

ВОССТАНОВЛЕНИЕ ПРОПУСКОВ В ДАННЫХ МЕТЕОРОЛОГИЧЕСКИХ СТАНЦИЙ МЕТОДОМ АНАЛОГОВ

Восканян К.Л.¹, Крюкова С.В.¹, Кузнецов А.Д.¹, Сероухова О.С.¹, Симакина Т.Е.¹

¹ – *Российский государственный гидрометеорологический университет, Санкт-Петербург, Россия. kvosia@mail.ru*

Аннотация. Рассмотрена возможность использования метода поиска аналогов для восстановления пропущенных данных в измерениях метеорологических автоматических станций.

Ключевые слова: временные ряды метеорологических величин, заполнения пропусков во временных рядах, метод аналогов.

Анализ и прогноз погоды и климата базируется на данных измерений, точность которых обусловлена метрологической надежностью средств измерений. Метеорологическая техника эксплуатируется на открытом воздухе в условиях меняющейся агрессивности внешней среды [1, 2]. Сбои в работе аппаратуры приводят к «выбросам» значений временных рядов или к пропускам данных. Исключение и игнорирование пропущенных значений искажает статистические свойства выборки и, как следствие, ухудшает результаты, получаемые на ее основе [3]. С другой стороны, неудачный выбор метода заполнения пропусков может не только не улучшить, но и сильно ухудшить результаты.

Существует множество методов восстановления, базирующихся, в частности, на использовании аналитических, статистических и эвристических методов [4, 5]. В данной работе для восстановления пропусков в ряду метеоизмерений предлагается использовать метод аналогов [6]. Он заключается в поиске в пределах обучающей выборки отрезков – аналогов заданному на основе использования количественных характеристик их «близости». При этом мера сходства («похожести», «адекватности») определяется с помощью задания соответствующей метрики. В данной работе для оценки близости двух отрезков временного ряда использовались: Евклидова метрика, Манхэттенское расстояние, коэффициент корреляции, сумма относительных абсолютных отклонений и др.

Для восстановления интервала с пропусками данных в исходном ряду ищутся два аналога: первый подбирается к отрезку временного ряда, предшествующему разрыву в данных, второй – следующему за разрывом. По первому выполняется восстановление левой половины пропущенного интервала, по второму – правой. Другими словами, осуществляется прогноз вперед на пол-интервала с пропусками данных и назад на пол-интервала. Степень «похожести» аналогов эталонным отрезкам выбиралась в результате численных экспериментов с временным рядом приземной температуры воздуха, полученного автоматической метеорологической станцией «Погода», установленной на территории РГГМУ [1]. При этом из целого исходного ряда температур окружающего воздуха, идущих с шагом 15 минут, удалялась серия наблюдений, количество измерений в которой варьировалось от 10 до 20. Далее методом аналогов осуществлялось восстановление пропущенных значений, которые сравнивались с результатами измерений. Наилучшие результаты были получены при использовании Манхэттенского расстояния для оценки меры близости двух отрезков временного ряда. В качестве примера в табл. 1 приведены статистические характеристики точности восстановления пропуска с использованием различных методов. Средняя ошибка восстановления, которая зависит от продолжительности пропущенного интервала и его положения во временном ряду, для разрывов длительностью в 2-3 часа колеблется от 0.03⁰С до около 0.5⁰С.

Для сравнения с предлагаемым методом аналогов пропуски восстановлены методом линейной интерполяции и повторением результата последнего наблюдения. Повтор последнего измерения является по сути инерционным прогнозом. Ошибки восстановления данными методами, также приведенные в табл. 1, свидетельствуют о преимуществе метода аналогов.

Таким образом, метод аналогов рекомендуется для восстановления пропусков во временных рядах метеовеличин, измеряемых автоматическими станциями.

Таблица 1. Пример точности восстановления пропуска временного ряда температуры

Показатели	Метод аналогов	Метод линейной интерполяции	Инерционный прогноз комбинированный	Инерционный прогноз (слева)	Инерционный прогноз (справа)
Среднеарифметическая ошибка восстановления, °С	0.07	-0.48	-0.48	0.97	-1.93
СКО ошибки восстановления, °С	0.22	0.34	1.03	0.79	0.79

Литература

1. Восканян К.Л., Кузнецов А.Д., Сероухова О.С. Автоматические метеорологические станции. Часть 1. Тактико-технические характеристики. // Учебное пособие. – СПб.: РГГМУ, 2016. – 195 с. ISBN 978-5-86813-421-0
2. Восканян К.Л., Кузнецов А.Д., Сероухова О.С. Автоматические метеорологические станции. Часть 2. Цифровая обработка данных автоматических метеорологических станций. // Практикум. — СПб.: РГГМУ, 2016. — 99 с.
3. Головкин В.В. Об истории автоматизации первичной обработки и накопления текущей режимной метеорологической информации. // Труды ВНИИГМИ-МЦД, 2000. Вып. 166. С.
4. Восканян К.Л., Дивинский Л.И., Кузнецов А.Д., Сероухова О.С., Симакина Т.Е. Применение сплайн-экстраполяции для текущего прогнозирования приземной температуры воздуха // Труды военно-космической академии им. А.Ф. Можайского, вып. 647, июнь 2015, с. 88-921.
5. Алдошкина Е.С., Кузнецов А.Д., Пугачев А.А., Сероухова О.С., Симакина Т.Е., Чукин В.В. Опыт использования аппарата нейронных сетей для анализа и прогноза временного ряда температуры воздуха // Ученые записки РГГМУ. — СПб: изд-во РГГМУ, 2009. — №11. — С. 91–100.
6. Восканян К.Л., Кузнецов А.Д., Сероухова О.С., Симакина Т.Е. Текущее прогнозирование экологических измерений на основе поиска аналогов // Сборник тезисов XI научно-практической международной конференции «Естественные и антропогенные аэрозоли», СПб., 16-18 октября 2018, с.16, ISBN 978-5-00045-613-2

RECOVERY OF GAPS IN THE DATA OF METEOROLOGICAL STATIONS USING THE ANALOG METHOD

Voskanyan K.L.¹, Kryukova S.V.¹, Kuznetsov A.D.¹, Serouhova O.S.¹, Simakina T.E.¹

¹ – Russian State Hydrometeorological University, Saint-Petersburg, Russia. kvosia@mail.ru

Abstract: the possibility of using the analog search method for recovering missing data in measurements of meteorological automatic stations is considered.

Key words: time series of meteorological variables, filling gaps in the time series, the method of analogues.