

АНАЛИЗ МЕТОДОВ ВОССТАНОВЛЕНИЯ ПРОПУЩЕННЫХ ЗНАЧЕНИЙ ВО ВРЕМЕННЫХ РЯДАХ

Кухаревич М. Ф.

Учреждение образования «Брестский государственный технический университет», г. Брест, Республика Беларусь, kukharevichmikhail@gmail.com
Научный руководитель – Шешко Н. Н., к.т.н., доцент

The paper deals with the methods of restoring the missing values on the example of time series of water temperatures of the Lan river in the period from 1999 to 2017. On the basis of the data obtained as a result of recovery, the analysis of the restored values is performed.

Процесс исследования некоторых характеристик водного объекта зачастую предусматривает под собой обработку большого количества данных, описывающих тем или иным образом рассматриваемые характеристики. Ранее этот процесс был достаточно трудоемким и требовал значительных затрат времени. В настоящее время использование современных технологий позволяют полностью или частично решить некоторые проблемы, связанные с обработкой данных, и таким образом получить более достоверные результаты исследований.

Особый интерес при работе с данными представляет проблема обработки и восстановления пропущенных значений. Данной проблеме посвящено немалое количество работ [1–3]. В некоторых работах [1,2] утверждается, что наиболее простым решением данной проблемы является исключение пропущенных значений, однако это может приводить к искажению величин статистических характеристик и, как результат, привести к осложнению их математической обработки [1].

В качестве исходного материала для выполнения восстановления в данном исследовании использовались данные среднемесячных температур воды за период с 1999 по 2017 год. Дополнительно для целей восстановления температур воды использовались среднемесячные температуры почвы и воздуха за аналогичный период. Значения температур воды были взяты из «Гидрологического ежегодника» и Государственного водного кадастра «Ежегодные данные о режиме и ресурсах поверхностных вод суши» по реке Лань. Значения температур воздуха и почвы были получены из Климатического кадастра Республики Беларусь «Метеорологический ежемесячник» по двум метеорологическим станциям: Житковичи и Полеская.

Причины появления пропусков могут иметь различный характер, что необходимо учитывать при выборе метода восстановления. Так, пропуски могут быть обусловлены неполадкой измерительного оборудования, наличием неблагоприятных погодных условий, присутствием ошибок, потерей данных и другими причинами [2]. Анализ пропусков исходного временного ряда температур воды показал, что имеющиеся пропуски не имеют случайного характера и выпадают на период ноябрь – март. При этом в некоторых годах имеются пропуски за несколько месяцев подряд (рисунок 1).

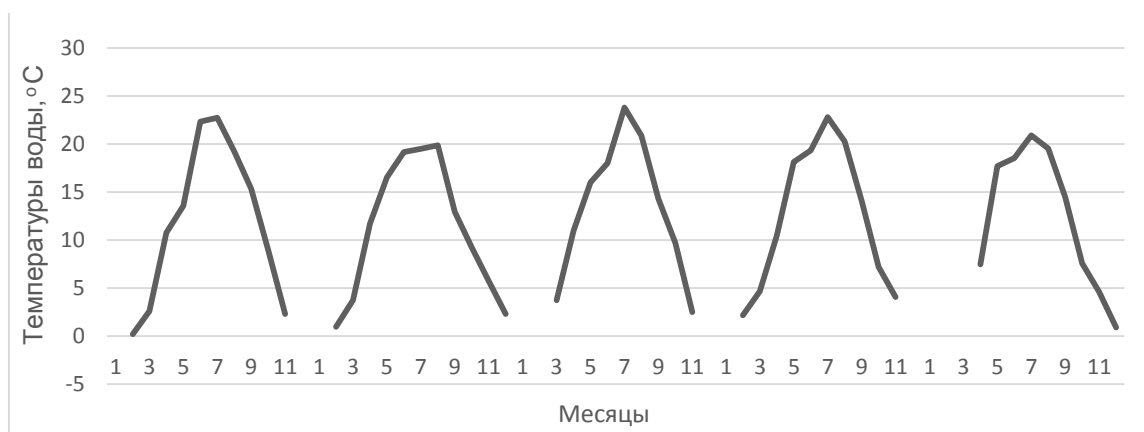


Рисунок 1 – График изменения температуры воды

В настоящее время существует множество методов восстановления пропусков во временных рядах. Так, во многих исследованиях в качестве методов восстановления использовались методы интерполяции [1,2] и регрессии [3]. Каждый из них имеет свою специфику, определяющую их преимущества и недостатки.

Интерполяция. В ходе исследования в качестве метода восстановления была использована интерполяция сплайном 1–го порядка (линейная интерполяция), 2–го и 3–го порядка (рисунок 2). Линейная интерполяция в некоторых случаях дает весьма хорошие результаты, однако данное исследование показало ее неприменимость, что обусловлено двумя причинами:

1. Пропущенные значения выпадают на пиковое значение [1].
2. Имеется несколько пропущенных значений подряд (характерно для сплайн-интерполяции) [2].

Более реальный ход изменения температуры воды показали значения восстановленные методами сплайн интерполяции 2-го и 3-го порядка. Однако некоторые восстановленные значения имели значения ниже 0 °С, что явно не соответствует действительности из-за специфики охлаждения вод в водных объектах.

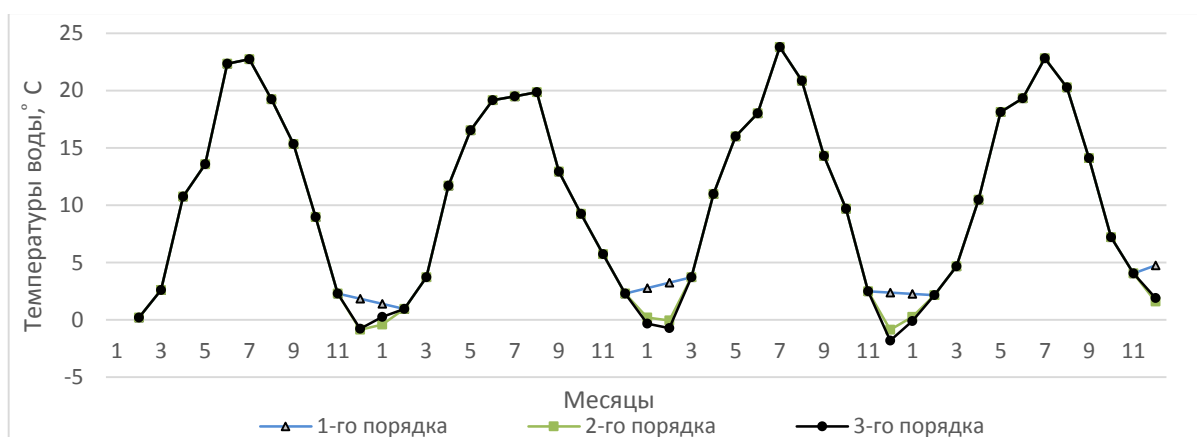


Рисунок 2 – График изменения температуры воды с восстановлением пропущенных значений сплайн-интерполяцией n-го порядка

Регрессия. В процессе исследования были выполнены попытки восстановления пропущенных значений с помощью линейной однофакторной и

множественной регрессии, а также авторегрессии. Наиболее предпочтительной является однофакторная регрессия, а после – многофакторная и авторегрессия [3]. В качестве зависимых факторов для однофакторной и многофакторной регрессии использовались температуры воздуха и почвы. Результаты восстановления представлены на рисунке 3.

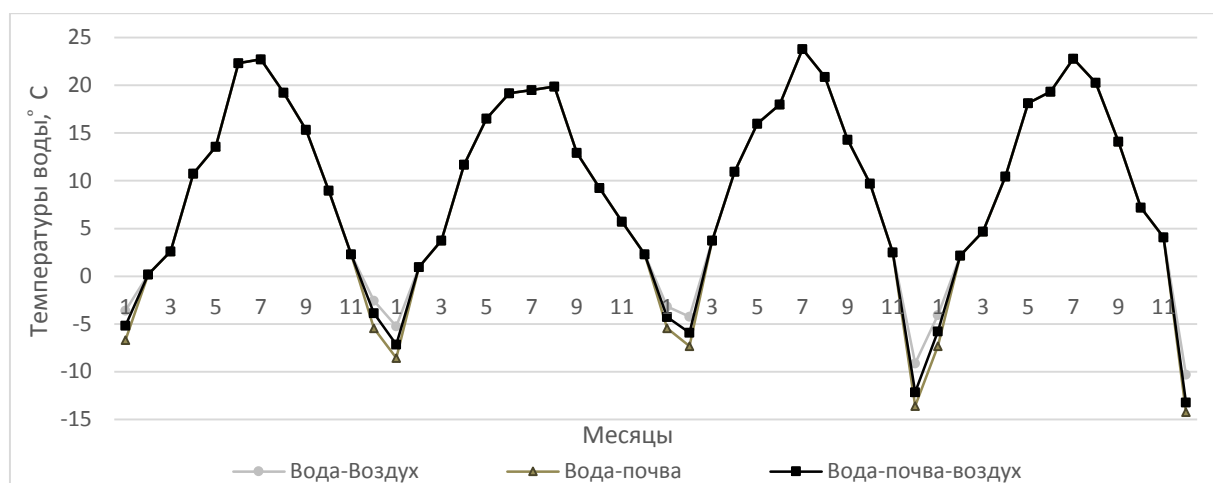


Рисунок 3 – График изменения температуры воды с восстановлением пропущенных значений регрессионными моделями

Наиболее правдоподобные результаты были получены в результате восстановления методом линейной однофакторной регрессии по параметрам вода – воздух. Однако, как и в случае со сплайн-интерполяциями 2–го и 3–го порядка восстановленные значения местами имеют значения ниже 0 °C, что обуславливают неприменимость данной модели. Восстановление с помощью однофакторной и многофакторной регрессии по температурам воздуха и почвы дает очень достоверные результаты лишь при восстановлении пропусков за теплый период года, так как значения температур воды за этот период зачастую очень хорошо коррелируют с температурами почвы и воздуха. По этой причине наиболее приемлемым видится использования метода авторегрессии. Однако по причине того, что пропущенные значения выпадают на одни и те же месяца использование авторегрессии 1–го и более высоких порядков является невозможным.

Список использованных источников

1. Радчикова, Е. С. Анализ применения способов заполнения пропусков в данных во временных рядах в экологических исследованиях / Е. С. Радчикова // Экология и защита окружающей среды: сб. тез. докл. Междунар. науч.-практ. конф., 19-20 марта 2014 г. – Минск, 2014. – С. 112–116.
2. Круглов, В. В. Методы восстановления пропусков в массивах данных / В. В. Круглов, И. В. Абраменкова // Программные продукты и системы. – 2005. – № 2. – С. 16–20.
3. Волчек, А. А. Методика и алгоритм приведения временных рядов уровней грунтовых вод к расчетным (на примере природно–территориального комплекса "Беловежская пуца") / А. А. Волчек, Н. Н. Шешко // Вестник Брестского государственного технического университета. Сер. Водохозяйственное строительство, теплоэнергетика и геоэкология. – 2010 – № 2. – С. 2–7