

Utilizando R para manipular dados de projeção climática

Graciela Oliveira¹
Alan Massaru Nakai²

O Laboratório de Modelagem Agroambiental (LMA) da Embrapa Informática Agropecuária utiliza dados de projeção climática para realização de simulações de cenários agrícolas futuros. Grande parte desses dados é disponibilizada no formato NetCDF.

A NetCDF é uma biblioteca livre, que possui funções de manipulação de dados armazenados em matrizes que contêm dimensões, variáveis e metadados. É utilizada em linguagens como C, C++, Fortran, Java, R entre outras. Essa combinação entre linguagem, biblioteca e arquivo possibilita criação, acesso e compartilhamento de dados científicos.

Os arquivos NetCDF possuem informações descritivas sobre os dados que contêm (os chamados metadados). Podem armazenar diferentes tipos de variáveis numéricas e caracteres. Além disso, o formato é eficiente, permitindo ao usuário acessar um pequeno subconjunto de uma grande base de dados, visualizar simultaneamente o mesmo arquivo por mais de um leitor e anexar dados a um arquivo NetCDF, sem ter que copiar sua base de dados ou redefinir sua estrutura.

Antes de serem utilizados, esses dados passam por diversos processamentos, tais como recorte temporal e espacial, reamostragem de grade e extração de dados. O objetivo deste trabalho é demonstrar como tais processamentos podem ser realizados utilizando o R.

O R é um software livre, que possui um ambiente destinado a cálculos matemáticos, estatísticos e confecção de gráficos, possibilitando o desenvolvimento de novos algoritmos em sua plataforma.

¹ Universidade Estadual de Campinas (Unicamp)

² Embrapa Informática Agropecuária

Por ser extensível, o R apresenta diversas bibliotecas que permitem manipular arquivos NetCDF, como “ncdf”, “ncdf4” e “raster”.

Neste trabalho, adotou-se o pacote “raster” para processamento dos dados NetCDF. Este pacote contém funções de criação, leitura e manipulação de dados geográficos espaciais, devidamente convertidos para o formato raster. Um arquivo raster é uma estrutura de dados espaciais que divide uma região em retângulos, chamados ‘pixels’, que podem armazenar valores.

A biblioteca “raster” fornece funções de tratamento de dados que podem ser utilizadas no desenvolvimento de funções mais específicas e complexas.

O trecho de código da Figura 1 ilustra a utilização do R para extração de dados climáticos em série temporal, em um ponto determinado, originalmente apresentados em um arquivo NetCDF.

```
1.library(raster)
2.q=stack('arquivo.nc')
3.xy=matrix(nrow=1,ncol=2)
4.xy[1,1]==-46
5.xy[1,2]==-23
6.extracao=extract(q,xy)
7.plot(extracao[1,],type='l',xlab='Tempo (meses de 12/1859a 11/2005)',
      ylab='Temperatura (°C)')
8.writeRaster(extracao,'pastadestino/nomedoarquivo.nc',type='CDF',overwrite=T, varname='tas',varunit='graus',xname='lon',yname='lat',zname='tempo',zunit='mês')
```

Figura 1. Extração de uma série temporal de um arquivo ncdf.

Inicialmente, a biblioteca “raster” é carregada, como demonstrado na linha 1. Em seguida, na linha 2, o arquivo NetCDF (`arquivo.nc`) é carregado e convertido para o formato raster. Nas linhas 3, 4 e 5, define-se uma matriz com a coordenada do ponto a ser extraído, e a extração é realizada na linha 6. Um gráfico mostrando o resultado é gerado na linha 7 e, finalmente, o resultado é salvo, no destino de interesse, e com o nome definido pelo usuário, como mostra a linha 8.

O resultado dessa extração, apresentado na Figura 2, permite a visualização do comportamento dos dados e pode ser utilizado para análise de tendência e variabilidade de determinada variável com relação ao tempo.

O uso da biblioteca raster permitiu tratar conjuntos de dados que são grandes demais para serem carregados na memória RAM da máquina, uma

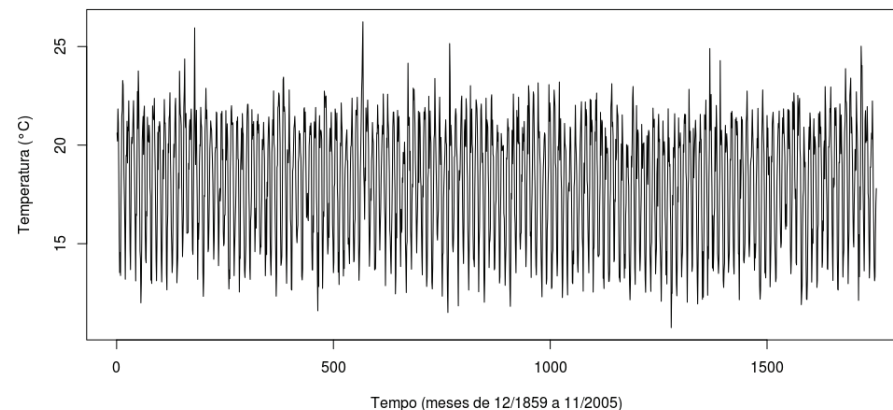


Figura 2. Série histórica extraída de um pixel de um arquivo NetCDF.

vez que cria objetos, a partir dos arquivos, que contêm apenas informações sobre a estrutura dos dados, como número de linhas e colunas, extensão espacial e nome do arquivo, sem tentar ler todos os valores das células simultaneamente em sua memória. Quando manipulados, os dados são processados em pequenos blocos, otimizando o tempo de processamento e uso de memória.

Palavras-chave: R, raster, NetCDF, modelos de projeção climática.