

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ МЕТЕОРОЛОГИЧЕСКИХ ДАННЫХ

Н. А. Солтаганов, И.А Ботыгин
Томский политехнический университет
nas38@tpu.ru

Введение

Климат планеты подвержен постоянным изменениям. Анализ наблюдений климата позволяют понять причины его изменений и прогнозировать. Фактические сведения об атмосфере, погоде и климате дают метеорологические наблюдения. Метеорологические наблюдения ведутся на станциях по единым методикам, с использованием однотипных приборов и в определенные часы суток [1-3].

Разработка программного обеспечения

Спектр измеряемых параметров стандартными метеостанциями достаточно широкий. С привязкой к сроку наблюдения измеряются свыше трех десяткой параметров (температура, давление, влажность, ветер, облачность и другие). Указанные параметры могут быть связанными между собой (взаимосвязанными), либо независимыми. Как правило, зависимость между указанными параметрами носит статистический характер. Изменение одного из параметров или нескольких влечет изменение распределения других параметров, которые принимают свои значения с определенными вероятностями.

В метеорологии чаще всего корреляционный анализ включает изучение связей между множеством параметров на одной выборке. То есть вычисления корреляции производятся для каждой парой из множества рассматриваемых параметров. В настоящей работе исследуется взаимосвязь (корреляционная зависимость) между значениями следующих метеорологических параметров: горизонтальная дальность видимости, общее количество облачности, количество облачности нижнего яруса, форма облаков верхнего яруса, форма облаков среднего яруса, форма облаков вертикального развития, слоистые и слоисто-кучевые облака, слоисто-дождевые, разорванно-дождевые облака, высота нижней границы облачности, погода между сроками, погода в срок наблюдения, направление ветра, средняя скорость ветра, максимальная скорость ветра, сумма осадков, температура поверхности почвы, минимальная температура поверхности почвы между сроками, максимальная температура поверхности почвы между сроками, температура воздуха по сухому термометру, минимальная температура воздуха между сроками, максимальная температура воздуха между сроками, парциальное давление водяного пара, относительная влажность воздуха, дефицит насыщения водяного пара, температура точки

росы, атмосферное давление на уровне станции, атмосферное давление на уровне моря, характеристика барической тенденции, величина барической тенденции, температура поверхности почвы, минимальная температура поверхности почвы между сроками, максимальная температура поверхности почвы между сроками, температура воздуха по сухому термометру, минимальная температура воздуха между сроками, максимальная температура воздуха между сроками, парциальное давление водяного пара, относительная влажность воздуха, дефицит насыщения водяного пара, температура точки росы, атмосферное давление на уровне станции, атмосферное давление на уровне моря, характеристика барической тенденции. В работе корреляционная зависимость между указанными параметрами определяется для взаимосвязи двух случайных величин только в виде коэффициента корреляции. Так как корреляционный момент (коэффициент ковариации) имеет большую размерность, по которой трудно оценивать зависимость параметров.

Для вычисления коэффициента корреляции использовалось следующее выражение:

$$r_{xy} = \frac{\sum (x_i - \bar{x}) (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}}$$

Программные эксперименты проводились для данных, полученных со станции с синоптическим индексом станции 29430 (Томск) со сроками наблюдения, указанным в таблице.

Таблица. Описание сроков наблюдения

Год по Гринвичу	2016
Месяц по Гринвичу	1-12 (интервал 1)
День по Гринвичу	1-7 (интервал 1)
Срок по Гринвичу	1-12 (интервал 1)
Год источника (местный)	2016
Месяц источника (местный)	1-12 (интервал 1)
День источника (местный)	7
Срок источника (местный)	1-12 (интервал 1)
Номер срока в сутках по ПДЗВ	1-7 (интервал 1)
Время местное	1(интервал 3)
Номер часового пояса	6
Начало метеорологических суток по ПДЗВ	18

В таблице под термином «интервал №» подразумевается точки отсчета наблюдений (или месяц или день или трехчасовой период).

Структура и содержание метеорологических данных

Файлы с данными, были взяты с сайта ФГБУ "ВНИИГМИ-МЦД" Росгидромета (г. Обнинск Калужской области).

Массивы содержит записи фиксированной длины в коде ASCII с символом конца строки. Каждая запись содержит определенное количество полей с фиксированным положением в записи и фиксированной длиной. Дополнительно, все поля записи отделены друг от друга символом «пробел».

Такое «двойное» форматирование записи позволяет использовать любой из наиболее употребительных способов разделения полей («comma separate» или «fixed length fields») при вводе данных в программные средства хранения (СУБД), электронные табличные процессоры, статистические пакеты и средства графического представления информации [4].

Файлы с данными были переведены в текстовый формат CSV, предназначенный для представления табличных данных.

Используемая среда разработки

В качестве языка программирования использовался Python [5]. Был использован дистрибутив Anaconda с наборами библиотек для научных и инженерных расчетов [6]. Дистрибутив был взят с сайта Кембриджского университета [7, 8], включающий в себя набор библиотек для научных и инженерных расчетов, менеджер пакетов и интерактивную оболочку IPython.

Результаты вычислений коэффициентов попарной корреляции между метеорологическими параметрами приведены на рис. 1.

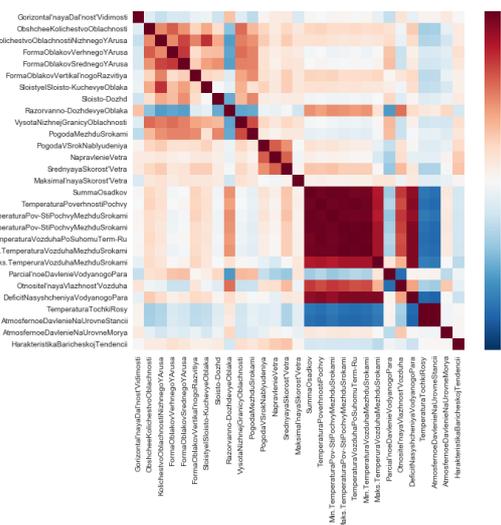


Рис. 1. Результаты тестирования

Расчеты показали, что между многими

параметрами существует сильная взаимосвязь. Так, за указанный срок наблюдений коэффициент корреляции между температурой точки росы и суммой осадков равен -0.7.

Знак «-» свидетельствует о том, что большие значения температуры точки росы соответствуют малым значениям сумме осадков. Это означает о наличии обратной зависимости между этими величинами.

Коэффициент корреляции температурой поверхности почвы и дефицитом насыщения водяного пара равен +0.7. Знак «+» означает, что большие значения температуры поверхности почвы соответствуют большим значениям дефицита насыщения водяного пара. И наоборот, малые значения температуры поверхности почвы соответствуют малым значениям дефицита насыщения водяного пара.

Заключение

Дальнейшие исследования предполагают использование многомерных методов статистического анализа (факторного, кластерного, множественного регрессионного анализа), т.к. только одна корреляционная матрица не дает возможности сделать статистически обоснованный вывод о значениях коэффициентов корреляций между её элементами.

Список использованных источников

1. Хромов С.П., Петросянц М.А. Метеорология и Климатология: учебник для университетов. – М.: МГУ, 2001. – 528 с.
2. Сторм Данлоп. Атлас погоды. Атмосферные явления и прогнозы. – СПб.: Амфора, 2010. – 192 с.
3. Стехновский Д.И., Васильев К.П. Справочник по навигационной гидрометеорологии. – М.: Транспорт, 1976. – 168 с.
4. ВНИИГМИ-МЦД [Электронный ресурс]. URL: <http://meteo.ru/data> (дата обращения: 12.05.17)
5. Python. [Электронный ресурс]. URL: <https://www.python.org/> (дата обращения: 12.05.17)
6. pandas 0.20.2 documentation [Электронный ресурс]. URL: <http://pandas.pydata.org/pandas-docs/stable/index.html> (дата обращения: 12.05.17)
7. Anaconda Scientific Python Distribution. [Электронный ресурс]. URL: <http://www.ch.cam.ac.uk/computing/software/anaconda-scientific-python-distribution> (дата обращения: 12.05.17)
8. Открытый курс OpenDataScience по машинному обучению. [Электронный ресурс]. URL: https://github.com/Yorko/mlcourse_open (дата обращения: 12.05.17)