rsnodas vignette

Logan Schneider

May 16, 2023

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. A summary of the variables and resolution assoicated with each data source is provided in the table below. 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 yeoly data. Their 30 year normal (i.e. average) maps are free at the 800m resolution and are currently from 1991-2020. These 30 year maps change from year to year and are free at the 800m resolution.

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

*maps available at 800 meter resolution for a fee.

This vignette uses functions from `tidyverse`, `ggplot2`, `sf`, and the `stars` package and will be loaded now.

```
library(tidyverse)
library(sf)
library(stars)
'%>%' <- magrittr::'%>%'
```

1.1 Installation

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

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

The vignette will first proceed with a demonstration of how to download SNODAS, SNOTEL, and PRISM data. After downloading all of those sources of data, the vignette will show the process of creating a raster of predictions and calculating station density. The blending of SNODAS and a 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` allows users to easily create 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
snodas_2021 <- download_snodas(dates = format_dates(day = 1,
                                                     month = 4,
                                                     year = 2021),
                masked = TRUE,
                overwrite = TRUE,
                remove_zip = TRUE,
                data_saved = "swe",
                out_dir = paste0(getwd(), "/snodas_data"),
                GTiff = FALSE) #
```

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 "SWE Value"
snodas_ut_2021 <- sf::st_crop(snodas_2021[[1]], ut_map)
names(snodas_ut_2021) <- "SWE\nValue"
```

```
# Plot SNODAS April 1st 2021 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", limits = c(0, 850), na.value = "white") +
  labs(x = "Longitude", y = "Latitude") +
  theme(plot.title = element_text(hjust = 0.5, size = 16),
        axis.text.x = element_text(angle = 45, hjust = 1),
        text = element_text(size = 13),
        legend.position = "bottom")
```

2021 SNODAS SWE predictions
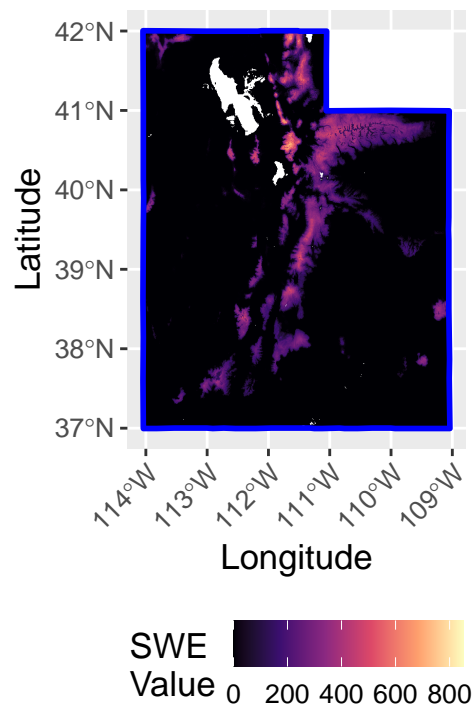
2.2 Download PRISM climate data

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. This package has three separate functions to download daily, monthly, and yearly data and store them in different file paths to keep them organized.

In comparison, we've created the `download_prism` function to seamlessly 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 alternatives are to manually or by use `prism` functions to download PRISM data and move all PRISM files into one directory. Currently, `download_prism`

is limited to only download daily, monthly, or yearly data and can't be used to download PRISM's 30 year normal maps.

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

```r
# Example of downloading 4km monthly precipitation maps from Jan. 2021 - March 2021
download_prism(sp_res = "4km", data = c("ppt"), t_res = "monthly",
               start_date = as.Date("2021-01-01"), end_date = as.Date("2021-03-15"),
               out_dir = paste0(getwd(), "/prism"))
```

```
## [1] "Downloading month 1 of 3"
## [1] "Downloading month 2 of 3"
## [1] "Downloading month 3 of 3"
```

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` which is included as a dataset internal to the `rsnodas` package

```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 hour to download
# download_all_ghcnd_stations(directory = ".")
```

```r
# april 1st data for SNOTEL stations in Utah and subset to just April 1st, 2021.
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 variable maps from PRISM 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 the Longitude, Latitude, and extract PRISM climate variable information for every grid cell in the area of interest. This is accomplished by using the `gam_to_df` function. This function only looks 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 Longitude, Latitude, and monthly precipitation values from March 2021 in Utah.

```r
# creation of a data frame with all the PRISM variables of precipitation
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"),
                      model_y <- c("VALUE"),
                      coords = c("LONGITUDE", "LATITUDE"))
head(gam_2021, 5)
```

```
##    LONGITUDE LATITUDE ppt_2021_03
## 1 -114.0458 42.00417     18.2794
## 2 -114.0375 42.00417     18.5215
## 3 -114.0292 42.00417     18.5215
## 4 -114.0208 42.00417     18.5215
## 5 -114.0125 42.00417     18.5215
```

The second step is to build a model with the SNOTEL stations data. Any model that can use the `predict` function will work here. 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. GAMs were calculated by using the function `gam` from the package `mgcv` and more information is available at this link.

The following code creates a GAM with the SNOTEL station information using March 2021 PRISM monthly precipitation.

```r
# This allows users to explore using different models with the monthly precipitation 2021.
model <- mgcv::gam(data = snotel_ut_2021,
                   VALUE ~ s(LONGITUDE, LATITUDE, bs = "sos", k = 25) + s(ppt_2021_03),
                   method = "REML")

# This is another example of a model that uses elevation, slope and precipitation.
# NOTE: In order to run this You MUST download elevation and store it in the same folder.
# model <- mgcv::gam(data = snotel_ut_2021,
#                    VALUE ~ s(LONGITUDE, LATITUDE, bs = "sos", k = 25) + s(ppt_2021_02) +
#                        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 snow water equivalent (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.

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

# Crop the map to just the state of utah and name the values to be "SWE Value"
gam_rast <- sf::st_crop(gam_rast, ut_map)
names(gam_rast) <- "SWE\nValue"
```

```
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", limits = c(0, 850), na.value = "white") +
  theme_void() +
  theme(legend.position = "bottom")
```


2021 GAM SWE predictions
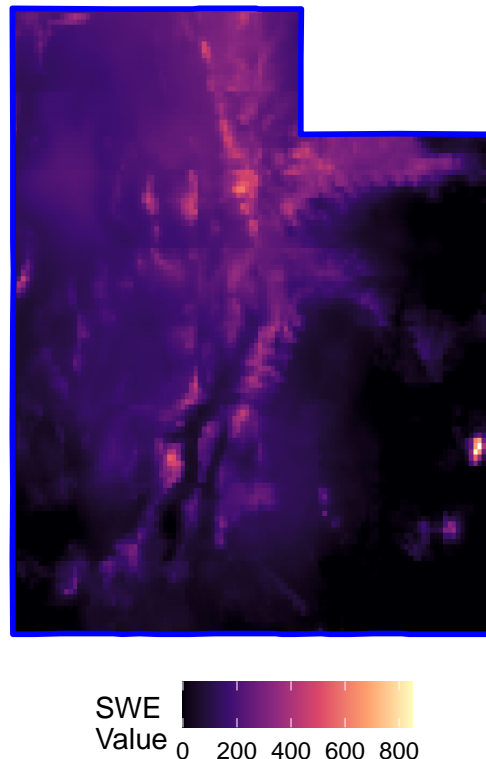
3.1 Station Density

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. SNODAS currently provides the best SWE estimates but is known to struggle predicting in mountainous regions. This map of SWE estimates created using SNOTEL stations has the potential to provide better information in mountain regions. The final map generated by blending SNODAS and the SNOTEL map would not rely solely on one prediction but on SNOTEL information and SNODAS. The goal of blending these maps is to use local information from SNOTEL sites to improve the global predictions that SNODAS provides. The combining of these maps should reduce the variance of the error of the predictions. In order to blend these maps together, we need to blend based on some criteria. This vignette will blend these two maps based on station density.

We are going to blend and weight each map by observation or station density. This uses the `points_to_density_stars` function to produce a raster of weights. The `sigma` argument will influence the kernel density calculation to include a larger or smaller distance. 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 won't allow 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")

# Crop the map to just the state of utah and name the values to be "SWE Value"
dens_2021 <- sf::st_crop(dens_2021, ut_map)
names(dens_2021) <- "Weights"

ggplot() +
  stars::geom_stars(data = dens_2021) +
  geom_sf(data = ut_map, fill = "NA", size = 1, color = "blue") +
  ggtitle("2021 Station Density") +
  scale_fill_viridis_c(option = "D", na.value = "white") +
  theme_void() +
  theme(legend.position = "bottom")
```

2021 Station Density
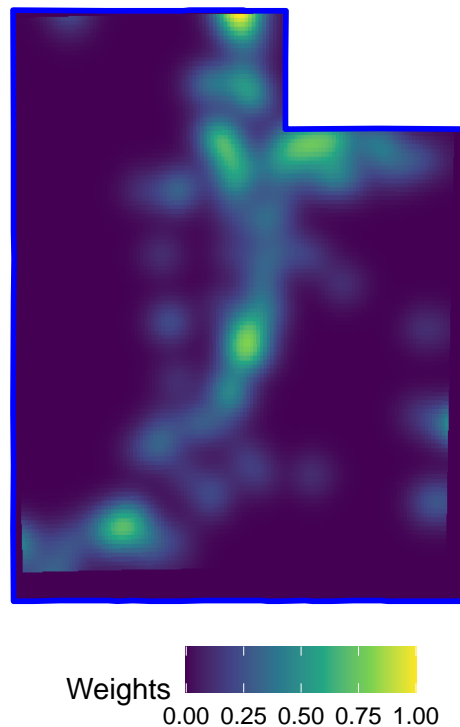
It is important to note that the station density follows the Wasatch Mountains. The data collected from SNOTEL stations are important to predicting how much water is contained in the snowpacks for flood forecasting and water management. Therefore, it is important that we understand how much water is in the mountain snowpacks.

3.2 Ensemble SNODAS and GAM predictions

Lastly, the `blend_raster` function will blend the rasters together with the weights associated. This function could be used to apply weight to any two raster and can be applied to multiple types of spatial problems. It isn't necessary to use the GAM for the land-based predictions if that doesn't apply to your problem. Currently, this function only allows the blending of two maps together. We are going to use all the previous maps to create a final prediction or estimate of SWE.

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)

# Crop the map to just the state of utah and name the values to be "SWE Value"
comb_map <- sf::st_crop(comb_map, ut_map)
names(comb_map) <- "SWE\nValue"


ggplot() +
  stars::geom_stars(data = comb_map) +
  geom_sf(data = ut_map, fill = "NA", size = 1, color = "blue") +
  ggtitle("2021 Ensemble Map") +
  scale_fill_viridis_c(option = "B", limits = c(0, 850), na.value = "white") +
  labs(x = "Longitude", y = "Latitude") +
  theme_void() +
  theme(legend.position = "none")
```

2021 Ensemble Map
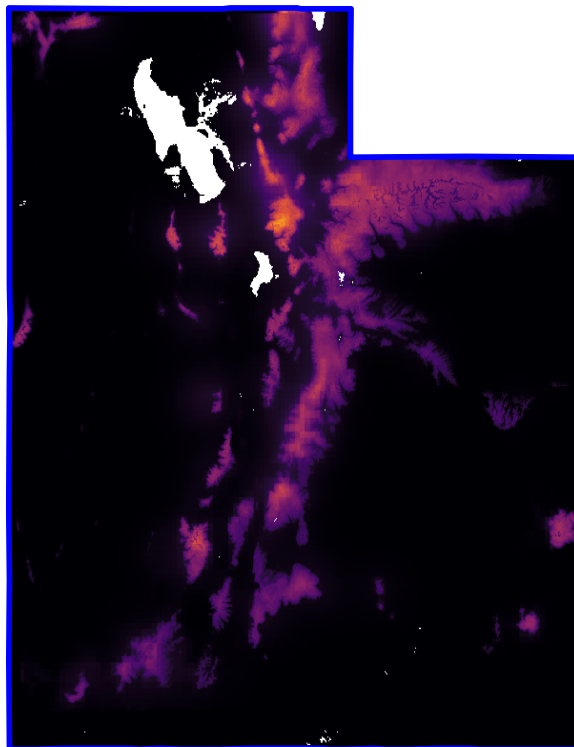
4   Conclusion

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 objective to provide local improvements to currently available national gridded climate and snow products.