

Детектирование аномальных значений временных рядов вегетационных индексов

Л.Ф. Спивак

*Международный университет природы, общества и человека «Дубна»
Московская область, г. Дубна 141980, Россия
E-mail: levspivak@mail.ru*

При формировании сезонных рядов вегетационных индексов одна из проблем заключается в выявлении и устранении ошибочных значений. В частности, важно уметь отличать ложные аномалии, связанные с устойчивым облачным покровом, от реальных, вызванных экстремальными погодными условиями. В статье описывается алгоритм обнаружения аномальных элементов временного ряда вегетационных индексов, основанный на сравнении «ожидаемых» (априорных) и зарегистрированных (апостериорных) значений измеряемого параметра. Предложена классификация аномалий с учетом специфики пространственных и временных особенностей их проявления. Аномалии подразделяются на локализованные (Л-аномалии), положение которых в пространстве не изменяется, и мигрирующие (М-аномалии). В зависимости от режима различаются устойчивые аномалии (У-аномалии), имеющие постоянно действующий источник, и ситуационные (С-аномалии), возникающие при определенных обстоятельствах. Приведены примеры аномалий различных типов. Ситуационной аномалией следует считать любое отклонение от прогнозируемого значения, которое превышает определенный порог. Что бы обнаружить ситуационную мигрирующую аномалию (СМА), нужно не только регулярно сканировать всю территорию их возможного появления, но и уметь распознавать аномалии на фоне экстремальных сезонных условий. Часто обнаружить аномальные значения недостаточно, нужно еще понять являются ли они достоверными или ошибочными. Описана процедура вычисления «ожидаемых» значений и критерии распознавания ложных аномалий. Рекомендованы схемы восстановления ошибок измерения.

Ключевые слова: временные ряды, вегетационные индексы, устойчивые аномалии, ситуационные аномалии.

Введение

Для корректного распознавания и анализа долговременных изменений растительного покрова необходимо сравнивать многолетние сезонные ряды вегетационных индексов, построенные на основе данных дистанционного зондирования Земли (ДЗЗ). Разумеется, для получения объективных результатов ряды должны быть сформированы по единой технологии, откалиброваны и очищены от ошибок.

При формировании сезонных рядов вегетационных индексов одна из проблем заключается в выявлении и устранении ошибочных значений. В частности, важно уметь отличать ложные аномалии, связанные с устойчивым облачным покровом, от реальных, вызванных экстремальными погодными условиями. Часто обнаружить и исправить ошибки измерений удастся только в результате специального анализа по завершению сезона.

Аналогичные проблемы возникают и при формировании временных рядов других параметров, в частности, температуры поверхности моря или суши.

Основные определения

Традиционно слово «аномалия» интерпретируется как отклонение от нормы, общей закономерности (Даль, онлайн версия). Природные аномалии возникают в пространстве и времени. Особенности проявления аномалий определяются спецификой их источника. Пространственные аномалии - отражают неоднородности распределения значений измеряемого параметра в пределах заданной территории. Следует различать *локализованные* аномалии (Л-аномалии), положение которых в пространстве не изменяется, и *мигрирующие* аномалии (М-аномалии), пространственное положение которых может изменяться.

В свою очередь, временные аномалии характеризуют отклонения в режиме контролируемого процесса и подразделяются на *устойчивые* (У-аномалии), имеющие постоянно действующий источник, и *ситуационные* (С-аномалии), возникающие при определенных внешних обстоятельствах. Таким образом, можно построить следующую классификацию аномалий.

Пространственные Временные (режимные)	Локализованные (Л-аномалии)	Мигрирующие (М-аномалии)
Устойчивые (У-аномалии)	УЛА	УМА
Ситуационные (С-аномалии)	СЛА	СМА

Методы выделения устойчивых локализованных аномалий (УЛА) геофизических полей хорошо разработаны и широко используются при поиске полезных ископаемых (Никитин, 1979).

Зоны повышенных температур в районах глубинных разломов еще один пример УЛА. Отметим, что несмотря на постоянное действие источника, дистанционно обнаружить УЛА бывает не просто. В частности, известная УЛА в районе Семипалатинского испытательного полигона (СИП), приуроченная к Большому Чингизскому разлому, доступна для регистрации из космоса только в короткий период с февраля по апрель. После повышения температуры приземного воздуха аномалию уже не возможно распознать (Sultangazin et al., 1998; Спивак и др., 2006).

Вулканические процессы являются хорошим примером ситуационных локализованных аномалий (СЛА).

Примерами ситуационных мигрирующих аномалий (СМА) могут служить изменения температуры поверхности земли перед землетрясением или резкое изменение значений вегетационных индексов из-за погодных условий. Обнаружить их гораздо сложнее, чем УЛА, особенно когда есть дополнительные маскирующие факторы. В частности, изменения вегетационных индексов могут регистрироваться как в годы с благоприятными,

так и в годы с неблагоприятными погодными условиями. Таким образом, чтобы обнаружить СМА, нужно не только регулярно сканировать всю территорию их возможного появления, но и уметь распознавать аномалии на фоне экстремальных сезонных условий.

С учетом вышеизложенного, аномалией следует считать не только отклонение от типичного (штатного) хода контролируемого процесса, но и отклонение от прогнозного сценария развития ситуации, превышающее некоторое пороговое значение.

Ниже описан алгоритм выделения ситуационных аномалий в процессе формирования сезонного ряда вегетационных индексов на основе разницы между «ожидаемым» (априорным) и зарегистрированным (апостериорным) значениями измеряемого параметра.

Описание алгоритма выделения аномальных значений

Схему расчета «ожидаемых» значений измеряемого параметра, проиллюстрируем *рисунком*, на котором приведены графики сезонной динамики условного вегетационного индекса. Для определенности положим, что это декадные композиты нормализованного дифференциального вегетационного индекса (NDVI). По оси абсцисс откладывается время, по оси ординат – значение параметра.

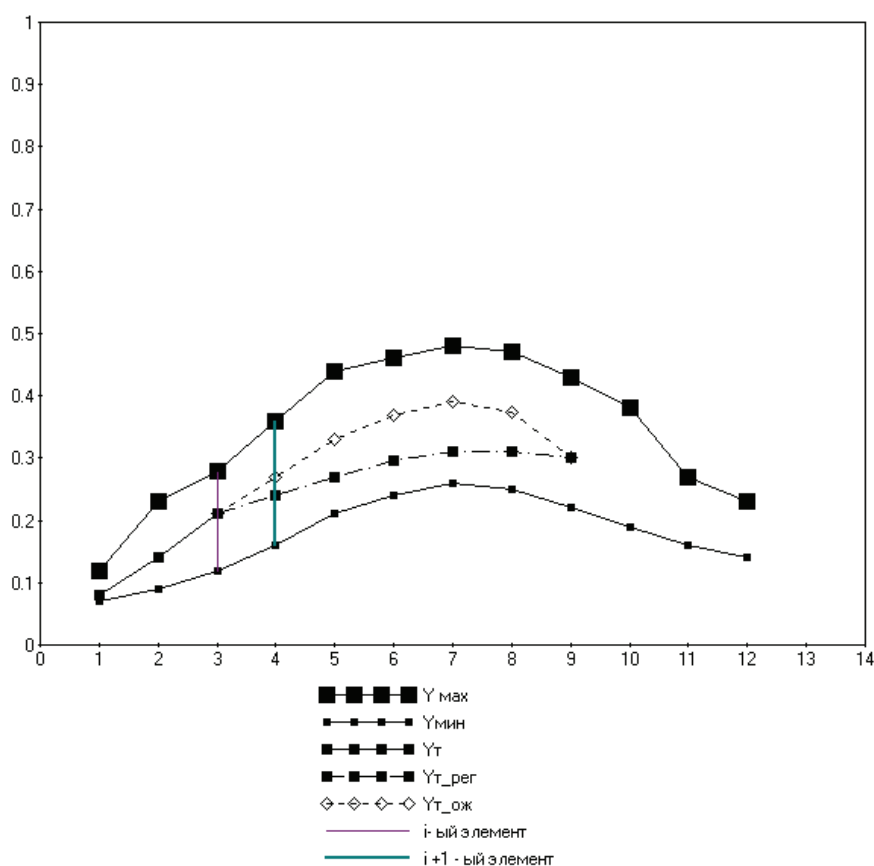


Рис. Иллюстрация к алгоритму расчета «ожидаемых» значений вегетационных индексов

Положим, что в течение сезона регистрируется N значений параметра $Y(i)$ с фиксированным шагом ($i=1, \dots, N$). На каждый временной шаг формируется матрица значений, покрывающая всю область наблюдения. Для простоты будем рассматривать измерения в одной и той же точке (пикселе) и считать, что тип растительного покрова не изменяется.

Для корректной работы алгоритма необходимо соблюдение следующих условий.

1) Известно не менее 2 следующих подряд значений за текущий год, предшествующих аномалии, в достоверности которых нет сомнений.

2) В базе данных многолетних сезонных рядов можно выбрать два сезонных ряда, таких что первый является «ближайшим сверху», а второй - «ближайшим снизу» (см. рисунок).

Введем следующие обозначения:

- $Y_{T_ож}(i+1)$ – ожидаемое (априорное) значение параметра за $i+1$ - период (декаду) текущего сезона;

- $Y_{T_рег}(i+1)$ – зарегистрированное (апостериорное) значение параметра за $i+1$ - период текущего сезона;

- $Y_{max}(i)$ – значение параметра за i - период «ближайшего сверху» года, т.е., $Y_{max}(i) > Y_{T_рег}(i)$ для всех предшествующих периодов от 1 до i ;

- $Y_{min}(i)$ – значение параметра за i - период «ближайшего снизу» года , т.е. $Y_{min}(i) < Y_{T_рег}(i)$ для всех предшествующих периодов от 1 до i .

Нетрудно видеть, что «ожидаемое» значение в момент времени $i+1$ можно вычислить по формуле

$$Y_{T_ож}(i+1) = Y_{min}(i+1) + (Y_{T_рег}(i) - Y_{min}(i))(Y_{max}(i+1) - Y_{min}(i+1)) / (Y_{max}(i) - Y_{min}(i)).$$

Допустим, что в результате измерений в момент времени $i+1$ мы получили зарегистрированное значение параметра $Y_{T_рег}(i+1)$. Что бы определить является ли это значение аномальным нужно поверить ряд условий. В первую очередь необходимо вычислить разницу между ожидаемым и зарегистрированным значением.

Если $|Y_{T_ож}(i+1) - Y_{T_рег}(i+1)| \leq \lambda_{min}$, т.е. расчетное и зарегистрированное значения параметра отличаются незначительно, то в качестве итогового значения можно выбрать любое из них или взять среднее арифметическое.

Если же $|Y_{T_ож}(i+1) - Y_{T_рег}(i+1)| > \lambda_{max}$, т.е. зарегистрированное значение выходит за допустимые пределы, то это свидетельствует об ошибке измерения и в качестве итогового следует принять расчетное значение параметра.

Ситуация, когда $\lambda_{\min} \leq |Y_{T_ож}(i+1) - Y_{T_рег}(i+1)| \leq \lambda_{\max}$, требует дополнительного рассмотрения. При этом необходимо учитывать характер сезонной динамики вегетационных индексов, значения которых вначале возрастают, а затем убывают (см. рис.).

Если мы находимся на восходящей ветви графика, то должно выполняться условие $Y_{T_ож}(i) \leq Y_{T_рег}(i+1)$. При этом возможны два варианта.

1) $Y_{T_ож}(i+1) < Y_{T_рег}(i+1)$, т.е. зарегистрированное значение больше «ожидаемого» на величину превышающую λ_{\min} . В этом случае мы имеем *положительную аномалию*, которая (при условии достоверности) свидетельствует о резком улучшении состояния растительности.

2) $Y_{T_ож}(i+1) - Y_{T_рег}(i+1) > \lambda_{\min}$, т.е. мы имеем *отрицательную аномалию*. Ее причинами могут быть как облачный покров, так и резкое ухудшение состояния растительности.

При $Y_{T_ож}(i) > Y_{T_рег}(i+1)$, мы также имеем отрицательную аномалию.

Аналогичным образом можно определить аномальные значения для нисходящей ветви графика.

Но обнаружить аномальные значения недостаточно, нужно еще понять, являются ли они достоверными или ошибочными. Один из признаков достоверности – длительность регистрации аномалии. Температурные аномалии предвестники крупных землетрясений появляются за одну-две недели до толчка. СМА, связанные с облачным покровом, могут длиться месяц и более. Чтобы учесть фактор времени, введем понятие длины компактного аномального подмножества временного ряда. Обозначим через m количество подряд идущих значений, для которых выполняется условия аномальности. Тогда ограничения на длительность СМА можно записать в виде неравенства: $1 \leq m < M$. Разумеется, ограничения на длины положительных и отрицательных аномалий могут быть разными.

Будем считать аномалию «ложной», если длина компактного аномального подмножества меньше M , т.е. условия аномальности выполняется для точек $i+1, i+2, \dots, i+(M-1)$ и не выполняется для точек i и $i+M$. При этом зарегистрированные наблюдения признаются ошибочными, а в качестве итоговых значений параметра принимаются расчетные значения $Y_{T_ож}(i)$. В противном случае аномалия считается реальной, и в качестве итоговых значений параметра принимаются $Y_{T_рег}(i)$.

Выводы и рекомендации

1. При формировании временных рядов вегетационных индексов аномалией следует считать не отклонение сезонных значений от типичного (среднепогодного) графи-

ка, а отклонение «ожидаемых» (априорных) и зарегистрированных (апостериорных) значений измеряемого параметра, превышающее некоторый порог.

2. В ряде случаев для расчета «ожидаемых» значений достаточно использовать только элементы «ближайшего сверху» или «ближайшего снизу» сезонных рядов.

3. Следует отметить, что для восстановления значений, искаженных из-за сплошной облачности, можно построить более сложные схемы расчета, использующие в частности:

- значения рядом расположенных пикселей из текущей матрицы $Y_T(i)$;
- значения тех же или рядом расположенных пикселей из матриц $Y_T(i-1)$ и $Y_T(i+1)$ текущего сезона;
- значений тех же или рядом расположенных пикселей из матриц $Y_(i-1)$, $Y_(i)$ и $Y_(i+1)$ за другие сезоны;
- различные комбинации этих схем.

4. Не исключено, что процесс восстановления может потребовать привлечения дополнительных наземных данных, включая метеонаблюдения.

Литература

1. *Даль В.* Толковый словарь живого великорусского языка (онлайн версия).
2. *Никитин А.А.* Статистические методы выделения геофизических аномалий. М.: Недра, 1979. 280 с.
3. *Спивак Л.Ф., Витковская И.С., Батырбаева М.Ж., Сагатдинова Г.Н.* Космический мониторинг температурного режима района Семипалатинского испытательного полигона: 10 лет наблюдений // Вестник Национального Ядерного Центра Республики Казахстан. Алматы, 2006. Вып. 2. С.97–105.
4. *Sultangazin U., Zakarin E., Spivak L., Arkhipkin O., Muratova N., Terekhov A.* Monitoring of temperature anomalies in the former Semipalatinsk Nuclear Test Site // Proceedings of the Academy of Sciences. Paris. 1998. T.326. Serie IIb Metodologie, instrumentation. pp.135–140.

Detection of anomalous values of time series of vegetation indices

L.F. Spivak

*“DUBNA” International University, Moscow region 141980, Russia
E-mail: levspivak@mail.ru*

When compiling the seasonal series of vegetation indices, it is a challenge to identify and eliminate erroneous values. It is very important to distinguish false anomalies due to sustainable cloud cover from real ones caused by extreme weather conditions. This paper describes the methodology for detection of anomalous elements of vegetation indices time series, based on comparison of “expected” and registered values of measured parameter. Moreover, the classification of anomalies specific to spatial and temporal characteristics of their manifestation is proposed. Spatially, anomalies are divided into localized (L-anomalies) and migratory (M-anomalies). Depending on the mode of appearance, there are two types of anomalies: permanent anomalies (P-anomaly), having a stable source, and situational anomalies (S-anomaly), arising under certain circumstances. Examples of different types of anomalies are presented. Situational anomaly is considered as any deviation from the predicted value, which exceeds the specified

threshold. To detect situational-migratory anomaly (SMA), apart from regular observations of probable territory of occurrence, it is necessary be able to recognize anomalies on a background of extreme seasonal conditions. Frequently it is not enough to detect the anomalies, we must also understand whether they are valid or erroneous. The calculation procedure of “expected” values and criteria for recognition of false anomalies are described. Recommendations for recovery measurement errors are provided.

Keywords: time series, vegetation indices, permanent anomalies, situational anomalies.

References

1. Dal' V., *Tolkovyj slovar' zhivogo velikoruskogo jazyka: onlajn versija* (Explanatory Dictionary of living Russian Language: on-line version).
2. Nikitin A.A., *Statisticheskie metody vydelenija geofizicheskijh anomalij* (Statistical methods of allocation of geophysical anomalies), Moscow: Nedra, 1979, 280 p.
3. Spivak L.F., Vitkovskaja I.S., Batyrbaeva M.Zh., Sagatdinova G.N., Kosmicheskij monitoring temperaturnogo rezhima rajona Semipalatinskogo ispytatel'nogo poligona: 10 let nabljudenij (Space monitoring of temperature regime Semipalatinsk test site: 10 years of observations), *Vestnik Natsional'nogo Yadernogo Centra Respubliki Kazahstan*, Almaty, 2006, Vol. 2, pp.97–105.
4. Sultangazin U., Zakarin E., Spivak L., Arkhipkin O., Muratova N., Terekhov A., Monitoring of temperature anomalies in the former Semipalatinsk Nuclear Test Site, *Proceedings of the Academy of Sciences*, Paris, 1998, Vol. 326, Serie IIb Metodologie, instrumentation, pp.135–140.