиальной функции, приводящей к неконтролируемому скачку значений полинома вне точек аппроксимируемого временного ряда.
 - 3.2. Недостатком двойной логистической функции является фиксированное количество параметров, что отражается на удобстве её использования по сравнению со сплайном и кусочно-линейной функцией: у последних имеется возможность улучшения качества аппроксимации за счет увеличения числа параметров, а также ускорения расчетов за счет уменьшения количества параметров.
 - 3.3. Профиль временного ряда, формируемый с помощью сплайнов, более соответствует наблюдаемому в реальности процессу плавного изменения значений вегетационного индекса NDVI по сравнению с кусочно-линейной функцией.

Таким образом, из всех функций, предлагаемых для моделирования временных рядов, по выбранным критериям качества наилучшей является сплайн третьего порядка (9). Учитывая большое количество использованных в исследовании исходных данных за продолжительный временной диапазон (5 лет), этот вывод можно экстраполировать и на предшествующие и последующие временные диапазоны (для них также наилучшей функцией будет являться предложенный сплайн 3-го порядка).

4. Моделирование временных рядов на основе сплайна третьего порядка

4.1. Анализ модели

Рассматриваемая в данной работе модель временного ряда – это функция $g(c_1, \dots, c_m, x)$, описывающая полезный сигнал без шумовой составляющей. Проверим, подходит ли сплайн 3-го порядка для описания такого сигнала.

В результате аппроксимации значения временного ряда разбиваются на две части – кривая аппроксимации и невязки в точках. Если невязки представляют собой шум с некоррелированными значениями, значит, они не несут в себе никакой полезной информации и вся полезная информация содержится в значениях кривой.

Исследуем некоррелированность невязок аппроксимации сплайном 3-го порядка с различным числом параметров (от 3 до 10). Для каждого сплайна с заданным числом параметров сформируем единый вектор невязок $v(t)$, составленный последовательным присоединением друг за другом невязок аппроксимации всех временных рядов из набора исходных данных. Используем автокорреляционную функцию $K(\tau)$ (далее – АКФ), показывающую корреляцию значений исходной последовательности $v(t)$ и ее смещенной на τ отсчетов копии:

$$K(\tau) = E\{[v(t) - \bar{v}(t)][v(t - \tau) - \bar{v}(t - \tau)]\} / D, \quad (13)$$

где $\bar{v}(t) = E[v(t)]$ – математическое ожидание

Расчет АКФ производился с использованием встроенной в Matlab функции `autocorr`. Далее на рисунках приведены графики автокорреляционных функций, построенных по невязкам аппроксимации сплайнами с различным числом параметров. На рисунке 7 приведены графики зависимостей значений АКФ в некоторых отсчетах, соответствующих смещению на 1, 2, 3, 4, 5 от количества параметров метода аппроксимации. На каждом графике АКФ дополнительно приведен порог, равный 0,2.

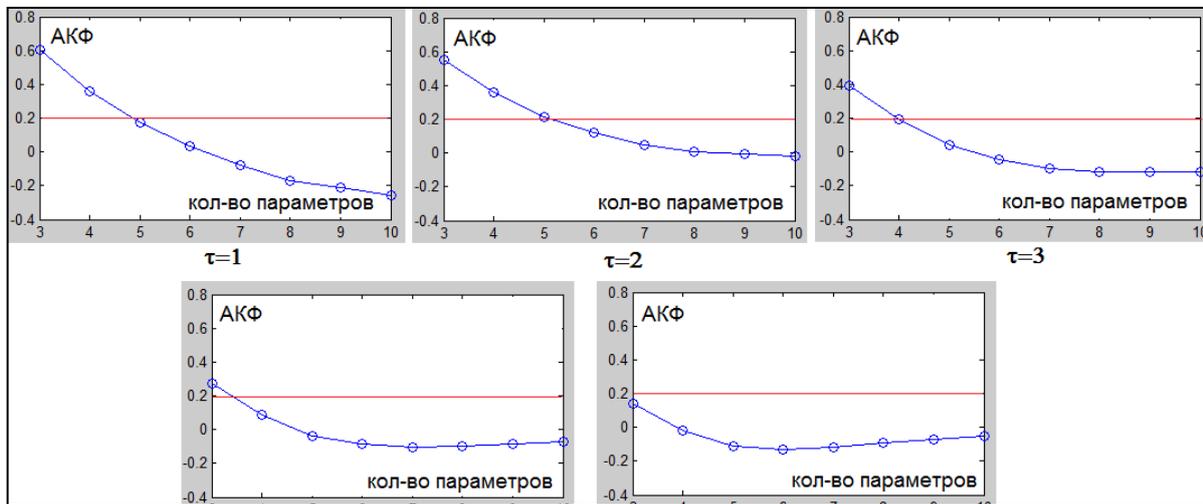


Рис. 7. Зависимость значений АКФ в некоторых отсчетах, соответствующих смещению на 1, 2, 3, 4, 5 от количества параметров метода аппроксимации.

Будем считать, что корреляция в значениях невязок отсутствует, если значение АКФ менее 0,2. Это предположение позволяет сделать вывод о том, что начиная с 5 параметров, сплайн 3-го порядка (9) подходит для описания полезного сигнала и функции $g(c_1, \dots, c_m, x)$.

Далее приведен пример генерации временных рядов с использованием модели временного ряда, задаваемой сплайном (9).

4.2. Пример генерации временных рядов

Рассмотрим набор временных рядов, построенных для полей из некоторого класса, расположенных в области с одинаковыми природно-климатическими условиями произрастания культур. Примером деления территории региона на такие области является агроклиматическое зонирование (например, территория Самарской области делится на три такие зоны: северную, южную и центральную).

Для каждого временного ряда из набора проведем аппроксимацию сплайном 3-го порядка с параметрами c_1, \dots, c_m . В результате, набору временных рядов будет соответствовать набор реализаций вектора параметров $\mathbf{c} = (c_1, \dots, c_m)$. Предположим, что вектор \mathbf{c} имеет многомерный нормальный закон распределения с вектором математических ожиданий \mathbf{M} и ковариационной матрицей B . Оценим параметры \mathbf{M} и B этого закона с использованием полученного набора реализаций вектора параметров \mathbf{c} . Далее с использованием полученного многомерного нормального закона распределения, заданного характеристиками \mathbf{M} и B , будем генерировать временные ряды выбранного класса, соответствующие рассмотренной области с одинаковыми природно-климатическими условиями. Указанным способом можно генерировать временные ряды различных классов, расположенных в любых других областях произрастания культур.

В качестве иллюстраций на рисунках 8 и 9 приведены соответственно гистограммы значений и диаграммы рассеяния параметров c_1, \dots, c_5 сплайна с 5-ю параметрами для класса «Поздние яровые» в южной зоне Самарской области в 2015 году. На каждую гистограмму (рис. 8) наложены аппроксимирующие функции нормального распределения.

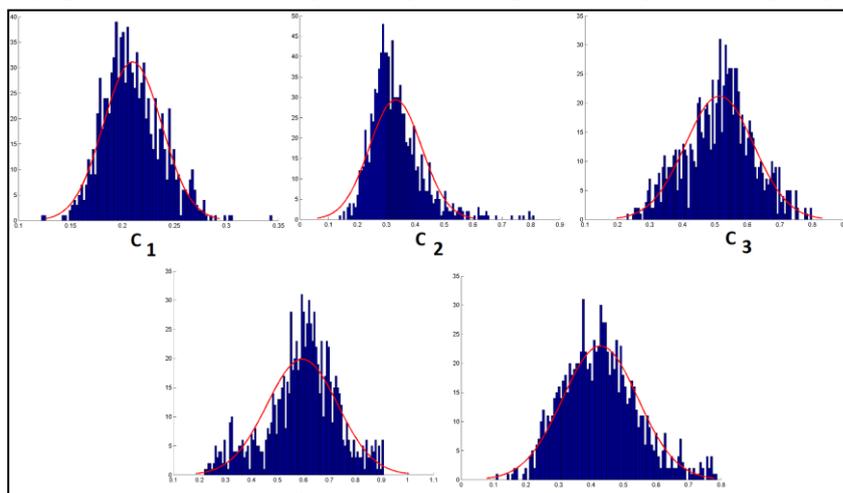


Рис. 8. Гистограммы значений параметров c_1, \dots, c_5 класса «Поздние яровые» в южной зоне в 2015 году для сплайна с 5-ю параметрами.

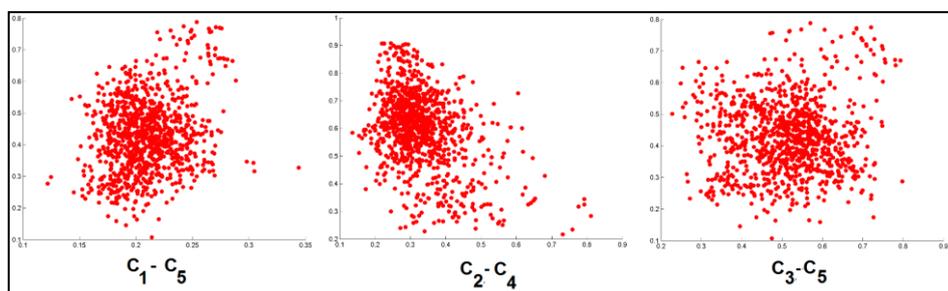


Рис. 9. Диаграммы рассеяния некоторых параметров класса «Поздние яровые» в южной зоне в 2015 году для сплайна с 5-ю параметрами.

На рисунке 10 приведены примеры моделей временных рядов (черная линия), сгенерированных для класса «Поздние яровые» для южной зоны 2015 года с использованием сплайна с параметрами c_1, \dots, c_5 . На графиках дополнительно изображен сглаженный усредненный тренд (красная кривая) зависимости NDVI от времени, рассчитанный по всем временным рядам исходных данных класса «Поздние яровые» южной зоны 2015 года.

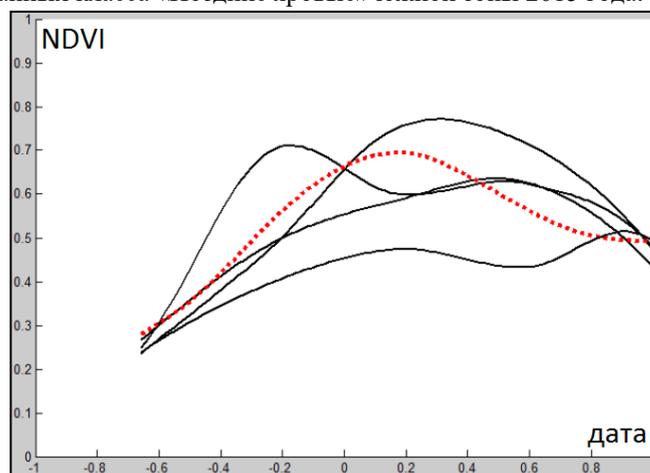


Рис. 10. Пример сгенерированных временных рядов для класса «Поздние яровые» с использованием сплайна с 5-ю параметрами.

5. Заключение

В работе рассматривалась задача поиска функции, подходящей для описания моделей временных рядов, для использования в информационной технологии раннего распознавания видов с/х культур по космическим снимкам. Для анализа был сформирован набор функций, состоящий из кусочно-линейной, двойной логистической, асимметричной гауссовой функций, ряда Фурье, полинома и сплайна. По заданным критериям качества, включающим среднеквадратическую и максимальную ошибки, а также количество выполненных аппроксимаций, в качестве наилучшей функции был выбран сплайн 3-го порядка. В работе приведен пример генерации временных рядов с использованием данной функции.

Благодарности

Работа выполнена при финансовой поддержке РФФИ: гранты № 16-37-00043_мол_а «Разработка методов использования данных геоинформационных систем для обработки данных дистанционного зондирования Земли» и № 16-29-09494_офи_м «Методы компьютерной обработки мультиспектральных данных дистанционного зондирования Земли для определения ареалов растений в специальных криминалистических экспертизах».

Литература

- [1] Султангазин, У.М. Использование космического мониторинга в планировании и прогнозировании параметров зернового производства / У.М. Султангазин, Н.Р. Муратова, А.Г. Терехов // Современные проблемы дистанционного зондирования Земли из космоса. – 2004. – Т. 1, № 1. – С. 291-297.
- [2] Schmedtmann, J. Reliable Crop Identification with Satellite Imagery in the Context of Common Agriculture Policy Subsidy Control / J.Schmedtmann, M.L. Campagnolo // Remote Sensing of Environment. – 2015. – Vol. 7(7). – P. 9325-9346. DOI: 10.3390/rs70709325.
- [3] Плотников, Д.Е. Экспериментальная оценка распознаваемости агрокультур по данным сезонных спутниковых измерений спектральной яркости / Д.Е. Плотников, С.А. Барталев, В.О. Жарко, В.В. Михайлов, О.И. Просянникова // Современные проблемы дистанционного зондирования Земли из космоса. – 2011. – Т. 8, № 1. – С. 199-208.
- [4] Wardlow, B.D. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains / B.D.Wardlow, S.L.Egbert, J.H.Kastens // Remote Sensing of Environment. – 2007. – Vol. 108. – P. 290-310. DOI: 10.1016/j.rse.2006.11.021.
- [5] Wardlow, B.D. Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U.S. Central Great Plains / B.D.Wardlow, S.L.Egbert // Remote sensing of environment. – 2008. – Vol. 112(3). – P. 1096-1116. DOI: 10.1016/j.rse.2007.07.019.
- [6] Черепанов, А.С. Вегетационные индексы / А.С. Черепанов // Геоинформатика. – 2011. – Т. 2. – С. 98-102.
- [7] Воробьёва, Н.С. Информационная технология раннего распознавания видов сельскохозяйственных культур по космическим снимкам / Н.С. Воробьёва, В.В. Сергеев, А.В. Чернов // Компьютерная оптика. – 2016. – Т. 40, № 6. – С. 929-938. DOI: 10.18287/2412-6179-2016-40-6-929-938.
- [8] Lui, M.W. Crop Type Classification using satellite images of different resolutions. A Thesis for degree of Master of Science / M.W.Liu – University of Wisconsin-Madison, 2011. – 73 p.
- [9] Lui, M.W. Crop Type Classification by Simultaneous Use of Satellite Images of Different Resolutions / M.W.Liu, M.Ozdogan, X.Zhu // IEEE Transactions on geoscience and remote sensing. – 2014. – Vol. 52(6). – P.3637-3649.
- [10] Ozdogan, M. The spatial distribution of crop types from MODIS data: Temporal unmixing using Independent Component Analysis / M.Ozdogan // Remote Sensing of Environment. – 2010. – Vol. 114(6). – P. 1190 – 1204. DOI: 10.1016/j.rse.2010.01.006.
- [11] Fischer, A. A Model for the Seasonal Variations of Vegetation Indices in Coarse Resolution Data and Its Inversion to Extract Crop Parameters / A.Fischer // Remote Sensing of Environment. – 1994. – Vol. 48(2). – P. 220-230. DOI: 10.1016/0034-4257(94)90143-0.
- [12] Fischer, A. A simple model for the temporal variations of NDVI at regional scale over agricultural countries. Validation with ground radiometric measurements / A.Fischer // International Journal of Remote Sensing. – 1994. – Vol. 15(7). – P. 1421-1446. DOI: 10.1080/01431169408954175.
- [13] Itzerott, S. Crop Classification based on spectral standard curves / S.Itzerott, K.Kaden // Proceedings of the second Workshop of the EARSeL SIG on Remote Sensing of Land Use & Land Cover "Application and Development". – Bonn: Center for Remote Sensing of Land Surfaces. – 2006. – P. 229-238.
- [14] Майорова, В.И. Контроль состояния сельскохозяйственных полей на основе прогнозирования динамики индекса NDVI по данным космической мультиспектральной и гиперспектральной съёмки / В.И. Майорова, А.М. Банников, Д.А. Гришко, И.С. Жаренов, В.В. Леонов, А.Г. Топорков, А.А. Харлан // Наука и образование: научное издание МГТУ им. Н.Э. Баумана. – 2013. – В. 07. – С. 199-228.
- [15] Hird, J.N. Noise reduction of NDVI time series: An empirical comparison of selected techniques / J.N.Hird, G.J.McDermid // Remote Sensing of Environment. – 2009. – Vol. 113(1). – P. 248 – 258. DOI: 10.1016/j.rse.2008.09.003.
- [16] Michishita, R. Empirical comparison of noise reduction techniques for NDVI time-series based on a new measure / R. Michishita, J.Chen, B.Xu // ISPRS Journal of Photogrammetry and Remote Sensing. – 2014. – Vol. 91. – P. 17-28. DOI: 10.1016/j.isprsjprs.2014.01.003.
- [17] Beck, P.S.A. Improved monitoring of vegetation dynamics at very high latitudes- a new method using MODIS NDVI / P.S.A.Beck, C.Atzberger, K.A.Hogda, B.Johansen, A.K. Skidmore // Remote Sensing of Environment. – 2006. – Vol. 100(3). – P. 321-334. DOI: 10.1016/j.rse.2005.10.021.
- [18] Victoria, D.C. Cropland area estimates using Modis NDVI time series in the state of Mato Grosso / D.C.Victoria, A.R.Paz, A.C.Coutinho, J.C.Brown // Brazil. Pesq. agropec. bras. – 2012. – Vol. 47(9). – P. 1270-1278. DOI: 10.1590/S0100-204X2012000900012.
- [19] Son, N.T. A Phenology-Based Classification of Time-Series MODIS Data for Rice Crop Monitoring in Mekong Delta / N.T.Son, C.F.Chen, C.R.Chen, H.N.Duc, L.Y.Chang // Vietnam. Remote Sensing. – 2014. – Vol. 6(1). – P. 135-156. DOI: 10.3390/rs6010135.
- [20] Geerken, R.A. An algorithm to classify and monitor seasonal variations in vegetation phenologies and their inter-annual change / R.A.Geerken // ISPRS Journal of Photogrammetry and Remote Sensing. – 2009. – Vol. 64(4). – P. 422-431. DOI: 10.1016/j.isprsjprs.2009.03.001.
- [21] Mingwei, Z. Crop discrimination in Northern China with double cropping systems using Fourier analysis of time-series MODIS data / Z. Mingwei, Z. Qingbo, C. Zhongxin, L. Jia, Z. Yong, C.Chongfa // International Journal of Applied Earth Observation and Geoinformation. – 2008. – Vol. 10(4). – P. 476 – 485. DOI: 10.1016/j.jag.2007.11.002.
- [22] Chen, J. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter / J.Chen, P. Jönsson, M.Tamura, Z.Gu, B.Matsushita, L.Eklundh // Remote Sensing of Environment. – 2004. – Vol. 91(3-4). – P. 332 – 344. DOI: 10.1016/j.rse.2004.03.014.
- [23] Madden, H.H. Comments on the Savitzky-Golay Convolution Method for Least-Squares Fit Smoothing and Differentiation of Digital Data / H.H.Madden // Analytical chemistry. – 1978. – Vol. 50 (9). – P. 1383 – 1386. DOI: 10.1021/ac50031a048.
- [24] Keys, R.G. Cubic Convolution Interpolation for Digital Image Processing / R.G.Keys // IEEE Transactions on acoustics, speech and signal processing. – 1981. – Vol. 29(6). – P. 1153 – 1160. DOI: 10.1109/TASSP.1981.1163711.