Современные проблемы дистанционного зондирования Земли из космоса. 2014. Т. 11. № 2. С. 103-110

Восстановление временных рядов данных дистанционных измерений методом полиномиальной аппроксимации в скользящем окне переменного размера

Д.Е. Плотников, Т.С. Миклашевич, С.А. Барталёв

Институт космических исследований РАН, Москва, Россия
E-mail: dmitplot@d902.iki.rssi.ru

Использование временных рядов спутниковых данных для мониторинга земной поверхности сопряжено с проблемой учёта разного рода мешающих факторов, приводящих к частичной потере или искажению информации о динамике спектрально-отражательных характеристик объектов наблюдения. К таким факторам относятся непрозрачные в видимом и ближнем ИК диапазоне атмосферная дымка и облачность, тени от неё, а также шумы прибора. Существующие статистические подходы к реконструкции временных рядов измерений решают, как правило, только проблему коррекции зашумленных данных, предполагая задачу восстановления пропущенных значений уже решенной. Такие методы также не предусматривают обнаружение и полное исключение из дальнейшего рассмотрения заведомо зашумленных измерений. Однако, с учетом того, что корректное восстановление пропущенных значений временной серии предполагает использование свободных от шумов измерений, эти проблемы тесно связаны. В статье представлено описание алгоритма скользящей оконной аппроксимации на основе полиномов второй степени, позволяющего при восстановлении временных рядов спутниковых данных одновременно решать задачи коррекции искаженных и восстановления пропущенных измерений. Предварительный итеративный процесс анализа временной серии предусматривает исключение заведомо зашумленных измерений.

Ключевые слова: временные ряды, спутниковые данные, восстановление пропусков, фильтрация выбросов.

Спутниковые наблюдения предоставляют наиболее объективную и оперативную информацию о состоянии земной поверхности на больших масштабах пространственного охвата. В зависимости от целей мониторинга, спутниковые измерения могут находиться под влиянием различных мешающих факторов. При дистанционном мониторинге растительного покрова в видимом и ближнем ИК-диапазоне, к мешающим факторам относятся облачный и снежный покров, тени от облаков, дымка и аэрозоли, солнечные блики и шумы прибора.

Проводимые с использованием ряда современных спутниковых систем (NOAA-AVHRR, SPOT-Vegetation, Terra/Aqua-MODIS, Proba-V и др.) регулярные измерения позволяют формировать временные ряды данных о спектрально-отражательных характеристиках наблюдаемых объектов, используемых, в частности, для распознавания типов и оценки динамики земного покрова.

В Институте космических исследований РАН создан постоянно пополняемый в оперативном режиме многолетний архив данных измерений прибором MODIS на территорию Северной Евразии (Балашов и др., 2008). Необходимость обработки огромных объемов получаемых данных, исключающих непосредственное участие в этом процессе человека, требует разработки соответствующих автоматических алгоритмов и технологий.

Предварительная обработка спутниковых данных MODIS включает в себя фильтрацию облаков и отбрасываемых ими на земную поверхность теней, маскирование снежного покрова, а также исключение измерений, выполненных при неблагоприятных геометрических условиях съёмки или характеризуемых наличием высокого уровня аппаратных шумов (Барталев и др., 2011). В результате проведения предварительной обработки и комплексирования прошедших фильтрацию рядов последовательных измерений за временные интервалы фиксированной длины формируются в значительной степени освобожденные от влияния мешающих факторов композитные спутниковые изображения. Как правило, с увеличением длины вышеуказанного временного интервала сокращается доля пропусков во временных сериях композитных изображений и достигается более высокий уровень их очистки от шумов. При этом необходимо учитывать, что с увеличением длины временного интервала формирования композитных изображений информативность составленных из них временных серий может снизиться при их использовании для решения ряда задач распознавания типов и мониторинга динамики земного покрова. В частности, это особенно важно при использовании временных рядов спутниковых данных для мониторинга сезонных изменений растительного покрова. Использование более частых наблюдений растительного покрова позволяет обнаруживать и точнее определять даты важных фенологических фаз (начало роста, цветение), оценивать скорость роста и прогнозировать урожайность сельскохозяйственных культур, улучшать качество их классификации, используя незаметные на больших временных масштабах отличия сезонной динамики развития (Плотников и др., 2011).

Появление под влиянием мешающих факторов пропусков и искажений во временных рядах измерений спектрально-отражательных характеристик земной поверхности делает актуальным поиск эффективных методов восстановления временных рядов. При этом существующие статистические подходы к восстановлению временных серий измерений решают, как правило, только проблему коррекции серии частично зашумленных данных без пропусков, никак не затрагивая вопрос восстановления пропущенных значений во временной серии с искаженными значениями. Однако эти проблемы тесно связаны, поскольку корректное восстановление пропущенных значений предполагает использование неискаженных измерений. В статье представлено описание алгоритма скользящей оконной аппроксимации на основе полиномов второй степени, позволяющего при реконструкции временных рядов спутниковых данных одновременно решать задачи коррекции искаженных и заполнения пропущенных измерений. Разработанный метод восстановления временных рядов данных дистанционных измерений обеспечивает исключение зашумленных и заполнение пропущенных значений на основе алгоритма оконной аппроксима-

ции локальной серии измерений полиномами второй степени вида $f(t) = at^2 + bt + c$, где *f(t)* – значение некоторой спектрально-отражательной характеристики земной поверхности в момент времени *t*, а *a, b* и *c* – коэффициенты полинома. Алгоритм предусматривает использование скользящего вдоль оси времени окна, включающего некоторое фиксированное количество значащих измерений для вычисления коэффициентов полинома. Поиск значений коэффициентов происходит с помощью метода наименьших квадратов с нахождением локальных экстремумов соответствующего функционала методом оптимизации Нелдера-Мида (метод деформируемого многогранника) (Nelder, Mead, 1965). Данный метод не использует значения градиентов функции и пригоден для негладких функций, выгодно отличаясь быстродействием и надежностью. Полином второй степени является оптимальным для локальной аппроксимации фенологической динамики растительности, являясь относительно простым с вычислительной точки зрения и одновременно позволяющим корректнее в рамках данного подхода описывать фенологические экстремумы и точки перегиба фенологической кривой по сравнению с линейной функцией. Расчет коэффициентов полинома происходит на основе значений локального временного интервала переменной длины, выбираемой в общем случае таким образом, чтобы обеспечить включение в него фиксированного количества значащих измерений. Используя такой подход можно показать (*рис. 1*), что значения искаженных измерений зачастую заметно отличаются от соответствующих значений локальных аппроксимирующих функций, несмотря на то, что они также используются при их построении, за счет стабилизирующего влияния массы чистых измерений. Кроме этого, стабилизирующее действие оказывает нормальность распределения зашумленных значений около кривой тренда.

Пусть имеется исходный временной ряд $Y = \{y_1, y_2, ..., y_n\}$, состоящий из *n* измерений y_i, включая зашумленные, полученных в соответствующие моменты времени t_i. Измерение, полученное в момент времени t, обозначим $Y(t)$. Для построения локальных полиномов пусть выбран размер *p* скользящего окна локализации. Таким образом, входящие в окно локализации с началом на *k*-ом элементе из ряда Y элементы можно обозначить $Y_k = \{y_k, y_{k+1}, ..., y_{k+p-1}\}$. Принцип работы алгоритма заключается в создании множества оценок A_k для каждого *k*-го измерения временного ряда, на основе которого можно судить о валидности данного измерения. После вычисления на основе Y_k коэффициентов полинома $f_k(t)$, становится возможным вычислить соответствующую ему оценку для каждого элемента Y_k, а, значит, и каждого элемента Y. Скользящее окно шириной *p* включает в себя каждый элемент исходного временного ряда Y не более *p* раз. Таким образом,

каждый элемент Y получает не более p оценок на основе всех локальных функций $f_k(t)$, проходящих через этот элемент. Более наглядно распределение оценок для каждого элемента исходного временного ряда можно представить в виде матрицы M, в которой строки соответствуют значениям оценок элементов Y, а столбцы – различным оценкам одного элемента:

$$M = \left\| \begin{matrix} a_{11} & a_{12} & a_{13} & \dots & a_{1p} & & & & & \\ & a_{21} & a_{22} & a_{23} & \dots & a_{2p} & & & & \\ & & a_{31} & a_{32} & a_{33} & \dots & a_{3p} & & & \\ & & & & & & \dots & & & \\ & & & & & & & a_{n1} & \dots & a_{np} \end{matrix} \right\| .$$

Таким образом, k-ый элемент временного ряда получит в общем случае (когда $k \geq p$ и $k \leq n-p$) множество оценок $A_k = \{a_{(k-m)m}\}, m = 1,.., p$, мощностью p. Учёт краевых эффектов происходит за счет экстраполяции полиномов $f_k(t)$ при $k \geq p$ и $k \leq n-p$ за пределы окна локализации.

Работу программно реализованного алгоритма можно представить в виде следующих шагов.

После расчета p различных оценок для текущего элемента временного ряда, заполняющих множество A_k, вычисляется среднее значение $\overline{y}_k$ и дисперсия этих оценок σ_k. Процесс получения восстановленной оценки значения текущего элемента можно разделить на два этапа. На первом этапе проводится классификация всех элементов временного ряда на следующие три группы:

- достоверно искаженные измерения;
- вероятно искаженные измерения;
- неискаженные измерения.

Отнесение элементов временного ряда $Y(t_k)$ к категории достоверно искаженных измерений предполагает совместное удовлетворение критериев $|Y(t_k) - \overline{y}_k| > 2\sigma_k$ и $\frac{\sigma_k}{\overline{y}_k} > 0{,}05$, либо только критерия $\frac{\sigma_k}{\overline{y}_k} > 1$. При совместном соблюдении условий $|Y(t_k) - \overline{y}_k| > 2\sigma_k$ и $\frac{\sigma_k}{\overline{y}_k} < 0{,}05$ элемент $Y(t_k)$ относится к категории вероятно искаженных. В случае невыполнения ни одного из вышеуказанных условий, элемент временного ряда считается неискаженным и остается неизменным. Искаженные измерения исключаются из временной серии, а вероятно искаженные заменяются их расчетной средней оценкой: $Y(t_k) = \overline{y}_k$.

На втором этапе производится оконная полиномиальная аппроксимация очищенной от искаженных значений временной серии данных. Описанные этапы при необходимости могут итеративно повторяться с заданным числом итераций.

Пропущенные значения исходной временной серии также заменяются их средними оценками, полученными на основе локальных полиномов, рассчитанных для лежащих в их окрестности окон. С вычислительной точки зрения перед расчетом локальных полиномов исходная временная серия модифицируется линейным преобразованием так, чтобы измерения были строго положительными и занимали диапазон от 0 до 10000. Все относительные и абсолютные пороги, указанные в используемых алгоритмом критериях, распространяются только на модифицированный временной ряд. После окончания работы алгоритма к восстановленной временной серии применяется обратное преобразование.

Пример расположения локальных полиномов на отрезке исходной временной серии полученных по данным MODIS значений перпендикулярного вегетационного индекса PVI представлен на *рис. 1*.

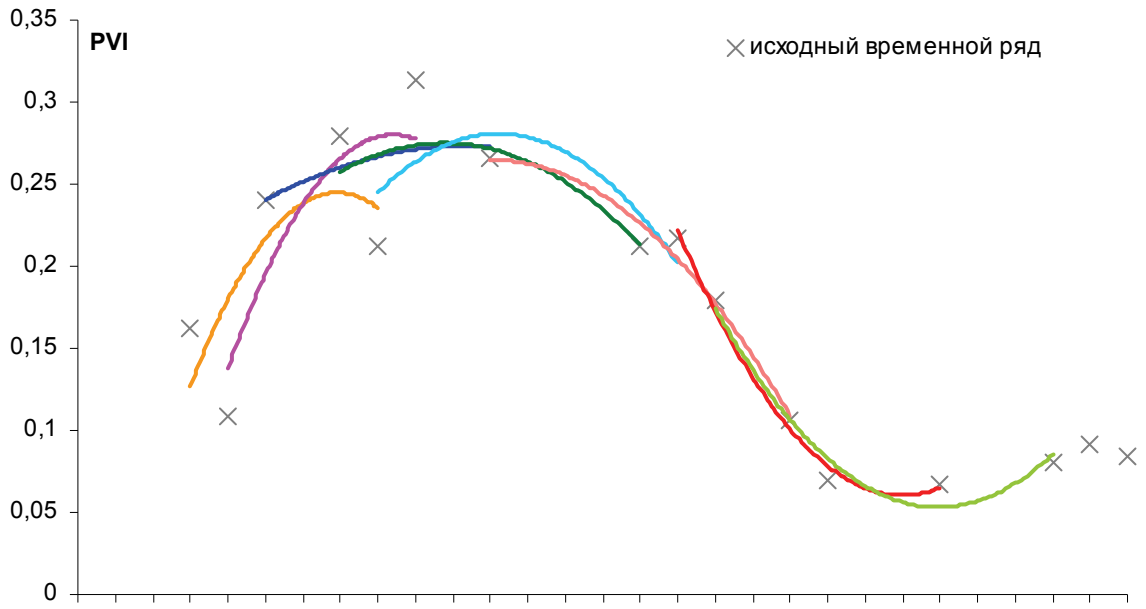

Рис. 1. Пример полиномиальной аппроксимации зашумленного временного ряда PVI с наличием пропусков при ширине окна p=5. Полиномы, полученные при разных положениях скользящего окна, показаны разными цветами

Пример временных рядов исходных и восстановленных значений PVI с помощью скользящей оконной аппроксимации с шириной окна локализации p=5 приведен на *рис. 2*.

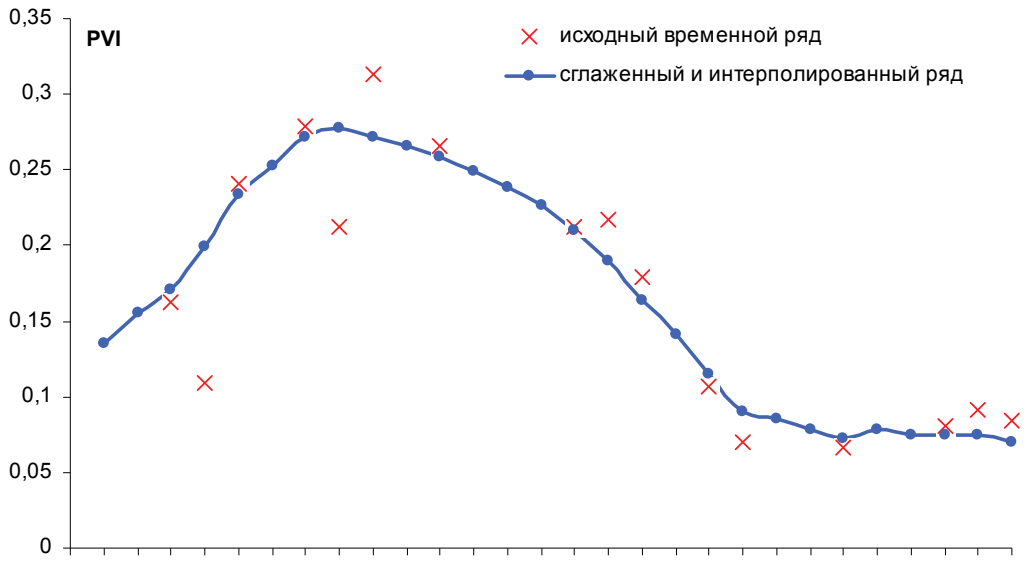

Рис. 2. Пример сглаживания и заполнения пропущенных значений временного ряда PVI алгоритмом полиномиальной аппроксимации. Исходный и сглаженный временные ряды имеют продолжение в обе стороны вдоль оси абсцисс

Разработанный алгоритм выгодно отличается от других разработок, использующих скользящие окна и аппроксимацию полиномами разных порядков, тем, что он одновременно решает задачи интерполяции пропущенных значений, поиска и исключения измерений, сделанных под влиянием мешающих факторов и сглаживания временных серий. Другие распространенные алгоритмы, в частности, алгоритм полиномиального сглаживания Голая-Савицкого (Savitzky, Golay, 1964), а также статистические алгоритмы скользящих окон предполагают обработку зашумленных измерений, для которых задача заполнения пропущенных значений не стоит.

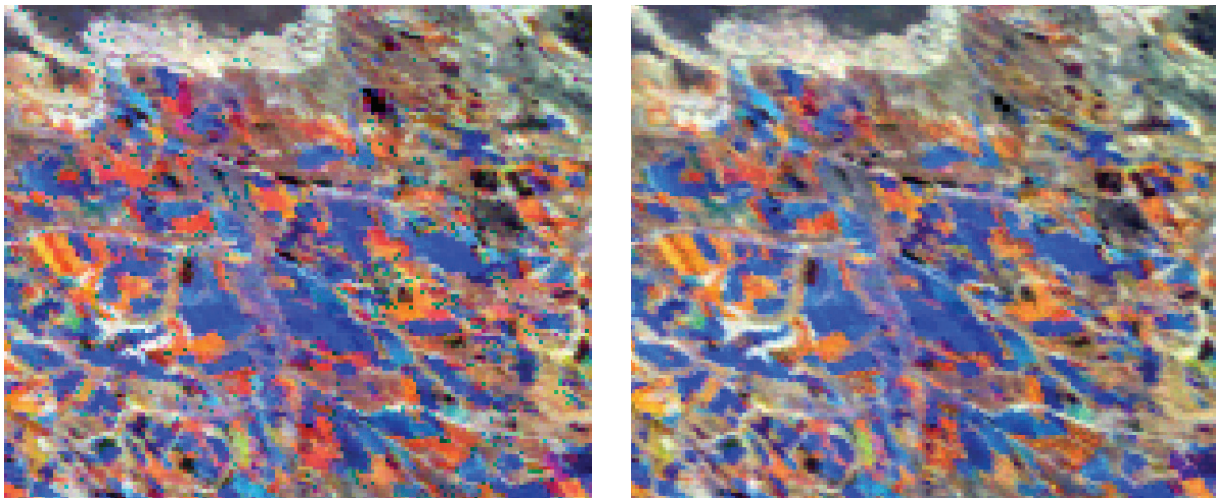

Рис. 3. Пример снижения пространственных шумов изображения в результате сглаживания временного ряда данных алгоритмом полиномиальной аппроксимации, слева – до сглаживания, справа – после сглаживания

К преимуществам использования разработанного алгоритма можно отнести также и обеспечиваемое им повышение пространственной четкости обработанных изображений (*рис. 3*). Этот эффект объясняется коррекцией в процессе сглаживания временной серии зашумленных измерений, сделанных под влиянием облачности и теней.

Алгоритм реализован на языке C++ в виде двух версий, первая из которых позволяет осуществлять восстановление стационарных временных рядов спутниковых данных. Модифицированная версия алгоритма позволяет обрабатывать динамически пополняемые временные ряды спутниковых данных, позволяя уточнять результаты восстановления по мере поступления новых данных.

Получаемые в результате работы алгоритма временные ряды спутниковых данных информационные продукты используются для распознавания и оценки сельскохозяйственной растительности (Плотников, Барталёв, Лупян, 2010; Плотников, 2011), оценки последствий пожаров (Стыценко, 2013), а также включены в состав информационного сервиса спутникового мониторинга растительного покрова «Вега» (Лупян и др., 2011).

Литература

1. *Балашов И.В., Бурцев М.А., Ефремов В.Ю., Лупян Е.А., Прошин А.А., Толпин В.А.* Построение архивов результатов обработки спутниковых данных для систем динамического формирования производных информационных продуктов // Современные проблемы дистанционного зондирования Земли из космоса, 2008. Выпуск 5. Т. I. С. 26-32.
2. *Барталев С.А., Егоров В.А., Лупян Е.А., Плотников Д.Е., Уваров И.А.* Распознавание пахотных земель на основе многолетних спутниковых данных спектрорадиометра MODIS и локально-адаптивной классификации // Компьютерная оптика, 2011. Самара. ИСОИ РАН. Т. 35. № 1. С. 103-116.
3. *Лупян Е.А., Савин И.Ю., Барталев С.А., Толпин В.А., Балашов И.В., Плотников Д.Е.* Спутниковый сервис мониторинга состояния растительности ("Вега") // Современные проблемы дистанционного зондирования Земли из космоса. 2011. Т. 8. № 1. С. 190-198.
4. *Плотников Д.Е., Барталев С.А., Жарко В.О., Михайлов В.В., Просянникова О.И.* Экспериментальная оценка распознаваемости агрокультур по данным сезонных спутниковых измерений спектральной яркости // Современные проблемы дистанционного зондирования Земли из космоса. 2011. Т. 8. № 1. С. 199-208.
5. *Плотников Д.Е., Барталев С.А., Лупян Е.А.* Признаки распознавания пахотных земель на основе многолетних рядов данных спутникового спектрорадиометра MODIS // Современные проблемы дистанционного зондирования Земли из космоса. 2010. Т. 7. № 1. С. 330-341.
6. *Стыценко Ф.В., Барталев С.А., Егоров В.А., Лупян Е.А.* Метод оценки степени повреждения лесов пожарами на основе спутниковых данных MODIS // Современные проблемы дистанционного зондирования Земли из космоса. 2013. Т. 10. № 1. С. 254-266.
7. *Nelder J.A., Mead R.* A simplex method for function minimization // Computer Journal. 1965. V. 7. P. 308–313.
8. *Savitzky A., Golay M.J.E.* Smoothing and Differentiation of Data by Simplified Least Squares Procedures // Analytical Chemistry. 1964. V. 36 (8). P. 1627–1639.

Using local polynomial approximation within moving window for remote sensing data time-series smoothing and data gaps recovery

D.E. Plotnikov, T.S. Miklashevich, S.A. Bartalev

Space Research Institute RAS, Moscow, Russia
E-mail: dmitplot@d902.iki.rssi.ru

Remote sensing data provide operative and unbiased information about vegetation state and dynamics. Remote sensing data time series analysis facilitates vegetation types discrimination. However, haze, clouds and their shadows distort spectral reflectance values in Visible, Red and NIR bands. Presently, these hindering factors are well dealt with during data pre-processing and multi-temporal image composing, resulting in data gaps appearing within time series, but residual impurities and instrument noises still disarrange data in time series. Current approaches for time series smoothing deal mostly with data disturbances due to noises and hindering factors, while time series are considered gapless. Besides, these methods do not imply detection and following exclusion of certainly disturbed data. However, data gaps and distortions must be considered jointly to avoid drawbacks of data recovery process through distorted data. This paper describes the use of local polynomial approximation within moving window of variable size jointly for time-series smoothing and data gaps filling.

Keywords: remote sensing, time series, gaps filling, data smoothing.

References

1. Balashov I.V., Burtsev M.A., Efremov V.Yu., Loupian E.A., Proshin A.A., Tolpin V.A. Postroenie arkhivov rezul'tatov obrabotki sputnikovykh dannykh dlya sistem dinamicheskogo formirovaniya proizvodnykh informatsionnykh produktov (Archives of remote sensing data processing results for use in dynamic systems for derivative products creation), *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2008, Issue 5, Vol. 1, pp. 26-32.
2. Bartalev S.A., Egorov V.A., Loupian E.A., Plotnikov D.E., Uvarov I.A. Raspoznavanie pakhotnykh zemel' na osnove mnogoletnikh sputnikovykh dannykh spektroradiometra MODIS i lokal'no-adaptivnoi klassifikatsii (Recognition of arable lands using multi-annual satellite data from spectroradiometer MODIS and locally adaptive supervised classification), *Komp'yuternaya optika*, 2011, Vol. 35, No. 1, pp. 103-116.
3. Loupian E.A., Savin I.Yu., Bartalev S.A., Tolpin V.A., Balashov I.V., Plotnikov D.E. Sputnikovyi servis monitoringa sostoyaniya rastitel'nosti ("Vega") (Satellite service for vegetation monitoring VEGA), *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2011, Vol. 8, No. 1, pp.190-198.
4. Plotnikov D.E., Bartalev S.A., Zharko V.O., Mikhailov V.V., Prosyannikova O.I. Eksperimental'naya otsenka raspoznavaemosti agrokul'tur po dannym sezonnykh sputnikovykh izmerenii spektral'noi yarkosti (An experimental assessment of crop types recognisability using time-series of intra-seasonal spectral reflectance measurements by satellite sensor), *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2011, Vol. 8, No 1, pp. 199-208.
5. Plotnikov D.E., Bartalev S.A., Loupian E.A. Priznaki raspoznavaniya pakhotnykh zemel' na osnove mnogoletnikh ryadov dannykh sputnikovogo spektroradiometra MODIS (The recognition features to map arable lands based on multi-annual MODIS Earth observation data), *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2010, Vol. 7, No. 1, pp. 330-341.
6. Stytsenko F.V., Bartalev S.A., Egorov V.A., Loupian E.A. Metod otsenki stepeni povrezhdeniya lesov pozharami na osnove sputnikovykh dannykh MODIS (Post-fire forest tree mortality assessment method using MODIS satellite data), *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2013, Vol. 10, No. 1, pp. 254-266.
7. Nelder J.A., Mead R. A simplex method for function minimization, *Computer Journal*, 1965, Vol. 7, pp. 308–313.
8. Savitzky A., Golay M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures, *Analytical Chemistry*, 1964, Vol. 36 (8), pp. 1627–1639.