



Ссылка на статью:

// Ученые записки УлГУ. Сер. Математика и информационные технологии. УлГУ. Электрон. журн. 2021, № 1, с. 137-143.

Поступила: 23.05.2021

Окончательный вариант: 31.05.2021

© УлГУ

УДК 004.4

Программа для чтения и обработки данных атмосферных реанализов

Шугурова М. А. *, Галушкина Д. В.

[*m.a.shugurova@gmail.com](mailto:m.a.shugurova@gmail.com)

ФГБОУ ВО «УлГПУ им. И.Н. Ульянова», Ульяновск, Россия

Атмосферные реанализы являются важным источником данных для диагностики и анализа изменений климата. Результаты реанализов обычно приводятся в формате NetCDF для узлов некоторой широтно-долготной гауссовой сетки с переменным шагом по широте, что осложняет для исследователей непосредственную работу с данными и требует применения специальных программных средств. Целью статьи является описание разработанного программного обеспечения на языке MATLAB, позволяющее по заданным географическим координатам извлекать данные реанализов температуры воздуха для ближайшей точки гауссовой сетки, осуществлять их статистическую обработку и строить графики. Разработанное программное обеспечение может быть использовано для моделирования годовых колебаний температуры воздуха.

Ключевые слова: реанализ, математическое моделирование, среднесуточная температура воздуха, гауссова сетка, NetCDF, MATLAB.

Введение

В современном мире одним из основных инструментов исследования климата и понимания его изменений как в прошлом, так и в будущем являются математические модели. Исследование климата уходит своими корнями в глубокую древность. Первая книга об атмосферных явлениях под названием «Метеорология» была написана Аристотелем (384–322 гг. до н.э.). На протяжении своей истории человечество накапливало знания об изменениях температуры и других метеорологических параметров, как сезонных, так и на протяжении длительного времени. Объемы измерительных данных постоянно увеличивались и со временем стал необходим надежный инструмент, алгоритм для объединения и структурирования этих данных. Этим инструментом стал реанализ.

Реанализ — это данные численных расчетов с некоторой моделью атмосферы, в которых используются все доступные наблюдения в течение анализируемого периода. При этом

моделируемая динамика атмосферы «притягивается» к данным наблюдений [7]. В настоящее время на каждом временном шаге используется примерно 7–9 миллионов наблюдений. На сегодняшний день существует более десятка различных реанализов, данные которых охватывают период с 1948 г. и позднее.

Атмосферные реанализы являются важным источником данных для диагностики изменений климата и понимания механизмов этих изменений. Результаты реанализов принято приводить для узлов некоторой широтно-долготной сетки. Таких сеток существует достаточно много. На рисунке 1 приведен пример гауссовой сетки T62, имеющей 192 узла по долготе и 94 узла по широте, причем шаг узлов по долготе является постоянным, а по широте — переменным.

Без данных атмосферных реанализов, которые невозможно получить без тесной международной кооперации, было бы практически нереально проследить глобальные изменения климата, поскольку имеются большие разрывы и неточности при слиянии данных.

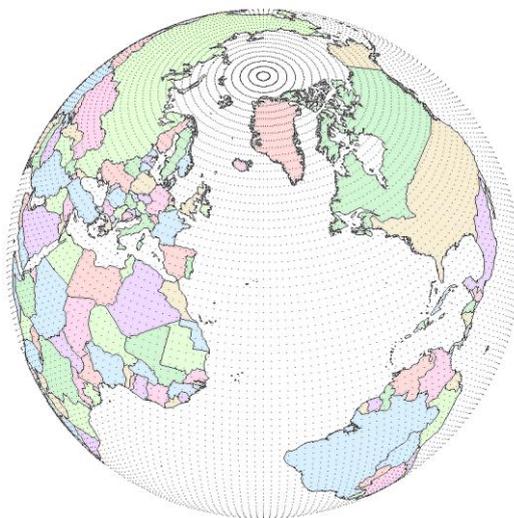


Рис. 1. NCEP T62 Gaussian grid (M. W. Toews, ORCID: 0000-0003-3657-7963)

Использование устойчивых систем интернализации данных позволило:

- создать науку климатологию;
- сформировать временные ряды различных метеорологических параметров;
- достаточно точно установить аномалии;
- усовершенствовать системы прогнозов;
- иметь надежные архивы данных и сделать системы наблюдений более совершенными.

Представление о климате будущего обеспечивается с помощью математического моделирования планетарной циркуляции атмосферы и океана, термического режима и состояния увлажнения. Для этой цели используют климатические модели, основанные на уравнениях геофизической гидродинамики. Их решение возможно только численными методами

[8]. Эта задача требует не только математического и физического обоснования, но и развития технологии компьютерного эксперимента, что делает ее реализацию достаточно сложной и требует использования суперкомпьютеров [6].

1. Описание программы

В данной статье представлен программный продукт, реализованный на языке MATLAB, позволяющий считывать данные реанализов американского Национального центра прогнозов окружающей среды (NCEP), основанного в 2001 году [1]. За основу взяты данные реанализа NCEP-DOE R2, охватывающие временной промежуток с 1979 г. по настоящее время.

1.1. Формат NetCDF

Данные реанализов обычно приводятся в формате NetCDF (Network Common Data Form) — машинезависимом двоичном формате файлов, являющимся стандартом для обмена научными данными. Заголовок формата содержит информацию о содержимом файла. Формат является открытым стандартом. Данный формат в основном используется в метеорологии и климатологии, например, при прогнозировании погоды, изучении изменения климата и геоинформационных системах [2].

```
netcdf filename {
dimensions:
  lat = 3 ;
  lon = 4 ;
  time = UNLIMITED ; // (2 currently)

variables:
  float lat(lat) ;
    lat:long_name = "Latitude" ;
    lat:units = "degrees_north" ;
  float lon(lon) ;
    lon:long_name = "Longitude" ;
    lon:units = "degrees_east" ;
  int time(time) ;
    time:long_name = "Time" ;
    time:units = "days since 1895-01-01" ;
    time:calendar = "gregorian" ;
  float rainfall(time, lat, lon) ;
    rainfall:long_name = "Precipitation" ;
    rainfall:units = "mm yr-1" ;
    rainfall:missing_value = -9999.f ;

// global attributes:
  :title = "Historical Climate Scenarios" ;
  :Conventions = "CF-1.0" ;

data:
  lat = 48.75, 48.25, 47.75;
  lon = -124.25, -123.75, -123.25, -122.75;
  time = 364, 730;
  rainfall =
    761, 1265, 2184, 1812, 1405, 688, 366, 269, 328, 455, 524, 877,
    1019, 714, 865, 697, 927, 926, 1452, 626, 275, 221, 196, 223;
}
```

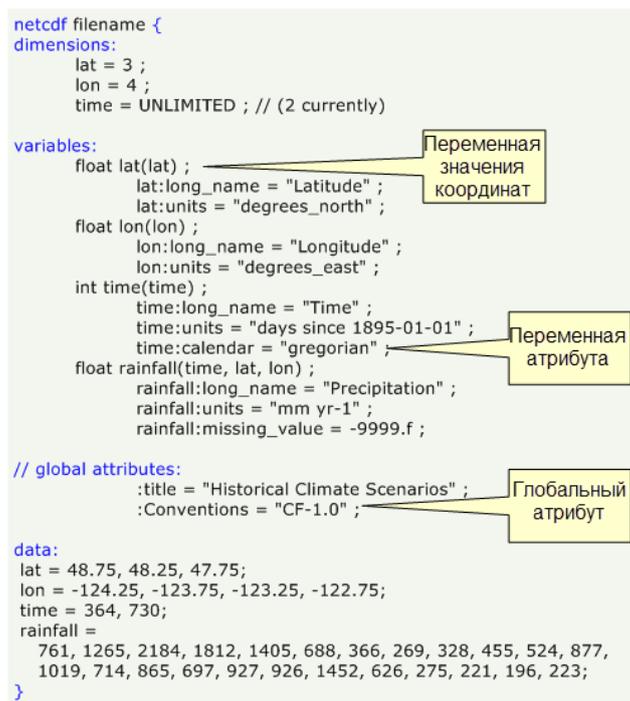


Рис. 2. Пример файла CDL [3]

Данные в файле NetCDF хранятся в форме массивов. Например, температура, меняющаяся со временем, хранится в виде одномерного массива. Температура на площади за ука-

занное время хранится в виде двумерного массива. Трехмерные данные, например, температура на площади, изменяющаяся со временем, или четырехмерные данные, такие как: температура на площади, меняющаяся во времени и с изменением высоты, хранятся в последовательностях двумерных массивов [3].

Файл NetCDF содержит измерения, переменные и атрибуты. Совместно эти компоненты используются для сохранения значения данных, а также отношений между полями данных в наборе данных, состоящем из массивов. На рис. 2 показана структура файла NetCDF с помощью формата CDL — формата ASCII, с помощью которого описывается содержимое файла NetCDF [3].

Файлы NetCDF имеют расширение nc. Известны два вида NetCDF — 32-битный и 64-битный для поддержки файлов размером более 2 гигабайт (введен в netCDF версии 3.6 [4]).

Кроме того, исходный файл NetCDF формата представляет собой двоичный код, с которым нельзя работать напрямую или посредством стандартных программных инструментов (MS Excell, Access и аналогичного программного обеспечения). Но возможности поддержки netCDF-файлов имеют язык MATLAB (начиная с версии 7.01), библиотеки для доступа к NetCDF языков программирования C, C++, Fortran, Perl, Java [4].

Используя возможности, открывающиеся при работе с файлами формата NetCDF, обязательно необходимо помнить, что в них представлена не первичная информация, а это базы данных реанализа — базы, информация в которых частично восстановлена с использованием физико-математических моделей. Это требует разумных подходов и особой осторожности в использовании подобных данных в научных исследованиях [4].

1.2. Вспомогательная утилита

В нашей разработке была использована утилита gauss2lats, адаптированная Томом Холтом, для MATLAB из программы NCAR Fortran, 23/10/2002 года. Саму программу можно найти на электронном ресурсе [5].

Эта функция вычисляет широты узлов гауссовой сетки, исходя из их количества.

Формат вызова утилиты:

$$[xlat,dlat,sinc] = gauss2lats(nlat)$$

где nlat — количество узлов по широте, например, 94 для сетки T62.

На выходе функция дает следующие значения:

- xlat — широты узлов;
- dlat — расстояния между узлами;
- sinc — синусы дополнительных углов.

Хотя в свободном доступе можно найти координаты широтного шага гауссовой сетки T62 в текстовом файле и пользоваться ими, плюс этой функции состоит в том, что она возвращает значения с высокой точностью.

1.3. Описание программы

Программа выполняет поиск ближайшей к заданным пользователем географическим координатам точки гауссовой сетки, считывает необходимые параметры из файлов реанализа (в нашем случае для примера берется среднесуточная температура воздуха, полученная на высоте 2 метров над поверхностью земли, и файлы реанализа NCEP-DOE R2), строит график и рассчитывает основные статистические характеристики температуры, что облегчает процесс построения математических моделей.

Разберем работу нашей программы на примере вычисления среднесуточной температуры воздуха за период 2000 по 2020 год в точке с географическими координатами [54.3282400, 48.3865700], что соответствует месту — Ульяновск, Россия.

Передадим эти координаты в переменную `GeoCoord = [54.3282400, 48.3865700]`.

Далее программа вызывает функцию `getdataset`, которая отвечает за получение температурных данных за указанный промежуток времени.

Формат вызова:

```
[T, Data, GridRowCol, GaussCoord] = getdataset(GeoCoord, folder, prefix, TimeRange);
```

входные данные:

- `GeoCoord = [lat, lon]` — географические координаты в градусах с десятичной дробью, `lat` — широта: (Юж. полюс) $-90 \leq lat \leq +90$ (Сев. полюс), `lon` — долгота: (Гринвич) $0 \leq lon < +360$ (к востоку от Гринвича), отрицательная долгота обрабатывается по формуле: $lon = 360 + lon$;
- `folder` — папка с данными (строка);
- `prefix` — префикс имени файла реанализа (строка);
- `TimeRange = [startyear, endyear]` — период наблюдений (у нас `TimeRange = [2000, 2020]`), `startyear` — начальный год (2000), `endyear` — конечный год (2020);

выходные данные:

- `T = [1, 2, ..., N]` — моменты времени (`N` — количество наблюдений);
- `Data` — значения температуры в градусах Цельсия;
- `GridRowCol = [row, col]` — строка и столбец гауссовой сетки;
- `GaussCoord = [lat, lon]` — географические координаты узла гауссовой сетки;

Функция `getdataset` в свою очередь вызывает функцию `latlon2rowcol`:

```
[row, col, latitudes, longitudes] = latlon2rowcol(lat, lon, nlat),
```

которая возвращает номер строки (`row`) и номер столбца (`col`) гауссовой сетки, ближайшей к указанным пользователем географическим координатам, а также широты (`latitudes`) и долготы (`longitudes`) узлов гауссовой сетки. Данная функция в качестве вспомогательной использует утилиту `gauss2lats`.

В результате работы программы мы получаем, что ближайшая, к заданным нами координатам (`GeoCoord = [54.3282400, 48.3865700]`), точка на гауссовой сетке `T62` это `GridRowCol = [27, 19]`, ей соответствуют координаты `GaussCoord = [54.2846, 48.7500]` точки, находящейся в Чердаклинском районе Ульяновской области. Далее рассчитываются средняя температура за выбранный период `meantemp = 4.7227 °C`, а также минимальная `mintemp`

$= -35.5200$ °C и максимальная температура за этот период $\text{maxtemp} = 34.3500$ °C. Поскольку данные в файлах реанализа представлены в градусах Кельвина, программа выполняет их автоматическое преобразование в градусы Цельсия. После этого строится график среднесуточных значений температуры в точке T62(27,19) за выбранный период времени, приведенный на рисунке 3.

Это неполный пример того, что можно получить при помощи разработанного программного обеспечения. Основной его задачей является удобная работа с данными при построении математических моделей, созданных на основе данных реанализа, хранящихся в формате netCDF и расчета необходимых для этого математических характеристик.

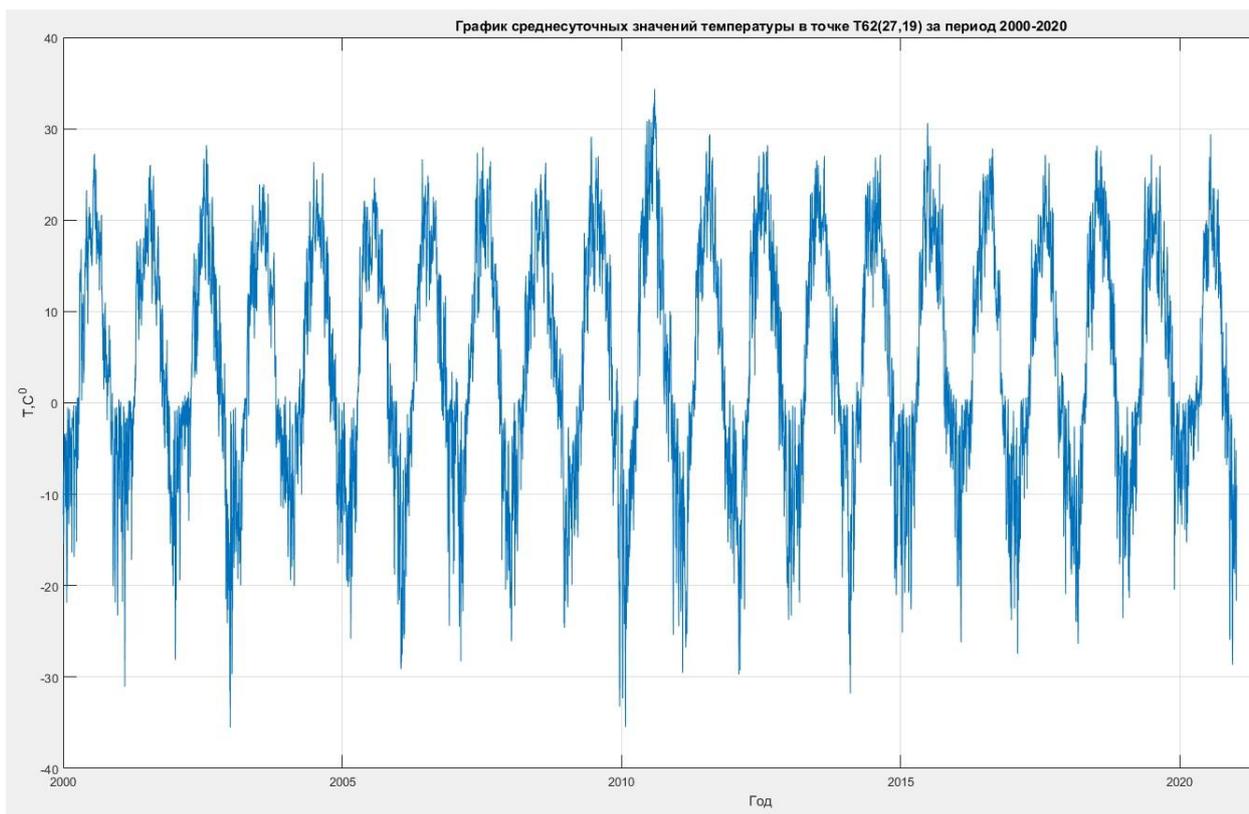


Рис. 3. График среднесуточных значений температуры в точке T62(27,19) за период с 2000 по 2020 год.

Заключение

В статье представлена реализация программного обеспечения для чтения и обработки температурных данных атмосферных реанализов, американского Национального центра прогнозов окружающей среды (NCEP). За основу взяты данные реанализа NCEP-DOE R2. Файлы данных получены с FTP сервера [9]. Разработанное программное обеспечение может быть использовано при построении математических моделей годовых колебаний температуры воздуха.

Список литературы

1. NOAA Earth System Research Laboratories. Режим доступа: <http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.html> (дата обращения 24.05.2021).
2. NetCDF — Википедия. Режим доступа: <https://ru.wikipedia.org/wiki/NetCDF> (дата обращения 24.05.2021).
3. Базовые сведения о хранилище данных netCDF. Режим доступа: <https://desktop.arcgis.com/ru/arcmap/10.3/manage-data/netcdf/fundamentals-of-netcdf-data-storage.htm> (дата обращения 25.05.2021).
4. Тесленок С. А., Тесленок К. С. Программное обеспечение для работы с пространственно-временными данными формата NetCDF // *“Проблемы гидрометеорологического обеспечения хозяйственной деятельности в условиях изменяющегося климата” (материалы Международной научн. конф., 5–8 мая 2015 г.)* / Белорус. гос. ун-т; редкол.: П.С. Лопух (отв. ред.) [и др.]. – Минск, Электрон. журн. 2015. Режим доступа: <https://elib.bsu.by/handle/123456789/118428> (дата обращения 25.05.2021).
5. Tom Holt (2021). Gauss2lats (<https://www.mathworks.com/MATLABcentral/fileexchange/2586-gauss2lats>), MATLAB Central File Exchange. Retrieved May 24, 2021.
6. National Centers for Environmental Prediction. Режим доступа: <https://www.weather.gov/ncep> (дата обращения 24.05.2021).
7. Kalnay E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell M., Saha S., White G., Woollen J., Zhu Y., Leetmaa A., Reynolds R. 1996. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American meteorological Society*, vol. 77(3), pp. 437-471.
8. Кислов А.В. *Климатология: учебник для студ. учреждений высш. проф. образования* / А.В.Кислов. М.: Издательский центр «Академия», 2011. 224 с.
9. NOAA Physical Sciences Laboratory. Режим доступа: ftp://ftp2.psl.noaa.gov/Datasets/ncep.reanalysis2.dailyavgs/gaussian_grid (дата обращения 24.05.2021).