Smoothing

Marlin Lee

2023-06-15

Introduction

Overview of Smoothing

In this vignette, we will explore the smoothing techniques supported in the DSIWastewater package. Smoothing is a widely used technique in data analysis and visualization that helps reduce noise, reveal underlying trends, and highlight important patterns in the data. DSIWastewater provides several powerful functions to perform smoothing operations, and this vignette will guide you through their usage and showcase their capabilities.

Data

We will be using the long range Madison Covid data in the vignettes with the occasional use of other data to show a more complete picture of the different smoothing options.This data set has a clear trend but a ton of noise.

```
library(DSIWastewater)
library(dplyr)
library(ggplot2)

data("Example_data", package = "DSIWastewater")

smoothing_df <- Example_data%>%
  select(site, date, N1, N2)%>%
  filter(N1 != 0, N2 != 0)%>%
  mutate(N1 = log(N1), N2 = log(N2))

base_plot <- smoothing_df%>%
  ggplot(aes(x = date))+
  geom_point(aes(y = N1, color = "N1"))+
  facet_wrap(~site)

base_plot
```

Smoothing Techniques

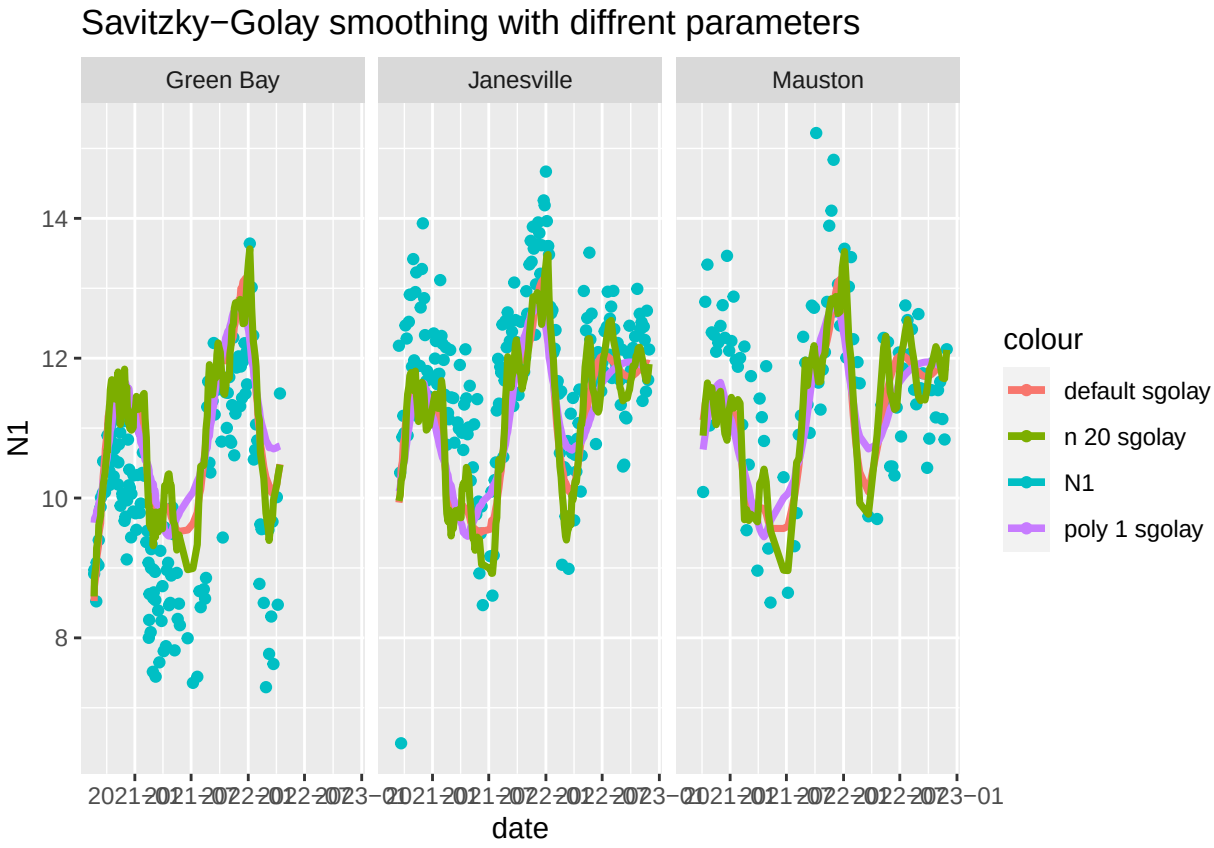
Each smoothing function shares the same framework. The first three arguments are the dataframe, the name of the column to be smoothed, and the name of the column the smoothing is placed. The last argument "Filter" control if some data points should be removed before the filtering happening. Other arguments are parameters for how the smoothing should be done.

Savitzky-Golay Filter

The Savitzky-Golay filter is a widely used smoothing technique that performs local polynomial regression on the data. It provides excellent noise reduction while preserving important features such as peaks and valleys. The DSIWastewater package includes the sgolaySmoothMod() function, which allows you to apply this filter to your data. sgolaySmoothMod() has 2 parameters. "poly" controls the degree of the polynomial fit on the local data whereas "n" control the number of points considered "local". by default its set to "guess" which uses the density of the data to make guesses about the ideal n.

```
# Example code demonstrating sgolaySmoothMod() function
sgolay_smooth <- DSIWastewater:::sgolaySmoothMod(smoothing_df, "N1", "N1_sgolay", poly=5,n="guess", Fil

sgolay_plot <- base_plot +
  geom_line(aes(y = N1_sgolay, color = "default sgolay"), data = sgolay_smooth, linewidth = 1.25)

param_sgolay_smooth <- DSIWastewater:::sgolaySmoothMod(sgolay_smooth, "N1", "N1_sgolay_2", poly=2, n =
```

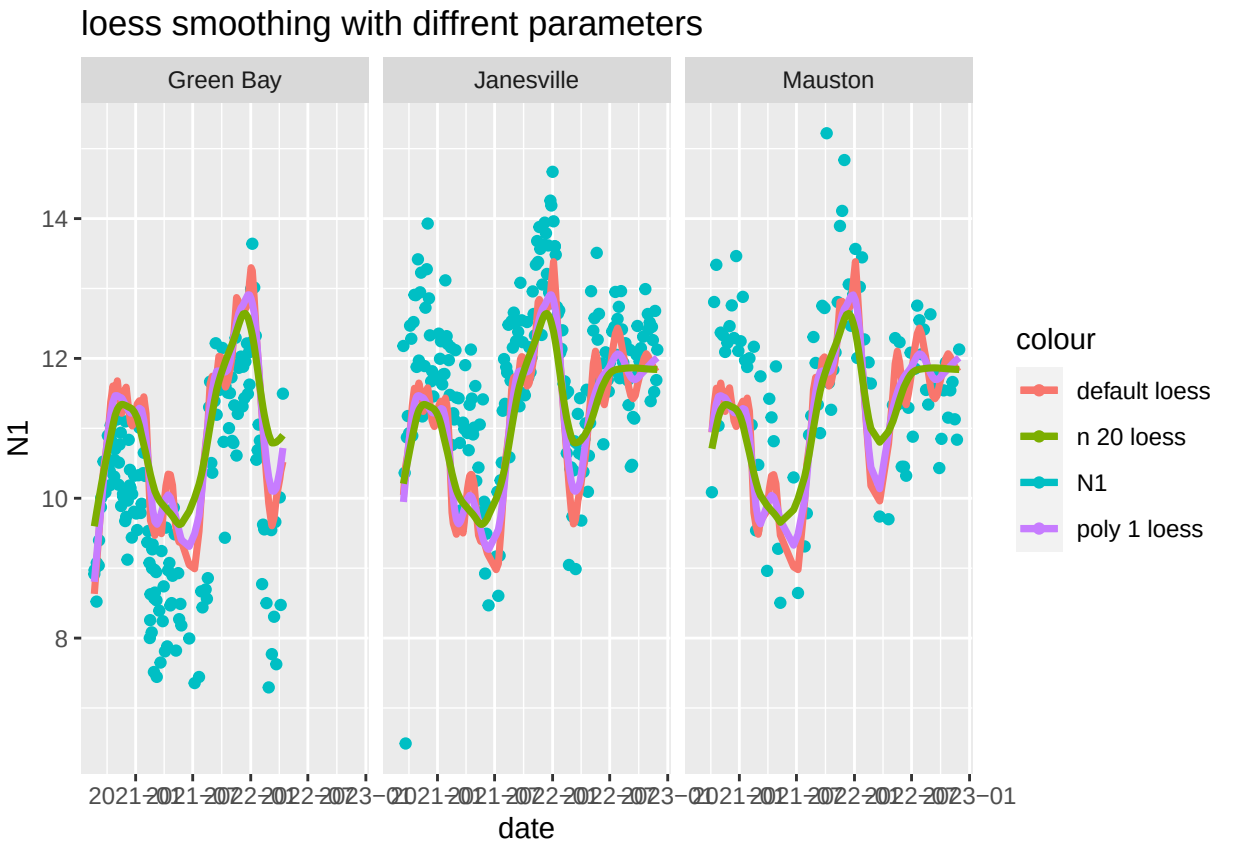
```
alt_plot <- sgolay_plot +
  geom_line(aes(y = N1_sgolay_2, color = "poly 1 sgolay"), data = param_sgolay_smooth, linewidth = 1.25)


param_sgolay_smooth_2 <- DSIWastewater:::sgolaySmoothMod(sgolay_smooth, "N1", "N1_sgolay_3", poly = 5, n


show_plot <- alt_plot +
  geom_line(aes(y = N1_sgolay_3, color = "n 20 sgolay"), data = param_sgolay_smooth_2, linewidth = 1.25)

show_plot+
  ggtitle("Savitzky-Golay smoothing with diffrent parameters")
```



Loess Smoothing

Loess (locally estimated scatterplot smoothing) is a non-parametric regression technique that fits a smooth curve to the data by locally weighted polynomial regression. It is particularly useful when dealing with complex or nonlinear relationships. The DSIWastewater package provides the loessSmoothMod() function to perform loess smoothing on your data. The only parameter is "span" which controls what portion of the data is looked at. This is set to "guess" by default where the package picks what it thinks is optimal based on data density.

```r
# Example code demonstrating loessSmoothMod() function
loess_smooth <- loessSmoothMod(sgolay_smooth, "N1", "N1_loess", span="guess", Filter = NULL)

all_smooth_plot <- sgolay_plot + geom_line(aes(y = N1_loess, color = "default loess"), data = loess_smoo

loess_plot <- base_plot + geom_line(aes(y = N1_loess, color = "default loess"), data = loess_smooth, si

param_loess_smooth <- loessSmoothMod(loess_smooth, "N1", "N1_loess_2", span = .1, Filter = NULL)

alt_plot <- loess_plot + geom_line(aes(y = N1_loess_2, color = "poly 1 loess"), data = param_loess_smoo


param_loess_smooth_2 <- loessSmoothMod(loess_smooth, "N1", "N1_loess_3", span= .2, Filter = NULL)


show_plot <- alt_plot + geom_line(aes(y = N1_loess_3, color = "n 20 loess"), data = param_loess_smooth_

show_plot+
  ggtitle("loess smoothing with diffrent parameters")
```



Exponential Smoothing

Exponential smoothing is a technique used to smooth time series data by assigning exponentially decreasing weights to observations. It is particularly useful for capturing trends and seasonality in time series data.It has two parameters alpha and beta which control the change of the weights.

Everything

Finaly this shows each of the methods with the default parameters on the same dataset. This shows that exponential smoothing has a systemic lag and loess by default is a little less stable.

```
all_smooth_plot
```