

Análise de Dados Ambientais com R

Jônatan Tatsch

2018-04-16

Contents

Apresentação	5
1 Introdução	7
1.1 Análise de dados meteorológicos	7
1.2 Ciência de dados	7
1.3 Etapas para abordagem de um problema	7
1.4 Programação computacional	8
1.5 R	8
1.6 Por que o R?	8
1.7 Pacotes da comunidade do R	8
1.8 Por que um meteorologista usaria o R?	8
1.9 R não é perfeito!	9
1.10 RStudio	9
1.11 Para saber mais sobre o R	9
2 Instalação do R e RStudio	11
2.1 Sistema Hidroclimático Terrestre	11
2.2 Ciclo Hidrológico	12
2.3 Questão da escala	14
2.4 Propriedades da água	15
2.5 Bacia Hidrográfica	16
3 Interface do Usuário	17
3.1 Pluviometria	17
4 Operações básicas	19
4.1 Formação do solo	19
4.2 Propriedades do solo	19
4.3 Textura e Estrutura	19
4.4 Movimento da água no solo	19
4.5 Infiltração	19
5 Tipos de dados	21
5.1 Evapotranspiração de referência	21
5.2 Evapotranspiração da cultura	21
5.3 Evapotranspiração real	21

Apresentação

Este material é uma composição das notas de aula da disciplina **Análise de Dados Ambientais com R** do curso de Graduação em METEOROLOGIA oferecido no Departamento de Física da Universidade Federal de Santa Maria (UFSM).

O livro é designado para quem não tem experiência em programação, ou qualquer um com interesse em aprender o R para manipular dados ambientais. O objetivo é prover uma material para ensinar os conceitos básicos de programação necessários para o processamento, a visualização e a análise de dados ambientais com o sistema computacional R. Estes procedimentos são potencializados com o uso do software RStudio, uma interface de desenvolvimento integrado (IDE) para o R.

Neste livro o leitor aprenderá a sintaxe básica da linguagem R (R Core Team, 2018), a importação e exportação de dados, a criação de gráficos, funções, a padronização e organização de conjunto de dados; e finalmente, a confecção de relatórios dinâmicos e reproduzíveis.

O material do livro inclui o uso de dados ambientais de diferentes áreas (meteorologia, climatologia, hidrologia, sensoriamento remoto) em exemplos práticos e em exercícios, para estimular a prática da programação.

O texto é intercalado com trechos de códigos que podem ser reproduzidos e os resultados visualizados no computador do leitor.

Após a introdução ao R apresenta-se as capacidades específicas do R para manipulação de dados. Baseado na experiência do autor são empregados os pacotes mais adequados para cada finalidade, como `dplyr` e `tidyr` para o processamento de dados e o `ggplot2` para visualização de dados.

A intenção do livro é que após a leitura, o leitor tenha o conhecimento suficiente para desenvolver códigos que automatizem tarefas repetitivas, assim reduzindo o tempo na etapa de preparação de dados. Esta programação mais efetiva permitirá focar mais na análise de dados e na comunicação dos resultados, seja ela na forma de documentos acadêmicos, ou relatórios técnicos em empresas públicas e privadas.

O texto está em formato html para tirar o melhor proveito de recursos de multimídia, da capacidade de busca de texto e links para websites.

O texto é organizado em 5 capítulos:

- 1 Introdução
- 2 Instalação do R e Rstudio
- 3 Interface do Usuário
- 4 Operações Básicas
- 5 Tipos de dados

Chapter 1

Introdução

Breve intro.

1.1 Análise de dados meteorológicos

Processo pelo qual adquire-se conhecimento, compreensão e percepção dos fenômenos meteorológicos a partir de observações (dados) qualitativas e quantitativas.

1.2 Ciência de dados

1.3 Etapas para abordagem de um problema

1. **Questão científica/problema**
2. **Obtenção de dados:** coleta/medida do(as) estado/condições da atmosfera
 - Instrumentos e sensores
3. **Processamento de dados:** *download* —> limpeza —> formatação —> transformação —> controle de qualidade
 - ferramenta/software
 - conhecimento em programação
4. **Análise de dados**
 - ferramenta/software
 - conhecimento em programação
5. **Solução para o problema**
 - Proposta de um modelo
 - estatístico, empírico, ou fisicamente baseado
 - conhecimento em programação
6. **Apresentação/divulgação/publicação**

1.4 Programação computacional

1.5 R

- R é o termo usado para se referir a linguagem de programação e ao software que interpreta os scripts escritos usando esta linguagem.
- Comunidade fantástica
- Contribuidores (R-core Team)
- milhares de pessoas usam o R diariamente e ajudam outras pessoas
- **Software Livre** (GPL), Código aberto e multiplataforma
- Ambiente para Análise de dados interativa

1.6 Por que o R?

- R não é uma GUI (Interface gráfica do usuário) e isso é bom
 - há uma natural resistência e dificuldade ao uso de códigos e scripts
 - scripts favorecem a **automatização** e **reprodutibilidade**
 - força você a ter um conhecimento mais aprofundado do que está fazendo
- Reprodutibilidade
 - qualquer pessoa (inclusive você mesmo no futuro) pode obter os mesmos resultados do mesmo conjunto de dados
 - R é integrado com outras ferramentas de que permitem atualizar seus resultados, figuras e análises automaticamente
- Relatório dinâmicos e interativos
- Acesso ao estado da arte da ciência de dados (*Big Data*, *Data Mining*, *Machine Learning*)
- Interface com Fortran, C, C++, Python
- Visualização de dados
- R produz gráficos de alta qualidade
- R trabalha com dados de todas formas e tamanhos
- Extensões para Manipulação de dados

1.7 Pacotes da comunidade do R

Evolução do n^o de pacotes disponíveis no CRAN

1.8 Por que um meteorologista usaria o R?

A meteorologia é 4D:

```
meteorologia <- function(x, y, z, t){  
  ...muita coisa para caber em um slide...  
}
```


Logo, requer ferramentas específicas para:

- manipulação de dados espaciais
- análise de séries temporais
- importação e ferramentas de SIG
- leitura de dados em formatos específicos (netcdf, binários, grib2, ...)

1.9 R não é perfeito!

- Muitos códigos em R são escritos para resolver um problema;
 - foco nos resultados e não no processo
 - usuários não são programadores
 - códigos deselegantes, lentos e difíceis de entender
- Como o nosso idioma, há muitas exceções para serem lembradas
- R não é muito rápido e códigos mal escritos serão lentos
- São apenas ~20 anos de evolução
- Há muito o que melhorar

1.10 RStudio

RStudio é um **ambiente de desenvolvimento integrado** livre e de código aberto.

- Multiplataforma (Windows, Linux e Mac)
- ênfase da sintaxe do R, auto-preenchimento de código, indentação inteligente
- execução do R diretamente do editor
- manejo de diretórios e projetos
- histórico de gráficos, zoom, atalhos para exportar imagens
- Integrado com knitr
- Integrado com GitHub para controle de versões

1.11 Para saber mais sobre o R

Documentação oficial - Manuais do R traduzidos

Lista de Livros relacionados ao R

- Livros gratuitos (em inglês)

Fóruns:

- lista Brasileira de discussão do programa R: **R-br**
- stackoverflow

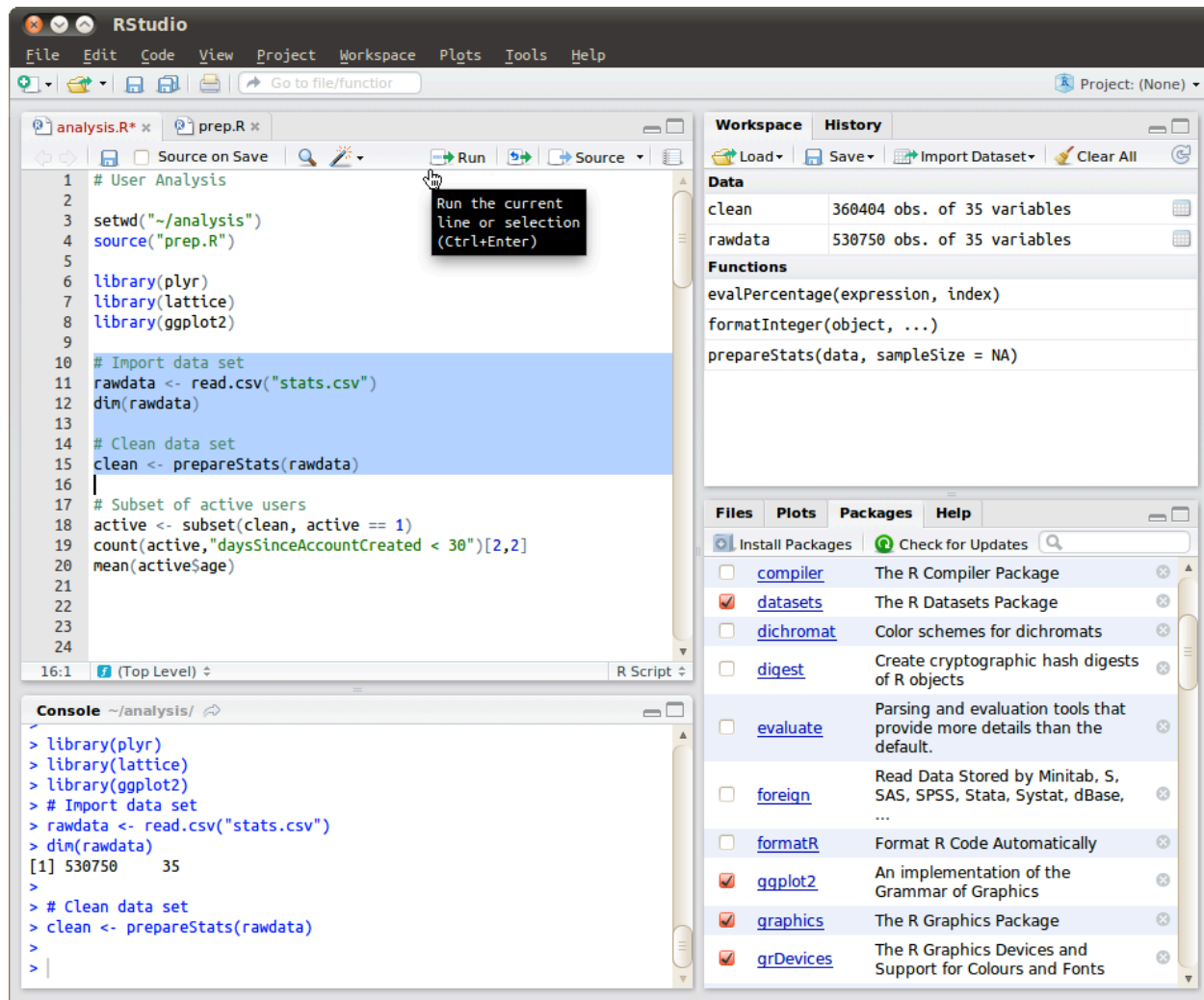


Figure 1.1: RStudio IDE

Chapter 2

Instalação do R e RStudio

Objetivos de aprendizagem:

1. Descrever os principais reservatórios no ciclo hidrológico global e seus tamanhos relativos
2. Descrever os principais fluxos conectando os reservatórios do ciclo hidrológico global
3. Descrever as propriedades da água relevantes para HIDROMETEOROLOGIA
4. Converter massa, volume e fluxos de energia
5. Aplicar a equação do balanço de massa para um volume de controle particular
6. Informar as taxas médias do ciclo hidrológico global e os tempos de residência da água nos diferentes reservatórios
7. Definir bacia hidrográfica e explicar sua relevância para HIDROMETEOROLOGIA
8. Construir uma bacia hidrográfica a partir de dados topográficos para uso em análises hidrológicas

2.1 Sistema Hidroclimático Terrestre

O clima da Terra é entendido como um **sistema** composto de diversas esferas que interagem entre si através de trocas de energia, água e dos ciclos biogeoquímicos. As principais componentes do sistema climático são:

- Atmosfera (ar)
- Hidrosfera (água)
- Criosfera (porções congeladas da Terra)
- Biosfera (organismos vivos)
- Pedosfera (solo)
- Antroposfera (humanos)

A componente humana é um importante agente de mudança ambiental através das mudanças do uso e cobertura da superfície terrestre que afetam diretamente componentes do ciclo hidrológico e dos ciclos biogeoquímicos. Diversos processos físicos, biológicos e químicos dentro do sistema terrestre geram *feedbacks* no sistema climático que acentuam ou mitigam mudanças no clima. Muitos desses *feedbacks* estão associados aos ecossistemas terrestres e as atividades humanas. Um entendimento mais abrangente sobre o clima da Terra requer que todas as componentes do sistema climático (físicas, químicas, biológicas e sócio-econômicas) sejam consideradas.

Table 2.1: Água armazenada no planeta Terra. Adaptado de @Trenberth2007.

Reservatório	Volume (Km ³)	Total de água (%)	Total de água doce (%)
Oceano	1.335.040.000	97,000	NA
Água doce	41.984.700	3,000	NA
Áreas congeladas	26.372.000	1,900	62,80
Água subterrânea	15.300.000	1,100	36,40
Lagos e Rios	178.000	0,013	0,42
Água no solo	122.000	0,009	0,29
Atmosfera	12.700	0,001	0,03



Como leitura complementar sobre as esferas do sistema climático recomenda-se a leitura do Capítulo 1 do Livro *Terrestrial Hydrometeorology* [Shuttle2012].

A esfera de grande interesse é a Hidrosfera que descreve a água contida no planeta Terra armazenada em rios, lagos, oceanos, solo e ar. A Terra armazena cerca de 1377 milhões de km³ de água. Esta quantidade espalhada sobre a área superficial da Terra (510 milhões de km²) seria equivalente a uma lâmina de água de profundidade média de 2700 m.

Os oceanos comportam 97% daquele volume de água (Tabela 2.1). Isto equivale a uma lâmina d'água de profundidade média de 3700 m espalhada sobre a área superficial dos oceanos.

Outros 1,9% estão congelados nas calotas polares, geleiras, solo permanentemente congelado (permafrost). Apenas 1,1 % da água da Terra é líquida e encontrada na superfície continental. Rios, lagos e áreas alagáveis (pântanos, banhados, manguezais) contém 178000 km³ de água.

Mais de 15 milhões de km³ de água estão abaixo da superfície. Aquíferos profundos que incluem a água subterrânea armazenam maior parte desse volume. O solo próximo a superfície (umidade do solo), armazena muito pouca água (122000 km³). A atmosfera tem a menor quantidade de água ~13000 km³ ou ~25 mm de água espalhados sobre a área superficial da Terra.

A pequena fração de água armazenada nos reservatórios de água doce (lagos, rios, umidade do solo), na atmosfera e a biosfera podem subestimar a importância deles uma vez que tendem a ser os mais dinâmicos do ciclo hidrológico, conforme descrito em mais detalhes a seguir.

A água é uma parte importante do sistema climático. Os oceanos armazenam e transportam calor, redistribuindo o aquecimento geograficamente desigual da Terra pelo Sol. O armazenamento de carbono nos oceanos regula a concentração do CO₂ atmosférico. O vapor d'água é o gás de efeito estufa mais importante em termos de concentração e aquecimento radiativo. O vapor d'água condensa para formar nuvens. As nuvens podem gerar precipitação e também afetar o balanço global de radiação refletindo radiação solar, absorvendo e emitindo radiação de onda longa. O calor latente liberado durante a condensação fornece considerável energia para abastecer as tempestades. O ciclo hidrológico entre a atmosfera, o oceano e a superfície terrestre continental regula a quantidade de vapor d'água no ar. As taxas de precipitação e evaporação dependem da temperatura do ar e de outros fatores climáticos de forma que quando o clima muda a quantidade de vapor d'água na atmosfera também muda.

2.2 Ciclo Hidrológico

PAREI AQUI A água se move entre os reservatórios via fluxos. Some of the key hydrologic fluxes include: precipitation (either in liquid or solid form), evaporation and transpiration (together referred to as evapotranspiration), infiltration, recharge, and runoff. Precipitation and evapotranspiration are the key fluxes

between the atmosphere and surface (land and oceans). Precipitation may accumulate when it falls as snow while rainfall is partitioned at the surface into infiltration and surface runoff. Percolation of water through the unsaturated soil zone recharges groundwater aquifers which ultimately feeds surface water bodies via lateral flow and runoff. The atmospheric water is replenished via evaporation from the soil and open water surfaces and transpiration from vegetation. Fluxes and storage are directly linked via mass balance as described in more detail below.

A key aspect of the hydrologic cycle is the fact that it is driven by energy inputs (primarily from the sun; Figure 1.3). At the global scale, the system is essentially closed with respect to water; negligible water is entering or leaving the system. In other words, there is no external forcing in terms of a water flux. Systems with no external forcing will generally eventually come to an equilibrium state. So what makes the hydrologic cycle so dynamic? The solar radiative energy input, which is external to the system, drives the hydrologic cycle. Averaged over the globe, 342 W m^{-2} of solar radiative energy is being continuously input to the system at the top of the atmosphere. This energy input must be dissipated, and this is done, to a large extent, via the hydrologic cycle. Due to this fact, the study of hydrology is not isolated to the study of water storage and movement, but also must often include study of energy storage and movements.

O ciclo hidrológico é o movimento vertical e horizontal da água em suas diferentes fases (vapor, líquida ou sólida) entre a os oceanos, a atmosfera e a superfície terrestre. Clicando na figura abaixo você pode visualizar uma animação do ciclo hidrológico produzida pela NASA.



O ciclo da água é um processo global de circulação fechada controlado pela energia solar, a força de coriolis e a força da gravidade.

A EVAPORAÇÃO é processo físico pelo qual a água líquida nos oceanos ou no continente muda para vapor no ar. Isto ocorre quando o ar não saturado entra em contato com uma superfície úmida. A evaporação

fornece umidade para atmosfera que retorna para a superfície como chuva ou neve. A EVAPORAÇÃO também consome uma grande quantidade de calor, o que ajuda a resfriar a superfície evaporante. O vapor d'água na atmosfera, pode condensar formando nuvens e dependendo das condições atmosféricas essa água pode retornar a superfície como precipitação. Quando o vapor d'água condensa na atmosfera é liberado calor. Esse calor é uma importante fonte de energia para a circulação atmosférica e as tempestades. Os oceanos são a maior fonte água para EVAPORAÇÃO Os solos contêm menos de 1% da água doce não congelada na terra. Entretanto, a água do solo é um importante determinante dos fluxos do balanço de energia e do clima próximo a superfície. Além disso, A descarga de água dos rios para os oceanos, dilui a água dos oceanos, evitando torná-la mais salgada, o que por sua vez influencia o transporte de calor pelos oceanos.

Os rios também carregam nutrientes para o oceano o que influencia a produtividade dos ecossistemas marinhos e o ciclo de carbono entre a atmosfera, terra e oceano.

A água é a fonte de vida na terra. Entre 60-90% do corpo humano e da biomassa vegetal é água, e sua disponibilidade determina onde as plantas crescem e quão bem elas crescem. Dessa forma, o ciclo de carbono está intimamente vinculado ao ciclo hidrológico. O balanço de energia global é fortemente influenciado pela alta capacidade da água armazenar energia térmica e pela grande quantidade de energia requerida para sua mudança de fase. A abundância de água na atmosfera e oceanos exerce um importante papel como regulador do clima, sendo o vapor d'água um dos mais importantes gases-estufa.

Os recursos hídricos foram sempre considerados como ilimitados, entretanto o século vinte testemunhou um grande crescimento no uso da água, assim como um aumento no risco de sua contaminação. Está previsto que a população global irá dobrar de valor nos próximos 50 anos, o que causará uma pressão maior sobre os recursos hídricos.

Diante dos cenários de mudanças climáticas projetados deve ocorrer uma intensificação do ciclo hidrológico evidenciada pelo aumento da variabilidade da precipitação (maior frequência de eventos extremos e estia- gens), o que ressalta ainda mais a importância de um planejamento do uso dos recursos hídricos desde já. Estimativas recentes sugerem que as mudanças climáticas quantificarão por um aumento de 20% na escassez de água global.

2.3 Questão da escala

O ciclo da água no sistema terrestre e sua variabilidade na escala global, regional e local são influenciados por uma variedade de processos, interações mútuas, mecanismos de *feedback* e processos antropogênicos. As escalas em que esses processos interagem espacial e temporalmente variam para cada componente do sistema terrestre (atmosfera, hidrosfera, criosfera e biosfera) de uma forma complexa.

<https://www.youtube.com/embed/8oRjP8yj2Wo>



Source: http://www.nasa.gov/vision/earth/environment/warm_wetworld.html

teste 2

“And here’s to you, Mrs. Robinson, Jesus loves you more than you will know.”

ENFASE COM COR FNS

DONATE

2.4 Propriedades da água

A água não é a molécula mais abundante no planeta, mas é a mais importante para todas formas de vida. A vida surgiu na água e é impossível sobreviver sem ela. Aproximadamente 60% do tecido animal e 90% do tecido das plantas é constituído de água.

Em geral a *água* é um termo usado para se referir a parte da água doce que é renovável anualmente e inclui a água superficial, a água no solo e no sub-solo.

capítulo 1.

2.5 Bacia Hidrográfica

Para compreender melhor os processos hidrológicos é necessário definir um volume de controle mais apropriado do que áreas continentais ou politicamente definidas. A maior parte dos problemas requerem volumes de controle de escala muito inferior a continental.

[menor escala, mais aberto o ciclo]

Qualquer volume pode ser usado para aplicar a equação do balanço hídrico, mas em hidrologia o volume de controle ou a unidade hidrológica fundamental é a BACIA HIDROGRÁFICA[[^]].



BACIA HIDROGRÁFICA (BH): Área de captação natural sobre a superfície terrestre que faz convergir os escoamentos para um ponto chamado exutório. A BACIA HIDROGRÁFICA é constituída de um conjunto de superfícies vertentes e uma rede de drenagem formada por cursos d'água que confluem até resultar em um leito único no exutório.

A BH é definida a partir de uma localização específica, geralmente ao longo de um curso d'água e a área terrestre associada pode ser considerada **captar** a água que escoar por aquele ponto. Então é possível definir inúmeras BH para um rio específico.

O exutório é o ponto que defini a delimitação de uma BH, para onde toda precipitação captada à montante converge por escoamento superficial e sub-superficial.

watershed: pequena BH

basin: grande BH

catchment: termo genérico para BH (qualquer escala)

Chapter 3

Interface do Usuário

```
knitr::include_graphics('images/SatCloud.gif')
```

3.1 Pluviometria

Em construção.

Chapter 4

Operações básicas

Os solos são caracterizados por uma típica sequência de horizontes que constituem o perfil do solo.

4.1 Formação do solo

4.2 Propriedades do solo

4.2.1 Materiais

4.2.2 Hídricas

4.3 Textura e Estrutura

4.4 Movimento da água no solo

4.5 Infiltração

4.5.1 Modelo de frente de umedecimento de Green-Ampt

asd

Chapter 5

Tipos de dados

A EVAPORAÇÃO ocorre quando uma superfície úmida está exposta a ar relativamente mais seco. Quando parcelas de ar movem-se sobre a superfície elas carregam umidade daquela superfície.

A água evaporada da superfície aumenta a quantidade de vapor d'água no ar. Quando o ar está saturado com vapor d'água a evaporação cessa.

A TRANSPIRAÇÃO é a evaporação da água das folhas das plantas quando esta move-se do solo, através das plantas e através das folhas para o ar.

As plantas consomem grandes quantidades de água durante o crescimento. Um lavouro de milho de 4000 m² pode consumir 10000-15000 litros de água (2,5 - 3.75 mm) em um dia. Uma árvore sem restrição hídrica pode transpirar 100-150 litros por dia. Os processos meteorológicos próximo a superfície controlam a EVAPORAÇÃO e TRANSPIRAÇÃO. A TRANSPIRAÇÃO é também regulada pela fisiologia das plantas. Quando as plantas cobrem uma pequena porção do solo a evaporação do solo é o fluxo dominante. A TRANSPIRAÇÃO torna-se mais importante com o aumento da cobertura de área das plantas aumenta. Entretanto é difícil distinguir EVAPORAÇÃO de TRANSPIRAÇÃO e os dois termos são frequentemente combinados em EVAPOTRANSPIRAÇÃO.

EVAPORAÇÃO TRANSPIRAÇÃO EVAPOTRANSPIRAÇÃO

5.1 Evapotranspiração de referência

5.2 Evapotranspiração da cultura

5.3 Evapotranspiração real

Bibliography

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.