

МАТЕМАТИЧЕСКИЙ ПОДХОД К АНАЛИЗУ АНОМАЛИЙ И КВАЗИПЕРИОДИЧЕСКИХ ВАРИАЦИЙ НА ВРЕМЕННЫХ РЯДАХ

Н.Р. Зелинский¹, Р.Ю. Архипов², Н.Г. Клейменова^{1,3}

¹Институт Физики Земли им. О.Ю. Шмидта РАН, Москва

²Национальный исследовательский Университет «Высшая школа экономики», Москва

³Институт космических исследований РАН, Москва

e-mail: hello_nikita@mail.ru

Аннотация. В анализе временных рядов основными задачами являются: (а) выделение крупномасштабных явлений – трендов и крупных событий, (б) выделение аномалий, в том числе отдельных выбросов, участков резкой смены статистических параметров, частотных аномалий и др., (в) выбор допустимого уровня аномальности, (г) определение статистических параметров шума (например, проверка на нормальность и коррелированность). В работе рассматривается подход по выделению тренда временного ряда в целях прогнозирования. Разделение временного ряда на трендовую и аномальную части позволяет изучать физически разные явления отдельно, прогнозировать временные ряды.

Такой подход может быть применен к анализу данных наблюдений любых геофизических явлений. В данной работе с помощью адаптации методов, применяемых в геофизике, решается задача прогнозирования числа посещения пациентами, не стоящими на динамическом наблюдении, врачей-кардиологов в городских поликлиниках г. Москвы по будням на период 13.01 – 02.04 2016.

Введение

В различных областях науки используются свои “классические” методы выделения трендов. Например, в геофизике это низкочастотная фильтрация, фильтры Калмана и т.п., в эконометрике часто используется конструирование априорной модели тренда и поиск её параметров с помощью регрессии (модель Тейла-Вейджа, модель Хольта-Уинтерса.), автокорреляционный анализ для выделения низкочастотных периодических трендов.

Как в геофизике, так и в эконометрике для выделения трендов применяются различные сглаживания (нелинейные низкочастотные фильтры): осреднения по N точкам, сплайновые сглаживания, экспоненциальное скользящее среднее, вейвлет-сглаживания, сглаживания на основе методов нечеткой логики и т.п.

К задачам выделения тренда можно отнести задачу восстановления импульсного отклика среды в сейсморазведке, которая, в частности, решается с применением деконволюции по форме импульса (свертка с фильтром, обратным к форме импульса).

В данной работе пойдет речь о выделении трендов и прогнозировании временных рядов. С помощью адаптации методов, применяемых в геофизике, решается задача прогнозирования числа посещения пациентами, не стоящими на динамическом наблюдении, врачей-кардиологов в городских поликлиниках г. Москвы (более 300 поликлиник) в период 13.01 – 02.04 2016 на основе анализа статистических данных московских поликлиник за два предшествующих года - 2014-2015. Использовались данные системы ЕМИАС (Единая медицинская информационная система г. Москвы).

Обработка данных

На рис. 1 приведен исходный временной ряд числа посещений (около 900 тыс. посещений в год) пациентами, не стоящими на динамическом наблюдении, врачей-кардиологов в городских поликлиниках г. Москвы в период 01.01.2014 – 31.12.2015. Минимальные значения на графике соответствуют выходным и праздничным дням.

В работе рассматривается свёрточная модель временного ряда

$$S = S_y * w_r, w_r = \alpha \left(\sum_{i=1}^5 k_i + \varepsilon_i \right), \varepsilon \sim N(0, D)$$

Годовая кривая числа дневных посещений (рис. 1) рассматривается как множество недельных кривых, свернутых с одинаковым недельным профилем, осложненным нормально распределенным шумом. Недельный профиль рассматривается как результат среднего по всем 88 неделям, содержащим данным по всем пяти рабочим дням (т.е. без праздников). Для модели нормально распределенного шума такое осреднение дает повышение соотношения сигнал/шум в $\sqrt{88}$ (более чем в девять) раз. Дисперсия

недельного профиля достаточно небольшая (см. доверительный интервал на рис. 2), чтобы можно было использовать его в качестве эталонного импульса для деконволюции.

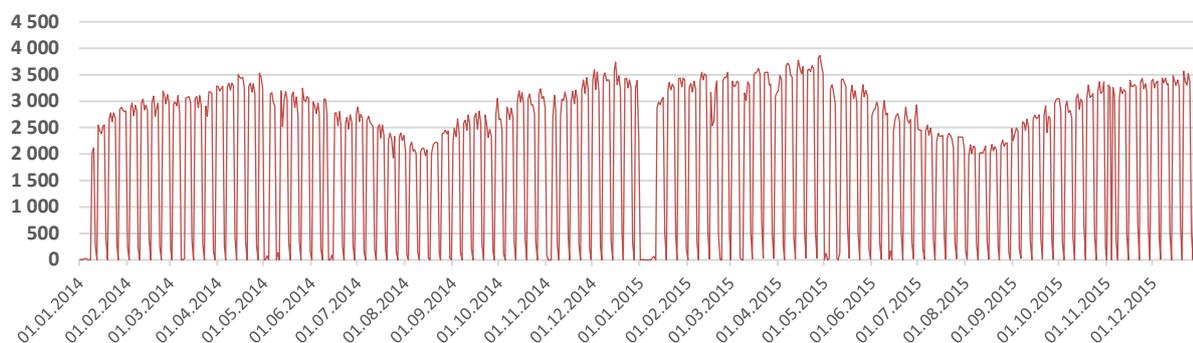


Рисунок 1. Исходный временной ряд – число дневных посещений пациентами, не стоящими на динамическом наблюдении, врачей-кардиологов в городских поликлиниках г. Москвы в период 01.01.2014–31.12.2015.

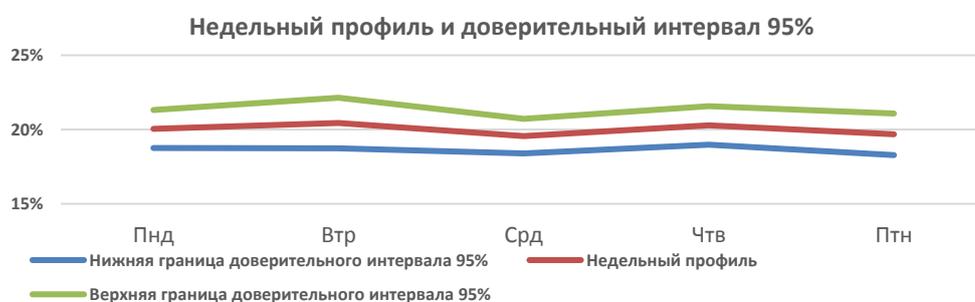


Рисунок 2. Недельный профиль и доверительный интервал 95%, для каждого дня недели распределение веса нормальное по тесту Skewness-Kurtosis [Mardia, 1970; Lin and Mudholkar, 1980].

Для восстановления временного ряда найдем множитель каждой недели методом наименьших квадратов:

$$\alpha = \frac{\sum_{i=1}^5 x_i k_i}{\sum_{i=1}^5 k_i^2}$$

Здесь α – множитель каждой недели, т.е. фактически искомый тренд на недельном уровне. Для недель, содержащих один или два выходных, значения заполняются по той же формуле на αk_i . На рис. 3 приводится сравнение ряда до и после деконволюции (нулевые значения – праздничные дни, когда поликлиники не работали). На рис. 4 показан эффект применения деконволюции для наиболее посещаемого дня недели – вторника.

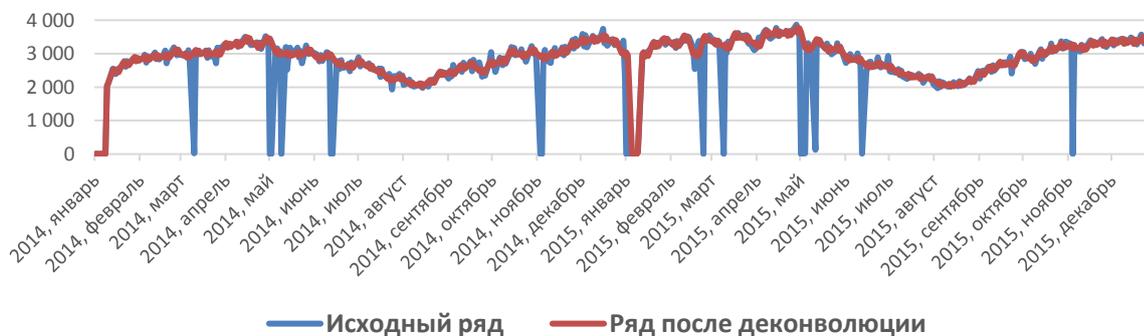


Рисунок 3. Сравнение исходного временного ряда и ряда после деконволюции. По вертикальной оси – число посещений.



Рисунок 4. Сравнение временного ряда из вторников исходного и этого же ряда после деконволюции. По вертикальной оси – число посещений.

Прогнозирование

Прогнозный ряд на 2016 год был получен как полусумма взвешенных 2014 и 2015 годов исходного ряда после деконволюции. Веса определялись числом поликлиник и мощностью, в которых на момент наблюдения был информатизирован кабинет кардиолога, т.е. в систему ЕМИАС поступали данные о приемах пациентов.

На рис. 5 представлены три временных ряда: фактический (синяя кривая), прогнозный по взвешенному ряду (МАРЕ – средняя абсолютная процентная ошибка прогноза – составил 3,49%), прогнозный по ряду после применения деконволюции и взвешивания ряду (МАРЕ 3,41%). МАРЕ определялся только для будних дней. «Недопрогноз» 25.02 и 10.03 объясняется отложенным спросом (невозможность приема кардиологом в праздники при имеющемся спросе на такие приемы приводит к накоплению спроса и увеличению числа приемов в послепраздничные дни) из-за праздников 23.02 и 08.03, что, скорее всего, справедливо и для предпраздничного спроса.

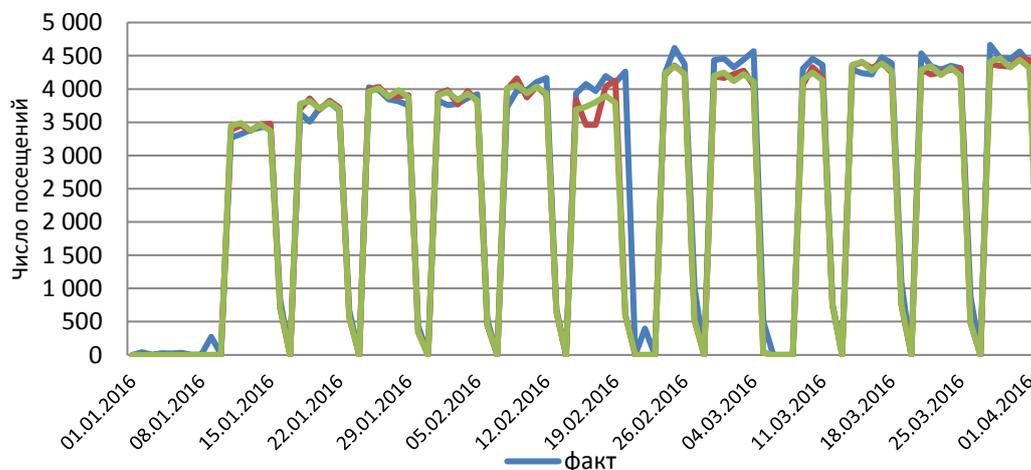


Рисунок 5. Сравнение прогноза на 2016 год с применением деконволюции к исходному ряду и без.

Несмотря на практически одинаковый показатель средней относительной процентной ошибки прогноза (МАРЕ), применение деконволюции снижает количество грубых ошибок (рис. 6).

Выводы

В работе показано, что применение традиционно геофизического инструмента – деконволюции по форме импульса – оправдано для использования в анализе медицинских данных. В частности, на основе анализа предшествующего реализованного спроса (число ежедневных обращений больных к врачам-кардиологам в поликлиниках Москвы в 2014-2015 гг.) был разработан прогноз таких обращений на 2016 г. Сопоставление прогнозируемых величин с реальным числом обращений в первом квартале 2016 г. показало высокую степень достоверности прогноза, средняя абсолютная процентная ошибка прогноза составила всего 3,49%. Применение деконволюции по форме недельного профиля снижает число грубых (более 10%) ошибок прогноза.

Полученные распределения сезонного положения максимумов и минимумов числа обращений к врачам-кардиологам хорошо согласуется с многолетними наблюдениями смертности от инфаркта миокарда в Болгарии и вызовов скорой помощи по поводу инфарктов в Москве [Иванова и др., 2002, Kleimenova et al., 2007; Клейменова и Козырева, 2008, Клейменова, 2013].

Установлено, что наиболее посещаемым днем недели оказался вторник, однако, на кривой посещаемости во вторник также наблюдаются летние минимумы и зимние максимумы.

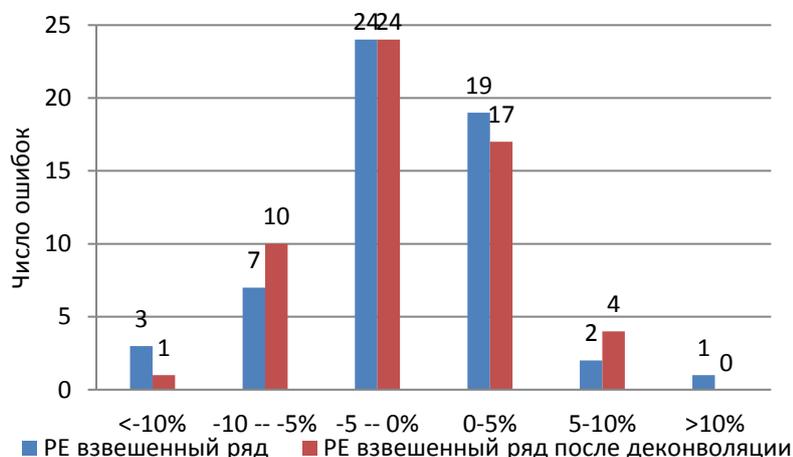


Рисунок 6. Распределение ошибок прогноза на 2016 год с применением деконволюции к исходному ряду и без. Применение деконволюции позволило снизить количество больших ошибок прогноза (более 10%) с четырех до одной.

Благодарности. Исследование было поддержано грантом РФФИ № 16-35-00069 (Н.З.) и частично Программой № 7 Президиума РАН (Н.К.). Авторы также благодарят ДИТ г. Москвы и ИАЦ ДЗМ г. Москвы за разрешение использовать данные системы ЕМИАС.

Список литературы

- Иванова П., Клейменова Н.Г., Гамбурцев А.Г. Смертность от инфаркта миокарда в Болгарии // Атлас временных вариаций природных, антропогенных и социальных процессов. Т.3 “Природные и социальные сферы как части окружающей среды и объекты воздействий” М.: “Янус-К”, 561-563, 2002.
- Канасевич Э.Р. Анализ временных последовательностей в геофизике. - М.: Недра, 1985.- 300 с.
- Клейменова Н.Г. Влияние космической погоды на человека // Земля и Вселенная, № 6, 74-82, 2013.
- Клейменова Н.Г., Козырева. О.В. Магнитные бури и инфаркты. Всегда ли бури опасны? // Геофизические процессы и биосфера, Т.7, №3, 5-24, 2008.
- Kleimenova N.G., Kozyreva O.V., Breus T.K., Rapoport S.I. Pc1 geomagnetic pulsations as a potential hazard of the myocardial infarction // J. Atmos. Solar-Terr. Phys. 69, 1759–1764. 2007.
- Lin C.C., Mudholkar G.S. A simple test for normality against asymmetric alternatives // Biometrika. 67 (2), 455–461. doi:10.1093/biomet/67.2.455. 1980.
- Mardia K.V. Measures of multivariate skewness and kurtosis with applications //Biometrika 57, 519–530. 1970.