МЕТОДИКА ОПРЕДЕЛЕНИЯ ТРЕНДОВ ВО ВРЕМЕННЫХ РЯДАХ В ПРИМЕНЕНИИ К АНАЛИЗУ МНОГОЛЕТНЕГО МАССИВА ДАННЫХ НАБЛЮДЕНИЙ ЗА ОСАДКАМИ

В.Л. Посошков

Морской гидрофизический институт
г. Севастополь, ул. Капитанская, 2
E-mail: vla_po@mail.ru

В статье выполнен анализ трендов во временных рядах на основе объективного метода идентификации тренда. Примененная методика позволила определить характер тренда в рядах осадков XX столетия в 22 пунктах исследуемого региона.

Введение. В связи с глобальным потеплением климата все возрастающий интерес приобретает исследование временных трендов метеорологических параметров. При этом особое внимание уделяется поиску антропогенной составляющей в трендах, связанной с прогрессирующей энергоемкой деятельностью человечества.

Как правило, исследование трендов начинается с визуального просмотра временных рядов. Это позволяет выявить долговременные тенденции, а также другие особенности. Однако, в этих оценках много субъективного. Для того чтобы минимизировать этот фактор, существуют объективные критерии для оценки трендов во временных рядах наблюдений. Одной из таких методик является статистика Манн-Кендалла (МК) [1, 2]. Она широко используется, в частности, в гидрологии [3] и метеорологии [4]. В данной работе эта методика используется для идентификации трендов во временных рядах наблюдений месячных сумм осадков в ряде пунктов Юго-Восточного сектора Европы за период 1900 – 2000 гг.

Краткое описание метода. МК тест является непараметрическим тестом для определения тренда и основан на ранжировании данных наблюдений. Этот тест сравнивает в большей степени относительные величины в выборке данных, нежели сами данные. Тот факт, что тест является непараметрическим (и это – одно из преимуществ данного теста) означает, что не требуется каких-либо особых предположений о распределении вероятностей в данных (в том числе о нормальности распределения). Использование рангов данных вместо самих данных наблюдений помогает минимизировать потенциальную возможность вклада за счет ошибок измерений и позволяет распознать монотонно возрастающий или уменьшающийся тренд, который априори не обязательно линейный. Более того, пропуски в наблюдениях можно присоединить к данным, присвоив им одинаковые величины, меньше, чем наименьшие (из измеренных) в имеющемся ряде данных. Процедура предполагает, что существует лишь единственное измерение в дискретном временном интервале. Когда существует множество таких измерений за наблюдаемый период, используется медианное значение. Итак, МК тест широко используется благодаря тому, что он является простым, робастным, аккумулирующим ошибочные данные, а также данные, не требующие подчинения каким-либо статистическим распределениям. Следует подчеркнуть, что МК тест не оценивает саму величину тренда, а только его наличие.

Величины данных оцениваются как значения в упорядоченном (ранжированном) временном ряде данных. Каждая цифра ряда сравнивается со всеми величинами ряда. Первоначальной величине статистики МК S присваивается значение 0 (т.е. нет тренда). Если величина за более поздний временной период выше, чем за предшествующий период, S увеличивается на единицу. С другой стороны, если последующая величина меньше, чем предшествующая, S уменьшается на единицу. Итоговый результат таких увеличений и уменьшений дает окончательную величину S. Итак, имеем

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sign(x_j - x_i), \quad (1)$$

где x_j – соответствующие величины данных; n – длина ряда;

$$sign(\theta) = \begin{cases} 1, \text{при } \theta > 0, \\ 0, \text{при } \theta = 0, \\ -1, \text{при } \theta < 0 \end{cases}$$

– знаковая функция. Статистика S при $n > 7$ приближенно описывается нормальным распределением со средним значением E и дисперсией VAR:

$$E(S) = 0, \tag{2}$$

$$VAR(S) = \frac{n(n-1)(2n+5) - \sum_{p=1}^{g} t_p(t_p - 1)(2t_p + 5)}{18}, \tag{3}$$

где n – число точек данных; g – число совпадающих групп (совпадающая группа – это часть выборки данных, имеющая равные значения); t_p – число точек данных в p-той группе.

Нормализованная статистика (Z) МК-теста (с нулевым средним значением и единичной дисперсией), а также статистическая значимость одностороннего теста (p) даются соответственно формулами:

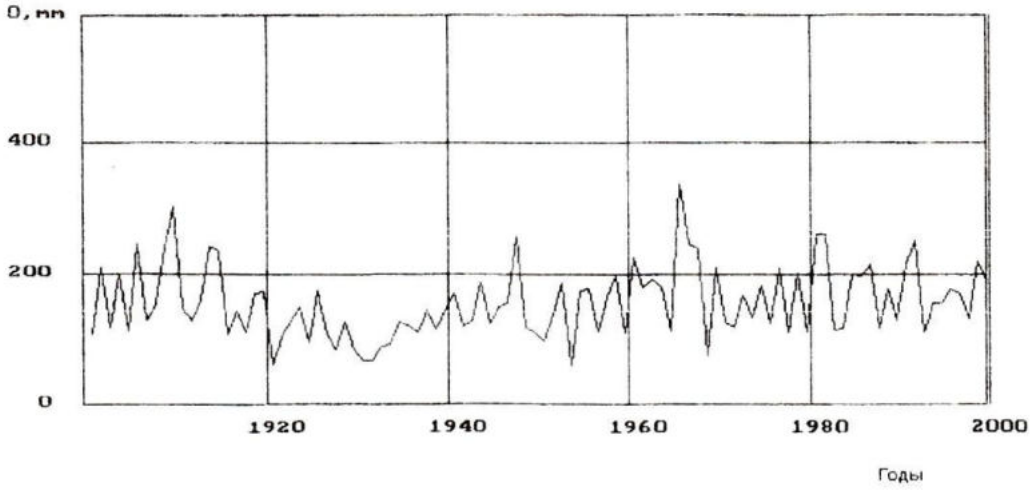
$$Z = \begin{cases} \dfrac{S-1}{\sqrt{VAR(S)}}, S > 0 \\ 0, S = 0 \\ \dfrac{S+1}{\sqrt{VAR(S)}}, S < 0 \end{cases}, \tag{4}$$

$$p = 0{,}5 - \Phi(|Z|), \tag{5}$$

где $\Phi(Z)$ – функция Лапласа [5].

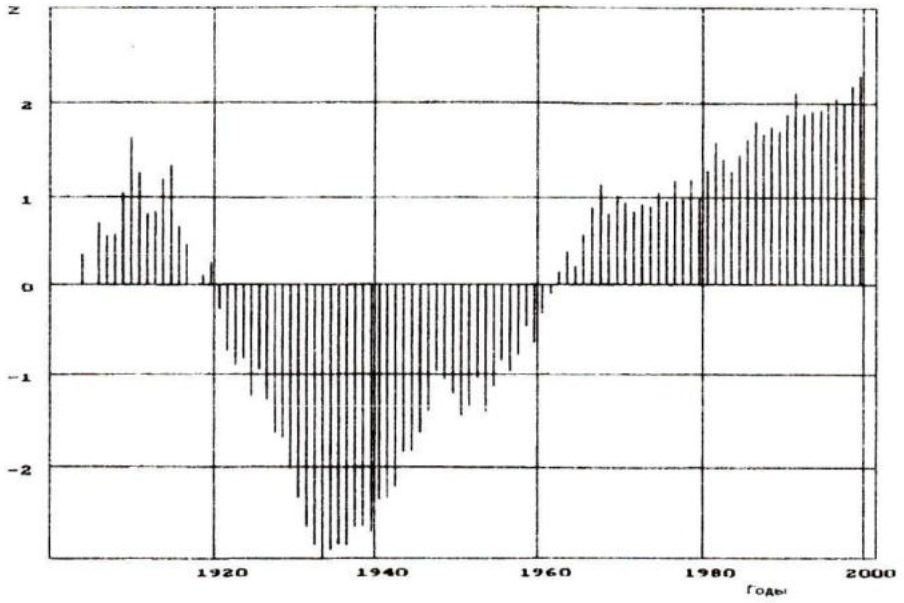
Положительное или отрицательное значение Z свидетельствует о присутствии соответственно положительного или отрицательного тренда. Обычно выбирается уровень значимости $p = 0{,}05$ (95% доверительный интервал).

Результаты. В качестве иллюстрации, разберем изложенную методику на конкретном примере. На рис. 1 показан временной ряд (1900 – 2000 гг.) суммы осадков за холодный период (ХП, ноябрь – март), полученный на основе наблюдений в Луганске. Это типичный случай, когда визуально трудно определить, существует ли тренд за 100-летний период. Хотя на протяжении века видны отрезки порядка 30 лет (период, принятый как приемлемый для оценки климатических характеристик), на протяжении которых отчетливо видны тенденции роста или уменьшения количества осадков.



Р и с. 1. Временной ряд суммы осадков за холодный период

Перейдем теперь к объективным числовым оценкам. На рис. 2 показан временной ход за исследуемый период нормированной статистики МК (функция Z, формула (4)), рассчитанной для временного ряда рис. 2. Максимальная абсолютная величина Z до 1940 г. (тренд отрицательный) достигает трех среднеквадратичных отклонений. Для нормального распределения обеспеченность нулевой гипотезы (что величина тренда обусловлена естественным выборочным разбросом) составляет при этом более 99 % (правило « трех сигм»).



Р и с. 2. Нормированная функция *Z* статистики Манн Кендалла в зависимости от длины рассматриваемого периода. Луганск

Таблица 1

Сводная таблица оценок присутствия тренда (по Манн Кендаллу) во временных рядах атмосферных осадков в различных пунктах исследуемого региона. «+» – положительный тренд, «–» – отрицательный, «0» – отсутствие тренда (с доверительной вероятностью 95 %). ХП – холодный период года, ТП – теплый период, ГОД – весь год

Географический пункт	ХП	ТП	ГОД
Донецк	0	0	0
Евпатория	0	0	0
Золотоноша	+	0	+
Керчь	+	0	0
Киев	0	0	0
Лозовая	0	0	0
Луганск	+	0	0
Львов	+	0	0
Мелитополь	0	0	0
Одесса	+	0	+
Полтава	+	0	0
Рава Русская	–	0	0
Симферополь	0	0	0
Ужгород	0	–	–
Херсон	+	0	+

Величина Z за весь период превышает «две сигмы», что укладывается в пределы 92 % доверительной вероятности. Здесь нулевая гипотеза H_0 (что тренд образуется за счет случайного характера процесса) также отвергается и можно говорить о присутствии положительного тренда.

В табл. 1 приводятся результаты идентификации тренда сезонных и годовых сумм осадков в 15 пунктах региона, полученных согласно методике, описанной выше.

В таблице проведено разбиение сезонов на два периода: холодный (ноябрь–март, 5 месяцев) и теплый (апрель–октябрь, 7 месяцев). Кроме того, приведены результаты и за суммарные годовые величины.

Прежде всего, обращает на себя внимание тот факт, что в теплый период во всех рассмотренных пунктах тренды отсутствуют. В одной трети данных тренды присутствуют в холодный период. При этом тренды имеют положительный знак (возрастание количества осадков за столетие), за исключением Равы Русской, где тренд отрицательный. Возможно, это объясняется особенностью физико-географического расположения данного пункта (близость к Карпатам), где определяющую роль играет орография местности. То же самое можно сказать об Ужгороде, где годовая сумма осадков также имеет отрицательный тренд.

В табл. 1 использованы не все данные, имеющиеся в распоряжении. В нее не включены данные, имеющие множество пропусков в наблюдениях. В первую очередь эти пропуски связаны с отсутствием или неполным объемом наблюдений в годы Великой Отечественной Войны. Некоторая часть данных исключена из-за недостаточной организации наблюдений в предвоенный период. Таким образом, регион доступных наблюдений в послевоенное время (1945 – 2000 гг.) становится шире. Кроме того, этот период интереснее с той точки зрения, что на него приходится более весомое влияние антропогенного фактора.

Таблица 2

Сводная таблица оценок присутствия тренда (по Манн Кендаллу) во временных рядах атмосферных осадков в различных пунктах исследуемого региона за период 1945 – 2000 гг.

Географический пункт	ХП	ТП	ГОД
Винница	0	0	0
Донецк	–	+	0
Евпатория	0	0	0
Запорожье	0	0	0
Золотоноша	+	0	+
Керчь	0	0	0
Киев	0	0	0
Кировоград	0	0	0
Лозовая	0	0	0
Луганск	+	0	+
Львов	+	0	0
Мелитополь	0	0	0
Николаев	0	0	0
Одесса	0	0	0
Полтава	0	0	0
Рава Русская	–	0	0
Симферополь	0	0	0
Сумы	0	0	0
Ужгород	0	0	0
Умань	0	+	0
Херсон	0	+	+
Шепетовка	0	+	+

Поэтому тестирование на тренд было проведено отдельно за послевоенный период (в 22 пунктах). Результаты приведены в табл. 2.

Вопреки ожиданиям, здесь картина более пестрая. Положительный тренд в холодный период года сохранился лишь во Львове и Золотоноше, зато в теплый период года наблюдается в 4-х пунктах. Уместно подчеркнуть, что критерий для идентификации тренда выбран достаточно жесткий, а именно: 95 % доверительная вероятность. Если снизить требования, хотя бы до 90 %, статистика трендов (табл. 2) может измениться.

Была предпринята попытка более жесткого разбиения на календарные сезоны: «зима/лето» вместо «ХП/ТП» (период 1945-2000 гг.). При таком разбиении явных трендов не обнаружено, за исключением Николаева, где проявляется заметный отрицательный тренд в летнее время с доверительной вероятностью 99 %. Во временных рядах зимних осадков более или менее положительный тренд в Луганске (доверительная вероятность 92 %), Шепетовке (76 %) и Золотоноше (78 %). К ним частично примыкает Херсон. По таким же признакам тренда эти пункты группируются и в летнее время.

Учитывая кластерную структуру пространственного распределения осадков, было проведено разбиение региона на три группы (по качественным признакам трендов) с последующим осреднением внутри каждой группы. В первую группу объединили упомянутые выше пункты: Шепетовка, Луганск, Золотоноша, Херсон. Ввиду особого географического положения (близость к Карпатам) и больших годовых норм осадков, в отдельную группу вошли Львов, Рава Русская и Ужгород. Оставшиеся 14 пунктов были объединены в третью. Анализ трендов по этим трем кластерам дал следующие результаты.

Первая группа – *зима*. Положительный тренд с вероятностью 76 %. Максимальный рост 1945 – 1970 гг. Вероятность 99,8 %. Далее без тренда. *Лето*. Положительный тренд с вероятностью 87 %.

Вторая группа – *зима*. Отрицательный тренд с вероятностью 34 %. Положительный на отрезке 1950 – 1970 гг. *Лето*. Отрицательный тренд с малой обеспеченностью в 28 %. Рост на периоде 1960 – 1980 гг.

Третья группа – *зима*. Отрицательный тренд. Вероятность 32 %. Заметный рост за период 1950 – 1970 гг. Здесь доверительная вероятность составляет 98 %. *Лето*. Отрицательный тренд. С малой обеспеченностью 11 %.

Заключение. Итак, был выполнен анализ трендов во временных рядах месячных сумм осадков в 22 пунктах территории, примыкающей к Черному морю и севернее вплоть до горных Карпат. Чтобы исключить годовую периодичность использовалось разбиение на сезоны. В качестве объективного критерия для идентификации трендов применялась методика Манн Кендалла. Следует подчеркнуть, что данная методика лишь идентифицирует тренд, но не определяет саму величину тренда. Анализ проводится на двух временных интервалах все столетие и только послевоенный период. При этом явного антропогенного влияния на тренды, связанного с бурным развитием технического прогресса в послевоенное время, не обнаружено. Это подтверждается, в частности, и зарубежными авторами на основании результатов, полученных в других регионах, в частности при анализе трендов во временных рядах атмосферных осадков на территории Турции [4].

СПИСОК ЛИТЕРАТУРЫ

1. *Kendall, M.G.* Rank correlation methods // 4th. Charles Griffin, London. – 1975.
2. *Mann, H.B.* Nonparametric tests against trend // Econometrica. – 1945. – 13 – P. 245 – 259.
3. *Yue, S., Pilon, P. & Cavadias, G.* Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series // J. Hydrol. – 2002. – 259. – P. 254 – 271.
4. *Turgay Partal and Ercan Kahya.* Trend analysis in Turkish precipitation data // Hydrol. Process. – 2006. – 20. – P. 2011 – 2026.
5. *Gilbert, R.O.* Statistical methods for environmental pollution monitoring. // Van Nostrand Reinhold, New York. – 1987. – 336 pp.