rsnodas vignette

Logan Schneider

July 12, 2022

Contents

1   Introduction

This rsnodas package allows users to access, clean, visualize, and analyze data from the following sources:

- Snow Data Assimilation System (SNODAS)

- University of Arizona (UA)

- Parameter-elevation Regressions on Independent Slopes Model (PRISM)

- Historical Climatology Network daily (GHCND) which provides access to Snow Telemetry (SNOTEL) stations

SNODAS, UA, and PRISM provide data in the form of gridded rasters for the contiguous United States. Each source can have different spatial resolutions like 800m, 1km, or 4km and multiple temporal scales of daily, monthly, annual, and 30 year normals. GHCND provides in-situ measurements, spatial point data, with different temporal scales. This package focuses on downloading the daily maps and measurements from SNODAS, PRISM, and GHCND. UA data currently can't be downloaded in this package and must be downloaded manually after creating and signing into an Earthdata account. Please refer to their links for more information about each data product.

The following table lists the data products and spatial resolution of each data source. Note that like PRISM allows users to download data at the 4km resolution for free but charge a fee for their maps at a 800m when it is their daily, monthly, or yealy data. Their 30 year normal maps are free at the 800m.

Data Source	Resolution	Variables available
SNODAS	1 km	Snow water equivalent (SWE), Snow Depth (SnD) Snow Melt runoff (SM), Sublimation from Snow pack (SSP) Sublimation of blowing snow (SBS), Precipitation (Ppt), Snow pack Temperature (SPT)
University of Arizona (UA)	4 km	SWE, SnD
PRISM	4 km 800 m*	PPT, Elevation (Elev), mean dew point (TDmean) Minimum and Maximum vapor pressure deficit (VPmin and VPDmax) Minimum and Maximum Temperature (Tmin and Tmax) total global shortwave solar radiation (SolTol)
Daymet	1 km	PPT, SWE, Shortwave radiation (SR) Tmin, Tmax, Water vapor pressure (VP)

1.1   Installation

rsnodas is available on GitHub and can be installed with devtools:

```
# install.packages("devtools")
library(devtools)
install_github("lschneider93/rsnodas")
```

The outline will be Downloading SNODAS, SNOTEL, and PRISM. After downloading all of those sources of data, the process of creating a raster of predictions and calculating station density will be shown. The blending of SNODAS and the Generalized Additive model can then be explored.

2 Downloading SNODAS, PRISM, and SNOTEL data

2.1 SNODAS

The function `download_snodas` allows us to choose the variables we want to download. Note that the dates inputted in this function need to be in 'YYYY-MM-DD' or 'YYYY-M-D' format. The function `format_date` is included in this package to allow users capabilities of easily creating character vectors with dates in the needed format. `format_date` function can be applied inside of the `download_snodas` for mass downloading. The following example shows how to download snow water equivalent for the masked area of the US into a folder called snodas_data in our working directory.

```
# Download Snodas SWE data for April 1st in 2021 and 2022
snodas_2021 <- download_snodas(dates = format_dates(day = 1,
                                                     month = 4,
                                                     year = 2021),
               masked = TRUE,
               overwrite = TRUE,
               remove_zip = TRUE,
               data_saved = c('swe'),
               out_dir = paste0(getwd(), "/snodas_data")) #
```

After downloading SNODAS, you can crop it to the area or state of interest. These next portion of code gets a shape file of Utah from the `maps` package, crops the SNODAS map to the state of Utah, and creates a visual by using `ggplot2`.

```
# get the pipe function from magrittr
"%>%" <- magrittr::"%>%"

# Get an shape of utah from the maps package
ut_map <- maps::map("state", plot = FALSE, fill = TRUE) %>%
  sf::st_as_sf() %>%
  dplyr::filter(ID == "utah") %>%
  sf::st_transform(crs = sf::st_crs(snodas_2021[[1]]))

# Crop the maps to just the state of utah and name the values to be "Value"
snodas_ut_2021 <- sf::st_crop(snodas_2021[[1]], ut_map)
names(snodas_ut_2021) <- "Value"

# Plot SNODAS April 1st 2015 SWE map of Utah with blue outline
ggplot() +
  stars::geom_stars(data = snodas_ut_2021) +
  geom_sf(data = ut_map, fill = "NA", size = 1, color = "blue") +
  ggtitle("2021 SNODAS SWE predictions") +
  scale_fill_viridis_c(option = "A") +
  theme(plot.title = element_text(hjust = 0.5, size = 24),
        text = element_text(size = 22))
```

2.2 Download PRISM climate data

The PRISM climate group has been provided climate maps since 1895 for the Precipitation (PPT), Minimum and maximum temperature (Tmin and Tmax), Minimum and maximum vapor pressure deficit (vpdmin and vpdmax), mean dew point (tdmean), Elevation (Elev), and total global shortwave solar radiation.

Most of this data from PRISM can be downloaded by utilizing the R package, `prism`, and more examples and information is available at [this link](). The difference is that they have three separate functions to download daily, monthly, and yearly data and store them in different file paths to keep them organized.

In comparison, `download_prism` was created because it is one function that can download daily, monthly, and yearly PRISM climate data and stores all the data files in one directory. This will be optimal for decreasing time and effort later to create a model for prediction. The alternative using `prism` and moving all the data files into one directory.

The follow example will download monthly precipitation from December 2021 until Feb 2022 and store in the folder called 'prism_data' in your working directory.

```r
## NOT RUN, This was ran earlier
# Example of downloading 4km monthly precipitation maps from Dec. 2013 - Feb. 2014
download_prism(sp_res = "4km", data = c("ppt"), t_res = "monthly",
               start_date = as.Date("2021-02-01"), end_date = as.Date("2021-03-15"),
               out_dir = paste0(getwd(), "/prism"))
```

Note that all PRISM files need to be in the same location in order to use the function `gam_to_df` in a later section.

2.3   Download SNOTEL

The `data-raw` folder contains an script titled `DATASET` that shows how to download the all SNOTEL data by using `download_all_ghcnd_stations`. Note that there is more than 30 gigabytes of data and this will take time to download. The functions `download_all_ghcnd_stations`, `get_station_data`, and `get_state_data` come from a currently private `snowload2` package (authors Jadon Wagstaff and Brennan Bean) and is replicated here with the permission from the authors. `get_station_data` and `get_state_data` are for the purpose of sifting through all the data to get the stations in a specific state. These functions were used to create the dataset `april_1_snotel_data`.

```r
# Code to download all stations in your working directory. A file path could
#  have been used instead of ".".
### This code downloads 30 GB of data and will take time to download
download_all_ghcnd_stations(directory = ".")
```

```r
# april 1st data for SNOTEL stations in Utah and subset to just April 1st, 2014.
snotel_ut <- rsnodas::april_1_snotel_data
snotel_ut_2021 <- snotel_ut[snotel_ut$DATE == "2021-04-01", ]
```

3   GAM to Data frame to Raster

This section will use the point data provided by SNOTEL sites and climate variables from PRISM gridded products to create estimates. The process of creating gridded raster estimates will be demonstrated in a three-step process. The first step is to create a data frame with all the Longitude and Latitude with the corresponding PRISM climate information. This is accomplished by using the `gam_to_df` function. This function will only look in one directory and it is vital that all PRISM data files are in one directory.

This example shows creating a data frame with the Long/Lat of Utah and the variables of monthly precipitation data from March 2022, elevation, and slope.

```
# creation of a data frame with all the PRISM variables of precipitation and elevation
gam_2021 <- gam_to_df(model_data = snotel_ut_2021,
                      raster_template = snodas_ut_2021,
                      path_to_prism = paste0(getwd(), "/prism"),
                      model_x <- c("ppt_2021_03", "elevation", "slope"),
                      model_y <- c("VALUE"),
                      coords = c("LONGITUDE", "LATITUDE"))
head(gam_2021, 5)
```

The second step is to build a model with the SNOTEL station. This could be any model that can have the `predict` function from the `stats` package applied afterwards. A commmon model used for spatial data is a Generalized Additive Model (GAM). This is because GAMs account for non-linear effects and can use splines on the sphere (SOS) which account for the spherical shape of the earth.

The following code creates a GAM with the SNOTEL station information using elevation, slope, and March 2022 PRISM monthly, precipitation

```
# This allows users to explore using different models with the same climate variables.
model <- mgcv::gam(data = snotel_ut_2021,
                   VALUE ~ s(LONGITUDE, LATITUDE, bs = "sos", k = 25) +
                   s(ppt_2021_03)  + s(slope) +
                   s(elevation),
                 method = "REML")
```

The final step predict the output with the variables throughout the area of interest and create a gridded output of predictions. This is accomplished by using the `df_to_raster` that takes the previously created model and data frame to create a raster.

Below is an example of creating a raster predicting SWE using the previous model using elevation, slope and precipitation from March and the data frame with all the points in the area of Utah.

```
# After creating a model, we can make predictions of SWwith the information available.
gam_rast <- df_to_raster(model = model,
                         data_frame = gam_2021,
                         raster_template = snodas_ut_2021)

ggplot() +
  stars::geom_stars(data = gam_rast) +
  geom_sf(data = ut_map, fill = "NA", size = 1, color = "blue") +
  ggtitle("2021 GAM SWE predictions") +
  scale_fill_viridis_c(option = "A") +
  theme(plot.title = element_text(hjust = 0.5, size = 24),
        text = element_text(size = 22))
```

3.1   Station Density

We have created a map of SWE estimates throughout Utah using point data and climate information. There are already products, like SNODAS, that estimate SWE throughout the US or global information. The hope is using local information (SNOTEL) to improve the global information (SNODAS). By combining these maps of estimate should reduce the variance of the predictions. In order to blend these maps together, we need to blend based on some criteria. This blending allows our predictions to not rely solely on one prediction.

We are going to blend and weight each map by observation or station density. This uses the `points_to_density_stars` function. The `sigma` argument will extend the distance and is in meters while the gridded product is in kilometers. Changing the `max_weight` argument gives more weight to the Generalized Additive model or the model from the in-situ measurements. The max weight needs to be within the range of 0-1. This weight is a percentage and don't want negative percentages or percentages over 100.

This creates a density map of all the SNOTEL stations in 2022.

```
dens_2021 <- points_to_density_stars(sp_points = snotel_ut_2021,
                                     coords = c("LONGITUDE", "LATITUDE"),
                                     raster_template = snodas_ut_2021,
                                     sigma = 15000,
                                     max_weight = 1,
                                     flat_crs = "+proj=utm + zone=12 + datum=WGS84")

ggplot() +
  stars::geom_stars(data = dens_2021) +
  geom_sf(data = ut_map, fill = "NA", size = 1, color = "blue") +
  ggtitle("2014_Station_Density") +
  scale_fill_viridis_c(option = "B") +
  theme(plot.title = element_text(hjust = 0.5, size = 24),
        text = element_text(size = 22))
```

3.2   Ensemble SNODAS and GAM predictions

Lastly, the `blend_raster` function with blend the rasters together with the weights associated. We are going to use all the previous maps to create a final prediction or estimate of SWE. This process can be applied to multiple types of spatial problems.

This is blending the SNODAS and GAM model based on the SNOTEL station density.

```
comb_map <- blend_raster(raster_sate = snodas_ut_2021,
                         raster_land = gam_rast,
                         weights = dens_2021)

ggplot() +
  stars::geom_stars(data = comb_map) +
  geom_sf(data = ut_map, fill = "NA", size = 1, color = "blue") +
  ggtitle("2022 Ensemble Map") +
  scale_fill_viridis_c(option = "B") +
  theme(plot.title = element_text(hjust = 0.5, size = 24),
        text = element_text(size = 22))
```

4   Conclusion

This package allows users to download, store, and access multiple types of information available from SNODAS, PRISM, and SNOTEL. These data sources can explore multiple types of modeling techniques for spatial data with the potential to improve current gridded products.